



(12) 发明专利

(10) 授权公告号 CN 113393521 B

(45) 授权公告日 2023.05.05

(21) 申请号 202110545613.3

G06V 20/52 (2022.01)

(22) 申请日 2021.05.19

G06V 10/44 (2022.01)

(65) 同一申请的已公布的文献号

G06V 10/764 (2022.01)

申请公布号 CN 113393521 A

G06V 10/82 (2022.01)

(43) 申请公布日 2021.09.14

G06N 3/0464 (2023.01)

(73) 专利权人 中国科学院声学研究所南海研究  
站

G06N 3/048 (2023.01)

地址 570105 海南省海口市龙华区滨海大  
道63号

G06N 3/08 (2023.01)

(56) 对比文件

CN 110210485 A, 2019.09.06

(72) 发明人 李松斌 晏黔东 刘鹏 张遥

审查员 魏兰

(74) 专利代理机构 北京方安思达知识产权代理  
有限公司 11472

专利代理师 陈琳琳 杨青

(51) Int. Cl.

G06T 7/73 (2017.01)

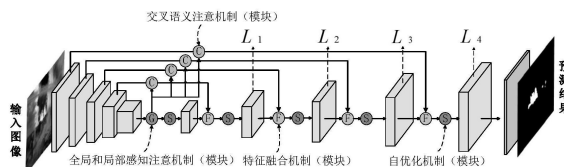
权利要求书4页 说明书10页 附图2页

(54) 发明名称

一种基于双语义注意力机制的高精度火焰  
定位方法及系统

(57) 摘要

本发明公开了一种基于双语义注意力机制的高精度火焰定位方法及系统,该方法包括:通过编码器对输入图像进行多层次特征提取,得到对应的空间细节和高级上下文特征。解码器用于逐步聚合高级上下文特征和低级空间细节,并生成火焰分割图。在解码操作之前,对编码器得到的层级特征进行全局和局部感知注意机制处理,来捕获更完整的高级语义和全局上下文信息,之后采用具有双线性插值的四个交叉语义注意机制获取优化后的局部空间细节信息,采用特征融合策略来聚合高级上下文特征和低级的空间细节特征,并通过自优化机制来捕获每个解码层内部特征图之间的异同并逐步生成火焰分割图,实现对火焰像素级别的精确定位。



1. 一种基于双语义注意力机制的高精度火焰定位方法,所述方法包括:

步骤1) 数据预处理,将训练所需图片按每组a张随机分成m组;

步骤2) 随机读取一组图片,由具有n层结构的编码器对输入图像进行多层次特征提取,得到每一层对应的层级特征 $f_i, i \in [1, n]$ ;

步骤3) 将步骤2) 中最后一层的层级特征 $f_n$ ,进行全局和局部感知注意机制处理,捕获更丰富的高级语义和全局上下文信息,得到全局及局部语义融合特征 $f'_n$ ;

步骤4) 将编码器第i层输出的所述层级特征 $f_i, i \in [1, n-1]$ ,与所述全局及局部语义融合特征 $f'_n$ 进行交叉语义注意机制处理,优化局部空间细节信息,得到第i层的局部空间细节优化特征 $f_i^{cc}$ ;

步骤5) 将编码器第n-i层的局部空间细节优化特征 $f_{n-i}^{cc}$ 与对应的第i层解码器的输出 $f_{i-d}$ 进行融合处理后,得到解码器每层的输出特征 $\tilde{o}_i, i \in [1, n-1]$ ,对 $\tilde{o}_i$ 进行自优化机制处理,得到自优化输出结果 $o_i$ ,其中,所述解码器第1层的输入为所述全局及局部语义融合特征 $f'_n$ 经自优化机制处理后的输出 $f_{in-d}$ ;

步骤6) 对所述解码器最后一层的输出 $f_{n-d}$ ,进行上采样,得到与原始输入图像相同分辨率的输出特征;

步骤7) 对输出特征进行特征映射,并采用梯度下降法对模型中的参数进行更新;反复迭代,直至训练出最优参数组合;

步骤8) 基于最优参数模型实现火焰定位;

所述步骤5) 具体包括:

步骤5-1) 将编码器部分得到的所述全局及局部语义融合特征 $f'_n$ 经自优化处理后得到特征 $f_{in-d}$ ,作为解码器第一层的输入;

步骤5-2) 将 $f_{in-d}$ 经解码器第一层解码后的输出 $f_{1-d}$ 与编码器第n-1层得到的局部空间细节优化特征 $f_{n-1}^{cc}$ 进行特征融合机制处理,得到特征 $\tilde{o}_1$ ,将 $\tilde{o}_1$ 进行自优化机制处理,得到第一层的自优化输出结果 $o_1$ ,作为解码器第二层的输入;

步骤5-3) 将解码器前一层的自优化输出结果作为后一层的输入,执行与步骤5-2) 相同的计算。

2. 根据权利要求1所述的方法,其特征在于,所述全局和局部感知注意机制包括全局上下文分支和局部上下文分支;

所述步骤3) 具体包括:

步骤3-1) 利用卷积层对所述层级特征 $f_n$ 进行特征压缩,得到密集特征 $\tilde{f}_n$ ,所述卷积层的卷积核大小为 $3 \times 3$ ,步长为1,通道数为256;

步骤3-2) 所述全局上下文分支包括自适应最大池化层和卷积核大小为 $1 \times 1$ 的卷积层,由所述自适应最大池化层将 $\tilde{f}_n$ 转换为语义向量 $v_s, v_s$ 与 $\tilde{f}_n$ 进行通道乘法及卷积运算,得到全局上下文信息 $f_n^{gc}$ ,计算公式如下:

$$v_s = gp(\tilde{f}_n)$$

$$f_n^{gc} = \delta(conv(v_s \odot \tilde{f}_n))$$

其中，“gp”表示自适应最大池化处理，“ $\odot$ ”表示通道乘法，“conv”表示卷积操作，“ $\delta$ ”表示ReLU激活函数；

步骤3-3)所述局部上下文分支由n个卷积核大小为3\*3的卷积层堆叠而成,由所述局部上下文分支将 $\tilde{f}_n$ 转换为局部上下文权重 $w_{lc}$ ,计算公式如下:

$$w_{lc} = \delta(\text{stacked\_conv}(\tilde{f}_n))$$

其中,stacked\_conv表示堆叠的卷积层;

