基于区域全卷积网络结合残差网络的火焰检测方法

洪伟，李朝锋

江南大学物联网工程学院，江苏 无锡 214122

摘要　传统的火焰检测方法大多基于火焰的物理信号手动设计火焰特征，根据其使用模式进行识别。这类方法容易被外部环境干扰，且手动设计的火焰特征泛化性不强，当火焰形态或场景变化剧烈时，会降低识别精度。针对这一问题，提出了一种基于区域全卷积网络(R-FCN)结合残差网络(ResNet)的深度学习方法对火焰进行检测。通过特征提取网络自动提取特征，利用 R-FCN 确定火焰位置，并使用 ResNet 对该位置的火焰进行二次分类，以进一步降低误报率。该方法实现了端到端自动获取火焰特征并进行相应检测的过程，省去了传统火焰特征提取的过程。本文方法在 Bilkent 大学的火焰视频数据集上平均识别精度达到 98.25%。

关键词　图像处理；火焰检测；深度学习；区域全卷积网络；残差网络

中图分类号　TP391　　**文献标识码**　A　　**doi**：10.3788/LOP55.041011

Flame Detection Method Based on Regional Fully Convolutional Networks with Residual Network

Hong Wei, Li Chaofeng

School of Internet of Things Engineering, Jiangnan University, Wuxi, Jiangsu 214122, *China*

Abstract　The flame pattern is artificially designed by most of the traditional flame detection methods based on physical signal of the flame, and is identified according to the pattern recognition method. These methods are easy to be interfered by the external environment. Because the generalization of artificially designed flame feature is not strong, the recognition accuracy will reduce when the flame shape or scene changes violently. To solve this problem, a method of deep learning for detecting the flame based on the regional full convolution network (R-FCN) with the residual network (ResNet) is proposed. The feature is extracted automatically by the feature extraction network, and the flame position is determined by R-FCN, and it is secondary classified by ResNet for further reducing the false alarm rate. The proposed method, which eliminates the feature extraction process of the traditional flame, realizes end-to-end automatic acquisition of flame characteristics and performs corresponding detection processes. An average recognition accuracy reaches to 98.25% in the flame video data set of Bilkent University.

Key words　image processing; flame detection; deep learning; regional full convolution network; residual network

OCIS codes　100.4996; 100.5010; 140.1135

1　引　　言

火灾是一种常见且容易发生的灾害，在各类工业控制过程中火灾对公众的生命与财产构成了重大威胁。由于火灾具有突发性而且危害大，所以构建准确有效的火焰识别算法对预防火灾拥有重要的意义。

传统的火焰检测算法大致可归纳为两类：基于传感器的火焰检测算法和基于图像的火焰检测算法。基于传感器的火焰检测算法是通过感温和感烟的方式进行火焰检测，如徐炀[1]构建了智能火灾自动报警系统，通过使用火灾报警控制器、感温感烟探测器等物理装置侦测周围温度和空气变换进行火焰

收稿日期：2017-09-29；**收到修改稿日期**：2017-11-01

基金项目：国家自然科学基金(61771223)

作者简介：洪伟(1992—)，男，硕士研究生，主要从事目标检测、深度学习方面的研究。E-mail：517332051@qq.com

导师简介：李朝锋(1971—)，男，博士，教授，硕士生导师，主要从事人工智能与模式识别方面的研究。E-mail：wxlichaofeng@126.com(通信联系人)

检测；胡幸江[2]针对常规的感温感烟探测器等物理装置探测距离短、响应慢等缺点，提出使用多波段红外火焰探测器进行火焰检测。但是，这类方法对环境依赖性强、覆盖范围小、成本较高。随着高清摄像机、图像处理技术的不断发展，基于图像的火焰检测算法逐渐成为主流，张进华等[3]根据火焰的物体特性，利用火焰区域边缘点到中心质点的高度，并使用离散傅里叶变换排除非火焰区域；荣建忠等[4]和Chen等[5]利用RGB和HIS两个颜色空间获取疑似火焰区域，并通过判断其面积增长和中心稳定情况获取火焰的动态像素，引入并改进了统计地形特征的纹理描述方法，采用神经网络算法对火焰候选区域进行分类；李文辉等[6]使用日夜两用型红外摄像机，通过获取火焰在红外和可见光两种状态下的颜色模型进行火焰检测；严云洋等[7]通过RGB彩色空间建立了火焰的彩色模型，并提出了基于连通区域面积阈值化的单像素宽度目标轮廓特征的抽取方法。但是上述传统基于图像的火焰检测识别方法受制于手工设计火焰特征，随着场景变换和场景复杂程度加深，手工设计的火焰特征无法很好描述火焰目标并有效区分疑似火焰目标和火焰目标，造成识别精度的下降。

