

文章编号:1007-2780(2022)07-0900-13

融合细节特征与混合注意力机制的火灾烟雾检测

汪睿卿, 王慧琴*, 王 可

(西安建筑科技大学 信息与控制工程学院, 陕西 西安 710311)

摘要:针对卷积神经网络高层特征图中细节特征被削弱造成烟雾图像底层特征丢失的问题,提出一种融合细节特征与混合注意力机制的 YOLOv4 改进算法。设计了细节特征融合模块,将主干网络中的底层细节特征引入高层特征图,得到具有丰富多尺度信息的融合特征。在通道和空间维度上采用混合注意力机制对融合特征的图权重进行重新赋值,在增强烟雾目标特征的同时抑制无关区域特征,使烟雾特征表达具有更好的鲁棒性。实验结果表明,本文算法的平均精确率、精确率和召回率相比 YOLOv4 算法分别提高了 4.31%, 1.21%, 9.86%, 同时保持了较快的检测速度。本文算法能够有效提取烟雾目标的整体特征,对于复杂背景下的火灾烟雾检测任务更为适用。

关键词:烟雾检测;深度学习;YOLOv4 算法;特征融合;混合注意力机制

中图分类号:TP391 **文献标识码:**A **doi:**10.37188/CJLCD.2021-0325

Fire smoke detection combined with detailed features and hybrid attention mechanism

WANG Rui-qing, WANG Hui-qin*, WANG Ke

(College of Information and Control Engineering, Xi'an University of Architecture and Technology,
Xi'an 710055, China)

Abstract: To solve the problem that the detailed features of the high-level feature map are weakened and the low-level features of the smoke image are lost, an improved YOLOv4 algorithm that combines the detailed features and the attention mechanism is proposed. The detail feature fusion module is designed, and the low-level features in backbone network are introduced into high-level features to obtain the fusion feature map with extensive multi-scale information. Then, a hybrid attention mechanism in two separate dimensions of channel and spatial is adopted, to reassign the weight of the fusion feature map. The smoke target features are enhanced the background features are suppressed, hence the proposed algorithm is robust in feature expression. The experimental results show that the average precision, precision and recall rate of

收稿日期:2021-12-14;修订日期:2022-01-11.

基金项目:陕西省自然科学基金计划(No. 2021JM-377);陕西省科技厅科技合作项目(No. 2020KW-012);陕西省教育厅智库项目(No. 18JT006);西安市科技局高校人才服务企业项目(No. GXYD10.1)

Supported by Natural Science Basic Research Plan in Shaanxi Province (No. 2021JM-377); Science and Technology Cooperation Projects in Department of Science and Technology of Shaanxi Province (No. 2020KW-012); Think Tank Project in Department of Education of Shaanxi Province (No. 18JT006); Talent Service Enterprise Project of Science and Technology University in Xi'an City (No. GXYD10.1)

*通信联系人, E-mail: hqwang@xauat.edu.cn

the algorithm in this paper are increased by 4.31%, 1.21% and 9.86% respectively compared with the YOLOv4 algorithm, while maintaining a high detection speed. The proposed algorithm can effectively extract the overall features of smoke targets, and is suitable for smoke detection tasks in complex backgrounds.

Key words: smoke detection; deep learning; YOLOv4; feature fusion; hybrid attention mechanism

1 引言

当前,主流的火灾探测方法仍然依靠传统的传感器来监测特定物理量,如烟雾颗粒、环境温度、相对湿度和辐射光强等。由于火灾燃烧产物的生成和传播需要一定时间,因此,此类探测器可能会产生响应延迟,且难以在大空间建筑和室外环境正常工作^[1]。火灾的早期阴燃阶段常伴有烟雾生成,实时检测火灾烟雾可对火灾的产生发出早期预警,及时发现早期火情,避免了因燃烧产物的传播造成的检测滞后,其检测范围相较传统传感器也更广,适用于大空间建筑和室外环境的火灾检测。此外,这种检测方式可节约大量人力资源并消除了人的主观错误判断对火灾预警的影响。

基于传统计算机视觉算法的视频烟雾检测主要可分为如下几个阶段:图像预处理、疑似区域提取、烟雾特征描述和烟雾识别^[1]。Kim^[2]等人利用高斯混合模型(Gaussian Mixture Model, GMM)作为背景估计算法提取烟雾区域,采用Adaboost算法检测候选区域的烟雾,在室外烟雾检测中取得了较好效果。Zhao^[3]等人利用卡尔曼滤波分割候选烟雾区域并采用局部二元模式(Local Binary Motion Pattern, LBMP)定义烟雾的纹理特征,提出了一种基于动态纹理特征的烟雾检测算法。Wang^[4]等人采用基于模糊逻辑的数据增强策略增强了图像的灰度动态范围,提取并融合了烟雾图像的静态和动态特征,将这些特征向量归一化并输入到支持向量机(Support Vector Machine, SVM)模型中进行识别。为有效提取烟雾图像的纹理特征,Ye^[5]等人将视频帧序列视作独立的多维数据,设计了一种融合Surfacelet变换和3D-隐马尔可夫树模型(Hidden Markov Tree, HTM)的烟雾动态特征描述子,这种特征描述子更接近烟雾图像动态纹理的本质特征。

上述视频烟雾检测方法基于人工设计复杂

的烟雾特征,人工设计的烟雾特征在很大程度上依赖于设计者的先验知识,具有较强的可解释性但是泛化性很差,仅适用于简单的火灾场景^[6]。卷积神经网络(Convolutional Neural Network, CNN)作为最重要的深度学习模型之一,具有良好的特征提取能力和泛化能力,随着硬件算力的提升和大规模数据集的出现,卷积神经网络工程化应用得以实现,成为计算机视觉领域的主流方法之一^[7-8]。Tao^[9]等人基于AlexNet^[10]设计了一个端到端的火灾烟雾识别网络,在测试集上实现了99.4%的检测率,优于传统检测算法。为缓解训练样本不平衡造成的过拟合,Yin^[11]等人将批量归一化(Batch Normalization, BN)引入烟雾识别网络,有效提高了识别精度。为了提高特征提取效果,Gu^[12]等人提出了一个由两个子网络组成烟雾检测网络,并以串行形式融合两个子网络的输出。随着卷积神经网络的发展,诸多性能优异的目标检测网络也被应用到火灾烟雾检测中。He^[13]等人将Faster R-CNN^[14]中的特征提取网络替换为深度残差网络(Residual Network, ResNet)^[15]并将特征金字塔网络(Feature Pyramid Network, FPN)集成到区域候选网络(Regional Proposal Networks, RPN)中,相比原始算法在烟雾和火焰的识别精度上均有提高。端到端的单阶段检测算法结构简单,具有较强的实时性。Cai^[16]等人通过将通道注意力模块和正则化模块引入YOLOv3^[17]提出了YOLO-SMOKE算法,在提高了精确率的同时增强了算法的鲁棒性。Gagliardi^[18]等人将卡尔曼滤波器和改进的轻量级SSD^[19]算法相结合,设计了一个级联的端到端烟雾检测算法,在嵌入式设备上实现了较快的检测速度。吴凡^[20]等人通过改进三维卷积神经网络,提出了一种时空域烟雾检测算法,引入了时域变化特征,提高了烟雾检测准确率。

