

基于空 - 频域联合表征的YOLOv11改进模型及初期火灾烟雾检测应用

陈 漓, 崔 燕, 邓乾涛

湖南科技大学资源环境与安全工程学院, 湖南 湘潭

收稿日期: 2026年5月10日; 录用日期: 2026年6月3日; 发布日期: 2026年6月10日

摘 要

针对复杂环境下初期火灾烟雾小目标、动态性检测难题, 提出一种基于改进YOLOv11的高精度轻量化火灾烟雾检测算法。将空 - 频域聚合Mamba引入YOLOv11架构, 构建MixFreSBlock模块替代原C3K2模块; 通过MSPLCK模块在空域构建多尺度大感受野, 同时利用MFCA模块在频域实现噪声压缩与纹理强化, 二者拼接形成互补的空 - 频联合表征; 采用SS2D模块沿四维扫描路径对联合特征进行线性复杂度全局建模, 突破卷积局部性限制; 最后嵌入坐标注意力机制, 沿水平 - 垂直方向重校准特征, 精准定位火焰、烟雾细长结构。实验结果表明, 该模型在自建与开源数据集上mAP50达94.7%, 较原YOLOv11提升0.9%, 参数量仅3.1 M, 计算量仅8.0 G, 在精度与轻量化之间实现更优平衡。本研究为实时火灾预警系统提供了高性能、低计算成本的解决方案, 可广泛应用于安防监控、智慧城市等多场景, 具有重要的实际应用价值。

关键词

火灾烟雾检测, YOLOv11, 空 - 频域聚合网络, 注意力机制, 轻量化

YOLOv11 Improved Model Based on Joint Spatial-Frequency Domain Representation and Its Application in Early Fire Smoke Detection

Li Chen, Yan Cui, Qiantao Deng

School of Resources, Environment and Safety Engineering, Hunan University of Science and Technology, Xiangtan Hunan

Received: May 10, 2026; accepted: June 3, 2026; published: June 10, 2026

文章引用: 陈漓, 崔燕, 邓乾涛. 基于空-频域联合表征的YOLOv11改进模型及初期火灾烟雾检测应用[J]. 安防技术, 2026, 14(2): 57-72. DOI: 10.12677/jsst.2026.142006

Abstract

To address the challenges of detecting small and dynamic early fire smoke targets in complex environments, a high-precision lightweight fire smoke detection algorithm based on an improved YOLOv11 is proposed. The spatial-frequency domain aggregation Mamba is introduced into the YOLOv11 architecture, and the MixFreSBlock module is designed to replace the original C3K2 module: the MSPLCK module constructs a multi-scale large receptive field in the spatial domain, while the MFCA module achieves noise compression and texture enhancement in the frequency domain, and the two are concatenated to form complementary spatial-frequency joint representations; the SS2D module performs linear complexity global modelling of joint features along a four-dimensional scanning path, overcoming the locality limitation of convolution; finally, a coordinate attention mechanism is embedded to recalibrate features along horizontal and vertical directions, accurately locating flame and slender smoke structures. Experimental results show that the model achieves an mAP50 of 94.7% on both custom and open-source datasets, an increase of 0.9% compared to the original YOLOv11, with only 3.1 M parameters and 8.0 G computations, achieving a better balance between accuracy and lightweight design. This study provides a high-performance, low-computation solution for real-time fire warning systems and can be widely applied in security monitoring, smart cities, and other scenarios, with significant practical value.

Keywords

Fire Smoke Detection, YOLOv11, Spatial-Frequency Domain Aggregation Network, Attention Mechanism, Lightweight

Copyright © 2026 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

火灾作为全球范围内频发的重大公共安全事件，具有突发性强、扩散迅速、破坏性大等特点。火灾不仅对人类生命和财产安全构成严重威胁，还可能对环境和生态系统造成长期不可逆的影响[1]。有效识别火灾初期现象是火灾防控的关键核心环节，及时准确的初期火灾预警不仅能迅速启动灭火措施，更能为人员安全疏散和救援行动争取宝贵的黄金时间[2]。传统基于烟雾传感器或温度阈值的探测方法存在响应滞后、误报率高等问题。随着计算机视觉的发展，基于深度学习的火灾检测方法通过直接分析火焰与烟雾特征，展现出实时性强、适应性高的优势，已成为研究热点。然而，当前主流深度学习模型仍面临挑战：复杂环境下小目标和动态烟雾的检测能力不足，且难以兼顾精度与实时性。因此，研究高性能火灾检测算法，提升泛化能力与检测效率，对早期火灾防控具有重要的学术价值与应用意义。

目前，One-Stage 目标检测是一种新兴的目标检测框架，以 YOLO 系列算法为代表，能够同时进行物体位置回归和类别分类，实现端到端的处理，推理速度较快，更符合火灾实时预警的需求。随着深度学习技术的不断发展，众多学者针对火灾检测算法展开了深入研究。易冠霖等[3]提出一种改进 DeepLabv3+ 模型的语义分割算法，有效提升了船舶机舱初期火灾中小火焰及稀薄烟雾的检测精度。胡久松等[4]设计了一种融入轻量级网络和卷积块注意力机制的 YOLOv8 烟雾分类算法，显著提升了烟雾分类的精度与效率。李永福等[5]在 YOLOv7 中引入 EPSANet 与 SANet 注意力机制，增强了多尺度特征表达能力；高均益等[6]提出一种可部署到终端设备上的基于改进 YOLOv10 的新型火灾检测模型 YOLO-BFEPS，实现了对烟雾与火灾的快速准确检测。曲英伟等[7]将轻量级网络 MobileNetV3 引入 YOLOv5，并结合自适应非

极大值抑制(Adaptive-ElIoU-NMS)优化目标框筛选过程,显著降低了模型复杂度并提升了性能。Su 等[8]提出一种基于 YOLOv5s 网络的改进火焰和烟雾检测算法,引入全注意力网络(FLAnet)增强通道注意力,加入空间感知能力模块 CoordConv 提升模型识别效率,使 mAP@0.5 提升 3.5%。李敏学等[9]基于 YOLOv8 设计融合部分卷积与点态卷积的 FasterC2fBlock 模块的 FireLight-YOLO 轻量化架构,实现了轻量化条件下对森林火灾的精准检测。Phan 等[10]提出一种基于视觉的智能火灾和烟雾检测方法,将无参数注意模块 SimAM 和 ResNetGSimMix 模块集成到 YOLOv7 的主干和头部,添加额外的预测头提升了检测精度。

尽管现有的火焰或烟雾检测模型已取得阶段性进展,但当前火灾与烟雾检测研究领域仍面临若干核心挑战:

1) 环境干扰的复杂性:检测精度易受多变光照条件及复杂背景噪声的干扰,如强光、阴影、相似颜色物体等都会影响模型对烟雾和火焰的准确识别。

2) 烟雾表征的高度动态性与异质性:火灾初期产生的烟雾,其形态、尺度、色度及密度呈现显著的环境依赖性动态变化,极大地提升了检测模型的泛化难度。例如,在室内封闭空间和室外开阔空间,烟雾的扩散形态和特征差异巨大。

3) 对系统实时响应能力的严苛需求:火灾预警的时效性要求检测系统具备极高的处理速度,以支撑快速报警与应急处置,这就需要模型在保证检测精度的同时,具备较低的计算复杂度和较快的推理速度。

4) 小目标与动态背景下的鲁棒性瓶颈:对微小火焰或远距离烟雾的识别能力不足,以及在动态场景中维持检测稳定性,构成显著技术障碍。小目标烟雾的特征信息少,容易被背景噪声淹没,而动态背景会使模型难以区分目标和背景的变化。