针对上述基于传统图像火焰检测算法的缺陷，本文提出一种通过训练特征提取网络自动提取火焰特征，并在该特征上使用区域全卷积网络(R-FCN)定位火焰目标，最后使用残差网络(ResNet)进行二次分类端到端的火焰检测算法。

2　R-FCN＋ResNet的火焰检测模型介绍

如图1所示，R-FCN＋ResNet算法模型分为三个部分：第一部分为自动提取特征的深度卷积神经网络，目的是自动提取特征，并将网络最后一层的卷积特征谱图(高×宽为35 pixel×63 pixel，约为原图的1/16)提供给R-FCN；第二部分为R-FCN直接用在特征谱图上检测火焰，并回归出火焰在原图中的位置；第三部分为二次分类器，根据R-FCN给出的火焰位置，将原图中的目标截取下来并调整成224 pixel×224 pixel的图片，再对该目标进行二次分类，以进一步降低误报率。

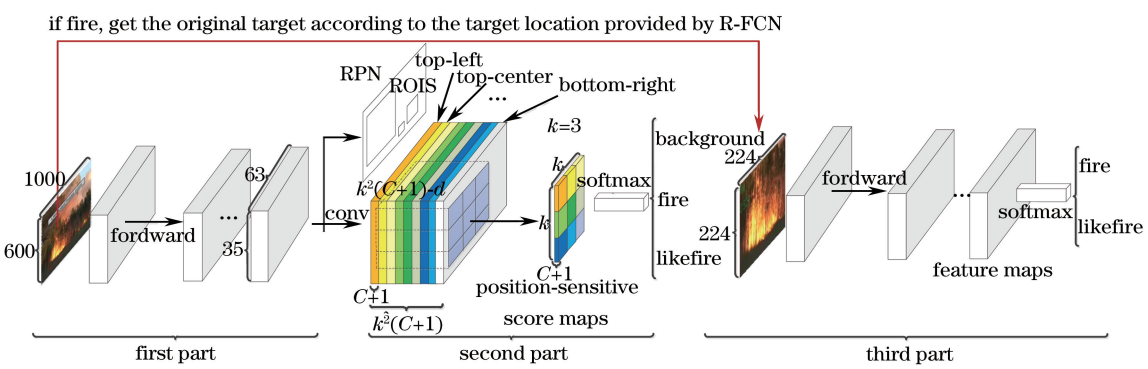

图1　火焰检测算法模型示意图

Fig. 1　Flame detection algorithm model diagram

2.1　特征提取网络模型结构

本文算法的特征提取网络为ResNet。其图像输入尺寸为600 pixel×1000 pixel，图像通道数为3，网络层数为50，并且已经在ImageNet数据集上做过预训练。

2.2　R-FCN模型介绍

R-FCN效仿R-CNN[10]，采用流行的物体检测策略，分为两步：1) 由RPN[11]根据特征提取网络提供的特征谱图求出疑似火焰区域集合(ROIS)；2) 对RPN求出的ROIS逐个进行火焰识别(ROI为ROIS中任意一个疑似火焰区域)。

图1中位于特征谱图右边彩色卷积层是对其进行卷积操作后得出的卷积层，并用来生成位置敏感分数图。位于特征谱图右边彩色卷积层在整幅图像上为每类生成k^2个位置敏感分数图，有C类物体外加1个背景，因此其通道数为$k^2(C+1)$。与PASCAL VOC数据集中的20个类别相比，火焰的特征比较单一，若k值太大，则位置敏感分数图的检测速度必然会下降，但是识别精度未必有明显的提高，反而可能造成过拟合。所以，本文算法模型针对火焰检测设置$k=3$。检测类别分别是疑似火焰目标和火焰目标，其类别数$C=2$。

