

文章编号: 2096-1618(2023)03-0281-10

基于烟雾区域和轻量化模型的视频烟雾检测

蒲建飞, 魏维, 吴帝勇, 程鹏, 袁定胜

(成都信息工程大学软件工程学院, 四川 成都 610225)

摘要:烟雾是早期火灾发生的典型特征,针对烟雾的智能检测能有效降低森林火灾造成的破坏。为了对监控视频中烟雾的及早检测,提出一种基于烟雾区域及轻量化模型的烟雾检测算法。首先通过残差帧堆叠获取视频中的运动区域,然后再利用自适应暗通道掩码对运动区域进一步筛选获得疑似烟雾块。在此基础上,又设计了一个轻量化的神经网络模型用于烟雾识别,模型利用卷积局部感知的特性提取烟雾的浅层特征,而在网络深层则将卷积和 self-attention 相结合,通过比较全局相似度,在浅层特征图的基础上获取烟雾的全局信息。实验结果表明,算法具有较强的鲁棒性,无论是远距离烟雾还是近距离烟雾均有良好的检测效果。

关键词:早期烟雾;暗通道;自注意力机制;卷积神经网络

中图分类号:TP391.4

文献标志码:A

doi:10.16836/j.cnki.jcui.2023.03.006

0 引言

森林是一种重要的生态资源,为人类生产生活提供重要的原材料。森林火灾是一种极具破坏性的灾难,不仅使生态环境受到严重破坏,而且给人类造成巨大的经济损失,甚至危及生命安全。而发生森林火灾时灭火的难度远远比城市火灾大,因此对森林火灾进行有效的预警就显得尤为重要。传统的火灾探测系统一般采用温度或烟雾传感器,当火灾发生时,生成的烟、温度和光等物理量达到一定数值时,发出报警信号^[1]。但这种探测系统仅适用于建筑物、隧道等的火灾预警中,在高大空间的火灾报警中,上述方法不能很好地适用。

近年来随着机器视觉与图像处理技术的不断发展,基于视频图像的火灾检测技术引起了关注。这种新型火灾检测技术可以忽略地形地势的限制,其检测覆盖面大,不存在检测死角和空白,并且拥有高检测率和能够远程实时预警的特点,因此逐渐成为森林火灾预警的主要方案。其中,由于烟雾和火焰是火灾的两种主要表现形式,而烟雾产生于火焰之前,相比对火焰的检测,对烟雾的有效识别能更及时地预警和预防森林火灾的发生,因此对烟雾的研究越来越多。Russo A U等^[2]首先利用背景减法从输入帧中去除背景,然后采用基于形状的滤波方法获取感兴趣区域。根据感兴趣区域的像素点计算局部二值模式的值并构建直方图形成特征向量,最后利用支持向量机对形成的特征向量进行判别。Tang等^[3]则通过背景差法提取运动区域,并对每个运动区域进行处理以获得局部信息。利用局部二值化模型提取每个块的纹理特征。然后利用

支持向量机对提取的特征进行分类。Liu等^[4]使用小型烟雾图像块与视频中运动区域的图像特征匹配进行烟雾识别。首先利用 Vibe 算法提取视频帧中的运动区域,然后提取运动区域的局部向量化特征和中心对称的局部二值模式特征进行特征融合,将融合特征输入支持向量机进行烟雾识别。Gao等^[5]在帧处理阶段不提取火灾烟雾特征(如纹理、颜色、频率信息等),仅使用连续帧提取动态区域中的稳定点作为烟雾根候选点。在扩散模型模拟阶段,采用所有烟雾根候选点信息生成模拟烟雾。最后实现基于颜色、动态区域和模拟烟雾的匹配算法辨别烟雾。Wang等^[6]设计了一种识别圆锥几何特征的算法,用于区分烟雾圆锥区域和动态区域。然后,使用颜色滤波算法对这些圆锥形区域进行滤波处理。最后,通过纹理滤波算法将真实烟雾与候选烟雾区域区分。Zhou等^[7]采用最大稳定极值区域算法提取烟雾的局部极端区域。通过跟踪极端区域获取疑似烟雾区域,后利用提出的累积区域方法根据烟雾的运动特征来检测烟雾。Wang等^[8]则利用 RGB 和 HSV 颜色模型获取烟雾颜色特征,利用小波变换获取背景模糊特征,计算像素比获取烟雾轮廓特征,利用光流法获取主运动方向的特征。然后,将获取的各烟雾特征融合后输入支持向量机进行烟雾检测。上述方法均依靠人工设计的特征来进行烟雾检测。但人工设计的特征容易受到环境的影响,当环境复杂人工设计的特征在进行烟雾检测时往往出现高误报和高漏报的情况。

随着深度学习技术的发展,2015年,卷积神经网络(convolutional neural network, CNN)在 ImageNet 图像分类竞赛上超越了所有传统图像特征提取方法,首次超越受过训练的人类的识别率,证明卷积神经网络在提取图像特征上拥有很强的能力。越来越多的研究

者开始将基于卷积神经网络的深度学习技术应用在视频烟雾检测上。Zhang^[9]提出了一种使用迁移学习来检测烟雾图像的双通道卷积神经网络(DC-CNN):一个通道上使用带有迁移学习的 AlexNet 网络提取广义特征,一个通道使用纯卷积神经网络提取特定特征。Yuan 等^[10]受到 GoogleNet, Inception 模块网络结构的启发,设计了更加适合烟雾识别的网络模块,提出 DM-CNN 网络结构。相比 ZF-Net、VGG16、Inception-v3 等网络有着更好的识别效果。Yin 等^[11]提出了一种基于视频的深度学习神经网络烟雾检测方法,该方法有两个独立的网络分别学习空间特征和运动特征。其中,空间网络从原始帧学习特征,运动网络学习相邻两帧的运动特征。之后,将其输入 RNN 网络进行特征融合借此识别烟雾。Xu 等^[12]提出了一种基于 SSD 和多尺度深度卷积神经网络的烟雾检测器,并通过使用合成烟雾图像样本进行网络训练。对比基于 ZFNet 的 SSD,基于 VGGNet 的 SSD,该方法在烟雾检测性能上显著提升。利用卷积神经网络这类的深度学习方法,虽然可以自动提取烟雾特征并拥有较好的烟雾识别率,但该类方法需要巨大的训练数据且计算开销大,无法做到对野外火灾进行实时预警。