5) 高质量数据集的稀缺性与局限性:精确的火焰与烟雾检测严重依赖高质量、多样化的标注数据集,但目前公开的火灾数据集规模相对较小,且场景覆盖不够全面,难以满足模型训练的需求。

为此,本文提出一种基于改进 YOLOv11 的火灾检测算法。核心创新是将空-频域聚合 Mamba 引入 YOLOv11,构建 MixFreSBlock 模块替代 C3K2。该模块并行级联 MSPLCK 与 MFCA,形成空-频联合表征;经 SS2D 全局建模后,嵌入坐标注意力沿水平-垂直方向重校准特征,精准捕捉细长烟焰结构。该方案在保持实时性的同时,显著提升了对复杂烟焰目标的判别能力。

2. YOLOv11 算法与改进

2.1. YOLOv11 算法

YOLO (You Only Look Once) [11]-[15]系列算法作为经典的单阶段(One-stage)目标检测算法,它将图像划分为网格单元,并同时预测每个单元中目标的边界框和类别,实现了目标检测的实时高效和多尺度检测,使其在快速目标检测和实时应用场景中表现良好,目前在各个行业都有广泛的应用场景。

YOLOv11 [16]是 Ultralytics 公司于 2024 年 9 月底推出的模型,采用单阶段检测架构,整体流程为主干网络→颈部网络→检测头。主干网络基于 CSPDarknet 改进,引入 C3k2 模块替代 YOLOv8 中的 C2f 模块,计算效率更高;颈部采用 PANet 构建双向特征金字塔,实现深浅层特征融合;检测头采用解耦头设计,支持 Anchor-Free 预测。图 1 展示了 YOLOv11 的网络结构图。

2.2. 改进 YOLOv11 算法

本文在传统 YOLOv11 算法的基础上进行了改进,改进后的网络结构如图 2 所示。本工作针对 C3K2 网络进行改进,引入并集成了一个空-频域聚合 mamba 网络,将空-频域聚合 Mamba [17]引入 YOLOv11,提出以 MSPLCK 与 MFCA 并行级联的空-频耦合模块 MixFreSBlock 替代 C2K3。首先 MSPLCK 在空域提供多尺度大感受野,其次 MFCA 在频域压缩噪声并强化纹理,二者拼接后形成互补的空-频联合表征;

随后 SS2D 沿四维扫描路径对联合特征进行线性复杂度全局建模，弥补卷积局部性缺陷；最终嵌入坐标注意力，沿水平 - 垂直方向重校准，精确定位火焰、烟雾细长结构。

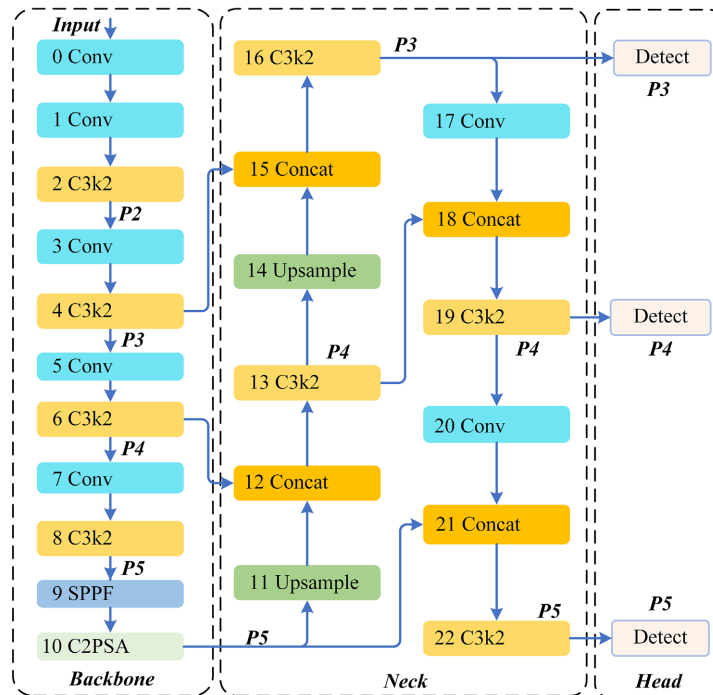


Figure 1. YOLOv11 network architecture diagram

图 1. YOLOv11 网络结构图

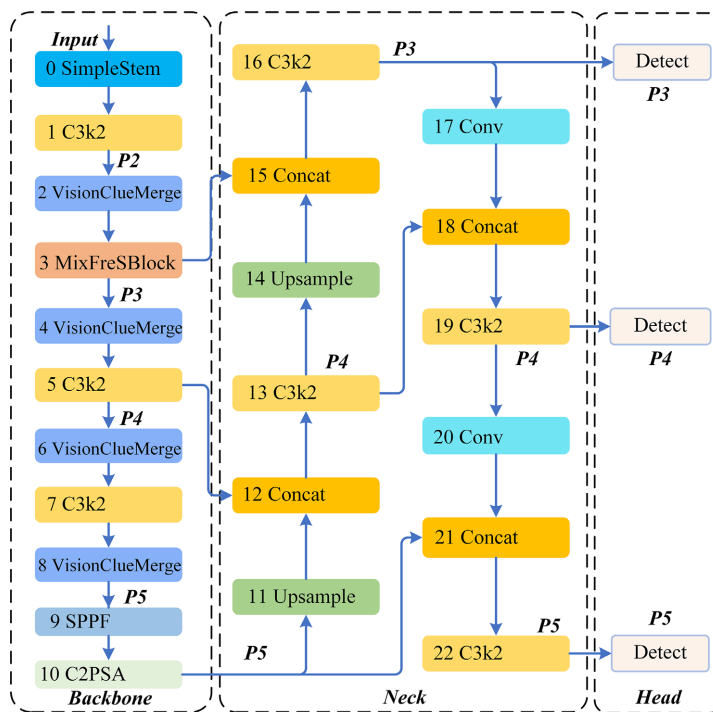


Figure 2. Improved YOLOv11 network architecture diagram

图 2. 改进 YOLOv11 网络结构图

该设计在保持实时性的同时，显著增强对不规则、半透明、多尺度烟焰目标的特征判别力，实现检测精度与效率的新平衡。下图 3 为 MixFreSBlock 混合模块结构图。

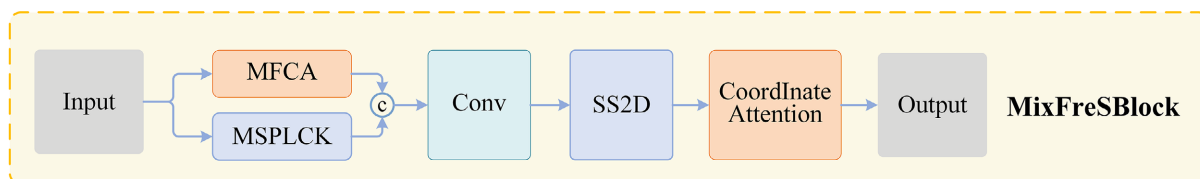


Figure 3. Structure diagram of MixFreSBlock hybrid module

图 3. MixFreSBlock 混合模块结构图

2.3. 主干网络下采样

为优化 YOLOv11 在复杂场景下的特征提取效率与多尺度感知能力，本文对其主干网络进行了重构。首先通过 Simple Stem 模块进行下采样，得到一个分辨率为 $H/4$ 、 $W/4$ 的二维特征图。随后，基础特征与 Vision Clue Merge 模块和 C3K2 模块交替堆叠替换传统卷积进行下一步下采样。