如图1所示，RPN给出感兴趣的ROIS后，对每个ROI做位置敏感池化得出通道数为$k^2(C+1)$位置敏感分数图。当$k=3$时，ROI可分割成9个矩形块，则对应的位置敏感分数图也就分割成9块，

分别为{top-left，top-center，top-right，…，bottom-right}。再对位置敏感分数图做均值池化，从而得出一个长度为 $C+1$ 的向量，并对向量进行 softmax 分类以得出该 ROI 的类别。

位置敏感 ROI 池化具体操作是将 1 个尺寸为 $w\times h$ 的 ROI 分割成 k^2 个大小为 $w/k\times h/k$ 的矩形方块，并使用 (i,j) 表示 ROI 和位置敏感分数图中每个分块的位置，且 (i,j) 的取值范围为 $(0\leqslant i\leqslant k-1, 0\leqslant j\leqslant k-1)$。再将图 1 中 top-left 箭头指向的通道数为 k^2 的特征图对应的 ROI 中第(0,0)的方块放到位置敏感分数图的第(0,0)；top-center 箭头指向的通道数为 k^2 的特征图对应的 ROI 中第(1,0)的方块放到位置敏感分数图的第(1,0)，以此类推，最后将 bottom-right 箭头指向的通道数为 k^2 的特征图对应的 ROI 中第(2,2)的方块放到位置敏感分数图的第(2,2)，得到完整的位置敏感分数图。得到位置敏感分数图中第 (i,j) 个矩形方块的具体操作为

$$r_c(i,j\mid\theta)=\frac{1}{n}\sum_{(x,y)\in \mathrm{bin}(i,j)} Z_{i,j,c}(x+x_0,y+y_0\mid\theta),\tag{1}$$

式中：$r_c(i,j\mid\theta)$ 为第 c 类的位置敏感分数图 k^2 个分块中的第 (i,j) 个分块；$z_{i,j,c}$ 为特征谱图右边彩色卷积层的一个特征图；$\mathrm{bin}(i,j)$ 为 ROI 和位置敏感分数图中的位置集合；(x_0,y_0) 为 ROI 的左上角坐标；(x,y) 为在 ROI 中以 (x_0,y_0) 为原点时每个元素的坐标值取值范围，其中 $i(w/k)\leqslant x\leqslant(i+1)w/k$，$j(h/k)\leqslant y\leqslant(j+1)h/k$；$n$ 为第 (i,j) 个块里的像素总数；θ 为网络参数。

对位置敏感分数图生成的长度为 $C+1$ 的向量均值池化操作为

$$r_c(\theta)=\frac{1}{k^2}\sum_{i,j} r_c(i,j\mid\theta)。\tag{2}$$

对向量做柔性最大值传输函数分类操作为

$$s_c(\theta)=\exp[r_c(\theta)]/\sum_{c=0}^{C}\exp[r_c(\theta)]。\tag{3}$$

R-FCN 选用的训练方式是梯度下降，代价函数为每个 ROI 的损失函数交叉熵损失与边界框回归损失的和，可表示为

$$\begin{cases} L(s,t_{x,y,w,h})=L_{\mathrm{cls}}(s_c^*)+\lambda[c^*>0]L_{\mathrm{reg}}(t,t^*) \\ L_{\mathrm{cls}}(s_c^*)= \\ -\log\left\{\exp[r_c(\theta)]/\sum_{c=0}^{C}\exp[r_c(\theta)]\right\} \\ L_{\mathrm{reg}}(t,t^*)=R(t-t^*) \end{cases},\tag{4}$$

