

RT-DETR 火焰检测模型优化研究^{*}

熊子健

(四川理工技师学院 信息服务处, 四川 成都 611130)

【摘 要】针对火焰检测任务 RT-DETR (Real-Time Detection Transformer) 模型多尺度特征捕捉不足、复杂背景干扰大及推理速度受限等问题, 提出改进方案: 采用多尺度特征融合模块 DFM 增强尺度适应性, 通过 EMA (Efficient Multi-Scale Attention) 注意力机制优化特征聚焦, 替换骨干网络为轻量化模型 StarNet 提升实时性. 实验显示, 改进后模型 mAP 提升 1.1%, 推理速度达 33.7 f/s, 优于原始模型及主流对比模型, 可满足工业消防、森林防火等场景实时监测需求.

【关键词】火焰检测; RT-DETR; StarNet; 特征融合; 注意力机制

【中图分类号】TP312 【文献标志码】A 【文章编号】1674-5639 (2025) 06-0107-07

DOI: 10.14091/j.cnki.kmxyxb.2025.06.014

火焰检测作为公共安全与工业防控领域的关键技术, 其核心目标是通过视觉传感器实时识别火焰目标, 实现火灾的早期预警与快速响应, 从而最大限度减少人员伤亡和财产损失. 当前, 城市化进程加快与工业生产规模扩大, 使火灾隐患点增多^[1], 这对火焰检测技术的精度、速度及抗干扰能力提出了更高要求. 传统火焰检测方法依赖温度传感器、烟雾报警器等设备, 存在检测范围有限、易受环境干扰等缺陷, Zhang 等^[2]指出, 此类设备检测半径 ≤ 8 m, 且在高温高湿环境下误报率超 25%. 而基于计算机视觉的火焰检测技术凭借非接触式、大范围监测的优势, 成为当前研究热点.

现有火焰检测方法可分为传统视觉与深度学习两类. 传统方法依赖人工设计特征 (如 RGB 颜色阈值、动态轮廓等), 在简单场景 (如实验室受控火焰) 中检测精度可达 75%, 但面对光照突变 (如正午强光)、烟雾遮挡时, 鲁棒性显著下降, Li 等^[3]实验表明, 基于颜色阈值的方法在烟雾环境下漏检率骤升至 68.7%.

在基于深度学习的火焰检测方法中, 主流算法可分为两类: 基于卷积神经网络 (CNN)^[4]的方法和基于 Transformer^[5]的方法. 基于 CNN 的方法以 YOLO 系列^[6]、Faster R-CNN^[7]为代表, 通过多尺度检测头和锚框机制实现目标定位与分类. 例如, YOLOv8 通过改进特征金字塔网络 (FPN) 增强多尺度特征融合, 在火焰检测中实现了较快的推理速度 (约 60 f/s), 但对小尺度火焰 (如初期火苗) 的检测精度不足, 且在复杂背景 (如强光、红色物体干扰) 下易出现误检. Faster R-CNN 虽通过区域提议网络 (RPN) 提升了定位精度, 但两阶段检测机制导致推理速度较慢, 难以满足实时监测需求.

RT-DETR 由百度 PaddlePaddle 团队^[8]于 2023 年提出, 是基于 DETR (DEtection TRansformer) 架构进行轻量化设计的实时目标检测模型, 旨在克服传统 DETR 模型在部署时面临的高延迟、难训练、推理慢等难题. 与常见的基于卷积神经网络 (CNN) 架构的实时检测器 (如 YOLO 系列) 不同, RT-DETR 属于无 anchor、无 NMS 的纯 Transformer 检测器, 其在推理速度上进行了大量工程优化. 吴栋等^[9]针对原始 RT-DETR 在火焰检测中参数量大难适配边缘设备, 复杂背景下检测精度不足的问题, 通过替换轻量骨干网络并剪枝注意力机制实现模型轻量化, 新增火焰特征增强模块提升复杂场景检测精度, 引入 focal loss 与 CloU loss 优化损失函数, 改进出适配边缘设备且精度更高的轻量级火焰检测模型. 吴晓宁等人^[10]采用 Adown 子采样模块对原卷积模块进行改进, 提高了检测精度, 减少了参数值的个数. 在骨干网络上使用

^{*} 【收稿日期】2025-08-14

【作者简介】熊子健, 男, 湖北天门人, 四川理工技师学院教师, 硕士, 研究方向为计算机视觉.

