基于 YOLOv7-tiny 的轻量化火焰检测算法

朱康跃 [1]
ZHU Kangyue

摘 要 针对目前火焰检测算法具有参数量大、检测效果差等问题，提出一种基于 YOLOv7-tiny 的火焰检测算法。首先，为减少模型的参数量和计算量，在 Backbone 层和 Head 层分别融入 PConv 和 GSConv。然后，为降低轻量化网络造成的特征损失，引入 CBAM 注意力机制。实验结果表明，改进后的参数量和计算量分别下降 47.5% 和 45%，而 mAP 提升 2.1%。结果表明，改进后算法在参数量、计算量、检测精度和检测速度等方面具有明显优势。

关键词 火焰检测；YOLOv7-tiny；轻量化；注意力机制

doi：10.3969/j.issn.1672-9528.2024.08.009

0 引言

火灾具有突发性、传播速度快、破坏性强等特点 [1]。火灾一旦发生，若不能及时发现并扑灭，就容易造成严重的人员伤亡和财产损失 [2]。针对火灾频发，如果能提高火焰检测算法性能，更好地预防火灾，则能充分保障人民群众的生命财产安全。

作为火灾早期预警手段所依据的一种信息，火灾场景中的火焰由于具有鲜明的可视特性，与周围的环境易区分等特点，在火灾早期预警中具有突出的优势，已成为国内外研究热点。

基于深度学习的火焰检测方法是目前火焰检测研究的热点领域。近几年，由于计算机硬件性能的提升，目标检测算法取得一系列重大突破，例如 SSD[3]（single shot multibox detector）、YOLO[4-6]（you only look once） 系 列 算 法。YOLO 系列中多个版本用于火焰检测。如张为等人 [7] 基于 YOLOv3 改进，采用 DenseNet 结构和空洞卷积模块改进 YOLOv3 火焰检测算法，并与许多传统算法在相同的数据集进行实验对比。王冠博等人 [8] 提出一种基于 YOLOv4-tiny 的火焰实时检测方法，首先扩大网络模型的感受野，再生成具有更高分辨率的特征，最后对生成的多重感受野进行融合，最终 FPS 达到 49.6。Li 等人 [9] 利用 MobileNetv3 改进的 YOLOv4 主干网络，提出一种 Yolo-Edge 火焰检测算法，显著降低了模型的内存占用，达到良好的轻量化效果。卞苏阳等人 [10] 提出一种基于 CXANet-YOLO 的火焰检测方法，利用 XSepConv、大卷积核、Mish 激活函数作为 YOLOv5 的主干网络，并引入 CBAM 注意力机制，提高模型的鲁棒性和泛化能力。常丽等人 [11] 提出一种基于 YOLOv5s 的实时火焰检测算法，首先利用 K-meansv 聚类算法降低误检率，与 SRGAN 模型结合，再引入 CBAM 注意力机制和梯度均衡机制，解决了难易样本和正负样本不平衡的问题。Lian 等人 [12] 提出基于 YOLOv7 的火灾烟雾检测算法，将部分卷积层整合到 YOLOv7 的 E-ELAN 模块中，同时引入 Focal-Eiou 损失函数，以实现对火焰的快速精准检测。

本文对 YOLOv7-tiny 进行改进，在大幅减少模型参数量的同时保证检测的准确率和实时性。主要改进包括以下三点。

（1）在 Backbone 层融入 PConv，降低模型的参数量和计算量。

（2）在 Head 部分融入 GSConv，进一步轻量化的同时提高检测精度。

（3）为减少轻量化网络造成的特征损失，引入注意力机制 CBAM 对特征进行进一步提取，提高模型对火焰的检测准确率。

1 YOLOv7-tiny 网络结构与原理

YOLOv7-tiny 算法是由 YOLOv7 算法的一个变体在 YOLOv7 算法基础上简化而来的，是一种小型网络，保留了基于级联的模型缩放策略，在一定程度上保证了模型的检测精度、参数量和检测速度，适应于火焰检测场景。本文以 YOLOv7-tiny 算法为基础模型进行改进，使其火焰检测效果更好。YOLOv7-tiny 网络结构主要由 4 个部分组成，分别为：输入端（Input）、主干网络（Backbone）、颈部特征融合网络（Neck）、输出端（Output）。