步骤3-4)利用 $w_{lc}$ 对 $f_n^{gc}$ 进行特征增强,得到增强后的特征 $f_{glc}$ ,将 $f_{glc}$ 进行卷积核大小为1\*1的卷积运算,进行特征压缩,同时根据 $\tilde{f}_n$ 生成局部偏差b,由卷积运算的输出结果与局部偏差b计算得到所述全局及局部语义融合特征 $f'_n$ ,计算公式如下:

$$f_{glc} = w_{lc} \odot f_n^{gc}$$

$$f'_n = \delta(\text{conv}(f_{glc})) + b$$

其中，“ $\odot$ ”表示点乘。

3.根据权利要求1所述的方法,其特征不在于,所述步骤4)具体包括:

步骤4-1)将所述编码器第i层的输出 $f_i$ 进行卷积运算得到特征 $f_{i-256}$ ,卷积核大小为1\*1,通道数为256;

步骤4-2)将 $f'_n$ 进行卷积运算,卷积核大小为1\*1,通道数为256,之后进行双线性插值上采样计算,得到特征 $f'_{n-256}$ ;

步骤4-3)对 $f_{i-256}$ 和 $f'_{n-256}$ 按照通道维度进行级联,得到能够表示空间细节及全局上下文的特征 $f_{i-cc}$ ,计算公式如下:

$$f_{i-cc} = [\text{conv}(f_i), u(\text{conv}(f'_n))]$$

其中,“u”表示线性插值上采样,“[]”表示级联;

步骤4-3)对 $f_{i-cc}$ 进行卷积运算,卷积核大小为1\*1,通道数为256,将卷积运算结果进行有效通道注意机制处理,将所得结果与 $f_{i-256}$ 进行点乘运算,得到第i层的局部空间细节优化特征 $f_i^{cc}$ ,计算公式如下:

$$f_i^{cc} = f_{i-256} \odot \text{ECA}(f_{i-cc})$$

其中,ECA表示通道注意机制,“ $\odot$ ”表示点乘。

4.根据权利要求1所述的方法,其特征不在于,所述自优化处理机制,包括:

步骤5-2-1)对解码器第i层的输出 $\tilde{o}_i$ 经过两个卷积运算,分别得到位置权重 $W_i$ 和位置偏置 $b_i$ ;其中两个卷积运算的卷积核大小均为3\*3,通道数为256;

步骤5-2-2)利用位置权重 $W_i$ 和位置偏置 $b_i$ 对 $\tilde{o}_i$ 进行自优化,得到输出结果 $o_i$ ,计算公式如下:

$$W_i = \text{conv}_w(\tilde{o}_i)$$

$$b_i = \text{conv}_b(\tilde{o}_i)$$

$$o_i = \delta(W_i \odot \tilde{o}_i + b_i)$$

其中,“ $\odot$ ”表示点乘。

5.根据权利要求1所述的方法,其特征不在于,所述编码器包括:多尺度特征提取、隐式深

度监督和通道注意力机制三个部分；

多尺度特征提取计算过程如下：

$$B=M(A)=[h_{1*1}(A),h_{3*3}(A),h_{5*5}(A),h_{\text{pooling}}(A)]$$

其中,A为输入图像, $A \in \mathbb{R}^{C*H*W}$ ,C表示通道数,H为图像的高度,W为图像的宽度,[ ]表示级联, $h_{n*n}$ 表示卷积核大小为n\*n的卷积,每个卷积之后都进行批归一化和激活函数处理, $h_{\text{pooling}}$ 表示3\*3最大池化；

隐式深度监督处理的输出特征设置为256；

通道注意力机制的计算过程如下：

$$x' = gp(x) = \left( \sum_{i,j=0}^{H-1,W-1} x(i,j) \right) / (H * W)$$

$$x_{lb} = \sigma(w_2 \otimes \delta(w_1 \otimes x'))$$

$$f_i = x' * x_{lb}$$

其中,gp表示全局平均池化,x(i,j)表示x的特征值, $w_1$ 和 $w_2$ 表示权重矩阵, $\delta$ 表示ReLU激活, $\sigma$ 表示Sigmoid激活, $\otimes$ 表示卷积运算, $x_{lb}$ 表示可学习参数。

6.根据权利要求1所述的方法,其特征在于,所述特征映射包括1个全连接层和1个Sigmoid函数;所述全连接层包括两个神经元,实现高层抽象特征向输出层输出结果的特征映射;采用Sigmoid归一化函数进行归一化处理,得到图像的像素级分类结果,表示为:

$$p_i = \frac{e^i}{\sum_j e^j}$$

其中i表示第i个得分值, $\sum_j e^j$ 表示所有特征值的总和;选取概率最大值对应的类别作为像素级火焰检测的判别结果。

7.根据权利要求1所述的方法,其特征在于,所述步骤7)具体包括:

步骤7-1)将模型输出的结果与真实标签进行误差求取;所述误差求取采用交叉熵,其可以表示为:

$$Loss(y, \bar{y}) = H(p, q) = - \sum_{i=1}^n p(x_i) \log q(x_i)$$

其中, $Loss(y, \bar{y})$ 表示真实标签 $\bar{y}$ 与预测结果y之间的误差, $p(x_i)$ 表示模型的前向输出结果, $q(x_i)$ 表示真实标签, $\Sigma$ 表示求和;

步骤7-2)采用步骤7-1)得到的参数作为本次迭代的权重值;从剩余的图片中随机选取一组图片,经步骤2)、步骤3)、步骤4)、步骤5)、步骤6)和步骤7-1),得到新的参数组合;反复迭代,直至完成一个迭代周期;

步骤7-3)对训练图片进行重新洗牌,转至步骤1);反复执行,直至训练出最优参数组合。

8.根据权利要求1所述的方法,其特征在于,所述步骤8)具体包括:

步骤8-1)对于待检测的图片,利用双线性插值将待测图片的尺寸调整至模型输入所需大小 $224 \times 224 \times 3$ ;

步骤8-2)将调整后的图片输入至最优参数模型,经模型前向传输得到预测结果;

步骤8-3)将每一像素的预测结果最大概率值所对应的类别与真实标签进行比较,若一

致,则预测正确,反之,则预测错误;

