

基于循环神经网络的森林火灾识别研究^{*}

曹毅超¹ 吴泽鹏² 周宇飞² 魏书精² 封晓强¹

(1. 南京恩博科技有限公司, 江苏 南京 210007; 2. 广东省森林培育与保护利用重点实验室 / 广东省林业科学研究院, 广东 广州 510520)

摘要 森林火灾初期的烟雾检测对于预防森林火灾有着重要意义。由于长距离视频监控中通常烟雾运动速度慢、特征不明显, 可靠的烟雾检测仍然是一项十分具有挑战性的任务。以林火视频监控场景为研究对象, 提出了一种基于循环神经网络的森林火灾识别算法。该算法由空间特征提取网络和时序神经网络组成。空间特征提取网络通过深度学习来高效提取烟和火的静态特征。循环神经网络通过中间隐藏层的记忆内容来对空间特征进一步融合, 进而挖掘烟火动态特征。该方法不仅可以从图像局部区域中提取具有区分性的空间特征, 还可以通过循环神经网络来提取时序特征。实验表明, 基于循环神经网络的森林火灾识别算法能够在多种多样的场景下实现 96.3% 的准确率和 4.5% 的误报。

关键词 森林火灾; 烟雾检测; 深度学习; 循环神经网络; 时序特征

中图分类号: S762.3+2 **文献标志码:** A **文章编号:** 2096-2053 (2020) 05-0034-07

Forest Fire Recognition Method Based on Recurrent Neural Network

CAO Yichao¹ WU Zepeng² ZHOU Yufei² WEI Shujing²
FENG Xiaoqiang¹

(1. Nanjing Enbo Technology Co., Ltd, Nanjing, Jiangsu, 210007, China; 2. Provincial Key Laboratory of Silviculture, Protection and Utilization/Guangdong Academy of Forestry, Guangzhou, Guangdong 510520, China)

Abstract Smoke in the early stage of forest is very important to prevent forest fires. Due to the slow moving speed and nonobvious features of smoke in long-distance video surveillance, reliable smoke detection is still a very challenging task. We present a novel forest fire recognition method based on recurrent neural network (RNN) for forest fire video monitoring scenes. This architecture consists of spatial feature extraction network and temporal neural network. Spatial feature extraction network extracts static features of fire smoke efficiently through deep learning. Through the memory of the hidden layer, RNN can further fuse the spatial features and mine the dynamic features of fire smoke. It can not only extract distinguishing spatial features from images, but also extract temporal features by recurrent neural networks. Experiments show that this method can achieve higher accuracy and less false positives in a variety of complex scenes.

Key words forest fire; smoke detection; deep learning; RNN; temporal features

森林火灾是一种突发性强、破坏性高、处置类活动的影响, 森林防火行业面临着愈发严峻的困难的自然灾害。近年来, 受全球气候变化与人 考验, 对森林火灾识别技术提出了新的需求与挑

^{*} 基金项目: 中央财政林业科技推广示范项目 ([2018] GDTK-04)、国家林业科技发展项目 (林业知识产权项目 - 转化应用) (KJZXZC2018006)。

第一作者: 曹毅超 (1993—), 男, 研究员, 主要从事森林火灾监测预警技术研究, E-mail: 745021926@qq.com。

通信作者: 吴泽鹏 (1967—), 男, 正高级工程师, 主要从事森林防火及应急管理研究, E-mail: wzpeng1998@163.com。



图1 林火视频监控系统捕获到的典型早期森林火灾图像

Fig.1 Typical early forest fire image from forest fire video surveillance

战^[1-3]。基于林火视频监控系统的火灾监测方式以其报警时效性强、稳定可靠、信息灵活多样等独特的优势，在森林火灾检测中占据着重要的位置。林火视频监控系统涉及了计算机视觉、嵌入式、模式识别等技术^[4-5]。近年来，随着大数据、机器学习、深度学习为代表的人工智能技术在语音识别、计算机视觉、推荐系统等领域得到成功应用，人工智能在算法、模型、架构等方面取得了巨大进展。以深度学习为代表的人工智能技术也为森林火灾识别提供了新的思路。作为林火视频监控系统核心构件的森林火灾识别算法近年来受到研究人员的广泛关注^[6]。