输入端（Input）功能是对输入图片进行预处理。输入端的预处理主要采用了 mosaic 数据增强方法与自适应锚框等方

1. 三峡大学计算机与信息学院 湖北宜昌 443002

式，将四张图片随机进行缩放、剪裁、拼接，丰富数据集的同时确保图片被处理为统一的尺寸，从而满足特征提取网络的输入要求。

主干特征提取网络（Backbone）主要由 CBL 卷积块、MPConv 模块和 ELAN 模块 3 个部分组成。CBL 卷积块提取原始特征，由 Conv 层、BN 层和激活函数 LeakyReLU 构成；MPConv 模块用于下采样，由两个分支构成：一个分支是由最大池化和 CBL 模块构成，另一个分支是由两个 CBL 模块构成。ELAN 模块利用大量的普通卷积来学习原始的特征。ELAN 层对原 YOLOv7 中的两个特征块进行剔除，在加快特征提取速度的同时，也导致了特征提取性能的下降。

颈部特征融合网络（Neck）采用特征金字塔（PAFPN）网络结构，由大量的 CBL 模块、ELAN 模块和上采样模块组成，将特征金字塔网络顶层的强语义信息与路径聚合网络自下而上传递的强定位信息张量相结合，通过特征信息融合实现多尺度学习。YOLOv7-tiny 的特征融合网络中的张量拼接对于融合相邻层的特征信息不够全面，同时融合特征没有优先考虑目标特征信息，会导致特征信息提取的丢失，从而导致检测精度下降。

输出端（Output）采用输出大中小三种目标尺寸的检测头，分别输出三种大小的像素特征图，最后利用特征图对不同尺寸的火焰目标进行预测。检测头使用了普通卷积，使得特征融合的结果并未集中于预测目标，且缺少有针对性的方法提高小目标检测性能。

本文面向火焰检测问题，围绕 YOLOv7-tiny 算法，降低网络的参数量和计算量，提高网络的特征提取能力，为最后提高模型检测性能做出改进。

2 改进 YOLOv7-tiny 火焰检测算法

2.1 ELAN-PC 结构

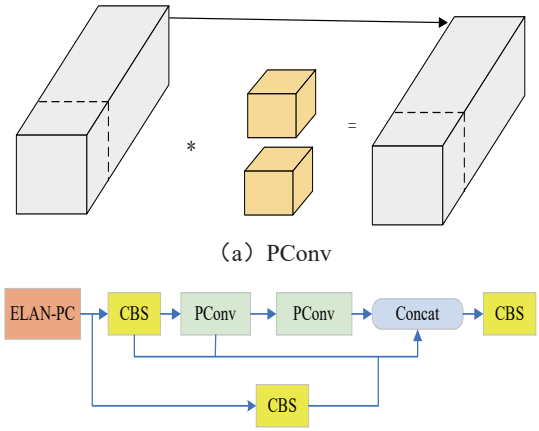
如何降低浮点操作的运算次数，是实现快速神经网络设计的关键。但是，浮点操作次数的降低并不意味着能带来延迟的减少，这主要是由于浮点运算的低效率所致。为提高网络的速度，将 PConv 卷积 [13] 模块引入骨干网络中。

使用 PConv 可以更好地减少卷积过程中冗余计算和内存访问次数，更好地平衡检测延迟（Latency）和浮点运算（FLOPs），它们之间的公式为：

$$\text{Latency}=\frac{\text{FLOPs}}{\text{FLOPS}} \tag{1}$$

式中：FLOPS 指的是浮点每秒运算次数。PConv 卷积降低 FLOPs 的同时优化 FLOPS，它只需要在输入通道的一部分上应用常规 Conv 进行空间特征提取，并保持其余通道数不变。工作原理如图 1（a）所示。

