

# 改进 YOLOv3 的火灾检测<sup>①</sup>

王 林, 赵 红

(西安理工大学 自动化与信息工程学院, 西安 710048)

通信作者: 赵 红, E-mail: ZhaoHong5054@163.com



**摘 要:** 针对火灾检测中小目标检测率低、复杂场景下检测精度低和检测不及时等问题, 提出了一种改进 YOLOv3 的火灾检测算法. 首先, 通过改进的 K-means 聚类算法重新获取更符合火焰和烟雾尺寸的 anchor; 其次在 Darknet-53 后添加空间金字塔池化, 提升了网络的感受野进而增强了网络对小尺度目标的检测能力; 然后通过 CIoU 改进损失函数, 在计算坐标误差时考虑中心和宽高坐标两者的相关性, 加快了损失函数的收敛; 最后使用 mosaic 数据增强丰富了待检测物体的背景. 在自制的数据集上训练并测试, 实验结果表明: 改进后的算法比 YOLOv3 火焰的 AP 从 94% 提升至 98%, 烟雾的 AP 从 82% 提升至 94%, 平均检测速度从 31 fps 提升至 43 fps, 相比 Faster R-CNN、SDD 等算法也有更高的 mAP 和更快的检测速度. 因此, 改进后的算法能够更有效地进行火灾预警.

**关键词:** 火灾检测; YOLOv3; 空间金字塔池化; CIoU; mosaic 数据增强; 目标检测; 深度学习

引用格式: 王林, 赵红. 改进 YOLOv3 的火灾检测. 计算机系统应用, 2022, 31(4): 143-153. <http://www.c-s-a.org.cn/1003-3254/8404.html>

## Fire Detection Based on Improved YOLOv3

WANG Lin, ZHAO Hong

(School of Automation and Information Engineering, Xi'an University of Technology, Xi'an 710048, China)

**Abstract:** Given the low detection rates of small targets, low detection accuracy in complex scenes, and delayed detection in fire detection, an improved you only look once v3 (YOLOv3)-based fire detection algorithm is proposed. Firstly, an improved K-means clustering algorithm is used to retrieve anchors that are more in line with the sizes of the flames and smoke. Secondly, spatial pyramid pooling is added after the Darknet-53, which improves the network receptive field and enhances the detection ability of the network on small-scale targets. Thirdly, the loss function is improved through complete intersection over union (CIoU), and the convergence of the loss function is sped up by taking into consideration the correlations of the center with the width and height coordinates when calculating the coordinate error. Finally, mosaic data enhancement is employed to enrich the background of the object to be detected, and the improved algorithm is trained and tested on a self-made data set. The experimental results show that compared with the YOLOv3 algorithm, the improved algorithm improves the flame AP from 94% to 98%, increases the smoke AP from 82% to 94%, and promotes the average detection speed from 31 fps to 43 fps. Compared with the Faster R-CNN, SDD, and other algorithms, it also has a higher mAP and a faster detection speed. Therefore, the improved algorithm is more effective in fire warning.

**Key words:** fire detection; YOLOv3; spatial pyramid pooling (SPP); complete intersection over union (CIoU); mosaic data enhancement; target detection; deep learning

火灾比其它自然灾害更为常见, 一旦发生, 不仅会造成严重的人员伤亡和财产损失, 还会产生有毒气体

污染环境<sup>[1]</sup>. 2019 年 3 月四川凉山火灾事件导致 31 人遇难, 大面积的森林土地丢失. 若能及时发现火情并报

<sup>①</sup> 基金项目: 陕西省科技计划重点项目 (2017ZDCXL-GY-05-03)

收稿时间: 2021-06-22; 修改时间: 2021-07-14; 采用时间: 2021-07-20; csa 在线出版时间: 2022-03-22

警, 则能减少损失<sup>[2]</sup>。随着科学技术的发展, 基于传感器的设备被用于火灾检测, 但探测范围有限, 对场地开阔的林火等大空间来说安装成本高很难做到均匀覆盖, 且对物理量的观测很容易受环境因素的干扰。随着计算机视觉和图像处理技术的发展及监控的普及, 基于视频图像的火灾检测出现<sup>[3]</sup>。李红娣等人<sup>[4]</sup>将图像特征金字塔的纹理特征和边缘特征组合, 用机器学习方法支持向量机训练达到 94% 以上的检测率和 3.0% 以下的误报率。Kim 等人<sup>[5]</sup>通过火焰的 RGB 颜色模型检测火灾但是检测的准确率不高。李巨虎等人<sup>[6]</sup>提出了基于颜色和纹理特征的森林火灾图像识别, 通过提取不同空间的纹理特征, 将得到的特征向量送入 SVM 分类器, 获得了 94.55% 的准确率。但传统的视频火灾检测大都基于人工提取火焰和烟雾的静态特征及运动区域的动态特征来研究的, 人工提取特征容易造成检测的精度较低。

2012 年, 深度学习的出现促使人工智能飞速发展, 为目标检测注入新的活力, Krizhevsky 等人<sup>[7]</sup>提出 Alexnet 并将它用在图片分类中, 将错误率从 25.8% 降到了 16.4%, 2014 年 Girshick 等人<sup>[8]</sup>提出将具有 CNN 特征的区域 (RCNN) 用于目标检测, 开启了目标检测的新时代。此后, 目标检测在计算机视觉<sup>[9]</sup>语音识别<sup>[10,11]</sup>多个领域表现出了良好的效果。2015 年, 相关学者提出具有历史意义的一步检测 YOLO (you only look once), 可以达到每秒 45 帧的速度, Redmon 等人<sup>[12,13]</sup>前后提出了将 YOLO1、YOLO2 用于火灾检测, 精度和速度逐渐提高。杜晨锡等人<sup>[14]</sup>提出了基于 YOLOv2 网络的火焰检测方法, 但模型可应用的场景简单。杨剑等人<sup>[15]</sup>提出了基于图像处理技术的烟雾检测在家庭火灾防范中应用, 实验取得了较高的准确率, 但对小区域烟雾存在漏检情况。针对以上问题, 本文提出了一种改进 YOLOv3<sup>[16]</sup>算法的火焰和烟雾检测, 通过网络爬虫、视频截取等手段获得场景复杂, 数量较多的火焰和烟雾的数据集, 人工标注区域的位置信息和类别信息。针对 YOLOv3 算法对小目标区域检测精度低的问题做出改进, 实验表明, 本文的改进相比原来的 YOLOv3 算法准确率和误检率有了较大幅度的提升, 能够在火灾初始阶段及时发现并报警。