针对上述问题,本文提出一种基于疑似烟雾区域的神经网络早期森林烟雾识别方法。该方法拥有以下优势:(1)将残差帧应用在运动物体检测的同时,与提出的自适应暗通道掩码相结合用于确定疑似烟雾区域,这样可以减少后续处理的数据量,同时保证未丢失后面待识别的烟雾特征。(2)所提出的烟雾识别模型不厚重,参数量很小,能够做到对捕获到的疑似烟雾区域进行实时识别。(3)在所设计的轻量级模型中将自注意力和卷积相融合,使模型在提取烟雾局部特征的同时关注烟雾的全局信息。

总的来说,通过该方法能够同时兼顾林火烟雾检测的准确性和实时性。方法流程如图 1 所示。

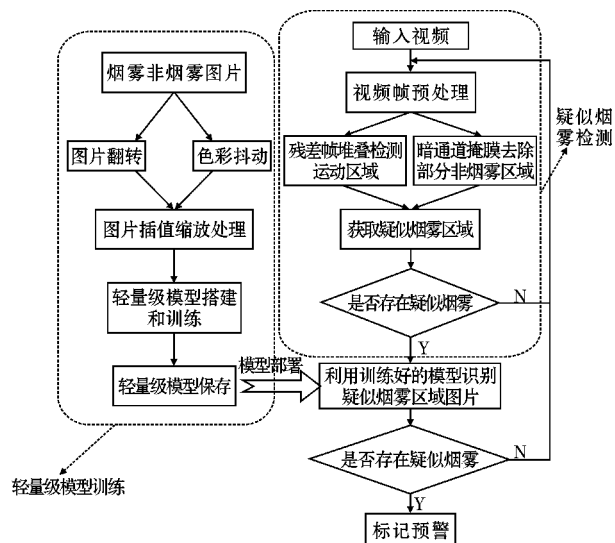


图1 早期森林火灾烟雾识别方法流程

1 疑似烟雾目标区域检测

1.1 残差帧堆叠检测运动区域

常见的运动检测算法有光流法,背景差分法和帧差法等。光流法利用图像序列中像素在时间域上的变化以及相邻帧之间的相关性,找到上一帧跟当前帧之间存在的对应关系,计算相邻帧之间物体的运动信息。但大多数的光流计算方法计算量巨大,结构复杂,且易受光照、物体遮挡或图像噪声的影响,鲁棒性差,不能做到对运动目标进行实时检测。背景差分法是指将图像序列中的当前帧和已经确定好背景图像做减法,像素值差异超过一定阈值的区域为运动区域。其不足之处在于当场景变得复杂、不可预知时,如光照突然变化、运动物体进出场景时,选择性的更新背景成为背景差分法的一大难点。帧差法是利用相邻帧(或相邻三帧)之间求取两两帧之间图像灰度差的绝对值,当绝对值超过一定阈值时,可判断为运动目标。其优点是算法实现简单,程序设计复杂度低,能够适应各种动态环境,有着比较强的鲁棒性。缺点是对慢速运动的物体,特别是当物体在前后两帧中几乎完全重叠时,检测不到物体。由于烟雾形成初期运动缓慢,单纯地使用帧差法将很难检测到缓慢运动的烟雾。

本文提出利用残差帧堆叠^[13]来获取视频中的运动物体。通常的残差帧通过对相邻两帧进行帧差法获取,但由于早期烟雾扩散缓慢,相邻两帧获取的残差帧不能很好地捕捉到疑似烟雾区域,因此本文提出的算法扩大帧间距离,利用相隔 20 帧的两帧来获取残差帧。获取残差帧的公式如下:

$$\text{ResFrame}_i = \begin{cases} 1, & |\text{Frame}_i - \text{Frame}_{i-20}| > T \\ 0, & \text{else} \end{cases}$$

其中 Frame_i , Frame_{i-20} 分别代表第 i 帧图像和第 $i-20$ 帧的背景参考帧图像, ResFrame_i 代表获取的残差帧图像,在实验中发现 T 值为 20 时效果较好。利用帧差法来获取残差帧虽然速度较快,但由于其固有的缺点不能提取出对象的完整区域,只能提取出边界。所以,本文提出利用残差帧堆叠来弥补该缺陷。受 vibe 算法的启发,首先将连续的多个残差帧进行保存得到残差帧集合 $M(x)$:

$$M(x) = \{\text{ResFrame}_i, \text{ResFrame}_{i-1}, \dots, \text{ResFrame}_{i-t}\}$$

后对集合 $M(x)$ 中的残差帧堆叠得到当前视频帧的运动前景。其具体的实现方法如下:

$$\text{MoveFrame}_i(x, y) = \begin{cases} 1, \sum_{i=0}^t \text{ResFrame}_i(x, y) > 2 \\ 0, \text{else} \end{cases}$$

其中 $\text{MoveFrame}_i(x, y)$ 表示当前视频帧的运动前景, t 值为 20 代表集合 $M(x)$ 中保存的残差帧个数。

利用残差帧堆叠虽能实时获取视频帧中的运动物体,但当捕捉到运动的烟雾后,由于烟雾扩散缓慢和帧差法固有的缺陷,烟雾后期将很难检测到运动的烟雾,因此提出若有连续 5 帧检测到烟雾,则冻结前文提出的背景参考帧 Frame_{i-20} 。尽管残差帧堆叠能有效去除视频帧噪声和镜头晃动对运动检测的干扰,但是户外环境往往是非常复杂的,如野外环境树叶晃动、飞鸟突入镜头、行人、汽车均会对视频帧的运动烟雾检测形成较严重的干扰。因此,本文在利用残差帧堆叠进行运动烟雾检测的基础上,引入自适应暗通道掩膜来进一步筛选出运动的烟雾。

1.2 自适应暗通道掩膜

通常彩色图像包括 3 个通道,即 RGB 三通道,可以理解成一张图片由 3 层同样大小的 RGB(光学三原色:红绿蓝)三色堆叠而成,而图片实质上是由一个个像素组成的,像素点上的每一种颜色由这三原色组合而成。暗通道先验^[14]指清晰无雾的彩色图片中除天空区域(因为天空区域或者白色区域和雾的特征较为接近)外的任一局部区域像素至少有一个通道值很低,几乎趋近于零。即在一个 RGB 图像中所有像素点取最低的颜色通道值所构成的新图像称之为暗通道。通过以上描述,一张图像的暗通道的计算公式如下:

$$J^{\text{dark}}(X) = \min_{y \in \omega(x)} \left(\min_{c \in \{r, g, b\}} J^c(Y) \right)$$