PConv 相对于常规卷积，更好地利用了设备的计算能力，在特征空间提取上也很有效。本文在 Backbone 部分中的 ELAN 引入 PConv 卷积模块，得到 ELAN-PC 结构，降低算法的参数量和计算量，以达到轻量化目的。其结构如图 1（b）所示。



（a）PConv

（b）ELAN-PC

图 1 PConv 原理图和 ELAN-PC 结构图

2.2 ELAN-GC 结构

GSConv 原理为：设输入的通道数为 C_1，输出通道数为 C_2。首先经过一个标准卷积，通道数变为 $C_2/2$，再经过一个深度可分离卷积，通道数不变，最后将两次卷积的结果进行拼接和混洗。最后的混洗操作，能够将通道信息进行均匀打乱，增强提取到的语义信息，提高图像特征的表达能力。GSConv 结构如图 2 所示。

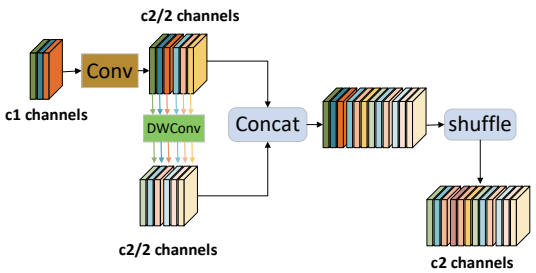


图 2 GSConv 模块

如果在网络结构的每个阶段使用 GSConv 模块，网络层会更深，深层会加剧对数据流的阻力，显著增加推理时间，增加网络参数量和计算量。因此，本文在 YOLOv7-tiny 网络的 Output 部分引用 GSConv 卷积模块，替换卷积核为 3 网络层的标准卷积进行上采样和下采样，同时改进 Neck 部分的 ELAN 结构，降低模型的参数量和计算量，有效提取特征值，提高检测效果，改进后结构为 ELAN-GC，如图 3 所示。

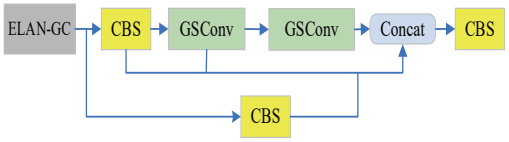


图 3 ELAN-GC

2.3 CBAM 注意力模块

注意力机制是通过神经网络自主学习得到权重，根据权重使网络更多地关注重要信息，减少无用信息的影响，提高网络的性能。

CBAM 注意力机制 [14] 是一种简单、高效且轻量级的注意力机制，由通道注意力机制（CAM）和空间注意力机制（SAM）组成，沿着空间和通道两个维度推断特征图，同时使用全局平均池化和全局最大池化，进行特征提取，起到防止信息丢失的作用。CBAM 注意力机制中的通道维度集中于输入图像的相关特征，空间维度侧重于输入图像的位置信息。CBAM 注意力机制同时关注了空间特征和通道特征，可以提高特征提取能力。其结构图如图 4 所示。

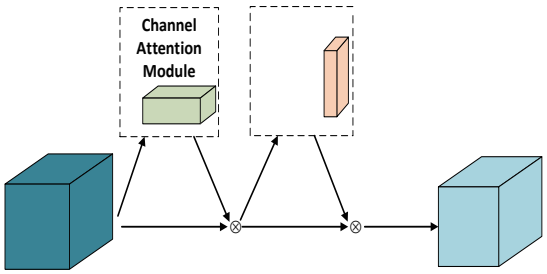


图 4 CBAM 注意力机制结构图

上述在 Backbone 部分，使用 PConv 对 ELAN 部分进行改进。由于 PConv 在对参数量和计算量进行优化的同时，丢失了部分特征，随着网络层数的增加，特征丢失更加严重。因此，本文在每个 ELAN-PC 后添加 CBAM 注意力机制，提高特征提取能力，以提高对火焰的检测精度。

