

基于 YOLOv3 改进的火焰目标检测算法

张彬彬, 帕孜来·马合木提*

新疆大学电气工程学院, 新疆 乌鲁木齐 830047

摘要 针对火灾中小目标、多目标、边缘模糊等导致的检测精度下降问题, 提出了一种基于 YOLOv3 (You only look once, v3) 改进的火焰目标检测算法。首先, 通过改进的特征金字塔网络对局部信息进行二次利用。然后, 设计大尺度全卷积模块以获取不同尺度的全局空间信息, 用改进的通道空间注意力机制提高有效信息, 并抑制无用信息。最后, 用完全交并比和 Focal Loss 作为损失函数, 以提高难识别目标的检测精度和缓解数据集不平衡的问题。在自制火焰数据中的实验结果表明, 本算法具有较高的检测精度和较快的检测速度, 平均精度均值高达 89.82%, 检测速度可达到 20.2FPS (每秒传输帧数), 能满足火灾检测的实时性和高效性要求。

关键词 成像系统; 目标检测; 特征金字塔; 全卷积模块; 注意力机制; 损失函数

中图分类号 TP391.4

文献标志码 A

doi: 10.3788/LOP202158.2411003

Improved Flame Target Detection Algorithm Based on YOLOv3

Zhang Binbin, Pazilai·Mahemuti*

School of Electrical Engineering, Xinjiang University, Urumqi, Xinjiang 830047, China

Abstract This paper proposes an improved flame target detection algorithm based on YOLOv3 (You only look once, v3) to solve the causes of reduced detection accuracy caused by fire on small and medium-sized target, multi-target, and fuzzy edge. First, the improved feature pyramid network makes use of local information twice. Then, a large-scale full convolution module is designed to obtain global spatial information of various scales, and an improved channel space attention mechanism is used to improve effective information and suppress useless information. Finally, as loss functions, complete intersection-over-union and Focal Loss are used to improve the detection accuracy of difficult-to-recognise targets and alleviate the problem of data set imbalance. Experimental results in self-made flame data show that this algorithm has higher detection accuracy and faster detection speed. The average accuracy is up to 89.82%, and the detection speed can reach 20.2FPS (Frames per second), enabling it to meet real-time and high-efficiency fire detection requirements.

Key words imaging systems; target detection; feature pyramid; full convolution module; attention mechanism; loss function

OCIS codes 110.2970; 100.3008; 110.2960; 150.1135

1 引言

火灾是人类面临的主要自然灾害之一, 2018 年美国大约发生了 1318500 起火灾, 直接导致 720 人死亡和 256 亿美元损失。2019 年四川凉山发生的森林火灾造成了 31 人死亡以及大量财产损失, 且森

林灭火工作难度大于城市灭火工作。因此, 对于火灾早期的有效预防、预警和应对方法变得尤为重要。

早期火灾预警大多采用传统的火灾探测器, 即通过火灾产生的烟雾、温度或其他因素组合触发探测器并启动预警和灭火系统^[1-4]。这种探测方法受多种因素影响, 响应速度慢, 且存在漏报、误报等问

收稿日期: 2021-01-20; 修回日期: 2021-02-05; 录用日期: 2021-03-04

基金项目: 新疆维吾尔自治区自然科学基金(2016D01C038)

通信作者: *294625876@qq.com

题,可能会导致火势迅速蔓延,甚至无法控制。机器学习的发展使通过监控视频或图像数据进行火灾探测成为可能^[5],但该方法的效果取决于人工选取特征的好坏^[6]。因火灾类型难以界定,且受边缘模糊和环境的影响,难以提取到理想特征,导致该方法存在检测精度低、延迟和计算量大等问题。深度学习方法可将大量标注好的图像输入深层非线性网络中自动学习特征,对图像的变形具有高度不变性,能解决传统机器学习在特征提取中面临的问题,在火灾检测领域中的发展空间极大^[7]。目前,主流检测算法包括以区域卷积神经网络(RCNN)^[8]为代表的两阶段算法和以 YOLO(You only look once)系列网络^[9]为代表的单阶段算法。Zhang 等^[10]设计了一种结合 Faster-RCNN 模型与软件平台的火灾检测算法,检测精度较高,但不能达到实时检测的效果。富雅捷等^[11]提出了一种以 AlexNet 模型为基础,以多个支持向量机(SVM)为分类器的森林火灾检测网络,并用迁移学习训练整个网络模型,取得了较高的检测精度,但 SVM 的二分类特性和数据集背景单一导致该网络的泛化能力较差。Fang 等^[12]基于 Tinier-YOLO 设计了一种微型检测网络,在检测速度方面取得了较大优势。Frizz 等^[13]提出了一种用于识别视频中火灾的卷积神经网络(CNN),将火灾和烟雾的运动信息进行编码,大大降低了时间成本,但其仅在目标分类领域的表现较好。