其中 J^c 表示图像三通道中的一个, $\omega(x)$ 表示以像素 x 为中心的区域,两次最小值滤波的滤波半径为

$$\text{WindowSize} = 2 \cdot \text{Radius} + 1$$

做近似计算,其中 WindowSize 为最小值滤波窗口大小。

在自然环境中,早期的烟雾通常为白色,表现在暗通道上则是存在烟雾的区域拥有很高的像素值。因此,暗通道像素值的强度能够在一定程度上作为检测烟雾是否存在的方法。获得暗通道后,可通过阈值判断得到暗通道掩膜来划分真实图像中的疑似烟雾区域。具体计算公式为

$$\text{DarkMask} = \begin{cases} 1, J^{\text{dark}}(X) > T_1 \\ 0, \text{else} \end{cases}$$

其中 DarkMask 即为所求的暗通道掩膜,但一般阈值 T_1 是固定的,得到的掩膜无法适应野外的复杂环境,做到去除烟雾检测的大部分干扰。因此,本文提出自适应掩膜来解决该难题,使掩膜的阈值随环境变化而

变化。具体方案是,利用残差帧堆叠获取到的运动前景和暗通道来改变阈值 T_1 。首先将运动前景与暗通道并操作获取在运动区域在暗通道上的映射,再对得到的映射图像求像素平均值,该像素平均值即可视为暗通道掩膜的阈值。其计算公式如下:

$$\text{MoveDark}_i = \text{MoveFrame}_i(x, y) \cap J_i^{\text{dark}}(X)$$

$$T_1 = \frac{\sum_{x=0y=0}^h \sum_{x=0y=0}^w \text{MoveDark}(x, y)}{\sum_{x=0y=0}^h \sum_{x=0y=0}^w \text{MoveFrame}(x, y)}$$

其中 MoveDark_i 代表第 i 帧图像运动区域在其暗通道上的映射, $\text{MoveFrame}_i(x, y)$ 代表第 i 帧图像的运动前景, $J_i^{\text{dark}}(X)$ 代表第 i 帧图像的暗通道, T_1 表示第 i 帧图像暗通道掩膜的阈值, h, w 表示视频帧的高和宽。

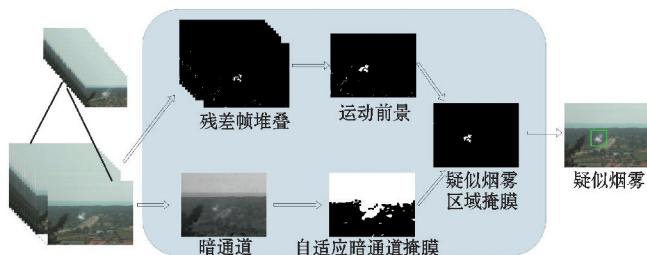


图2 疑似烟雾目标区域检测

通过残差帧堆叠和自适应暗通道掩膜进行疑似烟雾区域检测的效果如图 2 所示。由图 2 可以看到,残差帧堆叠能有效去除视频帧上的噪声,并且弥补了帧差法获取运动目标易形成的“孔洞”的缺点。自适应暗通道掩膜也能在获取的运动前景的基础上进一步去除如树叶晃动、运行车辆等运动物体,从而得到较少且准确的疑似烟雾块。

2 融合自注意力机制的轻量化深度学习网络设计

2.1 整体网络设计

本文提出的融合自注意力机制的轻量化深度学习网络如图 3 所示。该网络模型主要包含两大模块:浅层特征网络模块,深层特征网络模块。具体来说,模型由一个卷积核为 7×7 的卷积层、一个最大池化层、3 个浅层特征网络模块、3 个深层特征网络模块、1 个平均池化层、2 个全连接层组成。纯卷积的浅层特征网络模块负责利用卷积操作对烟雾图像的局部特征进行提取和组合,形成较长较广的深层特征;深层特征网络模块将卷积和自注意力机制相融合,凭借其较好的抽象能力更好地提取烟雾的深层抽象特征,并增强整个网络模型对特征的提取能力。卷积和自注意力机制相结

合共同提取物体特征,使模型能够在网络模型不厚重的基础上,对烟雾特征拥有更好的表征能力,并在视频烟雾识别任务中取得较好的识别效果。

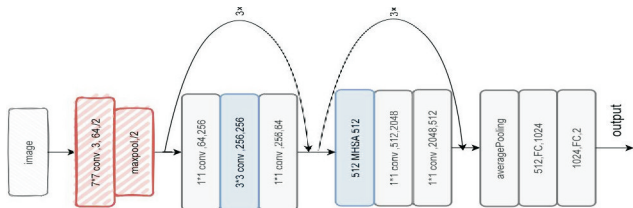


图3 浅层和深层网络相结合的轻量型的深度学习网络

2.2 浅层特征提取模块

受 MobileNetV2^[15] 的启发,提出嵌入位置注意力机制^[16] (coordinate attention) 的倒残差模块 (inverted residual block, IRB), 利用该模块进行烟雾的浅层特征提取。通常使用的注意力机制往往关注于哪些通道对整个网络的识别性能提升最大,但都忽略了位置信息对物体特征提取的影响。位置注意力机制将位置信息融合在通道注意力中,在不丢失通道信息的同时关注不同位置的像素点的重要信息。其结构如图4所示。

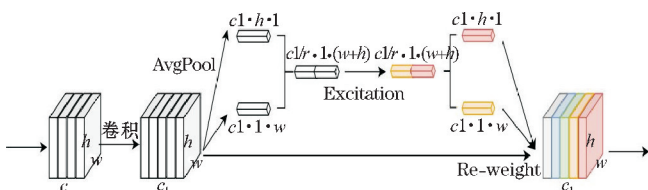


图4 位置注意力模块结构图

位置注意力使用 AvgPool 操作将多维特征图沿高和宽分解为两个一维特征向量,并分别沿两个空间方向通过 Excitation 操作聚合特征。通过这种方式,位置注意力可以沿一个空间方向捕获远程依赖关系,同时沿另一个空间方向保留精确的位置信息。然后将得到的特征图单独编码成一对方向感知和位置敏感的注意力图,得到的注意力图分别加权到输入特征图中从而增强神经网络对感兴趣对象的描述。