LSKA 注意力模块进一步提高了检测精度, 与原始 RT-DETR 模型相比, FD-DETR 火焰检测的精确度和 mAP 分别提高了 0.8% 和 0.1%。

但结合实际火焰检测场景特性, RT-DETR 仍存在三方面局限亟待优化: 一是火焰尺度变化剧烈 (从几像素小火苗到占据图像大半区域的大面积火焰), 而模型对多尺度特征的融合能力不足, 导致小火焰漏检率较高; 二是火焰常伴随烟雾、灰尘等干扰因素, 且易与红色物体 (如广告牌、衣物) 混淆, 而模型注意力机制难以精准聚焦火焰区域; 三是原模型采用 ResNet-50 作为骨干网络, 计算成本高, 在嵌入式设备 (如边缘摄像头) 等资源受限场景中难以部署。RT-DETR 的网络结构图如图 1 所示。

为解决上述问题, 本文针对火焰检测任务的特性, 从模型轻量化、多尺度特征融合及注意力优化三个维度对 RT-DETR 展开改进。首先, 替换骨干网络为轻量化模型 StarNet^[11], 在降低模型参数量与计算成本的同时, 保留对火焰动态形态的捕捉能力; 然后, 在交叉尺度特征融合模块 (CCFM) 中引入 DFM 特征融合模块^[12], 强化不同尺度火焰特征的关联性, 提升对多尺度火焰的适配能力; 最后, 采用 EMA 注意力^[13] 机制优化特征聚焦效果, 增强模型对火焰区域的关注度并抑制背景干扰。

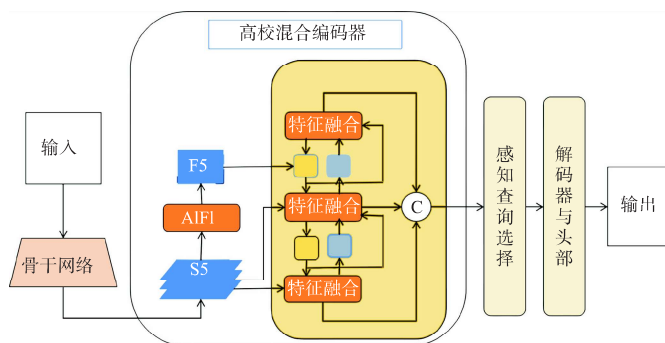


图 1 RT-DETR 的网络结构

1 RT-DETR 的改进策略

1.1 轻量化骨干网络 StarNet

RT-DETR 原骨干网络 (ResNet-50)^[14] 参数量大、计算成本高, 难以满足火焰检测的实时性需求。本研究采用轻量化的 StarNet 作为替代骨干。StarNet 在实时检测任务中表现突出, 其核心优势在于采用星形拓扑结构的特征聚合机制, 通过多分支并行提取不同尺度特征, 同时引入轻量化卷积单元 (如深度可分离卷积与动态扩张卷积结合), 在减少冗余计算的同时增强对火焰动态形态 (如跳动、扩散) 的捕捉能力。StarNet 输出的特征层级 (对应 C3 ~ C7) 与 RT-DETR 的混合编码器兼容性良好, 其中浅层特征保留了小火焰的细节信息 (如边缘轮廓), 深层特征经内置的轻量注意力模块强化后, 能有效抑制烟雾、强光等干扰, 提升复杂场景下火焰的识别精度, 与火焰检测任务的特性高度匹配。StarNet 的网络结构如图 2 所示。

