

基于深度学习的火焰检测算法设计

姚一^{1*}, 鞠全勇², 吕章锬¹

¹盐城工学院机械工程学院, 江苏 盐城

²金陵科技学院机电学院, 江苏 南京

收稿日期: 2024年4月16日; 录用日期: 2024年5月22日; 发布日期: 2024年5月31日

摘要

为了更好地实现消防机器人对于火源的自主识别, 结合YOLOv5s提出一种改进的火焰检测方法。在YOLOv5s的主干网络和检测头中添加CBAM注意力模块, 能够自适应地对特征图进行调整, 使得网络能够更好地关注重要的特征, 从而降低火焰检测的小目标漏检率。使用C3-DBB模块替换卷积层, C3-DBB模块采用了密集的连接和多尺度的感受野, 从而能够更好地捕捉目标的细节信息, 减少信息损失。使用ConvNeXt V2模块替换主干网络模块, 从而获得更大的感受野和更好的特征提取能力, 提高目标检测的精度。实验结果表明, 与YOLOv5s原始模型相比, 改进后的模型平均精度提升5.8%, 同时降低了模型大小。

关键词

火焰检测, 注意力机制, YOLOv5s算法, 深度学习

Design of Flame Detection Algorithm Based on Deep Learning

Yi Yao^{1*}, Quanyong Ju², Zhangkun Lv¹

¹School of Mechanical Engineering, Yancheng Institute of Technology, Yancheng Jiangsu

²School of Mechanical and Electrical Engineering, Jinling Institute of Technology, Nanjing Jiangsu

Received: Apr. 16th, 2024; accepted: May 22nd, 2024; published: May 31st, 2024

Abstract

In order to better realize the autonomous recognition of fire sources by firefighting robots, an improved flame detection method is proposed in combination with YOLOv5s. Adding the CBAM at-

*第一作者。

attention module to the backbone network and detection head of YOLOv5s can adaptively adjust the feature map, enabling the network to better focus on important features, thus reducing the small target miss detection rate of flame detection. Replacing the convolutional layer with the C3-DBB module, the C3-DBB module employs dense connectivity and a multi-scale receptive field, which enables better capture of detailed information about the target and reduces information loss. The ConvNeXt V2 module is used to replace the backbone network module, thus obtaining a larger receptive field and better feature extraction capability to improve the accuracy of target detection. The experimental results show that the improved model improves the average accuracy by 5.8% compared to the original YOLOv5s model, while reducing the model size.

Keywords

Fire Detection, Attention Mechanism, YOLOv5s Algorithm, Deep Learning

Copyright © 2024 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

随着经济的快速发展,城市建筑物规模的增大,火灾的复杂性也在增加,使得灭火工作的难度越来越大。因此,研发更为先进的消防手段,如使用消防机器人进行灭火,已经成为研究的趋势。消防机器人对火焰目标检测有着较强的实时性和准确性要求,而通过研究基于图像的火焰目标检测方法,可以提高消防机器人的灭火效率和准确性,减少消防员的伤亡风险。

目标检测在计算机视觉领域是一个重要的研究方向,用于检测图像或视频中的特定物体或目标。目前存在两种主要的目标检测算法:双阶段检测算法和单阶段检测算法。双阶段检测算法首先通过一种 region proposal 方法生成一系列候选区域,这些区域可能包含待检测的物体。然后,这些候选区域通过卷积神经网络进行物体检测。双阶段检测算法包括 Faster R-CNN [1]、Mask R-CNN [2]和 Cascade R-CNN [3]等。这些算法的优点是检测精度较高,但由于需要两个阶段的处理,检测速度相对较慢。单阶段检测算法将候选区域生成和物体检测合并为一个阶段。这类算法具有简单的网络结构和快速的检测速度。常见的单阶段检测算法包括 RetinaNet [4]、CenterNet [5]和 YOLO [6]系列算法。然而,这些算法在检测准确性和鲁棒性上仍有提升的空间,尤其对于小目标的检测能力相对较差。这些限制导致在城市高层建筑灭火和小火点灭火等场景中,火焰检测算法的应用局限性较大。

为解决以上问题,本文基于 YOLOv5s 算法提出一种改进模型的火焰检测算法。首先,使用 ConvNeXtv2 模块替换主干网络模块,能够提供更高的感受野和更好的特征提取能力,从而在目标检测任务中获得更高的精度。其次,在 YOLOv5s 的主干网络和检测头中添加 CBAM 注意力模块,提升网络检测小目标的能力,降低火焰检测的小目标漏检率。最后,使用 C3-DBB 模块替换原始的卷积层,提高模型对目标的检测能力。