卷积神经网络可以获取丰富的图像特征,其中底层特征图包含了图像的纹理、边缘等细节特

征,高层特征图则保留了抽象的语义信息^[21]。在火灾烟雾检测任务中,底层细节特征有利于烟雾目标准确定位,高层语义特征则有利于烟雾目标正确分类。随着网络层数的增加,特征图包含的高层语义信息愈发抽象,同时分辨率降低,特征图中关于烟雾目标的底层细节特征将会被削弱。

本文提出了一种基于 YOLOv4(You Only Look Once version4)^[22]网络的火灾烟雾检测算法,通过融合特征提取网络 CSPDarknet53 各层级特征图,得到同时具有细粒度特征和语义信息的特征表示,并结合 CBAM 混合注意力机制^[23]在融合后特征图的空间和通道维度上进行权重重新分配。实验结果表明,本文算法在保持较快检测速度的同时提高了火灾烟雾的检测精度。

2 YOLOv4 算法

作为 YOLO 系列算法的后续版本,YOLOv4

算法继承了前代算法的思想,将目标检测任务视作回归问题构建了一个端到端的网络模型,在检测速度和检测精度上均有优异表现。YOLOv4 的网络结构如图 1 所示,其主要由 CSPDarknet53 (Cross Stage Partial Connections Darknet53)、空间金字塔池化(Spatial Pyramid Pooling, SPP)^[24]、路径聚合网络(Path Aggregation Network, PANet)^[25]和 YOLO 检测头(YOLO Head)等部分组成。其中,基于 Darknet53^[17]特征提取网络并引入 CSPNet(Cross Stage Partial Network)^[26]思想提出了新的骨干网络 CSPDarknet53。CSPDarknet53 共有 5 个 CSPDarknet 模块,生成的 5 个层级特征图对应原始输入的下采样率分别为 2, 4, 8, 16, 32, 特征图对应的通道数分别为 64, 128, 256, 512, 1 014。YOLOv4 算法的颈部(Neck)由 SPP 和 PANet 结构组成,PANet 使用了更短的路径对特征图进行融合,SPP 模块通过最大池化(Max Pooling)操作增大了感受野,在高层特征图融合了全局特征和局部特征。

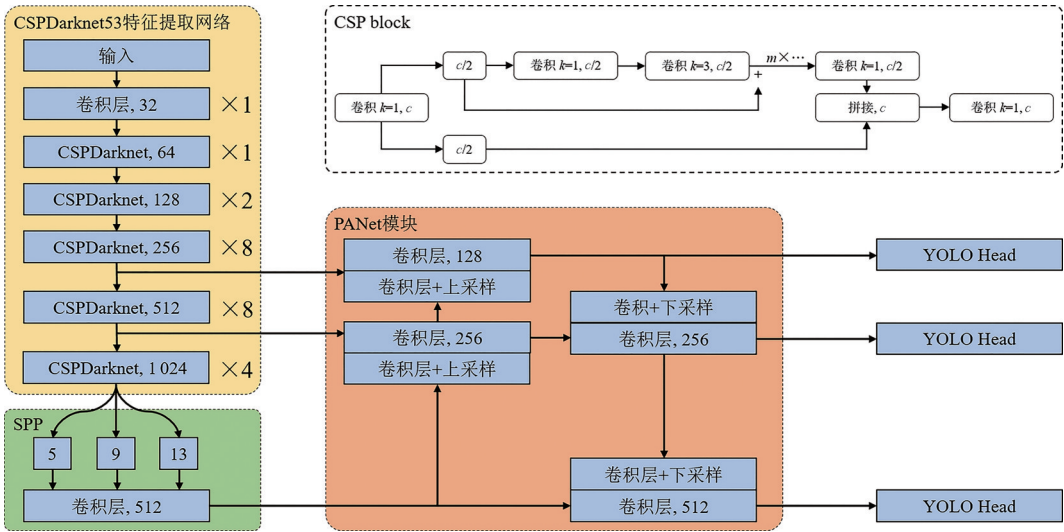


图 1 YOLOv4 网络结构图

Fig. 1 Structure of YOLOv4

3 融合细节特征与混合注意力机制的火灾烟雾检测算法

更深层的卷积可提高卷积神经网络的语义表达和特征提取能力,然而高层特征图中关于图像的细节特征将会被削弱,不利于目标的有效定

位,而且真实火灾烟雾场景中的复杂背景和遮挡将会对烟雾检测造成干扰。本文通过提取 CSPDarknet53 特征提取网络生成的 5 个层级的特征图,将浅层细节特征融合至高层特征图,获得具有丰富多尺度信息的高质量烟雾特征。随后,使用混合注意力机制对融合后特征图在通道域和

空间域进行权重重新分配,使得无效背景信息被进一步抑制,提高了烟雾目标区域的特征表达效果,增强了网络特征表达的鲁棒性。

3.1 烟雾细节特征提取与融合

CSPDarknet53 特征提取网络由 Darknet53 网络改进而来,其中包括 5 个 CSPDarknet 结构,对应输出 5 个层级不同尺寸的特征图。YOLOv4 算法中提取 3 个高层的特征图将其输入至 PANet 中进行融合,最后经过 YOLO Head 结构得到输出结果。通用的目标检测算法旨在处理多类别的目标检测任务,底层的细节特征如纹理、边缘和轮廓等不足以有效区分不同类目标,因此也较少被直接用于表征目标特征信息。火灾烟雾检测是一种单类别目标检测,图像底层的细节特征有利于区分烟雾目标和图像背景;此外,常见的烟雾场景多样复杂,烟雾的扩散也会造成特征的不明显。为增强烟雾区域特征,设计了一个细节特征融合模块(Detailed Feature Fusion Module, DFF),通过统一特征图的尺寸将底层特征图和高层特征图在通道维度进行拼接,实现了细节特征与语义特征的融合,如图 2 所示。

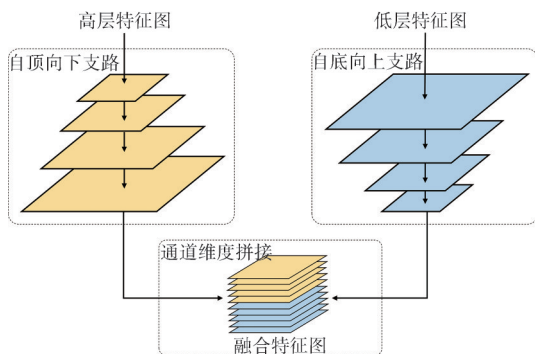


图2 细节特征融合模块

Fig. 2 Detailed feature fusion module

细节特征融合模块是由两个特征金字塔(Feature Pyramid Networks)^[27]结构组成的特征金字塔网络。两个特征金字塔分别为自顶向下(Top-down)支路和自底向上(Bottom-up)支路,同时接收底层特征图和高层特征图,对尺寸较大的底层特征图采用最大池化(Max Pooling)操作进行下采样,对尺寸较小的高层特征图以双线性插值法(Bilinear Interpolate)进行上采样,分别生成两个尺寸一致的特征图。常见的特征融合方

