

基于YOLOv11的火灾检测算法研究

陆电朋, 杨丽娟*, 张一斌

北华航天工业学院计算机系, 河北 廊坊

收稿日期: 2025年9月1日; 录用日期: 2025年9月30日; 发布日期: 2025年10月10日

摘要

由于火灾事故的频繁发生,火灾检测技术在保障生命财产安全和推动社会可持续发展方面具有重要意义。近年来,基于计算机视觉和深度学习的火灾检测方法因其高效性和较强的适应能力,逐渐成为火灾检测领域的研究热点。鉴于现有火灾检测算法难以兼顾多尺度目标与小目标检测,本文基于YOLOv11提出一种改进的火灾检测模型YOLOv11_VAE。首先,为弥补现有数据集的不足,本文构建了一个包含7687张图像、专注于高危场景下初期微小火焰的专用数据集。其次,为提升对微小火焰的感知能力,设计了多尺度卷积与注意力增强模块(MCAE),通过“生成-选择”策略捕捉并增强不同尺度的关键火焰特征。为降低复杂背景导致的误报,提出了全局-局部特征聚合模块(GLFA),通过融合场景级上下文信息提升模型的判别能力。在自建数据集上的大量实验证明了所提方法的有效性。结果表明,YOLOv11_VAE的mAP@0.5值达到0.964,相较于YOLOv11n基线模型提升了3.3个百分点,同时模型参数量仅为2.70 M,计算量为6.6 GFLOPs,实现了高精度与高效率的平衡。YOLOv11_VAE为高危环境下的早期火灾预警提供了一种有效且高效的解决方案。

关键词

火灾检测, YOLOv11, 小目标检测, 多尺度特征

Research on Fire Detection Algorithm Based on YOLOv11

Dianpeng Lu, Lijuan Yang*, Yibin Zhang

Department of Computer Science, North China Institute of Aerospace Engineering, Langfang Hebei

Received: September 1, 2025; accepted: September 30, 2025; published: October 10, 2025

Abstract

Due to the frequent occurrence of fire accidents, fire detection technology is of great significance in

*通讯作者。

ensuring the safety of life and property and promoting the sustainable development of society. In recent years, fire detection methods based on computer vision and deep learning have gradually become a research hotspot in the field of fire detection due to their high efficiency and strong adaptability. In view of the fact that the existing fire detection algorithms are difficult to take into account the detection of multi-scale objects and small objects at the same time, this paper proposes an improved fire detection model YOLOv11_VAE based on YOLOv11. Firstly, in order to make up for the shortcomings of the existing data sets, this paper constructs a dedicated data set containing 7687 images focusing on the early tiny flames in high-risk scenes. Secondly, in order to improve the perception ability of tiny flames, a multi-scale convolution and Attention Enhancement module (MCAE) is designed to capture and enhance key flame features at different scales through the “generation-selection” strategy. In order to reduce false positives caused by complex backgrounds, a Global-local feature aggregation module (GLFA) was proposed to improve the discriminative ability of the model by fusing scene-level context information. Extensive experiments on self-built datasets demonstrate the effectiveness of the proposed method. The results show that the mAP@0.5 value of YOLOv11_VAE reaches 0.964, which is 3.3 percentage points higher than that of the YOLOv11n baseline model. At the same time, the model parameter number is only 2.70 M, and the calculation amount is 6.6 GFLOPs, achieving a balance between high accuracy and high efficiency. YOLOv11_VAE provides an effective and efficient solution for early fire warning in high-risk environments.

Keywords

Fire Detection, YOLOv11, Small Target Detection, Multi-Scale Features

Copyright © 2025 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. 绪论

1.1. 研究背景及意义

火灾作为一种突发性、破坏性极强的灾难，一直以来对社会、经济和人民生活造成了恶劣的影响。无论住宅区、工业园区还是森林火灾，其引发的损失都是巨大的。火灾不仅直接危害到人民生命财产安全，还会带来环境污染、生态破坏等间接影响，甚至影响国家的经济发展与社会稳定。火灾对环境的破坏也不容忽视，森林火灾对生态系统的破坏极为严重[1]。全球每年约有 100 万平方公里的森林受到火灾影响，火灾不仅摧毁大量树木，还可能导致生物多样性的丧失和土地荒漠化。2019 年亚马逊雨林的火灾如图 1 所示。

火灾防控和检测系统的改进，不仅能够减少火灾造成的生命财产损失，还能有效提升应急响应的效率。随着火灾场景的复杂性增加，传统方法的局限性逐渐显现，因此，基于深度学习的火灾检测方法逐步得到应用和研究。通过提高火灾检测系统的准确性和实时性，尤其是在早期火灾阶段和多尺度火灾目标检测方面的突破，将对火灾防控工作起到重要的促进作用。