Simple Stem 模块：现代视觉 Transformer 模型通常采用分块图像作为其初始模块，通过核大小为 4、步长为 4 的卷积操作将图像划分为非重叠区域。然而，近期研究(如 EfficientFormerV2 [18])指出，此方法可能限制 ViT 的优化潜力，从而影响整体性能。为平衡性能与效率，Simple Stem 模块采用了一种简化的主干层设计：摒弃非重叠分块方式，转而使用两个步长为 2、核大小为 3 的卷积层。下图 4 为 Simple Stem 模块结构图。

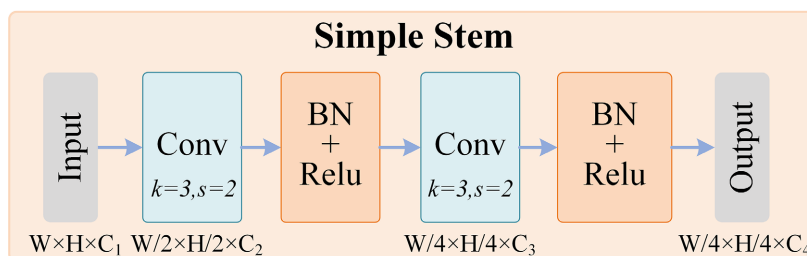


Figure 4. Structure diagram of the Simple Stem module

图 4. Simple Stem 模块结构图

Vision Clue Merge 模块：尽管卷积神经网络与视觉 Transformer 结构普遍采用卷积操作进行下采样，我们发现该方法会干扰 SS2D 在不同信息流阶段的选择性操作[19]。为解决此问题，Vision Clue Merge 模块采取了截然不同的思路，它不进行空间上的信息混合，而是通过维度的拆分与重组来实现分辨率的减半，并保留全部信息，此网络对二维特征图进行分割，并利用 1×1 卷积降低维度。结果表明，为空-频域聚合模型保留更多视觉线索有助于提升模型训练效果。与传统的维度减半方法不同，Vision Clue Merge 模块通过以下方式简化此流程：

- 1) 移除归一化层；
- 2) 分割维度特征图；
- 3) 将冗余特征图附加至通道维度；
- 4) 将用 4 倍压缩的逐点卷积进行下采样。

相较于采用步长为 2 的 3×3 卷积进行下采样的传统方法，此方法能够完整保留前一层所选取的特征图信息，从而避免因常规卷积操作导致的特征选择性损失。下图 5 为 Vision Clue Merge 模块结构图。

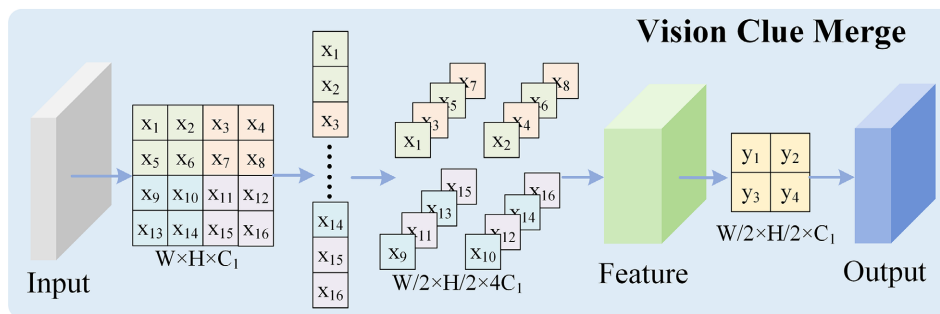


Figure 5. Structure diagram of the Vision Clue Merge module
图 5. Vision Clue Merge 模块结构图

2.4. MSPLCK

多尺度并行大卷积核心模块 MSPLCK [20]既具有多尺度特性，又具有较大的接收范围，它可以同时恢复纹理细节，同时捕获大片雾霾区域，并行膨胀卷积还具有较大的接收范围和远距离模拟能力，核模块用于获得单变压器[21] [22]多头自关注模块所不具备的图像的多尺度特征。MSPLCK 模块兼具多尺度特性与大感受野。

首先，设 x 为原始特征图，通过批归一化对其进行归一化： $x = \text{BatchNorm}(x)$ 。此处，PWConv 指逐点卷积(1×1 卷积)。Conv 指卷积核大小为 5 的标准卷积。DWConv19 指空洞卷积核大小为 19，具体为膨胀率为 3 的 7×7 深度可分离空洞卷积；DWConv13 指空洞卷积核大小为 13，具体为膨胀率为 3 的 5×5 深度可分离空洞卷积；DWConv7 指空洞卷积核大小为 7，具体为膨胀率为 3 的 3×3 深度可分离空洞卷积。最后，Concat 指沿通道维度进行特征拼接。三条并行且具有不同卷积核大小的空洞卷积支路能够提取多尺度特征。其中，大尺寸和中尺寸的空洞卷积具备长距离建模能力与广阔的感受野，其作用类似于 Transformer 中的自注意力机制，能够聚焦于大范围的烟尘区域；而小尺寸的空洞卷积则能专注于小范围烟尘区域并恢复纹理细节。将从通道维度拼接来自多尺度的信息，使得特征图 x_3 的通道维度数变为输入 x 的三倍。

随后，模块将特征图 x_3 输入一个多层感知机，将其特征维度转换至与原始输入 x 相同。该多层感知机包含两次逐点卷积，并使用 GELU 函数作为激活函数。最后，将多层感知机的输出与恒等捷径连接 x 相加。下图 6 为 MSPLCK 模块结构图。

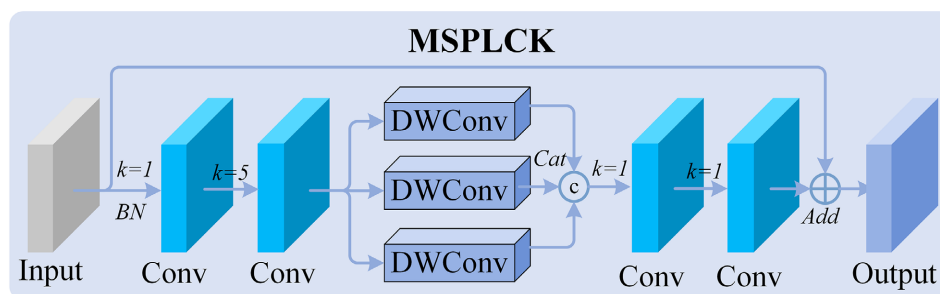


Figure 6. Structure diagram of the MSPLCK module
图 6. MSPLCK 模块结构图

2.5. MFCA

本文空 - 频耦合模块 MixFreSBlock 使用具有 2D 离散余弦变换(2D DCT)的多频率通道注意力(MFCA)

[23]通过提取频率统计来产生通道注意力图，擅长从特征图中提取可区分的特征，在频域压缩噪声并强化纹理。与此同时，整合多频率技术和注意机制的研究工作激增，旨在通过实施各种频率转换方法来增强从精细到粗略信息的局部和全局背景提取，尤其是 2D DCT 由于其压缩能力，被广泛应用于计算机视觉中以提取频率统计特征，以提高表示能力。

近年来，二维离散余弦变换(2D DCT)的压缩能力因其能够将图像表达为不同频率余弦函数振荡所产生的基图像的加权和，在频域特征提取领域受到广泛关注[24]-[26]。利用基本图像 D ，各尺度分支的特征可通过二维离散余弦变换表征如下：

$$X_i^{S,k} = \sum_{h=0}^{H_S-1} \sum_{w=0}^{W_S-1} (X_i^S)_{h,w} D_{h,w}^{u_k, v_k},$$