式主要包括通道拼接、逐元素相乘和逐元素相加 3 种形式,为避免特征金字塔获取的细节特征被高层特征覆盖造成细节特征丢失,对生成的特征图采用通道拼接的方式进行融合,得到信息更丰富的强鲁棒性多尺度特征。以底层特征图和高层特征图分别经过两次下采样和上采样为例,细节特征融合模块可表示为式(1):

$$F = C\left(M_s\left(M_s(F_l)\right), B_r\left(B_r(F_h)\right)\right), \quad (1)$$

其中: F_l 和 F_h 分别表示输入特征金字塔底层和高层特征图; $M(\cdot)$ 和 $B(\cdot)$ 分别表示最大池化运算和双线性插值,下标 s 和 r 表示下采样和上采样中的尺度因子; $C(\cdot)$ 表示沿通道维度拼接操作,融合过程如图 3 所示。

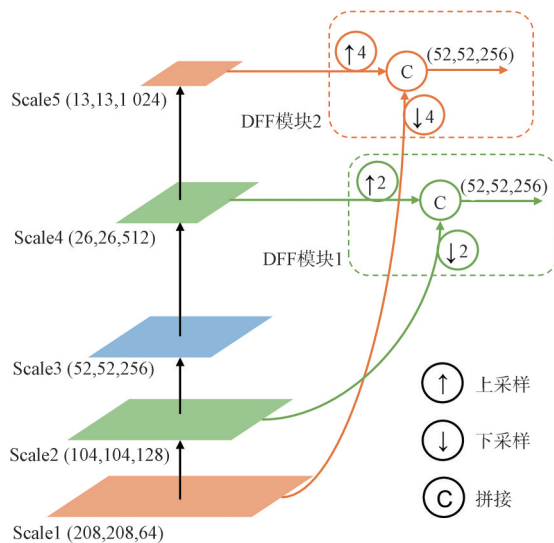


图3 特征融合过程

Fig. 3 Process of feature fusion

以实验中设置的输入图片尺寸(416, 416, 3)为例,CSPDarknet53 特征提取网络共生成了 5 个尺寸分别为(208, 208, 64), (104, 104, 128), (52, 52, 256), (26, 26, 512), (13, 13, 1024)的特征图,分别为 scale1, scale2, scale3, scale4, scale5。scale5 是 CSPDarknet53 网络经特征映射输出的最高层特征图,其通道维长度为 1024,具有丰富的语义信息;但是 scale5 特征图的分辨率较低,其长宽均为 13,缺乏足够的细节特征。因此本文选择将 scale5 特征图同最底层特征图 scale1 相融合,使用最短路径引入细节特征,以获得更丰富的具有尺度差异的特征表达并尽可能保留原始

特征,同理特征图 scale2 和 scale4 相融合。具体来看,DFF 模块 1 同时输入特征图 scale2 和特征图 scale4,首先分别对这两个特征图进行两次尺度因子为 2 的下采样和上采样,经过卷积操作将通道数调整为 128,最后在通道维度进行拼接生成了尺寸为 (52, 52, 256) 的融合特征图;DFF 模

块 2 则对 scale1 和 scale5 特征图进行了融合,得到的融合特征图尺寸同 DFF 模块 1。

具体参数设置如表 1 所示, s 和 r 分别表示池化层的步长和双线性插值的尺度因子, c 表示卷积层的输出通道数,括号内为当前的特征图尺寸。

表 1 细节特征融合模块参数设置

Tab. 1 Parameters setting of DFF modules

	DFF 模块 1	DFF 模块 2
	输入 (104, 104, 128)	输入 (208, 208, 64)
自底向上支路	最大池化层, $s=2$ (52, 52, 128)	最大池化层, $s=2$ (104, 104, 64)
	卷积层, $c=128$ (52, 52, 128)	卷积层, $c=128$ (104, 104, 128)
		最大池化层, $s=2$ (52, 52, 128)
		卷积层, $c=128$ (52, 52, 128)
	输入 (26, 26, 512)	输入 (13, 13, 1 024)
自顶向下支路	双线性插值, $r=2$ (52, 52, 512)	双线性插值, $r=2$ (26, 26, 1 024)
	卷积层, $c=256$ (52, 52, 256)	卷积层, $c=512$ (26, 26, 512)
	卷积层, $c=128$ (52, 52, 128)	双线性插值, $r=2$ (52, 52, 512)
		卷积层, $c=256$ (52, 52, 256)
		卷积层, $c=128$ (52, 52, 128)
拼接	通道维度拼接 (52, 52, 256)	通道维度拼接 (52, 52, 256)

3.2 基于混合注意力机制的烟雾特征鲁棒性增强

通过细节特征融合可将烟雾图像底层细节特征引入高层特征图,但是通过通道拼接方式得到的特征图可能会带来冗余信息及无关背景信息,因此引入注意力机制对融合特征图进行权重重新分配,以提升烟雾特征的表达效果,增强网络特征表达的鲁棒性。注意力机制通过调整特征图的权重达到抑制无关区域特征和增强目标区域特征的目的,提高了网络的特征表达效果。卷积神经网络中的注意力机制可分为通道注意力机制、空间注意力机制和混合注意力机制 3 种。通道注意力和空间注意力分别在通道域和空间域对特征图加以权重,混合注意力机制则兼顾这两点。本文在此处采用的 CBAM^[20]是一种典型的混合注意力机制,通过池化操作获得通道域和空间域权重,并以串行形式将这两部分权重与输入特征图进行逐元素相乘,如图 4 所示。

通道注意力模块对特征图的空间维度进行压缩获得各通道的注意力权重,不同于此前的注

意力机制,此处加入了最大池化以提取各通道间的差异性特征。首先通过对输入特征图进行最大池化和平均池化得到两个通道注意力描述子 $F_{\text{avg}}^c \in \mathbb{R}^{1 \times 1 \times C}$ 和 $F_{\text{max}}^c \in \mathbb{R}^{1 \times 1 \times C}$,两个注意力描述子依次经过一个共享的多层感知机(Multi Layer Perceptron, MLP)、逐元素相加和 Sigmoid 激活函数得到最终的通道注意力图 $M_c \in \mathbb{R}^{1 \times 1 \times C}$,如式(2)所示:

$$M_c = \sigma \left(\text{MLP}(\text{AvgPool}(F)) + \text{MLP}(\text{MaxPool}(F)) \right) = \sigma \left(W_1 \left(W_0(F_{\text{avg}}^c) \right) + W_1 \left(W_0(F_{\text{max}}^c) \right) \right), \quad (2)$$