式中：λ 为 1；c^* 为 ROI 的类标（$c^*=0$ 时，表示这个 ROI 是背景类）；L_{reg} 中的 $t^*=(t_x^*,t_y^*,t_w^*,t_h^*)$ 为真实的回归框的参数，$t=(t_x,t_y,t_w,t_h)$ 为 RPN 得出的回归框的参数，并且 t^* 和 t 与更快 R-CNN[11] 中的边界回归框的参数相同。损失函数为[11]

$$\mathrm{Smooth}_{L1}(x)=\begin{cases}0.5x^2, & \text{if}\,|x|<1\\ |x|, & \text{otherwise}\end{cases}。\tag{5}$$

2.3　二次分类模型描述

图 1 中第三部分为二次分类器，并且已经在 ImageNet 数据集上做过预训练，再用火焰数据对其训练后的残差网络 ResNet。其图像输入尺寸为 224 pixel×224 pixel，图像通道数为 3，网络层数为 50。

3　训练数据库建立

3.1　火焰区域全卷积网络训练数据库

由于 R-FCN 训练需要 PASCAL VOC[12] 格式的数据集，但是目前公开的 PASCAL VOC 的训练数据集的 20 类中没有火焰这一类。为了满足实验需求，通过网络和拍摄等途径收集相关视频，并将视频转换成图片格式，再根据 PASCAL VOC 的格式手工标注图片中的目标，训练数据如表 1 所示。

表 1　训练数据描述

Table 1　Training data description

Label	Number of pictures	Target number
Likefire	953	1000
Fire	7396	9790
Total	7861	10790

所建数据集共 21 个场景，7861 张图片。将原图中，红色不规则物体和灯光等目标标注为 likefire，火焰目标标注为 fire，未标注的区域默认为背景，表 1 中未描述背景类标；图片个数为各个类标所在图片的数量，目标个数为 7861 张图片中所含该类标的个数。

3.2　二次分类器 ResNet 火焰类别训练数据库

为了训练二次分类器，将用来训练 R-FCN 的数据中类标为 likefire 和 fire 的目标根据标注信息截取下来。并将其类标分别定义为 likefire、fire，其数目分别为 1000、9790。但是由于 likefire 和 fire 两个类别的数目不均衡，所以将 likefire 类别复制 9 份，最终 likefire、fire 的数目分别为 9000 和 9790。

4　实验结果以及分析

为了验证算法的有效性，将与传统的基于图像的典型火焰检测算法进行对比实验。本文实验软件

工具为 Windows 下 Caffe[13]框架，训练和测试机器基本配置：GPU 为 Titan X，CPU 为 Inter Core i7，内存为 8 G。

4.1　实验数据介绍

实验视频来源于 Bilkent 大学的公开火灾火焰视频库（http://Signal.ee.bilkent.edu.tr/VisiFire）分辨率为 320 pixel×240 pixel，共 6 段视频，并且此视频不在训练集中。图 2 为 6 段测试视频的截图，其中视频 1 是近处的森林大火和浓烟；视频 2 是远处的森林大火，浓烟滚滚；视频 3 是燃烧的火焰，以及背景颜色和火焰的颜色；视频 4 是人们在森林里四处走动，在森林中躲避火焰；视频 5 是夜晚车子经过一根高梁，横梁与火焰颜色相似；视频 6 是汽车穿过明亮灯光的隧道。