2. 改进 YOLOv5s 算法

2.1. YOLOv5s 算法简介

YOLOv5 (You Only Look Once)是一种单阶段目标检测算法,其主要特点是准确率高、效率高、支持多尺度检测等。自从 Ultralytics 版本的 YOLOv5 发行以来不断被更新,本文改进的是 7.0 版本。YOLOv5

根据宽度、深度的不同提供了 n、s、m、l 等 4 种尺寸模型。其中 YOLOv5s 是 YOLOv5 系列中最小的模型，它具有较少的参数和计算量，可以在保持较高检测速度的同时实现相对较高的检测精度。YOLOv5s 的网络结构和子模块结构如图 1 和图 2 所示。

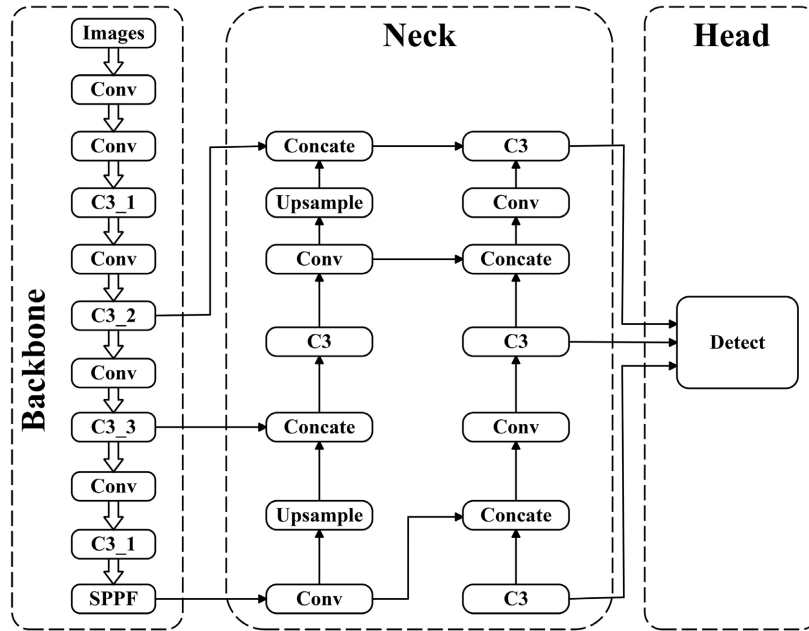


Figure 1. The network structure of YOLOv5s

图 1. YOLOv5s 网络结构

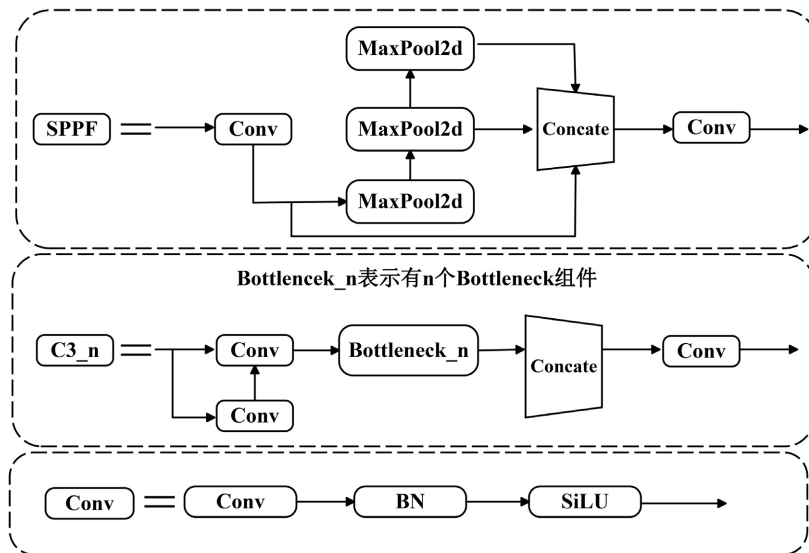


Figure 2. YOLOv5s submodule structure

图 2. YOLOv5s 子模块结构

2.2. 本文算法

本文使用 ConvNeXt V2 模块、CBAM 模块和 C3-DBB 模块改进原有的 YOLOv5s 网络，这些模块的引入旨在提高网络的感知能力、注意力机制和特征融合能力。不仅能够提高特征提取能力，还降低了小