其中, W_0 、 W_1 分别表示 MLP 的两层权重, σ 表示 Sigmoid 激活函数。

空间注意力模块则通过在通道维度利用两种池化方法生成两个相应的空间注意力描述子 $F_{\text{avg}}^s \in \mathbb{R}^{H \times W \times 1}$ 和 $F_{\text{max}}^s \in \mathbb{R}^{H \times W \times 1}$,并依次进行卷积和 Sigmoid 激活函数运算得到最终的空间注意力图 $M_s \in \mathbb{R}^{H \times W \times 1}$,如式(3)所示:

$$M_s = \sigma \left(f^{7 \times 7} \left(\left[\text{AvgPool}(F); \text{MaxPool}(F) \right] \right) \right) = \sigma \left(f^{7 \times 7} \left(\left[F_{\text{avg}}^s; F_{\text{max}}^s \right] \right) \right), \quad (3)$$

其中, $f^{7 \times 7}$ 表示卷积核大小为 7×7 卷积操作。整体可表示为式(4):

$$\begin{aligned} F' &= M_c \otimes F, \\ F'' &= M_s \otimes F', \end{aligned} \quad (4)$$

其中, $\otimes$ 表示逐元素相乘运算。

此外,为了缓解混合注意力机制中特征映射过程造成的特征丢失,避免出现网络退化问题,本文对输入混合注意力机制的特征图通过残差连接方式(Shortcut Connections)^[15]同混合注意力机制的输出特征图进行逐元素相加,如式(5)所示:

$$F''' = F'' \oplus F, \quad (5)$$

生成最终的残差混合注意力特征图 $F''' \in \mathbb{R}^{H \times W \times C}$ 。

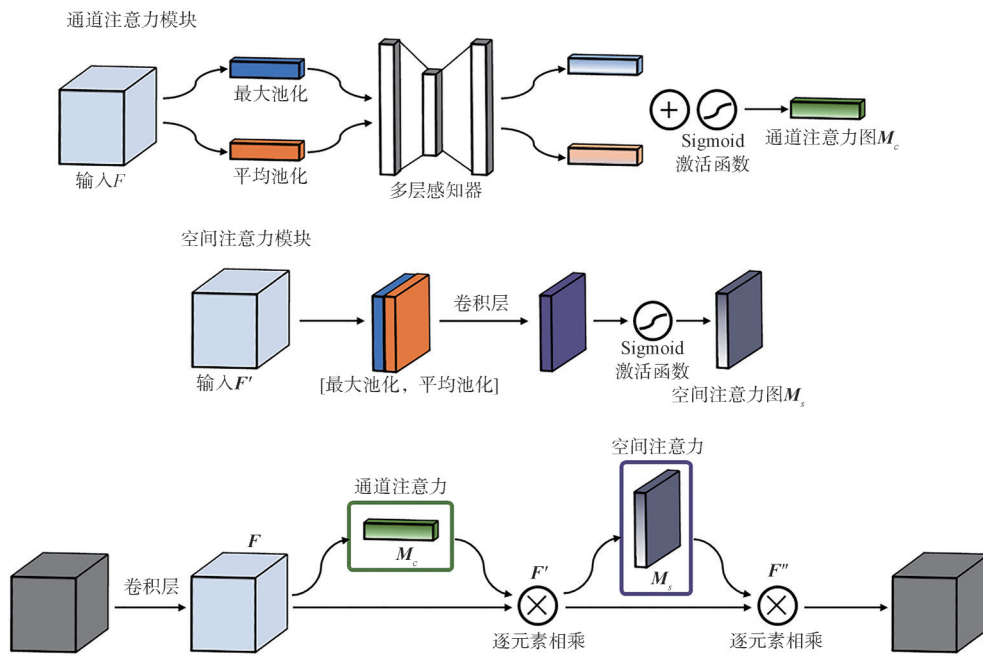


图4 CBAM结构图

Fig. 4 Diagram of CBAM

CBAM 机制的细节特征融合模块(Detailed Feature Fusion module with CBAM, DFF-CBAM)由 DFF 模块和 CBAM 模块串联组成,如图 5 所示。DFF-CBAM 模块中对 DFF 输出的特征图分别在通道域和空间域加以权重,融合了烟雾图

像底层细节特征并对特征图通道间和空间位置的特征分布进行了学习,在烟雾特征得到进一步强化的同时对背景干扰信息进行了抑制,使得图像中烟雾区域特征对网络输出结果具有更大的贡献,得到了更具鲁棒性的特征表达。

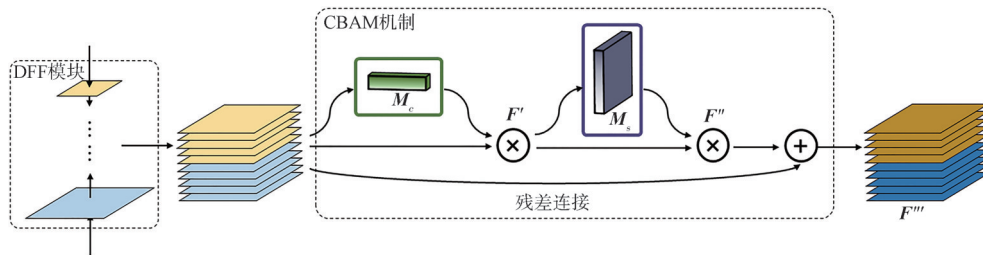


图5 引入CBAM的细节特征融合

Fig. 5 Detailed feature fusion module with CBAM(DFF-CBAM)

3.3 网络整体结构

本文算法的网络结构如图6所示,算法的主干网络为 CSPDarknet53 特征提取网络,颈部由引入 CBAM 的细节特征融合模块(DFF-CBAM)和 PANet 组成,头部由 YOLO Head 组成。输入图像经 CSPDarknet53 得到 5 个不同尺寸的特征图,底层的特征图 scale1 和 scale2 分别通过 DFF-CBAM 模块与高层的 scale5 和 scale4 融合,得到的融合特征图引入了底层细节特征同时在空间域和通道域

对权重进行了重新赋值,对图像烟雾区域特征进行增强,生成了信息更丰富且更具鲁棒性的特征表达。随后,融合特征图和 CSPDarknet53 输出的 scale3 特征图依次经过 PANet 和 YOLO Head 结构得到最终的多尺度预测结果。本文算法中提出的 DFF-CBAM 模块直接将底层特征图和高层特征进行融合,保留了更多的原始特征,得到了信息更丰富的融合结果。本文算法移除了 SPP 模块,减少了冗余操作并降低了模型参数量。

