

陈浩霖^{1,2} 高尚兵^{1,2} 相林¹ 蔡创新¹ 汪长春¹

FIRE-DET:一种高效的火焰检测模型

摘要

模型的效率在计算机视觉中变得越来越重要.本文通过研究用于火焰检测的神经网络结构,提出了几个关键的优化方案,以提高模型效率和检测效果.第一,提出一种由多卷积组合结构构建的主干网络(FIRE-Net),它能高效地从多个尺度上提取丰富的火焰特征;第二,提出一种改进的加权双向特征金字塔网络(BiFPN-mini)以快速实现多尺度特征融合;第三,提出一种新的注意力机制(FIRE-Attention),让检测器对火焰特征更敏感.基于上述优化,本文开发出了一种全新的火焰检测器 FIRE-DET,它在硬件资源有限的条件下能够取得比现有基于深度学习的火焰检测方法更高的检测效率.FIRE-DET模型在自建数据集上进行训练后,最终对火焰检测的准确率和帧率分别达到97%和85 FPS.实验结果表明,与主流算法相比,本文火焰检测模型检测性能更优.本文为解决火焰探测问题提供了一个更通用的解决方案.

关键词

特征提取;特征融合;注意力机制;火焰检测

中图分类号 TP391.41

文献标志码 A

收稿日期 2021-11-12

资助项目 国家重点研发计划(2018YFB1004904);江苏省高校自然科学研究重大项目(18KJA520001);2021年淮阴工学院研究生科技创新计划项目(HGYK202122)

作者简介

陈浩霖,男,硕士生,研究方向为深度学习及计算机视觉.1007256512@qq.com

高尚兵(通信作者),男,博士,教授,研究方向为机器学习、计算机视觉、模式识别和数据挖掘.luxiaofen_2002@126.com

1 淮阴工学院 计算机与软件工程学院,淮安,223001

2 淮阴工学院 江苏省物联网移动互联网技术工程实验室,淮安,223001

0 引言

火灾的发生会威胁人类的生命安全,如:2019年9月6日至2020年1月14日,澳大利亚森林火灾燃烧面积达1 860万 hm^2 ,伤亡人数34人,死于火灾的动物超过10亿只^[1];2020年3月30日,四川省凉山州西昌市经久乡发生森林火灾,造成19名消防战士牺牲.为了预防火灾,我国在“十三五”全国森林防火规划建设期间,全国规划投入森林防火资金450.95亿元,其中中央建设投资和财政经费分别投入189.89亿元和185.44亿元.

火焰是火灾中最明显的特征之一,因此基于图像处理提取火焰特征是常用的检测火灾的方法.目前该类方法主要分为基于传统图像处理的火焰检测、基于传统机器学习的火焰检测和基于深度学习的火焰检测.近年火焰检测的方法具体有:

1) 基于传统图像处理的火焰检测方法,使用手工提取的火焰特征识别火焰.如:张慧珍等^[2]提出一种基于超像素分割并结合闪频特征进行判断识别火焰的方法;Gong等^[3]通过计算图像每帧火焰区域的质心,然后通过提取火焰的空间变异性、形状变异性和面积变异性等特征检测火焰.

2) 基于传统机器学习的火焰检测方法,更注重特征的提取,特征可解释性较强.如:Ko等^[4]提出基于模糊有限自动机的火焰检测算法;Gunay等^[5]提出一种基于熵函数的在线自适应决策融合框架;Dimitropoulos等^[6]利用火焰的先验知识,提取火焰候选区域,然后使用两类SVM分类器对候选区域进行分类.

3) 基于深度学习的火焰检测方法,通过自动学习特征(卷积特征)检测火焰,无需设计指定特征,学习获得的特征只有计算机理解.如:Kim等^[7]将连续帧中边界框内的汇总特征由Long Short-Term Memory(LSTM)在短期时间序列中判断是否存在火并进行分类,然后将大多数连续的短期决策合并在一起,长期投票决定最终结果;Aslan等^[8]提出一种基于视觉的深度卷积生成对抗性神经网络火焰检测方法;Zhang等^[9]提出一种有效的非对称编解码U形结构ATT-Squeeze-Net来鉴别森林火灾;Dunnings等^[10]提出一种基于超像素分割结合深度卷积神经网络的火焰检测方法.

基于传统图像处理的火灾检测所依靠的特征和判定阈值规则是人为设定的,往往泛化能力略显不足.基于传统机器学习的火焰检测

同样是利用人工提取特征,只是在最后的检测时利用机器学习分类算法进行火焰分类,依然无法避免人工特征工程的缺点.基于深度学习的火焰检测,可以避免人工特征工程的繁杂,利用计算机自动提取特征,能够实现端到端的实时输出,并且能够保证检测精度和泛化能力,这就拥有了很好的智能性.但复杂的模型会降低模型的检测效率,所以本文利用深度学习方法中单步检测的框架实现火焰检测.

现有的目标检测器主要根据它们是否具有区域建议步骤,分为单步检测(One-Stage Detectors)^[11-15]和双步检测(Two-Stage Detectors)^[15-20].虽然双步检测器往往更灵活和更准确,但单步检测器通常被认为更简单和更高效.在本文中,主要以单步检测器的设计为主体,优化网络结构使其获得更高的效率和精度.深度卷积神经网络自提出起就被广泛应用于图像领域的分类、检测和分割等任务中.近些年,在深度卷积神经网络在图像领域任务中出现了一系列优化的方法:

1) 主干网络:现有的单步检测的主干网络主要使用 CenterNet^[20]、DarkNet52^[11]和 EfficientNet^[21]等,可以分为两种类型:一种是轻量型的,其目标在于开发出一种微型的主干网络并且能拥有较好的特征提取效果,如 EfficientNet;另一种是重量型的,在特征提取上具有较好的效果,如 DarkNet52.虽然重量型的网络在多目标检测过程中表现出优异的效果,但是在面对火焰目标时会出现鲁棒性低的情况,这是由于火焰非结构化、形态特异性高的特性决定的.因此本文提出一种轻量型主干网络使其获得更高的效率和精度.