目标的漏检率，提高模型对目标的检测能力，如图 3 所示。

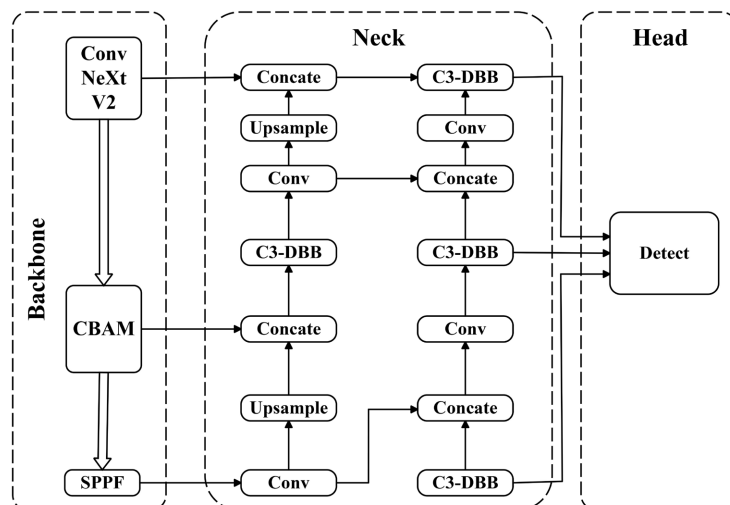


Figure 3. Improving the overall network structure of YOLOv5s

图 3. 改进 YOLOv5s 整体网络结构

2.3. ConNeXt V2

在改进的体系结构和更好的表征学习框架的推动下，视觉识别领域在本世纪 20 年代初获得了快速的现代化和性能提升。例如，以 ConvNeXt 为代表的现代 ConvNet 在各种场景中都表现出了强大的性能。虽然这些模型最初是为使用 ImageNet 标签的监督学习而设计的，但它们也可以潜在地受益于自监督学习技术，如掩蔽自动编码器(MAE)。然而，作者发现简单地将自监督学习技术与 ConvNeXt 结合在一起并不会带来比平均水平更好的性能。因此，作者 Woo S, Debnath S [7]提出了一种全卷积掩码自编码器框架(FCMA)，并引入了一个新的全局响应归一化(GRN)层，该层可以添加到 ConvNeXt 结构中来增强通道之间的特征竞争。

1) FCMA (Fully Convolutional Masked Autoencoder)是在输入图像上进行遮挡并生成学习信号的一种方法,如图所示。它由两个组件组成:ConvNeXt 编码器和 ConvNeXt 解码器,它们使用稀疏卷积进行操作。编码器负责将可见像素作为输入,并将其转换为潜在表示。潜在表示是编码器输出的特征图,其中包含了对可见部分的编码信息。解码器使用潜在表示以及遮挡的补丁(mask tokens),尝试重构原始图像。解码器的目标是通过使用潜在表示和补丁来产生尽可能接近原始图像的重建图像。在计算损失时,只在遮挡的补丁上计算。这意味着只有模型预测的遮挡区域才会对损失函数产生贡献,从而激励模型学习如何正确预测缺失部分。

2) GRN (Global Response Normalization)是一种新的全局响应规范化技术,以使 FCMAE 预训练与 ConvNeXt 架构结合起来更有效。通过给定输入特征 $X \in \mathcal{R}^{H \times W \times C}$, 所提出的 GRN 单元由三个步骤组成:全局特征聚合、特征归一化、特征校准。

首先使用全局函数 $G(\cdot)$ 将空间特征地图 X_i 聚合为向量 gx :

$$g(x) := x \in \mathcal{R}^{H \times W \times C} \rightarrow gx \in \mathcal{R}^C \quad (1)$$

其次,将响应归一化函数 $N(\cdot)$ 应用于聚合值。具体地说,使用标准的可视化规范化:

$$N(\|X_i\|) := \|X_i\| \in \mathcal{R} \rightarrow \frac{\|X_i\|}{\sum_{j=1, \dots, C} \|X_j\|} \in \mathcal{R} \quad (2)$$

最后使用计算的特征归一化分数来校准原始输入响应：

$$X_i = X_i * N(g(X)_i) \in R^{H \times W} \tag{3}$$

如图 4 所示，将 GRN 层引入原始 ConNeXt 中后，不再需要 LayerScale，但增加了参数和 FLOPS。相较于原 YOLOv5s 网络中的 CSP bottleneck 模块，ConvNeXt V2 模块具有自动编码器的能力。全卷积网络的灵活性，能够提供更高的感受野和更好的特征提取能力，从而在目标检测任务中获得更高的精度。