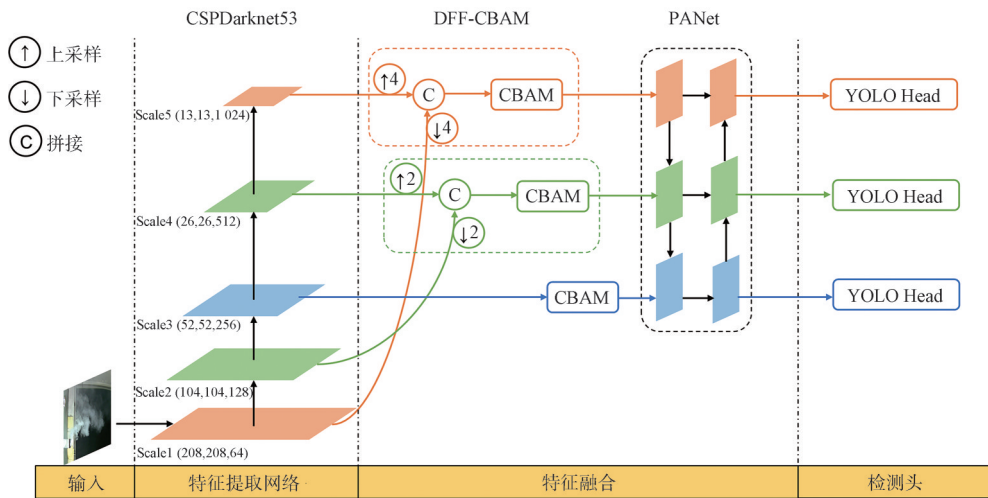


图6 网络整体结构

Fig. 6 Structure of the proposed network

4 实验设计与分析

本文实验的硬件环境为 AMD Ryzen9 5900x 处理器,32 GB 内存,NVIDIA GeForce RTX3090 显卡;软件环境为 Windows 10 操作系统,Pytorch 1.8.1 深度学习框架,CUDA 11.1,Python 3.6。

4.1 实验数据集和参数设置

本文实验数据集主要选自实验室拍摄的火灾烟雾视频、公开数据集和互联网图片,通过手工标注为 Psacal VOC 格式的标签文件。数据集样本共包括不同场景和视角的火灾烟雾图像共 9 097 张,实验中划分测试集 1 000 张,训练集 8 097 张,训练时以 0.2 的比例随机选取验证集,实验中输入图片大小设置为 416×416。

训练中采用自适应动量估计(Adaptive Moment Estimation, Adam)优化器,批量大小(Batch Size)设置为 16,初始学习率设置为 0.001,

每个训练周期(Epoch)结束后学习率调整为上一次的 0.92 倍。

4.2 评价指标

本文采用精确率(Precision)、召回率(Recall)和平均精确率均值(mAP)对模型进行评估,因本文检测目标为单类别目标,故 mAP 即为平均精确率(AP),公式如下:

$$\text{Precision} = \frac{TP}{TP + FP}, \quad (6)$$

$$\text{Recall} = \frac{TP}{TP + FN}, \quad (7)$$

$$\text{mAP} = \text{AP} = \frac{1}{R} \times \sum_{i=1}^M \left(I_i \times \frac{R_i}{i} \right), \quad (8)$$

式(6)是精确率的表达式,TP(True Positive)表示真正例,即预测正确的烟雾样本数量;FP(False Positive)表示假正例,即非烟雾样本被预测为烟雾的数量。式(7)是召回率的表达式,FN(False

Negative)表示假负例,即烟雾样本被预测为非烟雾的数量。式(8)是平均精确率的表达式, R 表示测试集中预测正确的烟雾样本总数; M 表示测试集中样本总数;当第 i 个样本为正样本时, I_i 为1,反之为0; R_i 表示前 i 个样本中预测正确的烟雾样本个数。另外,本文采用每秒帧数(Frames Per Second, FPS)来评估模型的检测速度。

4.3 检测结果

本文算法的检测结果示例如图7所示,包括了室内火灾烟雾、室内灯光干扰下烟雾、室外小目标烟雾、室外复杂背景烟雾、森林火灾烟雾和夜间火灾烟雾等常见火灾烟雾场景。结果表明,在多个实际场景中,本文算法均能以较高置信度准确检测出完整的烟雾目标。

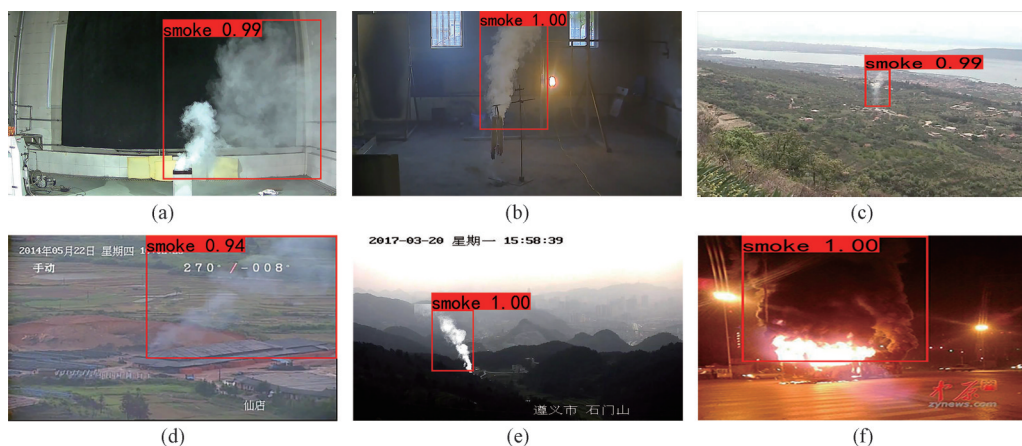


图7 检测结果示例。(a)室内烟雾;(b)室内灯光干扰烟雾;(c)室外小目标烟雾;(d)室外复杂背景烟雾;(e)森林火灾烟雾;(f)夜间火灾烟雾。

Fig. 7 Examples of detection results. (a) Indoor smoke; (b) Indoor smoke under light interference; (c) Outdoor small target smoke; (d) Outdoor smoke in complex background; (e) Forest fire smoke; (f) Night fire smoke.

