

# 改进 YOLOX-nano 的火灾火焰烟雾检测<sup>①</sup>



汪子健<sup>1,3</sup>, 高焕兵<sup>1,3</sup>, 侯宇翔<sup>1,3</sup>, 杜传胜<sup>1,3</sup>, 贝太学<sup>2</sup>

<sup>1</sup>(山东建筑大学 信息与电气工程学院, 济南 250101)

<sup>2</sup>(山东建筑大学 机电工程学院, 济南 250101)

<sup>3</sup>(山东省智能建筑技术重点实验室, 济南 250101)

通信作者: 高焕兵, E-mail: gaohuanbing@126.com

**摘要:** 针对在检测火焰和烟雾的火灾检测过程中存在火灾初期小目标难以检测的情况, 本文提出了一种基于自然指数损失 (eCIoU) 的改进 YOLOX-nano (ASe-YOLOX-nano) 目标检测算法. 首先, 提出一种新的目标检测函数 eIoU 损失函数来替代传统 IoU 损失, 解决在检测小目标时预测框和真实框易出现无交集的情况, 及无法反应宽高影响等问题. 其次, 在网络模型中引入注意力模块, 在网络初期模糊定位目标位置, 提高网络后期对目标尤其是小目标检测的准确性. 此外, 本文还采用软池化空间金字塔池化结构提取不同尺寸的空间特征信息, 可以提升模型对于空间布局和物体变性的鲁棒性, 因此目标较小时也可以提取充足的特征, 采用 Mosaic 增强技术预处理数据集, 提升模型的泛化能力, 以此进一步提高网络性能. 通过目标数据集进行对比验证, 其结果显示, mAP 指标达到 70.07%, 比原模型提高了 3.46%, 火焰的准确率达到 84.66%, 烟雾的达到 74.56%, FPS 能够稳定在 73, 相对于传统 YOLOX-nano 算法拥有更好的火灾检测能力.

**关键词:** 火灾检测; 注意力机制; 软池化空间金字塔; 损失函数; YOLOX

引用格式: 汪子健, 高焕兵, 侯宇翔, 杜传胜, 贝太学. 改进 YOLOX-nano 的火灾火焰烟雾检测. 计算机系统应用, 2023, 32(3): 265-274. <http://www.c-s-a.org.cn/1003-3254/9000.html>

## Flame and Smoke Detection of Fires Based on Improved YOLOX-nano

WANG Zi-Jian<sup>1,3</sup>, GAO Huan-Bing<sup>1,3</sup>, HOU Yu-Xiang<sup>1,3</sup>, DU Chuan-Sheng<sup>1,3</sup>, BEI Tai-Xue<sup>2</sup>

<sup>1</sup>(School of Information and Electrical Engineering, Shandong Jianzhu University, Jinan 250101, China)

<sup>2</sup>(School of Mechanical and Electrical Engineering, Shandong Jianzhu University, Jinan 250101, China)

<sup>3</sup>(Shandong Provincial Key Laboratory of Intelligent Building Technology, Jinan 250101, China)

**Abstract:** As small targets in the early stage of a fire are difficult to detect during flames and smoke detection of fires, this study proposes an improved YOLOX-nano (ASe-YOLOX-nano) object detection algorithm based on natural exponential loss (eCIoU). Firstly, a new object detection function, the eIoU loss function, is proposed to replace the traditional IoU loss, which solves the problems of no intersection between the prediction box and the real frame in small target detection and the inability to react to the influence of width and height. Secondly, the attention module is introduced in the network model to vaguely locate the target position in the early stage of the network and improve the accuracy of the detection of targets, especially small targets, in the later stage of the network. In addition, the soft pooled spatial pyramid pooling structure is employed to extract spatial feature information of different sizes, which can improve the robustness of the model for spatial layout and object degeneration. In this way, sufficient features can be extracted when the target is small. Moreover, the Mosaic enhancement technology is used to preprocess the dataset to improve the generalization ability of the model for further improvement in network performance. The comparative verification of the target data set shows that

① 基金项目: 国家自然科学基金 (61903227); 山东省重点研发计划 (2019GGX104105)

收稿时间: 2022-08-12; 修改时间: 2022-09-15; 采用时间: 2022-09-27; csa 在线出版时间: 2022-12-09

CNKI 网络首发时间: 2022-12-15

the mAP index reaches 70.07%, which is 3.46% higher than that of the original model, and the model enjoys accuracy of flame and smoke detection of 84.66% and 74.56%, respectively, and a stable FPS of 73, which has better fire detection ability than the traditional YOLOX-nano algorithm.

**Key words:** fire detection; attention mechanism; soft pooling spatial pyramid; loss function; YOLOX

随着全球二氧化碳排放量的增加,引起全球变暖,从而导致火灾的发生概率也逐渐增加,每年因火灾造成的经济损失和人员伤亡都是巨大的。其中,建筑物中或工厂设施中发生的火灾占很大的比重。2018年全国共发生23.7万起火灾,造成共计1407人死亡,其中因建筑火灾造成死亡的占比80%左右<sup>[1]</sup>。在国际上的火灾造成的损失也是巨大的。在美国,2018年公共消防部门共报告了131.85万起火灾,共造成3655人死亡,15200人受伤,其中建筑火灾约49.9万起,造成2910人死亡。同一年,韩国报告的火灾共42337起,66%的火灾发生在建筑物内<sup>[2]</sup>。因此,对于发生率较高的建筑类火灾,提前预防和及时应对对保护生命和财产安全起到至关重要的作用。