图 2 展现了 StarNet 的网络架构, 整体由多阶段特征提取与特定模块构成。输入的图片首先通过初始卷积 (Conv) 层, 经此提取初步特征后, 进入包含阶段 1 - 阶段 4 的结构, 每个阶段内 “Conv + Star Blocks” 组合, 通过卷积与 Star Blocks 模块协同, 逐步提取、强化不同层次特征。

图 2 中虚线连接的细节模块里, DW-Conv (深度可分离卷积)^[15] 负责高效特征提取, FC (全连接) 层参与特征的变换与交互, 助力网络在轻量化前提下, 精准捕捉火焰等目标的特征, 最终经 GAP (全局平均池化) 和 FC (全连接) 层输出, 适配火焰检测等任务对特征提取与识别的需求, 实现从基础特征到有效分类识别特征的转化。Star-

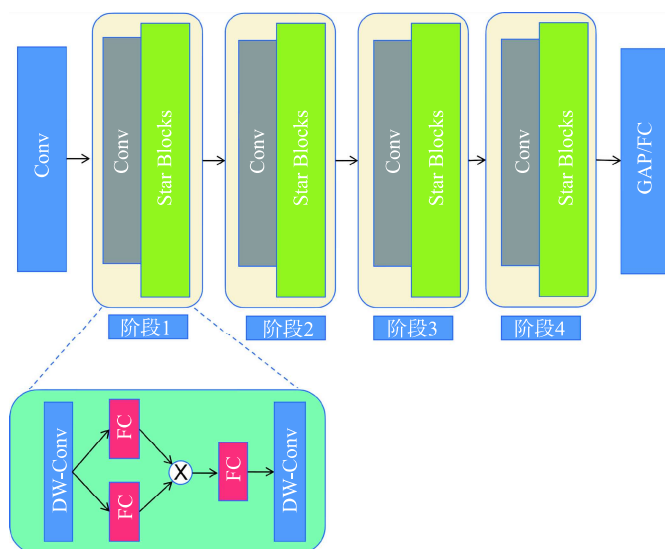


图 2 StarNet 网络结构

Net 最核心的 Star Blocks (星形模块) 包含深度可分离卷积 (DW-Conv) 和动态全连接 (FC) 交互, 可以用公式 (1) 表示:

$$F_{\text{star}} = FC_{\text{fusion}} (DW\text{-}Conv (F_{\text{in}}) \odot FC_{\text{adapt}} (F_{\text{in}})) . \quad (1)$$

其中, F_{in} 表示 Stage 内输入特征; DW-Conv 表示深度可分离卷积, 提取空间特征 (如火焰边缘、纹理); FC_{adapt} 是动态全连接层, 学习通道维度的注意力权重 (适配火焰的动态形态); $\odot$ 符号表示逐元素相乘 (动态特征交互, 强化关键火焰特征) 最后通过 FC_{fusion} 融合卷积, 输出增强后的特征.

1.2 引入多尺度特征融合模块 (DFM)

RT-DETR 的混合编码器中, 交叉尺度特征融合模块 (CCFM) 负责对骨干网络输出的多尺度特征进行初步融合, 但原始 CCFM 采用简单的自上而下特征传递方式, 存在浅层细节特征与深层语义特征融合不充分的问题, 难以适应火焰尺度剧烈变化 (从几像素火苗到大面积燃烧区域) 的检测需求. 为此, 本文在 CCFM 中嵌入密集特征融合模块 (DFM), 通过密集连接增强跨尺度特征关联性. 改进后的 CCFM 结构如图 3 所示.