本研究将致力于改进现有的基于深度学习的火灾检测方法，针对火灾早期阶段的小火苗和烟雾的检测问题以及火灾现场多尺度目标的检测问题，提出一种更高效、更精准的火灾检测模型。通过优化模型结构、构建更全面的数据集以及设计更具适应性的算法实现模型的改进，本研究期望能够提高现有检测模型的性能，进一步推动火灾检测技术的研究与应用，进而减少火灾带来的损失，为公共安全和环境保护贡献力量。



Figure 1. Fire scene in the Amazon rainforest
图 1. 亚马逊雨林火灾现场图

1.2. 核心技术挑战

将深度学习目标检测技术应用于高危环境的早期火灾预警，面临两大核心技术挑战：

在火灾初期，火焰通常以 5~15 像素的微小形态出现。而主流的卷积神经网络(CNN)为了获取高级语义信息，采用了分层下采样的结构[2]。在这一过程中，微小火焰目标的精细空间信息和纹理细节被严重压缩和稀释，极易与复杂的工业背景特征相混合，最终在深层特征图中变得与背景噪声无异，导致模型漏检。

高危工业场景中普遍存在与火焰颜色、形状相似的干扰物，如橙色的警示灯、红色的灭火器、反光金属表面等。缺乏对场景整体理解能力的模型，仅凭局部特征很难区分真实的火焰与这些干扰物，从而产生大量误报[3]。例如，YOLOv11 等通用模型中引入的自注意力机制，在计算全局依赖时，会自然地赋予信号强度高、占据面积大的静态工业设备更高的权重，反而可能将微小的火焰信号视为噪声并加以抑制。

1.3. 解决方案与主要贡献

为应对上述挑战，本文提出了一种专为高危环境早期火灾检测设计的模型——YOLOv11_VAE。本文的主要贡献可概括为以下四点：

构建了专用的高危环境火灾数据集：通过整合公开数据集与定向网络爬取，构建了一个包含 7687 张图像的数据集。与通用火灾数据集不同，该数据集经过精心筛选和标注，其数据分布呈现出初期小尺寸火焰目标压倒性占比的特点，为解决该特定问题提供了坚实的实证基础。

提出了用于微小目标感知的 MCAE 模块：为解决微小火焰特征丢失问题，本文设计了多尺度卷积与注意力增强模块(Multi-Scale Convolution and Attention Enhancement, MCAE)。该模块采用“生成 - 选择”架构，通过多尺度卷积分支生成丰富的特征表示，再利用并行的注意力机制筛选并增强关键的火焰特征[4]，从而显著提升模型对微小目标的检测灵敏度。

设计了用于上下文消歧的 GLFA 模块：为降低由背景干扰引发的误报，本文提出了全局 - 局部特征聚合模块(Global-Local Feature Aggregation, GLFA)。该模块通过双通路设计，协同融合了由空洞卷积[5]捕获的多感受野局部特征与由空间金字塔池化生成的全局场景上下文，使模型能够利用场景信息来区分真实火焰与视觉干扰物。

实现了高精度与高效率的平衡：通过大量的对比实验与消融实验，验证了 YOLOv11_VAE 的卓越性能。该模型在检测精度上全面超越了多种主流检测器，同时保持了轻量化的架构，适用于资源受限的边缘设备进行实时部署。

2. 相关工作

火灾检测技术经历了从传统到智能的演进过程，主要可分为传统检测范式和基于深度学习的检测范式。

2.1. 传统检测范式

传统的火灾检测技术主要包括基于传感器的物理检测和基于图像处理的视觉检测。基于传感器的技术，如烟雾和感温探测器，是应用最广泛的手段。然而，这类技术的根本缺陷在于其被动性，必须等待燃烧副产物(如烟雾、热量)扩散至传感器位置才能触发警报，这在大型、高顶空间中会导致严重的响应延迟[6]。此外，它们只能提供“有/无”的二元信号，无法提供火灾位置、规模等关键信息，且易受水蒸气、灰尘等非火灾因素干扰。

为克服传感器的局限性，研究人员转向基于图像处理的视觉检测技术。早期方法主要依赖于人工设计的规则，如利用 RGB、YCbCr 或 HSV 颜色空间中的颜色特征来分割疑似火焰区域。一些研究通过融合火焰的纹理、形状、动态属性等多种特征来提高鲁棒性。尽管这些方法在特定场景下取得了一定的成功，但其本质上是脆弱的。硬编码的规则对光照变化、复杂背景以及与火焰颜色相似的干扰物极其敏感，导致算法泛化能力差，难以适应多变的现实场景。

2.2. 基于深度学习的检测范式

近年来，以卷积神经网络(CNN)为代表的深度学习技术为火灾检测开辟了新路径。与传统方法不同，深度学习模型能够自动从海量数据中学习火灾的多层次特征，无需人工设计规则，从而展现出更强的鲁棒性和泛化能力。目标检测领域的两大主流框架——以 Faster R-CNN 为代表的双阶段检测器和以 YOLO 系列为代表的单阶段检测器，均被应用于火灾检测任务中。