当火灾发生时,往往会存在多种情况,有的只存在火焰,有的只存在烟雾,有的同时具有两者。因此,火灾检测的主要目标是火焰和烟雾。目前,常用的火灾检测方案主要包括物理传感器检测、传统的图像处理、深度学习图像检测等。

早期检测火灾的有效方法主要是通过物理探测器进行检测,例如,在大型商场里配有红外火焰探测器,烟雾传感器等检测火灾常用的设备。但是,这些传感器通常需要离火源很近才能够起到检测作用,并且,它们只能提供是否存在火源这一信息,火灾大小、火灾位置这些细节信息不能够体现出来。而视觉传感器的检测范围要远远大于物理探测器,因此,国内外专家学者逐渐开始尝试使用视觉传感器作为火灾检测的解决方案。

传统的视觉传感器火灾检测方式是采用人工设计火焰检测特征,例如:颜色、纹理、动态等特征。其中颜色是火灾检测的最主要的特征,现存较有效的有RGB、HIS、YCbCr等颜色模型。binti Zaidi等人<sup>[3]</sup>提出了基于RGB和YCbCr特征的火灾检测方法。Li等人<sup>[4]</sup>提出了基于Dirichlet过程高斯混合颜色模型的视频自主火焰检测模型。Wang等人<sup>[5]</sup>通过对森林火灾进行特征提取,其通过颜色、纹理、形状等多种特征的综合提取来建立识别火灾的模型。以上研究人员通过研究

火灾特征,人工设计特征提取模型,很大程度上满足了火灾检测的需求,为视觉检测火灾奠定了基础。但是人工设计的特征存在特殊性,不能满足复杂的火灾环境。

近年来,随着深度学习的发展,越来越多的领域开始结合深度学习,并取得良好的效果。尤其是卷积神经网络(convolutional neural network, CNN)在对图片和视频的分类和识别方面取得了重大的突破,使其越来越多地被用在视觉检测方面。受此影响,Sharma等人<sup>[6]</sup>结合VGG16和ResNet50,提出了一种火灾检测模型;Yang等人<sup>[7]</sup>参考MobileNet提出了用于火灾检测的CNN模型。Zhang等人<sup>[8]</sup>提出了加入SqueezeNet结构的Attention U-Net模型,并命名为ATT Squeeze U-Net,作为火灾检测系统。

上述火灾检测方法相对于传统图像处理方法,其泛化能力有了很大的提高。随着深度学习的发展,越来越精确的模型被提出,但随之而来的是模型所占内存的增长,已经不能满足轻量化的要求。考虑到火灾检测在移动设备检测的需要,在满足检测精度的要求下,如何尽量减小参数大小成为国内外专家学者关注的研究热点。

因此,本文提出了基于ASe-YOLOX-nano的目标检测算法,通过增加注意力机制(attention mechanism, AM),提高模型对火灾位置的兴趣,增强检测性能,并将空间金字塔池化(space pyramid pooling, SPP)<sup>[9]</sup>结构中的MaxPool换成SoftPool结构,提高模型保留特征的能力,获取更加准确的特征信息,用改进的损失函数(eCIoU)替换训练过程中的目标框回归损失函数IoU,提高目标框的回归准确性。

## 1 相关工作

### 1.1 目标检测算法

深度学习的方法可以通过大量图片提取目标特征,获得泛化信息,具有较好的学习能力和适应能力。基于深度学习的计算机视觉主要的一个分支就是目标检测。

目标检测算法中的杰出代表是Redmon等人<sup>[10]</sup>提

出的YOLO,其通过卷积神经网络对输入图像进行特征提取和分类,整体通过端对端的操作,在原始图像上对目标进行框定并给出其中的类别,具有较好的识别效果,但精度略差.为了提高精度,Redmon等人<sup>[11]</sup>随后提出了YOLOv2算法,其通过K-means对先验框进行聚类,通过将先验框分成3种尺寸,每个尺寸再细分为3个大小,分别对应大中小3种尺寸的目标,从而提高检测精度,但其对小目标的检测效果表现不佳.Redmon等人<sup>[12]</sup>随后又提出了YOLOv3算法,在YOLOv2的基础上,利用特征金字塔结构(feature pyramid networks, FPN)<sup>[13]</sup>通过对不同尺寸的特征融合来提高检测性能,在小目标检测方面检测效果提升明显.Wang等人<sup>[14]</sup>在YOLOv3的基础上提出了DSE-YOLO模型,用于草莓检测,取得很好的效果.但是,YOLOv3对于特征复杂的目标物的检测效果不尽人意.Bochkovskiy等人<sup>[15]</sup>在Redmon在YOLO系列的基础上进行多方面的优化,提出了YOLOv4,在应用方面取得了较好的性能指标.Jocher等人<sup>[16]</sup>在YOLOv4的基础上提出了YOLOv5,其修改了损失函数,加入了Focus结构,输入端采用自适应锚框计算等优化方法.与之前的YOLO模型相比,其具有更加轻量化的结构和更加精确的精度.但是由于其采用的是锚框作为初始预设框,因此计算量会增加,且会影响模型性能.因此,Ge等人<sup>[17]</sup>提出了YOLOX,采用Anchor-free技术,实现预测框通过网络自动生成,弥补了之前YOLO算法锚框的缺点.