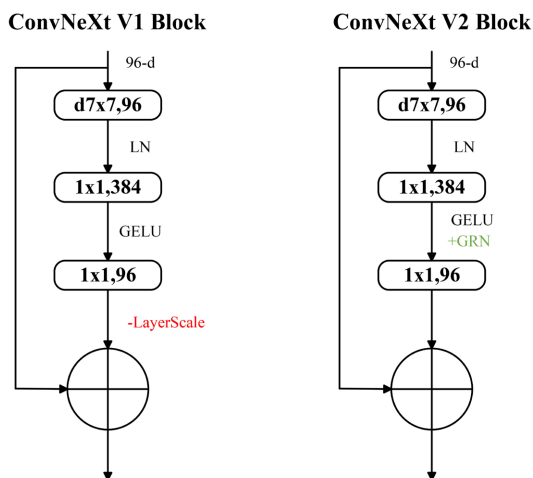


Figure 4. ConNeXt Block designs
图 4. ConNeXt Block 结构变化

2.4. CBAM 注意力

CBAM [8] (Convolutional Block Attention Module)是一种用于增强卷积神经网络(CNN)注意力机制的模块。它由两个子模块组成：通道注意力模块(Channel Attention Module)和空间注意力模块(Spatial Attention Module)，如图 5 所示。

通道注意力模块通过对输入特征图的通道维度进行注意力加权，来自适应地选择和强化重要的特征通道。它通过计算每个通道的全局平均池化和全局最大池化，然后通过两个全连接层产生一个通道注意力图。最后，将该通道注意力图与输入特征图相乘，以增强重要的特征通道。

空间注意力模块通过对输入特征图的空间维度进行注意力加权，来自适应地选择和强化重要的空间位置。它通过使用一个卷积层和一个全连接层产生一个空间注意力图。最后，将该空间注意力图与输入特征图相乘，以增强重要的空间位置。

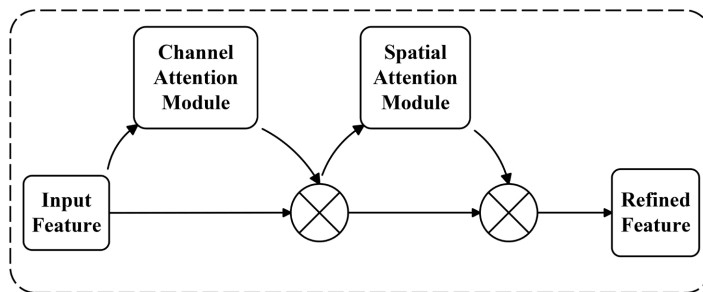


Figure 5. The network architecture of CBAM
图 5. CBAM 网络结构

通过将通道注意力模块和空间注意力模块结合起来，CBAM 模块可以同时考虑通道和空间的注意力信息，从而更好地捕捉重要的特征通道和空间位置。在原先的 YOLOv5s 网络中添加 CBAM 注意力模块，能够增强网络提取图像特征的能力，更好地提取小目标火焰特征，降低漏检率。

2.5. C3-DBB

C3-DBB 模块是由丁霄汉[9]提出的一个卷积神经网络(ConvNet)通用构建块，用于提高没有任何推理时间成本的性能。它通过结合不同尺度和复杂的不同分支来丰富特征空间，包括卷积序列、多尺度卷积和平均池，增强了单个卷积的表示能力。训练结束后，DBB 可以等价地转换为一个单一的转换层以进行部署。与新的 ConvNet 架构不同，DBB 在维护宏架构的同时使训练时的微观结构复杂化，因此它可以用作任何架构的常规转换层的临时替代。这样，该模型可以被训练以达到更高的性能水平，然后转换为原始的推理时间结构进行推理。作者在文章中总结了六个转换以将 DBB 转换为单个卷积，分别是批量归一化(BN)、分支相加、深度串联、多尺度操作、平均池化和卷积序列。

它采用 1×1 , $1 \times 1 - K \times K$, $1 \times 1 - AVG$ 等组合方式对原始 $K \times K$ 卷积进行增强。对于 $1 \times 1 - K \times K$ 分支，设置中间通道数等于输入通道数并将 1×1 卷积初始化为 Identity 矩阵，其他分支则采用常规方式初始化。此外，在每个卷积后都添加 BN 层用于提供训练时的非线性，这对于性能提升很有必要。