其中 (u_k, v_k) 是对应于 $X_i^{S,k}$ 的频率索引。此外，第 S 个尺度分支上的 2D DCT 基本图像定义为

$$D_{h,w}^{u_k, v_k} = \cos\left(\frac{\pi h}{H_S}\left(u_k + \frac{1}{2}\right)\right) \cos\left(\frac{\pi w}{W_S}\left(v_k + \frac{1}{2}\right)\right)$$

并采用了 top-K 项选择策略[25]。随后，通过全局平均池化、全局最大池化和全局最小池化，将每个 $X_i^{S,k}$ 压缩为 Z_{avg} ， Z_{max} 和 Z_{min} 。接着，通过两个全连接层 $W_1 \in R^{C_s \times \frac{C_s}{r}}$ 和 $W_2 \in R^{\frac{C_s}{r} \times C_s}$ (其中 r 为缩减比)，聚合每个频率的统计量，以生成第 s 个尺度分支上的通道注意力图，具体过程如下：

$$M_i^S = \sigma\left(\sum_{d \in \{\text{avg, max, min}\}} W_2(\delta(W_1 Z_d))\right) \in R^{C_s}$$

其中， δ 与 σ 分别表示 ReLU 与 Sigmoid 激活函数。最终，我们在第 S 个尺度分支上利用 M_i^S 对特征图 X_i^S 进行重新校准，即： $\hat{X}_i^S = X_i^S \times M_i^S$ 。下图 7 为 MFCA 模块结构图。

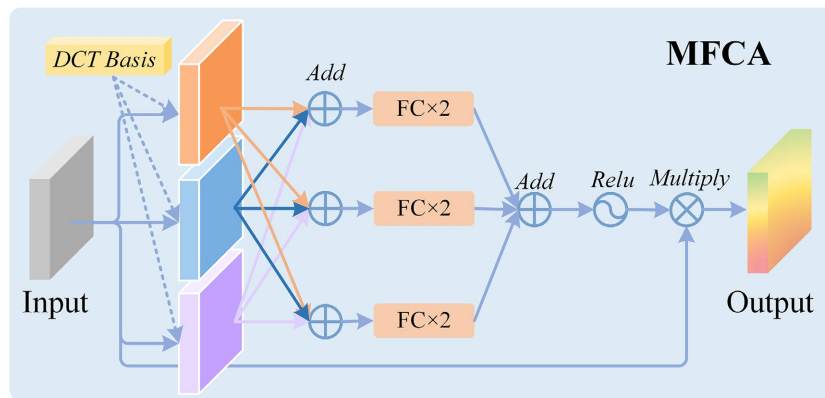


Figure 7. Structure diagram of the MFCA module
图 7. MFCA 模块结构图

2.6. SS2D

Scan Expansion (扫描扩展)、S6 Block 和 Scan Merge (扫描合并)是 SS2D 算法的三个主要步骤，其主要流程如图 8 所示。扫描扩展操作将输入图像扩展为一系列子图像，每一幅子图像代表一个特定方向。当从对角线视角观察时，扫描扩展操作沿着四个对称方向进行，分别是自上而下、自下而上、从左到右以及从右到左。这样的布局不仅全面覆盖了输入图像的所有区域，还通过系统性的方向变换为后续特征提取提供了丰富的多维信息基础，从而提升了图像特征多维度捕获的效率和全面性。SS2D 算法中的扫描

合并操作将以获得的序列作为 S6 模块[27]的输入，并将来自不同方向的序列进行合并，从而将特征提取并整合为全局特征。

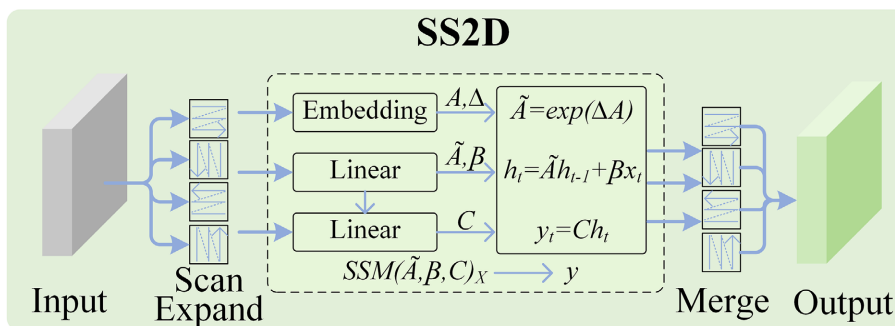


Figure 8. Structure diagram of the SS2D module
图 8. SS2D 模块结构图

2.7. 坐标注意力

为了增强神经网络的数据处理能力，注意力机制被广泛研究，如全局注意力机制(Global Attention Mechanism, GAM) [28]、自注意力机制(Self Attention, SA) [29]。本文在提出的耦合模块中引入一种坐标注意力机制(Coordinate Attention, CA) [30]。该机制旨在将精确的位置信息嵌入到通道注意力之中，以指导模型更有效地“关注”目标对象。坐标注意力模块可被视为一种计算单元，其目标在于增强移动网络所学习特征的表达能力，该模块能够接收任意中间特征张量 $X = [x_1, x_2, \dots, x_C] \in R^{C \times H \times W}$ 作为输入，并输出一个经过增强表征的变换张量 $Y = [y_1, y_2, \dots, y_C]$ ，其尺寸与输入张量 X 相同。其设计具备以下优势：首先，它同时捕获了跨通道关系以及方向感知的空间信息，有助于模型的精确定位；其次，该机制设计简洁、计算开销极小，可以灵活集成到各类经典移动网络模块(如 MobileNetV2 [31]的倒残差块)中；最后，作为预训练模型的一部分，它能显著提升下游任务(特别是目标检测、语义分割等密集预测任务)的性能。大量实验表明，坐标注意力不仅能提升 ImageNet 分类精度，更能有效推动移动网络在复杂视觉任务中的表现。下图 9 为坐标注意力模块结构图。

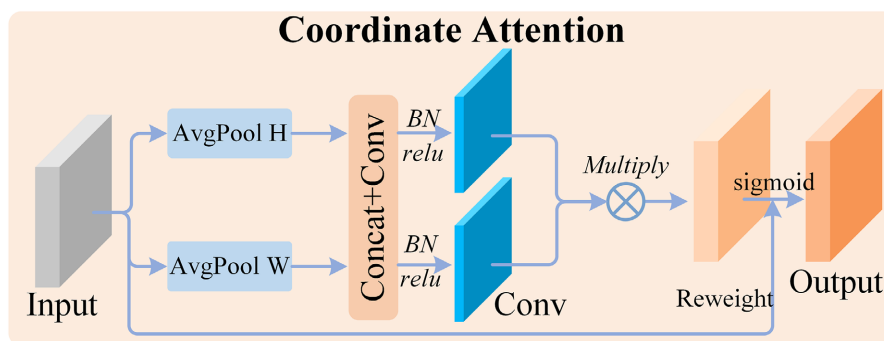


Figure 9. Structure diagram of the coordinate attention module
图 9. 坐标注意力模块结构图

3. 实验准备与评价指标

3.1. 实验环境

本实验在 Windows 11 专业版操作系统下进行，采用 Python 3.10 作为编程语言，PyTorch 2.5.1 作为

深度学习框架。具体实验环境与配置参数详见表 1。

Table 1. Experimental environment
表 1. 实验环境

环境项	Value
操作系统	Windows11 专业版
编程语言	Python3.10
深度学习框架	Pytorch2.5.1
CPU	13th Gen Intel(R) Core(TM) i5-13600KF 3.50GHz
GPU	NVIDIA GeForce RTX 4060
内存	32GB