2) 特征融合网络:融合多尺度特征是目标检测的主要难点之一.早期的检测器通常直接将骨干网络中提取的金字塔特征进行预测.特征金字塔网络(Feature Pyramid Network, FPN)^[22]作为其开创性工作之一,提出了一种自顶向下的多尺度特征组合路径;PANet^[23]在 FPN 的基础上增加了一个自下而上的路径聚合网络;NAS-FPN^[24]利用神经结构搜索来自动设计特征网络拓扑.最近, BiFPN^[25]利用简单的双向特征融合网络并进行加权融合. BiFPN 虽然性能更好,但面对火焰检测速度仍有不足.因此本文改进了 BiFPN 模块,提升输出特征对火焰语义的抽象程度与鲁棒性.

3) 注意力机制:注意力机制是近年提出的一种新观念.注意力机制主要分为两种:一种是通道注意

力机制,另外一种为空间注意力机制. SENet^[26]最早提出通道注意力机制,为每个通道加上一个权重,目的是为了体现每个通道的重要程度; CBAM^[27]则在通道注意力机制的基础上增加了空间注意力机制,体现特征图中每个位置的重要程度.传统的注意力机制是无监督的,它通过网络的反馈计算获得自适应的注意力矩阵,所以它会生成一些背离预期的结果,这是我们所不愿看见的.因此本文提出了一种监督式的空间注意力机制,通过检测器的标签值生成一组二值图像标签来监督获得的注意力矩阵.

基于上述分析,本文提出了一种基于 FIRE-DET 的快速火焰检测方法.主要的贡献如下:

1) 提出了一种 Multi-Conv 模块,并通过 Multi-Conv 模块建立 FIRE-Net 网络,并将其作为主干网络提取多个尺度的火焰特征,增强特征提取效果.

2) 受 BiFPN 网络的启发,本文将主干网络提取的特征进行双向连接融合,提取出含有多个尺度的特征,而不是单尺度特征.

3) 提出一种监督式的空间注意力机制 Fire-Attention,通过检测器的标签值生成一组二值图像标签来监督获得的注意力矩阵.

1 本文方法

1.1 FIRE-DET 模型

FIRE-DET 检测器是由主干网络、特征融合网络、注意力机制和检测头构成的一种端到端的深度卷积神经网络,并遵循了单步检测器的结构. FIRE-DET 检测器的整体架构如图 1 所示. FIRE-DET 的主干网络是 FIRE-Net,是由 Multi-Conv 模块建立的轻量级网络,可获取输入图像多个尺度的特征 $\{P_1, P_2, P_3\}$; FIRE-DET 的特征融合网络是 BiFPN-mini, BiFPN-mini 将主干网络提取的多个尺度的特征 $\{P_1, P_2, P_3\}$ 进行融合; FIRE-DET 中提出了一种注意力机制 FIRE-Attention,对融合后的特征进行反馈调节,增强检测器对火焰特征的敏感性; FIRE-DET 中使用现有的检测有 YOLO head 预测对象和对象位置.

1.2 FIRE-Net 网络

1.2.1 深度卷积模块

一个卷积层 i 可以定义为一个函数: $y = f_i(\mathbf{x}_{(h,w,c)})$, 其中 f 是算子, y 是输出张量, \mathbf{x} 是输入张量,具有张量形状 h, w, c , 其中 h 和 w 是空间维度, c 是通道维度. 两个算子连续作用于一个输入张量可以定义为

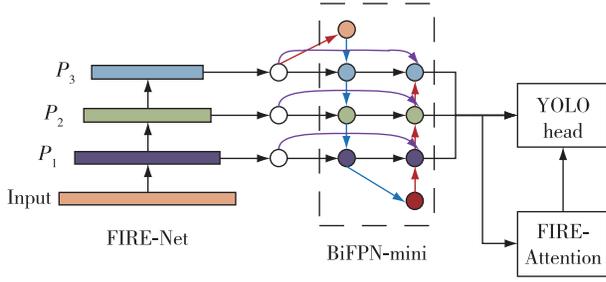


图1 FIRE-DET 体系结构

Fig. 1 The proposed FIRE-DET architecture

$$y = f_2(f_1(x_{(h,w,c)})) = f_2 \odot f_1(x_{(h,w,c)}),$$

一个卷积模块的结构可以被表示为

$$N = f_k \odot \dots \odot f_2 \odot f_1(x) = \bigodot_{j=1 \dots k} f_j(x_{(h,w,c)}). \quad (1)$$

在实践中,卷积模块层通常被划分为多个阶段,每个阶段中的所有层共享相同的体系结构.因此,本文可以将卷积模块定义为

$$N = \bigodot_{j=1, \dots, k} f_j(x_{(h,w,c)}). \quad (2)$$

图2为具有代表性的卷积模块.图2a为传统的卷积块,由卷积归一化与激活函数构成.传统卷积模块可以表示为

$$g_1(x) = \text{relu}(\text{BN}(\text{Conv}(x_{(h,w,c)}))), \quad (3)$$

其中,Conv是特征处理的卷积操作,BN为归一化操作,relu为激活函数, g_1 是传统卷积模块的算子.图2b显示了经典的残差卷积块,由两种传统卷积块组合构成.残差卷积块可以表示为