倒残差结构如图5(a)所示,由两种卷积操作组成,分别为深度卷积 (depthwise convolution, Dw) 和逐点卷积 (pointwise convolution, Pw)。深度卷积的一个卷积核负责一个通道,一个通道只被一个卷积核卷积,而常规卷积每个卷积核是同时操作输入特征图的每个通道。逐点卷积的运算与常规卷积运算非常相似,它的卷积核的尺寸为 $1 \times 1 \times M$, M 为上一层输出的通道数,逐点卷积会在深度方向上对上一步的特征图进行加权组合,生成新的特征图。倒残差结构接收到上一层的特征图后,先使用逐点卷积扩充特征图维度,将低

维空间映射到高维空间。因为在低维空间下,网络无法提取到整体的足够多的信息。而后使用深度卷积在高维空间下提取特征,卷积核的尺寸为 3×3 。最后,再次使用逐点卷积降低特征图维度,与输入倒残差结构的特征图维度一致,从而方便后续的残差操作。

相比于传统的残差结构,倒残差结构特征提取时不会损失信息,且计算量也大幅降低。这主要源于倒残差结构使用深度可分离卷积 (depthwise separable convolution) 替换了标准卷积。

假设一个标准卷积的输入特征图维度为 $D_F \times D_F \times N$, 其内部卷积核尺寸为 $N \times F_w \times F_h \times M$, 输出的特征图维度为 $D_C \times D_C \times N$, 则标准卷积的计算量如下:

$$F_C = F_w \times F_h \times D_C \times D_C \times N \times M$$

深度可分离卷积包含逐点卷积和深度卷积两个操作,进行逐点卷积时其计算量为

$$F_{Pw} = F_w \times F_h \times D_C \times D_C \times M$$

进行深度卷积时,其计算量为

$$F_{Dw} = 1 \times 1 \times D_C \times D_C \times M \times N$$

所以,利用深度可分离卷积进行运算的总计算量为

$$F_{Pw+Dw} = F_w \times F_h \times D_C \times D_C \times M + 1 \times 1 \times D_C \times D_C \times M \times N$$

深度可分离卷积和标准卷积的计算量之比为

$$\frac{F_{Pw+Dw}}{F_C} = \frac{1}{N} + \frac{1}{F_w \times F_h}$$

本文采用的嵌入位置注意力机制的倒残差模块的结构如图5(c)所示,位置注意力作用于倒残差模块的最后一个逐点卷积后,进行特征图内部信息的权重再分配。文献[16]将位置注意力作用于倒残差模块中深层卷积后,其结构如图5(b)所示。本文在实验阶段设置了3组对比实验用于验证有无位置注意力和位置注意力模块嵌入位置对模型烟雾识别效果的影响。

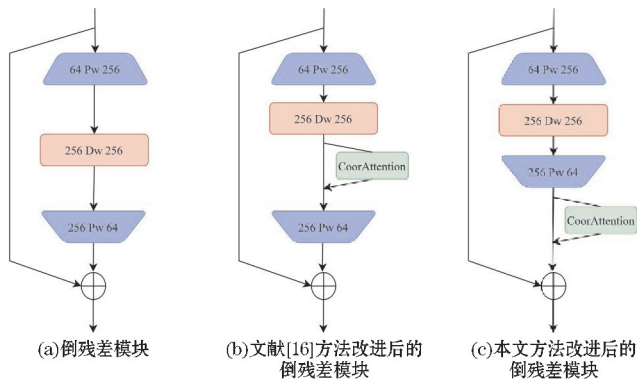


图5 改进的倒残差模块

2.3 深层特征提取模块

通过浅层特征网络模块可以获得较长较广的深层

特征,但由于卷积层的设计需通过受限的感受野来确保局部性,以及通过权重共享来确保平移等效性,其卷积核固有的局部性使卷积操作无法得到图像中的全局信息。为更好地识别图像中的对象,全局信息又必不可少。

自注意力机制能直接获取图像上任意两点的远距离依赖关系,在一定程度上与同特征图大小一致的卷积核结构类似,避免了纯卷积的过深网络结构。而本文使用的多头注意机制^[17] (multi-head self-attention, MHSA)的基础单元为点积型自注意力机制(dot product attention),其实现方式如图 6 所示,可表达为

$$\text{Attention}(Q,K,V)=\text{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V$$

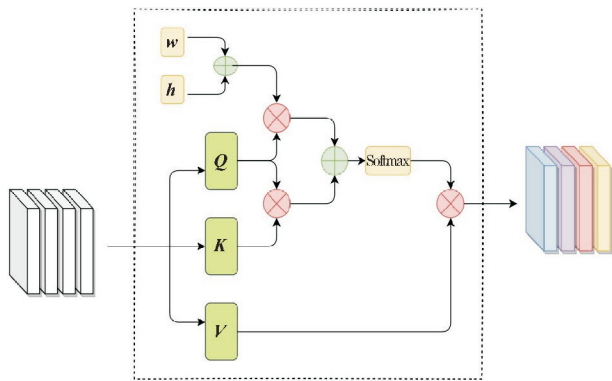


图 6 自注意力机制实现图

其中的 Q,K,V 是 3 个向量由特征图经一层全连接层转换得来,用于分析图像各区域间的联系。 d_k 为矩阵 Q,K 的维度。

自注意力的计算可分 3 步:通过 Q,K 计算各区域间的相似度,亦可称为注意力的值;使用 Softmax 函数对注意力值进行归一化;将归一化后的注意力值同 V 进行点积运算最终得到自注意力值。为防止 d_k 过大导致 Q,K 进行点积运算的结果方差过大,从而让后续的 Softmax 函数无法进行梯度计算,引入 $\sqrt{d_k}$ 进行内积缩放。另外,本文提及的自注意力机制引入相对位置编码 h,w ^[18-19],使模型在聚焦物体特征的同时考虑不同特征之间的空间位置关系。