## 1.2 本文方法

为了在火灾检测任务中检测出火焰和烟雾,并保证小目标检测效果,本文做出以下贡献.1)采用YOLOX的Darknet-53作为骨干特征提取网络,并在主干中增加了注意力机制,以提高火焰和烟雾检测精度及对小目标的感知精度.2)采用改进的eDIoU函数来加强小目标预测目标框的回归效果,从而提高检测效果.3)采用空间金字塔池化结构,并将其中所有MaxPool池化层替换为更加有效的SoftPool,提升模型对于空间布局 and 物体变性的鲁棒性.4)通过采用Mosaic增强技术来增强数据,增强模型泛化能力.此外,通过采用深度卷积来替代传统的卷积结构,以降低计算量,降低设备要求.

## 2 算法设计和实现

### 2.1 主干网络

YOLOX是旷世科技提出的YOLO系列的一种强

大目标检测框架,其最大的优点是抛弃了之前的锚框的机制,采用Anchor-free技术对目标进行解析识别.本文在YOLOX的nano版本基础上提出了改进版的ASe-YOLOX-nano的目标检测结构.为了提高网络的检测性能,ASe-YOLOX-nano通过融合多种现存的高性能的网络改进的方法,并修改其中的损失函数,以提高网络对于目标的检测性能.YOLOX网络结构主要分为骨干特征提取、多尺度预测、解耦预测结构3个部分.本文以其为基础网络提出的整体框架图如图1所示.

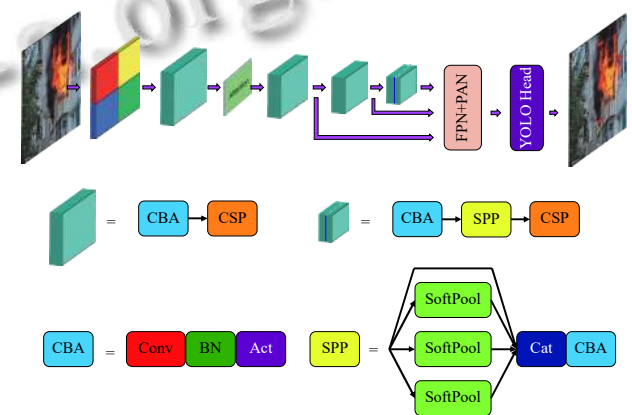


图1 ASe-YOLOX-nano的网络结构图

### 2.2 注意力机制的特征加强结构

注意力机制是聚焦于局部信息的机制,通过仿照人类的注意力方式,对目标区域投入更多的注意力资源,以获得更多的目标信息,并且抑制背景信息.其中主要有Hu等人<sup>[18]</sup>提出的(squeeze-and-excitation networks, SE)、Wang等人<sup>[19]</sup>提出的(efficient channel attention module, ECA)以及Woo等人提出的卷积块注意模块(convolutional block attention module, CBAM)<sup>[20]</sup>.本次采用的是CBAM,其是将通道和空间注意力机制进行结合,通过对通道进行注意力操作获取更多目标所在通道信息,对空间进行注意力操作,保证目标位置信息被更好获取,提高目标检测的准确性.CBAM注意力机制如图2所示.

根据Foggia等人<sup>[21]</sup>的研究表明,火焰在不同的通道上的特征信息是不同的,而且每个特征图上的火焰的位置信息也不同,因此,采用通道注意力和空间注意力机制结合的CBAM对其进行处理,提高模型对于目标的兴趣,增强获取特征目标的有效性.

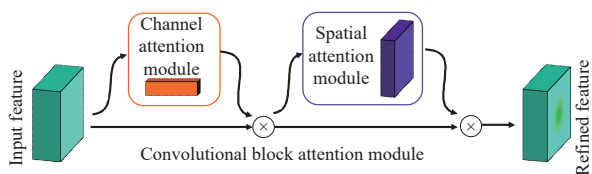


图2 注意力机制 CBAM 模块

### 2.3 改进的空间金字塔池化结构

空间金字塔池化是一种多尺度特征融合的池化方法,能够很好地保留物体特征并维持特征图形状,可以在输入任何大小的特征图像的情况下,输出固定大小的特征。池化操作运算是深度学习领域中对图像处理的一种最基本的算法,通过保留部分特征,减小模型的特征图的大小,同时减少计算成本,防止过拟合,提高模型的泛化能力。在 CNN 中,常见的池化方法是平均池化和最大池化。平均池化是对邻域特征点进行平均运算,可以很好地保留背景信息,使图像更加平滑,但会造成特征信息丢失。最大池化是获取邻域特征点的最大值,可以很好地提取明显的纹理特征信息,因此更适合作为卷积神经网络中保留突出特征和加快模型响应速度的部分,但其也很容易忽略一些详细的特征信息。SoftPool 是池化结构的一个变种,它具有原始的池化层功能,同时保留更多的特征信息,减小池化过程中详细特征丢失的风险,并且对小目标有良好的特征保留效果。因此,本文考虑使用 SoftPool 将空间金字塔池化中的 MaxPool 取代,以达到更好地保留火焰的详细特征,提高检测效果的目的。SoftPool 的计算过程如图 3 所示。

定义局部特征区域为  $m$ ,  $R$  为池化核的大小为  $k \times k$ , 维数表示为  $C \times H \times W$ , 其中  $C$  表示为通道数,  $H$  表示为特征高度,  $W$  表示为特征宽度。根据每个特征点的值非线性计算其对应的特征权重, 计算表达式如下所示:

$$\omega_i = \frac{e^{a_i}}{\sum_{j \in R} e^{a_j}} \quad (1)$$

其中,  $\omega_i$  表示权重,  $a_i$  表示某点的特征值,  $e^{a_i}$  表示激活值。计算权重的目的是保证特征纹理信息能够传递,并且激活区域  $m$  中的特征值向后传递,并在获得权重后与区域  $m$  中的特征值进行加权求和获得输出结果。

$$\tilde{a} = \sum_{i \in R} \omega_i \cdot a_i \quad (2)$$

其中,  $\tilde{a}$  表示特征点经过 SoftPool 后的输出值, 其是通过对内核邻域  $R$  内所有加权激活的标准求和得到的。如图 4 所示。