在模型训练过程中，设置其特定的参数和超参数以确保模型训练最佳性能，具体参数设置见表 2。

Table 2. Training hyperparameters
表 2. 训练超参数

参数	Value
Learning rate	0.01
Image size	640 × 640
Momentum	0.937
Optimizer	SGD
Batch size	64
Epoch	150
Weight decay	0.0005

3.2. 数据集构建

本研究从中国科学技术大学火灾科学国家实验室的数据集中精心挑选了 6500 张不同环境下的初期烟雾火灾图像作为实验数据集，涵盖室内、室外、森林等场景，包含小火、轻烟等典型初期火灾烟雾特征的图像。为确保数据集规范化，对所有图像进行了统一格式化命名，并逐幅标注烟雾和火焰目标。此外，为避免模型在训练过程中出现过拟合或欠拟合现象，将收集的数据按照 6:2:2 的比例划分为训练集、验证集和测试集，并统一缩放为 640 × 640 的分辨率。部分场景的烟雾火焰图像如图 10 所示。

3.3. 评价指标

在火灾与烟雾检测算法的性能评估中，选择合适的评价指标对于全面、客观地衡量算法效能至关重要。为综合评估模型在多方面的表现，本研究选取了以下关键指标：

模型复杂度与效率：采用参数量(Params)反映模型的规模与复杂程度，十亿浮点运算数(GFLOPs)衡量计算效率。参数量越少，模型越轻量化，GFLOPs 越低，计算效率越高。

检测精度：使用精确度(Precision)与召回率(Recall)分别评估算法识别正类样本的准确性与完整性。其计算依赖于真正例(TP)、假正例(FP)和假负例(FN)等基本判别结果，计算公式如下：Precision、Recall 的表达式如公式(1)和公式(2)所示：

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \quad (1)$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \quad (2)$$

综合性能：引入 mAP@0.5 (即 mAP50)，即交并比(IoU)阈值为 0.5 时的平均精度均值(AP)，以及 mAP@0.5:0.95 (即 mAP95)，即在 IoU 阈值从 0.5 至 0.95 (步长 0.05)范围内计算的平均 mAP 值。二者共同反映模型在不同定位精度要求下的综合识别能力，计算公式如公式(3)和公式(4)所示。

$$\text{AP} = \int_0^1 P(r) dr \quad (3)$$

$$\text{mAP} = \frac{1}{n} \sum_{i=1}^n \text{AP}_i \quad (4)$$

通过这些指标的系统结合，能够从计算效率、模型轻量化与检测准确性等多个维度对火灾与烟雾检测算法进行全面而客观的评价。

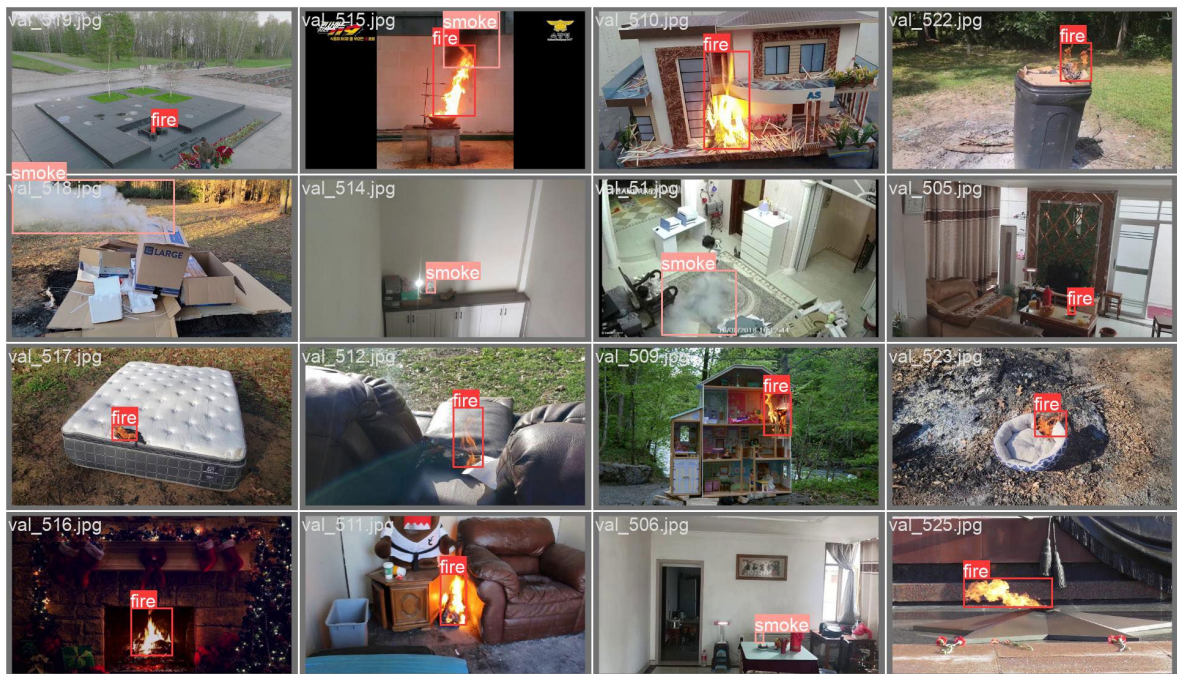


Figure 10. Partial dataset

图 10. 部分数据集

4. 改进前后算法模型识别结果分析

4.1. 损失函数分析

在实验过程中，改进 YOLOv11 模型在火灾烟雾数据集上进行了 150 轮训练，其训练集和验证集的性能指标如图 11 所示。

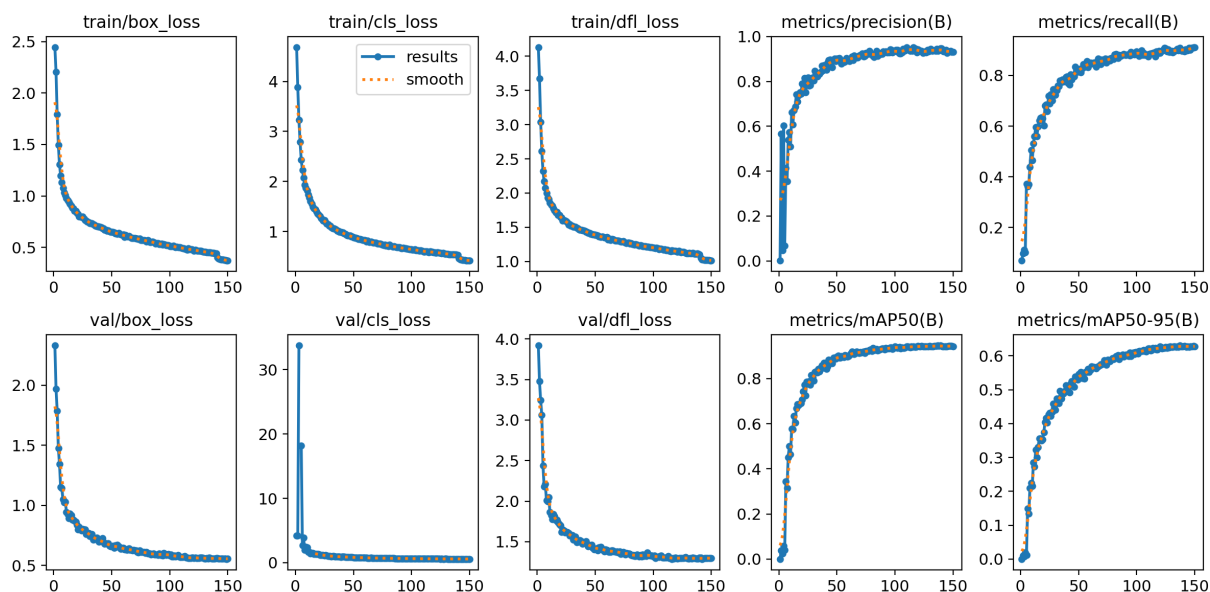


Figure 11. Model training evaluation metric curves

图 11. 模型训练评价指标曲线

图 11 中，前 3 列图表展示了训练集与验证集损失随训练轮次的变化情况。横轴为训练轮次，纵轴为损失值。由图可知，两条曲线均呈现明显的收敛趋势：训练损失在初期迅速下降，显示模型对数据特征的高效学习；验证损失随后也逐渐降低并趋于平稳，表明模型泛化能力持续增强。实验结果表明，改进 YOLOv11 模型在 150 轮训练中展现出优秀的收敛性能，兼顾了高准确率与正样本识别能力，验证了所提轻量化架构的训练稳定性与有效性。

