

基于跨尺度特征增强与多层注意力机制的火灾检测方法

顾成杰¹, 高紫莲², 朱东郡¹, 张俊军¹, 郑亚兵¹

(1. 安徽理工大学公共安全与应急管理学院, 合肥 231100; 2. 安徽理工大学计算机科学与工程学院, 淮南 232000)

摘要: 针对视频图像火灾检测在低光照条件下背景与烟雾难以区分、火焰与烟雾形状多变以及复杂背景干扰等问题, 给出了一种基于跨尺度特征增强与多层注意力机制的火灾检测方法, 以提升火灾检测的准确性和鲁棒性. 首先给出了一种跨尺度特征增强模块(CSFE), 使模型更加聚焦于关键特征提取, 提高特征的判别性. 其次, 给出一种下采样模块(CG-Adown), 通过扩张卷积和全局上下文引导, 增强模型对全局上下文信息的捕捉能力, 并利用残差连接, 防止深层网络中的梯度消失和信息丢失. 此外, 设计了一种多层注意力模块(MLA), 自适应地调整通道和空间维度的权重, 提升模型在处理复杂视觉任务时的特征表达能力. 最后, 在 EIoU 损失函数的基础上引入缩放因子, 来提升对模型的优化效果. 实验结果表明, 相较于基准模型, 所给出的火灾检测模型在 FAS-CVP 数据集以及 FSD-CVP 数据集在检测精度上分别提升了 6.5%、1.3%, 参数量分别下降了 2.3%、1.7%.

关键词: 火灾检测; 特征增强; 注意力机制; 全局上下文

中图分类号: TP751.1

文献标志码: A

文章编号: 1006-8740(2026)01-0095-14

Fire Detection Method Based on Cross-Scale Feature Enhancement and Multi-Layer Attention Mechanism

Gu Chengjie¹, Gao Zilian², Zhu Dongjun¹, Zhang Junjun¹, Zheng Yabing¹

(1. School of Public Safety and Emergency Management, Anhui University of Science and Technology, Hefei 231100, China; 2. School of Computer Science and Engineering, Anhui University of Science and Technology, Huainan 232000, China)

Abstract: Aiming at the problems in video image fire detection under low light conditions, such as the difficulty in distinguishing background from smoke, the variable shape of flame and smoke, and complex background interference, a fire detection method based on cross-scale feature enhancement and multi-layer attention mechanism was proposed to improve the accuracy and robustness of fire detection. Firstly, a cross-scale feature enhancement module (CSFE) was proposed to make the model focus more on the key feature extraction and improve the discrimination of features. Secondly, a down-sampling module (CG-Adown) was proposed to enhance the model's ability to capture global context information through dilated convolution and global context guidance, and the residual connection was used to prevent gradient disappearance and information loss in the deep network. In addition, a multi-layer attention module (MLA) was designed to adaptively adjust the weights of channels and spatial dimensions to improve the feature represen-

收稿日期: 2025-05-26.

基金项目: 国家重点研发计划资助项目(国家长三角科技创新共同体联合攻关项目 2023CSJGG1103); 国家重点研发计划资助项目(2022YFB2901305); 安徽省高等学校科学研究基金资助项目(2023AH051197); 安徽理工大学引进人才科研启动基金资助项目(2023yjrc33).

作者简介: 顾成杰(1985—), 男, 博士, 教授, cjgu@aust.edu.cn.

通信作者: 朱东郡, 男, 博士, 讲师, dongjun.zhu@aust.edu.cn.

tation ability of the model when dealing with complex visual tasks. Finally, a scaling factor was introduced on the basis of the EIoU loss function to improve the optimization effect of the model. The experimental results show that compared with the baseline model, the proposed fire detection model has the detection accuracy (mAP0.5) increased by 6.5% and 1.3%, respectively, and the number of parameters decreased by 2.3% and 1.7%, respectively, on FAS-CVP dataset and FSD-CVP dataset.

Keywords: fire detection; feature enhancement; attention mechanism; global context

随着城市化进程的加速及气候变化的日益严峻,火灾灾害的发生频率与复杂性逐渐加剧。在城市环境中,建筑密集、人口集中以及基础设施复杂等特点,使得火灾蔓延速度极快且具有较高的致命性^[1]。而在森林及其他自然环境中,气候变化引发的极端天气^[2],如高温、干旱等,显著提高了火灾发生的风险及蔓延的范围^[3]。因此,传统的火灾烟雾检测技术在面对这些复杂环境时,暴露出灵敏度不足、适应性差、响应速度慢等明显局限性。

目前,大多数火灾烟雾检测技术依赖于基于传感器的监测系统,这些传感器通常基于光学、化学或热成像技术^[4]。然而,在复杂的环境中,烟雾的传播受多种因素的影响,如空气流动、气温、湿度及建筑结构等,这使得烟雾的检测愈加困难。例如,在城市高楼林立的区域,烟雾可能被建筑物遮挡或改变传播路径,导致传感器未能及时捕捉到火灾发生的初期信号;而在森林火灾中,干燥的气候、复杂的植被层次以及多变的地形条件,也使得火灾的检测面临更大的挑战^[5]。此外,火灾区域可能被浓密的烟雾遮挡,使得传统基于传感器的检测系统往往无法在火灾初期迅速识别,进而影响早期报警的准确性与响应速度。

近年来,随着人工智能、大数据分析 with 物联网技术的快速发展,新的火灾检测方法逐渐兴起^[6],尤其是在火灾烟雾检测领域。这些技术的融合为提高火灾检测的准确性、效率和实时性提供了新的解决方案。例如,基于深度学习的图像识别技术能够从大量监控视频中提取烟雾特征,显著提升火灾检测的精准度^[7]。深度学习的优势在于,它能有效应对烟雾在不同环境中的复杂表现,如遮挡效应、烟雾传播路径变化等,从而避免了传统方法中的盲区。而物联网技术通过联网传感器实时采集环境数据,结合大数据分析,实现对火灾发展态势的全方位监控与预警^[8]。然而,这些方法在复杂场景中仍然面临诸多挑战,特别是在低光照条件下背景与待检测物难以区分、火焰与烟雾形状多变、火灾蔓延速度快等情况,检测的准确性和实时性仍然存在一定的局限性。

在火灾探测领域,卷积神经网络(convolutional

neural networks, CNN)已被广泛应用^[9],凭借其卓越的特性,CNN 技术显著提升了火灾探测的精准度、运行效率和实时能力,也推动了其从传统方法向更为智能化的方向发展。例如,Ba 等^[10]通过 SmokeNet 强化卫星数据中的烟雾探测,显著提高了火灾响应的效率,Barmoutis 等^[11]利用 Faster R-CNN 模型来识别潜在的火灾区域。其他研究如 Huang 等^[12]和 Li 等^[13]提出的基于 CNN 的方法,进一步提高了火灾检测的态势感知和灵活性。

尽管 CNN 在火灾检测中具有较高的精度,YOLO (you only look once)算法因其检测快速和高效性,尤其适用于需要实时反应的火灾烟雾检测场景^[14]。YOLO 将目标检测问题转化为回归问题,通过一次前向传播同时完成物体的定位与分类任务,显著提高了检测速度和效率。在火灾烟雾检测中,YOLO 能够迅速检测到烟雾的存在并精确定位火灾位置,尤其在复杂的环境下如高楼大厦或森林火灾中,表现出较强的鲁棒性。Talaat 等^[15]提出了一种基于 YOLOv8 模型的智能火灾检测系统,旨在为提高智能城市中的火灾检测准确性并及时发出警报,以便及时疏散。其优势在于可以快速响应并且具有较好的实时性,但仍面临在复杂场景下的误检与漏检问题。

此外,为了进一步提升火灾烟雾检测的精度,研究者们提出了结合不同技术的创新方法。例如,Celik^[16]结合火灾颜色建模与运动检测,成功减少了传统系统中的误报。而 Xu 等^[17]提出的集成学习方法,整合了 YOLOv5 和 EfficientDet 模型,通过增强局部特征提取能力并结合 EfficientNet 学习全局信息,减少了误报并提升了检测准确性。尽管该方法在实时性上具有优势,但集成学习的计算复杂度可能影响检测速度,特别是在硬件资源有限的情况下,仍需进一步优化。李欣健等^[18]基于 YOLO 提出了一种火焰检测方法,通过融合和分析火焰特征有效实现火灾的检测。喻丽春等^[19]则在 Mask R-CNN 基础上,通过自下而上的特征金字塔融合,进一步提升了火灾检测的鲁棒性。然而,这些方法在复杂环境下,尤其是火灾蔓延迅速、烟雾扩散的情况下,仍面临低光照、背