针对上述火灾检测算法的不足,本文以自制火焰数据集为研究对象,提出了一种改进的 YOLOv3 检测算法。首先,重新设计了特征金字塔网络(FPN),实现了局部信息与全局信息的二次融合。然后,在深层特征提取路径上通过注意力机制模块实现通道和空间维度中的信息交互,使模型能筛选出不重要的信息特征,提高网络对有效信息的提取能力和抗干扰能力。最后,设计了大尺度全卷积模块,使用对称的卷积组合实现大卷积功能,减少了计算量并丰富了特征图的全局信息。在此基础上,用完全交并比(CIoU)损失函数替换坐标误差损失函数,并考虑到目标与预测框的中心距离、重叠率和长宽比信息,提高了算法对难识别图像的检测精度。采用 Focal Loss 替换类别损失函数,缓解了二分类数据集不平衡的问题。实验结果表明,改进后的算法在中小目标、多目标和复杂背景图像上具有较高的检测精度和较快的检测速度。

2 YOLOv3 算法概述

2.1 YOLOv3 的结构

YOLOv3 用 Darknet53 为主干特征提取网络^[14],Darknet53 包含 53 个卷积层,并结合残差网络(ResNet)^[15]的思想加入了残差结构,该结构能使网络进行深层卷积提取特征时避免梯度消失和梯度爆炸问题。 X 个残差模块(X Res)由残差单元(Res unit)堆叠而成,如图 1 所示。其中,DBL 为深度卷积,包含了卷积(Conv)、批标准化(BN)和带泄露修正线性单元(Leaky ReLU)激活函数操作,每次卷积均使用了 L2 正则化处理,以减轻网络训练中出现的过拟合情况。残差单元先将输入特征图进行一次大小为 3×3 、步长为 2 的卷积,并保存该卷积的层(Layer)再进行一次 1×1 的卷积压缩(2 倍)和 3×3 的卷积扩张(2 倍)操作;最后将得到的特征和保存的层进行堆叠。最终的残差模块由 DBL 单元和若干个残差单元组合而成,Darknet53 中分别使用了残差单元数量为 1、2、8、8、4 的 5 个残差模块。

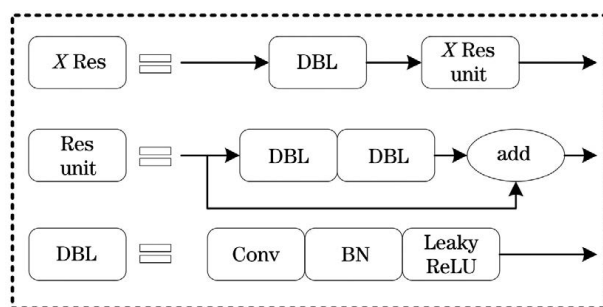


图 1 残差模块的结构

Fig. 1 Structure of the residual module

YOLOv3 网络的整体结构如图 2 所示。首先,将输入图像重塑(Reshape)成尺寸(C, W, H)为(3, 416, 416)的图像,其中, C, W, H 分别表示图像的通道数、长、宽;然后,经过 5 个步长为 2 的残差模块进行下采样,得到 5 个尺度的特征图;最后,用 YOLOv3 网络将后 3 个残差模块得到的特征图输入 FPN 中。其中,尺寸为(1024, 13, 13)的特征经过 5 次二维卷积(Conv2D block 5L)后送到检测网络中进行卷积(压缩通道数和上采样扩充尺寸操作)并与尺寸为(512, 26, 26)的特征图进行融合(Concat)。将融合后尺寸为(768, 26, 26)的特征再次进行 5 次尺寸为(256, 26, 26)的卷积并送到检测网络中。按相同的方法,将尺寸为(256, 26, 26)的特征经过卷积和上采样后与尺寸为(256, 52, 52)的特征图进行融合,再经 5 次卷积后送到检测网络。最终检测网络

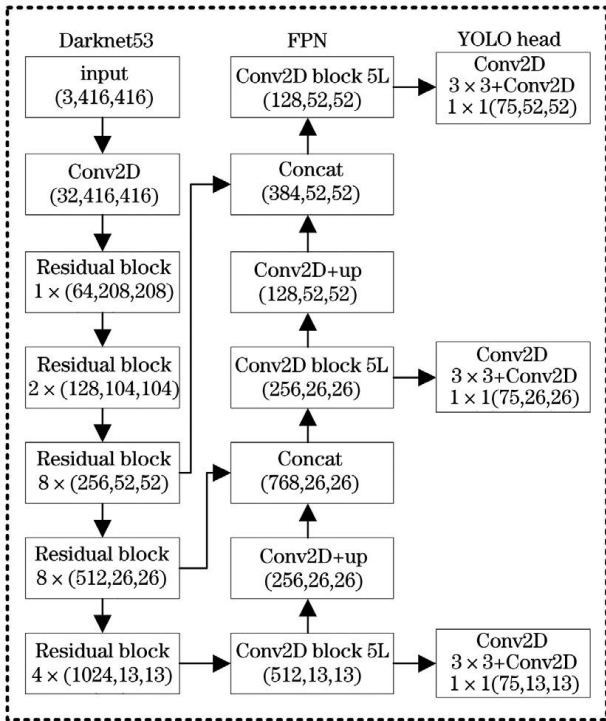


图 2 YOLOv3 算法的结构

Fig. 2 Structure of the YOLOv3 algorithm

得到尺寸为(75, 13, 13)、(75, 26, 26)和(75, 52, 52)的 3 个不同尺度特征进行多尺度检测,分别对应大、中、小物体。采用 YOLOv3 算法检测物体前,先用 K-means 聚类算法根据数据集标注框计算出 9 组锚框(Anchor),并将其平均分配到 3 个不同的尺度上。每个尺度中的每个锚点分别对应 3 组锚框,总计 10647 个锚框,从而得到 3 个特征层先验框的位置^[16]。