在原先的 YOLOv5s 网络中添加 C3-DBB 模块，改善了网络的特征表示能力，以提高模型的性能。通过引入自适应特征融合机制，以在不同的分支中融合特征，进一步的提高了网络的性能。

3. 实验与结果分析

3.1. 实验数据

通过网络获取火焰整体数据集，其中设置了训练图片 1482 张，验证图片 165 张，测试图片 412 张。数据集中包含了多种场景的火灾类型，有大火-小火、建筑起火、草原起火、森林起火、车辆起火、白天-黑夜起火、室内-室外起火等。数据集采用标注工具 Labelimg 进行标注。

3.2. 实验环境

实验平台的搭建如表 1 所示，在模型训练过程中，初始学习率为 0.01，最终学习率为 0.1，使用随机梯度下降(SGD)优化算法，并设置动量为 0.937。优化器的权重衰减为 0.0005，训练批次(batch size)是 8，最大迭代次数(epoch)是 100。在训练完成后，我们将对改进的 YOLOv5s 火焰检测模型和原始 YOLOv5s 模型进行性能对比。我们将使用训练好的改进模型对测试集中的火焰图像进行识别测试，以证明算法的有效性。此外，我们还将对本文提出的改进模型与目前主流的目标检测模型进行性能比较，以进一步综合分析本文方法在火焰检测方面的性能。

Table 1. Experimental platform configuration

表 1. 实验平台配置

项目	配置
操作系统	Windows 11
显卡	NVIDIA GeForce RTX 3050 Laptop GPU
CUDA 版本	12.2
Python	3.8.17
深度学习框架	Py Torch 1.9.0

3.3. 评价指标

采用 IoU (交并比)、平均精度均值(mAP)、精准率(P)、召回率(R)等评价指标, 具体如下:

1) IoU: IoU 是指预测框与真实框之间的交集与并集的比值。在目标检测中, IoU 用于判断预测框与真实框之间的重叠程度。通常, IoU 越高表示检测结果越准确。

2) 平均精度均值(mAP): mAP 是衡量目标检测算法性能的重要指标之一。它是对不同 IoU (交并比) 阈值下的精度进行平均得到的。通常, mAP 越高表示算法的性能越好, 公式如下:

$$m_{\text{mAP}} = \frac{1}{m} \sum_{i=1}^m A_{AP_i} \quad (4)$$

3) 准确率(P): 精确率是指检测出的正样本中真正正样本的比例。在目标检测中, 精确率表示检测出的目标中真正存在的目标的比例。精确率越高表示算法的误检率越低, 公式如下:

$$P = \frac{T_{TP}}{T_{TP} + F_{FP}} \quad (5)$$

其中: T_{TP} 代表 IoU 大于设定阈值的数量; F_{FP} 代表 IoU 小于设定阈值或者对于同一个真实框检测到的多余预测框数量之和。

4) 召回率(R): 召回率是指正确检测出的正样本占有所有正样本的比例。在目标检测中, 召回率表示检测到的目标占有所有真实目标的比例。召回率越高表示算法的漏检率越低。

3.4. 结果分析

在本文中, 原始的 YOLOv5s 网络和改进后的 YOLOv5s 网络使用了相同的数据集、训练参数和训练方法进行训练, 以便有效地分析改进后模型的性能。通过比较每轮训练后的检测精度、准确率等评价指标, 可以评估改进后的 YOLOv5s 网络相对于原始网络的性能提升。从图 6 中也可以看出, 改进后的火焰检测模型在迭代次数到达 50 时逐渐趋向稳定, 在阈值设定为 0.5 的前提下, 检测精度最高达到了 68%, 比 YOLOv5s 原始模型的检测精度提升了 6%。