实验中发现,部分室外烟雾目标易受类烟物体干扰,训练集的完备性也对网络训练有较大的影响,因此算法在少数测试样本上的检测结果并未达到预期。典型的未成功检测图像如图8所



图8 未成功检测图像

Fig. 8 Image of failed to detection

示,图中烟雾目标整体轮廓较模糊,且和昏暗的天空背景相连,难以有效区分;烟雾区域内多呈半透明的白色,缺少足够的纹理特征,易受植被和裸露的岩石等干扰。因此,在这种情况下,算法对于烟雾目标的整体把握较差,即使检测出的部分烟雾目标其置信度也较低,不能判定为成功检测。

4.4 消融实验

为充分说明本文算法中设计和改进的模块对火灾烟雾检测效果的影响,设计消融实验以验证各模块的有效性,实验结果如表2所示。

消融实验中除 YOLOv4 算法以外均删去

表2 消融实验结果

Tab. 2 Experimental results of ablation study

	精确率/%	召回率/%	平均精确率/%	检测速率/(frame·s ⁻¹)
YOLOv4	96.16	81.59	91.91	51.9
YOLOv4+DFF	94.52	88.79	92.87	51.2
YOLOv4+CBAM	97.32	85.31	93.47	52.7
本文算法	97.37	91.45	96.22	50.8

SPP 模块。加入 DFF 模块后,在引入烟雾图像细节特征的同时也带来部分背景无关信息,使得算法精确率相比于 YOLOv4 算法下降了 1.62%,但召回率提升了 7.2%。加入 CBAM 模块后,算法更聚焦于烟雾目标区域,使得精确率相较于 YOLOv4 算法提升了 1.16%,同时召回率也得到了提升。值得注意的是,实验中在删去 SPP 模块并加入 CBAM 模块之后,检测速度

比原始 YOLOv4 算法更快,这也表明了 CBAM 机制在引入较少参数的情况下有效提升了模型性能。

4.5 对比实验

为全面评估本文算法的有效性,设计实验与当前主流的深度学习目标检测算法和基于深度学习的火灾烟雾检测算法进行对比,实验结果如表 3 所示。

表 3 与其他检测算法性能对比
Tab. 3 Performance comparison with other algorithms

	精确率/%	召回率/%	平均精确率/%	检测速率/(frame·s ⁻¹)
Faster-RCNN	73.82	81.72	82.55	20.3
SSD	65.31	71.46	69.89	72.8
RetinaNet	91.49	79.56	89.93	52.7
YOLOv3	89.89	79.34	86.77	67.3
EfficientNet ^[28]	82.91	83.71	87.66	49.5
文献[29]	74.82	82.21	80.31	70.5
YOLOv4	96.16	81.59	91.91	51.9
SE-YOLOv4 ^[30]	93.32	86.47	93.64	48.5
本文算法	97.37	91.45	96.22	50.8

实验结果中,本文算法的精确率、召回率和平均精确率均优于对比算法,检测速度相较于改进前的 YOLOv4 算法略有下降,但优于两阶段目标检测算法。文献[29]和文献[30]分别基于单阶段算法 SSD 和 YOLOv4 针对烟雾检测任务进行改进,检测性能相较原始算法均有提升。其中文献[29]中采用了递进池化和反卷积操作指导特征提

取和融合。文献[30]则采用嵌入通道注意力机制的方式增强了 YOLO 检测头的信息提取能力。

图 9 为 YOLOv4 算法和本文算法的 PR 曲线图,图中阴影面积即为平均精确率。通过细节特征融合和引入注意力机制,本文算法在相同实验条件下相比于 YOLOv4 算法的平均精确率提高了 4.31%。

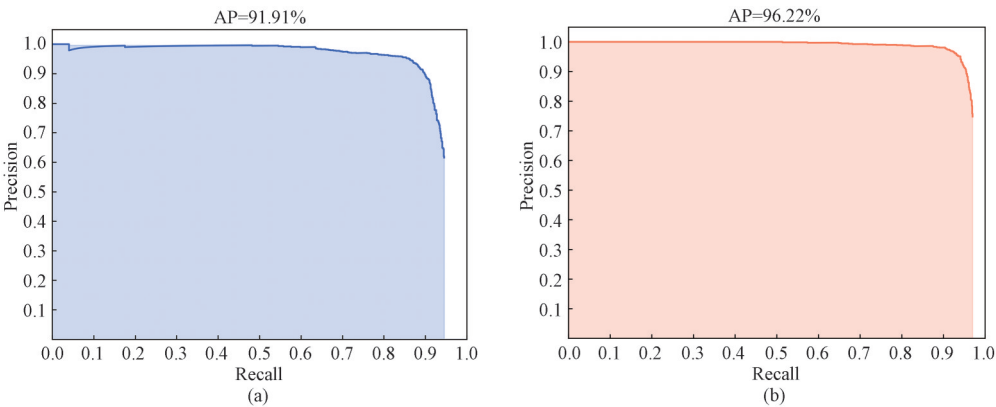


图 9 PR 曲线。(a) YOLOv4 算法 PR 曲线;(b) 本文算法 PR 曲线。
Fig. 9 PR curves. (a) PR curve of YOLOv4; (b) PR curve of proposed algorithm.

图 10 和图 11 分别为精确率曲线和召回率曲线图,计算精确率和召回率时设置置信度阈值为

0.5。本文算法的精确率和召回率相比于 YOLOv4 算法分别提高 1.21% 和 9.86%。图 11 的召回率

曲线显示,随着置信度阈值的提高,本文算法的召回率相比于YOLOv4有了更大的提升,这表明本文算法对于难分烟雾样本有更好的检测效果。综合上述对比结果可知,本文算法在烟雾检测方面具有更好的性能。细节特征融合增强了烟雾图像的特征表达效果,提升了算法对烟雾目标的定位能力;同时注意力机制可抑制无关的图像背景特征并对烟雾特征赋予更大的权重,减少了检测结果中假负例的数量,从而使得召回率相比于改进前有明显提高。

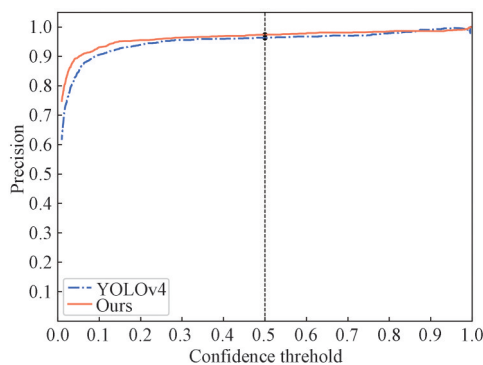


图 10 精确率曲线

Fig. 10 Precision curves

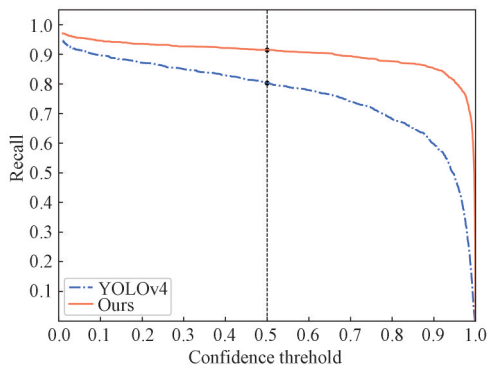


图 11 召回率曲线

Fig. 11 Recall curves

4.6 可视化分析

为了进一步分析本文算法在烟雾图像特征表达上相较于YOLOv4算法的优势,设计了特征图可视化对比实验并利用CAM^[31]方法绘制热力图。

图 12 是网络中 5 个层级特征图的可视化对比实验结果,分别为 scale1、scale2、YOLO Head1、YOLO Head2 和 YOLO Head3,对应尺寸大小为 (208,208), (104,104), (52,52), (26,26), (13,13)。