2.2 目标边界框预测

YOLOv1 算法的预测框坐标和宽高(x, y, w, h)是直接预测出来的,当坐标出现微小变化时,将其映射到实际图像中会造成很大变化,导致预测失败。针对该问题,YOLOv3 算法对预测框公式进行了改进,可表示为

$$b_x = \sigma(t_x) + c_x, \quad (1)$$

$$b_y = \sigma(t_y) + c_y, \quad (2)$$

$$b_w = p_w \exp(t_w), \quad (3)$$

$$b_h = p_h \exp(t_h), \quad (4)$$

式中, t_x, t_y, t_w, t_h 为网络输出的坐标, p_w, p_h 为锚框的宽、高,如图 3 所示。 c_x, c_y 为图 3 中左上角的位置坐标,图 3 中每个小格的长度为 1, t_x 与 t_y 经过 Sigmoid 函数输出 0~1 之间的偏移量,然后与 c_x, c_y 相加得到预测框的中心点。 p_w 和 p_h 分别与 t_w 和 t_h 作用后得到预测框的宽和高,通过置信度阈

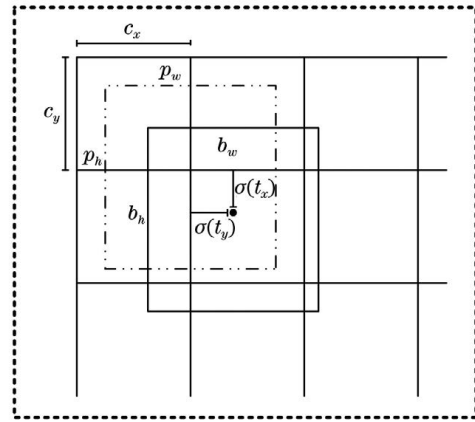


图 3 YOLOv3 算法的锚框预测示意图

Fig. 3 Schematic diagram of anchor prediction of YOLOv3 algorithm

值和非极大值抑制(NMS)处理后得到最终的预测框。相对于 YOLOv1 算法,YOLOv3 算法的边框回归计算方式可以有效防止偏移过大带来的预测失败等问题。

3 改进的 YOLOv3 算法

3.1 大尺度全卷积模块

已有的目标检测网络大多将特征图(Feature map)直接送入特征融合网络,然后对特征图进行多尺度融合。虽然可以减少计算量,但会丢失特征图的部分全局信息。针对该问题,设计了一种大尺度全卷积模块(GCM),在尽可能不增加计算量的基础上,提取不同层级特征图的全局信息,其计算方式如图 4 所示。卷积操作中的卷积核大小会直接影响全局信息的获取,因此,将卷积核大小增加到特征图的空间尺寸,使在原始图像上产生的有效域尽可能地包含目标,从而获取有效的全局信息。但增大卷积核的同时会增加计算量,因此,通过 $K \times 1$ 和 $1 \times K$

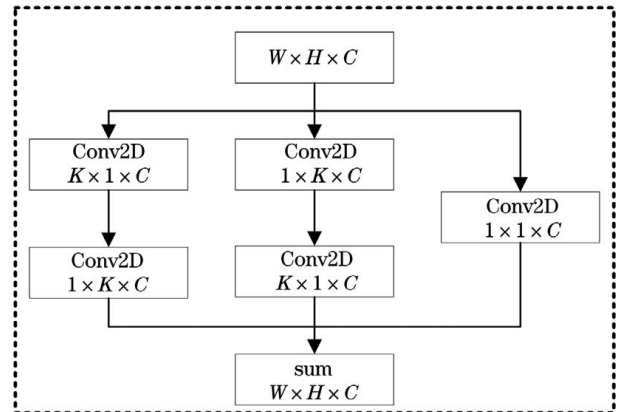


图 4 大尺度 GCM 的结构

Fig. 4 Structure of the large-scale GCM

K 的对称独立卷积组合,在增加少量计算量的基础上实现了尺寸为 $K \times K$ 的大卷积操作。此外,通过取消非线性操作避免了部分特征丢失,并将得到的特征与原始特征进行叠加,丰富了特征的上下文全局信息,同时也能防止网络退化。