云台相机非常适合大面积监控，它们可以被设置为自动巡检模式中，自动扫描识别预定的位置。本文提出了一种利用云台摄像机检测森林火灾的新方法。图1显示了由林火视频监控相机捕获的典型森林火灾初期烟雾的图片。早期森林火灾由于树木遮挡和地形缘故，通常以烟雾为主要特征。因此，森林火灾监测系统的核心是对燃烧造成的烟雾进行识别。针对烟雾的识别，研究人员已经做了大量的尝试，研究重点在于描述烟雾的特征。

现有的烟雾探测方法可分为两类：基于图像的烟雾探测^[7-9]和基于视频的烟雾探测^[10-11]。一般的烟雾检测算法通常结合运动检测、特征提取和分类方法。基于图像的烟雾检测方法通常在帧间信息上是独立的，即不会利用上下文信息。基于视频的方法通常不仅分析单帧图像的空间特征，而且提取帧间的时间特征。在某些场景下，我们很难得到稳定可靠的连续图像序列，这时候基于

单帧的检测方法是一种很好的选择。Tian 等人^[7]最近提出用凸优化的方法对图像子块进行分类。Yuan 等人^[12]提出了一种融合三种直方图特征和四种密度特征的双映射框架，将边缘方向直方图、边缘强度直方图和 LBP 直方图、边缘强度分布密度、LBP 位密度、颜色密度和饱和度密度等特征结合起来进行烟雾检测。Yuan 等人^[9]也提出了一种结合 LBP 特征、KPCA 和 GPR 的烟雾检测方法进一步提高识别模型的鲁棒性。李巨虎等人^[13]采用分块 LBP 直方图特征结合 LPQ 直方图直方图特征，来降低森林火灾识别的误报率。近年来，深度学习在图像分类和目标检测方面取得了巨大的成功，也有研究人员尝试使用深度归一化卷积神经网络（DNCNN）来实现特征的自动提取和分类^[8]。然而，运动特征是烟火的本质特征之一。当人眼从视频中分辨烟雾时，动态特征往往被用作关键的参考信息。如果能够有效利用视频序列中的烟火动态特征信息，将有助于进一步提升模型精度。Dimitropoulos 等人^[9]提出了一种用于多维动态纹理分析的高阶线性动力系统（H-LDS）描述方法。也有研究者将深度学习应用到森林烟火检测中。Lin 等人^[11]提出了一种基于 Faster RCNN^[13] 和 3DCNN^[15] 的检测框架。然而，该算法因为计算量较大，实际应用的成本较高。

本文的研究重点在于候选序列的分类。主要创新点如下：本文提出一种用于森林火灾识别的深度学习序列框架；本文采用了基于块的检测方案，使用卷积神经对候选块提取空间特征，使用循环神经网络对序列提取动态特征；通过在大量

实际监控视频数据集上的对比测试,证明了本文提出的方法取得了最好的识别精度和鲁棒性。

1 研究方法及流程

本文使用前景检测算法从视频中提取运动区域的序列,为了扩大烟雾的识别范围(有效地获取运动像素周围的参考信息)和定位,本文采用了基于块的检测方案。针对森林火灾烟雾识别中的候选图像序列分类问题,提出了一种基于循环神经网络的森林火灾识别方法。图 2 展示了本文方法的流程图。

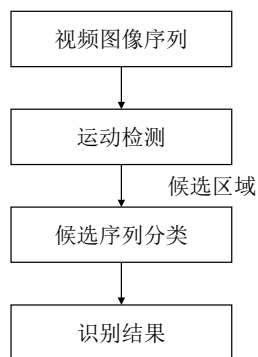


图 2 识别方法流程

Fig.2 Flow chart of the proposed method

本文所提出的识别模型框架如图 3 所示,其包括静态特征提取网络和时序网络两部分,其中时序网络通过循环神经单元进行构建。其中空间特征提取网络依次提取单帧图像的空间特征,时序网络从时间域对空间特征进行融合,学习输入序列的动态特征,最终对输入序列做出分类。

1.1 空间特征提取网络

CNN 已经在多种计算机视觉任务中取得了

优异的成绩。其中 Inception 是常用的通用主干模型^[16]。研究人员后续还提出了改进的 Inception V2^[16] 和 Inception V3^[18]。本文中使用的在 ImageNet 上预训练的 Inception V3 作为主干网络,进行空间特征的提取。Inception V3 是一个精心设计的网络模型,其中将 5×5 卷积分解为两个 3×3 卷积运算,以提高计算速度;并且将 $N \times N$ 卷积分解为 $1 \times N$ 和 $N \times 1$ 卷积的组合,将空间卷积分解为不对称卷积,进一步提高模型性能。在模型优化时使用了 Batch-Normalization 和标签平滑等正则化方式。