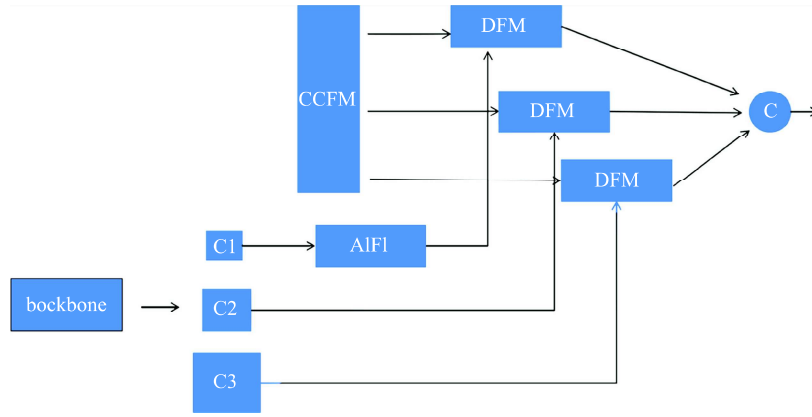


图3 融入DFM后的CCFM结构

图3展示了将DFM融入CCFM的核心逻辑, 骨干网络 (Backbone) 输出基础特征, 部分经AIFI模块增强. CCFM先初步融合多尺度特征, 再将其与AIFI增强特征、骨干网络部分特征, 一同输入DFM模块. DFM通过密集连接, 让不同来源、尺度特征充分交互, 最后聚合输出, 助力模型更好检测火焰. 其中核心逻辑流程可以用公式 (2) ~ (5) 表达.

$$F_{\text{backbone}} = \{ F_{\text{backbone}}^1, F_{\text{backbone}}^2, F_{\text{backbone}}^3 \}, \quad (2)$$

$$F_{\text{aifi}} = AIFI(F_{\text{backbone}}^k), \quad (3)$$

$$F_{\text{dfm}}^i = \text{Conv}_{3 \times 3} \left(F_i^o + \sum_{j=1, j \neq i}^N \text{UP/DOWN}(F_j^o) \right), \quad (4)$$

$$F_{\text{OUT}} = DFM(CCFM(F_{\text{backbone}}^k)) \cup AIFI(F_{\text{backbone}}^k). \quad (5)$$

其中, F_{backbone}^k 表示从骨干网络 (Backbone) 中输出的第 k 级 (或第 k 种尺度) 特征. AIFI 表示 AIFI 模块的特征变换; Up/Down() 表示根据 F_{dfm}^i 分辨率, 对 F_j^o 上采样或下采样, 确保尺寸对齐; $\cup$ 表示多源特征 (CCFM 处理后特征加 AIFI 增强特征) 的集合与交互, 最后通过 DFM() 操作, 实现了从多尺度对齐到密集残差融合到卷积增强.

1.3 高效多尺度注意力 (EMA) 模块

RT-DETR 的 Transformer 编码器采用自注意力机制建模全局特征关系, 但在复杂背景中, 对火焰区域的关注易被干扰 (如烟雾、红色物体). 李泽森等^[16] 在主干网络部位添加 EMA 模块, 提升了模型精度, mAP50 (IoU 阈值为 50% 时模型的平均精度均值) 提升了 6.3% 本研究引入 EMA (Efficient Multi-Scale Attention) 注意力机制, 优化特征图的通道和空间注意力分配. 在火焰检测中, EMA 能通过跨空间学习聚焦火焰区域 (如高温橙红色区域), 抑制背景干扰 (如灰色烟雾). 将 EMA 模块嵌入 RT-DETR 的 Transformer 编码器中, 替换原自注意力模块的部分计算过程, 可增强模型对火焰特征的响应强度. EMA 的网络结构图如图 4 所示.