$$g_2(x) = g_1 \odot g_1(x_{(h,w,c)}) + x_{(h,w,c)}. \quad (4)$$

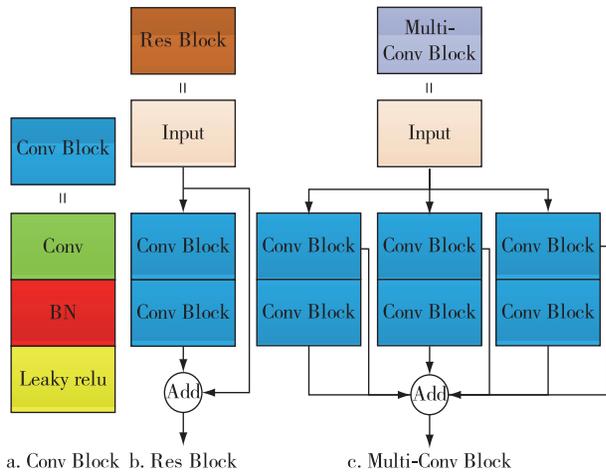


图2 代表性卷积模块

Fig. 2 Representative convolution blocks

1.2.2 Muti-Conv 模块

传统的 Conv Block 在感受野上受到限制,未能

融合不同感受野的信息,所以 Res Block 增加一个 Conv Block 的路径来扩展网络的感受野,如图 2b 所示.本文发现 Res Block 比 Conv Block 获得了更好的准确性,但仍然不能适应检测火焰的计算.如果卷积的路径只有一条,那么它获得的感受野比较有限,所以本文首先增加了卷积的路径数,然后将每次卷积的结果融合后输出,以避免特征损失,最后减少卷积的通道数,以提高模型的效率.通过上述优化,本文将新的特征网络命名为 Multi-Conv Block,如图 2c 所示.Multi-Conv 卷积块可以表示为

$$g_3(x) = g_1 \odot g_1(x_{(h,w,c)}) + g_1 \odot g_1(x_{(h,w,c)}) + g_1 \odot g_1(x_{(h,w,c)}) + g_1(x_{(h,w,c)}) + g_1(x_{(h,w,c)}) + g_1(x_{(h,w,c)}). \quad (5)$$

1.2.3 构建 FIRE-Net 网络

由于本文的模型主要用于火焰检测,因此提出针对火焰检测的特征提取网络.为了增加提取火焰的有效性,本文还提出一个新的主干骨网络,称为 FIRE-Net.该网络通过增加卷积的路径与卷积核的大小,提高网络对火焰特征的提取.表1中每一行描述了一个线性层,每层中介绍了算子 f_i 、输入分辨率 $h_i \times w_i$ 和输出通道 C_i .

表1 FIRE-Net 网络层结构

Table 1 FIRE-Net network layer structure

| i | f_i | $h_i \times w_i$ | C_i |
|-----|-----------------------------|------------------|-------|
| 1 | 4×4MaxPooling, stride 4 | 512×512 | 3 |
| 2 | Multi-Conv(1×1, 9×9, 27×27) | 128×128 | 16 |
| 3 | 4×4MaxPooling, stride 4 | 128×128 | 16 |
| 4 | Multi-Conv(1×1, 3×3, 9×9) | 32×32 | 32 |
| 5 | 4×4MaxPooling, stride 4 | 32×32 | 32 |
| 6 | Multi-Conv(1×1, 3×3) | 8×8 | 64 |

虽然 FIRE-Net 的 Muti-Conv 模块结构复杂,但是 FIRE-Net 网络通道数减少至 Darknet53 层结构的 1/4,并且下采样次数减少一半,使网络参数变少.最终模型变得更小,但由于 Muti-Conv 模块多卷积路径的特点,使模型对火焰特征提取的能力仍有提升.网络以 512×512 大小的图像作为输入进行 3 次最大池化和 Muti-Conv 操作,在 FIRE-Net 中第 1 次使用 Muti-Conv 采用 1、9、27 三种卷积核提取特征,第 2 次使用 Muti-Conv 采用 1、3、9 三种卷积核提取特征,第 3 次使用 Muti-Conv 采用 1、3 两种卷积核提取特征.这样设置卷积核基于以下原因:1)使用卷积核 1 的主要目的是形成一种无监督的注意力机制,让网络自动调整网络;2)本文模型的输入图像像素大小

为 512×512 ,若使用 2 的等比数列感受野过小,若使用 4 的等比数列会导致感受野过大,所以本文卷积核为 3 的等比数列中的值.FIRE-Net 的性能与最先进的分类器相比,其浮点运算更少,表明网络结构能更好地利用处理器,使其更有效地检测,从而更快获取火焰特征.

1.3 BiFPN-mini 网络

多尺度特征融合旨在聚集不同分辨率的特征.形式上,给定一个多尺度特征序列 $\mathbf{P}^{\text{in}} = (\mathbf{P}_{l_1}^{\text{in}}, \mathbf{P}_{l_2}^{\text{in}}, \dots)$,其中 \mathbf{P}_{l_i} 表示 $(l_i + 1)$ 倍下采样后的特征图,本文的目标是找到一个可以有效聚合不同特征的变换 h ,并输出一个新特征列表: $\mathbf{P}^{\text{out}} = h(\mathbf{P}^{\text{in}})$.