本文中,我们将 InceptionV3 模型中的 avg_pool 层作为空间特征的输出层。该层的输出维度为 2048 维,随后会将每帧图像提取得到的 2048 维空间特征输入到循环神经网络中提取动态特征。

1.2 时序网络的设计

循环神经网络是前向神经网络的一种扩展,其特点在于模型增加了隐含层的有向环结构,用于对序列数据进行时序建模。循环神经网络前向计算时,对于每个输入状态都会结合当前模型中的隐含状态进行输出,其当前时刻的状态是根据上一时刻的状态和当前时刻的输入来决定。从其结构特点可以看出,循环神经网络是处理这类序列数据最自然的网络结构。本文中,空间特征提取网络提取了高维抽象的空间特征,按时间顺序排列输入到循环神经网络中。

如图 4 所示,左侧的循环神经网络相比一般的全连接神经网络,在隐含层上增加了有向环 W_{ah} 的循环计算。其中黑色的块代表时间维度一个步

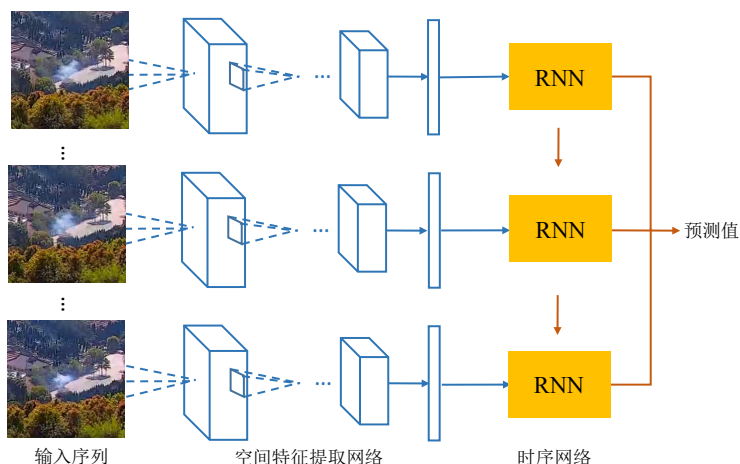


图 3 基于循环神经网络的多帧识别模型

Fig.3 Sequence recognition model based on recurrent neural network

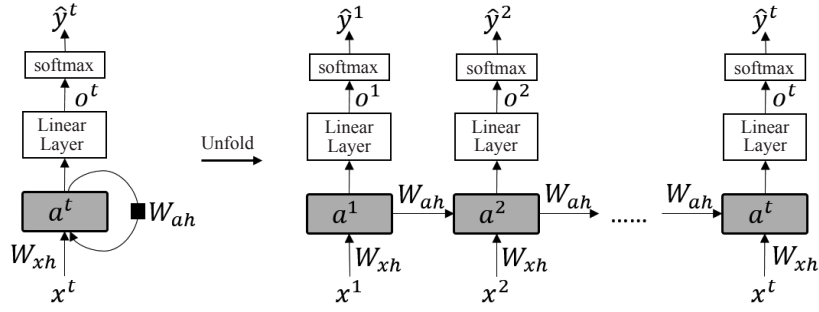


图4 循环神经网络结构示意图

Fig.4 Structure of recurrent neural network

长的延迟。每个步长的输入信号 x^t 结合上一时刻的隐含层状态 a^t ，最终通过全连接层和 softmax 层得到分类结果。对于图 4 所示的循环神经网络，按照时间顺序展开后，输入信号 x^t 是按照时间与隐含层状态 a^t 和输出状态 y^t 一一对应。对模型评估时，可以逐个时间步的输出计算损失。