步骤8-4) 执行步骤8-1)、步骤8-2) 和步骤8-3),直到全部待测图片检测完毕。

9. 一种基于双语义注意力机制的高精度火焰定位系统,根据权利要求1-8任一项所述的方法构建的双重情境注意网络模型实现,所述系统包括:

数据预处理模块,用于对输入图像进行尺寸调整;

编码器,用于对输入图像进行多层次特征的提取;

全局和局部感知注意模块,用于对输入特征进行全局和局部感知注意机制处理,捕获更丰富的高级语义和全局上下文信息,得到全局及局部语义融合特征;

交叉语义注意模块,用于对输入特征进行交叉语义注意机制处理,优化局部空间细节信息,得到局部空间细节优化特征;

特征融合模块,用于对输入特征进行融合,实现多层次特征的融合;

自优化模块,用于进行自优化机制处理,优化融合特征;

解码器,用于聚合高级上下文特征和低级空间细节特征。

## 一种基于双语义注意力机制的高精度火焰定位方法及系统

### 技术领域

[0001] 本发明涉及机器视觉及深度学习技术领域,特别涉及一种基于双语义注意力机制的高精度火焰定位方法及系统。

### 背景技术

[0002] 火灾作为一个高度频发、破坏性极大的自然灾害,严重威胁着人们的生命财产安全,大规模的火灾还会给生态平衡造成巨大的创伤。如2019年以来亚马孙森林火灾数量超过7.2万起,火灾频发严重影响着周边居民的生产生活。2109年澳大利亚大火事件对全球经济和生态环境造成重大影响,截至2020年7月28日,澳大利亚丛林大火或已致30亿动物死亡。2019年4月,法国巴黎圣母院发生火灾,其标志性尖顶被烧断,坍塌倒下,整座建筑损毁严重。因此,为避免火灾大规模蔓延造成的严重损失,及时准确的火灾探测和定位至关重要。

[0003] 随着科学技术的进步,现有的火灾探测方法可以分为传统的接触式传感器和计算机视觉两大类。前者包括烟雾传感器,颗粒传感器和温度传感器等。受部署密度的限制,基于接触传感器的方法无法覆盖较大的区域。基于视觉的火灾探测方法具有覆盖范围广和响应速度快的优点,受到社会各界的广泛关注。在2015年之前,基于视觉的火灾探测主要依靠手工设计的特征,例如颜色、形状、纹理和动态特征等。该类方法的主要缺点手工设计特征非常复杂繁琐,较高的错误检测率以及缺乏鲁棒性。近年来,随着将深度学习技术引入基于视觉的火灾探测中,使得火灾探测技术取得了长足的进步。但是,现有方法均属于图像级别的火灾检测,仅能识别图像中是否有火灾,而无法获得火灾的方向、位置和范围大小等信息。

### 发明内容

[0004] 本发明的目的在于克服现有火灾检测方法存在的技术缺陷,提出了一种基于双语义注意力机制的高精度火焰定位方法及系统,该方法能够进行像素级别的火焰检测,实现对火焰的定位,大大提高了火灾的检测精度,也更适用于复杂情况下的火灾检测。

[0005] 为了实现上述目的,本发明提出了一种基于双语义注意力机制的高精度火焰定位方法;所述方法包括:

[0006] 步骤1) 数据预处理,将训练所需图片按每组a张随机分成m组;

[0007] 步骤2) 随机读取一组图片,由具有n层结构的编码器对输入图像进行多层次特征提取,得到每一层对应的层级特征 $f_i, i \in [1, n]$ ;

[0008] 步骤3) 将步骤2) 中最后一层的层级特征 $f_n$ ,进行全局和局部感知注意机制处理,捕获更丰富的高级语义和全局上下文信息,得到全局及局部语义融合特征 $f'_n$ ;

[0009] 步骤4) 将编码器第i层输出的所述层级特征 $f_i, i \in [1, n-1]$ ,与所述全局及局部语义融合特征 $f'_n$ 进行交叉语义注意机制处理,优化局部空间细节信息,得到第i层的局部空间细节优化特征 $f_i^{cc}$ ;

[0010] 步骤5) 将编码器第n-i层的局部空间细节优化特征 $f_{n-i}^{cc}$ 与对应的第i层解码器的输出 $f_{i-d}$ 进行特征融合机制处理后,得到解码器每层的输出特征 $\tilde{o}_i, i \in [1, n-1]$ ,对 $\tilde{o}_i$ 进行自优化机制处理,得到自优化输出结果 $o_i$ ,其中,所述解码器第1层的输入为所述全局及局部语义融合特征 $f'_n$ 经自优化机制处理后的输出 $f_{in-d}$ ;

[0011] 步骤6) 对所述解码器最后一层的输出 $f_{n-d}$ ,进行上采样,得到与原始输入图像相同分辨率的输出特征;

[0012] 步骤7) 对输出特征进行特征映射,并采用梯度下降法对模型中的参数进行更新;反复迭代,直至训练出最优参数组合;

[0013] 步骤8) 基于最优参数模型实现火焰定位。

[0014] 进一步的,所述全局和局部感知注意机制包括全局上下文分支和局部上下文分支;所述步骤3) 具体包括:

[0015] 步骤3-1) 利用卷积层对所述层级特征 $f_n$ 进行特征压缩,得到密集特征 $\tilde{f}_n$ ,所述卷积层的卷积核大小为 $3 \times 3$ ,步长为1,通道数为256;

[0016] 步骤3-2) 所述全局上下文分支包括自适应最大池化层和卷积核大小为 $1 \times 1$ 的卷积层,由所述自适应最大池化层将 $\tilde{f}_n$ 转换为语义向量 $v_s, v_s$ 与 $\tilde{f}_n$ 进行通道乘法及卷积运算,得到全局上下文信息 $f_n^{gc}$ ,计算公式如下:

$$[0017] \quad v_s = gp(\tilde{f}_n)$$

$$[0018] \quad f_n^{gc} = \delta(\text{conv}(v_s \odot \tilde{f}_n))$$

[0019] 其中,“gp”表示自适应最大池化处理,“ $\odot$ ”表示通道乘法,“conv”表示卷积操作,“ $\delta$ ”表示ReLU激活函数;