特征图随着网络层数的加深尺寸逐渐变小,并在运算过程中逐渐将细节特征映射到高层语义特征。图 12 中本文算法的底层特征图相较于YOLOv4算法提取了更多明显的细节信息,如纹理和边缘轮廓等。随着感受野的增大,细节信息逐渐稀疏化,特征图分辨率降低,纹理和轮廓等愈发难以辨认,但是语义信息得到进一步增强,在目标区域中心形成聚焦并逐步映射至目标整体区域。在高层特征图中,得益于混合注意力机制对特征图空间维度和通道维度权重的重新分配,本文算法在烟雾目标中心区域产生了更高的激活值,且激活值的分布范围更符合实际图像中烟雾的分布。

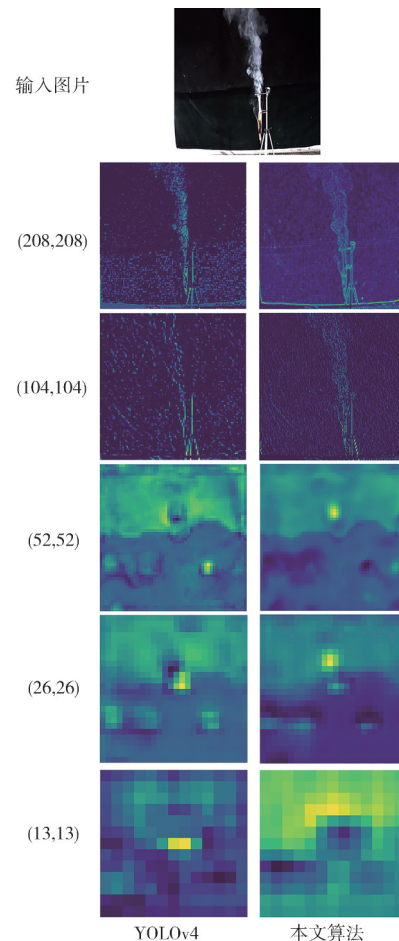


图 12 特征图可视化

Fig. 12 Visualization of feature map

图 13 是两种算法的热力图对比实验结果,通过热力图可以看出,由于烟雾目标的视觉特征较差且烟雾背景复杂,两种算法在检测过程中均不

同程度地受到了影响,YOLOv4算法在烟雾目标区域的激活值较低且分布较为杂乱,在背景区域的激活范围和强度明显高于本文算法,难以有效反映出真实烟雾目标位置;本文算法则相对更准确地地在烟雾目标区域形成聚焦产生较高的激活

值,并对背景区域的激活进行了抑制。由此可见,本文提出的基于注意力机制的细节特征融合方法有效增强了烟雾目标的整体特征,能够改善烟雾检测中烟雾目标特征不明显和背景复杂、干扰信息较多的问题。

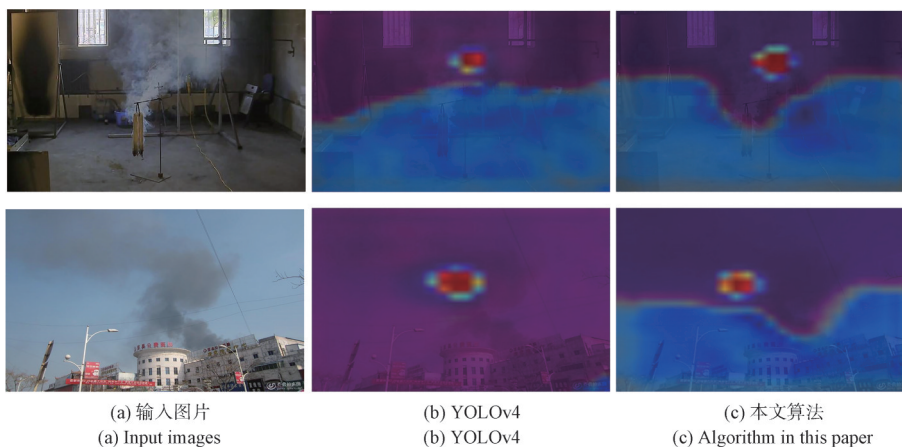


图 13 热力图对比

Fig. 13 Comparison of heatmaps

5 结 论

本文针对火灾烟雾检测中高层特征图的细节特征被削弱且背景干扰较多的问题,基于YOLOv4提出了一种融合细节特征与注意力机制的火灾烟雾检测算法。通过引入细节特征,提高了网络的特征提取和表达能力;利用CBAM

机制对融合特征图进行权重重新赋值,得到了更具鲁棒性的特征表达。与YOLOv4算法相比,在保持较高检测速率的同时,本文算法的平均精确率、精确率和召回率分别提高了4.31%,1.21%,9.86%。下一步工作将优化特征融合过程,减少冗余操作并降低模型参数量,设计适用于嵌入式设备的轻量化火灾烟雾检测网络。

参 考 文 献:

- [1] 史劲亭,袁非牛,夏雪. 视频烟雾检测研究进展[J]. 中国图象图形学报,2018,23(3):303-322.
SHI J T, YUAN F N, XIA X. Video smoke detection: a literature survey [J]. *Journal of Image and Graphics*, 2018, 23(3): 303-322. (in Chinese)
- [2] KIM H, RYU D, PARK J. Smoke detection using GMM and adaboost [J]. *International Journal of Computer and Communication Engineering*, 2014, 3(2): 123-126.
- [3] ZHAO Y Q, ZHOU Z, XU M M. Forest fire smoke video detection using spatiotemporal and dynamic texture features [J]. *Journal of Electrical and Computer Engineering*, 2015, 2015: 40.
- [4] WANG Y B. Smoke recognition based on machine vision [C]//*Proceedings of 2016 International Symposium on Computer, Consumer and Control*. Xi'an, China: IEEE, 2016: 668-671.
- [5] YE W, ZHAO J H, WANG S, et al. Dynamic texture based smoke detection using Surfacelet transform and HMT model [J]. *Fire Safety Journal*, 2015, 73: 91-101.
- [6] 李欣健,张大胜,孙利雷,等. 复杂场景下基于CNN的轻量火焰检测方法[J]. 模式识别与人工智能,2021,34(5): 415-422.
LI X J, ZHANG D S, SUN L L, et al. CNN-based lightweight flame detection method in complex scenes [J]. *Pattern*