然而，现有研究存在一个明显的研究空白：直接将通用目标检测任务(如 COCO 数据集)设计的模型应用于早期工业火灾检测，存在严重的“水土不服”。这些通用模型在架构设计上，为追求对中大型物体的识别精度，往往采用激进的下采样策略，这恰恰是导致微小火焰特征丢失的根源。虽然一些研究尝试通过引入标准的注意力模块或特征金字塔网络来改进 YOLO 等模型，但这些通常是通用性的改进，并未从根本上解决本任务的核心矛盾：如何在强噪声(复杂工业背景)中检测到极微弱的信号(初期小火焰)[7]。因此，本研究认为，必须设计专门的、能够感知微小目标并理解场景上下文的定制化模块，才能在该特定领域取得突破。

3. YOLOv11_VAE 火灾检测模型

3.1. 总体架构

YOLOv11_VAE 的整体架构如图 2 所示，依然遵循主干(Backbone)-颈部(Neck)-头部(Head)的经典设

计。选择 YOLOv11n 作为基线模型，是因为它代表了 YOLO 系列在精度和效率方面的一个先进平衡点，为针对性改进提供了坚实的基础。

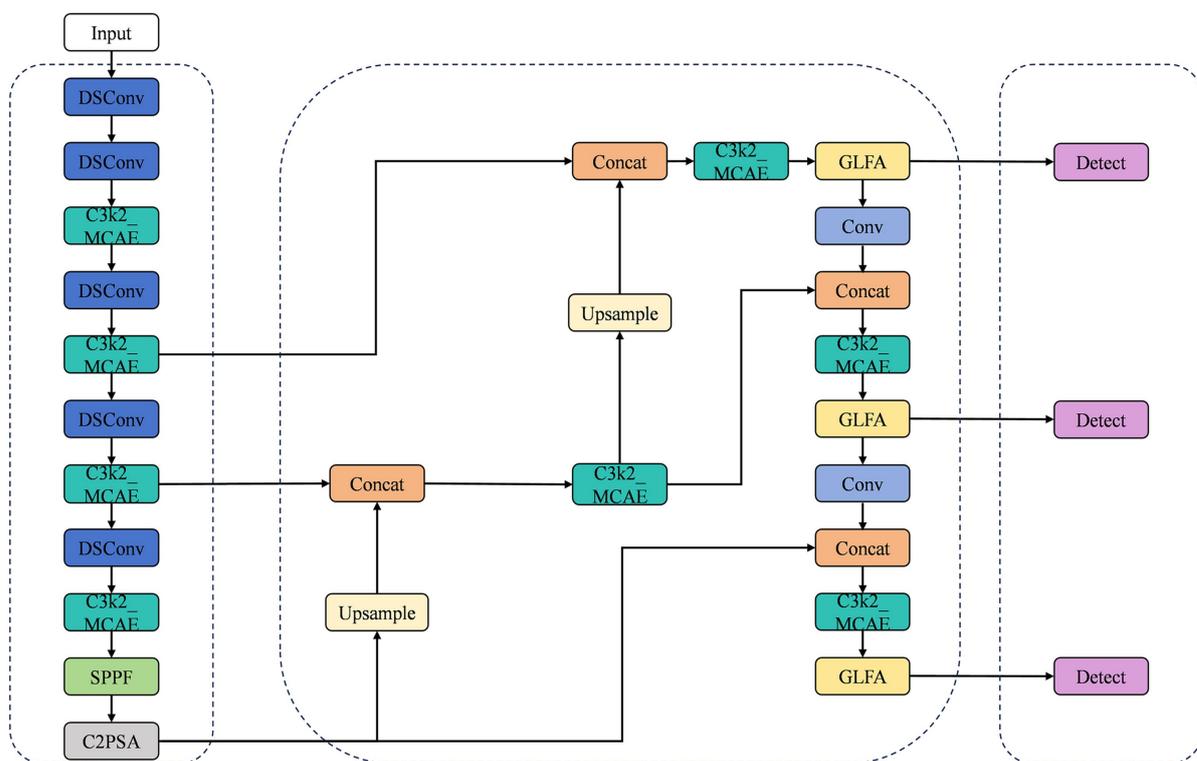


Figure 2. YOLOv11 VAE network structure diagram

图 2. YOLOv11_VAE 网络结构图

本模型的核心设计哲学是策略性资源分配。首先，通过在计算最密集的主干网络中引入深度可分离卷积(Depthwise Separable Convolution, DSCConv)，大幅降低基础特征提取的计算开销。然后，将节约下来的计算资源“再投资”于网络中更关键的阶段，即集成两个为解决核心难题而设计的、功能更强大的定制化模块：多尺度卷积与注意力增强模块(Multi-Scale Convolution and Attention Enhancement Module, MCAE)和全局-局部特征聚合模块(Global-Local Feature Aggregation Module, GLFA)。这种设计使得模型在整体计算复杂度可控的前提下，将“算力”用在“刀刃”上，实现了性能的最大化。

3.2. 基于深度可分离卷积的轻量化主干网络

标准卷积操作同时处理空间相关性与跨通道相关性，计算成本高昂[8]。深度可分离卷积通过将其分解为深度卷积(Depthwise Convolution)和逐点卷积(Pointwise Convolution)两个独立步骤，显著降低了计算量和参数量。