## 1 YOLOv3 原理概述

为避免模型优化中网络加深出现退化的问题<sup>[17]</sup>,

YOLOv3 的作者借鉴了残差块 (Residual block) 的思想形成了 Darknet-53 网络结构<sup>[18]</sup>。如图 1 所示, 网络由 3 部分组成, 特征提取网络 Darknet-53 和多尺度预测 FPN 及检测网络 YOLO Head。其中, Darknet-53 没有使用池化层, 而是借鉴残差网络 (residual network, ResNet) 的结构形成跳跃连接对输入的图片特征提取, YOLOv3 算法的卷积块大多采用卷积层 (convolution layers) 后跟批标准化层 (batch normalization, BN) 和激活函数 (leaky ReLU) 的模式来加速收敛避免梯度消失。输入图像经过一个普通卷积模块 (Conv2D+BN+Leaky\_relu) 后, 使用步长为 2 大小为  $3 \times 3$  卷积进行 5 次下采样, 如图 1 中 (a) 部分所示, 保留每次卷积后的 layer, 且分别在每一次下采样后使用  $1 \times 1$  和  $3 \times 3$  大小的卷积核再次进行卷积操作。将此次连续卷积后的结果和保留的 layer 相加。5 次下采样后图片的高和宽不断压缩变成  $13 \times 13$ , 通道数不断扩张变成  $1024 \times 1024$ , 这 5 次下采样使用残差块的结构堆叠起来, 残差块结构如图 1 中的 (b) 所示, 每个残差块的拼接次数依次是 1、2、8、8、4。

为了检测出不同尺寸的物体, YOLOv3 借鉴 SSD 的特征金字塔 (feature pyramid network, FPN) 结构进行多尺度检测, 选取后面 3 个 32 倍、16 倍、8 倍 ( $13 \times 13$ 、 $26 \times 26$ 、 $52 \times 52$ ) 的特征层预测物体, 为了更好的利用深层语义特征信息, 将最后一个特征尺度 ( $13 \times 13$ ) 用线性插值的方式上采样与前面一个特征层进行拼接, 这样可以在较低层特征上获得深层特征的信息检测尺寸较小的物体, 保证检测的效果。依次如此, 直到和  $52 \times 52$  大小的特征层拼接。

尽管 YOLOv3 是一个经典、高效的网络, 在目标检测领域有良好的表现力, 但火灾检测的公开数据集较少, 且相对形状固定的目标来说火焰和烟雾具有运动性、复杂性, 特别是针对小火焰区域和复杂背景下的火焰和烟雾, YOLOv3 容易发生漏检的情况, 失去了火灾预警的意义。本文针对这一问题对 YOLOv3 算法进行改进, 并将改进后的算法与 YOLOv3 原算法进行比较, 通过对比验证了本文算法在火灾检测方面的可行性。

## 2 YOLOv3 改进相关工作

### 2.1 主干网络 Darknet-53 的改进

YOLOv3 网络通过基本的卷积组件和一系列残差块对火焰和烟雾的图片特征提取, 使用 8 倍、16 倍、

32倍后输出的3种特征图进行多尺度预测,  $13 \times 13$  的尺度提取深层语义信息, 用来预测尺寸较大的物体,  $26 \times 26$  尺度的特征图预测尺寸相当的物体, 同理,  $52 \times 52$  的特征尺度因网格划分较细提取浅层位置信息, 用来对较小尺度目标预测. 但经8倍下采样得到的特征图和原始  $416 \times 416$  的输入图像来说依然不够精细, 当对更小尺度目标检测时会出现效果不佳的问题. 除非  $13 \times 13$  尺度的特征图能够给它提供更多的深层特征信息以供它检测小目标. 但 YOLOv3 网络对输入图像经过5次下采样后并没有其它操作.

本文决定改进 YOLOv3 网络的 Darknet-53 结构, 在其后添加一个 SPP (spatial pyramid pooling) 模块来加大网络的特征接受范围, 以充分利用上下文特征信

息, 增大感受野 (receptive field), 提升对小尺度目标的检测效果. SPP 由 He 等<sup>[19]</sup> 提出, 解决了卷积神经网络 (CNNs) 输入图像尺寸限制的问题. 对图像进行一次特征映射来避免特征重复提取, 加快目标检测速度. 如图 2 所示. 对  $13 \times 13$  尺度特征图继续处理, 首先对此特征图做 3 次卷积 (Conv2D+BN+Leaky\_relu), 分别是  $512 \times 1 \times 1$ 、 $1024 \times 3 \times 3$ 、 $512 \times 1 \times 1$ . 然后将得到的结果用  $1 \times 1$ 、 $5 \times 5$ 、 $9 \times 9$ 、 $13 \times 13$  四个不同尺寸的池化核进行最大池化操作 (max-pooling), 且并联一个输入到输出的直连操作. 最后将得到的特征图堆叠 (concat), 将不同尺度下提取的特征汇聚在一起. 最大池化核的尺寸和特征图尺寸一样, 保证融合了图片的局部特征信息和全局特征信息.