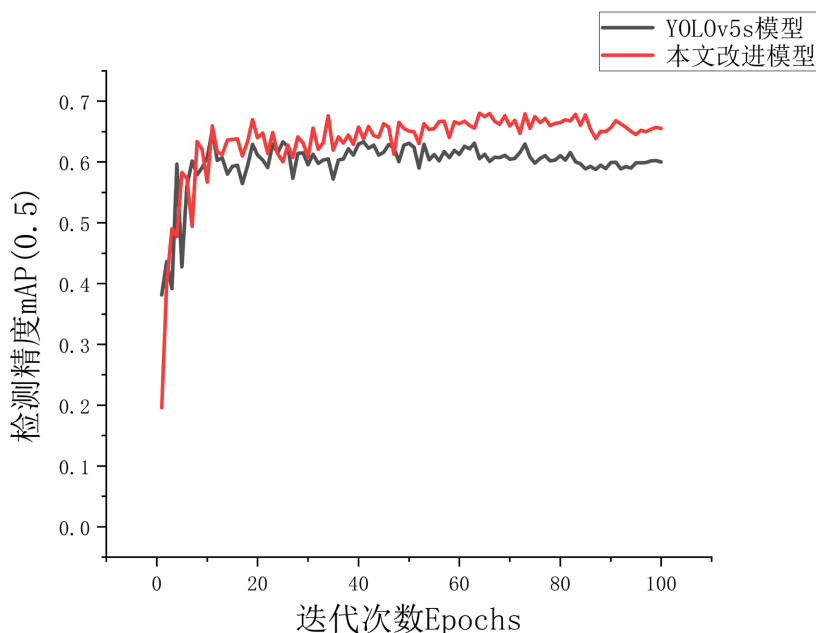


Figure 6. mAP (0.5) Comparison
图 6. mAP (0.5)对比

在训练过程中，损失函数是用来衡量模型预测结果与真实结果之间的差异的指标。网络模型的目标是通过不断调整参数，最小化损失函数的值。随着迭代次数的增加，如果网络模型能够稳定收敛，损失函数的值会逐渐减小并趋向于一个较小的稳定值。本文改进后的模型与 YOLOv5s 模型的损失函数曲线对比如图所示。图 7 中和图 8 中可以看出，改进后的模型具有更快的收敛速度，最低损失值可以降低到 0.015 以下。