深度卷积：为每个输入通道分配一个独立的空间滤波器，仅负责提取空间特征，通道间不交互。逐点卷积：使用 1×1 卷积核对深度卷积的输出进行线性组合，负责通道间的特征融合。通过在主干网络中用 DSCConv 替代标准卷积，模型在处理高分辨率输入特征图时获得了巨大的效率增益。这一基础性优化是后续集成更复杂模块而不过度增加模型整体重量的关键，为实现模型的实时部署奠定了基础。卷积网络结构对比情况如图 3 所示。

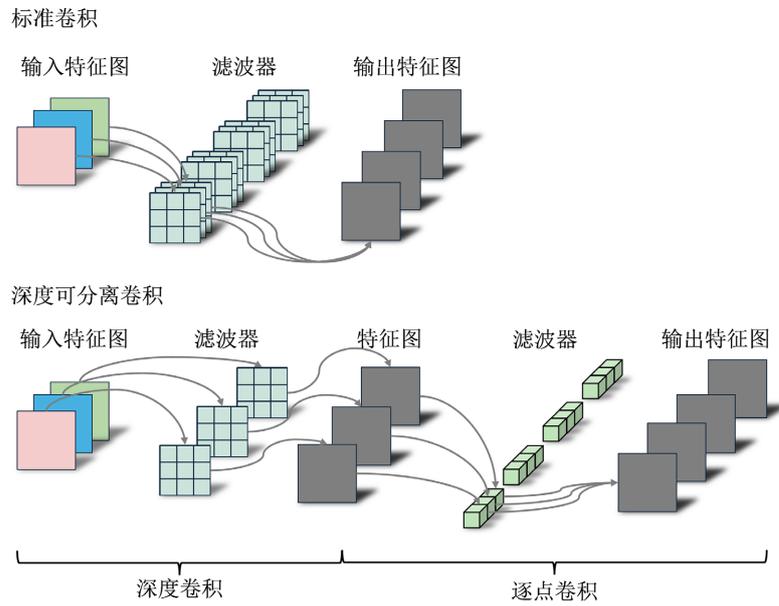


Figure 3. Comparison of convolutional network architecture
图 3. 卷积网络结构对比图

3.3. MCAE: 增强微小火焰感知

为解决下采样过程中微小火焰特征易被稀释的问题[9], 本文设计了 MCAE 模块。其核心思想是一种“生成 - 选择”策略。MCAE 模块结构如图 4 所示。

多尺度卷积与注意力增强模块

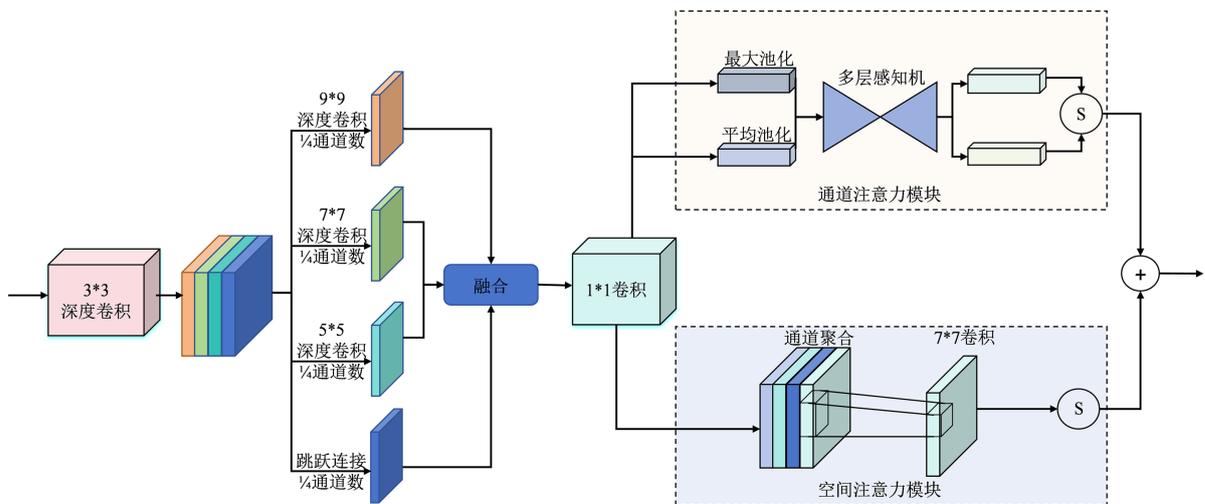


Figure 4. MCAE module structure diagram
图 4. MCAE 模块结构图

生成阶段(多尺度卷积): 该阶段旨在生成一个对不同尺度特征都敏感的丰富表示。输入特征图被送入四个并行的分支: 三个分支分别采用不同尺寸的卷积核(如 5×5 , 7×7 , 9×9)进行深度卷积, 以捕捉从