图 3a 为传统的自上而下的 FPN,需要 5 个尺度的特征 $\mathbf{P}^{\text{in}} = (\mathbf{P}_{l_1}^{\text{in}}, \mathbf{P}_{l_2}^{\text{in}}, \dots, \mathbf{P}_{l_5}^{\text{in}})$,其中 $\mathbf{P}_{l_1}^{\text{in}}$ 表示分辨率为输入图像 $1/2$ 的特征级别.例如,如果输入分辨率为 512×512 ,则 $\mathbf{P}_{l_1}^{\text{in}}$ 代表的 2 倍下采样后分辨率为 256×256 的特征图 ($512 \div 2^1 = 256$),而 $\mathbf{P}_{l_5}^{\text{in}}$ 代表 16 倍下采样后分辨率为 32×32 .FPN 自上而下的多尺度特征融合方法可以表示为

$$\begin{aligned} p_5^{\text{out}} &= \text{Conv}(p_5^{\text{in}}), \\ p_4^{\text{out}} &= \text{Conv}(p_4^{\text{in}} + \text{Resize}(p_5^{\text{in}})), \\ &\dots \\ p_1^{\text{out}} &= \text{Conv}(p_1^{\text{in}} + \text{Resize}(p_2^{\text{in}})). \end{aligned} \quad (6)$$

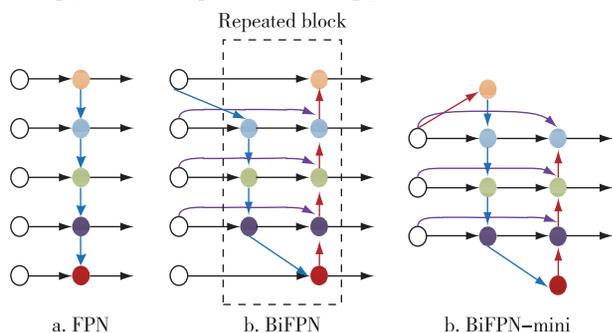


图 3 特征融合网络

Fig. 3 Feature fusion network

如图 3a 中传统的自上而下的 FPN 本质上受到单向信息流的限制,所以 BiFPN 提出了删除那些只有一条输入边的节点形成简化的双向网络,且从原始输入到输出节点添加额外的边融合更多的特征,如图 3b 所示.为了提高模型效率,本文提出几种跨尺度连接的优化方法:首先,沿用 BiFPN 双向输入的思想,但是并不是删除单向节点,而是为单向节点增加不同尺度输入而形成双向网络;其次,减少特征融合网络的输入,以便降低成本.本文将新特性网络命

名为简化版双向特性金字塔网络 (BiFPN-mini),如图 3c 所示.BiFPN-mini 定义如下:

$$\begin{aligned} p_5^{\text{td}} &= \text{Conv}(p_5^{\text{in}} + \text{Resize}(\text{Resize}(p_5^{\text{in}}))), \\ p_3^{\text{td}} &= \text{Conv}(p_3^{\text{in}} + \text{Resize}(p_5^{\text{td}})), \\ p_1^{\text{td}} &= \text{Conv}(p_1^{\text{in}} + \text{Resize}(p_3^{\text{td}})), \\ p_1^{\text{out}} &= \text{Conv}(p_1^{\text{in}} + p_1^{\text{td}} + \text{Resize}(\text{Resize}(p_1^{\text{td}}))), \\ p_3^{\text{out}} &= \text{Conv}(p_3^{\text{in}} + p_3^{\text{td}} + \text{Resize}(p_1^{\text{out}})), \\ p_5^{\text{out}} &= \text{Conv}(p_5^{\text{in}} + p_1^{\text{td}} + \text{Resize}(p_3^{\text{out}})). \end{aligned} \quad (7)$$

1.4 FIRE-Attention 网络

注意力机制的主要作用是增强目标的特征.注意力机制主要分为通道注意力机制、空间注意力机制和混合注意力机制等.

如图 4 通道注意力机制主要作用是将特征图每个通道与不同权重进行相乘,可以忽略无关通道信息而关注重点通道信息.空间注意力机制的作用是将特征图每个位置的特征与不同权重进行相乘,可以忽略无关空间信息而关注重点空间信息.通道注意力机制与空间注意力机制更关注区域或者通道,而且注意力是确定性的注意力,学习完成后直接可以通过网络生成.注意力是可微的,可以微分的注意力就可以通过神经网络算出梯度,并且通过反向传播来学习获得注意力的权重.

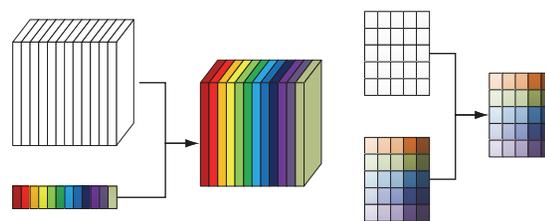


图 4 通道注意力机制与空间注意力机制
Fig. 4 Channel attention mechanism and spatial attention mechanism

传统的注意力机制是无监督的,它通过网络的反馈计算获得自适应的注意力矩阵.因为它是无监督的,所以会生成一些背离预期的结果.针对这个问题,本文提出一种监督式的空间注意力机制 FIRE-Attention,通过检测器的标签值生成一组二值图像标签来监督获得的注意力矩阵.

FIRE-DET 模型的损失函数在 YOLO-head 的损失函数的基础上增加了 FIRE-DET 模型中 FIRE-Attention 的损失值.图 5 的右图为图像分割的注意力图.