景干扰等问题。

在处理复杂环境中的火灾检测问题时，低光照和动态环境下的烟雾识别、烟雾遮挡效应等因素，导致实时性不足，难以在火灾初期就做出及时预警是当前技术面临的主要挑战。传统的火灾检测算法可能在这些条件下表现不佳，尤其是在待检测物与背景难以区分时。例如，Xu 等^[20]提出的基于深度显著网络的视频烟雾检测方法，在低光照和动态环境中仍难以保持较高的检测精度和鲁棒性。El-Madafri 等^[21]提出了一种基于多任务知识蒸馏的轻量级森林火灾实时检测方法，该方法通过将 DenseNet201 教师模型的深度知识转移到 MobileNetV3 学生模型中，该方法优化了火灾检测的准确性和鲁棒性，但仍面临硬件资源限制和在极端环境下表现不佳的挑战。

综上所述，现有的火灾烟雾检测方法尽管在精度和效率上有所突破，但仍面临复杂环境下的挑战，尤其是在低光照、背景与待检测物难以区分、火焰与烟雾形状多变等情况下，检测的准确性和鲁棒性需要进一步提升。未来的研究需要更加注重算法的实时性与环境适应性，结合多模态数据和轻量化模型，以应对日益复杂的火灾检测需求。

本研究中采用的基准模型为 YOLOv9^[22]，是一种单阶段目标检测算法，采用了更为精简的网络架构，显著提高检测速度的同时保持了较高的检测精度。它在处理动态场景火灾检测时，表现出更强的适应性和鲁棒性，但在低光照环境和复杂背景干扰时仍面临一定的挑战。因此，基于上述火灾检测的问题，本文基于基准模型进行了改进，以进一步提高火灾烟雾检测的精确度和准确性，尤其是在复杂环境下的

表现。

设计了一种跨尺度特征增强模块 (cross-scale feature enhancement module, CSFE)，应用于模型的骨干网络，取代原来的特征提取结构。这一创新旨在保持检测精度稳定的同时，提升模型的通用性和可扩展性，确保在多种任务场景中的稳定性能。

设计了一种新的下采样模块 (context guided adown, CG-Adown)，在保证高效特征提取的同时，提升了模型对火灾特征 (如火焰、烟雾) 的敏感度和准确性。并且有效改善了火灾图像中复杂环境和低光照下的表现，增强了模型的鲁棒性与泛化能力。

给出一种多层注意力机制 (multi-layer attention, MLA)，自适应地调整各个通道的重要性，有效提升了特征提取和表达能力，从而增强了模型在复杂图像处理中的精度和泛化性能，在处理复杂场景时更具鲁棒性和有效性，有利于提升模型整体性能。

给出一种改进的 Inner-ElIoU 方法，通过引入区域分割与局部优化技术，结合多尺度辅助框机制和缩放因子，优化了目标框匹配精度，加速了模型的收敛，尤其在目标密集或背景复杂的场景中表现突出。

1 本文模型

本文所给出的模型结构如图 1 所示。首先，在特征提取部分设计了 CSFE 模块，该模块能有效提升了模型的特征提取能力、表达能力和泛化性能，从而能够更精确地聚焦关键特征并适应不同任务需求。同时将部分卷积模块替换为 CG-Adown 下采样模块，通过扩张卷积和全局上下文引导有效增强全局信息

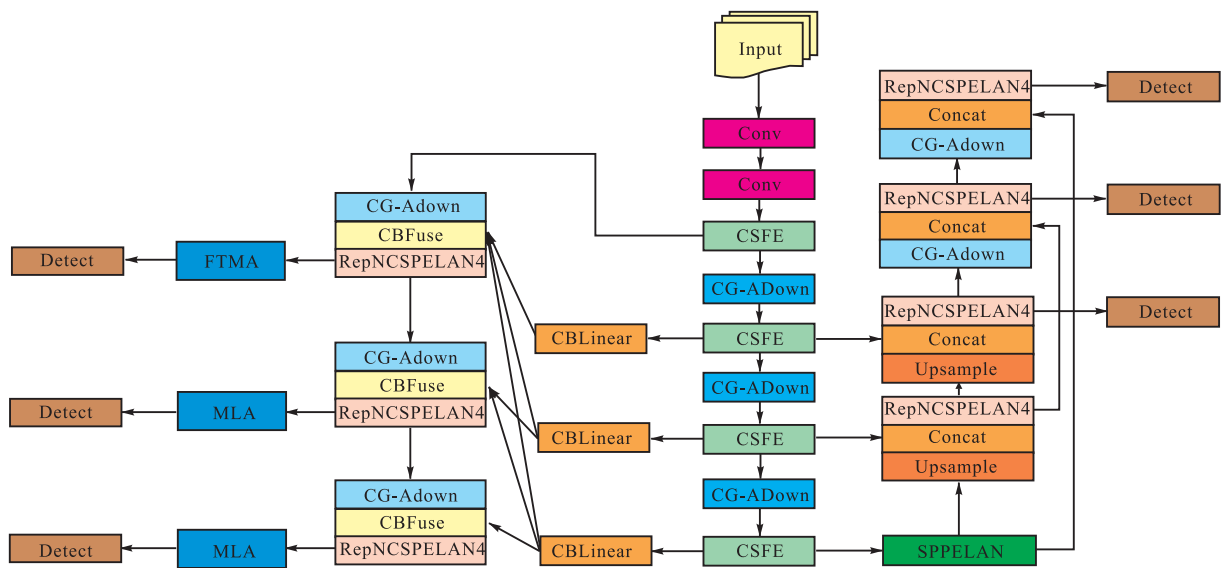


图 1 本文模型结构示意图
Fig.1 Ours structure chart

的捕捉能力. 随后, 设计了 MLA 模块, 有效提升了特征表达的精确度和空间信息的捕捉能力, 从而增强了模型在处理复杂图像时的性能. 最后, 利用改进的 Inner-EIoU 损失函数, 加快模型收敛速度并提高对于火灾目标的检测准确性, 有效提高模型鲁棒性.

1.1 CSFE 模块

CSFE 模块为特征提取模块, 首先使用注意力分配功能使模型更加聚焦于关键特征, 提高特征的利用效率, 另外在结构中加入了解压缩操作能够对特征进行更精细的加工和调整, 提取出更具代表性的信息, 结合 Dropout 层提高了模型的泛化能力, 减少过拟合. 同时在模型中加入了通道维度的自适应注意力机制, 能够使模型根据不同任务的需求, 智能地关注不同通道的重要性, 从而提高对重要特征的提取能力, 增强模型的表达能力和泛化性能. 最后将增强后的特征图与输入特征图拼接, 形成多尺度特征金字塔以捕捉来自不同层次的信息. CSFE 增强特征提取模块结构图如图 2 所示.

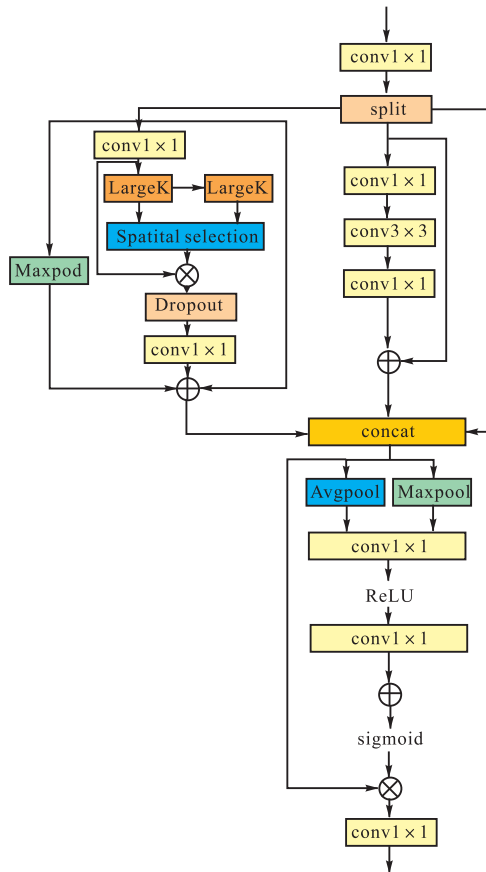


图 2 CSFE 结构示意图

Fig.2 CSFE structure chart

在该模块中, 首先输入一个形状为 (N, C, H, W) 的张量 X , 其中 N 是批量大小, C 是通道数, H 和 W 是特征图的高度和宽度. 首先使用 1×1 的卷积层

