文章编号: 1671-6833(2021)05-0013-06

# 基于多尺度特征融合的火灾检测模型

张坚鑫, 郭四稳, 张国兰, 谭 琳

(广州大学 计算机科学与网络工程学院,广东 广州 510006)

**摘 要**: 对双阶段目标检测模型 Faster R-CNN 进行火灾检测应用的改进。采用 Resnet101 模型作为特征提取网络,使用特征金字塔结构 FPN 提取了 Resnet101 的浅层特征和高层特征,将 Resnet101 的浅层特征图输入 Inception Module 结构提取多种尺寸的卷积特征,使用像素注意力机制和信道注意力机制对目标位置进行强化并弱化其余部分,使得检测目标更加精确。该网络避免了主干网络特征提取不充分的问题,融合了多种尺度的特征来区分火灾区域和非火灾区域,有效提高了火灾图像数据集的检测准确率,最终得到的平均检测准确率 *MAP* 为 0.851。

**关键词**: 深度学习; 火灾检测; 卷积神经网络; 多尺度特征; 特征金字塔结构

**中图分类号**: TP183 **文献标志码**: A **doi**: 10.13705/j.issn.1671-6833.2021.05.016

## 0 引言

火灾是全球范围内的灾难性破坏事件,会在短时间内对人民的生命和财产安全造成重大的损害。随着人工智能技术的不断发展,深度卷积神经网络已经在图像识别和检测方面展示了最先进的性能[1]。图像分类识别任务是对图像是否有该物体进行分类识别,图像检测任务是对物体的对应区域进行类别的标记。黄文锋等[2]将深度卷积神经网络应用于视频监控中火灾烟雾和火焰的识别和检测。Frizzi 等[3]提出了用于视频火焰和烟雾识别的卷积神经网络,结果表明,基于卷积神经网络的方法比一些常规的视频火灾探测方法具有更好的性能。Sharma 等[4]探讨了将 Vgg16[5]和 Resnet50[6]用于火灾识别。林作永等[7]使用Inception[8]、ResNet[6]、MobileNet[9]作为特征提取器 Backbone,使用 Faster R-CNN[10]、SSD[11]、R-FCN[12]作为深度学习检测框架进行火灾烟雾检测,结果发现,使用 Inception V2[13]特征提取网络能增加检测精度,Faster R-CNN 检测效果最好,SSD 速度最快,但定位精度不够,R-FCN 获得了速度与精度的平衡[14]。基于这些考虑,本文对双阶段目标检测网络 Faster R-CNN 在火灾图像的应用进行改进和研究,以实现对火灾区域的精准定位。

## 1 相关工作

在至今为止的目标检测算法中,基于深度学习的单阶段和双阶段目标检测算法都能够实现较高的检测准确率,但是两种算法各有优势: 双阶段的精度更准,单阶段的速度更快。双阶段的 Faster R-CNN 系列方案第 1 步是先训练好特征提取网络; 第 2 步是使用该网络生成很多候选框来检测目标。单阶段方案是将特征提取网络和生成候选框在同一个网络里完成。这里选用双阶段 Faster R-CNN 的目标检测方案进行改进。

在传统的目标检测任务中一般使用特征提取和机器学习结合的算法,特征提取方法往往有 3 种: 第 1 种是最基本的,一般是颜色、纹理等底层的特征; 第 2 种是中层特征,一般是经过特征挖掘学习之后得到的特征,包括 PCA 特征和 LDA 学习得到的特征; 第 3 种是高层的特征,是将前面 2 种特征结合进一步挖掘计算得到的特征。基于这个考虑,双阶段 Faster R-CNN 目标检测任务中的特征提取网络能够融合多种尺度的特征来提升网络的效果。特征金字塔结构 FPN 方法[15]是对不同层中的信息进行融合,浅层网络包含更多的细节信息,高层网络则包含更多的语义信息。将卷