[0020] 步骤3-3) 所述局部上下文分支由n个卷积核大小为 $3 \times 3$ 的卷积层堆叠而成,由所述局部上下文分支将 $\tilde{f}_n$ 转换为局部上下文权重 $w_{lc}$ ,计算公式如下:

$$[0021] \quad w_{lc} = \delta(\text{stacked\_conv}(\tilde{f}_n))$$

[0022] 其中,stacked\_conv表示堆叠的卷积层;

[0023] 步骤3-4) 利用 $w_{lc}$ 对 $f_n^{gc}$ 进行特征增强,得到增强后的特征 $f_{glc}$ ,将 $f_{glc}$ 进行卷积核大小为 $1 \times 1$ 的卷积运算,进行特征压缩,同时根据 $\tilde{f}_n$ 生成局部偏差b,由卷积运算的输出结果与局部偏差b计算得到所述全局及局部语义融合特征 $f'_n$ ,计算公式如下:

$$[0024] \quad f_{glc} = w_{lc} \odot f_n^{gc}$$

$$[0025] \quad f'_n = \delta(\text{conv}(f_{glc})) + b$$

[0026] 其中,“ $\odot$ ”表示点乘。

[0027] 进一步的,上述步骤4) 具体包括:

[0028] 步骤4-1) 将所述编码器第i层的输出 $f_i$ 进行卷积运算得到特征 $f_{i-256}$ ,卷积核大小为 $1 \times 1$ ,通道数为256;

[0029] 步骤4-2) 将 $f'_n$ 进行卷积运算,卷积核大小为 $1 \times 1$ ,通道数为256,之后进行双线性插值上采样计算,得到特征 $f'_{n-256}$ ;

[0030] 步骤4-3) 对 $f_{i-256}$ 和 $f'_{n-256}$ 按照通道维度进行级联,得到能够表示空间细节及全局上下文的特征 $f_{i-cc}$ ,计算公式如下:

$$[0031] \quad f_{i-cc} = [\text{conv}(f_i), u(\text{conv}(f'_n))]$$

[0032] 其中,“u”表示线性插值上采样,“[]”表示级联;

[0033] 步骤4-3) 对 $f_{i-cc}$ 进行卷积运算,卷积核大小为 $1*1$ ,通道数为256,将卷积运算结果进行有效通道注意机制处理,将所得结果与 $f_{i-256}$ 进行点乘运算,得到第i层的局部空间细节优化特征 $f_i^{cc}$ ,计算公式如下:

$$[0034] \quad f_i^{cc} = f_{i-256} \odot ECA(f_{i-cc})$$

[0035] 其中,ECA表示通道注意机制,“ $\odot$ ”表示点乘。

[0036] 进一步的,所述步骤5) 具体包括:

[0037] 步骤5-1) 将编码器部分得到的所述全局及局部语义融合特征 $f'_n$ 经自优化处理后得到特征 $f_{in-d}$ ,作为解码器第一层的输入;

[0038] 步骤5-2) 将 $f_{in-d}$ 经解码器第一层解码后的输出 $f_{1-d}$ 与编码器第n-1层得到的局部空间细节优化特征 $f_{n-1}^{cc}$ 进行特征融合机制处理,得到特征 $\tilde{o}_1$ ,将 $\tilde{o}_1$ 进行自优化机制处理,得到第一层的自优化输出结果 $o_1$ ,作为解码器第二层的输入;

[0039] 步骤5-3) 将编码器前一层的自优化输出结果作为后一层的输入,执行与步骤5-2) 相同的计算。

[0040] 进一步的,所述自优化处理机制,包括:

[0041] 步骤5-2-1) 对解码器第i层的输出 $\tilde{o}_i$ 经过两个卷积运算,分别得到位置权重 $W_i$ 和位置偏置 $b_i$ ;其中两个卷积运算的卷积核大小均为 $3*3$ ,通道数为256;

[0042] 步骤5-2-2) 利用位置权重 $W_i$ 和位置偏置 $b_i$ 对 $\tilde{o}_i$ 进行自优化,得到输出结果 $o_i$ ,计算公式如下:

$$[0043] \quad W_i = \text{conv}_w(\tilde{o}_i)$$

$$[0044] \quad b_i = \text{conv}_b(\tilde{o}_i)$$

$$[0045] \quad o_i = \delta(W_i \odot \tilde{o}_i + b_i)$$

[0046] 进一步的,所述编码器包括:多尺度特征提取、隐式深度监督和通道注意力机制三个部分;

[0047] 多尺度特征提取计算过程如下:

$$[0048] \quad B = M(A) = [h_{1*1}(A), h_{3*3}(A), h_{5*5}(A), h_{\text{pooling}}(A)]$$

[0049] 其中,A为输入图像, $A \in \mathbb{R}^{C*H*W}$ ,C表示通道数,H为图像的高度,W为图像的宽度,[]表示级联, $h_{n*n}$ 表示卷积核大小为 $n*n$ 的卷积,每个卷积之后都进行批归一化和激活函数处理, $h_{\text{pooling}}$ 表示 $3*3$ 最大池化;

[0050] 隐式深度监督处理的输出特征设置为256;

[0051] 通道注意力机制的计算过程如下:

$$[0052] \quad x' = gp(x) = \left( \sum_{i,j=0}^{H-1,W-1} x(i,j) \right) / (H*W)$$



[0053]  $x_{lb} = \sigma(w_2 \otimes \delta(w_1 \otimes x'))$

[0054]  $f_i = x' * x_{lb}$

[0055] 其中, gp表示全局平均池化,  $x(i, j)$  表示x的特征值,  $w_1$ 和 $w_2$ 表示权重矩阵,  $\delta$ 表示ReLU激活,  $\sigma$ 表示Sigmoid激活,  $\otimes$ 表示卷积运算,  $x_{lb}$ 表示可学习参数。

[0056] 进一步的, 所述特征映射包括1个全连接层和1个Sigmoid函数; 所述全连接层包括两个神经元, 实现高层抽象特征向输出层输出结果的特征映射; 采用Sigmoid归一化函数进行归一化处理, 得到图像的像素级分类结果, 表示为:

[0057] 
$$p_i = \frac{e^i}{\sum_j e^j}$$

[0058] 其中i表示第i个得分值,  $\sum_j e^j$ 表示所有特征值的总和; 选取概率最大值对应的类别作为像素级火焰检测的判别结果。

[0059] 进一步的, 所述步骤7) 具体包括:

[0060] 步骤7-1) 将模型输出的结果与真实标签进行误差求取; 所述误差求取采用交叉熵, 其可以表示为:

[0061] 
$$Loss(y, \bar{y}) = H(p, q) = -\sum_{i=1}^n p(x_i) \log q(x_i)$$

[0062] 其中,  $Loss(y, \bar{y})$ 表示真实标签 $\bar{y}$ 与预测结果y之间的误差,  $p(x_i)$ 表示模型的前向输出结果,  $q(x_i)$ 表示真实标签,  $\Sigma$ 表示求和;

[0063] 步骤7-2) 采用步骤7-1) 得到的参数作为本次迭代的权重值; 从剩余的图片中随机选取一组图片, 经步骤2)、步骤3)、步骤4)、步骤5)、步骤6) 和步骤7-1), 得到新的参数组合; 反复迭代, 直至完成一个迭代周期;

[0064] 步骤7-3) 对训练图片进行重新洗牌, 转至步骤1); 反复执行, 直至训练出最优参数组合。

[0065] 进一步的, 所述步骤8) 具体包括:

[0066] 步骤8-1) 对于待检测的图片, 利用双线性插值将待测图片的尺寸调整至模型输入所需大小 $224 \times 224 \times 3$ ;

[0067] 步骤8-2) 将调整后的图片输入至最优参数模型, 经模型前向传输得到预测结果;

[0068] 步骤8-3) 将每一像素的预测结果最大概率值所对应的类别是与真实标签进行比较, 若一致, 则预测正确, 反之, 则预测错误;

[0069] 步骤8-4) 执行步骤8-1)、步骤8-2) 和步骤8-3), 直到全部待测图片检测完毕。

[0070] 本发明还提供了一种基于双语义注意力机制的高精度火焰定位系统, 根据上述任一方法构建的双重情境注意网络模型实现, 所述系统包括:

[0071] 数据预处理模块, 用于对输入图像进行尺寸调整;

[0072] 编码器, 用于对输入图像进行多层次特征的提取;

[0073] 全局和局部感知注意模块, 用于对输入特征进行全局和局部感知注意机制处理, 捕获更丰富的高级语义和全局上下文信息, 得到全局及局部语义融合特征;

[0074] 交叉语义注意模块, 用于对输入特征进行交叉语义注意机制处理, 优化局部空间细节信息, 得到局部空间细节优化特征;

- [0075] 特征融合模块,用于对输入特征进行融合,实现多层次特征的融合;
- [0076] 自优化模块,用于进行自优化机制处理,优化融合特征;
- [0077] 解码器,用于聚合高级上下文特征和低级空间细节特征。
- [0078] 与现有方法相比,本发明的技术优势在于:
- [0079] 本发明提供的火焰定位技术方案,通过编码器对输入图像进行多级别特征提取,得到对应的层级特征。解码器用于逐步聚合高级上下文特征和低级空间细节,并生成火焰分割图。在解码操作之前,对编码器得到的层级特征进行全局和局部感知注意机制处理,来捕获更完整的高级语义和全局上下文信息,之后采用具有双线性插值的四个交叉语义注意机制获取跨上下文的重构特征,还采用了简单的特征融合策略来聚合高级上下文特征和低级的空间细节特征,并通过自优化机制来捕获每个特征图的内部特征逐步生成火焰分割图,使得深度学习模型对火焰具有更强的学习能力,可以实现对火焰像素级别的精准定位。在测试基准数据集时,本发明取得了超过现有方法的性能。

### 附图说明

- [0080] 图1为本发明提供的基于双语义注意力机制的高精度火焰定位网络结构示意图;
- [0081] 图2为本发明提供的全局和局部上下文注意模块结构示意图;
- [0082] 图3为本发明提供的交叉语义注意模块结构示意图;
- [0083] 图4为本发明提供的自优化模块结构示意图。

### 具体实施方式

- [0084] 下面结合附图和具体实施来对本发明作进一步描述。
- [0085] 图1为本发明提供的基于双语义注意力机制的高精度火焰定位网络结构示意图,图1中“G”表示全局和局部感知注意机制(模块)，“C”表示交叉语义注意机制(模块)，“F”表示特征融合机制(模块)，“S”表示自优化机制(模块)， $L_i$ 代表损失值，“U”代表上采样，“+”表示拼接，“\*”表示乘法计算。
- [0086] 如图1所示,本发明提出的火焰定位网络结构由对称的编码器-解码器、全局和局部上下文注意模块G,四个交叉语义注意模块C,四个特征融合模块F和五个自优化模块S构成。其中,图1中编码器与解码器均以 $n=5$ 为例进行展示。该模型的数据处理过程具体包括:
- [0087] 步骤1) 数据预处理,将训练所需图片按每组 $a$ 张随机分成 $m$ 组;
- [0088] 步骤2) 随机读取一组图片,由具有 $n$ 层结构的编码器对输入图像进行多层次特征提取,得到每一层对应的层级特征 $f_i, i \in [1, n]$ ;
- [0089] 步骤3) 将步骤2) 中最后一层的层级特征 $f_n$ ,进行全局和局部感知注意机制处理,捕获更丰富的高级语义和全局上下文信息,得到全局及局部语义融合特征 $f'_n$ ;
- [0090] 步骤4) 将编码器第 $i$ 层输出的所述层级特征 $f_i, i \in [1, n-1]$ ,与所述全局及局部语义融合特征 $f'_n$ 进行交叉语义注意机制处理,优化局部空间细节信息,得到第 $i$ 层的局部空间细节优化特征 $f_i^{cc}$ ;
- [0091] 步骤5) 将编码器第 $n-i$ 层的局部空间细节优化特征 $f_{n-i}^{cc}$ 与对应的第 $i$ 层解码器的输出 $f_{i-d}$ 进行特征融合机制处理后,得到解码器每层的输出特征 $\tilde{o}_i, i \in [1, n-1]$ ,对 $\tilde{o}_i$ 进行自

优化机制处理,得到自优化输出结果 $o_i$ ,其中,所述解码器第1层的输入为所述全局及局部语义融合特征 $f'_n$ 经自优化机制处理后的输出 $f_{in-d}$ ;