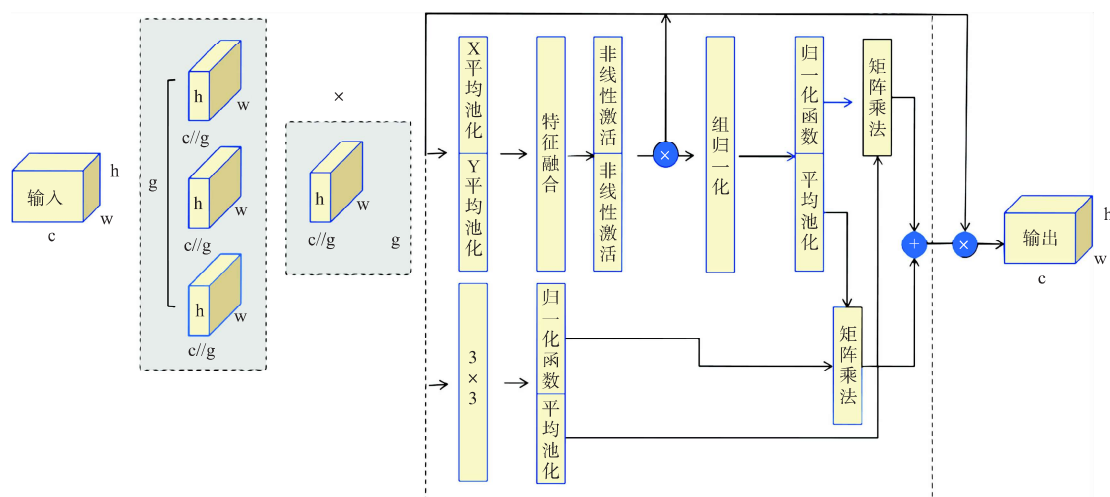


图 4 EMA 网络结构

EMA 的工作流程为：先在通道维度将输入特征图划分为 G 个子特征以降低计算复杂度，接着采用三个并行分支处理子特征，即两个 1×1 卷积分支分别沿水平和垂直方向做全局平均池化来捕获空间方向通道信息、一个 3×3 卷积分支提取多尺度空间特征，之后将三个分支输出拼接，经 1×1 卷积生成注意力权重，再与原始特征图相乘完成特征增强，EMA 注意力权重更新如公式 (6) 所示。

$$\alpha_{EMA,t} = \beta \cdot \alpha_{EMA,t-1} + (1 - \beta) \cdot \alpha_t. \quad (6)$$

$\alpha_{EMA,t}$ 表示第 t 时刻经 EMA 平滑后的注意力权重； $\alpha_{EMA,t-1}$ 表示第 $t-1$ 时刻的 EMA 平滑权重； α_t 表示第 t 时刻的原始注意力权重； β 为平滑系数。将 EMA 注意力机制嵌入到 RT-DETR Transformer 编码器的过程如图 5 所示。

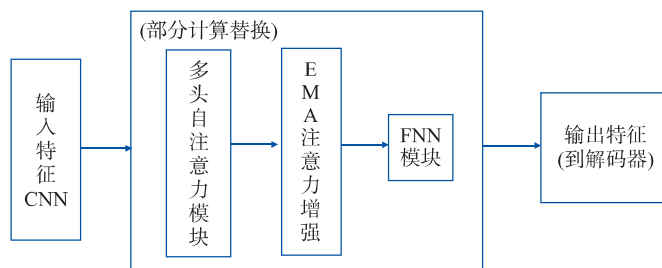


图 5 EMA 融合 Transformer 编码器过程

2 实验结果及分析

2.1 实验环境及超参设置

本实验在 Windows10 系统上运行，CPU 为 12 核 Intel (R) Xeon (R) Platinum 8352 V @ 2.10 GHz，GPU 为 NVIDIA GeForce RTX 4060 的高性能显卡，GPU 的内存大小是 8 GB，深度学习框架为 PyTorch 1.12，编程语言为 Python 3.8，GPU 计算环境为 CUDA 11.8，所有输入图像的分辨率统一转换为 640×640 (像素)。在深度学习领域，超参数特指那些在模型训练启动前就需要预先定义的配置参数，它们与模型在训练过程中自动学习得到的参数存在本质区别。超参数的设置直接影响模型的学习效率与最终性能，因此需要通过系统优化来确定最佳取值。表 1 详细列出了本研究模型训练过程中涉及的关键超参数设置。

2.2 实验数据集

本研究从多个渠道收集火灾的图像共 7 000 张，构建了一个全面的数据集，特别关注了各地区，各种地形环境和各种天气条件下的火灾情况。这些图像主要来源于新闻报道截图、消防队员提供的现场照片，以及社会各界人士在实地收集的图片，数据集示例如图 6 所示。