**收稿日期**: 2020-12-21; **修订日期**: 2021-03-20

**基金项目**: 国家重点研发计划项目(2018YFB1005104)

**通信作者**: 郭四稳(1963—),男,湖北天门人,广州大学教授,博士,主要从事图像处理、自动推理研究,E-mail: 2667690005@qq.com。

积网络中浅层和高层的特征图进行累加,使其不会丢失过多信息,从而提升检测的效果。Inception系列的方法[16-17]就是通过增加网络的宽度来提升卷积神经网络的性能。不同大小的卷积核会提取到不同的信息:较大的卷积核会提取到图像中的全局性的特征图信息;较小的卷积核会提取到图像中局部的特征图信息。视觉注意力机制主要是模仿人的大脑信号处理机制,获取全局图像中需要重点关注的目标区域的细节信息。使用通道注意力机制和像素注意力机制可以将前景和背景之间的细节信息进行区分,从而增强检测的效果。

## 2 多尺度特征融合网络

本文引入了一种改进的 Faster R-CNN 检测火灾区域的目标检测框架,如图 1 所示。该方法主要是基于 Resnet101 网络进行特征提取,得到生成的每个候选框的类别分数,从而找出最佳结果。为充分利用火灾场景的火焰特征,使用了预训练效果较好的 Resnet101 作为特征提取网络,加入特征金字塔结构 FPN,对浅层和高层特征进行提取,使用 Inception Module 结构提取多种卷积特征,并加入像素注意力机制和信道注意力机制,通过这些操作来突出特征使得生成候选框操作更加精准,以提高火灾区域定位的性能。

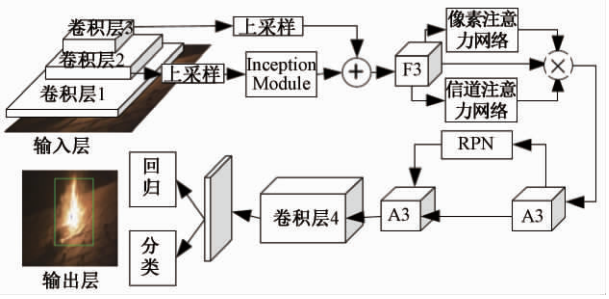


**图 1 多尺度特征融合网络结构**

**Figure 1 Multi-scale feature fusion network structure**

### 2.1 基础网络结构

一般认为浅层特征映射可以保留小对象的位置信息,而高层特征映射可以包含更高级别的语义提示。因此将语义信息充分的高层特征映射回分辨率较高、细节信息充分的底层特征,将两者以合适的方式融合来提升检测的效果。输入层的特征图尺寸为 600×600,使用特征提取网络Resnet 101 的网络架构[5]。在 Resnet101 网络基础上采用了特征金字塔结构 FPN,将 Resnet101 网络的卷积层 2 输出特征图和卷积层 3 输出特征图进行融合,如图 1 所示。通过这样的操作可以平衡语义信息和位置信息,同时忽略其他不太相关的特征。

### 2.2 Inception Module 网络结构

将 Resnet101 中的卷积层 2 输出特征图输入到 Inception Module 结构中可以抽取更多的特征。Inception Module 架构如图 2 所示。采用多个 1×1 卷积的作用是在相同尺寸的感受野中叠加更多的卷积,能提取到更丰富的特征,并且降低了计算复杂度。在上一层网络之后首先使用 1×1 卷积再进入较大的二维卷积的作用是当某个卷积层输入的特征数较多,对这个输入进行卷积运算将产生巨大的计算量;如果对输入先进行降维,减少特征数后再做卷积计算就会显著减少计算量。在这一层网络中还将较大的二维卷积分解成较小的一维卷积,例如将 3×3 卷积分解成 1×3 卷积和 3×1 卷积,将 5×5 卷积分解 1×5 卷积和 5×1 卷积,将 7×7 卷积分解成 1×7 卷积和 7×1 卷积。这个操作一方面可以减少参数、减轻过拟合,另一方面则增加了非线性扩展模型表达能力。非对称的卷积结构分解效果比对称地分解为几个相同的小卷积核效果更好,可以处理更丰富的空间特征,从而增加特征提取的多样性。