精细纹理到目标轮廓的多尺度信息；第四个分支是直连的跳跃连接，旨在无损地传递原始特征。这种设计确保了无论火焰以何种微小形态出现，其特征信号都至少在一个分支中被有效捕获，从而避免了在单一尺度处理中被“平均掉”的风险。深度卷积层中复合函数定义见式(1)。

$$F_k(x) = \delta(B(\phi_{dw,k}(x))) \quad (1)$$

其中， F_k 表示多尺度卷积模块； $\phi_{dw,k}$ 表示深度卷积层； B 表示批量归一层； δ 表示 ReLU6 激活函数。由公式(1)可得多尺度卷积分支整体计算公式，见式(2)。

$$F_{in} = \psi_{pw}(Concat(F_{k=9}(x_1); F_{k=7}(x_2); F_{k=5}(x_3); x_4)) \quad (2)$$

其中， F_{in} 表示注意力模块的输入特征图； ψ_{pw} 表示 1×1 逐点卷积； $Concat$ 表示特征融合； x_1, x_2, x_3, x_4 表示初始输入特征图的四等份分支。

选择阶段(并行注意力)：该阶段对生成的多尺度特征进行筛选和增强。与传统的串行注意力机制不同，MCAE 采用了并行的通道注意力和空间注意力机制。通道注意力负责建模“什么”是重要的(如火焰特有的颜色和亮度通道)，相关计算公式见式(3)。

$$M_c(F_{in}) = \sigma(W_1(\delta(W_0(avg_c^F))) + W_1(\delta(W_0(max_c^F)))) \quad (3)$$

其中， avg_c^F, max_c^F 表示经过平均池化与最大池化操作后得到的 $C \times 1 \times 1$ 大小特征描述符； $W_0 \in R^{(C/r) \times C}$ 为 C/r 行 C 列的权重矩阵； $W_1 \in R^{C \times (C/r)}$ 则为 C 行 C/r 列的权重矩阵； δ 表示 ReLU 激活函数，三者共同构成共享的多层感知机； σ 表示 Sigmoid 激活函数。

空间注意力负责建模“哪里”是重要的(如火焰的空间位置和形状)。具体计算公式见式(4)。

$$M_s(F_{in}) = \sigma(f^{7 \times 7}(Concat[AvgPool_c(F_{in}); MaxPool_c(F_{in})])) \quad (4)$$

其中， $AvgPool_c(\cdot), MaxPool_c(\cdot)$ 表示尺寸为 $1 \times$ 高度 \times 宽度的池化运算结果； $Concat[;]$ 表示拼接操作； $f^{7 \times 7}$ 表示尺寸为 7×7 的标准卷积层； σ 表示 Sigmoid 激活函数。

模块最终将经过通道注意力加权的特征图与经过空间注意力加权的特征图进行逐元素相加，计算公式见式(5)。

$$Y = (X \odot M_c(X)) + (X \odot M_s(X)) \quad (5)$$

其中， $M_c(X)$ 表示经过通道注意力调制后的结果； $M_s(X)$ 表示经过空间注意力调制后的结果； \odot 表示哈达玛积。

并行设计解除了两种注意力之间的固定依赖关系，允许模型根据数据自主学习如何最优地组合这两种信息。对于形态不规则、颜色和形状同等重要的火焰目标，这种灵活性至关重要，它确保了模型能够从不同维度捕捉火焰的关键特征[10]，从而根本上提升了对早期微小火焰的检测灵敏度。

3.4. GLFA：用于上下文消歧的特征聚合

MCAE 模块极大地提升了模型的检测敏感度，但这种敏感性也可能使其将背景中的干扰物误判为火焰。GLFA 模块的作用就是引入场景上下文信息，进行“二次确认”，从而提升模型的特异性，降低误报。