[0092] 步骤6)对所述解码器最后一层的输出 $f_{n-d}$ ,进行上采样,得到与原始输入图像相同分辨率的输出特征;

[0093] 步骤7)对输出特征进行特征映射,并采用梯度下降法对模型中的参数进行更新;反复迭代,直至训练出最优参数组合;

[0094] 步骤8)基于最优参数模型实现火焰定位。

[0095] 通过将解码器最后一层的输出 $f_{n-d}$ 上采样到与原始输入图像相同分辨率,则得到了与原始输入图像每个像素点一一对应的输出特征图像素点,通过后续对输出特征图上各像素点是否为火焰的判定,可以得到输入图像中每个像素点是否为火焰的判定结果,从而以像素点为单位,获取到火焰的形状、出现的位置以及火焰区域等信息。

[0096] 本发明设计的火焰定位方法,通过编码器对输入图像进行多层次特征提取,得到对应的空间细节和高层上下文特征。解码器用于逐步聚合高级上下文特征和低级空间细节,并生成火焰分割图。在解码操作之前,对编码器得到的层级特征进行全局和局部感知注意机制处理,来捕获更完整的高级语义和全局上下文信息,之后采用具有双线性插值的四个交叉语义注意机制获取跨上下文的重构特征,还采用特征融合策略来聚合高级上下文特征和低级的空间细节特征,并通过自优化机制来捕获每个解码层内部特征图之间的异同并逐步生成火焰分割图,实现对火焰像素级别的精准定位。

[0097] 其中,采用特征融合机制对编码器第 $n-i$ 层的局部空间细节优化特征 $f_{n-i}^{cc}$ 与对应的第 $i$ 层解码器的输出 $f_{i-d}$ 进行拼接,之后采用卷积核大小为 $1*1$ 的卷积层将拼接后的特征图压缩为256通道,得到解码器每层的输出特征 $\tilde{o}_i$ 。

[0098] 本发明提出了一种新的编码器,可以增强模型对火焰特征的提取。编码器主要由三部分构成:多尺度特征提取模块,隐式深度监督和通道注意力机制。多尺度特征提取计算过程如下:

$$[0099] \quad B = M(A) = [h_{1*1}(A), h_{3*3}(A), h_{5*5}(A), h_{pooling}(A)]$$

[0100] 其中, $A$ 为输入图像, $A \in \mathbb{R}^{C*H*W}$ , $C$ 表示通道数, $H$ 为图像的高度, $W$ 为图像的宽度, $[\ ]$ 表示级联, $h_{n*n}$ 表示 $n*n$ 卷积,每个卷积之后都进行批归一化和激活函数处理, $h_{pooling}$ 表示 $3*3$ 最大池化;

[0101] 对于编码器而言,丰富的空间细节不仅可以为高级语义特征和全局上下文提取建立良好的特征基础,还可以有效地还原火焰的边缘细节。本发明采用了三个多尺度特征提取模块来对输入图像进行特征提取,可以获得丰富的空间细节特征。

[0102] 隐式深度监督处理的输出特征设置为256。隐式深度监督处理可以使梯度直接流到较早的层,能够显著改善编码器的特征表示。

[0103] 通道注意力机制的计算过程如下:

$$[0104] \quad x' = gp(x) = \left( \sum_{i,j=0}^{H-1,W-1} x(i,j) \right) / (H*W)$$

$$[0105] \quad x_{ib} = \sigma(w_2 \otimes \delta(w_1 \otimes x'))$$

$$[0106] \quad f_i = x' * x_{1b}$$

[0107] 其中, gp表示全局平均池化,  $x(i, j)$  表示x的特征值,  $w_1$ 和 $w_2$ 表示权重矩阵,  $\delta$ 表示ReLU激活,  $\sigma$ 表示Sigmoid激活,  $\otimes$ 表示卷积运算,  $x_{1b}$ 表示可学习参数。

[0108] 图2为本发明提供的全局和局部上下文注意模块结构示意图,即图1中的“G”。如图2所示,所述全局和局部感知注意机制包括全局上下文分支和局部上下文分支;

[0109] 所述步骤3)具体包括:

[0110] 步骤3-1)利用卷积层对所述层级特征 $f_n$ 进行特征压缩,得到密集特征 $\tilde{f}_n$ ,所述卷积层的卷积核大小为 $3*3$ ,步长为1,通道数为256;

[0111] 步骤3-2)所述全局上下文分支包括自适应最大池化层和大小为 $1*1$ 的卷积层,由所述自适应最大池化层将 $\tilde{f}_n$ 转换为语义向量 $v_s$ , $v_s$ 与 $\tilde{f}_n$ 进行通道乘法及卷积运算,得到全局上下文信息 $f_n^{gc}$ ,计算公式如下:

$$[0112] \quad v_s = gp(\tilde{f}_n)$$

$$[0113] \quad f_n^{gc} = \delta(\text{conv}(v_s \odot \tilde{f}_n))$$

[0114] 其中,“gp”表示最大池化处理,“ $\odot$ ”表示通道乘法,“conv”表示卷积操作,“ $\delta$ ”表示ReLU激活函数;

[0115] 步骤3-3)所述局部上下文分支由n个大小为 $3*3$ 的卷积层堆叠而成,由所述局部上下文分支将 $\tilde{f}_n$ 转换为局部上下文权重 $w_{lc}$ ,计算公式如下:

$$[0116] \quad w_{lc} = \delta(\text{stacked\_conv}(\tilde{f}_n))$$

[0117] 其中,stacked\_conv表示堆叠的卷积层;

[0118] 步骤3-4)利用 $w_{lc}$ 对 $f_n^{gc}$ 进行特征增强,得到增强后的特征 $f_{glc}$ ,将 $f_{glc}$ 进行大小为 $1*1$ 的卷积运算,进行特征压缩,同时根据 $\tilde{f}_n$ 生成局部偏差b,由卷积运算的输出结果与偏差b计算得到所述全局及局部语义融合特征 $f'_n$ ,计算公式如下:

$$[0119] \quad f_{glc} = w_{lc} \odot f_n^{gc}$$

$$[0120] \quad f'_n = \delta(\text{conv}(f_{glc})) + b$$

[0121] 其中,“ $\odot$ ”表示点乘。

[0122] 与语义分割任务不同,由于气流的影响,火焰没有固定的形状信息。因此,编码器所提取特征中的全局上下文和局部语义信息对于火焰的语义分割是多余的。为此,本发明提出了一种全局和局部感知注意机制,对编码器最后一层输出的层级特征 $f_n$ 进行 $1*1$ 卷积运算,将其压缩为具有256个通道的特征,可以得到更紧密的特征表示,之后利用两个不同的分支来获取全局上下文特征和局部上下文特征,实现了火焰特征的增强。最后使用 $1*1$ 卷积将增强后的特征通道压缩为256,并采用矩阵加法实现对所提取火焰特征的进一步优化,可以提取适用于火焰检测的更具代表性的全局上下文和高级语义信息。