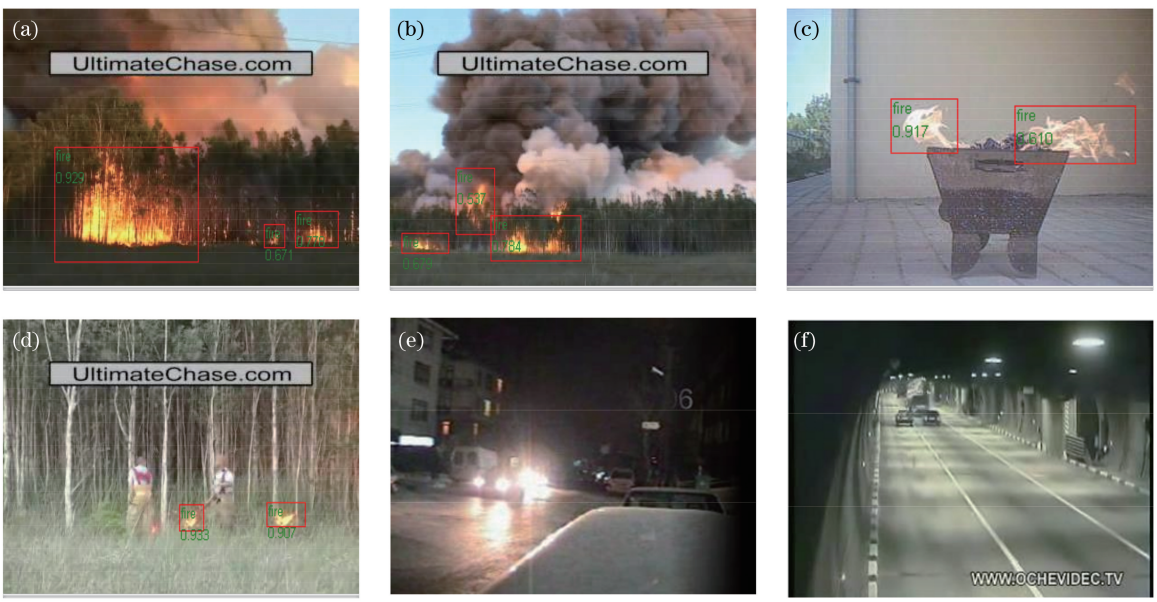

图 2　视频示意图。(a)视频 1；(b)视频 2；(c)视频 3；(d)视频 4；(e)视频 5；(f)视频 6

Fig. 2　Video schematic. (a) Video 1; (b) video 2; (c) video 3; (d) video 4; (e) video 5; (f) video 6

4.2　模型训练

R-FCN 的训练集为实验建立的 PASCAL VOC 格式的火焰目标数据集，类标分别为 likefire 和 fire，共 7861 张图片，类标个数分别为 1000 和 9790。

训练二次分类器 ResNet 的训练集是从训练 R-FCN 数据中根据标注信息在原图中截取下来的图片。类标分别为 likefire、fire。但是由于两个类别数目相差过大，故复制 likefire 将其数量扩充为原来的 9 倍，再用来训练二次分类器 ResNet。

4.3　实验结果对比

表 2 和表 3 分别给出了本文算法与其他算法在 4 段火焰视频和 2 段非火焰视频下的性能对比。其中，TP 和 FP 分别为火焰视频中包含火焰帧数的检测率和漏报率；TN 和 FN 分别为非火焰视频中不包含火焰帧数的正确率和误报率。

表 2　火焰视频检测结果

Table 2　Detection results of flame video

Video	Total Frame	The Flame	Ref. [14]		Ref. [15]		Ref. [16]		R-FCN		R-FCN+ResNet	
			TP /%	FP /%	TP /%	FP /%	TP /%	FP /%	TP /%	FP /%	TP /%	FP /%
1	200	171	85.5	14.5	91.0	9.0	93.5	6.5	100	0	99.60	0.4
2	216	164	75.9	24.1	91.2	8.8	91.2	8.8	100	0	100	0
3	439	420	95.7	4.3	96.1	3.9	88.6	11.4	93.40	6.6	93.40	6.6
4	170	148	87.1	12.9	76.5	23.5	91.2	8.8	100	0	100	0
Mean	256.25	225.75	86.05	13.95	88.7	11.3	91.13	8.87	98.35	1.65	98.25	1.75

文献[14-16]都是传统基于图像的火焰识别算法，文献[14]利用了疑似火焰区域的周长变换、面积变换、质心位移、圆形度以及圆形度变化等形状变化的特征，这种算法火焰形状变换剧烈时效果比较突出，所以在视频 3 中检测效果比较好。但火焰稳定燃烧时火焰形状变化并不太剧烈，如视频 1、2、4，所

以文献[14]的方法表现不是很理想。文献[15]的方法是利用傅里叶描述子提取轮廓特征，但是若出现遮挡情况，其不能完整提取检测物体的轮廓特征而造成漏报，所以该方法在视频 4 中表现不是非常理想。而文献[16]虽然融合了圆形度、矩阵度和重心高度系数等形状结构特征，但是火焰的形状不可控，如果所设计的特征主要依赖于火焰的形状，漏检率和误检率也不稳定的。如视频 3 中，由于火焰受到风的影响而形状变化剧烈，圆形度、矩阵度和重心高度系数不稳定，所以文献[16]在视频 3 中效果不理想。