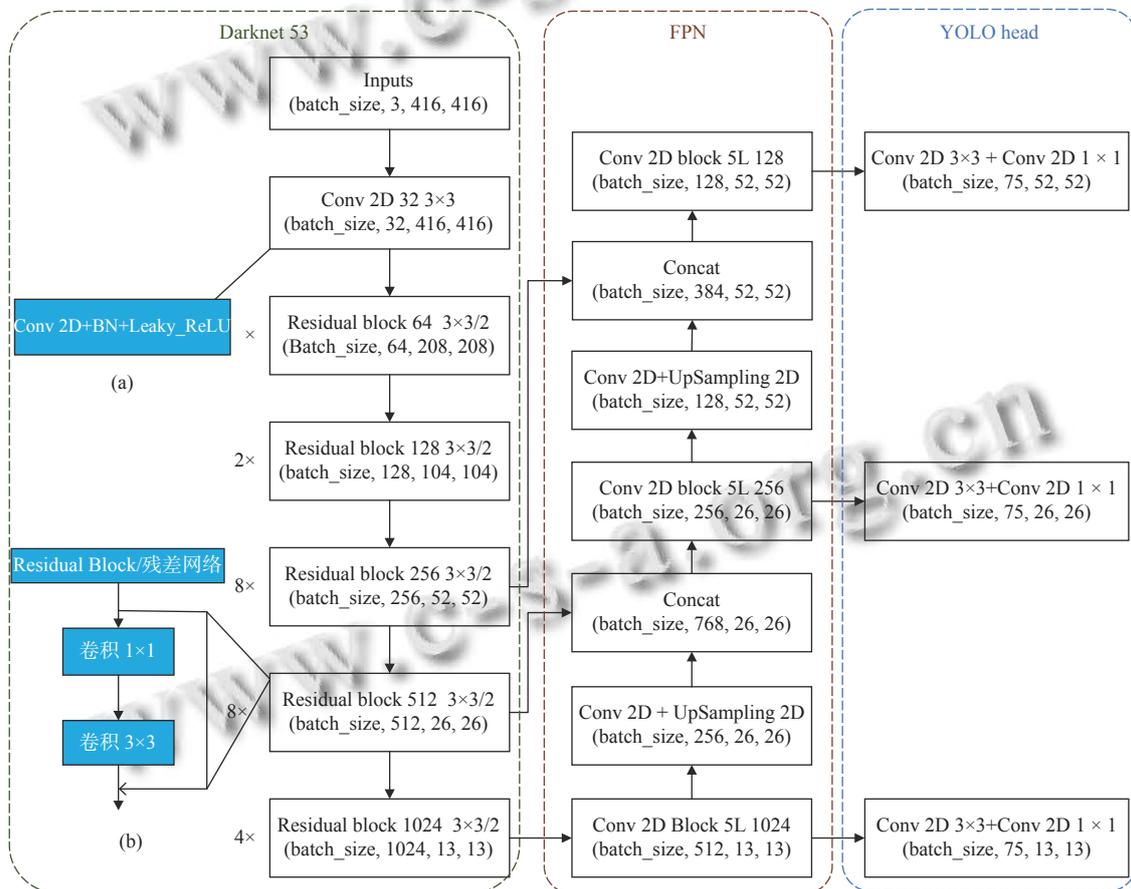


图1 YOLOv3 网络结构

## 2.2 先验框的改进

相比基于区域的两阶段 (two stage) 目标检测算法, YOLOv3 网络将输入图像直接送入一个深度卷积神经网络, 一次性的进行目标的类别和位置的判断. 将原始

图片划分为大小一样、互不重复的小网格, 通过 Darknet-53 网络得到 3 个不同尺度的特征图. 每个特征点代表原始图片的一小块元素, 当真实目标物体的中心落在这个网格内时, 则由这个网格负责检测该物体, 如图 3

所示。

YOLOv3 网络沿用 YOLOv2 采用 anchor boxes 机制, 给每个特征点预设一组面积不同, 宽高比大小不同的初始先验框, 这些先验框是通过 K-means 聚类算法得到的框. 对于目标检测算法来说, 一组合适的先验框将减少网络的调整使之更快的获得预测框. 为了实现

快速、准确的火焰和烟雾检测, 需要对 YOLOv3 本身的先验框优化使之适合自己的数据集. 本文决定通过改进的 K-means 聚类算法对标注火焰和烟雾的真实框 (true box) 分析, 使先验框的尺寸更符合本文火焰和烟雾数据集的尺寸. 聚类的个数  $K$  即为要选取的先验框的个数, 聚类中心框的宽和高即为先验框的宽和高.