3 实验结果与分析

3.1 实验数据集与实验环境

实验数据集为自建火焰数据集，数据集收集于互联网，共有图片 5500 张，训练集有 5000 张，验证集 500 张，并将图片大小统一设置为 640×640。使用 Labelimg 标注软件对数据集进行标注，数据集包含多种特征明显的火焰，含有建筑火灾、森林火灾等多种背景，基本包含日常生活中的大部分火灾场景，满足火焰检测的场景需求。

实验环境为 Windows 10（64 位）操作系统，CPU 为 i7-12700H，显卡为 NVIDIA GeForce RTX 3060，显存 6 GB，采用 PyTorch2.0.0 框架运行代码，CUDA11.6.134。

输入图片大小 Imgsize 为 640×640，批处理大小为 16，训练 300 个 Epoch，初始学习率为 0.01，momentum 为 0.973，Weight_decay（权重衰减）为 0.000 5。

3.2 评价指标

本文使用的评价指标有平均准确均值 mAP（Mean Average Precision）、参数量 Params、计算量 FLOPs 三种，分别介绍如下。

（1）平均准确均值 mAP 计算公式为：

$$\mathrm{mAP}=\frac{1}{N}\sum_{i=1}^{n}\mathrm{AP}_i\times 100\% \qquad (2)$$

式中：AP（average precision）是单一类别识别的平均精度；mAP 是反映不同类别的平均准确率并产生单个分数，mAP 越高表示模型对目标的检测越准确。

（2）Params 反映模型参数内存占用内存字节数，为模型轻量化指标，单位为 MB。

（3）FLOPs 是模型的计算量，反映模型复杂程度，单位为 10^9 个。

3.3 消融实验

为了验证本文改进方案的效果和程度，设计了本消融实验来验证每种方案的有效性。以改进前的 YOLOv7-tiny 网络模型为基准模型，实验的数据集、实验环境以及配置参数不变。

实验 A 是在模型的 Backbone 部分引入 PConv 模块进行轻量化；实验 B 是使用 GSConv 对模型的 Head 部分进行改进；实验 C 是使用 CBAM 注意力机制对 Backbone 部分的 ELAN 进行特征提取；消融实验结果如表 1 所示。

表 1 消融实验结果

排序	模型	mAP@0.5 /%	Params /MB	FLOPs /10^9 个
N1	YOLOv7-tiny	86.3	5.7	13.0
N2	YOLOv7-tiny+A	85.3	4.5	10.1
N3	YOLOv7-tiny+A+B	87.2	3.3	7.8
N4	YOLOv7-tiny+A+B+C	88.4	3.0	7.4

（1）由 N2 所示，在 Backbone 使用 ELAN-PC 模块，准确率下降 1%，参数量和计算量分别下降 21% 和 22.3%，验证 ELAN-PC 在轻量化的有效性，但是会造成准确率下降。

（2）在 Head 中使用 ELAN-GC 模块，如 N3 所示，mAP 提升 1.9%，参数量降低 42.1%，计算量降低 40%，

（3）由 N4 所示，在 ELAN-PC 后引入 CBAM 注意力机制，mAP 上涨 1.2%，参数量和计算量分别下降 5.3% 和 2.7%，在提高火焰检测准确率的同时，有一定的轻量化效果。

综上所述，改进后 YOLOv7-tiny 模型在火焰检测上，mAP 提升 2.1%，参数量和计算量分别下降 47% 和 43%。在轻量化的同时，提高了模型的准确率，满足火焰检测。

3.4 模型检测效果

为了更加直观地对比改进后网络模型和原 YOLOv7-tiny 网络模型的检测效果，采取随机抽样的方式在验证集中选取多张图片，如图 5 所示。其中，第 1 行是 YOLOv7-tiny 检测