表 1 模型训练过程中的超参数

类型	超参数	说明
Epoch	300	训练周期
Batchsize	32	批量大小
Learning rate1	0.01	初始学习率
Learning rate2	0.000 1	最小学习率
Workers	4	线程数

数据收集后, 被分为训练集、验证集和测试集. 训练集专用于模型训练阶段, 通过大量样本使模型充分学习数据特征; 验证集作为模型调优的关键环节, 既用于实时评估模型表现, 又辅助超参数优化, 同时计算精确率、召回率等核心评估指标, 以检验模型对新数据的适应能力; 测试集用于在模型训练和调优完成后, 对模型的最终泛化性能进行独立、客观地评价. 其构建旨在模拟真实应用场景中模型未曾受过训练的数据, 用于检验模型在火焰检测任务中的实际表现, 包括检测精度、鲁棒性和误检/漏检情况. 测试集不参与任何训练或调参过程, 确保评估结果公正、可信, 是衡量模型是否具备部署条件的关键依据.



(a) 夜间森林大火

(b) 日间森林火点

图6 数据集示例

2.3 实验评估指标

本研究从平均精度值 ($mAP@0.5$)、参数量 (M)、计算量、帧率 (f/s) 四个维度评估模型性能, 其中参数量的单位 M 表示 Million (百万). $mAP@0.5$ 以 IoU 阈值 0.5 衡量检测精度; 参数量反映模型复杂度与部署可行性; 计算量 (GFLOPs) 量化前向传播浮点运算次数; 内存大小 (MB) 体现存储与推理内存需求. 通过综合评估, 追求模型精度与效率最优平衡, 适配不同场景需求. 公式 (7) ~ (9) 分别展示了单个类别精度、平均精度和帧率.

$$AP_i = \int_0^1 P_i(R) dR, \quad (7)$$

$$mAP = \frac{\sum_{i=1}^Q Q_i AP_i}{Q} \times 100\%, \quad (8)$$

$$FPS = \frac{1\ 000}{t}. \quad (9)$$

AP_i 表示 (第 i 类的平均精度) 通过该类真正例 (TP 表示正确检测的目标数) 与总预测数 ($TP + FP$ 是真正例加假正例, 即错误检测为该类的数量) 的比值, 结合召回率等计算曲线下面积得到; mAP (平均精度均值) 是将所有 Q 个类别各自的 (AP_i) 求和后取平均并转化为百分比, 衡量多类别任务整体性能; FPS (帧率) 则是用 1 000 除以推理时间 (t , 通常指处理单张图像的时间), 反映模型处理图像的实时性, 数值越高处理速度越快.

2.4 结果分析

2.4.1 消融实验

目标识别中的消融实验, 是通过有计划地去除 (或 “消融”) 模型的某个组件 (如特定的特征提取模块、损失函数的部分设计、数据增强策略等)、结构或功能, 对比消融前后模型在目标识别任务上的性能变化 (常以 mAP 、参数量、计算量等指标衡量), 以此精准定位该部分对模型效果的贡献与影响.

表 2 为 RT-DETR 消融实验结果, 通过控制变量法探究不同模块对模型性能的影响. 实验设置 4 组对比: 在模型改进前时, $mAP50$ 为 84.2%、参数量 25.6M、GFLOPs 38.5、帧率 30.1 f/s ; 替换骨干为 StarNet 后, 参数量降至 18.5 M, GFLOPs 与帧率优化, $mAP50$ 几乎不变; 在此基础上加入 DFM 模块, 多尺度特征融合使 $mAP50$ 达 84.9%, 但因计算量增加, GFLOPs 上升、帧率下降; 再嵌入 EMA 模块强化特征, $mAP50$ 进一步到 85.3%, 参数量、GFLOPs 小幅增长, 帧率继续降低. 结果表明, StarNet、DFM、

EMA 模块协同可逐步提升火焰检测精度，同时需权衡计算成本与推理速度.