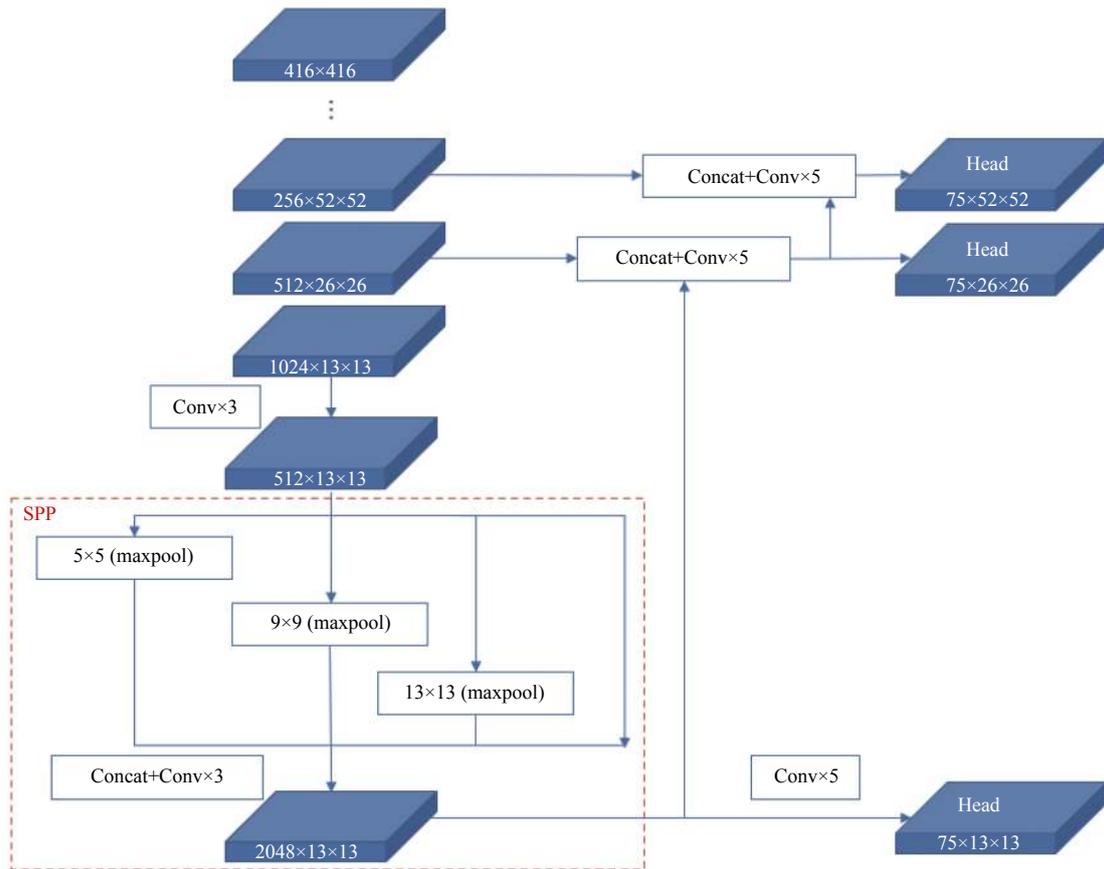


图2 加了 SPP 模块的网络结构

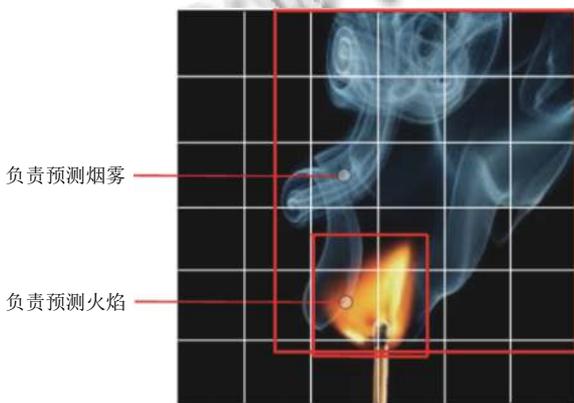


图3 网格预测示意图

传统的 K-means 聚类算法采用欧式距离计算, 但当聚类不是用于点的聚类而是用于边框的聚类时会产生误差. 边框的尺寸越大产生的误差也会越大, 因此本文决定用改进的交并比 (intersection over union, IoU)<sup>[20]</sup> 来进行 K-means 聚类. 聚类开始时随机选取  $k$  个真实框 (bounding box) 作为初始锚框 (anchor box), 计算每个 bounding box 与每个 anchor box 的 IoU, IoU 越大时误差越小. 依次将每个 bounding box 分配给误差最小的那个 anchor box. 最后, 将每个 anchor box 中所有的 bounding box 求宽高中值大小作为新的 anchor box. 重复迭代更新, 直至所有的真实框所属的 anchor box 和

上一次分配的一样, 停止更新. 保存最后得到的 anchor box 作为聚类后的先验框. 计算样本中 bounding box 与  $k$  个 anchor box 的平均 IoU (AvgIoU), 得到的结果如图 4 所示. 当值越来越大时, AvgIoU 的值先快速增加后平缓增加. 考虑到检测模型的计算精度和速度, 选取聚类个数  $k=9$ .

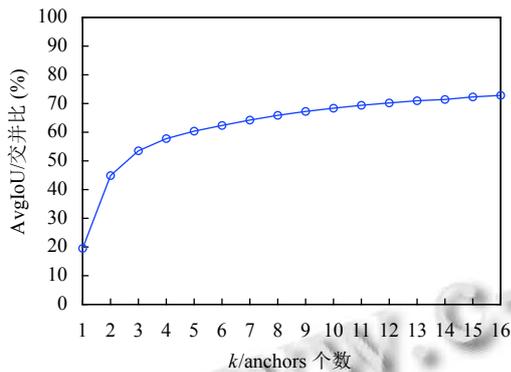


图 4 先验框个数  $k$  与 AvgIoU 关系图

聚类后的聚类中心坐标和先验框尺寸如表 1 所示, 其中坐标是相对于图像大小的比例, 将其与输入尺寸相乘得到先验框尺寸.

表 1 改进 K-means 聚类得到中心坐标和先验框尺寸

聚类中心坐标	先验框尺寸 (像素)
(0.031, 0.055)	13×23
(0.058, 0.106)	24×44
(0.084, 0.238)	35×99
(0.149, 0.137)	62×57
(0.178, 0.361)	74×150
(0.305, 0.219)	127×91
(0.339, 0.570)	141×237
(0.567, 0.361)	236×150
(0.750, 0.680)	312×283

沿用 YOLOv3 的做法, 将 9 种宽高比不同的先验框平均分配到 3 种不同尺度的特征图上, 经 32 倍下采样后的  $13 \times 13$  特征图感受野最大, 用来检测尺寸最大的火焰和烟雾, 将  $141 \times 237$ ,  $236 \times 150$ ,  $312 \times 283$  分配给它;  $26 \times 26$  特征图感受野一般, 将  $62 \times 57$ ,  $74 \times 150$ ,  $127 \times 91$  分配给它; 然后将剩余的  $13 \times 23$ ,  $24 \times 44$ ,  $35 \times 99$  分配给感受野最小的  $52 \times 52$  特征图, 用来检测小火焰烟雾目标.