对输入的特征图进行变换, 再将其输出分为多个不同尺度的特征提取层进行金字塔特征提取以捕捉不同层次的信息, 分别得到低尺度特征 F_{low} 、中尺度特征 F_{mid} 和高尺度特征 F_{high} , 其形状变为 $(N, C_{low}, H_{low}, W_{low})$ 、 $(N, C_{mid}, H_{mid}, W_{mid})$ 、 $(N, C_{high}, H_{high}, W_{high})$, 其中 C_{low} 、 C_{mid} 、 C_{high} 是各个尺度特征图的通道数, H_{low} 、 H_{mid} 、 H_{high} 和 W_{low} 、 W_{mid} 、 W_{high} 是各个尺度下的特征图的空间尺寸, 接下来分别对高、中、低尺度特征图同时进行不同的操作. 首先对低尺度特征图 x_i 进行压缩操作, 通过一个 1×1 的卷积层降低特征的维度, 得到压缩后的特征图 $F_{low-compressed} = Conv_{1 \times 1}(F_{low})$, 这一步操作旨在减少冗余信息, 从而提高计算效率. 随后特征处理管道 LargeK 接收通过卷积层处理后的低尺度特征图 $F_{low-compressed}$, 依次通过深度卷积、空间卷积和特征聚合模块进行特征提取、选择与聚合. 此时, 网络对低尺度特征图中的重要区域进行重点关注, 通过自注意力机制对不同部分的特征图计算出注意力权重 α , 该权重表示了每个位置在当前任务中的重要性. 将计算得到的注意力权重 α 应用于压缩后的低尺度特征图上, 得到注意力增强后的特征图 F'_{low} :

$$F'_{low} = \alpha \odot F_{low-compressed} \quad (\alpha \in R^{(N, \frac{C}{4}, H, W)}) \quad (1)$$

在这一过程中关键区域的注意力权重较大, 因此加权求和时特征贡献显著增强; 无关区域的注意力权重较小, 加权求和时特征贡献被削弱, 实现了对低尺度特征图的关键区域特征增强, 同时抑制了无关区域, 使网络能够更加专注于关键特征, 提升模型性能. 接着输出的特征图与压缩后的特征图进行残差连接, 增强网络的表达能力, 再经过 Dropout 层, 其在训练时随机丢弃部分神经元, 打破神经元间的固定依赖关系, 避免过拟合, 以提高模型的泛化能力. 经过处理后的特征图再经过一个 1×1 卷积层恢复维度, 同时保留重要的特征表示, 得到最终处理后的低尺度特征图, 其形状为 $(N, C_{low}, H_{low}, W_{low})$. 此时, 如果传播的步幅 $s = 1$, 则直接将处理后的特征图与原始低尺度特征图 F_{low} 相加, 生成增强后的低尺度特征图 $F_{low-enhanced}$; 若步幅 $s = 2$, 则对低尺度特征图 F_{low} 先进行最大池化处理, 然后再与原始低尺度特征图 F_{low} 相加生成增强后的低尺度特征图 $F_{low-enhanced}$. 这种设计可以有效地保持特征图的空间信息, 并通过跳跃连接增强梯度流.

接下来对中尺度特征 F_{mid} 进行 Bottleneck 处理, 先后使用 1×1 卷积、 3×3 卷积、 1×1 卷积提取空间

特征,再与中尺度特征图进行残差连接,以降低其特征维度、提取更深层次的空间特征,从而有效降低计算量并提高特征非线性表达能力:

$$F'_{\text{mid}} = \text{Bottleneck}(F_{\text{mid}}) \quad (2)$$

在获得增强后的低尺度特征图 $F_{\text{low-enhanced}}$ 和经过 Bottleneck 处理后的中尺度特征图 F'_{mid} 后,接下来将这两者与高尺度特征图 F_{high} 拼接,得到融合后的特征图 F_{fusion} ,其形状为 (N,C,H,W) ,将拼接后的特征图 F_{fusion} 输入通道注意力机制,首先通过平均池化和最大池化操作分别得到 F_{avg} 和 F_{max} ,表示了每个通道的全局平均值和全局最大值, F_{avg} 和 F_{max} 的形状分别为 $(N,C,1,1)$ 和 $(N,C,1,1)$. 然后将这两个特征图分别通过 1×1 卷积层中将通道数降维以提取通道间特征关系,经过 ReLU 激活函数进行非线性变换,最终再经过一个卷积层将其恢复到原始维度,得到两个经过处理的特征图 F'_{avg} 和 F'_{max} . 将这两个经过处理的特征图相加并通过 sigmoid 函数生成通道权重 W_{channel} ,表示每个通道在当前输入特征图中的重要性. 将生成的通道权重 W_{channel} 与融合后的特征图 F_{fusion} 相乘,得到加权后的特征图 F_w :

$$F_w = F_{\text{fusion}} \odot W_{\text{channel}} \quad (3)$$

这一过程可以显著增强关键特征,同时抑制了噪声和不相关信息,从而提升模型对重要特征的关注度. 最终加权后的特征图经过一个 1×1 卷积调整输出特征图的维度,同时通过卷积操作增强特征图的表达能力.

CSFE 模块通过多尺度特征提取、注意力机制和特征压缩等手段有效提升模型性能. 首先通过低、中、高尺度的特征提取,能够捕捉不同层次的图像信

息,增强对复杂场景的理解能力. 同时引入注意力机制可以让模型自动关注关键信息,并抑制噪声和无关信息,提高特征图的表达能力. 通道注意力机制通过加权通道的重要性进一步优化了特征图表示,增强了模型的判别力. 此外,特征压缩、Dropout 层和跳跃连接的设计减少了计算复杂度,并有效提升了模型的泛化能力,防止了过拟合. 通过跳跃连接保持了特征图的空间信息,确保了不同尺度之间的有效信息传递. CSFE 模块在提高模型表达能力的同时,保持了高效的计算和良好的泛化性能,适用于复杂的视觉任务.

1.2 CG-Adown 模块

在深度学习模型中,下采样是减少特征图空间维度的常用技术,其有助于模型在更高层次上捕捉图像的特征,有效减少计算量. 基准模型中 Adown 模块的缺点在于全局信息提取受限、特征融合不充分. 文献[23]中提出的 CG-Net 是一种通过引入全局上下文信息引导和扩张卷积^[24]来提升特征提取能力的网络结构,但其扩张卷积和特征拼接操作增加了计算复杂度和内存开销. 因此本文给出了一种 CG-Adown 下采样模块,相较于原始的 CG-Net,其在多尺度特征建模和计算效率方面具有显著的提升. CG-Adown 通过扩张卷积和全局上下文引导,有效增强了对远距离依赖和全局上下文信息的捕捉能力,同时采用残差连接解决了深层网络中常见的梯度消失和信息丢失问题. 与传统的下采样方法不同,CG-Adown 使用高效的 1×1 卷积进行通道压缩,同时保持了特征图的表达能力,避免了计算和内存的过度消耗. 此外,通道压缩和残差设计确保了模型在深层网络中依然能够稳定训练并加速收敛,提供了更高的参数和计算效率. CG-Adown 结构如图 3 所示.

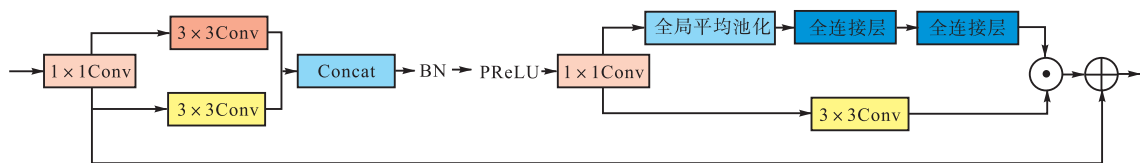


图 3 CG-Adown 结构示意图

Fig.3 CG-Adown structure chart

CG-Adown 模块首先使用 1×1 卷积对输入图像进行通道数压缩,降低特征图的维度并且有效过滤图像中的噪声和冗余信息. 然后将压缩后的图像分别通过两个分支,一个分支用于提取局部特征,另一个分支用于获取超距上下文信息. 这两个分支并行操作,旨在同时捕捉局部细节和全局结构,从而增强网络对多尺度信息的感知能力.

局部特征提取分支使用常规的 3×3 卷积操作进

行局部信息的捕捉,步幅为 1,此卷积操作在每个通道上独立执行,因此能够有效捕捉每个位置周围的细粒度局部信息,包括局部区域内的空间结构和纹理信息. 为了捕捉更长范围的上下文信息,另一个分支中使用扩张卷积,通过在卷积核之间增加卷积核的感受野,有效扩大感受区域,从而能够捕捉到更长程的依赖关系. 通过这种方式,网络能够识别更远距离的上下文信息,增强了对全局结构的建模能力,并且不增

加额外的计算开销。

在局部特征和超距特征提取之后,将输出特征图沿着通道维度进行拼接,形成一个新的特征图,拼接后的特征图包含了局部特征和超距特征的信息。接着应用批量归一化(BatchNorm)和 PReLU 激活函数对拼接后的特征图进行处理,从而提升特征图的表达能力和训练稳定性。随后为了减少计算负担和防止过多冗余特征,使用一个 1×1 卷积进行通道数压缩,同时保持其关键信息,该操作通过减少通道数,降低了模型的计算复杂度。随后,模块进一步通过全局上下文引导操作对特征图进行优化。首先,应用全局平均池化操作,对输入特征图的每个通道进行平均池化,得到每个通道的全局平均值。接下来,将池化后的特征图输入到两个全连接层中,使用全连接层提取高维特征,再通过 ReLU 激活函数进行非线性变换,能够有效引导网络关注图像中的全局信息,增强网络的全局感知能力,提高对复杂场景和多尺度目标的识别能力。同时将压缩后的特征图使用 3×3 扩张卷积进行处理,有效扩大了感受野,能够捕捉到更广泛区域内的特征信息,增强特征的丰富度。