表 2 RT-DETR 消融实验表

模型	StarNet	DFM	EMA	mAP50/%	参数量/M	计算量 (GFLOPs)	帧率/(f · s ⁻¹)
RT-DETR	×	×	×	84.2	25.6	38.5	30.1
	√	×	×	84.2	18.5	25.6	38.2
	√	√	×	84.9	19.2	27.8	35.2
	√	√	√	85.3	19.5	29.1	33.7

2.4.2 对比实验

本实验将原始版本 RT-DETR 以及改进后的 RT-DETR 与其他主流的目标检测算法进行了横向对比，从平均精度、参数量、计算量、帧率四个性能指标来综合评估模型的性能，对比结果如表 3 所示.

表 3 RT-DETR 对比实验表

模型	mAP50/%	参数量/M	计算量(GFLOPs)	帧率/(f · s ⁻¹)
RT-DETR(原始)	84.2	25.6	38.5	30.1
YOLOv9	83.5	22.1	41.3	28.4
SSD-lite	65.4	3.9	8.2	105.5
Faster-RCNN	79.8	32.5	45.6	22.7
RT-DETR(改进后)	85.3	19.5	29.1	33.7

表 3 为 RT-DETR 对比实验结果. RT-DETR（原始）mAP50 达 84.2%、参数量为 25.6M、GFLOPs 38.5、帧率 30.1；YOLOv9^[17]的参数量比 RT-DETR（原始）要低，但是在精度上低于 RT-DETR 原始版，且计算量（GFLOPs）更高；SSD^[18]虽然参数量和计算量很低，但是精度也很低，检测效果不如 RT-DETR（原始）和 YOLOv9；Faster-RCNN 属于两阶段模型，具有较大的参数量和计算量，同时它的帧率要低于表中的前三个模型；RT-DETR（改进后）经优化，mAP50 提升至 85.3%，参数量降至 19.5M、GFLOPs 为 29.1，帧率 33.7 f/s，无论是精度还是速度都有提升，实现了精度与轻量化的平衡.

3 结 语

本研究聚焦于火焰检测领域，针对复杂场景下火焰检测的挑战，对 RT-DETR 模型展开深入优化探索. 通过一系列创新改进，如以 StarNet 替换骨干网络、引入 DFM 模块以及嵌入 EMA 模块，成功实现了模型性能的显著提升. 实验结果清晰表明，改进后的 RT-DETR 模型在精度、参数量、计算量以及推理速度等多方面达到了更优的平衡. 其 mAP50 从原始的 84.2% 提升至 85.3%，参数量从 25.6M 降至 19.5M，计算量从 38.5 优化到 29.1，帧率从 30.1 f/s 提升至 33.7 f/s. 与其他主流模型相比，改进后的 RT-DETR 模型在精度与效率的平衡上表现出色，为火焰检测任务提供了更具竞争力的解决方案.

随着计算机视觉和深度学习技术的持续进步，火焰检测领域有望在以下方面取得更大突破. 一方面，针对现有模型在极端环境（如强光、浓烟等复杂场景）下的鲁棒性问题，可进一步探索更先进的特征提取与融合方法，增强模型对复杂背景的抗干扰能力；另一方面，在轻量化模型设计上，借助新兴的模型压缩和量化技术，在保持检测精度的同时，进一步降低模型的计算资源需求，以更好地适配资源受限的边缘设备和移动平台.

[参考文献]

[1] 陈刚, 朱方艳, 俞卫国. 仓储火灾隐患与防范对策 [J]. 江苏应急管理, 2025 (3): 39-41.
[2] ZHANG Y, LI X, CHEN Z. Performance analysis of traditional fire detection sensors in complex environments [J]. Journal of Fire Sciences, 2021, 39 (4): 321-338.