### 2.3 损失函数的改进

损失函数衡量的是真实值和预测值的误差. 神经网络反向传播的训练是通过计算损失函数对梯度的偏导来更新权重和参数, 因此损失函数的定义直接影响

到模型训练的好坏. YOLOv3 算法的损失函数<sup>[21]</sup>由 3 部分组成: 坐标损失 (coord loss)、类别损失 (class loss) 和置信度损失 (confidence loss), 损失函数表达式如下:

$$\begin{aligned}
 Loss = & \frac{1}{2} \sum_{i=1}^{10647} \lambda_{obj} * \left( (2 - truth_w * truth_h) \right. \\
 & * \sum_{r \in (x,y,w,h)} (truth_r - predict_r)^2 \\
 & + \sum_{r=0}^{k-1} ((r == truth_{class}) ? 1 : 0 - predict_{class_r})^2 \\
 & \left. + (truth_{conf} - predict_{conf})^2 \right) \quad (1)
 \end{aligned}$$

其中, 网格中存在物体为 1, 否则是 0,  $r$  为物体的种类, 如火焰或烟雾, 其中  $x, y$  为物体的中心坐标,  $w, h$  为物体的宽高.

YOLOv3 在计算坐标误差时将中心和宽高坐标作为单独的变量去计算, 忽视了两者的相关性. 考虑用真实框和预测框的重合程度去计算坐标损失, 但当两个框之间没有重合或尺寸相差较大时无法计算, 为加快损失函数的收敛, 采用 CIoU (Complete-IoU)<sup>[22]</sup> 作为位置坐标的损失函数. 计算公式如下:

$$Loss_{CIoU} = 1 - IoU(C, D) + p^2(b_p, b_t)/c^2 + \alpha v \quad (2)$$

其中,  $v = \frac{4}{\pi^2} \left( \arctan \frac{w_t}{h_t} - \arctan \frac{w_p}{h_p} \right)^2$ ,  $\alpha = \frac{v}{1 - IoU + v}$ . 下标  $t$  代表真实框, 下标  $p$  代表预测框.  $P^2(\cdot)$  为预测框和真实框中心坐标欧式距离的计算,  $C$  为预测框和真实框最小包围框对角线长度.  $\alpha, v$  是对预测框和真实框宽高比的惩罚项, 当宽和高很接近时,  $v$  等于 0.

## 3 实验结果及分析

### 3.1 实验环境

本文所用的深度学习框架为 PyTorch 1.2.0 版本, 实验基于 Anaconda 虚拟环境, 编辑器为 VScode, 编程语言为 Python 3.5.2, 操作系统是 Windows 10, GPU 为 GTX1080Ti, CPU 是 Inter i7-6700, 电脑内存为 128 GB, GPU 加速库 cuDNN 的版本是 7.4.1.5, 配套的 CUDA 的版本是 10.0. 实验程序将原始输入图像尺寸统一调整为  $416 \times 416$ .

### 3.2 实验数据集

目前基于深度学习的火灾检测公开的数据集较少且场景单一, 本文所用的实验数据集是通过网络爬虫和数据截取方式获得的, 选取了 10 825 张不同场景下

的火焰和烟雾图像,其中有山地、田野、森林、室内等场景,也有白天和夜晚情景下的图像,有3123张火焰图像,有2987张烟雾图像,4715张火焰和烟雾混合图像.本文将数据集按照8:1:1比例分别用作训练、验证和测试.数据集的部分展示如图5所示.

本文使用LabelImg软件来对数据集进行标注,标注的种类为火焰 fire 和烟雾 smoke 两类.注意标注时要完全框柱目标物体,不能缺少或超过以影响火灾检

测的准确率.将图像中标注物体的种类信息和位置信息保存至XML格式文件中,然后通过程序将这些信息转换为VOC格式进行训练.

为了提升模型的鲁棒性,本文对训练的数据集使用Mosaic在线数据增强<sup>[23]</sup>.此方法将4张图片进行拼接,它的一个巨大优点是丰富待检测物体的背景,且批量标准化计算的时一次性计算出4张图片的数据.过程如图6所示.

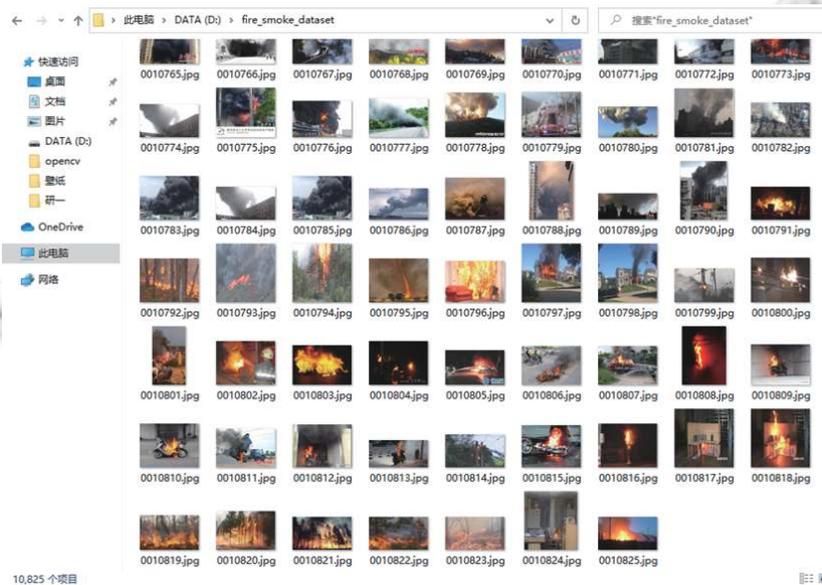


图5 火灾数据集部分图像



图6 Mosaic数据增强示例

### 3.3 模型训练

实验采用小批量梯度下降方法优化模型,批尺寸(batch size)设置为20,动量因子(momentum)设置为0.9,使用Mosaic进行在线数据增强.整个训练过程有100轮(epoch),前50次epoch是初始学习率为0.001的冻结训练,后50次epoch在冻结训练的基础上进行

初始学习率为0.0001的训练.模型所有的迭代次数为39000次.每次迭代训练结束后将得到的训练权重以及整体loss和验证集loss值保存在日志中,模型训练初期根据loss值收敛程度调整参数,最终将网络参数调整至合适大小,并将loss值收敛效果最好的一次的权重作为最终权重进行预测.损失函数随着迭代次数