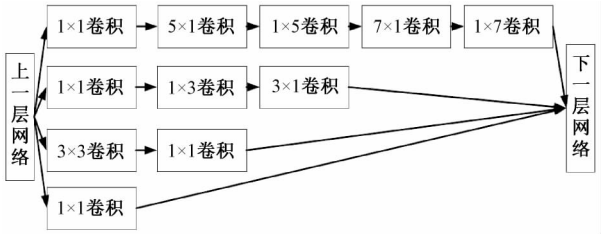


**图 2 Inception Module 网络结构**

**Figure 2 Inception Module network structure**

### 2.3 多维注意网络结构

将 Resnet101 的卷积层 2 和卷积层 3 融合得到的特征图 F3 输入一个多维注意力网络,如图 3 所示。多维注意力网络由像素注意力网络和信道注意力网络组成,用于抑制噪声和突出前景。像素注意力网络将更广泛的上下文信息编码为局部特征,从而增强其表示能力。在这个网络结构中的特征图 F3 需要进入 1 个新的 Inception Module,这里采用 5 个 3×3 的卷积核进行特征的抽取,输出 1 个两通道的显著特征图。显著特征图分别表示前景和背景的分数。由于显著特征图是连续的,非目标信息不会被完全消除,这有利于保留一定的上下文信息,提高鲁棒性。为了引导该网络学习到目标信息,这里采用了 1 个有监督学习的

方式,使用 1 个二值化图像作为真实标签,两通道的显著特征图作为预测标签,进而计算 2 个标签的交叉熵损失,并将其作为注意力损失添加到总损失的计算中。接着在显著特征图上执行 softmax 操作,并选择一个通道与 F3 相乘,获得新的信息特征图 A3。值得注意的是,在经过 softmax 函数操作之后的显著特征图的值在 [0,1 ],说明其可以减少噪声并相对增强对象信息。由于显著特征图是连续的,因此不会完全消除非对象信息,这对于保留某些细节信息并提高鲁棒性是有利的。信道注意力网络通过获取不同通道映射之间的相互依赖性,从而有效增强特征图对于特定语义的表征能力。在这个网络结构中主要采用 SE Block[18] 的结构,其重点是使用全局平均池化操作,将各个特征图全局感受野的空间信息置入特征图。全局平均池化操作可以通过减少模型中的参数总数来最小化过度拟合。与最大池化层类似,全局平均池化层用于减少三维张量的空间维度。然而全局平均池化层执行更极端的维数减少操作,将尺寸为 $h\times w\times d$ 的张量尺寸减小为 $1\times 1\times d$ 的尺寸。全局平均池化层通过简单地获取 $h$ 和 $w$ 的平均值,将每个 $h\times w$ 特征映射层减少为单个数字。每层卷积操作之后都接着一个样例特化激活函数,基于通道之间的依赖关系对每个通道进行一种筛选机制操作,以此对各个通道进行权值评比。

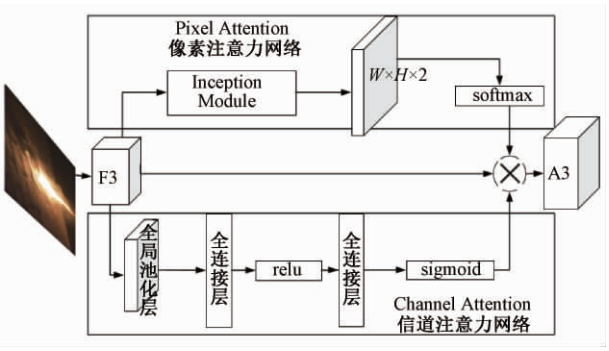


**图 3　多维注意力网络结构**

**Figure 3　Structure of multidimensional attention network**

### 2.4　损失函数

网络模型在进行训练时需要计算多个位置的损失: RPN 阶段的分类和回归损失、Faster R-CNN 阶段的分类和回归损失、多维注意力机制阶段的注意力损失,加起来的总损失就是整个网络的损失:

$$loss = loss_{\text{rpn}} + loss_{\text{fastrcnn}} + loss_{\text{attention}}。\quad (1)$$