为避免深层网络中的梯度消失问题,并确保信息的有效传递,模块在最终输出时使用残差连接。最终,优化后的全局上下文特征图与通过 1×1 卷积处理后的输入特征图相加,得到最终输出。这种设计确保了网络的深度和稳定性,避免了信息丢失,并使得模型能够高效地处理更加复杂的视觉任务。

CG-Adown 模块通过对输入特征图进行通道数压缩和多分支的特征提取,结合局部与全局信息的多层次融合,在提高特征表示能力的同时,显著减少了计算开销。通过扩张卷积和全局上下文引导的设计,CG-Adown 模块有效提升了模型对多尺度和长程依赖关系的建模能力。同时,残差连接的应用保证了信息流的稳定,使得深层网络在训练过程中能够保持高效的性能。

1.3 MLA 模块

MLA 注意力模块通过结合基于特征变换的通道注意力 CAFT 和基于多尺度空间注意力 MSSA 机制,旨在通过自适应地调整通道和空间维度的权重,从而显著提升神经网络在处理复杂视觉任务时的特征表达能力与整体性能。该模块通过精心设计的多层次注意力机制,能够有效利用输入数据在通道和空间维度上的多重信息,进而增强模型的感知能力、表达能力和泛化性能。

具体而言,MLA 模块采用了两种互补的注意力机制:通道注意力^[25](CAFT)以及空间注意力^[26]

(MSSA)。首先,基于特征变换的通道注意力机制通过建模通道之间的相互关系,自适应地调整各个通道的重要性,以便聚焦于对任务最具信息量的特征。这一过程通过结合 1×1 卷积与池化操作,从不同层次和尺度提取特征,并使用 Mish 激活函数进行非线性映射,从而增强网络的表示能力,并有效避免梯度消失问题。随后,通过 Sigmoid 激活函数生成的通道注意力图,将对输入特征图的各个通道进行加权调整,以突出有意义的通道,抑制冗余和无关的信息。其次,基于多尺度空间注意力机制通过多尺度卷积操作,聚焦于图像中的关键区域,尤其是在处理包含多尺度目标的复杂场景时,能够自动识别和强化重要空间区域。通过最大池化和平均池化的组合,MSSA 机制提取了局部显著特征与全局上下文信息,并通过多尺度卷积(如 1×1 、 3×3 、 5×5)捕捉不同尺度的空间特征。最终,经过 Sigmoid 激活函数处理的空间注意力图对输入特征图进行加权,突出图像中的关键区域,提升模型对细粒度信息的感知能力。MLA 注意力模块如图 4 所示。

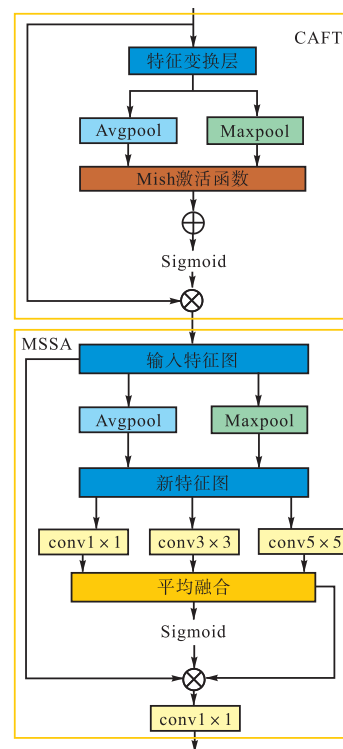


图 4 MLA 结构示意图
Fig.4 MLA structure chart

首先,模块接收一个形状为 $X \in R^{N \times C \times H \times W}$ 的输入特征图,其中 N 是批量大小, C 是通道数, H 和 W 分别是特征图的高度和宽度。接下来将输入特征图传入基于特征变换的通道注意力中进行处理,通过建模通道之间的相互关系,自适应地调整各通道的权重,

从而突出重要特征. 首先通过特征变换层对输入特征图 X 施加 1×1 的卷积操作, 以实现通道间的线性组合, 从而生成新的特征图 X' :

$$X' = W_{1 \times 1} * X + b \quad (4)$$

式中: $W_{1 \times 1}$ 表示 1×1 卷积核, $*$ 表示卷积操作, b 为偏置项.

对变换后的特征图 X' 分别进行最大池化和平均池化操作分别捕捉局部最显著特征和全局背景信息, 得到 $X'_{\max} = \text{MaxPool}(X')$ 和 $X'_{\text{avg}} = \text{AvgPool}(X')$, 对池化结果进行 Mish 激活函数处理后, 通过其非线性性质提升了模型的表达能力, 得到 F_{\max} 和 F_{avg} . 将处理后的特征图 F_{\max} 和 F_{avg} 相加并通过 Sigmoid 激活函数得到通道注意力权重图 A_c , 最后将输入特征图 X 乘以通道注意力权重图, 得到加权后的特征图 X_c :

$$X_c = X \odot \sigma(F_{\max} + F_{\text{avg}}) \quad (5)$$

其中, $\sigma(\cdot)$ 表示 Sigmoid 激活函数, 确保 A_c 的值在 $[0, 1]$ 之间.

将加权后的特征图 X_c 传入基于多尺度的空间注意力, 旨在通过多尺度的卷积操作捕捉空间上重要的区域信息, 从而增强模型对关键区域的关注能力. 首先对其进行最大池化和平均池化操作, 在不同尺度上对图像特征进行抽象, 从而捕捉到不同大小的特征, 再将最大池化和平均池化后的特征图拼接成一个新的特征图 $X_{c,\text{concat}}$, 可以有效地融合不同池化策略提取的信息, 增强空间特征表达能力. 接着对新特征图应用不同尺度的卷积操作, 以捕捉不同尺度的空间信息:

$$A_s^{(1)} = \text{Conv}_{1 \times 1}(X_{c,\text{concat}}) \quad (6)$$

$$A_s^{(2)} = \text{Conv}_{3 \times 3}(X_{c,\text{concat}}) \quad (7)$$

$$A_s^{(3)} = \text{Conv}_{5 \times 5}(X_{c,\text{concat}}) \quad (8)$$

通过均值融合多个尺度的空间注意力图, 得到融合后的空间注意力图 A_s :

$$A_s = \frac{1}{3} \sum_{i=1}^3 A_s^{(i)} \quad (9)$$

对融合后的空间注意力图应用 Sigmoid 激活函数, 得到最终的空间注意力图. 将加权后的特征图 X_c 与空间注意力图以及平均融合后的特征图相乘, 再通过 1×1 卷积对加权后的特征图进一步进行通道信息重建与融合, 最终得到模块的输出特征图 X_{out} :

$$X_{\text{out}} = \text{Conv}(X_c \odot \sigma(A_s)) \quad (10)$$

MLA 模块通过结合基于特征变换的通道注意力和多尺度空间注意力, 实现了在通道和空间两个维度

上的自适应加权, 极大增强了网络对重要特征的关注能力.

1.4 Inner-ElIoU 损失函数

在目标检测领域, 传统的 CIoU 与 ElIoU 损失函数虽改进了传统 IoU, 但尚存局限性. CIoU 基于预测框和真实框的中心点距离、长宽比及重叠区域优化回归, 然而面对复杂形状与尺度差异大的场景, 尤其框间几无重叠时, 梯度信息匮乏, 致使模型收敛慢. ElIoU 引入了最小包围物体及惩罚项以补 CIoU 之不足, 却易受目标形状复杂性和标注噪声干扰, 导致损失计算欠精确, 且缺乏框匹配自适应调整能力^[27].

鉴于此, 本文采用 Inner-ElIoU 替代传统的 CIoU. Inner-ElIoU 借助多尺度辅助边界框加速回归过程, 利用缩放因子优化尺度调整以提升检测精度^[28], 凭借最小包围物体实现精确重叠度计算并减少冗余影响, 通过自适应框匹配增强鲁棒性, 进而有效应对复杂场景, 优化目标检测效果.