的变化如图7所示,为了更加准确的观察 loss 的变化趋势,将第300次到第20000次的迭代放大并置于图中,可以看到模型的收敛速度和收敛效果均较好,可以看出,在第15000次迭代后,模型的 loss 值趋于稳定。

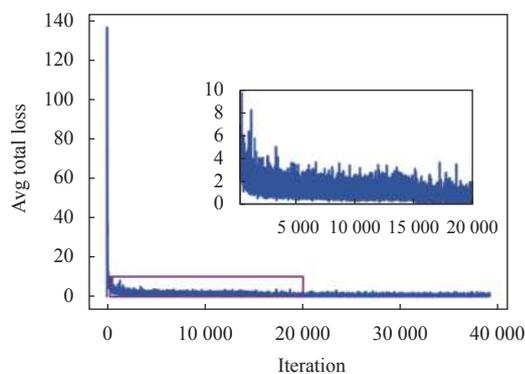


图7 训练模型损失函数曲线图

每次迭代预测框与真实框的平均 IoU 如图8所示,可以看到,随着迭代次数的增加, IoU 值上升并趋于稳定。

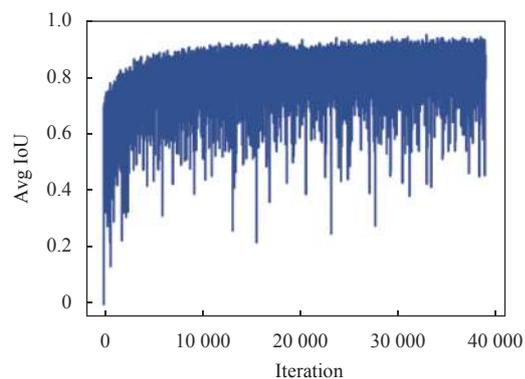
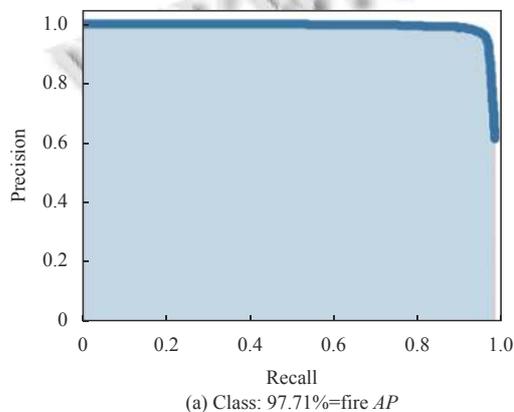


图8 训练模型 IoU 曲线图



### 3.4 评估指标

本文使用 mAP 作为模型的评价指标,下面是计算过程.  $TP$  是被模型分到正样本且分对了的个数,  $TN$  是被模型分到负样本且分对了的个数,  $FP$  是被模型分到正样本且分错了的个数,  $FN$  是被模型分到负样本且分错了的个数. 那么模型的准确率 (precision,  $P$ ) 计算公式为:

$$P = \frac{TP}{TP + FP} \quad (3)$$

模型的召回率 (recall,  $R$ ) 计算公式为:

$$R = \frac{TP}{TP + FN} \quad (4)$$

当取不同的置信度时可以得到不同的  $P$  和  $R$ , 由此, 得到平均准确率  $AP$  的计算公式为:

$$AP = \int_0^1 P(R) dR \quad (5)$$

检测目标火焰和烟雾对应的 P-R (Precision-Recall) 曲线如图9所示. 其中火焰的  $AP$  为 97.71%, 烟雾的  $AP$  为 94.32%, 当分类个数超过一个时, 便引入了  $mAP$ ,  $mAP$  的计算公式为:

$$mAP = \frac{\sum_{i=1}^n AP_i}{n} \quad (6)$$

其中,  $n$  为分类的种类数, 本文中有火焰和烟雾, 因此  $n = 2$ , 则本文改进后模型的  $mAP$  为 96.02%.

### 3.5 模型对比分析

#### 3.5.1 训练 loss 图对比

两个模型的损失函数如图10所示, 可以看到, 由于迁移学习 YOLOv3 初始权重, 模型一开始的整体损失较少, 随着迭代次数的增多, 改进后模型 loss 的收敛速度和最终收敛效果均比 YOLOv3 模型要好。

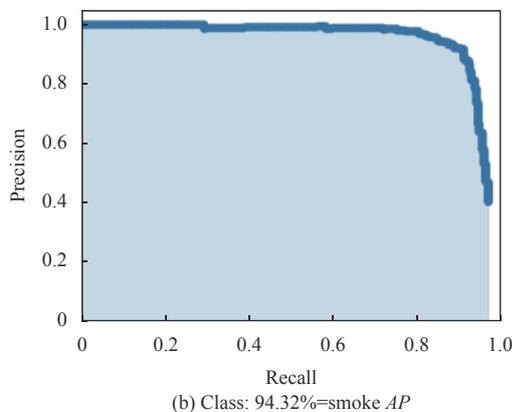


图9 火焰和烟雾 P-R 曲线图

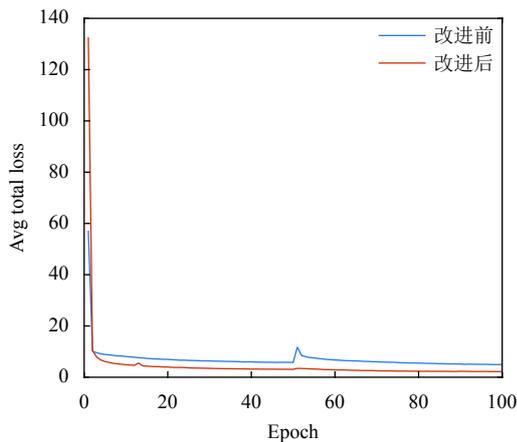


图10 模型训练 Loss 收敛对比图

### 3.5.2 $mAP$ 对比

两个模型的准确率 (precision)、召回率 (recall)、精度 (AP)、平均精度 ( $mAP$ ) 对比如表 2 所示。

表 2 模型训练结果参数对比 (%)