[0123] 图3为本发明提供的交叉语义注意模块结构示意图,即图1中的“C”。如图3所示,所述步骤4)具体包括:

[0124] 步骤4-1)将所述编码器第i层的输出 $f_i$ 进行卷积运算得到特征 $f_{i-256}$ ,卷积核大小

为1\*1,通道数为256;

[0125] 步骤4-2) 将 $f'_n$ 进行卷积运算,卷积核大小为1\*1,通道数为256,之后进行双线性插值上采样计算,得到特征 $f'_{n-256}$ ;

[0126] 步骤4-3) 对 $f_{i-256}$ 和 $f'_{n-256}$ 按照通道维度进行级联,得到能够表示空间细节特征及全局上下文的特征 $f_{i-cc}$ ,计算公式如下:

$$[0127] \quad f_{i-cc} = [\text{conv}(f_i), u(\text{conv}(f'_n))]$$

[0128] 其中,“u”表示线性插值上采样,“[]”表示级联;

[0129] 步骤4-3) 对 $f_{i-cc}$ 进行卷积运算,卷积核大小为1\*1,通道数为256,将卷积运算结果进行有效通道注意机制处理,将所得结果与 $f_{i-256}$ 进行点乘运算,得到第i层的局部空间细节优化特征 $f_i^{cc}$ ,计算公式如下:

$$[0130] \quad f_i^{cc} = f_{i-256} \odot ECA(f_{i-cc})$$

[0131] 其中,ECA表示通道注意机制,“ $\odot$ ”表示点乘。

[0132] 在解码过程中,通过逐步融合多层次特征可以增强不同分辨率特征的表达能力,从而优化火焰的语义分割。但是,低级空间特征不仅缺乏丰富的语义信息,还可能包括噪声。另外,上采样操作会导致高级语义特征被削弱。因此,本发明提出了交叉语义注意机制,通过双线性插值上采样计算,可以恢复 $f'_n$ 的特征分辨率,通过将 $f_{i-256}$ 和 $f'_{n-256}$ 按照通道维度进行级联,可以形成表示空间细节和全局上下文的新特征,通过加入有效的通道注意机制ECA,可以自适应地选择更有效的特征,进一步增强模型的跨上下文表示能力,实现抑制噪声并降低全局信息被削弱的影。

[0133] 在上述实施例的基础上,步骤5) 具体包括:

[0134] 步骤5-1) 将编码器部分得到的所述全局及局部语义融合特征 $f'_n$ 经自优化处理后得到特征 $f_{in-d}$ ,作为解码器第一层的输入;

[0135] 步骤5-2) 将 $f_{in-d}$ 经解码器第一层解码后的输出 $f_{1-d}$ 与编码器第n-1层得到的局部空间细节优化特征 $f_{n-1}^{cc}$ 进行特征融合机制处理,得到特征 $\tilde{o}_1$ ,将 $\tilde{o}_1$ 进行自优化机制处理,得到第一层的自优化输出结果 $o_1$ ,作为解码器第二层的输入;

[0136] 步骤5-3) 将编码器前一层的自优化输出结果作为后一层的输入,执行与步骤5-2) 相同的计算。

[0137] 图4为本发明提供的自优化模块结构示意图,即图1中的“S”。如图4所示,所述自优化处理机制,包括:

[0138] 步骤5-2-1) 对解码器第i层的输出 $\tilde{o}_i$ 经过两个卷积运算,分别得到位置权重 $W_i$ 和位置偏置 $b_i$ ;其中两个卷积运算的卷积核大小均为3\*3,通道数为256;

[0139] 步骤5-2-2) 利用位置权重 $W_i$ 和位置偏置 $b_i$ 对 $\tilde{o}_i$ 进行自优化,得到输出结果 $o_i$ ,计算公式如下:

$$[0140] \quad W_i = \text{conv}_w(\tilde{o}_i)$$

$$[0141] \quad b_i = \text{conv}_b(\tilde{o}_i)$$

$$[0142] \quad o_i = \delta(W_i \odot \tilde{o}_i + b_i)$$

[0143] 在解码器的解码过程中,本发明使用简单的特征聚合策略来融合多层次特征。但

由于网络不同层具有不同响应,以及每一层中特征的内部差异,使得融合后的特征表现欠佳。因此,本发明提出了自优化机制,进一步优化了融合特征,使得优化后的特征更适用于火焰定位任务。

[0144] 本发明中的特征映射包括1个全连接层和1个Sigmoid函数;所述全连接层包括两个神经元,实现高层抽象特征向输出层输出结果的特征映射;采用Sigmoid归一化函数进行归一化处理,得到图像的像素级分类结果,表示为:

$$[0145] \quad p_i = \frac{e^i}{\sum_j e^j}$$

[0146] 其中*i*表示第*i*个得分值, $\sum_j e^j$ 表示所有特征值的总和;选取概率最大值对应的类别作为火焰检测的判别结果。

[0147] 在上述实施例的基础上,步骤7)具体包括:

[0148] 步骤7-1)将模型输出的结果与真实标签进行误差求取;所述误差求取采用交叉熵,其可以表示为:

$$[0149] \quad Loss(y, \bar{y}) = H(p, q) = -\sum_{i=1}^n p(x_i) \log q(x_i)$$

[0150] 其中, $Loss(y, \bar{y})$ 表示真实标签 $\bar{y}$ 与预测结果 $y$ 之间的误差, $p(x_i)$ 表示模型的前向输出结果, $q(x_i)$ 表示真实标签, $\Sigma$ 表示求和;