为了提高图像分割网络提取的注意力图的准确



图5 原图与注意力图

Fig. 5 Original image and attraction map

性,将 BiFPN-mini 网络输出的特征图进行反卷积,与卷积后得到的注意力图和火焰对象的二值图像通过将图像与图像长宽 1/4 大小的卷积核进行平均卷积,来获取类似人眼关注事物的注意力图像之间定义一个损失函数,如图 6 所示。

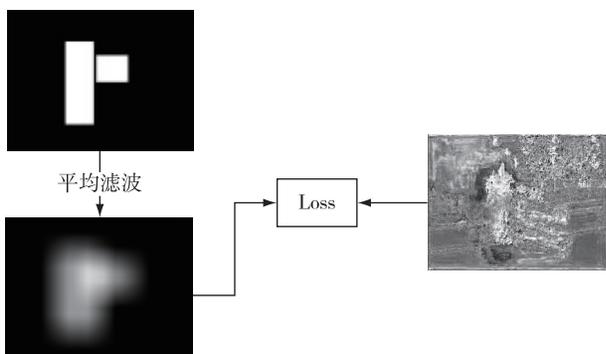


图6 图像分割损失值计算流程

Fig. 6 Flow chart of image segmentation loss calculation

损失函数 (L) 计算如下:

$$z(i, j) = \sum_{k, l} m(i+k, j+l) n(k, l),$$

$$L = \max(x, 0) - x \times z + \log(1 + e^{-\text{abs}(x)}), \quad (8)$$

其中 m 为火焰目标二值图像, n 为均值滤波, x 为图像分割网络输出的注意力图, i 和 j 分别代表图像的 x 轴与 y 轴坐标, k 和 l 分别代表卷积核的 x 轴与 y 轴坐标, z 为火焰目标二值图像进行均值滤波处理后的特征图.将图像分割网络输出的 3 个注意力图,分别与标记火焰的二值图像使用上述损失函数进行反馈计算,并将 L 与 YOLO-head 损失函数的和作为 FIRE-DET 模型的损失函数,可以使网络的损失函数快速下降并达到收敛。

2 实验与分析

2.1 火焰数据集

由于火灾检测研究尚处于初始阶段,目前没有较为完整和权威的公开火灾视频数据库.本文中的训练数据集是通过收集互联网上的图,然后根据 PASCAL VOC 的格式进行手工标注图片中的目标得到的,将其中火焰区域标注为 fire,未标注的区域为背景.图 7 为实验用的火焰场景的图片.火焰数据中含有小目标、多目标和类火焰数据.在研究过程中使用的数据,可联系笔者获取。

2.2 实验过程

本文实验所用机器的基本配置为 GPU: GTX 1080 Ti, CPU: i5-7500, 内存 16 GB, 操作系统为 Window10, 使用深度学习中的 tensorflow 框架.对模型使用动量为 0.9、权重衰减为 $1e-8$ 的 Adam 优化器进行训练,并使用裁剪、缩放、翻转与增减对比度增强数据。

在实验过程中,在遵循控制变量的原则下,各算法都在相同环境、相同训练集中训练 100 个周期,然

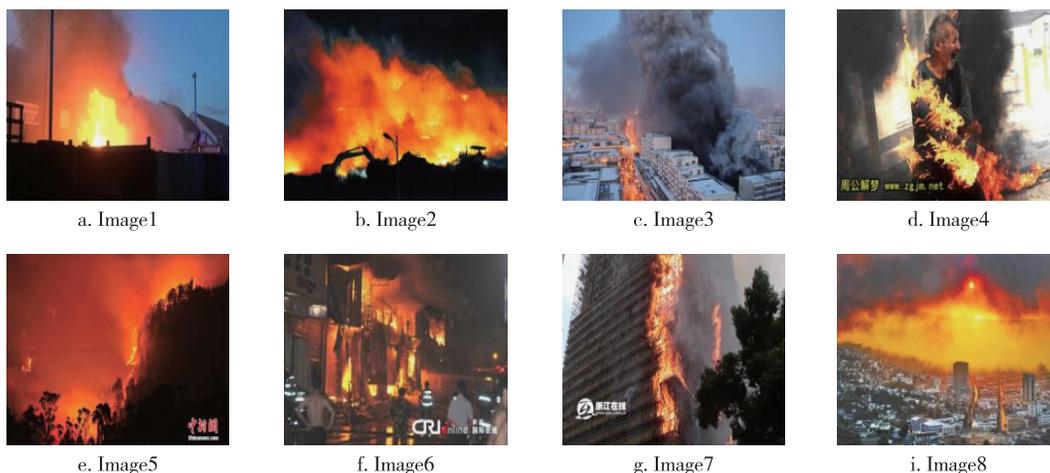


图7 训练图片范例

Fig. 7 Training picture examples

后进行检测实验验证 FIRE-DET 模型的鲁棒性.为了验证本文提出的贡献点和合理性,还进行了消融实验与分析.

2.3 评价指标

在模式识别任务中,真阳性和假阳性的定义如表 2 所示.如果预测框与地面真值框之间的交并比 (IOU) > 0.5,则预测框为“真阳性”(TP),否则,它是“假阳性”(FP).由于图像中没有预测到物体的每一部分都被认为是背景,所以测量“真阴性”(TN)是非必要的,故文中只测量“假阴性”(FN)作为模型遗漏的对象(图 8).预测的边界框是红色的,火焰目标的真实框是绿色的.

表 2 混淆矩阵

Table 2 Confusion matrix

| 预测类 | 真实类 | |
|-----|-----|----|
| | 正 | 负 |
| 正 | TP | FN |
| 负 | FP | TN |

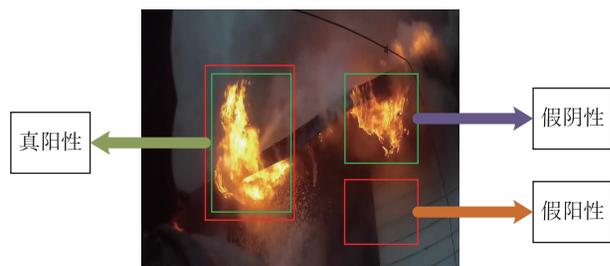


图 8 真阳性、假阳性及假阴性示意图

Fig. 8 Diagrams of true positive, false positive and false negative

准确率(A)表示预测正确的样本在所有样本中占的比例,一般准确率越高分类器越好.此外,精确度(P)和召回率(R)也是算法评价的两个重要指标.在分类任务中,一个类的精确度是真阳性数除以预测并标记为属于阳性类的元素总数.召回率的定义是真阳性数除以实际属于阳性类的元素总数.准确率、精确度和召回率的定义如下:

$$A = \frac{TP+TN}{TP+TN+FP+FN} \quad (9)$$

$$P = \frac{TP}{TP+FP} \quad (10)$$

$$R = \frac{TP}{TP+FN} \quad (11)$$

2.4 实验结果

实验中用到的验证数据集来源如表 3 所示,图 9 是在验证数据集下使用 FIRE-DET 模型对火焰进行检测后的部分效果图.由于模型的鲁棒性与数据增强的优点,因而对不同形态和不同颜色(图 9a—c))的火焰检测效果较好.由于模型本身采用中心点预测的方法,减少了回归过程中的震荡幅度,并且在特征提取网络中使用 Muti-Conv 结构构建了 FIRE-Net 网络结合 BiFPN-mini 提取多尺度特征,提升了检测器对火焰特征的提取效果,所以对于火焰形态较小的目标(图 9d—f)),该模型亦能够有效地提供更精确的定位.对拥有多个目标的图像(图 9g—i)也能快速检测.在面对类火焰的对象该模型也不会产生虚检(图 9j—l)),这是因为模型中 FIRE-Attention 结构能通过监督的方式抑制模型对非火焰物体误检的概率.FIRE-DET 模型对不同场景下不同长宽比的火焰展现出了较高的适应性,同时对类火焰对象不易产生误检,表明本方法对火焰和非火焰视频拥有良好的检测效果和鲁棒性.

本文将 FIRE-DET 与先进的检测器对比它们的准确率、精确率、召回率和检测速度(FPS),以验证目标测器的性能,结果如表 4 所示,可以发现 FIRE-DET 检测器比其他的检测器效率更高.YOLOv4-tiny 拥有极快的检测速度并可以部署于移动设备上,但其在其他评估指标中表现欠佳;Faster R-CNN 方法对火焰检测的效果不好;Dunnings 等^[10]的方法中虽然检测效果好,但是检测速度较低.

本文方法保持较快的速度和较高的准确度,主要是因为基于 FIRE-DET 模型的火焰检测方法,采用 FIRE-Net 增加模型的感受野,并且通过 BiFPN-mini 网络融合多尺度的特征加强了检测特征的语义信息,因而对不同场景下不同长宽比的火焰目标的检测,该模型亦能够有效地提供更精确的定位.同时,

表 3 火灾视频来源

Table 3 Fire video sources

| 机构名称 | 网址 |
|-------------------|---|
| 土耳其比尔肯大学实验室/视觉研究室 | http://signal.ee.bilkent.edu.tr/VisiFire/Demo/SampleClips.html |
| Github 公开数据集 | https://github.com/steffensbola/furg-fire-dataset |
| MIVIA | https://mivia.unisa.it/datasets/video-analysis-datasets/fire-detection-dataset/ |

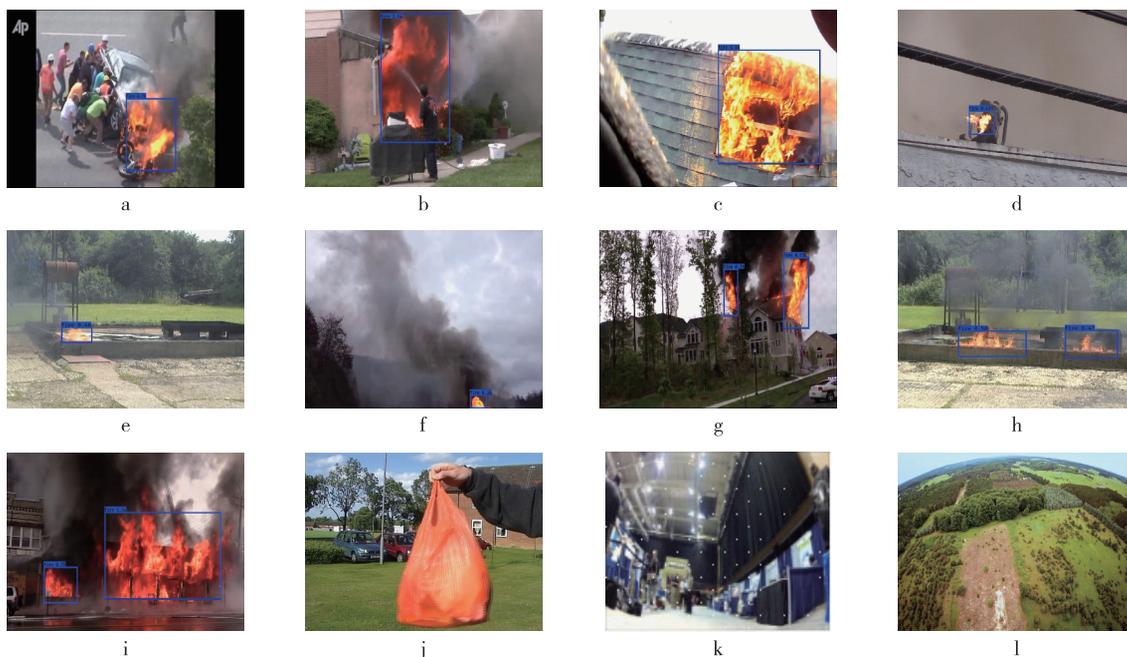


图9 测试结果范例

Fig. 9 Examples of test result

表4 FIRE-DET 实验对比

Table 4 Performance comparison between
FIRE-DET and other detectors

| 方法 | A/% | P/% | R/% | 检测速度/ FPS |
|------------------------------|-------|-------|-------|--------------|
| YOLOv3 ^[12] | 70.49 | 72.55 | 72.32 | 22 |
| YOLOv4 ^[13] | 76.83 | 77.29 | 71.68 | 24 |
| YOLOv4-tiny ^[14] | 74.29 | 72.43 | 70.29 | 73 |
| YOLOX ^[15] | 72.42 | 71.30 | 70.23 | 15 |
| CenterNet ^[20] | 65.07 | 90.31 | 36.99 | 56 |
| Faster R-CNN ^[19] | 67.62 | 34.41 | 76.99 | 34 |
| Dunnings 等 ^[10] | 95.41 | 96.41 | 95.11 | 17 |
| FIRE-DET | 97.42 | 95.84 | 98.21 | 85 |