RPN 阶段的分类和回归损失如式( 2) 所示,其中分类损失 $loss_{\text{r\_c}}$ 是判断方框中是否是火灾,也就是区分前景、背景两类物体的损失,如式( 3) 所示。$loss_{\text{r\_l}}$ 可以对方框位置进行评估和微调,也就是用于比较真实分类的预测参数和真实平移缩放参数的差别,如式( 4) 所示。$loss_{\text{r\_l}}$ 需要对损失进行 L1 平滑正则化处理,如式( 5) 所示,参数 $\sigma$ = 1。 Faster R-CNN 阶段的分类损失计算和 RPN 阶段一致,而回归损失的 L1 平滑正则化参数 $\sigma$ = 3。

$$loss_{\text{rpn}} = loss_{\text{r\_c}} + loss_{\text{r\_l}}。\quad (2)$$

$$loss_{\text{r\_c}} = \frac{1}{N_{\text{cls}}}\sum_i L_{\text{cls}}(p_i, p_i^*)。\quad (3)$$

$$loss_{\text{r\_l}} = \frac{1}{N_{\text{reg}}}\sum_i p_i^* L_{\text{reg}}(t_i, t_i^*)。\quad (4)$$

$$S_{\text{L1}}(x) = \begin{cases} 0.5x^2\times\sigma^{-2}, & |x| < 1/\sigma^2; \\ |x| - 0.5, & \text{其他。} \end{cases}\quad (5)$$

在多维注意力机制阶段,像素注意力损失如式( 6) 所示,其中 $w$ 和 $h$ 分别为方框的宽和高; $u_{ij}$ 为方框部分的真实标签,即原图的二值化图像,如图 4( b) 所示; $u_{ij}^{'}$ 为方框部分的预测标签,即输出的特征图,如图 4( c) 所示。

$$loss_{\text{attention}} = \frac{1}{h\times w}\sum_i^h\sum_j^w L_{\text{att}}(u_{ij}^{'}, u_{ij})。\quad (6)$$


**图 4　注意力图像**

**Figure 4　Attention image**

## 3　实验结果与分析

本实验选用了开源的 Tensorflow 框架,该框架可以快速地进行部署和实验,还可以对数据流图进行有效数据可视化,编程语言选用Python 3.6,硬件设备 CPU 为英特尔®酷睿™i7-7700HQ,内存大小为 8 G,GPU 型号为GeForce GTX 1060,显存大小为 6 G,系统环境为 Ubuntu 16.04。

### 3.1　火灾检测实验数据集

由于目前还没有比较完善的火灾图像检测数据集,本文自主构建了一个火灾图像检测的数据集,总共有 210 张图片。将所有图片使用标注软件labelImage进行标注,给图像中火灾区域标注 1 个方框,如图 5 所示。将标注信息保存到 xml 文件中,xml 文件包括图片文件的位置、名字、宽度 *width*、高度 *height* 和维度 *depth*,以及标注方框对

应的4个坐标 $x_{min}$、$y_{min}$、$x_{max}$ 和 $y_{max}$。将图片和标注文件制作成一一对应的 tfrecord 格式的数据集，然后按7:3的比例分成训练集和测试集，即训练集中有147张火灾图片，测试集有63张火灾图片，并将2个数据集均转化为 tfrecord 格式的数据。

(a) 实例1 (b) 实例2 (c) 实例3

**图5 标注实例图**

**Figure 5 Example drawing of annotation**

## 3.2 仿真实验

### 3.2.1 实验过程

实验首先将训练集数据 train.tfrecord 数据进行解析，得到图片数据和标签，加载卷积网络 resnet101.ckpt 作为特征提取网络，建立多尺度特征融合的目标检测网络，设置不同的参数来训练数据，使得网络得到最佳的训练效果。对训练集进行8万次迭代得到的损失曲线如图6所示，从图6可以看出，该模型在训练后达到收敛。