接下来，AP 值能全面地衡量准确率与召回率之间的关系，实际上，AP 值就是 PR 曲线与坐标轴所围成的面积，面积越大代表 AP 值越大，即模型对此类物体的检测精度越高。绘制了模型的精度 - 召回率曲线(简称 P-R 曲线)，如图 12 所示。

由图 12 可知，改进 YOLOv11 模型在数据集上取得了 94.7% 的 mAP50。“fire”类别曲线 AP 为 0.958，“smoke”类别曲线 AP 为 0.935，P-R 曲线整体更贴近右上角，说明模型在不同阈值下均具备稳健的检测性能。该曲线形态反映了模型在正负样本判别中的良好平衡，有效降低误检，增强实际应用可靠性。综合训练与评估曲线可见，改进 YOLOv11 在保持轻量化的同时，检测精度与稳定性均有显著提升。

4.2. 评价指标分析

1) 对比试验。为更全面地验证改进后的模型在目标检测任务中的有效性，本研究与当前主流的目标检测模型在相同实验条件下进行对比实验，评估各模型在精度、计算资源消耗以及实时性等性能指标上的表现，以确定改进模型在实际应用中的优势。选取了 6 种主流的 YOLO 和 RT-DETR 目标检测模型作为对比算法，包括 YOLOv5、YOLOv8、YOLOv9、YOLOv11、RT-DETR-R18 和 RT-DETR-R50。通过构建多模型对照实验体系，能够从多个维度客观评估所提架构的综合性能优势。对比试验结果见表 3。

由表 3 可知，不同算法在各项指标上呈现出明显的性能差异和特点，但本研究提出的改进 YOLOv11 算法在检测性能、推理效率及模型复杂度等多个关键性能指标上表现优异：

计算复杂度：改进 YOLOv11 算法展现出显著优势，其 GFLOPs 仅为 8.0 G，Params 仅为 3.1 M，表明其具备优异的计算轻量化特性。YOLOv5 与 YOLOv11 分别以 7.1 和 7.6 的 GFLOPs 位列其次，而 RT-

DETR-R50 的计算负担最重，达 134.8 G。在参数量上，改进 YOLOv11 算法的模型参数量为 3.1 M，远低于其他对比算法，成功地用相对复杂的结构实现了极低的参数量和计算量。

检测性能：改进 YOLOv11 算法不仅保持了极低的计算与存储开销，还具备优异的检测精度、召回率、mAP50 和 mAP95。较 YOLO 和 RT-DETR 系列模型均有所提高，显示出该算法在精度 - 效率平衡方面的优势。相较基准 YOLOv11，mAP50 提升 0.9%，mAP95 提升 0.9%，Recall 提升 1.6%，而参数量仅增加 0.2 M，计算量增加 0.3 G；与 YOLOv9 相比，mAP50 仍高 0.6%，但参数量压缩 87%，计算量压缩 92%。

综上可知，所提模块在火焰 - 烟雾这类半透明、多尺度、细长不规则目标上，能够以极低额外成本显著抑制虚警与漏检，验证了空 - 频耦合表征与线性复杂度全局建模的有效性，为轻量级火灾实时检测提供了新的 state-of-the-art 解决方案。

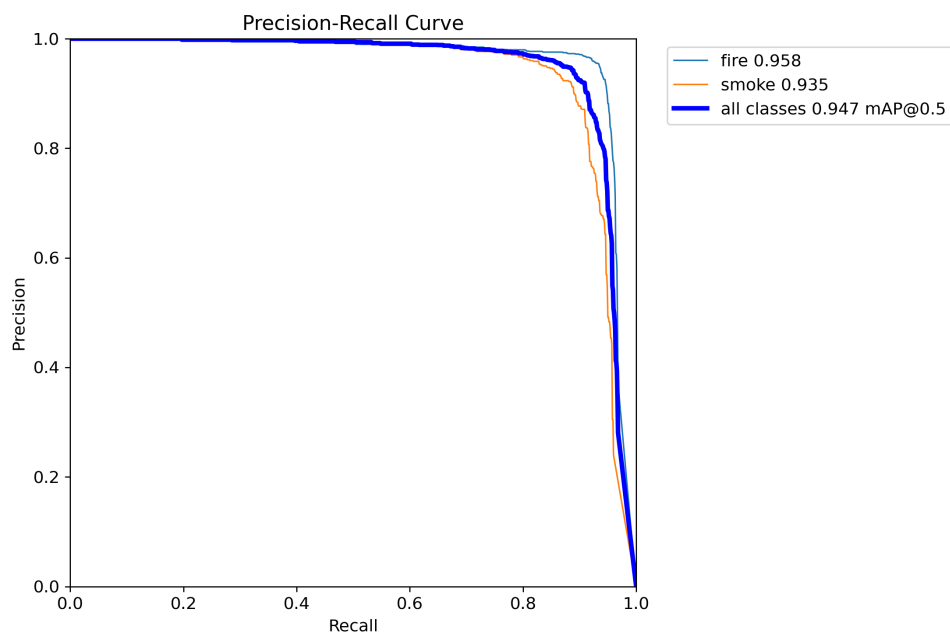


Figure 12. PR curve

图 12. 精度 - 召回率曲线

Table 3. Comparative experiment

表 3. 对比实验

	Precise	Recall	mAP50	mAP95	Param	GFLOPs
YOLOv5	0.931	0.894	0.937	0.614	2.5	7.1
YOLOv8	0.944	0.877	0.934	0.618	3.0	8.1
YOLOv9	0.949	0.907	0.941	0.625	25.3	102.3
YOLOv11	0.929	0.896	0.938	0.621	2.9	7.6
RT-DETR-R18	0.914	0.902	0.931	0.602	19.8	57.0
RT-DETR-R50	0.921	0.905	0.936	0.617	41.2	134.8
改进 YOLOv11	0.935	0.910	0.947	0.636	3.1	8.0

2) 消融实验。为评估 YOLOv11 模型改进的耦合模块 MixFreSBlock 的有效性,本研究设计了消融实验,逐一验证了耦合模块代替 C3K2 不同位置对模型性能的影响。通过对比不同位置下的实验结果,可以更直观地了解改进位置对模型性能的贡献。每项实验均保持其他结构与训练条件一致,并在相同测试集上评估性能。具体消融实验模型评价指标见表 4。

Table 4. Ablation experiments at different positions
表 4. 不同位置消融实验

	Precise	Recall	mAP50	mAP95
YOLOv11	0.929	0.896	0.938	0.621
YOLOv11_g1	0.930	0.893	0.937	0.622
YOLOv11_g2	0.935	0.910	0.947	0.636
YOLOv11_g3	0.936	0.892	0.938	0.634
YOLOv11_g4	0.941	0.89	0.939	0.625