模型中 FIRE-Attention 能提高检测模型的鲁棒性,减少检测过程中假阳性发生概率.实验结果表明,本文模型检测准确率高达 97.42%,检测速度为 85 FPS.基于 FIRE-DET 模型的检测方法检测准确率、召回率和检测速度均优于其他模型.

另外,本文对提出的 FIRE-DET 网络进行了消融实验,如表 5 中 α 表示 FIRE-Net 网络, β 表示 BiFPN-mini 网络, γ 表示 FIRE-Attention 网络,实验中将 YOLOv4 作为对照组.在第 1 组实验中将 FIRE-Net 网络加入实验中,在添加 FIRE-Net 网络后的模型比对照组模型的精确度提高了约 9 个百分点,且检测速度提高了 1 倍,这主要是因为 Muti-Conv 提高了模

表5 消融实验结果

Table 5 Ablation experiment results

| 组别 | α | β | γ | A/% | P/% | R/% | 检测速度/ FPS |
|-------|----------|---------|----------|-------|-------|-------|--------------|
| 对照组 | | | | 76.83 | 77.29 | 71.68 | 22 |
| 第 1 组 | ✓ | | | 85.12 | 88.31 | 81.51 | 53 |
| 第 2 组 | ✓ | | ✓ | 96.17 | 95.22 | 98.31 | 42 |
| 第 3 组 | ✓ | ✓ | ✓ | 97.42 | 95.84 | 98.21 | 85 |

型对火焰特征的提取效果,其次是因为 FIRE-Net 网络仅使用了 3 次下采样,提高了检测速度.第 2 组实验在第 1 组实验的基础上添加了 FIRE-Attention 网络,再次提高了模型的检测准确率,这是由于增加了监督式的注意力机制,但是也降低了模型的效率.第 3 组实验在第 2 组实验的基础上添加了 BiFPN-mini 网络,使模型的检测速度达到了 85 FPS,比第 2 组实验的检测效率提高了 1 倍.结果表明, FIRE-DET 模型通过 FIRE-Net 网络与 FIRE-Attention 网络可以增加特征提取效果,提高模型检测的准确率和检测效率,而 BiFPN-mini 能加快模型的检测速度.

3 结论

为了提高在自然场景下对火焰检测的检测速度,本文从模型检测效率出发,结合注意力机制和卷积神经网络在特征提取上的优势,提出了一种基于 FIRE-DET 的火焰检测方法.在主干网络中利用多个

卷积核进行卷积提高模型感受野,在特征融合网络中利用简单的双向结构提取更多的特征,最后利用提出的注意力机制提升模型的准确率.实验结果表明,在面对小目标、多形态和多目标时,本文模型均能准确预测.本文模型和现有主流的检测模型相比,在检测准确率和速率上表现更佳,达到预期实验效果.FIRE-DET在消融实验中也体现了FIRE-Net、BiFPN-mini和FIRE-Attention的优点,FIRE-Net、FIRE-Attention可以提高模型的准确率,BiFPN-mini可以提高模型的检测速度.本文的检测方法受限于检测数据集,在未来的工作中,将对小样本数据集训练的问题进行深入研究,进一步提高模型的检测效果.