双通路协同设计：采用双通路并行设计，功能互补。模块结构如图 5 所示。

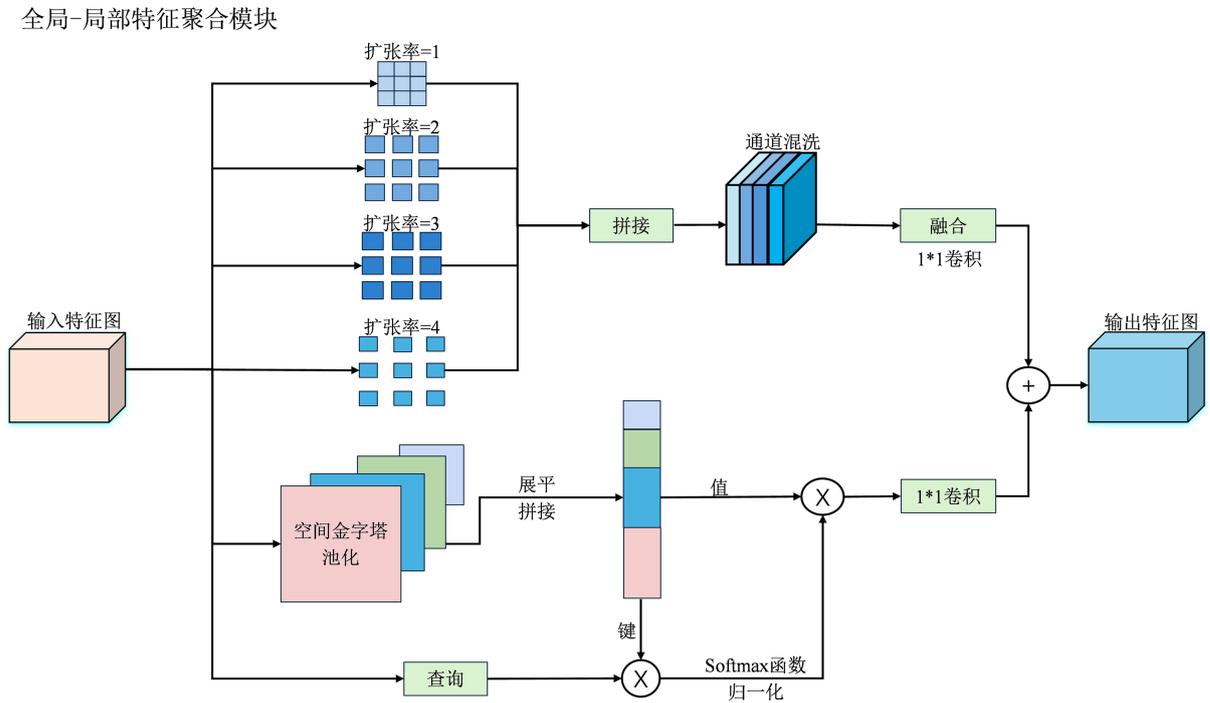


Figure 5. GLFA module structure diagram
图 5. GLFA 模块结构图

局部通路(多感受野特征提取): 该通路利用一组并行的空洞卷积来捕捉不同范围的局部上下文信息。空洞卷积能够在不增加计算量和不降低分辨率的前提下, 指数级地扩大感受野。计算公式见式(6)。

$$Y[i, j] = \sum_{k=1}^K \sum_{l=1}^K X[i+r \cdot k, j+r \cdot l] \cdot \omega[k, l] \quad (6)$$

其中, $Y[i, j]$ 表示输出特征图在坐标 (i, j) 处的像素值; X 表示输入特征图; ω 表示卷积核; K 表示卷积核尺寸; k, l 为循环变量, 用于遍历卷积核的每一个权重; r 表示扩张率, 控制卷积核采样点之间的距离。

关键在于, 在拼接不同空洞卷积的输出后, 引入了通道混洗操作。该操作能够打破因拼接带来的信息壁垒, 确保来自不同感受野的特征信息被充分交叉、融合, 从而生成一个集成度更高的局部特征表示。

全局通路(上下文感知注意力): 该通路负责生成一个场景级的理解。首先, 通过空间金字塔池化(Spatial Pyramid Pooling)将输入特征图压缩成一个紧凑的全局上下文描述符[4]。相关计算公式见式(7)。

$$pool_{(n \times n)}^X[i, j] = \max_{h, w} (X_c[h, w]) \quad (7)$$

其中, n 表示卷积核尺寸; $i, j \in [1, n]$, 表示网格索引; 而 (h, w) 的取值范围则对应于输入特征图上被第 (i, j) 网格所覆盖的区域。

然后采用一种高效的非对称自注意力机制。与传统的自注意力机制不同, 它从高分辨率的原始特征图中派生查询向量(Query), 保留了精确的空间位置信息; 而从紧凑的全局上下文描述符中派生键(Key)和值(Value)向量。这在大幅降低计算复杂度的同时, 也完美契合了其设计意图: 用全局上下文来调制(或重新加权)每一个局部特征。注意力机制相关公式见式(8)。

$$\text{Attention}(Q, K, V) = \text{Softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V \quad (8)$$

其中, $\sqrt{d_k}$ 表示缩放因子, d_k 表示键向量的维度; $\text{softmax}(\cdot)$ 表示使用 Softmax 函数进行归一化。

MCAE 和 GLFA 的组合形成了一个高效的认知系统。MCAE 扮演着一个“自底向上”的信号放大器角色, 对任何潜在的火焰信号都保持高度敏感。GLFA 则扮演着一个“自顶向下”的上下文过滤器角色, 它利用对整个场景的理解, 来验证或否决由 MCAE 识别出的候选目标。这种协同作用, 模拟了人类视觉系统先发现异常、再结合环境确认威胁的认知过程, 是 YOLOv11_VAE 能够同时实现高召回率和高精确率的核心所在。

4. 面向高危环境的火灾检测算法研究

为验证 YOLOv11_VAE 模型的有效性, 本文在自建的高危环境火灾数据集上进行了一系列全面的实验, 包括消融实验和与主流模型的对比实验。

4.1. 实验设置

本文采用的数据集通过筛选公开数据集 YOLO-HF [11]和定向网络爬取相结合的方式构建, 共包含 7687 张图像, 严格按照 8:1:1 的比例划分为训练集、验证集和测试集。该数据集的核心特点是其数据分布中初期小尺寸火焰目标占据压倒性比例, 能够真实反映并有效检验模型在目标场景下的性能。