以 YOLOv11 原始模型为基准(精度 0.929, 召回率 0.896, mAP50 为 0.938, mAP95 为 0.621),结合图 1、图 2,逐一验证了耦合模块 MixFreSBlock 代替 C3K2 不同位置对模型性能的影响。具体而言,将其分别置于骨干网络的四个关键位置,设计了四次对照实验,以系统性地评估模块位置变化对模型性能的影响。以下是四次实验及其结果的逐步分析:

YOLOv11_g1: 替换靠近输入端的第一个 C3K2 模块,模型在初始特征提取阶段细节信息增强,小目标检测 Precision 小幅提升约 0.1%,但整体 mAP 轻微下降约 0.1%。该位置替换增强了浅层特征的细节保留能力,但由于过早引入复杂融合,可能干扰了高层语义特征的梯度传播与形成。

YOLOv11_g2: 替换中间层的第二个 C3K2 模块(即本文图 2 中耦合模块 MixFreSBlock 代替 C3K2 的位置),小目标检测 Precision 提升 0.6%, mAP 提升 0.9%,特征融合效果最为显著。此处为多层次特征交汇的关键节点, MixFreSBlock 能同时融合浅层细节与深层语义,其跨频信息整合优势得到充分发挥,实现了鲁棒性与准确性的最佳平衡。

YOLOv11_g3: 替换深层网络的倒数第二个 C3K2 模块,小目标检测 Precision 提升 0.7%,但 mAP 未见显著变化。深层网络侧重于高级语义与全局上下文,模块在此处虽强化了上下文建模,但对多尺度特征融合的贡献有限,且引入了一定的计算开销。

YOLOv11_g4: 替换紧邻输出端的最后一个 C3K2 模块,模型平均精度 mAP 提升不明显,仅 0.1%,且因输出前特征重构过度,导致部分类别误检率上升。输出前的位置对特征分布极为敏感, MixFreSBlock 模块的复杂操作易与检测头产生兼容性问题,反而扰动已优化的判别特征。

为彰显改进 YOLOv11 的强大性能,在图 13 中展示了与图 10 对应示例图像的火灾烟雾检测图。

由图 13 可知,在多种复杂环境下,改进 YOLOv11 模型能够有效检测出多种场景下的火灾目标。整体检测效果优异,展现出该算法在复杂背景下的良好适应性与鲁棒性。此外,该模型能够较好地应对火灾不同发展阶段的尺度变化与形态多样性,体现出较强的泛化能力。上述实验结果不仅验证了改进 YOLOv11 算法的技术优势,也为其在更多复杂场景中的实际应用提供了有力支持。

5. 结论

1) 本文将空-频域聚合 Mamba 引入 YOLOv11 骨干网络,构建了 MixFreSBlock 模块,通过并行级

联空域多尺度感受野增强模块(MSPLCK)与频域降噪纹理强化模块(MFCA), 实现了空间细节与频率特征的互补性联合表征。随后, 利用 SS2D 进行线性复杂度的全局上下文建模, 并结合坐标注意力机制, 有效增强了模型对烟雾、火焰等细长、半透明目标的定位与判别能力。

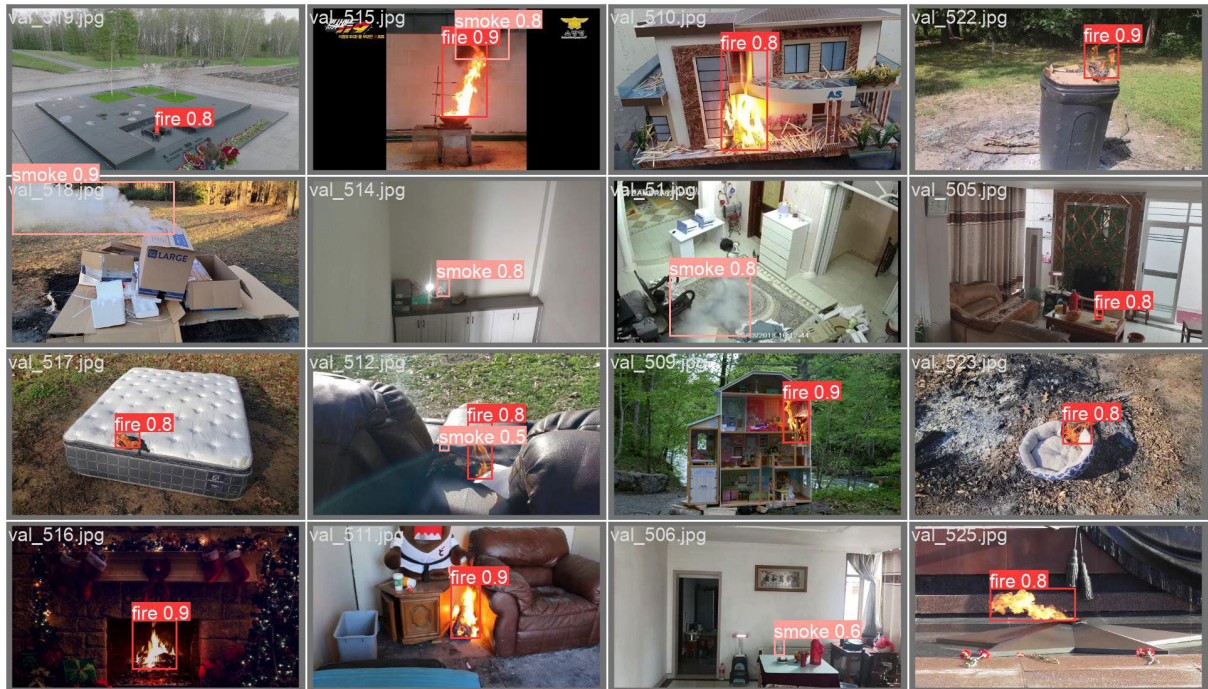


Figure 13. Fire and smoke detection diagram
图 13. 火灾烟雾检测图

2) 实验结果表明, 所提模型在综合性能上取得了显著提升。在包含多场景、多尺度目标的火灾数据集上, 其 mAP@0.5 达到 94.7%, 较基准 YOLOv11 提升 0.9%, 同时召回率提升 1.6%。在实现精度提升的同时, 模型保持了优异的轻量化特性, 参数量仅为 3.1 M, FLOPs 为 8.0 G, 计算效率显著优于 YOLOv9 等大型模型。消融实验进一步验证了 MixFreSBlock 模块置于网络中间层时对特征融合与性能提升的关键作用。

3) 接下来的研究工作将主要从两方面推进: 一是加强数据质量, 通过增补特定场景下的小尺寸火灾烟雾图像, 丰富样本多样性, 从而减少误判, 增强对小目标火灾烟雾的识别能力; 二是对模型结构进行优化, 降低计算复杂度与推理耗时, 进一步提升检测效率。

参考文献

- [1] Gajendiran, K., Kandasamy, S. and Narayanan, M. (2024) Influences of Wildfire on the Forest Ecosystem and Climate Change: A Comprehensive Study. *Environmental Research*, **240**, Article 117537. <https://doi.org/10.1016/j.envres.2023.117537>
- [2] Lian, J., Pan, X. and Guo, J. (2023) An Improved Fire and Smoke Detection Method Based on YOLOv7. *2023 32nd International Conference on Computer Communications and Networks (ICCCN)*, Honolulu, 24-27 July 2023, 1-7. <https://doi.org/10.1109/icccn58024.2023.10230135>
- [3] 易冠霖, 吴浩峻, 吴韵哲, 等. 基于语义分割的船舶机舱初期火灾识别算法[J]. *船海工程*, 2023, 52(5): 109-114.
- [4] 胡久松, 刘张驰, 余谦, 等. 融入 GhostNet 和 CBAM 的 YOLOv8 烟雾识别算法[J]. *电子测量与仪器学报*, 2024, 38(8): 201-207.
- [5] 李永福, 陈立斌, 惠君伟, 等. 基于 EPSA-YOLOv5 电力高空作业安全带佩戴检测[J]. *西安工程大学学报*, 2024,