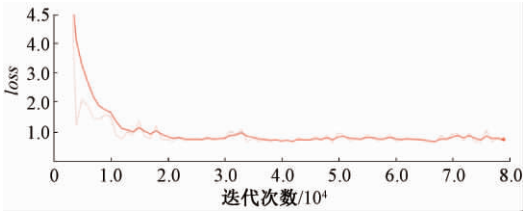


**图6 损失函数曲线**

**Figure 6 Loss function curve**

### 3.2.2 检测精度评估

在目标检测中，一般使用多种评价指标来进行模型的性能评价。二级指标准确率 $P$ 为在测试集的识别结果中预测正确的正例数目($TP$)占所有预测正例数目($FP+TP$)的比例，如式(7)所示。召回率 $R$ 为在测试集的识别结果中预测正确的正例数目($TP$)占所有真实正例数目($TP+FN$)的比例，如式(8)所示。利用准确率 $P$ 和召回率 $R$ 可以得到 $P$-$R$ 曲线，$P$-$R$ 曲线反映了分类器对正例的识别准确程度和对正例的覆盖能力之间的权衡。$AP$ 是 $P$-$R$ 曲线与 $X$ 轴围成的图形面积，如式(9)所示。平均检测准确率 $MAP$ 是对所有类别的 $AP$ 求均值，如式(10)所示。三级指标 $F_1$ 是准确率和召回率的调和平均值，如式(11)所示。$F_1$ 指标综合了 $P$ 与 $R$ 的产出结果。$F_1$ 的取值范围为0~1，1代表模型的输出结果最好，0代表模型的输出结果最差。

$$P = \frac{TP}{TP + FP}; \tag{7}$$

$$R = \frac{TP}{TP + FN}; \tag{8}$$

$$AP = \int_0^1 PR\mathrm{d}r; \tag{9}$$

$$MAP = \frac{\sum_{q=1}^{Q} AP(q)}{Q}; \tag{10}$$

$$F_1 = \frac{2 \times P \times R}{P + R}。 \tag{11}$$

本文所提出的多尺度特征融合网络模型的检测精度指标与其他的模型对比情况如表1所示。其中，所有模型都是以 Faster R-CNN 模型为基础网络模型，用于对比的模型特征提取网络分别使用了 Vgg16、Resnet101、Resnet101+FPN。从表1的对比结果可以发现，使用 Resnet101 相对于使用 Vgg16，$MAP$ 值提升了15.2%；使用 Resnet101+FPN 相对于仅使用 Resnet101，$MAP$ 值提升了3.6%；使用本文的多尺度特征融合网络模型相对于 Resnet101 特征提取网络有了8.5%的提升，相对于 Resnet101+FPN 也有4.9%的提升。多尺度特征融合网络模型的 $F_1$ 相对于 Resnet101 特征提取网络有2.3%的提高，相对于 Resnet101+FPN 也有1.7%的提高。

**表1 精度评价指标**

**Table 1 Accuracy evaluation index**

| 模型 | $R$ | $P$ | $MAP$ | $F_1$ |
|---|---|---|---|---|
| Vgg16 | 0.746 | 0.701 | 0.614 | 0.722 |
| Resnet101 | 0.825 | **0.928** | 0.766 | 0.873 |
| Resnet101+FPN | 0.873 | 0.887 | 0.802 | 0.879 |
| 多尺度 | **0.904** | 0.890 | **0.851** | **0.896** |

注：加粗数字为每列最优值。

各个模型对应的 $P$-$R$ 曲线如图7所示。从图7可以看出，多尺度特征融合火灾检测算法的 $P$-$R$ 曲线下方面积最大，即平均检测准确率 $MAP$ 最高。