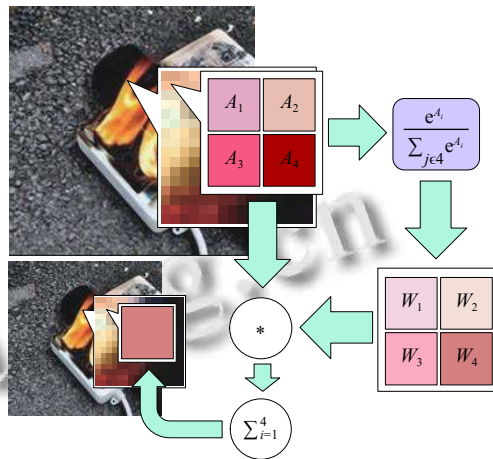


图3 SoftPool 计算流程

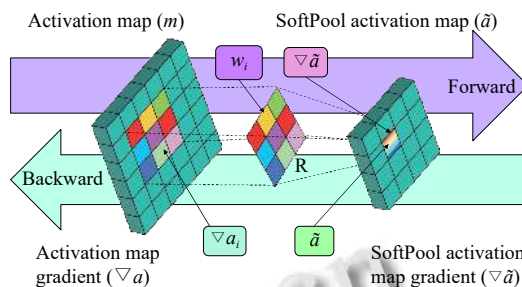


图4 SoftPool 传递过程

### 2.4 损失函数

YOLOX 的总损失计算分为 3 个部分, 采用二元交叉熵损失函数 (BCE) 计算置信度损失和分类损失, 坐标损失是利用交并比损失函数 (IoU) 来计算。边界框坐标损失采用的评价指标通常是交并比。

$$IoU = \frac{|B^{pr} \cap B^{gt}|}{|B^{pr} \cup B^{gt}|} \quad (3)$$

其中,  $B^{pr}$  表示预测框 (corresponding predicted box),  $B^{gt}$  表示真实框 (ground-truth box)。IoU 损失函数的损失函数如式 (4) 所示:

$$L_{IoU} = 1 - IoU \quad (4)$$

IoU 能够一定程度的反应预测框与真实框之间的重合关系, 但不能反映两者之间的距离关系, 并且在两者之间没有交集时, IoU 的数值为 0, 无法进行梯度传递, 实际应用中会在分子处增加一个非常小的值, 确保

$L_{IoU}$ 能够进行梯度回传. 但这样会导致 loss 计算存在误差, 影响最终结果. 为改善这一效果, 后续提出了很多  $IoU$  的优化方法, 例如 Rezatofighi 等人<sup>[22]</sup>提出了  $GIoU$  损失函数, 如式 (5) 所示:

$$GIoU = IoU - \frac{C - (B^p \cup B^{gt})}{C} \quad (5)$$

通过增加一个包裹真实框与预测框的最小框,  $C$  表示其面积, 根据式 (5) 计算出  $GIoU$  值.  $GIoU$  损失函数如式 (6) 所示:

$$L_{GIoU} = 1 - GIoU \quad (6)$$

尽管  $GIoU$  解决了在  $IoU$  作为损失函数时梯度无法计算的问题, 且加入了最小外包框作为惩罚项. 但当两个框是包含关系时,  $GIoU$  退化为  $IoU$ , 无法判断坐标损失的好坏. 对此, Zheng 等人<sup>[23]</sup>提出了  $DIoU$ , 其公式如式 (7) 所示:

$$DIoU = IoU - \frac{\rho^2(B^{pr}, B^{gt})}{c^2} \quad (7)$$

其中,  $\rho(B^F, B^{gt})$  表示真实框和预测框中心坐标的欧式距离,  $c$  表示包裹它们的最小框的对角线距离, 如图 5 所示.

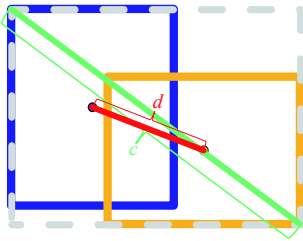


图 5  $DIoU$  距离定义图

当两框中心点重合,  $c$  与  $d$  值不变. 考虑到该种情况, Zheng 等人<sup>[23]</sup>引入宽高比来解决这一问题, 并将其定义为  $CIoU$ , 计算公式如式 (8) 所示:

$$CIoU = IoU - \frac{\rho^2(B^{pr}, B^{gt})}{c^2} - \alpha v \quad (8)$$

其中,  $\alpha$  是权重参数,  $v$  是度量纵横比的一致性.

$$\alpha = \frac{v}{(1 - IoU) + v} \quad (9)$$

$$v = \frac{4}{\pi} \left( \arctan \frac{w^{gt}}{h^{gt}} - \arctan \frac{w^{pr}}{h^{pr}} \right) \quad (10)$$