Inner-ElIoU 损失函数定义如下所示:

$$\text{IoU} = \frac{|b \cap b_{\text{gt}}|}{|b \cup b_{\text{gt}}|} \quad (11)$$

$$L_{\text{IoU}} = 1 - \text{IoU} \quad (12)$$

其中 b 和 b_{gt} 分别为预测框和真实框, L_{IoU} 用于衡量预测框与真实框的交并比差异.

$$L_{\text{dis}} = \frac{\rho^2(b, b_{\text{gt}})}{C_w^2} \quad (13)$$

$$L_{\text{asp}} = \frac{\rho^2(w, w_{\text{gt}})}{C_w^2} + \frac{\rho^2(h, h_{\text{gt}})}{C_h^2} \quad (14)$$

$$L_{\text{ElIoU}} = L_{\text{IoU}} + L_{\text{dis}} + L_{\text{asp}} \quad (15)$$

式中: w 和 h 分别代表预测框的宽度和高度; w_{gt} 和 h_{gt} 分别代表真实框的宽度和高度; $\rho(b, b_{\text{gt}})$ 是预测框和真实框中心点之间的欧几里得距离; $\rho(w, w_{\text{gt}})$ 和 $\rho(h, h_{\text{gt}})$ 分别表示预测框和真实框的宽度差异和高度差异; C_w 和 C_h 分别是预测框和真实框的最小外接矩形的宽度和高度. L_{dis} 是距离损失, 通过计算中心点之间的欧几里得距离来惩罚预测框与真实框中心点之间的距离. L_{asp} 为长宽比损失, 用于减少预测框和真实框的宽度和高度之间的差异, 优化边界框的形状. ElIoU 损失由 IoU 损失、距离损失以及长宽比损失构成.

$$b_{\text{gt}}^l = x_{\text{gt}}^c - \frac{w_{\text{gt}} \times \alpha}{2}, \quad b_{\text{gt}}^r = x_{\text{gt}}^c + \frac{w_{\text{gt}} \times \alpha}{2} \quad (16)$$

$$b_{\text{gt}}^t = y_{\text{gt}}^c - \frac{h_{\text{gt}} \times \alpha}{2}, \quad b_{\text{gt}}^b = y_{\text{gt}}^c + \frac{h_{\text{gt}} \times \alpha}{2} \quad (17)$$

$$b_l = x_c - \frac{w \times \alpha}{2}, \quad b_r = x_c + \frac{w \times \alpha}{2} \quad (18)$$

$$b_t = y_c - \frac{h \times \alpha}{2}, \quad b_b = y_c + \frac{h \times \alpha}{2} \quad (19)$$

$$inter = (\min(b_{gt}^r, b_r) - \max(b_{gt}^l, b_l)) \cdot (\min(b_{gt}^b, b_b) - \max(b_{gt}^t, b_t)) \quad (20)$$

$$union = (w_{gt} \times h_{gt}) \times \alpha^2 + (w \times h) \times \alpha^2 - inter \quad (21)$$

$$IoU_{inner} = \frac{inter}{union} \quad (22)$$

$$L_{inner-EIoU} = L_{IoU} + IoU - IoU_{inner} \quad (23)$$

真实框和预测框表示为 b_{gt} 和 b , 真实框的中心点和框内点分别由 (x_{gt}^c, y_{gt}^c) 和预测框的中心点和框内点 (x_c, y_c) 表示. α 是一个尺度因子, 用来调整辅助边界框的尺度.

2 实验结果与分析

本文模型训练所使用的计算机操作系统为 Windows 10, 计算机处理器为 16 vCPU Intel(R) Xeon(R) Platinum 8474C CPU @ 2.50GHz、显卡为 RTX 4090D, 运行内存 24GB, 实验深度学习框架为 Pytorch1.11, CUDA 版本为 11.3, 编程语言为 Python3.8. 本文中模型在 GPU 上训练 200 轮; 批次大小为 16; 初始学习率为 0.01; 训练和验证时输入图像的尺寸大小为 640×640 .

2.1 数据集与评价指标

本文所用数据集均包含城市、森林等各个场景以及不同角度的火灾烟雾情况, 数据集中包含两个类别, 分别为“Fire”、“Smoke”, 并且图像大小为 640×640 . 其中, FAS-CVP 数据集^[29]包括 9 815 张火灾图像, 对 FAS-CVP 数据集进行预处理, 将数据集格式转换为 txt 格式并将数据集分别划分为训练集、验证集, 比例为 8 : 2. FSD-CVP 数据集^[30]包括 7 468 张火灾图像, 类似地对数据集 FSD-CVP 进行预处理, 将数据集格式转换为 txt 格式, 然后将数据集按 7 : 2 : 1 的比例分别划分为训练集、验证集和测试集.

表 1 实验数据集构成

Tab.1 Composition of the experimental dataset

数据集	FAS-CVP	FSD-CVP
图像数量	9 815	7 468
图像大小	640 × 640	640 × 640
类别数	2	2

对于火灾目标检测任务, 本文实验结果使用目标检测中的部分常用评估指标^[31]来评价模型性能, 包

括精确度 P 、召回率 R 、平均精确度 (mean average precision, P_{ma}), 参数量 (params), 其中平均精确度采用 $P_{ma0.5}$ 、 $P_{ma0.5-0.95}$ 两种方式, $P_{ma0.5}$ 表示阈值为 0.5 时的平均精确度, $P_{ma0.5-0.95}$ 表示其阈值从 0.5 以步长 0.05 逐步上升为 0.95 时的平均检测精度. 计算方式如式 (24) ~ (27) 所示:

$$P = \frac{B_{TP}}{B_{TP} + B_{FP}} \quad (24)$$

$$R = \frac{B_{TP}}{B_{TP} + B_{FN}} \quad (25)$$

$$P_A = \int_0^1 P(R) dR \quad (26)$$

$$P_{ma} = \frac{1}{N} \sum_{i=1}^N P_{A,i} \quad (27)$$

式中: B_{TP} 表示真正例, 指模型正确地将真实目标识别为目标; B_{FP} 表示假正例, 指模型错误地将非真实目标识别为目标; B_{FN} 表示假负例, 指模型错误地将真实目标识别为非目标; P_A 表示单个类的精确度; $P_{A,i}$ 是指第 i 个类别的平均精度; N 表示总的类别数目.

精确度衡量算法相对于检测结果的准确性, 召回率衡量的是模型检测到所有真实目标的能力, 平均精确度综合评估模型在不同召回率下的表现以反映模型的总体效果. 参数量指的是模型训练过程中的参数总量, 包括权重、偏置值等, 用来衡量模型的复杂程度和规模. 较大的参数量表示提升了模型的表达能力和精度, 但需要更多存储空间和计算资源. 较小的参数量则表示模型更轻量化、效率更高, 适合资源受限的环境.

2.2 对比实验

为了进一步验证本文算法在检测火灾图像的有效性以及泛化性, 本文利用不同模型分别在两个数据集上进行了性能对比, 模型包括 Faster R-CNN、Tood、Atss、RT-DETR、YOLOv8、YOLOv9、YOLOv10、YOLOv11 以及本文算法, 使用精确度、召回率、 $P_{ma0.5}$ 、 $P_{ma0.5-0.95}$ 和参数量等指标来评估其性能, 对比结果如表 2 和表 3 所示, 其中, Ours 是本文的改进模型.

数据集 FAS-CVP 上的结果如表 2 所示. 结果显示本文所提算法相较于 YOLO 系列最新的 YOLOv11, P 和 R 分别提升了 3.1% 和 2.2%, $P_{ma0.5}$ 和 $P_{ma0.5-0.95}$ 分别提升了 4.0% 和 3.9%, 表明所提模型在准确性和全面性上均有所增强; 与 YOLOv10 相比, P 和 R 分别提升 3.9% 和 1.2%, $P_{ma0.5}$ 和 $P_{ma0.5-0.95}$ 分别提升 3.2% 和 3.8%, 这说明本文模型检测效果更加全面且准确; 与 RT-DETR 相比, P 和 R 分别提升 8.6% 和 8.7%, $P_{ma0.5}$ 和 $P_{ma0.5-0.95}$ 分别提升 11.8% 和

13.8%，表明本文模型在高 IoU 阈值下具有优势，其在细粒度检测性能较高；与两阶段模型 Faster R-CNN 相比， P 和 R 分别提升 9.8% 和 16.5%， $P_{ma0.5}$ 和 $P_{ma0.5-0.95}$ 分别提升 9.9% 和 18.6%，表明本文模型能够显著提高检测性能，并且在高 IoU 阈值下的提升尤为突出，显示了其在精细目标检测中的显著改进；与基准模型相比， P 和 R 分别提升了 6.5% 和 2.9%， $P_{ma0.5}$ 和 $P_{ma0.5-0.95}$ 分别提升 5.0% 和 3.6%，表明本文模型在检测任务中具有更高的准确性和更好的泛化能力。因此，本文模型在显著提升性能的同时，具有较好的检测精度和计算效率。数据集 FSD-CVP 上的结果如表 3 所示。相较于 RT-DETR，本文模型在精确度 P 提高了 7.9%，召回率 R 提高了 4.7%，表示本文