- 38(2): 18-25.
- [6] 高均益, 张伟, 李泽麟. YOLO-BFEPS: 一种高效注意力增强的跨尺度 YOLOv10 火灾检测模型[J]. 计算机科学, 2025, 52(S1): 424-432.
- [7] 曲英伟, 刘锐. 基于 YOLOv5-MobileNetV3 算法的目标检测[J]. 计算机系统应用, 2024, 33(6): 213-221.
- [8] Su, L., Zhang, S. and Ding, W. (2023) An Improved Real-Time Detection Method for Flame and Smoke Identification Based on YOLOv5. 2023 6th International Conference on Intelligent Autonomous Systems (ICoIAS), Qinhuangdao, 22-24 September 2023, 59-64. <https://doi.org/10.1109/icoias61634.2023.00017>
- [9] 李敏学, 张晓宇, 程英杰, 等. FireLight-YOLO:面向森林火灾实时监测的轻量化模型[J]. 北京林业大学学报, 2026, 48(1): 12-25.
- [10] Phan, D.T., Yap, K.H., Garg, K. and Han, B.S. (2023) Vision-Based Early Fire and Smoke Detection for Smart Factory Applications Using FFS-YOLO. 2023 IEEE 25th International Workshop on Multimedia Signal Processing (MMSP), Poitiers, 27-29 September 2023, 1-6. <https://doi.org/10.1109/mmisp59012.2023.10337640>
- [11] Redmon, J. and Farhadi, A. (2018) YOLOv3: An Incremental Improvement. arXiv: 1804.02767.
- [12] Li, C., Li, L., Jiang, H., et al. (2022) YOLOv6: A Single-Stage Object Detection Framework for Industrial Applications. arXiv: 2209.02976.
- [13] Wang, C., Bochkovskiy, A. and Liao, H.M. (2023) YOLOv7: Trainable Bag-of-Freebies Sets New State-of-the-Art for Real-Time Object Detectors. 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Vancouver, 17-24 June 2023, 7464-7475. <https://doi.org/10.1109/cvpr52729.2023.00721>
- [14] Bochkovskiy, A., Wang, C.Y. and Liao, H.Y.M. (2020) YOLOv4: Optimal Speed and Accuracy of Object Detection. arXiv: 2004.10934.
- [15] Wang, C.Y., Yeh, I.H. and Mark Liao, H. (2024) YOLOv9: Learning What You Want to Learn Using Programmable Gradient Information. In: Leonardis, A., Ricci, E., Roth, S., Russakovsky, O., Sattler, T. and Varol, G., Eds., *Lecture Notes in Computer Science*, Springer, 1-21. https://doi.org/10.1007/978-3-031-72751-1_1
- [16] Khanam, R. and Hussain, M. (2024) YOLOv11: An Overview of the Key Architectural Enhancements. arXiv: 2410.17725.
- [17] Wang, Z., Li, C., Xu, H., Zhu, X. and Li, H. (2025) Mamba YOLO: A Simple Baseline for Object Detection with State Space Model. *Proceedings of the AAAI Conference on Artificial Intelligence*, **39**, 8205-8213. <https://doi.org/10.1609/aaai.v39i8.32885>
- [18] Li, Y., Hu, J., Wen, Y., Evangelidis, G., Salahi, K., Wang, Y., et al. (2023) Rethinking Vision Transformers for Mobilenet Size and Speed. 2023 IEEE/CVF International Conference on Computer Vision (ICCV), Paris, 1-6 October 2023, 16889-16900. <https://doi.org/10.1109/iccv51070.2023.01549>
- [19] Jiao, J., Liu, Y., Liu, Y., Tian, Y., Wang, Y., Xie, L., et al. (2024) VMamba: Visual State Space Model. *Advances in Neural Information Processing Systems*, **37**, 103031-103063. <https://doi.org/10.52202/079017-3273>
- [20] Lu, L.P., Xiong, Q., Xu, B. and Chu, D. (2024) MixDehazeNet: Mix Structure Block for Image Dehazing Network. 2024 International Joint Conference on Neural Networks (IJCNN), Yokohama, 30 June 2024-5 July 2024, 1-10. <https://doi.org/10.1109/ijcnn60899.2024.10651326>
- [21] Dosovitskiy, A. (2020) An Image Is Worth 16x16 Words: Transformers for Image Recognition at Scale. arXiv: 2010.11929.
- [22] Liu, Z., Lin, Y., Cao, Y., Hu, H., Wei, Y., Zhang, Z., et al. (2021) Swin Transformer: Hierarchical Vision Transformer Using Shifted Windows. 2021 IEEE/CVF International Conference on Computer Vision (ICCV), Montreal, 10-17 October 2021, 10012-10022. <https://doi.org/10.1109/iccv48922.2021.00986>
- [23] Nam, J., Syazwany, N.S., Kim, S.J. and Lee, S. (2024) Modality-Agnostic Domain Generalizable Medical Image Segmentation by Multi-Frequency in Multi-Scale Attention. 2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Seattle, 16-22 June 2024, 11480-11491. <https://doi.org/10.1109/cvpr52733.2024.01091>
- [24] Gu, A.R., Nam, J.H. and Lee, S.C. (2022) FBI-Net: Frequency-Based Image Forgery Localization via Multitask Learning with Self-attention. *IEEE Access*, **10**, 62751-62762. <https://doi.org/10.1109/access.2022.3182024>
- [25] Qin, Z., Zhang, P., Wu, F. and Li, X. (2021) FcaNet: Frequency Channel Attention Networks. 2021 IEEE/CVF International Conference on Computer Vision (ICCV), Montreal, 10-17 October 2021, 783-792. <https://doi.org/10.1109/iccv48922.2021.00082>
- [26] Sang, M. and Hansen, J.H.L. (2022) Multi-Frequency Information Enhanced Channel Attention Module for Speaker Representation Learning. *Proceedings of Interspeech 2022*, Incheon, 18-22 September 2022, 321-325. <https://doi.org/10.21437/interspeech.2022-892>
- [27] Gu, A. and Dao, T. (2024) Mamba: Linear-Time Sequence Modeling with Selective State Space. arXiv: 2312.00752.
- [28] Liu, Y., Shao, Z. and Hoffmann, N. (2021) Global Attention Mechanism: Retain Information to Enhance Channel-Spatial

- Interactions. arXiv: 2112.05561.
- [29] Kelenyi, B., Domsa, V. and Tamas, L. (2024) Sam-Net: Self-Attention Based Feature Matching with Spatial Transformers and Knowledge Distillation. *Expert Systems with Applications*, **242**, Article 122804. <https://doi.org/10.1016/j.eswa.2023.122804>
- [30] Hou, Q., Zhou, D. and Feng, J. (2021) Coordinate Attention for Efficient Mobile Network Design. 2021 *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Nashville, 20-25 June 2021, 13713-13722. <https://doi.org/10.1109/cvpr46437.2021.01350>
- [31] Sandler, M., Howard, A., Zhu, M., Zhmoginov, A. and Chen, L. (2018) MobileNetV2: Inverted Residuals and Linear Bottlenecks. 2018 *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Salt Lake City, 18-23 June 2018, 4510-4520. <https://doi.org/10.1109/cvpr.2018.00474>