- [3] LI Y, WANG X, ZHANG Z. Performance analysis of color threshold-based flame detection methods in smoke-prone environments [J]. IEEE Transactions on Intelligent Transportation Systems, 2023, 24 (5): 4567-4578.
- [4] LE CUN Y, BOSER B, DENKER J S, et al. Handwritten digit recognition with a back-propagation network [M]//TOURETZKY D. Advances in neural information processing systems 2 (NIPS1989). New York: ACM, 1989: 396-404.
- [5] 杜丽霞, 梁晓萍. 基于 Swin Transformer 和 PJF 颜色空间的火灾检测算法 [J]. 现代电子技术, 2025, 48 (5): 147-152.
- [6] 马耀名, 张鹏飞, 谭福生, 等. 面向复杂背景下烟雾火焰检测的改进 YOLOv8s 算法 [J]. 计算机工程与应用, 2025, 61 (3): 121-130.
- [7] GIRSHICK R. Fast R-CNN [C]//2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Boston, MA, USA. IEEE, 2015: 1440-1448.
- [8] ZHAO Y, LV W, XU S, et al. DETRs beat YOLOs on real-time object detection [C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Seattle, WA, USA. IEEE, 2024: 16965-16974.
- [9] 吴栋, 范永胜, 桑彬彬. 基于改进 RT-DETR 的轻量级火焰检测算法 [J]. 计算机与现代化, 2025 (7): 106-111, 118.
- [10] 吴晓宁, 司占军. 基于改进 RT-DETR 的火灾探测模型 [J]. 印刷与数字媒体技术研究, 2024 (4): 107-114.
- [11] KARLINSKY L, SHTOK J, ALFASSY A, et al. StarNet: towards weakly supervised few-shot object detection [J]. Proceedings of the AAAI Conference on Artificial Intelligence, 2021, 35 (2): 1743-1753.
- [12] ZHENG X, WANG Y, YANG B, et al. Decision fusion networks for image classification [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2022, 45 (8): 9291-9305.
- [13] LI X, ZHONG Z, WU J, et al. Expectation-maximization attention networks for semantic segmentation [C]//2019 IEEE/CVF International Conference on Computer Vision (ICCV). Seoul, Korea. IEEE, 2019: 9166-9175.
- [14] HE K, ZHANG X, REN S, et al. Deep residual learning for image recognition [C]//2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Las Vegas, NV, USA. IEEE, 2016: 770-778.
- [15] CHOLLET F. Xception: deep learning with depthwise separable convolutions [C]//2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Honolulu, HI, USA. IEEE, 2017: 1800-1807.
- [16] 李泽森, 徐艳英. 基于改进 YOLOv8 的火灾探测报警系统的设计与实现 [J]. 科技与创新, 2025 (2): 92-95.
- [17] 葛梦莹, 陈广华, 邹新, 等. 基于 YOLOv9c 的隧道火灾检测算法研究 [J]. 国外电子测量技术, 2025, 44 (2): 25-32.
- [18] 严云洋, 杜晨锡, 刘以安, 等. 基于轻型卷积神经网络的火焰检测方法 [J]. 山东大学学报 (工学版), 2020, 50 (2): 100-107.

Research on RT-DETR Flame Detection Model Optimization

XIONG Zijian

(Information Service Department, Sichuan Polytechnic Technician College, Chengdu, Sichuan, China 611130)

Abstract: To address the shortcomings of the RT-DETR model in flame detection tasks, such as insufficient multi-scale feature capture, significant interference from complex backgrounds, and limited inference speed, this paper proposes an improved approach: a multi-scale feature fusion module (DFM) is introduced to enhance scale adaptability, the EMA attention mechanism is employed to optimize feature focusing, and the backbone network is replaced with the lightweight model Starnet to improve real-time performance. Experiments demonstrate that the improved model achieves a 1.1% increase in mAP and an inference speed of 33.7 f/s, outperforming both the original model and other mainstream comparative models. It thus meets the real-time detection requirements in industrial fire protection and forest fire prevention scenarios.

Key words: fire detection; RT-DETR; StarNet; feature fusion; attention mechanism

(责任编辑: 陈伟超)