在多头自注意力机制中,输入的特征图被均分成 h 块,每一个块形成一个子空间。然后,这均分的 h 块在各自的子空间下分别乘于不同的权重得到新的 Q,K,V ,各个子空间依靠新的 Q,K,V 计算各自空间下的自注意力值,将得到的自注意力值进行组合,然后映射成输入向量的维度。最终,获得多头自注意力的值。其计算公式如下:

$$\text{MultiHead}(Q,K,V)=\text{Concat}(\text{head}_1,\text{head}_2,\cdots,\text{head}_h)W^0$$

$$\text{head}_i=\text{Attention}(QW_i^Q,KW_i^K,VW_i^V)$$

使用多头注意力机制增加了模型获取特征信息的能力,并且在不同子空间下获取不同的特征信息,从而使得到的特征信息也更全面。

卷积只在局部小领域聚合上下文,避免了冗余的全局计算,但受限的感受野难以建模全局依赖。而自主注意力通过比较全局相似度,能自然地关联长距离目标,但同时其计算量过大且在浅层编码局部特征十分低效^[20]。因此,本文设计了一种新的特征学习模块,该模块将卷积和自注意力有机地统一,发挥两者的优势,解决了特征提取时局部冗余和全局依赖两大问题。该模块如图 7(a)所示,将倒残差结构中的深层卷积替换为多头自注意力。此时,自注意力机制的计算维度过高,大大增加模型参数量。因此,借鉴 ConvNeXt^[21]中倒残差的设计,将替换的多头自主注意力前移形成了图 7(b)结构。



(a)融合MHSA的倒残差结构 (b)融合MHSA的变型倒残差结构

图 7 深层网络模块

3 实验结果与分析

3.1 实验数据集及评估指标

由于森林火灾烟雾识别的研究者较少,环境复杂多变导致样本获取的难度较大,目前还没有针对森林火灾烟雾识别的标准数据集。所以,本文在公共数据集 <http://staff.ustc.edu.cn/~yfn/vsd.html> 的基础上构建了一个新的数据集,其中包含烟雾图片 4200 张,非烟雾图片 16000。由于在分类任务中,各类数据图片数量的相对平衡能显著提升模型表现^[10],因此本文通过水平翻转、随机旋转和色彩抖动等数据增强技术对数据集中的烟雾图片进行处理,得到新的烟雾图片,让烟雾和非烟雾图片数量基本一致。最终获得一个高质量的数据集用于模型训练,数据集中部分烟雾图片

如图8所示。数据集数据分配详情如表1所示,烟雾图片数量总计16016张,非烟雾图片数量总计16000。

表1 数据集中数据分配

DataSets	Smoke	Non-smoke	Total
Training Set	12812	12800	25612
Test Set	1602	1600	3202
Verification Set	1602	1600	3202
Total	16016	16000	32016

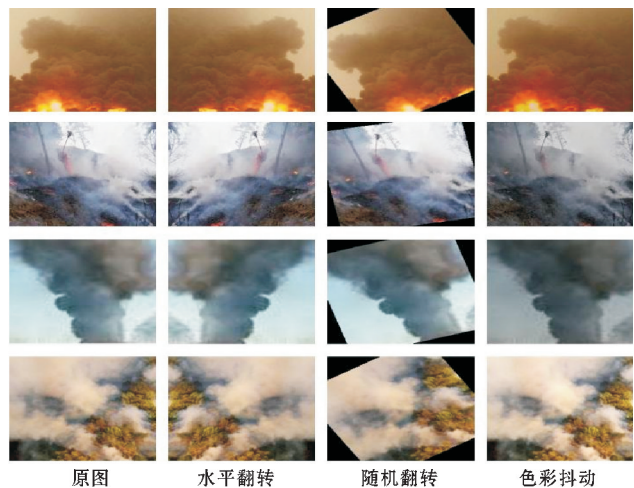


图8 数据集中的烟雾图片

为了评价神经网络模型的烟雾识别效果,本文采用准确率(accuracy rate, AR),检测率(detection rate, DR)和误警率(false alarm rate, FAR)作为模型评价指标,其计算公式如下:

$$AR=\frac{TP+TN}{TP+FP+FN+TN}$$
$$DR=\frac{TP}{TP+FN}$$
$$FAR=\frac{FP}{FP+TN}$$

其中 TP(TurePositive)代表实际标签为烟雾,同时被模型识别为烟雾的图片数量。FP(FalsePositive)代表实际标签为非烟雾,但被模型识别为烟雾的图片数量。TN(TrueNegative)代表实际标签为非烟雾,模型识别也为非烟雾的图片数量。FN(FalseNegative)代表实际标签为非烟雾,但被模型识别为烟雾的图片数量。若模型识别效果优秀,则在评价指标上表现为高准确率,高检测率以及低误警率。

3.2 疑似烟雾块缩放

在火灾发生的早期,产生的烟雾是逐渐弥漫扩散的,算法捕捉到的疑似烟雾区域的大小是不定的。而本文设计的神经网络模型的输入形状却是固定的,为

了方便网络模型提取烟雾特征并辨别疑似烟雾区域是否存在烟雾,需将疑似烟雾区域的图像数据转换为网络模型输入的大小 224×224。本文采用的缩放方法为双线性插值,如图9所示。虽然在检测过程中烟雾的形状一直变化,但烟雾的颜色和纹理特征以及其他特征在同一图像中总是相对一致的。因此,不管所获取到的疑似烟雾块形状如何,统一使用双线性插值将其转换为网络模型输入的大小。



(a)疑似烟雾区域 (b)缩放后疑似烟雾区域
图9 疑似烟雾区域缩放

3.3 网络模型训练

本文进行的实验基于 Window10 系统平台,实验环境为 Python3.8,开发环境为 Pycharm 和 VScode,使用的深度学习框架为 Pytorch1.7.0(GPU),CUDA 版本为 10.2。硬件设备为 Intel(R) Xeon(R) CPU E5-2603 v3 @ 1.60GHz,32GB RAM 和 24GB NVIDIA TITAN RTX。

在进行神经网络模型训练过程中使用交叉熵作为损失函数 Adam(Adaptive Moment Estimation)梯度下降算法,批处理图片数量为 16,初始学习率设定为 0.001,学习率变化策略设置为模型训练每经过 5 次迭代,学习率调整为上一次的0.7倍,而模型训练总的迭代次数为 100。

3.4 网络模型分析

3.4.1 位置注意力机制对特征提取影响分析
本文在浅层网络特征提取模块中引入位置注意力机制来提升位置信息对物体特征提取的效果。考虑到位置注意力机制作用于倒残差模块内部的不同位置会影响特征提取的效果,从而影响神经网络对烟雾的识别准确度。于是,本文针对位置注意力在倒残差模块中的位置设置了对比实验。首先,对位置注意力在倒残差模块中的位置进行设置,其中未嵌入位置注意力为算法1,嵌入位置注意力于深层卷积后作为算法2,嵌入位置注意力于逐点卷积后为本文算法,其各自的模块结构图见图7。然后,将3种结构的神经网络在