3.2 通道空间注意力机制

在有效的通道注意力(ECA)^[17]基础上添加空间注意力,设计出通道空间注意力机制(CSA),其结构如图 5 所示。其中,ECA 模块对特征图进行全局平均池化(GAP),并用大小为 K (局部跨通道交互的覆盖率)的快速 1D 卷积,实现每个通道和 K 个邻近通道的局部跨通道信息交互;再经过 Sigmoid 函数(σ)生成尺寸为 $1 \times 1 \times C$ 的通道注意力分支,并与原始特征图进行点乘。空间注意力部分:通过 1×1 卷积和 Sigmoid 函数,完成空间维度的权重与原始特征图的点乘。最后将两者的结果进行叠加(Add),从而实现通道和空间特征的权重分配。

3.3 改进的 FPN 层

图 2 中的 YOLOv3 算法 FPN 层没有充分利用

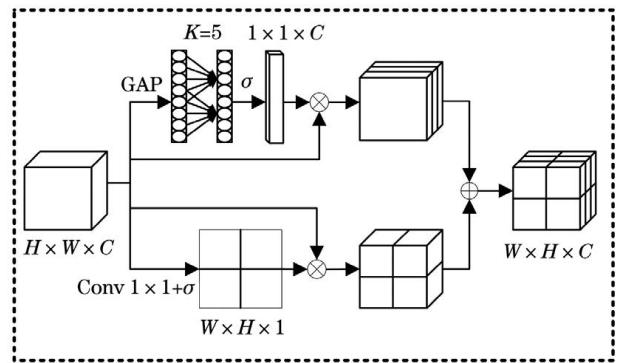


图 5 CSA 模块的结构

Fig. 5 Structure of the CSA module

局部信息,而局部信息有利于小目标的检测,全局信息则更侧重于对大目标的检测。从主干特征提取网络的浅层特征中能提取局部信息,深层特征中能提取全局信息,因此重新设计了 FPN 层,改进 YOLOv3 算法的流程图如图 6 所示。首先,在原始 YOLOv3 网络自下而上的融合策略基础上增加了自上而下的融合路径,并在此路径上通过 CSA 模块提取有效信息,抑制无用信息。同时在 4 个不同层