得到两个损失函数最终表示如下:

$$L_{DIoU} = 1 - DIoU \quad (11)$$

$$L_{CIoU} = 1 - CIoU \quad (12)$$

但是在计算过程中, 由于存在真实框与预测框中心距离太远, 有且是小目标时, 会通过增大预测框大小的方式来减小误差, 这会导致损失降低的过程更加繁琐, 增加模型迭代到稳定的次数. 因此, 本文提出了一种新的计算边界框坐标损失的函数, 以改善  $DIoU$  存在的计算过程复杂的问题, 其公式如式 (13) 所示:

$$eDIoU = IoU - \left( 1 - e^{-\left( \frac{\rho(B^{pr}, B^{gt})}{\rho(B^{pr}, B^{gt}) + 1} \right)} \right) \quad (13)$$

其中,  $\rho(B^{pr}, B^{gt})$  表示真实框与预测框的中心点欧式距离, 利用自然指数对中心点距离整合, 如式 (13) 所示, 通过  $y = 1 - e^{-x}$  的函数特性, 保证了取值范围在 0 至 1 之间, 能够在欧式距离较大时快速回归. 同时, 为了加快预测框与真实框之间的回归速度, 将其变成  $d/(d+1)$  的形式, 使其在中心点距离  $d$  较大时能够具有更大的下降梯度. 同时, 考虑到预测框和真实框中心点重合的情况, 在  $eDIoU$  的基础上增加了新的损失部分, 保证其能够反映宽高比的关系. 其公式如式 (14) 和式 (15) 所示.

$$\alpha = \frac{v}{(1 - IoU) + v} \quad (14)$$

$$v = \rho^2 \left( e^{-\left( \frac{w^{pr}}{h^{pr}} \right)} - e^{-\left( \frac{w^{gt}}{h^{gt}} \right)} \right) \quad (15)$$

其中,  $w^{pr}$  和  $h^{pr}$  表示预测框的宽高,  $w^{gt}$  与  $h^{gt}$  表示真实框的宽高, 采用自然指数特性计算损失项, 能够保证梯度反向传播时的偏导数的存在,  $eCIoU$  表达式如式 (16) 所示, 当中心距  $d$  较大时, 损失的主要由中心距来决定, 反之, 当中心距  $d$  较小时, 主要由  $IoU$  和两框之间的宽高比来决定.

$$eCIoU = eDIoU - \alpha v \quad (16)$$

最终的损失函数如式 (17) 和式 (18) 所示:

$$L_{eDIoU} = 1 - eDIoU \quad (17)$$

$$L_{eCIoU} = 1 - eCIoU \quad (18)$$

## 2.5 数据增强

Mosaic 增强技术是利用 4 张图片, 进行缩放、平移、翻转、色域变换等操作, 然后进行拼接, 每张图片上都有对应的目标框, 拼接后, 获得一张包含 4 张图片目标框的图片, 这样能够极大地丰富在目标出现的环境. 其实现过程如图 6 所示.

图 6(a) 为变换后的 4 张图片, 图 6(b) 为拼接过后的效果, 可以看出, 拼接完成的图片相对于原先 4 张图片来说背景更加丰富了, 目标出现的位置也更具有多样性, 因此可以很好地扩充原有数据集, 增强模型目标检测精度, 提高检测鲁棒性。

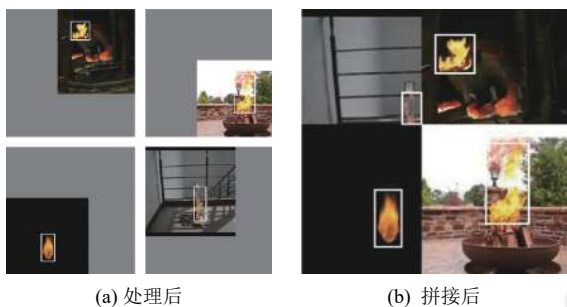


图 6 Mosaic 图片处理

### 3 实验和分析

#### 3.1 实验环境

为了实现对本文提出的方法和其他目标检测方法的训练和测试, 采用的实验设备是 CPU 为 11th Gen Intel(R) Core(TM) i5-11260H @ 2.60 GHz, 8 GB 的 RAM, 显卡设备为 NVIDIA GeForce RTX 3060 Laptop GPU. 训练参数如下: 采用 Adam 优化器, 共设置 100 个迭代周期, 其中 1-50 个周期对主干网络进行冻结训练, 初始学习率为 0.001, weight decay 为 0.0005, 批次为 8; 51-100 个周期进行解冻训练, 初始学习率为 0.0001, weight decay 为 0.0005, 批次为 2.

#### 3.2 评价指标

为了评价 ASe-YOLOX-nano 网络以及其他网络性能, 采用以下几种性能指标作为综合评价指标: 精确率、召回率、F1、平均精度 (mean average precision, mAP), 使用 Size 来衡量模型的大小, 使用 FPS 衡量模型实时性. 其部分公式定义如下所示.

$$P = \frac{TP}{TP + FP} \quad (19)$$

$$R = \frac{TP}{TP + FN} \quad (20)$$

$$F1 = 2 \times \frac{P \times R}{P + R} \quad (21)$$

$$AP_i = \int_0^1 P_i(R_i) dR_i \quad (22)$$

$$mAP = \frac{1}{n} \sum_{i=1}^n AP_i \quad (23)$$

其中, TP 表示真阳性样本, FP 表示假阳性样本, FN 表示假阴性样本. F1 是精确率和召回率的调和平均值. F1 值越高, 精确度和召回率都高. AP 表示每个目标对象类别的性能. mAP 指标表示平均精度的平均值, 被用作衡量目标检测算法模型的整体检测精度指标.