模型增强了正确识别目标的能力。同时， $P_{ma0.5}$ 和 $P_{ma0.5-0.95}$ 分别提高了 9.3% 和 7.9%，此数据表明本文模型增强了火灾检测的准确性和全面性。相较于 YOLO 系列最新的 YOLOv11， P 和 R 分别提升了 1.2% 和 0.7%， $P_{ma0.5}$ 和 $P_{ma0.5-0.95}$ 分别提升了 1.7% 和 2.8%，表明本文模型在高 IoU 阈值下具有优势，检测效果更加全面且准确；与 YOLOv10 相比， P 和 R 分别提升 3.1% 和 1.3%， $P_{ma0.5}$ 和 $P_{ma0.5-0.95}$ 分别提升 3.2% 和 3.5%，表明本文方法在 P 和 R 之间达成了较为良好的平衡。由表 2 和表 3 的结果可以看出，本文模型对火灾的检测效果不仅优于基准模型，而且相较于其他对比方法显著提高了火灾检测的精确性与鲁棒性。

表 2 基于 FAS-CVP 数据集的对比实验

Tab.2 Comparative experiments basing on FAS-CVP dataset

方法	$P/\%$	$R/\%$	$P_{ma0.5}/\%$	$P_{ma0.5-0.95}/\%$
Tood ^[32]	0.617	0.457	0.557	0.226
Atss ^[33]	0.613	0.448	0.551	0.225
Faster R-CNN	0.589	0.363	0.501	0.179
RT-DETR ^[34]	0.601	0.441	0.482	0.277
YOLOv11	0.656	0.506	0.560	0.326
YOLOv8	0.655	0.492	0.548	0.314
YOLOv10	0.648	0.516	0.568	0.327
Baseline	0.622	0.499	0.550	0.329
本文模型	0.687	0.528	0.600	0.365

表 3 使用数据集 FSD-CVP 进行的对比实验

Tab.3 Comparative experiments using the FSD-CVP dataset

方法	$P/\%$	$R/\%$	$P_{ma0.5}/\%$	$P_{ma0.5-0.95}/\%$
Tood	0.652	0.528	0.661	0.397
Atss	0.676	0.539	0.649	0.384
Faster R-CNN	0.589	0.535	0.642	0.375
RT-DETR	0.605	0.526	0.558	0.327
YOLOv11	0.672	0.566	0.634	0.378
YOLOv8	0.669	0.556	0.632	0.377
YOLOv10	0.653	0.560	0.619	0.371
Baseline	0.678	0.561	0.638	0.395
本文模型	0.684	0.573	0.651	0.406

2.3 消融实验

为了进一步分析验证本文算法的改进效果，本文进行一系列消融实验对各个改进方案进行验证。以 YOLOv9 为基线模型，在实验环境和参数设置保持一致的情况下，分别在基准模型上加入跨尺度特征增强模块 CSFE、多层注意力机制 MLA、CG-Adown 下采样模块以及改进 EIou 损失函数。消融实验结果如表 4 所示。

根据表 4 的结果可知，加入本文给出的 CSFE 特征提取模块后精确度提高了 6.1%，召回率提高了 1.0%， $P_{ma0.5}$ 提高了 4.2%， $P_{ma0.5-0.95}$ 提高 3.2% 并且参

数量下降了 1.5%，表明增强特征提取模块通过自适应的通道注意力机制帮助模型更加聚焦于关键特征，在提升对目标精确捕捉的同时减少了参数量。加入 CG-Adown 模块后精确度提高了 6.0%，召回率增加了 1.9%， $P_{ma0.5}$ 提高了 4.2%， $P_{ma0.5-0.95}$ 提高 3.4%，参数量增加了 7.4%，数据表明加入 CG-ADown 模块通过扩张卷积和全局上下文信息的引导，有效扩大感受区域，优化了模型对远距离目标和复杂场景的捕捉能力，使得模型在增加少量开销的同时大大提升了检测精度和召回率。MLA 注意力机制的加入使得精确度提高了 1.4%，召回率增加了 0.3%， $P_{ma0.5}$ 增加了

表 4 各模块对结果的影响

Tab.4 Influence of each module on the results

方法	CSFE	CG-Adown	MLA	Inner-ElIoU	P/%	R/%	$P_{ma0.5}/%$	$P_{ma0.5-0.95}/%$	参数量/M
Baseline+	✓	✓	✓	✓	62.2	49.9	55.0	32.9	57.98
					68.3	50.9	59.2	36.1	57.10
					68.2	51.8	59.2	36.3	62.27
	✓	✓	✓	✓	63.6	50.2	55.8	32.9	57.96
					63.8	50.8	56.7	33.2	57.70
					68.7	52.8	60.0	36.5	56.72

0.8%, $P_{ma0.5-0.95}$ 保持不变, 参数量降低了 0.2%, 表明 MLA 注意力机制优化了多尺度特征的融合能力, 使得模型能更好地应对不同尺度的目标, 并且, 通过减少模型的参数量, 降低了模型的复杂度, 从而降低计算资源的消耗和推理开销. 加入 Inner-ElIoU 损失函数后精确度提高了 1.6%, 召回率增加了 0.9%, $P_{ma0.5}$ 提高了 1.7%, $P_{ma0.5-0.95}$ 提高 0.3%, 并且参数量减少了 0.5%, 减少了模型参数量, Inner-ElIoU 损失函数提升了边界框回归精度, 优化了检测结果, 并通过减少不必要的参数优化了模型. 但总体而言, 本文模型在所有指标中表现最佳, 精确度为 68.7%, 召回率为 52.8%, $P_{ma0.5}$ 为 60.0%, $P_{ma0.5-0.95}$ 为 36.5%, 与基准模型相比, 本文模型在 P 、 R 、 $P_{ma0.5}$ 和 $P_{ma0.5-0.95}$ 指标上分别提升了 6.5%、2.9%、5.0%和 3.6%, 并且参数量下降了 2.3%, 该结果表明本文所提模型在提高精度和召回率的同时, 显著优化了计算效率和推理速度, 实现了模型轻量化, 在整体检测性能上具有显著优势.

2.4 数据分析

为了更好地证明本文方法的优越性, 本文还将根据 F_1 分数曲线, P - R 曲线等几个方面再次论证. P - R 曲线是一种用于评估二分类模型性能的重要指标, 特别是目标检测模型在处理类别不平衡时的情况, 主要用于衡量模型在不同阈值下的精确度和召回率之间的权衡关系. P - R 曲线的形状可以揭示模型的不同特征, 当模型在高精度和高召回率之间越平衡, 曲线越接近右上角. 如图 5 所示, 本文模型的 P - R 曲线比基准模型的 P - R 曲线更加靠近图的右上角, 该结果表明在相同召回率的情况下, 本文模型的精度相较于基准模型更高.

F_1 分数为模型精确率和召回率的调和平均值, 综合考虑了这两者的权衡, 是用来衡量二分类模型精确率的重要指标, 曲线的位置越接近 Y 轴的上方, 说明模型在多个阈值下的 F_1 分数越高, 表现越好. 基准模型的 F_1 分数曲线与本文模型的 F_1 分数曲线如图 6 所示. 可以明显看出本文模型的 F_1 分数曲线整体在基准模型的 F_1 分数曲线上方, 表示本文模型相

较于基准模型在各个种类均取得较高的 F_1 分数, 即意味着本文模型更好地平衡了精确度和召回率, 并且在多个置信度水平下表现更稳定. 结合图 6 的结果, 可以得到本文模型的整体性能优于基准模型.

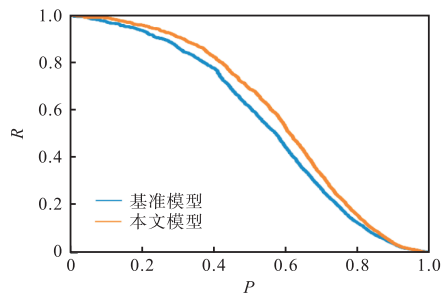


图 5 本文模型与基线模型的 P - R 曲线对比

Fig.5 Comparison of P - R curves between ours and baseline models

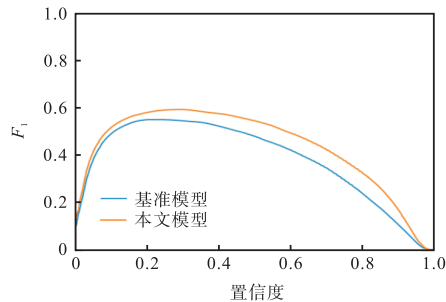


图 6 本文模型与基线模型的 F_1 曲线对比

Fig.6 Comparison of F_1 curves between ours and baseline models

2.5 可视化分析

热力图是一种用于展示数据空间分布的图形方式, 用于通过颜色的变化来表示二维数据的强度、频率或密度. 通常通过颜色渐变来表示数值的高低, 颜色的深浅反映了数值的强弱. 高数值或高密度区域使用暖色调表示 (如红色), 低数值或低密集区域则使用冷色调表示 (如蓝色), 而中间数值则使用黄色或绿色等颜色来表示. 每个单元格通过颜色深浅来表示该位置的数值大小. 在热力图中, 每个数据点或区域都被分配一个颜色值, 该值反映了该位置的数据密集