本文提出的数据集上进行烟雾识别实验。实验各项评价指标如表 2 所示。

表 2 位置注意力嵌入位置分析			单位: %
Model	AR	DR	FAR
算法 1	98.96	98.69	0.75
算法 2	99.13	98.75	0.50
本文算法	99.59	99.63	0.43

从表 2 可知,添加位置注意力机制能够提升模型的烟雾识别能力,位置注意力机制不仅关注特征图维度上的联系,而且也考虑了位置信息对物体特征提取的影响。如将本文算法的烟雾识别效果与算法 1 比较,其 AR 和 DR 分别提升了0.63%和0.94%,FAR 降低了0.32%。除此之外,从表 2 还可发现,位置注意力模块在模型中的作用位置也会影响模型的烟雾识别效果,将位置注意力作用于倒残差模块的逐点卷积后比作用于深层卷积后对模型效果的提升更好。如本文算法相比于算法 2,其 AR 和 DR 分别提升了0.46%和 0.88%,FAR 降低了0.07%。

3.4.2 Self-attention 与卷积融合方式分析

对于本文模型所采用的融合 MHSA 的倒残差结构的融合方式进行实验分析。融合方式包含:(1)单纯使用 MHSA 替换掉倒残差结构中的深层卷积,得到的模块结构如图 7(a)所示。(2)在方式 1 的基础上将 MHSA 前移,此时其结构如图 7(b)所示。然后,将这两种融合方式在本文数据集上进行实验,其实验结果的各项评价指标如表 3 所示。

表 3 Self-attention 与卷积融合方式分析				
融合方式	AR/%	DR/%	FAR/%	Params/M
方式 1	87.35	85.26	10.56	45.11
方式 2	99.59	99.63	0.43	9.14

从表 3 可知,当使用方式 1 对 MHSA 与倒残差结构进行融合时,其在测试集上的准确率仅为87.35%,检测率为85.26%。而使用方式 2 对 MHSA 与倒残差结构进行融合时,其准确率和检测率为 99.59%,99.63%,相比于方式 1 分别提升了12.24%,14.37%,同时其误检率也大幅降低了10.13%,从10.56%下降到0.43%,而其模型参数量更是从45.11 M大幅下降到9.14 M。由此可见,在训练数据有限的情况下,大幅提升输入自注意力模块的维度,不仅不能增强模型的特征提取能力,反而使模型参数量剧增。综合而言,本文模型所采用的融合方式 2 优于方式 1。

3.4.3 与近年来其他方法对比

为验证本文提出的神经网络模型的性能,将模型与近几年典型的神经网络模型进行对比实验。为公平比较,各个模型在训练过程中的相关超参数保持一致,训练过程中都使用本文提出的数据集,并且所有模型都从零开始训练,训练结果如图 10 所示。而训练完成后的各模型在测试集上的表现如表 4 所示。

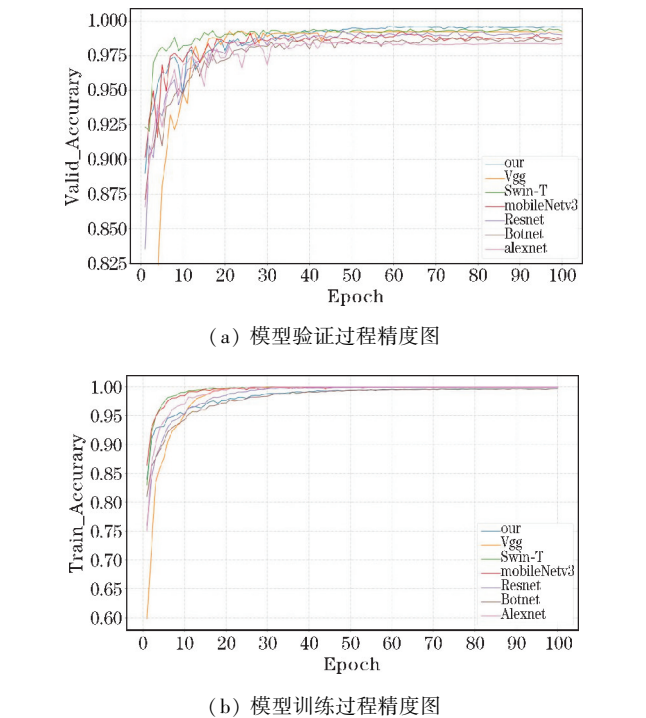


图 10 各网络模型训练和验证精度图

表 4 提出的模型法与其他经典模型比较

Model	AR/%	DR/%	FAR/%	Params/M
Alexnet ^[22]	98.96	98.69	0.75	61.10
Vgg ^[23]	96.45	97.06	2.49	15.66
Resnet ^[24]	98.65	98.37	1.06	25.56
MobileNetV3 ^[25]	98.93	98.56	1.69	5.40
Swin-transformer ^[26]	92.87	91.69	2.49	27.51
Botnet ^[18]	98.12	98.71	0.69	14.81
本文模型	99.59	99.63	0.43	9.14

从表 4 的数据结果可知,单纯使用自主注意力机制的神经网络模型在面临训练数据不足时,其模型表现远不如纯卷积模型。如 swin-transformer 在测试集上其 AR 和 DR 仅为92.87%,91.69%,FAR 为2.49%,各项指标均低于本文算法和卷积模型。但将卷积和自注意力相结合后,其模型表现又强于卷积模型,如本文提出的神经网络模型以及 Botnet 在烟雾识别任务上其准确率,检测率和误警率均优于其他模型。本文模型在测试集上其 AR 为99.59%,DR 为99.63%,FAR 为 0.43%,其烟雾识别效果优于与之比较的其他模型。

另外,本文所提出的神经网络模型参数量大小为 9.14 M。除 mobilenetV3 外,参数量远低于其他模型。但相比于 mobileNetV3,本文在 AR,DR 和 FAR 上均表现更好。