[0151] 步骤7-2)采用步骤7-1)得到的参数作为本次迭代的权重值;从剩余的图片中随机选取一组图片,经步骤2)、步骤3)、步骤4)、步骤5)、步骤6)和步骤7-1),得到新的参数组合;反复迭代,直至完成一个迭代周期(epoch);

[0152] 步骤7-3)对训练图片进行重新洗牌,转至步骤1);反复执行,直至训练出最优参数组合。

[0153] 在上述实施例的基础上,步骤8)具体包括:

[0154] 步骤8-1)对于待检测的图片,利用双线性插值将待测图片的尺寸调整至模型输入所需大小 $224 \times 224 \times 3$ ;

[0155] 步骤8-2)将调整后的图片输入至最优参数模型,经模型前向传输得到预测结果;

[0156] 步骤8-3)将每一像素的预测结果最大概率值所对应的类别是与真实标签进行比较,若一致,则预测正确,反之,则预测错误;

[0157] 步骤8-4)执行步骤8-1)、步骤8-2)和步骤8-3),直到全部待测图片检测完毕。

[0158] 本发明还提供一种基于双语义注意力机制的高精度火焰定位系统,基于上述实施例提供的任意方法构建的双重情境注意网络模型实现,该系统包括:

[0159] 数据预处理模块,用于对输入图像进行尺寸调整;

[0160] 编码器,用于对输入图像进行多层次特征的提取;

[0161] 全局和局部感知注意模块,用于对输入特征进行全局和局部感知注意机制处理,捕获更丰富的高级语义和全局上下文信息,得到全局及局部语义融合特征;

[0162] 交叉语义注意模块,用于对输入特征进行交叉语义注意机制处理,优化局部空间细节信息,得到局部空间细节优化特征;

[0163] 特征融合模块,用于对输入特征进行融合,实现多层次特征的融合;

[0164] 自优化模块,用于进行自优化机制处理,优化融合特征;

[0165] 解码器,用于聚合高级上下文特征和低级空间细节特征。

[0166] 另外,上述各模块可执行方法实施例种对应的方法流程,此处不再赘述。

[0167] 最后所应说明的是,以上实施例仅用以说明本发明的技术方案而非限制。尽管参照实施例对本发明进行了详细说明,本领域的普通技术人员应当理解,对本发明的技术方案进行修改或者等同替换,都不脱离本发明技术方案的精神和范围,其均应涵盖在本发明的权利要求范围当中。

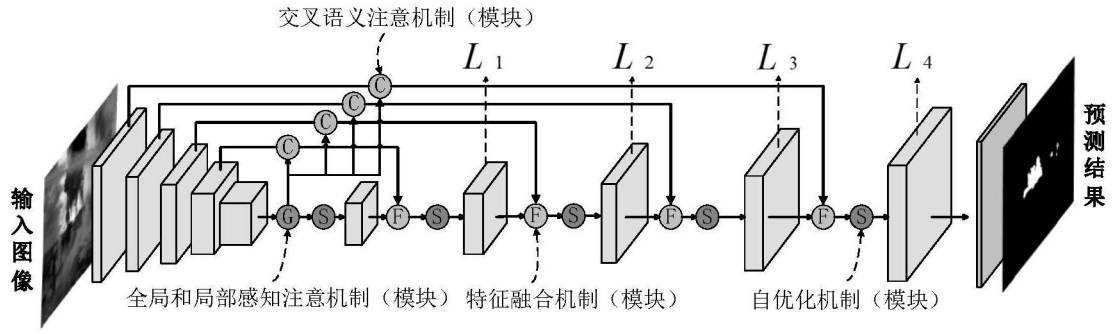


图1

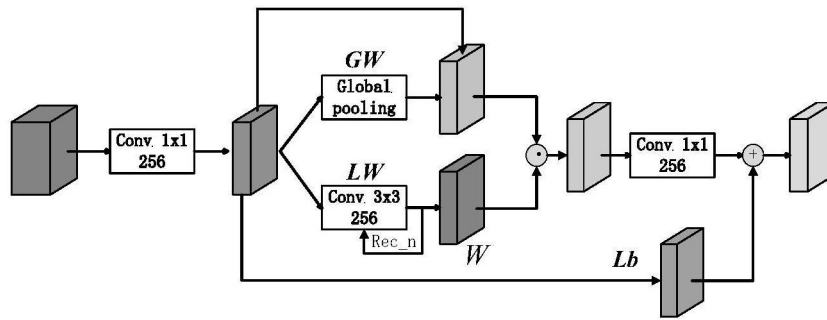


图2

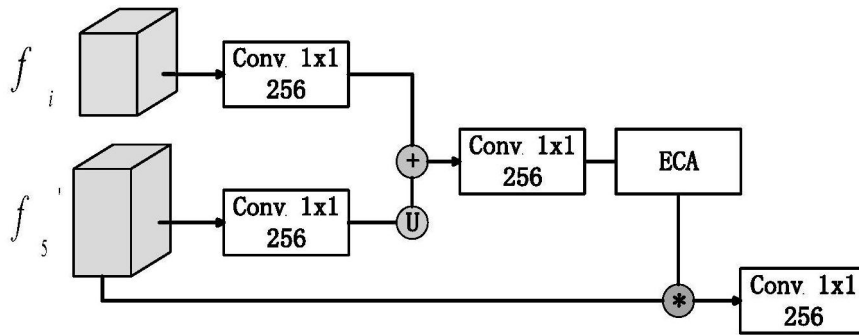


图3



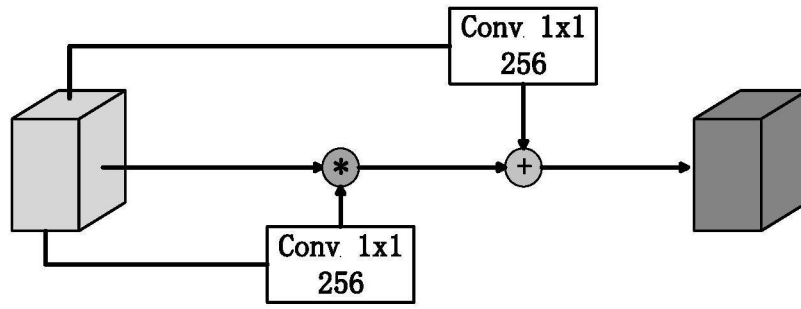


图4