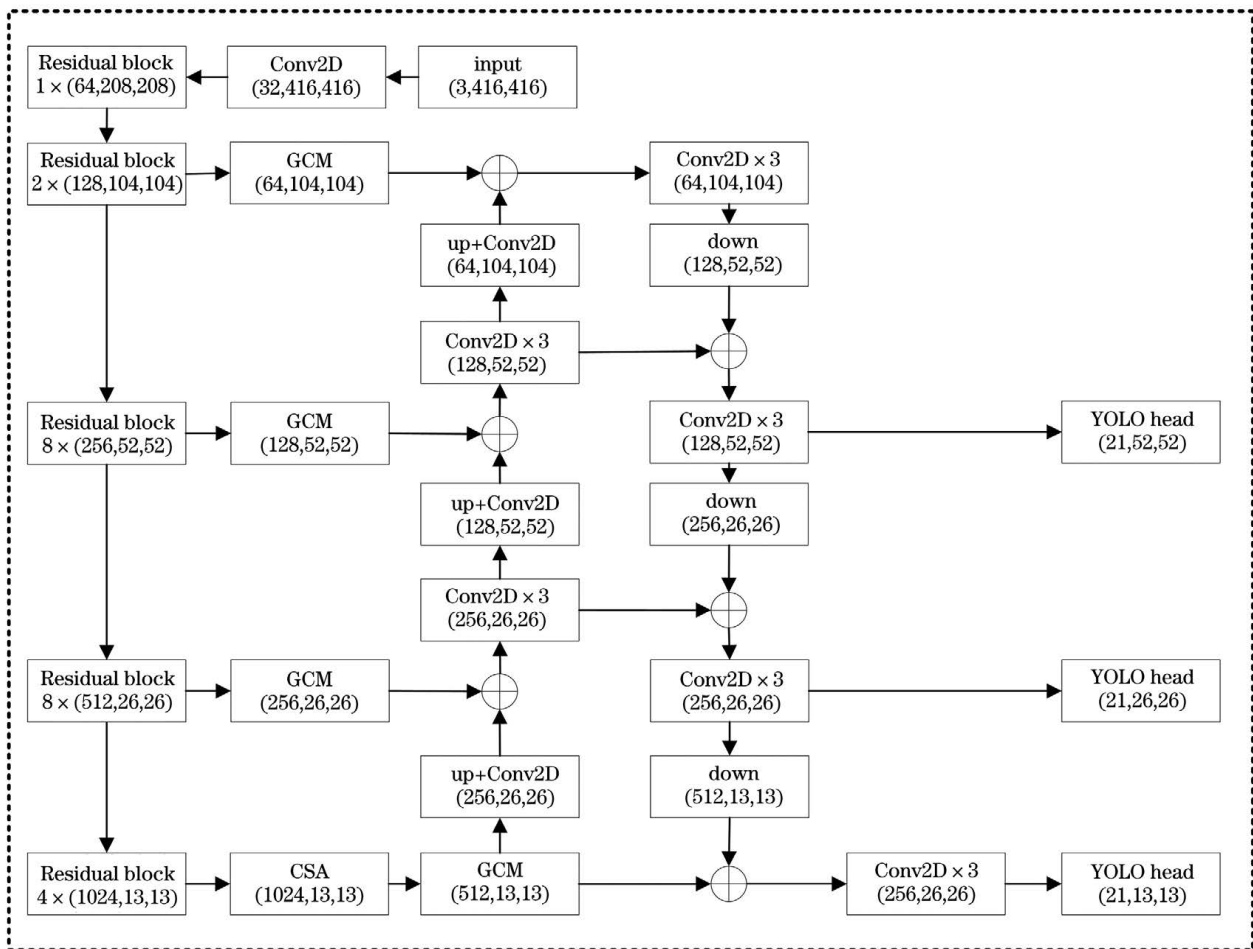


图 6 本算法的网络结构

Fig. 6 Network structure of our algorithm

级的特征图路径上,分别通过 GCM 丰富其全局信息,使最深层的特征图利用率最高。本算法的具体流程如下。

1) 最底层尺寸为(1024, 13, 13)的特征图经 2 倍上采样压缩尺寸和 1×1 的卷积扩充通道后保留原始特征图 1, 将其与尺寸为(256, 26, 26)的特征图进行融合后,再进行 3 次卷积操作,得到特征图 2。

2) 将特征图 2 经过 2 倍上采样和 1×1 的卷积后与尺寸为(128, 52, 52)的特征图进行融合,并经过 3 次卷积后得到特征图 3。

3) 将特征图 3 经过 2 倍上采样和 1×1 卷积与尺寸为(64, 104, 104)的特征图进行融合,并经过 3 次卷积和下采样后与特征图 3 进行融合,得到特征图 4。

4) 将特征图 4 经过下采样与特征图 2 进行融合,同样经过 3 次卷积后得到特征图 5,将其经过下采样与特征图 1 进行融合后,经过 3 次卷积得到特征图 6。最终将特征图 4、特征图 5、特征图 6 送入 YOLO head 得到三个不同尺寸的特征预测层,输出特征图的尺寸分别为(21, 13, 13)、(21, 26, 26)和(21, 52, 52)。

3.4 损失函数

YOLOv3 的损失函数由中心点的坐标和锚框的宽高误差、置信度误差和分类误差构成。其中,坐标误差用真实框与预测框的宽高差方计算,分类误差采用交叉熵损失函数计算。交并比(IoU)是目标检测中常见的评价指标,可表示为

$$L_{IoU} = \frac{|M \cap N|}{|M \cup N|}, \quad (5)$$

式中, M 为预测框坐标, N 为真实框坐标。IoU 可以直接反映真实框与预测框的距离,但在两个框没有相交时,产生的 Loss 值为 0,无法进行反向梯度更新。CIoU 考虑了真实框与预测框中心点的距离、重叠部分和长宽比信息,在训练过程中缓解了目标函数发散等问题,可表示为

$$L_{CIoU} = 1 - L_{IoU} + \frac{\rho^2(b, b^{gt})}{c^2} + \alpha v, \quad (6)$$

$$v = \frac{4}{\pi^2} \left(\arctan \frac{\omega^{gt}}{h^{gt}} - \arctan \frac{\omega}{h} \right)^2, \quad (7)$$

$$\alpha = \frac{v}{1 - L_{IoU} + v}, \quad (8)$$

式中, b, b^{gt} 为预测框与真实框的中心点坐标, ρ^2 为欧氏距离, c 为同时包含预测框与真实框最小闭包区域的对角线距离, v 可衡量长宽比的相似性, α 为权重。

实验使用的自制数据集二分类存在不平衡的问题,因此,分类误差采用二分类的 Focal Loss^[18],相

比交叉熵损失,Focal Loss 加入了平衡因子,可调节正负样本损失,加入的 Gamma 因子可减少易分类样本的损失。Focal Loss 可表示为

$$L_{cls} = \begin{cases} -\alpha(1-\hat{y})^\gamma \log(\hat{y}), & y=1 \\ -(1-\alpha)\hat{y}^\gamma \log(1-\hat{y}), & y=0 \end{cases}, \quad (9)$$

式中, $y \in 0, 1$ 为真实的标签, \hat{y} 为预测值,平衡因子 $\alpha=0.25, \gamma=2$ 。Focal loss 的引入可以有效解决单阶段目标检测中的分类样本不平衡问题。

4 实验结果与分析

4.1 数据集和实验平台

由于目前并没有公开且比较全面的火灾图像数据集,因此,利用人工拍摄的照片和网络图片制作了火焰数据集,以模拟火灾数据,使网络学习到更多的火焰特征。为了提高检测模型的实用性,加入了人造光源作为干扰项,使 CNN 学到多层特征中抽象的数据信息。其中,浅层提取的主要是物体的颜色和边缘信息,深层提取的主要是物体的轮廓信息。实验设备为 I5-9400F, GPU 为 NVIDIA2060S,使用 Pytorch 框架,优化器为 Adam,迭代轮次(Epoch)为 100 轮,初始学习率为 0.001,动量因子为 0.9。最终得到训练集图像 3816 张、测试集图像 562 张且火焰与人造光源图像数量比例为 3:1 的二分类数据集。