为更直观地分析模型在进行烟雾识别时的关注点,对神经网络由浅到深不同层部分通道的特征图进行可视化,如图 11 所示。上方为输入模型识别的疑似烟雾块,其中第一行展示的是神经网络最大池化层后输出的 5 个通道的特征图。由图 11 可以看出,在浅层,神经网络已经学习到了烟雾的的轮廓特征。第二行是融合位置注意力的反残差模块输出的部分通道特征图,在这一层神经网络不但注重烟雾的轮廓,也在学习烟雾的纹理特征。而第三行则是融合的 Self-Attention 的变型倒残差模块输出的部分通道特征图,这里的输出已经是神经网络的深层输出,从输出的特征图可以知道神经网络已经定位到烟雾在图片中的位置,并在图片的烟雾区域提取烟雾的深层特征。

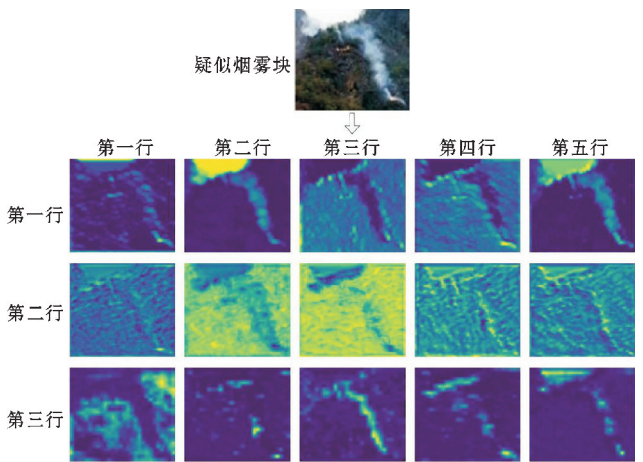


图 11 神经网络由浅到深不同层部分通道特征图

3.5 基于视频的烟雾检测结果分析

本文算法在多个实际烟火监控视频中进行测试,其视频数据来源于韩国启明大学 (<https://cvpr.kmu.ac.kr/>)、中国科学技术大学火灾科学国家实验室 (<http://smoke.ustc.edu.cn/datasets.htm>) 和土耳其比尔肯特大学 (<http://signal.ee.bilkent.edu.tr/VisiFire/>)。部分视

频数据的检测结果如图 12 所示。其中,video1 ~ video3 为烟雾视频,检测距离由远至近。video4 ~ video5 为干扰视频,主要包含自然环境下的云雾,晃动的树叶,运动的汽车和行人。

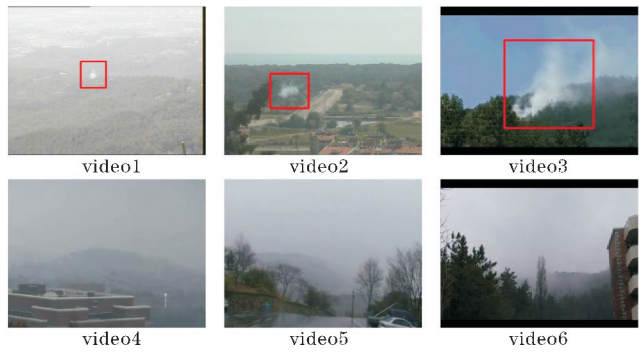


图 12 烟雾测试视频

为验证本文所提出的方法的有效性,将模型表现较为出色且为轻量级模型的 MobileNetv3 和 Botnet、文献[27]的方法、去除疑似烟雾检测步骤的本文方法与本文方法进行对比实验,实验结果如表 5 ~ 6 所示。从表 5 可看出,在 video1 ~ video3 等含烟雾视频中,本文方法相比 MobileNetv3、Botnet 和文献[27]的方法均能提前检测到烟雾的出现,且无论是远距离烟雾还是近距离烟雾的检测上,本文方法均表现出色。而从表 6 可得知,即使在烟雾类似视频 video4 ~ video6 中存在与烟雾颜色相近的云雾,频繁晃动的树枝,运动的白色车辆和行人等烟雾干扰物时,本文方法在进行烟雾检测时并未出现误检。这验证了本文方法拥有较强的鲁棒性。

在近距离烟雾检测视频 video3 中,去除疑似烟雾检测步骤的本文方法虽相较于其他 4 种方法表现优异,但在远距离烟雾检测视频 video1 中却无法检测到烟雾。在 video2 中,其他 4 种方法的检测速度慢于本文方法,且在烟雾类似视频 video4 ~ video6 中检测检测烟雾时出现了极多的误检行为。这证明了本文提出的疑似烟雾检测算法的有效性,该算法能极大地抑制云雾,晃动的树枝、运动的白色车辆和行人等物体对烟雾识别的干扰,在本文所提出的烟雾检测框架中发挥着巨大作用。

表 5 烟雾视频测试检测结果

视频 编号	视频 场景 描述	视频 总帧数	第几帧检测到烟雾			
			Mobile Netv3	Botnet	文献[27]	本文方法 (去除疑似烟雾检测)
Video1	远景烟雾缓慢扩散	6042	1165	1161	976	0
Video2	远景烟雾扩散有运动车辆	5956	492	497	523	1119
Video3	近处浓烟雾快速扩散,有风	2324	13	13	59	1

表 6 烟雾类似视频检测结果

视频 编号	视频 场景 描述	视频 总帧数	误检帧数				
			Mobile Netv3	Botnet	文献[27]	本文方法 (去除疑似烟雾检测)	本文方法
Video4	雨雪天有雾,远处有灯光闪烁	8208	0	0	0	2148	0
Video5	雨天大雾,有运动车辆,行人	7310	63	1	0	7310	0
Video6	雾天,微风,有树叶晃动	1992	0	0	0	45	0

5 结束语

频繁发生的森林火灾已经成为世界森林地区的一个主要威胁,而发生森林火灾时灭火的难度远远比城市火灾大。因此,对森林火灾进行有效的预警就显得尤为重要。烟雾产生于火焰之前,通过利用林区监控视频对烟雾的有效检测能更加及时地预警和预防森林火灾的发生。为提升检测效率,提出利用残差帧堆叠和自适应暗通道掩码进行先期的疑似烟雾区域获取。在此基础上,设计了一个融合自注意力机制的轻量化网络模型,用以对获取的疑似烟雾块进行快速识别。该模型相比于其他典型模型在烟雾识别时的准确率,检测率和误警率都表现出了更好的性能。实验结果表明,本文所提出的早期森林火灾检测算法相比于其他同类型算法,在检测速度上得到了极大的提升,且适用范围也更广阔,无论是远距离烟雾还是近距离烟雾都能做到准确实时地检测。不过,由于在进行烟雾识别时本文模型没有充分考虑到烟雾的时空特征,导致针对缓慢运动的云雾存在误检,未来的工作计划是构建更为有效的神经网络模型来降低烟雾识别的误检率。

参考文献:

[1] 严云洋,高尚兵,郭志波,等. 基于视频图像的火灾自动检测[J]. 计算机应用研究,2008(4): 1075-1078.