程度或数值大小。在目标检测过程中,热力图能够直观地展示模型对输入的关注区域,深色区域通常代表检测到目标的高概率区域,浅色区域则代表检测到目标的低概率区域,有助于模型理解图像中的重要区域,减少无关背景的干扰。图 7 展示了本文所提模型与基准模型的热力图对比。由图可以看出,本文模型

在火灾检测中的效果明显优于基准模型:背景检测覆盖较少且亮度较低,而在目标区域,热力图呈现较亮的颜色且覆盖面积较大,表明模型对火灾的检测效果较好。上述结果验证了本文模型在火灾检测中的优异表现,同时也展示了其在复杂背景下能有效抑制背景干扰,聚焦与重要特征区域特征。

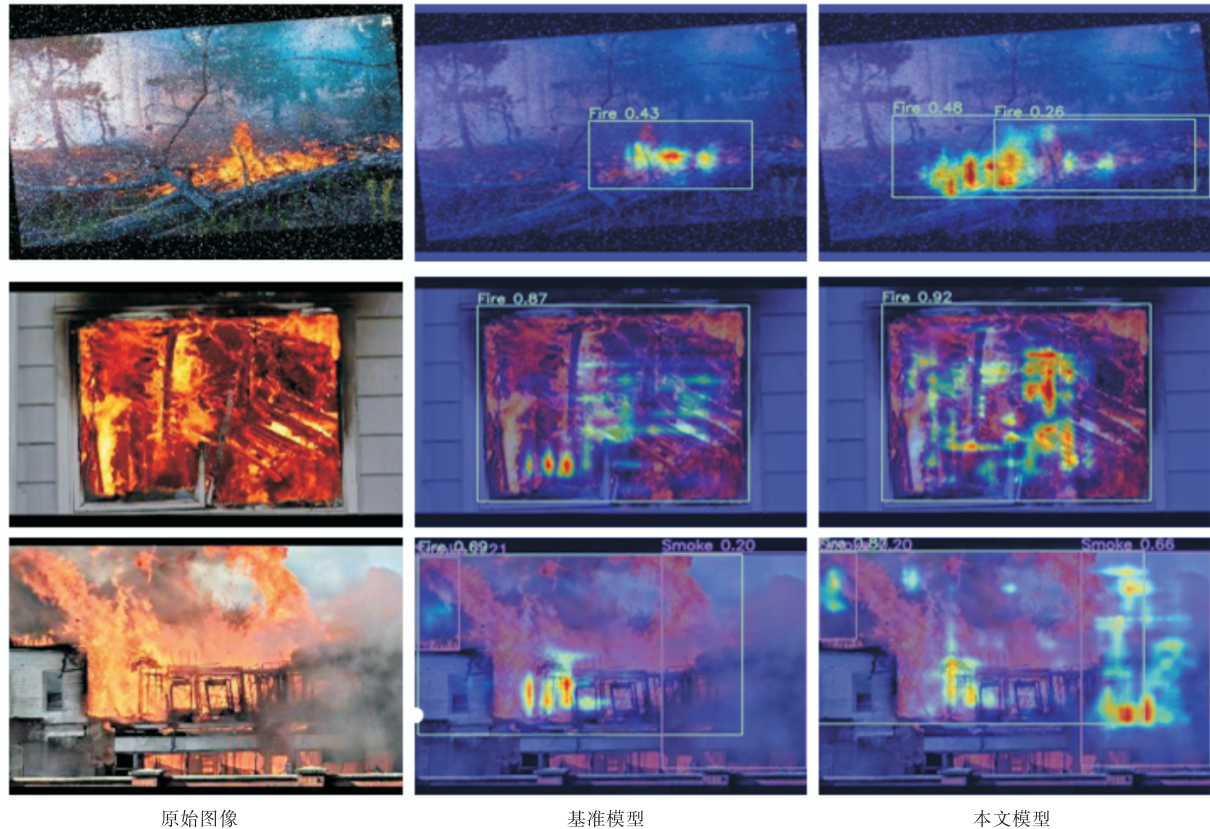


图 7 热力图对比

Fig.7 Heat map comparison

本文所提模型和基准模型在火灾场景下的检测结果对比如图 8 所示,包括了低光照时背景与烟雾难以区分、背景干扰情况下的火灾图像。由图 8 可知,基准模型未能准确识别到较小的火焰以及背景干扰情况下的火焰,而本文所提模型成功捕捉到了该难以察觉的目标。此外,在面临光线昏暗以及背景较为复杂的场景下,基准模型误将车辆的灯光检测为火焰以及未能准确识别出火焰目标,而所提模型则全面检测到了正确的火焰与烟雾目标。对比分析结果清晰地呈现出相较于基准模型,本文所提模型对复杂环境下的火灾识别和检测具有显著优势。同样与 YOLOV11 算法相比,本文模型在火灾目标较小的情况下效果显著,可以准确地识别出较为密集的火灾目标。本文所提基准模型以及其他模型在图像背景复杂以及图像质量较差、光照较低的情况下容易出现误检和漏检的

问题,而本文算法有效改善了这一问题,提高了火灾目标检测的准确性。

图 9 展示了在 FSD-CVP 数据集中背景复杂、目标比较密集的情况下检测结果的对比图。从图中可以观测得到,在基准模型的检测结果中存在较多的误检和漏检情况。由于背景元素的干扰和目标的密集分布,基准模型难以准确地分辨出目标物体,导致其检测框出现偏差,部分目标完全未被检测到。对于 YOLOv11 模型,虽然相较于基准模型有了一定的改善,但仍然存在问题。在目标密集区域, YOLOv11 出现了检测框重叠和部分目标误判的现象。而本文模型的检测框能够较为准确地识别目标物体,并且在火灾目标密集区域也能有效地避免重叠和误判现象。验证了本文模型在处理复杂场景以及目标密集的情况下的有效性和优越性。

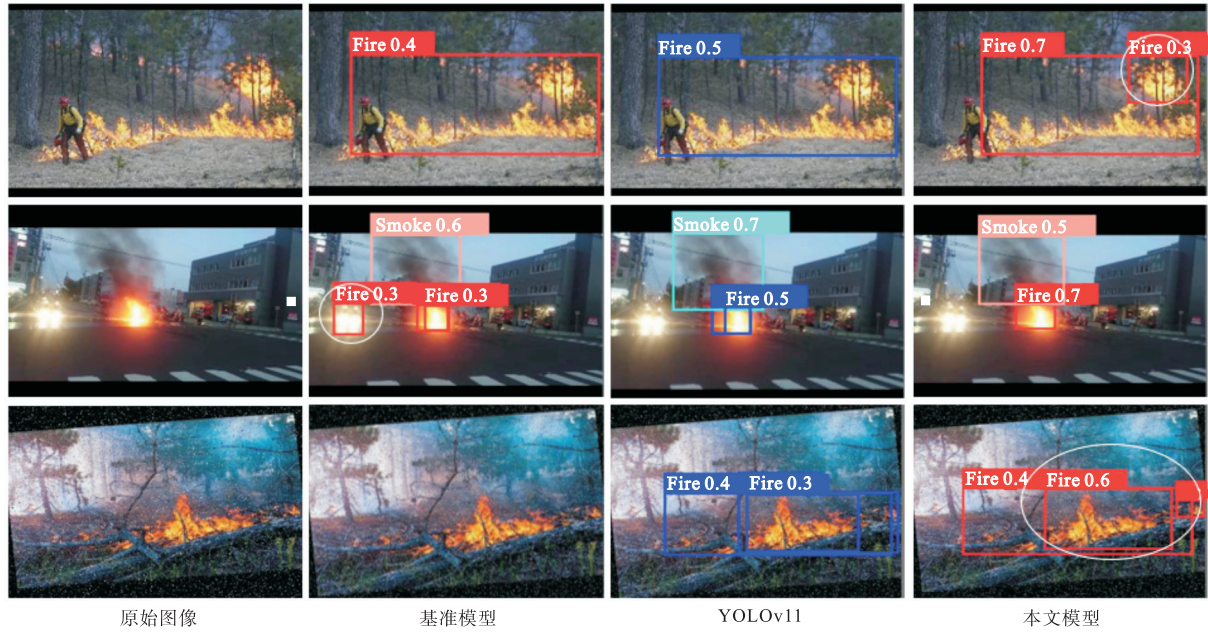


图 8 数据集 FAS-CVP 上的检测对比
Fig.8 Comparison plot of detection on the FAS-CVP dataset



图 9 数据集 FSD-CVP 上的检测对比
Fig.9 Comparison plot of detection on the FSD-CVP dataset

3 结 论

针对火灾检测中低光照背景烟雾难区分、火灾动态变化难以捕捉等难题, 本文通过对基准模型进行改进, 引入了 CSFE 增强特征提取模块、CG-Adown 模块和 MLA 模块, 从而给出了一种改进的火灾检测方法. 其中, CSFE 模块使用注意力分配功能引导模型更加聚焦于关键特征, 提高对重要特征的提取能力以提升火灾检测的性能. 并加入新的下采样模块 CG-