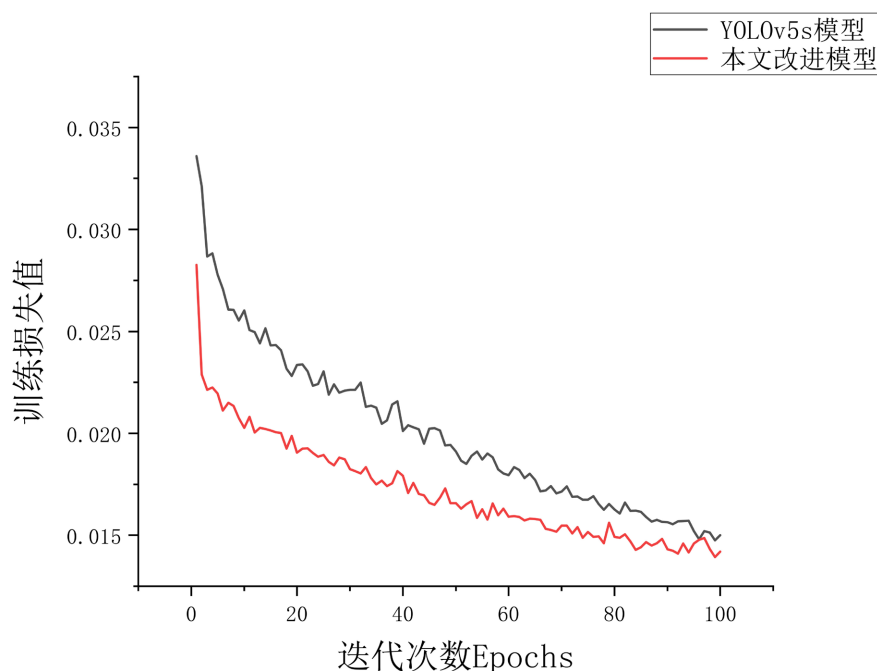


Figure 7. Comparison data of training loss values

图 7. 训练损失值对比数据

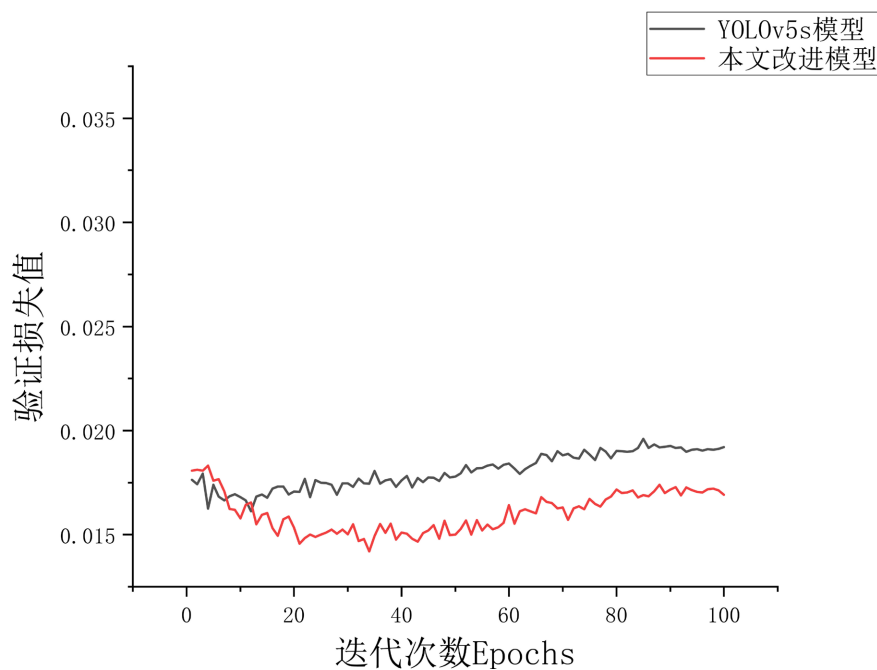


Figure 8. Verifying loss value comparison data

图 8. 验证损失值对比数据

此外, 本文为了分析了不同改进策略对模型检测性能的影响, 分别设计了 4 组实验, 如表 2 所示。每组实验均使用相同数据集和相同训练参数及方式完成训练, 实验结果如表 1 所示。分析可知, YOLOv5s 原始网络没有使用任何改进策略, 检测精度(mAP)为 62.2%; 改进网络 1 在 YOLOv5s 原始网络的基础上引入了协同注意力机制, 将空间位置信息嵌入到通道注意力中, 使得模型在完成依赖位置信息的火焰检测任务时取得了更好的预测效果, 检测精度(mAP)提升了 0.3%; 改进网络 2 在改进网络 1 的基础上增加了 C3-DBB 模块替换卷积层, 可以减少信息损失, 提高模型对目标的检测能力, 使模型检测精度(mAP)提升了 3.4%; 改进网络 3 在改进网络 2 的基础上使用 ConvNeXtv2 模块替换主干网络模块, 获得了更大的感受野和更好的特征提取能力, 从而提高目标检测的精度模型检测精度(mAP)提升了 2.4%。综上所述, 本文所提出的改进 YOLOv5s 模型比 YOLOv5s 原始模型的火焰检测精度(mAP)提升了 5.8%。

Table 2. Comparison of experimental results between the original network and different improvement strategies
表 2. 原始网络与不同改进策略的实验结果对比

网络模型	改进策略			mAP (0.5)	模型大小 (MB)
	注意力机制	C3-DBB 模块	骨干网络		
YOLOv5s 网络	×	×	×	62.2%	13.73
改进网络 1	√	×	×	62.5%	13.73
改进网络 2	√	√	×	65.6%	13.8
改进网络 3	√	√	√	68%	9.97

3.5. 测试结果

本文通过对测试集火焰图像进行测试, 如图 9 所示。测试证明了改进 YOLOv5s 火焰检测模型对陌生火焰图像的检测能力。由于测试集图像数据与训练集和验证集数据无重叠, 由检测结果可知, 基于改进的 YOLOv5s 模型可以准确地识别出陌生火焰图像中的火焰目标, 表明该模型具有良好的泛化性和鲁棒性。此外, 对于城市高层建筑灭火、小火点灭火等情境下的火焰目标检测, 基于该模型的检测方法仍具有良好的准确度。