[2] RUSSO A U,DEB K,TISTA S C,et al. Smoke Detection Method Based on LBP and SVM from Surveillance Camera[C]. proceedings of the 2018 International Conference on Computer, Communication, Chemical, Material and Electronic Engineering(IC4ME2),2018.

[3] TIAN TIAN T,LINHAN D,ZHIJIAN Y. Smoke Image Recognition Based on Local Binary pattern [C]. proceedings of the Proceedings of the 2017 5th International Conference on Mechatronics, Materials, Chemistry and Computer Engineering(ICM-

MCCE 2017),2017.

[4] LIU Z,YANG X,LIU Y,et al. Smoke-Detection Framework for High-Definition Video Using Fused Spatial- and Frequency-Domain Features[J]. IEEE Access,2019,7:89687-89701.

[5] GAO Y,CHENG P. Forest Fire Smoke Detection Based on Visual Smoke Root and Diffusion Model [J]. Fire Technology,2019,55(5):1801-1826.

[6] WANG S,HE Y,YANG H,et al. Video smoke detection using shape,color and dynamic features[J]. Journal of Intelligent & Fuzzy Systems,2017,33:305-313.

[7] ZHOU Z,SHI Y,GAO Z,et al. Wildfire smoke detection based on local extremal region segmentation and surveillance[J]. Fire Safety Journal,2016,85:50-58.

[8] WANG L,LI A. Early fire recognition based on multi-feature fusion of video smoke [C]. proceedings of the 2017 36th Chinese Control Conference (CCC),2017.

[9] ZHANG F,QIN W,LIU Y,et al. A Dual-Channel convolution neural network for image smoke detection[J]. Multimedia Tools and Applications,2020,79(45):34587-34603.

[10] YUAN F,ZHANG L,WAN B,et al. Convolutional neural networks based on multi-scale additive merging layers for visual smoke recognition[J]. Machine Vision and Applications,2019,30(2): 345-358.

[11] YIN M,LANG C,LI Z,et al. Recurrent convolutional network for video-based smoke detection [J]. Multimedia Tools and Applications,2019,78(1):237-256.

[12] XU G,ZHANG Q,LIU D,et al. Adversarial Adaptation From Synthesis to Reality in Fast Detector for Smoke Detection[J]. IEEE Access,2019,7:29471-29483.

[13] TAO L,WANG X,YAMASAKI T. Rethinking Motion Representation:Residual Frames With 3D

- ConvNets[J]. IEEE Transactions on Image Processing, 2021, 30: 9231–9244.
- [14] HE K, SUN J, TANG X. Single Image Haze Removal Using Dark Channel Prior[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2011, 33(12): 2341–2353.
- [15] SANDLER M, HOWARD A, ZHU M, et al. Mobilenetv2: Inverted residuals and linear bottlenecks[C]. proceedings of the Proceedings of the IEEE conference on computer vision and pattern recognition, F, 2018.
- [16] HOU Q, ZHOU D, FENG J. Coordinate attention for efficient mobile network design[C]. proceedings of the Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, F, 2021.
- [17] VASWANI A, SHAZEER N, PARMAR N, et al. Attention is all you need[J]. Advances in neural information processing systems, 2017: 30.
- [18] SRINIVAS A, LIN T Y, PARMAR N, et al. Bottleneck Transformers for Visual Recognition[C]. proceedings of the 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2021.
- [19] SHAW P, USZKOREIT J, VASWANI A. Self-Attention with Relative Position Representations[C]. proceedings of the NAACL, F, 2018.
- [20] LI K, WANG Y, PENG G, et al. UniFormer: Unified Transformer for Efficient Spatial-Temporal Representation Learning[C]. International Conference on Learning Representations, 2022.
- [21] LIU Z, MAO H, WU C-Y, et al. A convnet for the 2020s[C]. proceedings of the Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, F, 2022.
- [22] KRIZHEVSKY A, SUTSKEVER I, HINTON G E. ImageNet classification with deep convolutional neural networks[J]. Communications of the ACM, 2012, 60: 84–90.
- [23] SIMONYAN K, ZISSERMAN A. Very Deep Convolutional Networks for Large-Scale Image Recognition[J]. CoRR, 2015.
- [24] HE K, ZHANG X, REN S, et al. Deep Residual Learning for Image Recognition[C]. proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016.
- [25] HOWARD A, SANDLER M, CHEN B, et al. Searching for MobileNetV3[C]. proceedings of the 2019 IEEE/CVF International Conference on Computer Vision (ICCV), 2019.
- [26] LIU Z, LIN Y, CAO Y, et al. Swin Transformer: Hierarchical Vision Transformer using Shifted Windows[C]. proceedings of the 2021 IEEE/CVF International Conference on Computer Vision (ICCV), 2021.
- [27] 张斌, 魏维, 高联欣, 等. 基于时空域深度神经网络的野火视频烟雾检测[J]. 计算机应用与软件, 2019, 36(9): 236–242.

Video Smoke Detection based on Smoke Area and Lightweight Model

PU Jianfei, WEI Wei, WU Diyong, CHENG Peng, YUAN Dingshen

(College of Software Engineering, Chengdu University of Information Technology, Chengdu 610225, China)

Abstract: Smoke is a typical feature of early fires, and intelligent detection of smoke can effectively reduce the damage caused by forest fires. In order to achieve early detection of smoke in surveillance video, this paper proposes a smoke detection algorithm based on the smoke area and lightweight model. The algorithm first obtains the motion regions in the video by stacking the residual frames, and then uses the adaptive dark channel mask to further screen the motion regions to obtain suspected smoke blocks. On this basis, this paper proposes a lightweight neural network model for smoke recognition. This model uses the characteristics of convolutional local perception to extract the shallow features of smoke, and in the deep layer of the network, convolution and self-attention are related. The global information of smoke is obtained on the basis of shallow feature maps by comparing the global similarity. The experimental results show that the algorithm in this paper has strong robustness, and it has a good detection effect for long-distance smoke and short-distance smoke.

Keywords: early smoke; dark channel; self-attention mechanism; convolutional neural network