4.2 评价指标

用目标检测中常见的精确率(Precision)、召回率(Recall)、平均精度(AP)、平均精度均值(mAP)、F1 分数(F_1)和每秒检测帧数(FPS)评价算法的性能,其中,精确率和召回率可表示为

$$X_{Precision} = \frac{X_{TP}}{X_{TP} + X_{FP}}, \quad (10)$$

$$X_{Recall} = \frac{X_{TP}}{X_{TP} + X_{FN}}, \quad (11)$$

$$F_1 = 2 \times \frac{X_{precision} \times X_{recall}}{X_{precision} + X_{recall}}, \quad (12)$$

式中, X_{TP} 为正样本被正确识别的数量, X_{FP} 为负样本被错误识别为正样本的数量, X_{FN} 为正样本被错误识别为负样本的数量。AP 是不同置信度下精确率-召回率的曲线下面积,可衡量模型对每个类别的检测性能,mAP 则可评价整个模型的性能;F1 分数可作为精确率和召回率的调和平均数,也是多分类问题的参考指标;检测时间则用每秒生成图像的数量衡量。

4.3 实验结果与分析

本算法与原始 YOLOv3 算法的性能差异如图 7 所示,可以发现,本算法的 mAP 最终稳定在

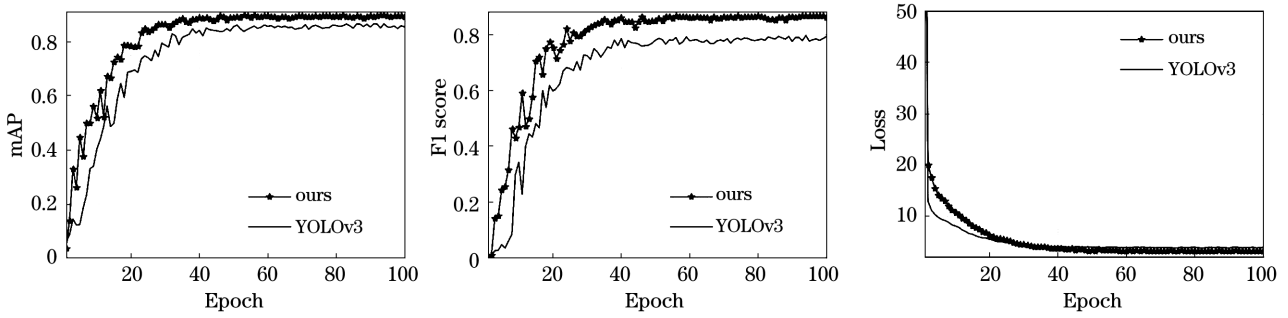


图 7 本算法与 YOLOv3 算法的性能对比。(a)mAP;(b)F1 分数;(c)Loss

Fig. 7 Performance comparison between our algorithm and the YOLOv3 algorithm. (a) mAP; (b) F1 score; (c) Loss

85.50%，最高可达到 89.82%，比 YOLOv3 算法提高了约 4 个百分点；F1 分数高达 86.59%，而 YOLOv3 算法的 F1 分数为 78.34%，这表明本算法的检测精度优于 YOLOv3 算法。

选取中小目标、多目标和复杂背景下的图像验证本算法与原始 YOLOv3 算法的检测性能，结果如图 8 所示。可以发现，相比原始 YOLOv3 算法，本算法在多目标、中小目标和复杂场景上的检测效果