采用精确率(Precision)、召回率(Recall)、F1 分数(F1-Score)和平均精度均值(mAP@0.5)作为模型精度的评价指标。同时, 使用模型参数量(Parameters, M)和浮点运算次数(GFLOPs)作为模型复杂度和效率的评价指标。

所有实验均在统一的软硬件环境下进行, 以保证结果的公平性和可复现性。关键训练超参数设置如表 1 所示。

Table 1. Model training hyperparameters

表 1. 模型训练超参数

序号	超参数	数值
1	学习率	0.01
2	批次大小	32
3	优化器	SGD
4	权重衰减系数	5e-4
5	训练周期	200

4.2. 对比实验

为证明 YOLOv11_VAE 的先进性, 本文将其与包括双阶段检测器 Faster R-CNN 以及 YOLOv5s、YOLOv8n 在内的多种主流模型在测试集上进行了性能对比。结果如表 2 所示。

Table 2. Comparing experimental results

表 2. 对比实验结果

模型	mAP@0.5	精确率	召回率	F1 分数	参数量	GFLOPs
Faster R-CNN	0.936	0.948	0.895	0.921	41.76	134.38
YOLOv5s	0.930	0.921	0.881	0.901	7.01	15.8
YOLOv8n	0.900	0.928	0.864	0.895	3.01	8.1
YOLOv11n	0.931	0.928	0.878	0.909	2.58	6.3
YOLOv11_VAE	0.964	0.959	0.925	0.942	2.70	6.6

与重量级的双阶段检测器 Faster R-CNN 相比, YOLOv11_VAE 不仅在 mAP@0.5 上高出 2.8 个百分点, 更在效率上呈现出压倒性优势, 参数量和计算量分别仅为前者的 6.5%和 4.9%, 更适用于实时检测任务。

与同为轻量级的 YOLO 系列模型(YOLOv5s, YOLOv8n, YOLOv11n)相比, YOLOv11_VAE 在所有精度指标上均实现了全面超越。特别是相较于其基线模型 YOLOv11n, mAP@0.5 实现了从 0.931 到 0.964 的显著提升。这一结果有力地证明, 通用的模型架构不足以应对本任务的特殊挑战, 而本文提出的针对性改进是实现性能突破的关键。

4.3. 消融实验

为逐一验证本文提出的各改进模块(DSConv, MCAE, GLFA)的有效性, 本文设计了消融实验。实验以 YOLOv11n 为基线, 逐步添加各个模块, 并在验证集上评估其性能。结果如表 3 所示。

Table 3. Results of ablation experiments

表 3. 消融实验结果

模型	mAP@0.5	精确率	召回率	F1 分数	参数量	GFLOPs
YOLOv11n	0.956	0.943	0.922	0.932	2.58	6.3
YOLOv11n + DSConv	0.952	0.942	0.912	0.927	2.10	4.9
YOLOv11n + MCAE	0.970	0.949	0.945	0.947	3.15	7.1
YOLOv11n + GLFA	0.969	0.947	0.941	0.944	2.62	6.4
YOLOv11_VAE	0.980	0.968	0.953	0.961	2.70	6.6

根据上表实验数据可以得出以下结论:

单独引入 DSConv 后, 模型的参数量和 GFLOPs 分别显著下降了 18.6%和 22.2%, 而 mAP 仅有微小下降。这证明了 DSConv 作为一种轻量化策略的有效性, 成功为后续更复杂的模块节约了计算预算。

单独引入 MCAE 模块带来了最为显著的性能飞跃, mAP 提升了 1.4 个百分点, 特别是召回率从 0.922 大幅提升至 0.945。这直接证明了 MCAE 的多尺度和并行注意力设计能够有效捕捉和增强微小火焰特征, 显著降低了漏检率。

单独引入 GLFA 模块同样带来了可观的性能增益(mAP 提升 1.3 个百分点), 且模型复杂度增加极小。其精确率的提升(从 0.943 到 0.947)验证了该模块通过融入全局上下文, 有效抑制了背景干扰, 降低了误报率。

集成了所有模块的完整模型 YOLOv11_VAE 在所有精度指标上均达到最优, mAP 达到了 0.980。这表明各个模块之间存在积极的协同作用, 共同构建了一个更精确、更鲁棒的检测模型, 并且最终模型的复杂度与基线模型相当, 验证了本文“策略性资源分配”设计哲学的成功。

5. 结论

5.1. 工作总结

本文围绕高危环境下火灾早期预警的迫切需求, 针对通用目标检测算法在识别初期微小火焰时存在的漏检和误报难题, 开展了系统性的研究。核心成果是设计并实现了一种轻量化、高精度的火灾检测模型 YOLOv11_VAE。该研究通过构建专用数据集和进行针对性的网络架构创新, 成功地解决了特定场景下的关键技术挑战。实验结果表明, YOLOv11_VAE 在自建的高危环境火灾数据集上取得了 0.964 的 mAP@0.5, 性能全面超越包括其基线模型在内的多种主流检测器, 同时模型参数量仅为 2.70 M, 计算量