Adown 模块, 通过扩张卷积拓展感受野范围, 结合全局上下文引导策略, 增强模型远距离依赖捕捉能力以及全局信息感知能力, 有效避免在火灾检测过程中因信息缺失而导致的误判风险. 此外, MLA 模块结合基于特征变换的通道注意力与基于多尺度空间注意力机制, 使用多层次的注意力架构设计增强模型对关键特征的关注, 提高火灾检测的准确性和鲁棒性. 最后, 给出改进的 Inner-ElIoU 损失函数, 精准度量框差异、加速模型收敛且提升泛化能力. 实验结果表明,

相较于基准模型, 本文的改进模型在多个评价指标上取得了显著提升, 包括精确度、召回率和平均精确度等, 增强了模型在复杂多变火灾场景下的实用性和适应性。未来的研究可以从进一步优化模型结构、图像增强、多模态数据融合等方面入手, 以提高对更广泛场景的适应能力, 同时提升模型的计算效率和实时性能。

但在复杂多火源场景下, 模型尚存缺陷, 常混淆多个火源的烟雾与火焰, 致使火源定位失准, 干扰后续救援指挥, 可以考虑扩充多火源场景火灾图像数据并且注重火源位置的精准判定, 提升其在复杂火灾场景下的可靠性与实用性, 为高效救援指挥提供有力支撑。

参考文献

- [1] Avazov K, Mukhiddinov M, Makhmudov F, et al. Fire detection method in smart city environments using a deep-learning-based approach[J]. *Electronics*, 2021, 11(1): 73.
- [2] Nolan R H, Collins L, Leigh A, et al. Limits to post fire vegetation recovery under climate change[J]. *Plant, Cell & Environment*, 2021, 44(11): 3471-3489.
- [3] Gallagher R V, Allen S, Mackenzie B D E, et al. High fire frequency and the impact of the 2019–2020 megafires on Australian plant diversity[J]. *Diversity and Distributions*, 2021, 27(7): 1166-1179.
- [4] Choi W, Jung I Y. Multi-sensor photoelectric fire alarm device implementation for early fire detection in campsites[J]. *Applied Sciences*, 2024, 14(21): 9965.
- [5] Duane A, Castellnou M, Brotons L. Towards a comprehensive look at global drivers of novel extreme wild-fire events[J]. *Climatic Change*, 2021, 165(3): 43.
- [6] Jiao Z, Zhang Y, Xin J, et al. A deep learning based forest fire detection approach using UAV and YOLOv3 [C]//2019 1st International Conference on Industrial Artificial Intelligence (IAI). IEEE. 2019: 1-5.
- [7] Seydi S T, Saeidi V, Kalantar B, et al. Fire net: A deep learning framework for active forest fire detection [J]. *Journal of Sensors*, 2022, 2022(1): 8044390.
- [8] Morchid A, Oughannou Z, El Alami R, et al. Integrated internet of things (IoT) solutions for early fire detection in smart agriculture[J]. *Results in Engineering*, 2024, 24: 103392.
- [9] Majid S, Alenezi F, Masood S, et al. Attention based CNN model for fire detection and localization in real-world images[J]. *Expert Systems with Applications*, 2022, 189: 116114.
- [10] Ba R, Chen C, Yuan J, et al. SmokeNet: Satellite smoke scene detection using convolutional neural network with spatial and channel-wise attention[J]. *Remote Sensing*, 2019, 11(14): 1702.
- [11] Barmpoutis P, Stathaki T, Dimitropoulos K, et al. Early fire detection based on aerial 360-degree sensors, deep convolution neural networks and exploitation of fire dynamic textures[J]. *Remote Sensing*, 2020, 12(19): 3177.
- [12] Huang L, Liu G, Wang Y, et al. Fire detection in video surveillances using convolutional neural networks and wavelet transform[J]. *Engineering Applications of Artificial Intelligence*, 2022, 110: 104737.
- [13] Li P, Zhao W. Image fire detection algorithms based on convolutional neural networks[J]. *Case Studies in Thermal Engineering*, 2020, 19: 100625.
- [14] Yin X, Yu Z, Fei Z, et al. Pe-yolo: Pyramid enhancement network for dark object detection[C]//International Conference on Artificial Neural Networks. Cham: Springer Nature Switzerland, 2023: 163-174.
- [15] Talaat F M, ZainEldin H. An improved fire detection approach based on YOLO-v8 for smart cities[J]. *Neural Computing and Applications*, 2023, 35(28): 20939-20954.
- [16] Celik T. Fast and efficient method for fire detection using image processing[J]. *ETRI Journal*, 2010, 32(6): 881-890.
- [17] Xu R, Lin H, Lu K, et al. A forest fire detection system based on ensemble learning[J]. *Forests*, 2021, 12(2): 217.
- [18] 李欣健, 张大胜, 孙利雷, 等. 复杂场景下基于 CNN 的轻量火焰检测方法[J]. 模式识与人工智能, 2021, 34(5): 415-422.
- Li Xinjian, Zhang Dasheng, Sun Lilei, et al. Lightweight flame detection method based on CNN in com-

- plex scenes[J]. *Pattern Recognition and Artificial Intelligence*, 2021, 34(5): 415-422 (in Chinese).
- [19] 喻丽春, 刘金清. 基于改进 Mask R-CNN 的火焰图像识别算法[J]. *计算机工程与应用*, 2020, 56(21): 194-198.
Yu Lichun, Liu Jinqing. Flame image recognition algorithm based on improved Mask R-CNN[J]. *Computer Engineering and Application*, 2020, 56(21): 194-198 (in Chinese).
- [20] Xu G, Zhang Y, Zhang Q, et al. Video smoke detection based on deep saliency network[J]. *Fire Safety Journal*, 2019, 105: 277-285.
- [21] El-Madafri I, Peña M, Olmedo-Torre N. Real-time forest fire detection with lightweight CNN using hierarchical multi-task knowledge distillation[J]. *Fire*, 2024, 7(11): 392.
- [22] Wang C Y, Yeh I H, Mark Liao H Y. Yolov9: Learning what you want to learn using programmable gradient information[C]//*European Conference on Computer Vision*. Springer, Cham, 2025: 1-21.
- [23] Wu T, Tang S, Zhang R, et al. Cgnet: A light-weight context guided network for semantic segmentation[J]. *IEEE Transactions on Image Processing*, 2020, 30: 1169-1179.
- [24] Basha R, Pathak P, Sudha M, et al. Optimization of quantum dilated convolutional neural networks: Image recognition with quantum computing[J]. *Internet Technology Letters*, 2025, 8(3): e70027.
- [25] Jin X, Xie Y, Wei X S, et al. Delving deep into spatial pooling for squeeze-and-excitation networks[J]. *Pattern Recognition*, 2022, 121: 108159.
- [26] Zhu X, Cheng D, Zhang Z, et al. An empirical study of spatial attention mechanisms in deep networks[C]//*Proceedings of the IEEE/CVF International Conference on Computer Vision*. Seoul, Korea (South), 2019: 6688-6697.
- [27] Yang Z, Wang X, Li J. EIou: An improved vehicle detection algorithm based on vehiclenet neural network[J]. *Journal of Physics*, 2021, 1924(1): 012001.
- [28] Wang M, Xie X, Qiu H, et al. GEI-YOLOv9-based leather defect detection algorithm research[C]//*Fourth International Conference on Image Processing and Intelligent Control (IPIC 2024)*. SPIE, 2024, 13250: 571-576.
- [29] Huazhong University of Science Technology. Fire and Smoke Detection System Dataset[EB/OL]. Roboflow Universe, <https://universe.roboflow.com/huazhong-university-of-science-and-technology-ws6fw/fire-and-smoke-detection-system-ycbil>, 2023-12-01.
- [30] Yang Nan. Fire and Smoke Detection Dataset[EB/OL]. Roboflow Universe, <https://universe.roboflow.com/yang-nan/fire-and-smoke-detection-yqrpi>, 2024-03-01.
- [31] Padilla R, Netto S L, Da Silva E A B. A survey on performance metrics for object-detection algorithms[C]//*2020 International Conference on Systems, Signals and Image Processing (IWSSIP)*. IEEE, Niteroi, Brazil, 2020: 237-242.
- [32] Feng C, Zhong Y, Gao Y, et al. Tood: Task-aligned one-stage object detection[C]//*2021 IEEE/CVF International Conference on Computer Vision (ICCV)*. IEEE Computer Society, 2021: 3490-3499.
- [33] Zhang S, Chi C, Yao Y, et al. Bridging the gap between anchor-based and anchor-free detection via adaptive training sample selection[C]//*Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2020: 9759-9768.
- [34] Zhao Y, Lü W, Xu S, et al. Detsr beat yolos on real-time object detection[C]//*Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2024: 16965-16974.