的结果，第 2 行是改进后检测的结果。从图 5 可以看出，原模型对火焰检测的置信度不高。改进后检测结果中准确率明显提高。

图 5 改进前后检测效果对比

4 结论

为解决火焰检测算法参数大、检测效果差等问题，本文提出基于 YOLOv7-tiny 的火焰检测算法。首先使用 PConv 和 GSConv 实现轻量化网络，降低算法的参数量和计算量；然后引入 CBAM 注意力机制来减少轻量化网络造成的特征损失，提高模型检测精度，同时进一步降低参数量。实验结果表明，相较于原 YOLOv7-tiny 算法，改进后 YOLOv7-tiny 算法在降低参数量的同时，提升了火焰检测精度。与其他目标检测模型相比，改进 YOLOv7-tiny 具有较好的性能。由于数据集的限制，YOLOv7-tiny 只针对火焰检测，后续将继续扩充带有烟雾的数据集，并进一步研究如何提高模型的检测精度。

参考文献：

[1]XAVIER K L B L, NANAYAKKARA V K.Development of an early fire detection technique using a passive infrared sensor and deep neural networks[J].Fire technology, 2022, 58(6): 3529-3552.

[2]BEYENE T, MURPHY V E, GIBSON P G, et al.The impact of prolonged landscape fire smoke exposure on women with asthma in Australia[J].BMC pregnancy and childbirth, 2022, 22(1): 254491437.

[3]WEI L, DRAGOMIR A, DUMITRU E, et al.SSD: single shot multibox detector[EB/OL].(2016-12-29)[2024-03-02].https://arxiv.org/abs/1512.02325.

[4]SHI Q, LI C, GUO B, et al.Manipulator-based autonomous inspections at road checkpoints: application of faster YOLO for detecting large objects[J].Defence technology, 2022, 18(6): 937-951.

[5]REDMON J, FARHADI A.YOLOv3: an incremental improvement[EB/OL].(2018-04-08)[2024-03-02].https://arxiv.org/abs/1804.02767.

[6]WANG C, BOCHKOVSKIY A, LIAO H M.YOLOv7: trainable bag-of-freebies sets new state-of-the-art for real-time object detectors[EB/OL].(2022-07-06)[2024-03-05].https://arxiv.org/abs/2207.02696.

[7] 张为，魏晶晶 . 嵌入 DenseNet 结构和空洞卷积模块的改进 YOLOv3 火灾检测算法 [J]. 天津大学学报 (自然科学与工程技术版),2020,53(9):976.

[8] 王冠博，赵一帆，李波，等 . 改进 YOLOv4-tiny 的火焰实时检测 [J]. 计算机工程与科学 ,2022,44(12):2196-2205.

[9]LI W, YU Z.A lightweight convolutional neural network flame detection algorithm[C]//2021 IEEE 11th International Conference on Electronics Information and Emergency Communication.Piscataway:IEEE,2021:83-86.

[10] 卞苏阳，严云洋，龚成张，等 . 基于 CXANet-YOLO 的火焰检测方法 [J]. 南京大学学报 (自然科学),2023,59(2):295-301.

[11] 常丽，张雪，蒋辉，等 . 融合 YOLOv5s 与 SRGAN 的实时隧道火灾检测 [J]. 电子测量与仪器学报 ,2022,36(8):223-230.

[12]LIAN J, PAN X, GUO J.An improved fire and smoke detection method based on YOLOv7[C]//2023 32nd International Conference on Computer Communications and Networks (ICCCN).Piscataway:IEEE,2023:1-7.

[13]CHEN J, KAO S, HE H, et al.Run, don't walk: chasing higher FLOPS for faster neural networks[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway:IEEE,2023:12021-12031.

[14]WOO S, PARK J, LEE J Y, et al.CBAM: convolutional block attention module[EB/OL].(2018-07-17)[2024-03-05].https://arxiv.org/abs/1807.06521.

【作者简介】

朱康跃（2000—），男，江苏徐州人，硕士研究生，研究方向：目标检测。

（收稿日期：2024-05-16）