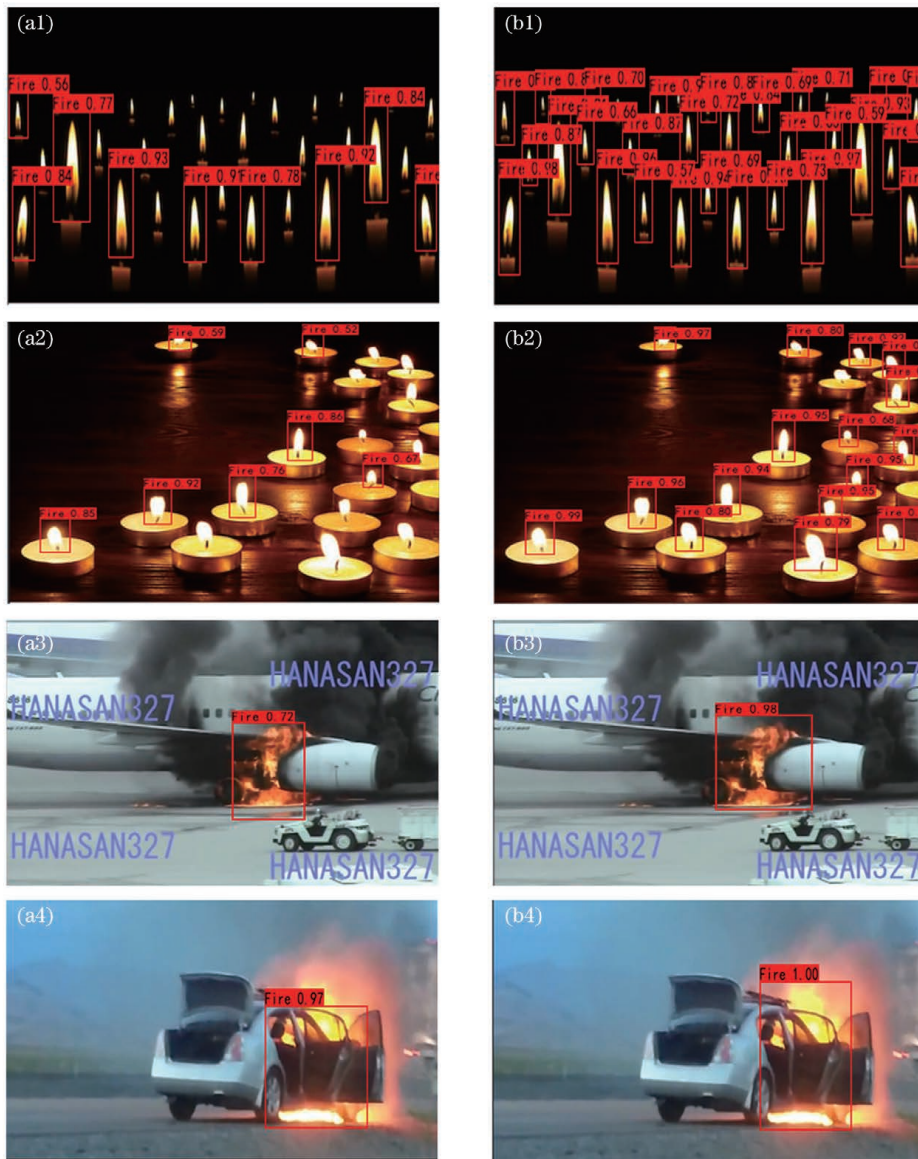


图 8 不同算法的检测结果。(a)原始 YOLOv3 算法;(b)本算法

Fig. 8 Detection results of different algorithms. (a) Original YOLOv3 algorithm; (b) our algorithm

均较好,在多目标和小目标的实际检测效果上,本算法也具有较大优势。

4.4 本算法与其他算法的对比

为了进一步验证本算法的性能,将本算法与其他先进算法进行了对比实验,结果如表 1 所示。可以发现,本算法对各个类别的精确率、召回率和 F1 分数均优于其他算法。由于实验平台的限制,Retinanet 采用 ResNet50 作为特征提取网络,其中,

表 1 不同算法的检测结果

Table 1 Detection results of different algorithms

Algorithm	Category	Precision / %	Recall / %	F1 score / %	mAP / %	FPS
YOLOv3	fire	91.72	69.76	79.00	85.35	26.9
	light	82.17	75.71	79.00		
Retinanet ^[18]	fire	90.71	78.43	84.00	86.22	18.5
	light	83.21	81.43	82.00		
Centernet-HG ^[19]	fire	93.57	77.95	85.00	88.19	15.3
	light	88.46	82.14	85.00		
Ours	fire	95.07	78.90	86.00	89.82	20.2
	light	89.39	84.29	87.00		

4.5 消融实验的结果

本算法在改变网络结构的同时增加了大尺度全卷积模块和注意力机制模块,并更改了原有 Loss 的计算方式。为了进一步分析不同模块的效果,将本算法拆分为 5 组对比实验,结果如表 2 所示。第 2 组实验中,通过在网络 FPN 层引入第 2 个 Residual block 提取局部信息,并加入自上而下的融合路径,提高了局部信息的利用率,输出的有效特征图更丰富,其 mAP 值比原始算法提升了 2.45 个百分点。在第 3 组实验中加入 CSA 机制,使网络更专注有用信息并抑制无用信息,相比第 2 组实验的 mAP 值提升 0.72 个百分点。第 4 组实验在 4 个 Residual block 提取特征图的路径上分别加

表 2 消融实验结果

Table 2 Ablation experiment results

Network	FPN	CSA	GCM	Loss	mAP / %
1	×	×	×	×	85.35
2	✓	×	×	×	87.80
3	✓	✓	×	×	88.52
4	✓	✓	✓	×	88.97
5	✓	✓	✓	✓	89.82

入 4 个 GCM,丰富了特征图的全局信息,且数据集中大目标占比较小,其 mAP 值比第 3 组实验提升了 0.44 个百分点。第 5 组实验则更换了损失函数的计算方式,引入了 CIoU 和 Focal Loss,缓解了数据集二分类不平衡的问题,相比第 4 组实验的 mAP 提升了 0.85 个百分点。这表明实验中对 YOLOv3 算法的改进策略均能提升算法对火焰目标的检测效果。

FPN 与 Focal Loss 的引入使 Retinanet 的检测精度略高于 YOLOv3 算法,略低于 Centernet-HG 算法。原因是 Centernet-HG 算法的特征提取网络 Hourglass Network 参数量较大(19000 万),且该网络利用高分辨率特征图能取得较好的检测效果,但巨大的网络参数使该网络在检测速度上不理想。而本算法在精度和速度上均有较大优势,相比 Centernet-HG 算法,其轻量化的结构更适合实时检测。