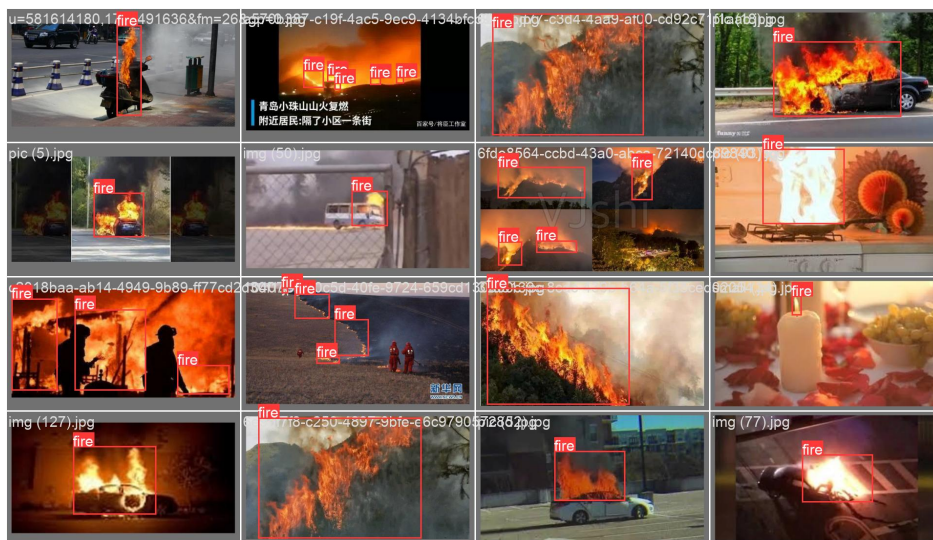


Figure 9. Flame detection test results
图 9. 火焰检测测试结果

4. 结束语

对于改进火焰检测的算法, 本文提出了以下几点改进: 首先添加 CBAM 注意力模块: 在 YOLOv5s 的主干网络和检测头中添加 CBAM 注意力模块, 可以提升网络检测小目标的能力。CBAM 注意力模块能够自适应地对特征图进行调整, 使得网络能够更好地关注重要的特征, 从而降低火焰检测的小目标漏检率; 其次使用了 C3-DBB 模块替换卷积层: 传统的卷积层在一定程度上存在信息损失的问题, 通过使用 C3-DBB 模块可以减少信息损失, 提高模型对目标的检测能力。C3-DBB 模块采用了密集的连接和多尺度的感受野, 从而能够更好地捕捉目标的细节信息; 最后使用 ConvNeXt V2 模块替换主干网络模块: 传统的主干网络采用的是卷积层, 通过使用 ConvNeXt V2 模块可以获得更大的感受野和更好的特征提取能力, 从而提高目标检测的精度。通过以上改进, 可以显著提高火焰检测算法的精度和性能, 提高小目标的检测准确率, 从而适用于更广泛的应用场景。

参考文献

- [1] Ren, S., He, K., Girshick, R., *et al.* (2015) Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. *Advances in Neural Information Processing Systems*. <https://doi.org/10.48550/arXiv.1506.01497>
- [2] Begum, S.R., Datta, S.Y. and Manoj, M.S.V. (2021) Mask R-CNN for Fire Detection. *International Research Journal of Computer Science*, **8**, 145-151. <https://doi.org/10.26562/irjcs.2021.v0807.003>
- [3] Cai, Z. and Vasconcelos, N. (2019) Cascade R-CNN: High Quality Object Detection and Instance Segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **43**, 1483-1498. <https://doi.org/10.1109/TPAMI.2019.2956516>
- [4] Lin, T.Y., Goyal, P., Girshick, R., *et al.* (2017) Focal Loss for Dense Object Detection. *Proceedings of the IEEE International Conference on Computer Vision*, Venice, 22-29 October 2017, 2980-2988. <https://doi.org/10.1109/ICCV.2017.324>
- [5] Duan, K., Bai, S., Xie, L., *et al.* (2019) CenterNet: Keypoint Triplets for Object Detection. *Proceedings of the IEEE/CVF International Conference on Computer Vision*, Seoul, 27 October-2 November 2019, 6569-6578. <https://doi.org/10.1109/ICCV.2019.00667>
- [6] Redmon, J., Divvala, S., Girshick, R., *et al.* (2016) You Only Look Once: Unified, Real-Time Object Detection. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Las Vegas, NV, 27-30 June 2016, 779-788. <https://doi.org/10.1109/CVPR.2016.91>
- [7] Woo, S., Debnath, S., Hu, R., *et al.* (2023) Convnext v2: Co-Designing and Scaling Convnets with Masked Autoencoders. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Vancouver, 17-24 June 2023, 16133-16142. <https://doi.org/10.1109/CVPR52729.2023.01548>
- [8] Woo, S., Park, J., Lee, J.Y., *et al.* (2018) CBAM: Convolutional Block Attention Module. *Proceedings of the European Conference on Computer Vision (ECCV)*, 3-19. https://doi.org/10.1007/978-3-030-01234-2_1
- [9] Ding, X., Zhang, X., Han, J., *et al.* (2021) Diverse Branch Block: Building a Convolution as an Inception-Like Unit. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Nashville, TN, 20-25 June 2021, 10886-10895. <https://doi.org/10.1109/CVPR46437.2021.01074>