参考文献

References

- [1] 张志,许文浩.澳大利亚2019—2020森林火灾对我国应急管理体系建设的启示[J].中国应急救援,2020(2):18-22
ZHANG Zhi, XU Wenhao. Emergency management system in China; lessons learned from Australia 2019 - 2020 bushfires[J]. China Emergency Rescue, 2020(2):18-22
- [2] 张慧珍,严云洋,刘以安,等.基于超像素分割与闪频特征判别的视频火焰检测[J].数据采集与处理,2018,33(3):512-520
ZHANG Huizhen, YAN Yunyang, LIU Yi'an, et al. Video flame detection based on super pixel segmentation and flash frequency feature discrimination[J]. Journal of Data Acquisition & Processing, 2018, 33(3):512-520
- [3] Gong F M, Li C T, Gong W J, et al. A real-time fire detection method from video with multifeature fusion[J]. Computational Intelligence and Neuroscience, 2019, 2019:1939171
- [4] Ko B C, Ham S J, Nam J Y. Modeling and formalization of fuzzy finite automata for detection of irregular fire flames[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2011, 21(12):1903-1912
- [5] Gunay O, Toreyin B U, Kose K, et al. Entropy-functional-based online adaptive decision fusion framework with application to wildfire detection in video[J]. IEEE Transactions on Image Processing, 2012, 21(5):2853-2865
- [6] Dimitropoulos K, Barmpoutis P, Grammalidis N. Spatio-temporal flame modeling and dynamic texture analysis for automatic video-based fire detection [J]. IEEE Transactions on Circuits and Systems for Video Technology, 2015, 25(2):339-351
- [7] Kim B, Lee J. A video-based fire detection using deep learning models [J]. Applied Sciences, 2019, 9(14):2862-2881
- [8] Aslan S, Gdkbay U, Treyin B U, et al. Deep convolutional generative adversarial networks for flame detection in video [C] // Computational Collective Intelligence, 2020:807-815
- [9] Zhang J M, Zhu H Q, Wang P Y, et al. ATT squeeze U-net: a lightweight network for forest fire detection and recognition [J]. IEEE Access, 2021, 9:10858-10870
- [10] Dunning A J, Breckon T P. Experimentally defined convolutional neural network architecture variants for non-temporal real-time fire detection [C] // 2018 25th IEEE International Conference on Image Processing (ICIP). October 7 - 10, 2018, Athens, Greece. IEEE, 2018:1558-1562
- [11] Redmon J, Farhadi A. YOLOv3: an incremental improvement [J]. arXiv e-print, 2018, arXiv:1804.02767
- [12] Bochkovskiy A, Wang C Y, Liao H Y Mark. YOLOv4: optimal speed and accuracy of object detection [J]. arXiv e-print, 2020, arXiv:2004.10934
- [13] Jiang Z C, Zhao L Q, Li S Y, et al. Real-time object detection method based on improved YOLOv4-tiny [J]. arXiv e-print, 2020, arXiv:2011.04244
- [14] Ge Z, Liu S T, Wang F, et al. YOLOX: exceeding YOLO series in 2021 [J]. arXiv e-print, 2021, arXiv:2107.08430
- [15] 陈西江,安庆,班亚.优化EfficientDet深度学习的车辆检测[J].南京信息工程大学学报(自然科学版),2021,13(6):653-660
CHEN Xijiang, AN Qing, BAN Ya. Optimized EfficientDet deep learning model for vehicle detection [J]. Journal of Nanjing University of Information Science & Technology (Natural Science Edition), 2021, 13(6):653-660
- [16] 梅舒欢,闵巍庆,刘林虎,等.基于Faster R-CNN的食品图像检索和分类[J].南京信息工程大学学报(自然科学版),2017,9(6):635-641
MEI Shuhuan, MIN Weiqing, LIU Linhu, et al. Faster R-CNN based food image retrieval and classification [J]. Journal of Nanjing University of Information Science & Technology (Natural Science Edition), 2017, 9(6):635-641
- [17] 金啸宇,尹媵,倪军,等.一种基于场景合成和锚点约束的SAR目标检测网络[J].南京信息工程大学学报(自然科学版),2020,12(2):210-215
JIN Xiaoyu, YIN Qiang, NI Jun, et al. SAR target detection network based on scenario synthesis and anchor constraint [J]. Journal of Nanjing University of Information Science & Technology (Natural Science Edition), 2020, 12(2):210-215
- [18] He K M, Gkioxari G, Dollár P, et al. Mask R-CNN [C] // 2017 IEEE International Conference on Computer Vision (ICCV). October 22 - 29, 2017, Venice, Italy. IEEE, 2017:2980-2988
- [19] 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
- [20] Zhou X Y, Wang D Q, Krhenbhl P. Objects as points [J]. arXiv e-print, 2019, arXiv:1904.07850
- [21] Tan M X, Le Q V. EfficientNet: rethinking model scaling for convolutional neural networks [J]. arXiv e-print, 2019, arXiv:1905.11946
- [22] Lin T Y, Dollár P, Girshick R, et al. Feature pyramid networks for object detection [C] // 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR).

- July 21 – 26, 2017, Honolulu, HI, USA. IEEE, 2017: 936-944
- [23] Liu S, Qi L, Qin H F, et al.Path aggregation network for instance segmentation[C]//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition.June 18–23, 2018, Salt Lake City, UT, USA. IEEE, 2018: 8759-8768
- [24] Ghiasi G, Lin T Y, Le Q V.NAS-FPN: learning scalable feature pyramid architecture for object detection [C] // 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). June 15 – 20, 2019, Long Beach, CA, USA.IEEE, 2019: 7029-7038
- [25] Tan M X, Pang R M, Le Q V.EfficientDet: scalable and efficient object detection[C] // 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). June 13 – 19, 2020, Seattle, WA, USA. IEEE, 2020: 10778-10787
- [26] Hu J, Shen L, Albanie S, et al.Squeeze-and-excitation networks[J].IEEE Transactions on Pattern Analysis and Machine Intelligence, 2020, 42(8): 2011-2023
- [27] Woo S, Park J, Lee J Y, et al.CBAM: convolutional block attention module[M] // Computer Vision — ECCV 2018. Cham: Springer International Publishing, 2018: 3-19

FIRE-DET : an efficient flame detection model

CHEN Haolin^{1,2} GAO Shangbing^{1,2} XIANG Lin¹ CAI Chuangxin¹ WANG Changchun¹

1 Faculty of Computer and Software Engineering, Huaiyin Institute of Technology, Huaian 223001

2 Laboratory for Internet of Things and Mobile Internet Technology of Jiangsu Province, Huaiyin Institute of Technology, Huaian 223001

Abstract In view of the increasing concern on model efficiency in computer vision, this paper proposed several optimization schemes to improve the flame detection models in model efficiency as well as the detection performance. A backbone network (FIRE-Net) was constructed from a multi-convolution combined structure, which can efficiently extract rich flame features from multiple scales. Then an improved weighted bidirectional feature pyramid network (BiFPN-mini) was used to quickly achieve multi-scale feature fusion. In addition, a new attention mechanism (FIRE-Attention) was proposed to make the detector more sensitive to flame characteristics. The above optimizations were combined to develop a new flame detector abbreviated as FIRE-DET, which was then trained on self-built dataset and tested on internet videos. The experimental results showed that the FIRE-DET outperformed mainstream algorithms by its flame recognition accuracy of 97% and frame rate of 85 FPS, thus provides a more common solution to solve the flame detection.

Key words feature extraction; feature fusion; attention mechanism; flame detection