### 3.2.3 检测速度评估

由于火灾检测的实时性要求，对于检测速度的评估也是性能评估的一个重要方面。对本文的 Resnet101 系列模型进行检测速度的评估，最终得到对比结果如表2所示。从表2可以看出，所有的模型均能检测出图像中较为明显的火焰。其中加载模型 Faster R-CNN 以 Resnet101 作为特征提

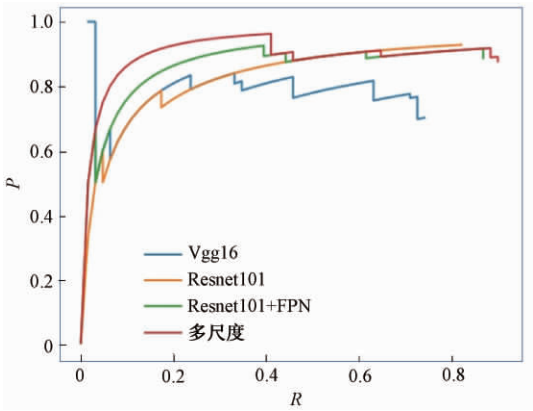


**图 7　*P-R* 曲线图**

**Figure 7　*P-R* curve**

取层时,不加上任何改进的检测效率最高,检测 1 张图片的时间 $t$ 为 0.83 s; 使用 Resnet101 为特征提取层加特征金字塔结构 FPN 的检测效率最低,检测 1 张图片的时间 $t$ 为 0.94 s; 使用 YOLOv3 网络[19] 的速度最快,时间只需要 0.05 s,但是 *MAP* 值相对较低为 0.762; 本文的多尺度特征融合网络得到了速度与精度的平衡,处于中间位置,检测一张图片需 0.86 s。

**表 2　不同方法在火灾检测数据集中的性能对比**

**Table 2　Performance comparison of different methods in fire detection dataset**

| 模型 | *MAP* | $t$/s |
|---|---|---|
| Vgg16 | 0.614 | 0.51 |
| Resnet101 | 0.766 | 0.83 |
| Resnet101+FPN | 0.802 | 0.94 |
| 多尺度 | **0.851** | 0.86 |
| YOLOv3 | 0.762 | **0.05** |

注: 加粗数字为每列最优值。

### 3.2.4　室内外火灾检测

火灾发生分室内和室外,使用本文多尺度特征融合火灾检测模型,输入图 8( a) 的室内火灾图像可以得到图 8( b) 的检测结果,输入图 8( c) 的室外火灾图像可以得到图 8( d) 的检测结果。通过对比可以看出,使用本文的多尺度特征融合火灾检测算法能够较好地检测出室内、室外火灾的目标区域。

## 4　结论

本文首先对火灾检测的现状进行了深入的分析,指出深度学习相对于手工提取特征的优势,并且参考了传统的机器学习的算法思想,在卷积神经网络中加入了多种特征的融合提取。首先将 Resnet101 的浅层特征和高层特征进行融合,接着使用 Inception Module 网络提取了多种尺寸的特征,最终加入了多维注意力机制,这些操作使得特征提取网络模型提取特征更加充分,更加接近人的视觉模型。该模型相对原始的 Resnet101 特征提取网络的检测精度提升了 8.5%,相比 Resnet101 加特征金字塔结构 FPN 也有 4.9%的提升,最终得到平均检测准确率 *MAP* 为 0.851。对 Resnet101 特征提取网络进行多种改进操作后检测 1 张火灾图片的时间为 0.86 s。经过验证可知,本文的火灾检测算法能够检测出室内和室外火灾的目标区域,在保证检测速度与原始模型[10] 接近的同时,精度相对于原始模型得到较大的提升。


**图 8　室内外火灾检测结果**

**Figure 8　Fire image input and output of indoor and outdoor**

## 参考文献:

[1]　杨文柱,刘晴,王思乐,等.基于深度卷积神经网络的羽绒图像识别[J].郑州大学学报(工学版),2018,39(2):11-17.

[2]　黄文锋,徐珊珊,孙燚,等.基于多分辨率卷积神经网络的火焰检测[J].郑州大学学报(工学版),2019,40(5):80-84.