5 结 论

通过引入主干特征提取网络第 2 个 Residual block 提取局部信息,并增加自上而下的特征融合结构,增加了局部信息的利用率;在 4 个 Residual block 特征路径上加入大尺度全卷积模块丰富了全局信息,并引入了注意力机制模块,使网络专注于有用信息;将损失函数更改为 CIoU 和 Focal Loss,增强了网络对难识别目标的识别精度并缓解了二分类数据集不平衡的问题。这些基于 YOLOv3 算法的改进,使算法在检测目标时具有较高的精度和检测速度,下一步还将考虑自行设计网络结构,以构造性能更优越的轻量级网络,从而提升模型的检测精度和实时性。

参 考 文 献

- [1] Wu S X, Guo C C, Yang J F. Using PCA and one-stage detectors for real-time forest fire detection[J]. *The Journal of Engineering*, 2020, 2020(13): 383-387.
- [2] Vikram R, Sinha D, De D, et al. EEFFL: energy efficient data forwarding for forest fire detection using localization technique in wireless sensor network[J]. *Wireless Networks*, 2020, 26(7): 5177-5205.
- [3] Dang J M, Yu H Y, Song F, et al. An early fire gas sensor based on 2.33 μm DFB laser[J]. *Infrared Physics & Technology*, 2018, 92: 84-89.
- [4] Toptaş B, Hanbay D. A new artificial bee colony algorithm-based color space for fire/flame detection[J]. *Soft Computing*, 2020, 24(14): 10481-10492.
- [5] Tian L, Wang J D, Zhou H M, et al. Automatic detection of forest fire disturbance based on dynamic modelling from MODIS time-series observations[J]. *International Journal of Remote Sensing*, 2018, 39(12): 3801-3815.
- [6] Fu T J. Forest fire image recognition algorithm and implementation based on deep learning[D]. Beijing: Beijing Forestry University, 2016.
傅天驹. 基于深度学习的林火图像识别算法及实现[D]. 北京: 北京林业大学, 2016.
- [7] Gaur A, Singh A, Kumar A, et al. Video flame and smoke based fire detection algorithms: a literature review[J]. *Fire Technology*, 2020, 56(5): 1943-1980.
- [8] Yin S L, Li H, Teng L. Airport detection based on improved faster RCNN in large scale remote sensing images[J]. *Sensing and Imaging*, 2020, 21(1): 1-13.
- [9] Redmon J, Farhadi A. YOLOv3: an incremental improvement[EB/OL]. (2018-04-08) [2021-01-15]. <https://arxiv.org/abs/1804.02767>.
- [10] Zhang J L, Jia Y, Zhu D, et al. Study on the situational awareness system of mine fire rescue using faster ross Girshick-convolutional neural network[J]. *IEEE Intelligent Systems*, 2020, 35(1): 54-61.
- [11] Fu Y J, Zhang H L. Forest fire detection method based on transfer learning of convolutional neural network[J]. *Laser & Optoelectronics Progress*, 2020, 57(4): 041010.
- [12] Fang W, Wang L, Ren P M. Tinier-YOLO: a real-time object detection method for constrained environments[J]. *IEEE Access*, 2019, 8: 1935-1944.
- [13] Frizzi S, Kaabi R, Bouchouicha M, et al. Convolutional neural network for video fire and smoke detection [C] // IECON 2016-42nd Annual Conference of the IEEE Industrial Electronics Society, October 23-26, 2016, Florence, Italy. New York: IEEE Press, 2016: 877-882.
- [14] Ju M R, Luo H B, Wang Z B, et al. Improved YOLOv3 algorithm and its application in small target detection[J]. *Acta Optica Sinica*, 2019, 39(7): 0715004.
鞠默然, 罗海波, 王仲博, 等. 改进的 YOLOv3 算法及其在小目标检测中的应用[J]. *光学学报*, 2019, 39(7): 0715004.
- [15] He K M, Zhang X Y, Ren S Q, et al. Deep residual learning for image recognition[C]//2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), June 27-30, 2016, Las Vegas, NV, USA. New York: IEEE Press, 2016: 770-778.
- [16] Zhao Q, Li B Q, Li T W. Target detection algorithm based on improved YOLOv3[J]. *Laser & Optoelectronics Progress*, 2020, 57(12): 121502.
赵琼, 李宝清, 李唐薇. 基于改进 YOLOv3 的目标检测算法[J]. *激光与光电子学进展*, 2020, 57(12): 121502.
- [17] Wang Q L, Wu B G, Zhu P F, et al. ECA-Net: efficient channel attention for deep convolutional neural networks[C]//2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), June 13-19, 2020, Seattle, WA, USA. New York: IEEE Press, 2020: 11531-11539.
- [18] Lin T Y, Goyal P, Girshick R, et al. Focal loss for dense object detection [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2020, 42(2): 318-327.
- [19] Zhou X Y, Wang D Q, Krähenbühl P, et al. Object as points [EB/OL]. (2019-04-16) [2021-01-15]. <https://arxiv.org/abs/1904.07850v1>.