评估指标	改进前-fire	改进前-smoke	改进后-fire	改进后-smoke
准确率	96.00	96.00	97.51	95.74
召回率	84.76	61.09	94.61	83.03
精度	93.61	82.50	97.71	94.32
平均精度	88.05		96.02	

YOLOv3 的  $mAP$  为 88.05%，改进后模型的  $mAP$  为 96.02%。可见改进后模型的 AP 和  $mAP$  表现均较好。

### 3.5.3 测试集检测结果对比

为了测试两个模型对复杂背景的检测效果以及小目标的检测效果,选取了测试集中几种有代表性的图像进行对比,检测结果如图 11 所示。

从图 11 的检测结果可以看到,改进后的模型在小目标检测方面比 YOLOv3 有更大的优势,对于图 11(a) 中未检测出来的小火焰区域,改进后的模型均能极好的检测出来,且对复杂背景下的烟雾来说因其运动不规则易受背景影响,从图 11(c)、图 11(e)、图 11(g) 可以看出,YOLOv3 存在多次漏检现象,但改进后的模型能很好的检测出复杂背景下的烟雾区域。因此,本文的改进是效果显著的,对小目标的查全率及复杂背景下目标的检测精度均比 YOLOv3 要好。

### 3.5.4 公开数据集检测结果对比

为了检验改进后算法的鲁棒性,本文在公开的火灾数据集上与其它算法进行对比检测。公开数据集来自于 Research Webpage about Smoke Detection for Fire Alarm (<http://signal.ee.bilkent.edu.tr/VisiFire/Demo/>

SampleClips.html), 本文对其中的视频进行截图后得到不同状态下的火焰和烟雾图片,部分图片如图 12 所示。

分别使用算法 Faster RCNN、SDD、文献 [24]、YOLOv3 及本文算法对公开数据集进行检测,并计算检测结果,算法的检测结果对比见表 3。

通过对比可知,本文算法在精度和速度方面均表现更好,可以更好的应用于火灾检测。其中, Faster RCNN 的  $mAP$  比 YOLOv3 好,但由于其在候选区处理时需要很大时间,所以检测时间较长。而火灾检测对实时性的要求比对精度的要求优先级更高,因此,改进后的本文算法更适用于火灾检测。

### 3.5.5 消融实验对比

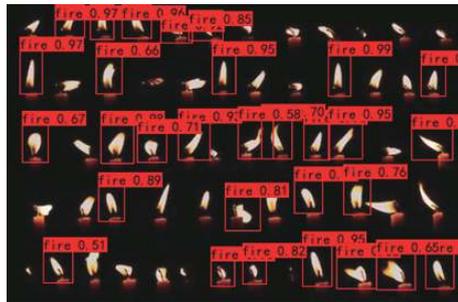
本文的算法改进有:在主干特征提取网络中加入 SPP 模块、使用 CIoU 计算新的损失函数、使用 Mosaic 进行数据增强。为深入了解这些改进对于算法性能提升的程度,将本文算法拆分为 4 组实验,实验结果如表 4 所示。

从表 4 可以看出,在只加入 SPP 模块的模型 2 中,由于提升了网络的感受野,融合了图片的局部特征信息和全局特征信息,使得  $mAP$  提升了 2.84 个百分点。在加入了 SPP 模块和 CIoU 改进损失函数的模型 3 中,由于计算坐标误差时考虑了中心和宽高坐标两者的相关性,加快了损失函数的收敛,使得  $mAP$  提升了 3.67 个百分点。在模型 4 中,增加了 Mosaic 数据增强,丰富待检测物体的背景,使得  $mAP$  提升了 1.46 个百分点。因此,本文的改进方法均可以提升火灾检测效果,其中,引入 CIoU 改进损失函数的改进效果最显著。

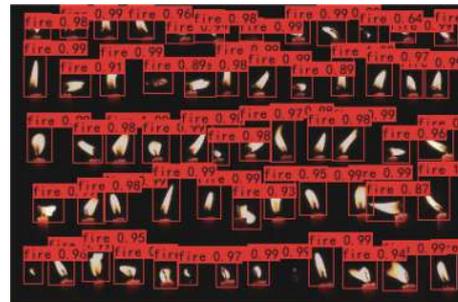
## 4 总结与展望

针对深度学习方法在火灾检测应用上的不足,本文提出了改进 YOLOv3 的视频图像火灾检测算法,通过特征提取网络的改进、先验框的优化、损失函数的改进和数据增强等一系列措施提升了火灾检测的精度和速度。

在自行标注的 Fire-Smoke-Detection-Dataset 上进行了实验。实验结果显示,本文改进的模型在火灾检测方面有更好的鲁棒性、更高的准确率,尤其是在复杂背景下的检测和对小目标的检测效果更优异。其中,对火焰的检测率从 94% 提升至 98%,对烟雾的检测率从 82% 提升至 94%, $mAP$  从 88.05% 提升至 96.02%,平均检测速度从 31 fps 加快到 43 fps。之后的工作里,将继续对网络进行优化改进,进一步改善检测效果和算法优化效果,增加对烟雾这一复杂对象的详细分析。