#### 3.3 数据集

数据集采用的是火焰及烟雾数据集 (<http://www.yongxu.org/databases.html>) 以及从网络和现实拍摄的图片, 并通过数据集标注工具 LabelImg 对获取的图片进行标注, 其数据集格式使用的 VOC2007 数据集格式, 共包括各种场景的火焰和烟雾图片共 3 688 张图片, 其中部分如图 7 所示.

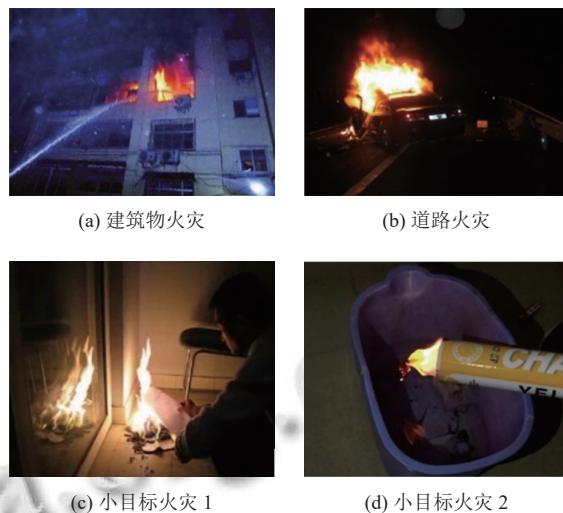


图 7 数据集部分图

此外, 数据集被按照的 9:1 的比例进行分割, 分别用作训练和验证, 其中训练部分又被分成训练集和验证集, 其比例为 9:1, 防止训练过程中过拟合的发生.

#### 3.4 损失比较

为了体现本文提出的损失函数效果, 我们设计了几种目标框和预测框之间的关系, 并做了可视化处理, 其效果如图 8 所示.

图 8 中蓝色框代表真实框, 红色框代表预测框, 图 8(a) 表示的是预测框和真实框重合的情况. 对以上 6 种情况分别计算 IoU、GIoU、eIoU、eCIoU 的值, 其结果如表 1 所示.

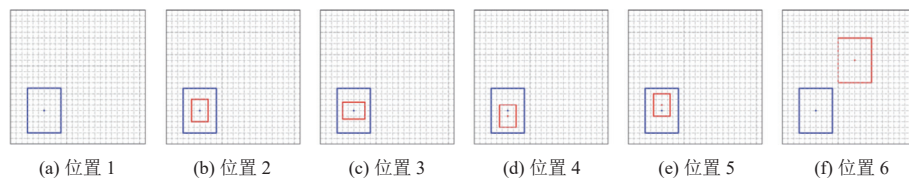


图8 目标框与预测框位置关系

表1 不同位置损失值

Value	1	2	3	4	5	6
<i>IoU</i>	1.000	0.250	0.250	0.250	0.250	0.000
<i>GIoU</i>	1.000	0.250	0.250	0.250	0.250	-0.565
<i>eDIoU</i>	1.000	0.250	0.250	-0.143	-0.143	-0.629
<i>eCIoU</i>	1.000	0.250	0.247	-0.143	-0.143	-0.629

表1中的值越接近1说明真实框和预测框越接近。从表中可以看出,当预测框和真实框没有交集时,*IoU*的值变为0,不能进行梯度回传。图8(d)和图8(e)的*IoU*相同,且图8(d)和图8(e)预测框的中心与真实框的中心的偏移量相同,因此其数值也相同。图8(b)和图8(c)预测框的大小相同,但是宽高比不同,所以此时*IoU*、*GIoU*和*eDIoU*不能展现其中它们之间的区别,而*eCIoU*能够很好地表现出宽高比对于值的影响。图8(f)与其他几幅图相比,中心距最大,因此其与1的差值也是最大,满足预测框与真实框位置相距越远,差值与1相差越大的规律,能够实现预测框的回归需求。

### 3.5 实验结果

为了直接和客观的体现网络模型和算法的性能,结合上述工作,综合提到5种评价指标对火焰烟雾检测算法 ASe-YOLOX 进行评估,并与原始的 YOLOX 模型在同一个数据集上进行比较,结果如表2所示。比较结果表明,本文所提出的 ASe-YOLOX 算法的 *mAP* 为 70.07%,比原始的算法的 66.61% 提高了 3.46%,这表明所提出的算法拥有更好的目标检测性能。在火焰和烟雾的 *AP* 值上,本文提出的算法比原始的 YOLOX-nano 算法分别提高了 5.34% 和 1.57% 的值, *P* 值火焰下降 0.96%, 烟雾提高了 1.84%, *R* 值分别提高了 8.57% 和 0.74%, *F1* 则分别提高了 5.3% 和 1.13%。对比结果表明,本文提出的算法火焰的 *P* 存在轻微下降,但检测性能有所提高,其能够更好地检测火焰和烟雾。与原始算法相比,其模型体积略有增长,检测的 FPS 有所下降,但其 FPS 为 73 的检测帧率能够满足实时性要求。综上所述,其总体性能远优于原始算法。

表2 原模型与改进模型对比实验结果

Model	class	<i>AP</i> (%)	<i>P</i> (%)	<i>R</i> (%)	<i>F1</i> (%)	<i>mAP</i> (%)	Size (KB)	FPS
YOLOX-nano	Fire	71.11	<b>85.00</b>	58.52	69.31	66.61	3754	75
	Smoke	62.11	73.60	50.15	59.65			
ASe-YOLOX-nano	Fire	<b>76.45</b>	84.04	<b>67.09</b>	<b>74.61</b>	<b>70.07</b>	3813	73
	Smoke	<b>63.68</b>	<b>75.44</b>	<b>50.89</b>	<b>60.78</b>			