表 3　非火焰视频检测结果

Table 3　Detection results of non-flame video

Video	Total Frame	The Flame	Ref. [14]		Ref. [15]		Ref. [16]		R-FCN		R-FCN+ResNet	
			TN /%	FN /%	TN /%	FN /%	TN /%	FN /%	TN /%	FN /%	TN /%	FN /%
5	144	0	44.4	55.6	60.4	39.6	75.7	24.3	94.3	5.7	95.9	4.1
6	120	0	69.2	30.8	87.5	12.5	95.8	4.2	95.6	4.4	96.7	3.3
Mean	132	0	56.8	43.2	73.95	26.05	85.75	14.25	94.95	5.05	96.3	3.7

文献[14-16]都是通过手动设计火焰特征，并使用模式识别或者增加阈值限定做分类或者定位，但是这类方法的一个通病就是在一些特定场景的效果比较理想，但是在另一些特殊和复杂场景，其效果就不理想。这是因为火焰的形状、颜色等特征不固定，易受到环境的干扰，所以手动设计的特征并不能完全抓取不同场景下的所有火焰，必然会在某些场景下造成大量漏报。而手动设计特征的不唯一性，造成在复杂场景下很多非火焰物体的形态满足了手动设计的特征条件，必然造成大量误报。

图 3 是特征提取网络在视频 3 的某一帧中提取出的特征谱图可视化后的示意图。可以看出特征提取网络有效地抑制了背景，所以火焰以外的区域呈现黑色；火焰区域则得到增强而呈现白色。R-FCN 是基于特征谱图进行的检测，当火焰和场景剧烈变换时，特征提取网络总是能抑制火焰以外的区域，并增强火焰区域，所以 R-FCN 可以适应于各种场景，并能检测出火焰目标。

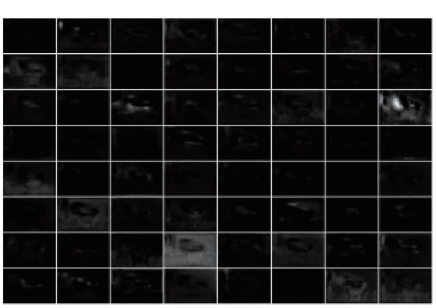

图 3　特征示意图

Fig. 3　Features schematic

但是 R-FCN 是基于特征谱图检测的，而特征谱图是由特征提取网络层提取出来，所以相对于原图必然会存在一些信息的丢失。故利用 R-FCN 对目标区域分类，不如直接使用残差网络在原图中进行分类更为有效。为了解决这个问题，本文算法在 R-FCN 基础上再次加入一个分类器进行二次分类，以降低误报率。具体做法是当 R-FCN 检测出火焰目标，并返回该目标在原图中的位置后，再将该位置的原图火焰截取下来送入二次分类器 ResNet 进行二次分类。由表 2、3 可以看出，当 R-FCN 加上二次分类器 ResNet 后虽然视频 1 平均漏报率上升至 0.1%，但是视频 5、6 平均误报率相对于没有二次分类器 ResNet 的模型下降了 1.35%。

为了验证 R-FCN+ResNet 算法在火焰检测应用领域的时效性，采用每秒帧数指标来衡量。实验结果可知，R-FCN+ResNet 算法与 R-FCN 算法的每秒帧数指标分别为 12 和 12.2，其数值基本持平，因此，该算法在确保低误报率的同时，提高了在火焰检测应用领域的实用性。

5　结　　论

通过深度学习 R-FCN 检测结合 ResNet 二次分类的方法，训练特征提取网络自动提取样本中火焰的特征特征谱图，在特征谱图上使用 R-FCN 检测出火焰在原图的位置，再将原图中火焰区域截取下来送入 ResNet 中进行二次分类以进一步降低误报率。实现了特征提取和火焰目标检测识别端到端的过程，避免了由于手工设计提取火焰特征而造成特征单一，及在特定情况和场景下无法有效检测并识别火焰的问题，确保算法模型的稳健性和可靠性。

参　考　文　献

[1]　Xu Y. The structure of automatic fire alarm system based on virtual instrument[D]. Tianjin: Tianjin University of Technology, 2013.
徐炀. 智能火灾自动报警系统的构建[D]. 天津: 天津理工大学, 2013.