[3]　FRIZZI S,KAABI R,BOUCHOUICHA M,et al.Convo-

lutional neural network for video fire and smoke detection[C]//IECON 2016-42nd Annual Conference of the IEEE Industrial Electronics Society. Piscataway: IEEE,2016: 877-882.

[4] SHARMA J,GRANMO O C,GOODWIN M,et al.Deep convolutional neural networks for fire detection in images [J]. Engineering applications of neural networks,2017,744: 183-193.

[5] SIMONYAN K, ZISSERMAN A. Very deep convolutional networks for large-scale image recognition [C]//International Conference on Learning Representations. San Diego,USA: ICLR,2015: 1-14.

[6] HE K M,ZHANG X Y,REN S Q,et al.Deep residual learning for image recognition[C]//2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR).Piscataway: IEEE,2016: 770-778.

[7] 林作永,谌瑶.基于深度卷积神经网络的火灾预警算法研究[J].信息通信,2018(5):38-42.

[8] SZEGEDY C,VANHOUCKE V,IOFFE S,et al.Rethinking the inception architecture for computer vision [C]//2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR).Piscataway: IEEE,2016: 2818-2826.

[9] HOWARD A G,ZHU M L,CHEN B,et al.MobileNets: efficient convolutional neural networks for mobile vision applications[EB/OL].(2017-04-17) [2020-12-01].https://arxiv.org/pdf/1704.04861.pdf.

[10] 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.

[11] LIU W,ANGUELOV D,ERHAN D,et al.SSD: single shot MultiBox detector[M]//Computer Vision-ECCV 2016. Cham: Springer International Publishing, 2016: 21-37.

[12] DAI J F,LI Y,HE K M,et al.R-FCN: object detection via region-based fully convolutional networks[C]// NIPS'16: Proceedings of the 30th International Conference on Neural Information Processing Systems.Barcelona,Spain: NIPS,2016: 379-387.

[13] IOFFE S,SZEGEDY C.Batch normalization: accelerating deep network training by reducing internal covariate shift[C]//32nd International Conference on Machine Learning 2015.Lille,France: ICML,2015: 448-456.

[14] 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.

[15] 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).Piscataway: IEEE,2017: 936-944.

[16] SZEGEDY C,LIU W,JIA Y Q,et al.Going deeper with convolutions [C]//2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE,2015: 1-9.

[17] SZEGEDY C, IOFFE S, VANHOUCKE V, et al. Inception-v4,Inception-ResNet and the impact of residual connections on learning[EB/OL].(2016-08-23) [2020-12-01].https://arxiv.org/pdf/1602.07261.pdf.

[18] HU J, SHEN L, SUN G. Squeeze-and-excitation networks[C]//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE,2018: 7132-7141.

[19] REDMON J, FARHADI A. YOLOv3: an incremental improvement[EB/OL].(2018-04-08) [2020-12-01]. https://arxiv.org/pdf/1804.02767.pdf.

## Fire Detection Model Based on Multi-scale Feature Fusion

ZHANG Jianxin, GUO Siwen, ZHANG Guolan, TAN Lin

(School of Computer Science and Cyber Engineering, Guangzhou University, Guangzhou 510006, China)

**Abstract**: This paper aims to modify the two-scene detection model Faster R-CNN. Specifically, this model uses Resnet101 to extract features which are processed by pyramid structure FPN to extract the shallow and high-level features of Resnet101. The shallow feature map of Resnet101 is input into Inception Module structure to obtain the convolutional features of multiple sizes, and finally the proposed model uses the pixel attention mechanism and channel attention mechanism to emphasize the target position and weaken the rest, which makes the detection target more accurate. This network avoids the problem of insufficient feature extraction of trunk network, and integrates features of various scales to distinguish fire area and non-fire area, thus effectively improves the detection accuracy of fire image data sets, and mean average precision *MAP* is 0.851.

**Key words**: deep learning; fire detection; convolutional neural network; multi-scale features; feature pyramid network