为 6.6 GFLOPs, 在检测精度与部署效率之间取得了当前最优的平衡点, 验证了其作为实际应用方案的巨大潜力。

5.2. 未来展望

尽管本研究取得了预期的成果, 但仍存在可进一步探索的方向。基于当前工作的局限性, 未来的研究可从以下几个方面深化与拓展:

当前模型仅依赖可见光图像, 其性能在烟雾遮挡或光照不足等恶劣条件下会受影响。未来的一个核心研究方向是融合热红外图像信息。热红外传感器能够直接捕捉温度异常, 甚至在可见火焰产生前就能发现过热征兆。构建能够同时理解颜色、纹理和温度分布的多模态融合模型, 将能极大提升检测系统的鲁棒性, 实现从“检测”到“预警”的范式转变。

火灾是一个动态过程, 火焰的闪烁、蔓延等时序特征包含了丰富的判别信息。当前模型对独立的视频帧进行分析, 未能利用这些动态线索。未来可以引入能够建模时间序列的神经网络架构, 如 3D CNN 或视觉 Transformer, 通过学习火焰在时间维度上的动态特征, 有望进一步降低误报率, 并为火灾风险评估提供更丰富的依据。

为适应资源高度受限的边缘计算设备, 未来可进一步探索模型压缩技术, 如知识蒸馏、网络剪枝等, 在追求极致性能的同时, 进一步降低模型的计算和存储开销, 推动算法在更广泛的工业场景中落地应用。

参考文献

- [1] Xia, K., Lv, Z., Zhou, C., Gu, G., Zhao, Z., Liu, K., *et al.* (2023) Mixed Receptive Fields Augmented YOLO with Multi-Path Spatial Pyramid Pooling for Steel Surface Defect Detection. *Sensors*, **23**, Article 5114. <https://doi.org/10.3390/s23115114>
- [2] Lin, T., Dollar, P., Girshick, R., He, K., Hariharan, B. and S. Belongie. (2017) Feature Pyramid Networks for Object Detection. 2017 *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Honolulu, 21-26 July 2017, 2117-2125. <https://doi.org/10.1109/cvpr.2017.106>
- [3] Yang, G., Lei, J., Zhu, Z., Cheng, S., Feng, Z. and Liang, R. (2023) AFPN: Asymptotic Feature Pyramid Network for Object Detection. 2023 *IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, Honolulu, 1-4 October 2023, 2184-2189. <https://doi.org/10.1109/smc53992.2023.10394415>
- [4] Zhang, X., Zhou, X., Lin, M. and Sun, J. (2018) Shufflenet: An Extremely Efficient Convolutional Neural Network for Mobile Devices. 2018 *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Salt Lake City, 18-23 June 2018, 6848-6856. <https://doi.org/10.1109/cvpr.2018.00716>
- [5] Huang, K. and Xu, Z. (2024) Lightweight Video Salient Object Detection via Channel-Shuffle Enhanced Multi-Modal Fusion Network. *Multimedia Tools and Applications*, **83**, 1025-1039. <https://doi.org/10.1007/s11042-023-15251-x>
- [6] Wei, X., Wang, G., Schmalz, B., Hagan, D.F.T. and Duan, Z. (2023) Evaluation of Transformer Model and Self-Attention Mechanism in the Yangtze River Basin Runoff Prediction. *Journal of Hydrology: Regional Studies*, **47**, Article 101438. <https://doi.org/10.1016/j.ejrh.2023.101438>
- [7] Zhang, Z., Li, B., Yan, C., Furuichi, K. and Todo, Y. (2025) Double Attention: An Optimization Method for the Self-Attention Mechanism Based on Human Attention. *Biomimetics*, **10**, Article 34. <https://doi.org/10.3390/biomimetics10010034>
- [8] Purkait, P., Zhao, C. and Zach, C. (2017) SPP-Net: Deep Absolute Pose Regression with Synthetic Views. arXiv: 1712.03452.
- [9] Pincott, J., Tien, P.W., Wei, S. and Calautit, J.K. (2022) Indoor Fire Detection Utilizing Computer Vision-Based Strategies. *Journal of Building Engineering*, **61**, Article 105154. <https://doi.org/10.1016/j.jobe.2022.105154>
- [10] Qi, Y., Yang, Z., Sun, W., Lou, M., Lian, J., Zhao, W., *et al.* (2021) A Comprehensive Overview of Image Enhancement Techniques. *Archives of Computational Methods in Engineering*, **29**, 583-607. <https://doi.org/10.1007/s11831-021-09587-6>
- [11] Peng, B. and Kim, T. (2025) YOLO-HF: Early Detection of Home Fires Using Yolo. *IEEE Access*, **13**, 79451-79466. <https://doi.org/10.1109/access.2025.3566907>