[2]　Hu X J. Research and product development of MIR

flame detector system [D]. Hangzhou: Zhejiang University, 2013.
胡幸江. 多波段红外火焰探测器系统研究与产品开发[D]. 杭州：浙江大学，2013.

[3] Zhang J H, Zhuang J, Du H F, *et al*. Flame detection algorithm based on video multi-feature fusion[J]. Journal of Xi′an Jiaotong University, 2006, 40(7): 811-814.
张进华，庄健，杜海峰，等. 一种基于视频多特征融合的火焰识别算法[J]. 西安交通大学学报，2006，40(7)：811-814.

[4] Rong J Z, Yao W, Gao W, *et al*. Fire video detection method based on muti-features fusion[J]. Journal of Combustion Science and Technology, 2013, 19(3): 227-233.
荣建忠，姚卫，高伟，等. 基于多特征融合技术的火焰视频探测方法[J]. 燃烧科学与技术，2013，19(3)：227-233.

[5] Chen T H, Wu P H, Chiou Y C. An early fire-detection method based on image processing [C]. International Conference on Image Processing, 2004, 3: 1707-1710.

[6] Li W H, Wang Y, Zhao R L, *et al*. Fire detection with day-night infrared camera[J]. Journal of Hunan University (Natural Science), 2012, 39(8): 73-80.
李文辉，王莹，赵瑞霖，等. 使用日夜两用型红外摄像机进行火焰检测[J]. 湖南大学学报(自然科学版)，2012，39(8)：73-80.

[7] Yan Y Y, Tang Y Y, Guo Z B, *et al*. Fusion of flame color and its contour for fire detection [J]. Microelectronics & Computer, 2011, 28(10): 137-141.
严云洋，唐岩岩，郭志波，等. 融合色彩和轮廓特征的火焰检测[J]. 微电子学与计算机，2011，28(10)：137-141.

[8] Dai J F, Li Y, He K M, *et al*. R-FCN: Object detection via region-based fully convolutional networks [C]. Advances in Neural Information Processing Systems, 2016: 379-387.

[9] He K M, Zhang X Y, Ren S Q, *et al*. Deep residual learning for image recognition[C]// Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, 2016: 770-778.

[10] Girshick R, Donahue J, Darrell T, *et al*. Region-based convolutional networks for accurate object detection and segmentation[J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2016, 38(1): 142-158.

[11] 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 & Machine Intelligence, 2015, 39(6): 1137-1149.

[12] Everingham M, Winn J. The PASCALvisual object classes challenge 2010 development kit contents[C]. International Conference on Machine Learning Challenges: Evaluating Predictive Uncertainty Visual Object Classification, 2011, 111(1): 117-176.

[13] Jia Y Q, Shelhamer E, Donahue J, *et al*. Caffe: Convolutional architecture for fast feature embedding [C]// Proceedings of the 22nd ACM International Conference on Multimedia, 2014: 675-678.

[14] Yang X, Wang J, He S, *et al*. A SVM approach for vessel fire detection based on image processing[C]// Proceedings of International Conference on Modelling, Identification & Control, 2012: 150-153.

[15] Zhang X, Xu F, Song Z, *et al*. Video flame detection algorithm based on multi-feature fusion technique [C]. 24th Chinese Control and Decision Conference, 2012: 4291-4294.

[16] Wu X Y, Yan Y Y, Du J, *et al*. Fire detection based on fusion of multiple features[J]. Transactions on Intelligent Systems, 2015, 10(2): 240-247.
吴茜茵，严云洋，杜静，等. 多特征融合的火焰检测算法[J]. 智能系统学报，2015，10(2)：240-247.