- Recognition and Artificial Intelligence*, 2021, 34(5): 415-422. (in Chinese)
- [7] 林景栋,吴欣怡,柴毅,等. 卷积神经网络结构优化综述[J]. 自动化学报,2020,46(1):24-37.
LIN J D, WU X Y, CHAI Y, *et al.* Structure optimization of convolutional neural networks: a survey [J]. *Acta Automatica Sinica*, 2020, 46(1): 24-37. (in Chinese)
- [8] 夏雪,袁非牛,章琳,等. 从传统到深度:视觉烟雾识别、检测与分割[J]. 中国图象图形学报,2019,24(10):1627-1647.
XIA X, YUAN F N, ZHANG L, *et al.* From traditional methods to deep ones: review of visual smoke recognition, detection, and segmentation [J]. *Journal of Image and Graphics*, 2019, 24(10): 1627-1647. (in Chinese)
- [9] TAO C Y, ZHANG J, WANG P. Smoke detection based on deep convolutional neural networks [C]//*Proceedings of 2016 International Conference on Industrial Informatics—Computing Technology, Intelligent Technology, Industrial Information Integration*. Wuhan, China: IEEE, 2016: 150-153.
- [10] KRIZHEVSKY A, SUTSKEVER I, HINTON G E. ImageNet classification with deep convolutional neural networks [C]//*Proceedings of the 25th International Conference on Neural Information Processing Systems*. Lake Tahoe: Curran Associates Inc., 2012: 1097-1105.
- [11] YIN Z J, WAN B Y, YUAN F N, *et al.* A deep normalization and convolutional neural network for image smoke detection [J]. *IEEE Access*, 2017, 5: 18429-18438.
- [12] GU K, XIA Z F, QIAO J F, *et al.* Deep dual-channel neural network for image-based smoke detection [J]. *IEEE Transactions on Multimedia*, 2020, 22(2): 311-323.
- [13] HE Z X, XIE L P, HUA B C, *et al.* Joint detection of smoke and flame in photovoltaic system based on deep learning [C]//*Proceedings of 2020 Chinese Automation Congress (CAC)*. Shanghai, China: IEEE, 2020: 6067-6071.
- [14] REN S Q, HE K M, GIRSHICK R, *et al.* Faster R-CNN: towards real-time object detection with region proposal networks [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2017, 39(6): 1137-1149.
- [15] HE K M, ZHANG X Y, REN S Q, *et al.* Deep residual learning for image recognition [C]//*Proceedings of 2016 IEEE Conference on Computer Vision and Pattern Recognition*. Las Vegas: IEEE, 2016: 770-778.
- [16] CAI W B, WANG C Y, HUANG H G, *et al.* A real-time smoke detection model based on YOLO-SMOKE algorithm [C]//*Proceedings of 2020 Cross Strait Radio Science & Wireless Technology Conference*. Fuzhou, China: IEEE, 2020: 1-3.
- [17] REDMON J, FARHADI A. YoLOv3: an incremental improvement [J]. *arXiv*, 2018:1804.02767.
- [18] GAGLIARDI A, DE GIOIA F, SAPONARA S. A real-time video smoke detection algorithm based on Kalman filter and CNN [J]. *Journal of Real-Time Image Processing*, 2021, 18(6): 2085-2095.
- [19] LIU W, ANGUELOV D, ERHAN D, *et al.* SSD: single shot MultiBox detector [C]//*Proceedings of 14th European Conference on Computer Vision*. Amsterdam: Springer, 2016: 21-37.
- [20] 吴凡,王慧琴,王可. 时空域深度学习火灾烟雾检测[J]. 液晶与显示,2021,36(8):1186-1195.
WU F, WANG H Q, WANG K. Spatio-temporal deep learning fire smoke detection [J]. *Chinese Journal of Liquid Crystals and Displays*, 2021, 36(8): 1186-1195. (in Chinese)
- [21] 李红光,于若男,丁文锐. 基于深度学习的小目标检测研究进展[J]. 航空学报,2021,42(7):024691.
LI H G, YU R N, DING W R. Research development of small object tracking based on deep learning [J]. *Acta Aeronautica et Astronautica Sinica*, 2021, 42(7): 024691. (in Chinese)
- [22] BOCHKOVSKIY A, WANG C Y, LIAO H Y M. YoLOv4: optimal speed and accuracy of object detection [J]. *arXiv*:2004.10934, 2020.
- [23] WOO S, PARK J, LEE J Y, *et al.* CBAM: convolutional block attention module [C]//*Proceedings of the 15th European Conference on Computer Vision*. Munich, Germany: Springer, 2018: 3-19.
- [24] HE K M, ZHANG X Y, REN S Q, *et al.* Spatial pyramid pooling in deep convolutional networks for visual recognition [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2015, 37(9): 1904-1916.
- [25] LIU S, QI L, QIN H F, *et al.* Path aggregation network for instance segmentation [C]//*Proceedings of 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Salt Lake City, UT, USA: IEEE, 2018: 8759-8768.

- [26] WANG C Y, LIAO H Y M, WU Y H, *et al.* CSPNet: a new backbone that can enhance learning capability of CNN [C]//*Proceedings of 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*. Seattle, WA, USA: IEEE, 2020: 1571-1580.
- [27] LIN T Y, DOLLÁR P, GIRSHICK R, *et al.* Feature pyramid networks for object detection [C]//*Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition*. Honolulu, USA: IEEE, 2017: 936-944.
- [28] TAN M X, LE Q V. EfficientNet: rethinking model scaling for convolutional neural networks [C]//*Proceedings of the 36th International Conference on Machine Learning*. Long Beach, CA, USA: PMLR, 2019: 6105-6114.
- [29] 刘丽娟, 陈松楠. 一种基于改进 SSD 的烟雾实时检测模型[J]. 信阳师范学院学报(自然科学版), 2020, 33(2): 305-311.
- LIU L J, CHEN S N. Real-time smoke detection model based on improved SSD [J]. *Journal of Xinyang Normal University (Natural Science Edition)*, 2020, 33(2): 305-311. (in Chinese)
- [30] 谢书翰, 张文柱, 程鹏, 等. 嵌入通道注意力的 YOLOv4 火灾烟雾检测模型[J]. 液晶与显示, 2021, 36(10): 1445-1453.
- XIE S H, ZHANG W Z, CHENG P, *et al.* Firesmoke detection model based on YOLOv4 with channel attention [J]. *Chinese Journal of Liquid Crystals and Displays*, 2021, 36(10): 1445-1453. (in Chinese)
- [31] ZHOU B L, KHOSLA A, LAPEDRIZA A, *et al.* Learning deep features for discriminative localization [C]//*Proceedings of 2016 IEEE Conference on Computer Vision and Pattern Recognition*. Las Vegas, NV, USA: IEEE, 2016: 2921-2929.

作者简介:



汪睿卿(1996—),男,安徽安庆人,硕士研究生,2019 于皖西学院获得学士学位,主要从事深度学习与计算机视觉等方面的研究。E-mail: wrq1216@163.com



王慧琴(1970—),女,山西长治人,博士,教授,2002 年于西安交通大学获得博士学位,主要从事智能信息处理、信息理论与应用、信息技术与管理、数字建筑等方面的研究。E-mail: hqwang@xauat.edu.cn