多次训练得到最好的模型数据后,本文使用实际场景中的火灾图像进行测试,其检测结果如图9所示。

图9中,前4幅图是建筑物火灾全面发展阶段的检测结果,后2幅图则是对小目标检测的结果。图9(a)~图9(d)所示是对建筑物火灾的检测结果,可以看出其能够准确识别火焰和烟雾。图9(e)和图9(f)是对小目标火灾得到识别结果,可以看出其不存在漏检的情况。通过检测结果可以看出,本文设计的网络对于火焰和烟雾有着较好的检测效果,并且对于小目标的检测有着良好的准确度。

为体现本文提出的网络结构在小目标检测上的性

能优于原始网络,我们对完全燃烧的小目标火焰以及火灾初期的小目标火焰进行对比测试,其对比检测结果如图10所示。

根据图10(a)、图10(b)可以看出在对充分燃烧的小目标火焰的检测其原始网络的检测精度是 0.67,改进网络的检测结果是 0.73,检测精度有了明显提升;对比图10(c)、图10(d)结果,证明改进的方法能够有效改善小目标火焰的漏检问题,图10(e)、图10(f)是对火灾初期小目标火焰的一个对比试验,原始网络检测精度为 0.56,改进网络的精度提升到 0.68,其结果表明能够在火灾初期的小目标检测上效果更好。综上,可

以证明本文提出的改进网络具有更好的小目标检测效果.

制. 综上所述, 每个优化算法的使用使得 ASe-YOLOX-nano 算法在火焰和烟雾检测方面能够取得了更好的结果.



图9 ASe-YOLOX-nano 检测结果

### 3.6 消融实验

为了体现出改进的算法对检测性能的直观影响, 进行一系列的消融实验, 使用原始 YOLOX-nano 作为骨干网络, 逐步增加优化算法, 测试其对网络检测性能所起到的作用, 其实验结果如表 3 所示.

表 3 中打勾表示使用了该部分的优化算法, 空白的部分表示有使用此优化算法. 在将  $IoU$  损失函数换成  $eDIoU$  和  $eCIoU$  后, 由于能够更准确地识别目标, 所以算法性能有一定的上升. 增加 Mosaic 和使用改进后的 SPP 后,  $mAP$  再次提升到 68.31% 和 68.62%, 在增加不同 AM 算法后, 网络性能有了不同的变化, 其中 CBAM 效果最好,  $mAP$  增长到 70.07%, ECA 次之, 增长到了 69.64%, SE 的  $mAP$  值反而下降到了 67.33%, 因此, 最终的模型采用的是 CBAM 注意力机

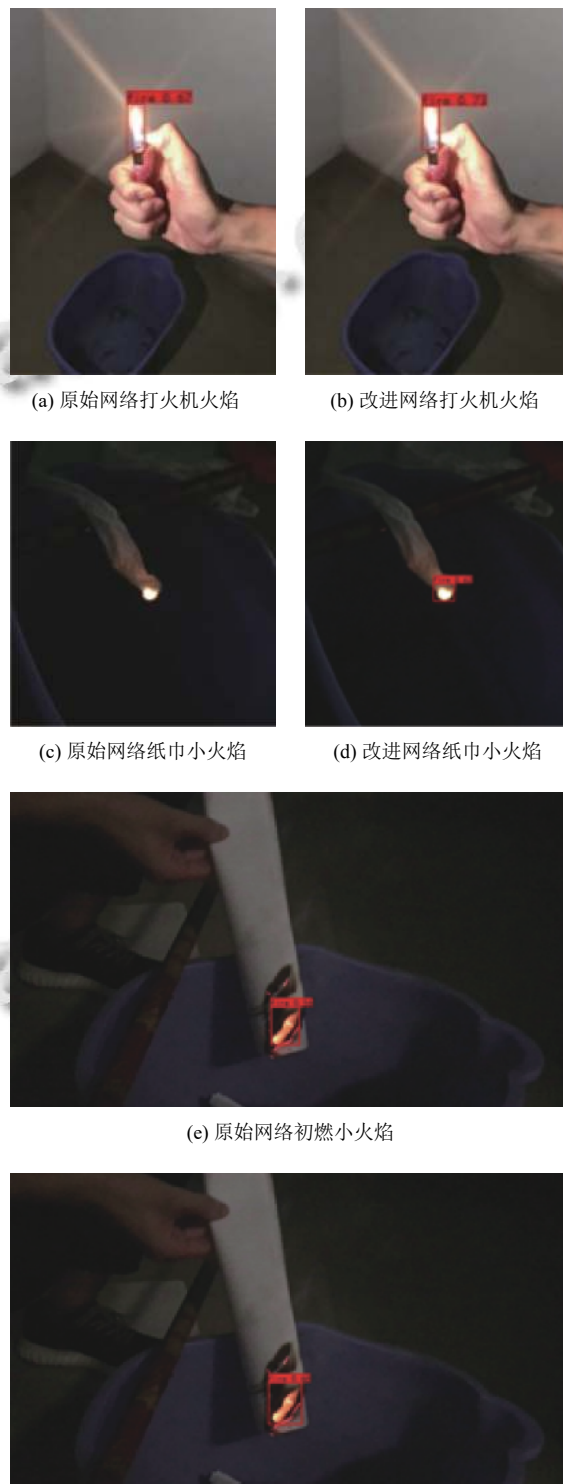


图10 小目标对比检测结果



表3 消融实验结果

<i>eDIoU</i>	<i>eCIoU</i>	Mosaic	SPP	AM			<i>mAP</i> (%)
				CBAM	SE	ECA	
							66.61
√							66.88
	√						66.90
	√	√					68.31
	√	√	√				68.62
	√	√	√	√			<b>70.07</b>
	√	√	√		√		67.33
	√	√	√			√	69.64