(a) 改进前



(b) 改进后



(c) 改进前



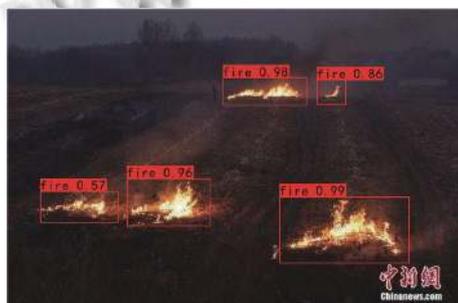
(d) 改进后



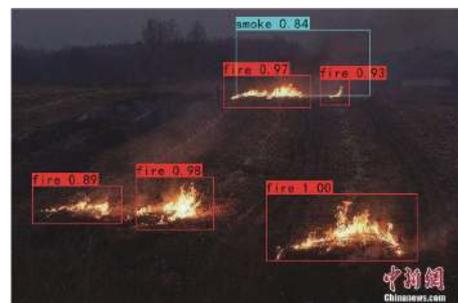
(e) 改进前



(f) 改进后



(g) 改进前



(h) 改进后

图 11 实际检测效果对比图



图 12 公开数据集部分图片

表 3 公开数据集检测结果对比

算法	<i>mAP</i> (%)	平均检测时间 (ms)
Faster RCNN	90.73	205.87
SDD	86.92	125.68
文献[24]	95.24	39.57
YOLOv3	88.87	32.26
本文算法	96.21	23.26

表 4 消融实验对比结果

Model	SPP	CIoU	mosaic	<i>mAP</i> (%)
1	×	×	×	88.05
2	√	×	×	90.89
3	√	√	×	94.56
4	√	√	√	96.02

参考文献

- 刘咏梅. 火灾环境影响及防治对策研究 [硕士学位论文]. 长春: 吉林大学, 2004.
- 赵心宇. 无线智能防火报警系统的研究与应用 [硕士学位论文]. 哈尔滨: 哈尔滨工程大学, 2011.
- 彭文健. 智能视频系统中的火焰检测算法研究 [硕士学位论文]. 成都: 电子科技大学, 2011.
- 李红娣, 袁非牛. 采用金字塔纹理和边缘特征的图像烟雾检测. 中国图象图形学报, 2015, 20(6): 772–780. [doi: 10.11834/jig.20150606]
- Kim YH, Kim A, Jeong HY. RGB color model based the fire detection algorithm in video sequences on wireless sensor network. International Journal of Distributed Sensor

- Networks, 2014, 2014: 923609. [doi: 10.1155/2014/923609]
- 李巨虎, 范睿先, 陈志泊. 基于颜色和纹理特征的森林火灾图像识别. 华南理工大学学报 (自然科学版), 2020, 48(1): 70–83. [doi: 10.12141/j.ssn.1000-565X.190181]
- Krizhevsky A, Sutskever I, Hinton GE. ImageNet classification with deep convolutional neural networks. Proceedings of the 25th International Conference on Neural Information Processing Systems. 2012. 1106–1114.
- Girshick R, Donahue J, Darrell T, et al. Region-based convolutional networks for accurate object detection and segmentation. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2016, 38(1): 142–158. [doi: 10.1109/TPAMI.2015.2437384]
- Girshick R, Donahue J, Darrell T, et al. Rich feature hierarchies for accurate object detection and semantic segmentation. Proceedings of 2014 IEEE Conference on Computer Vision and Pattern Recognition. Columbus: IEEE, 2014. 580–587. [doi: 10.1109/CVPR.2014.81]
- Sercu T, Puhersch C, Kingsbury B, et al. Very deep multilingual convolutional neural networks for LVCSR. Proceedings of 2016 IEEE International Conference on Acoustics, Speech and Signal Processing. Shanghai: IEEE, 2016. 4955–4959. [doi: 10.1109/ICASSP.2016.7472620]
- 甄斌, 吴玺宏, 刘志敏, 等. 语音识别和说话人识别中各倒谱分量的相对重要性. 北京大学学报 (自然科学版), 2001, 37(3): 371–378.
- Redmon J, Divvala S, Girshick R, et al. You only look once: Unified, real-time object detection. Proceedings of 2016

- IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas: IEEE, 2016. 779–788.
- 13 Redmon J, Farhadi A. YOLO9000: Better, faster, stronger. Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition. Honolulu: IEEE, 2017. 6517–6525.
- 14 杜晨锡, 严云洋, 刘以安, 等. 基于 YOLOv2 的视频火焰检测方法. 计算机科学, 2019, 46(6): 301–304.
- 15 杨剑, 陈世娥. 基于图像处理技术的烟雾检测在家庭火灾防范中应用. 电脑知识与技术, 2021, 17(6): 189–191.
- 16 Redmon J, Farhadi A. YOLOv3: An incremental improvement. Proceedings of the Conference on Computer Vision and Pattern Recognition. 2018. 1–4.
- 17 王旺. 基于生成对抗网络的图像超分辨率算法的设计与实现 [硕士学位论文]. 武汉: 武汉工程大学, 2018.
- 18 赵益, 张志梅. 基于 YOLOv3 的车辆检测算法. 青岛大学学报(自然科学版), 2020, 33(3): 57–64.
- 19 He KM, Zhang XY, Ren SQ, *et al.* Spatial pyramid pooling in deep convolutional networks for visual recognition. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2015, 37(9): 1904–1916. [doi: [10.1109/TPAMI.2015.2389824](https://doi.org/10.1109/TPAMI.2015.2389824)]
- 20 王国文. 多尺度特征融合改进 YOLOv3 网络的行人和车辆检测 [硕士学位论文]. 南京: 南京邮电大学, 2019.
- 21 吕铄, 蔡烜, 冯瑞. 基于改进损失函数的 YOLOv3 网络. 计算机系统应用, 2019, 28(2): 1–7. [doi: [10.15888/j.cnki.csa.006772](https://doi.org/10.15888/j.cnki.csa.006772)]
- 22 Zheng ZH, Wang P, Liu W, *et al.* Distance-IoU loss: Faster and better learning for bounding box regression. Proceedings of the AAAI Conference on Artificial Intelligence, 2020, 34(7): 12993–13000. [doi: [10.1609/aaai.v34i07.6999](https://doi.org/10.1609/aaai.v34i07.6999)]
- 23 Bochkovskiy A, Wang CY, Liao HYM. YOLOv4: Optimal speed and accuracy of object detection. arXiv: 2004.10934, 2020.
- 24 赵飞扬, 罗兵, 林国军, 等. 基于改进 YOLOv3 的火焰检测. 中国科技论文, 2020, 15(7): 820–826.