#### 4 结论与展望

本文提出了一种基于YOLOX-nano的神经网络模型,用于检测火灾。采用注意力机制使得模型一开始就能模糊定位到目标,然后采用池化金字塔结构对特征提取进一步加强,最后在训练过程中,采用改良的损失函数进行训练,并且采用VOC2007的预训练权重作为初始权重加快模型回归。提出的改进算法与原网络进行对比实验,结果表明,其性能要优于原网络,尤其是在对于小目标的检测方面,同时对每个优化部分进行消融实验,每个优化部分都能很好地提高算法识别性能。提出的改进方法可以用于工厂车间等其他建筑内,可以起到发现早期火情,加快反应速度,从而有助于减少财产损失。

本文算法虽然具有良好的火灾检测效果,但还存在优化空间。下一步工作继续优化模型与损失函数,并研究如何适用于移动设备,希望解决移动设备对设备性能的依赖的问题。

#### 参考文献

- Zhou XP, Li HR, Wang J, *et al.* CloudFAS: Cloud-based building fire alarm system using building information modelling. *Journal of Building Engineering*, 2022, 53: 104571. [doi: 10.1016/j.jobee.2022.104571]
- Jung S, Cha HS, Jiang SH. Developing a building fire information management system based on 3D object visualization. *Applied Sciences*, 2020, 10(3): 772. [doi: 10.3390/app10030772]
- binti Zaidi NI, binti Lokman NAA, bin Daud MR, *et al.* Fire recognition using RGB and ycbcr color space. *ARNP Journal of Engineering and Applied Sciences*, 2015, 10(21): 9786–9790.
- Li ZL, Mihaylova LS, Isupova O, *et al.* Autonomous flame detection in videos with a Dirichlet process Gaussian mixture color model. *IEEE Transactions on Industrial Informatics*, 2018, 14(3): 1146–1154. [doi: 10.1109/TII.2017.2768530]
- Wang Y, Dang L, Ren J. Forest fire image recognition based on convolutional neural network. *Journal of Algorithms & Computational Technology*. 2019, 13: 1–11. [doi: 10.1177/1748302619887689]
- Sharma J, Granmo OC, Goodwin M, *et al.* Deep convolutional neural networks for fire detection in images. *Proceedings of the 18th International Conference on Engineering Applications of Neural Networks*. Athens: Springer, 2017. 183–193.
- Yang HJ, Jang H, Kim T, *et al.* Non-temporal lightweight fire detection network for intelligent surveillance systems. *IEEE Access*, 2019, 7: 169257–169266. [doi: 10.1109/ACCESS.2019.2953558]
- Zhang JM, Zhu HQ, Wang PY, *et al.* ATT squeeze U-Net: A lightweight network for forest fire detection and recognition. *IEEE Access*, 2021, 9: 10858–10870. [doi: 10.1109/ACCESS.2021.3050628]
- 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]
- Redmon J, Divvala S, Girshick R, *et al.* You only look once: Unified, real-time object detection. *Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition*. Las Vegas: IEEE, 2016. 779–788.
- Redmon J, Farhadi A. YOLO9000: Better, faster, stronger. *Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition*. Honolulu: IEEE, 2017. 6517–6525.
- Redmon J, Farhadi A. YOLOv3: An incremental improvement. arXiv:1804.02767, 2018.
- Lin TY, Dollár P, Girshick R, *et al.* Feature pyramid networks for object detection. *Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition*. Honolulu: IEEE, 2017. 936–944.
- Wang Y, Yan G, Meng QL, *et al.* DSE-YOLO: Detail semantics enhancement YOLO for multi-stage strawberry detection. *Computers and Electronics in Agriculture*, 2022, 198: 107057. [doi: 10.1016/j.compag.2022.107057]
- Bochkovskiy A, Wang CY, Liao HYM. YOLOv4: Optimal speed and accuracy of object detection. arXiv:2004.10934, 2020.

- 16 Jocher G. YOLOv5. <https://github.com/ultralytics/yolov5>.
- 17 Ge Z, Liu ST, Wang F, *et al.* YOLOX: Exceeding YOLO series in 2021. arXiv:2107.08430, 2021.
- 18 Hu J, Shen L, Sun G. Squeeze-and-excitation networks. Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City: IEEE, 2018. 7132–7141.
- 19 Wang QL, Wu BG, Zhu PF, *et al.* ECA-Net: Efficient channel attention for deep convolutional neural networks. Proceedings of the 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle: IEEE, 2020. 11531–11539.
- 20 Woo S, Park J, Lee JY, *et al.* CBAM: Convolutional block attention module. Proceedings of the 15th European Conference on Computer Vision. Munich: Springer, 2018. 3–19.
- 21 Foggia P, Saggese A, Vento M. Real-time fire detection for video-surveillance applications using a combination of experts based on color, shape, and motion. IEEE Transactions on Circuits and Systems for Video Technology, 2015, 25(9): 1545–1556. [doi: 10.1109/TCSVT.2015.2392531]
- 22 Rezaeifighi H, Tsoi N, Gwak JY, *et al.* Generalized intersection over union: A metric and a loss for bounding box regression. Proceedings of the 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Long Beach: IEEE, 2019. 658–666.
- 23 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]

(校对责编: 牛欣悦)