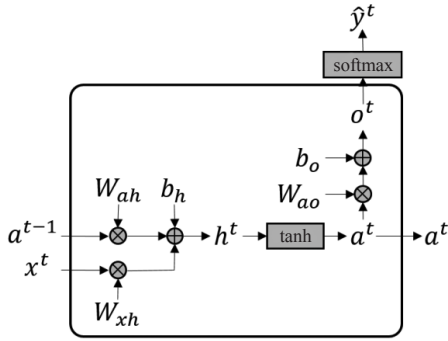


图5 循环神经单元的结构

Fig.5 Structure of recurrent neural network unit

在本文中，设循环神经网络的输入节点数为 K ，隐含层节点数为 N ，输出节点数为 L 。图 5 展示了循环神经单元的结构图，对于 t 时刻，假设输入向量为 $x^t_{K \times 1}$ ，隐含层状态为 $h^t_{N \times 1}$ ，通过激活函数和全连接层的变换后，通过 softmax 函数得到输出 $\hat{y}^t_{L \times 1}$ 。在任意时刻 t ，循环神经网络的信息流动可用如下公式计算：

$$h^t = W_{xh}x^t + W_{ah}a^{t-1} + b_h \quad (1)$$

$$a^t = \tanh(h^t) \quad (2)$$

$$o^t = W_{ao}a^t + b_o \quad (3)$$

$$\hat{y}^t = \text{softmax}(o^t) \quad (4)$$

其中， $(W_{xh})_{N \times K}$ 是连接 K 个输入节点和 N 个

隐含节点的权重矩阵。 $(W_{ah})_{N \times N}$ 是连接 $(t-1)$ 时刻 N 个隐含节点到 t 时刻 N 个隐含节点的权重矩阵。 $(W_{ah})_{N \times N}$ 是循环神经网络中的关键步骤，保证了当前信息与之前状态的信息融合。 $(a^t)_{N \times 1}$ 是激活后的隐含层状态向量。 $(o^t)_{L \times 1}$ 为输出节点的状态向量，经 softmax 计算后转换为输出标签向量 $(\hat{y}^t)_{L \times 1}$ 。本文研究的森林火灾识别是序列分类问题，因此最终类别数 L 为 2，对应发生火灾和不发生火灾的概率。

1.3 时序网络的训练方法

对于循环神经网络的参数优化，可由反向传播推广得到。模型对于不同的输入序列，会产生相应的预测概率，通过损失函数对输出进行评价后，由反向传播算法计算误差并对模型参数进行优化。与一般前向计算的神经网络区别在于，循环神经网络需要考虑时间维度的反向传播，因此该优化算法也被称为随时间反向传播 (Back Propagation Through Time, BPTT)。对于图 4 所示的网络结构，最终需要优化的参数有 W_{ao} 、 b_o 、 W_{ah} 、 b_h 和 W_{xh} 5 个参数矩阵。对于本文的分类问题，可以使用经典的交叉熵损失函数进行评估。与一般的分类任务相比，本文增加了时间维度，因此可以逐个时间点地统计损失函数：

$$L^t = -\sum_i y_i^t \ln(\hat{y}_i^t) \quad (5)$$

对于 k 类的图片序列，损失函数可以细化为：

$$L^t = -\sum_i y_i^t \ln(\hat{y}_i^t) = -y_i^t \ln(\hat{y}_i^t) + \sum_{i \neq k} -y_k^t \ln(\hat{y}_k^t) \quad (6)$$

其中， y_i^t 表示第 i 类的真实概率， $\hat{y}_i^t$ 表示第 i 类的预测概率， y_k^t 表示第 k 类的真实概率， $\hat{y}_k^t$ 表

示第 k 类的预测概率。

2 对比实验设置

2.1 数据集

据本文调研，目前还没有公开的大规模森林火灾数据集进行模型训练和测试。我们建立了一个大规模的森林火灾烟雾视频数据集，采集了大量真实的早期森林火灾视频来创建我们的数据集，所有视频都是从 1920 × 1080 的森林火灾监控系统中采集的。模型的输入是由 VIBE^[19] 前景检测算法从原始视频中截取的 299 × 299 大小的烟雾候选块序列。在林火视频监控系统中，对前景区域大于一定阈值的子块进行采样，本文中选择 5FPS 的采样频率。序列样本图如图 6 所示。

数据集中序列总数为 2 000 个，其中含烟序列 1 000 个，非烟序列 1 000 个。每个序列的长度为 20 帧，图像分辨率为 299 × 299，每秒 5 帧。为了进行训练和测试，数据集 80% 作为训练集，20% 作为测试集。数据集的详细信息如表 1 所示。

按照样本预测类别和真实类别的差异，可将

样本分为真阳性 (True Positive, TP)、假阳性 (False Positive, FP)、真阴性 (True Negative, TN) 和假阴性 (False Negative, FN) 四类。具体分类如下：真阳性：正样本被正确判别为有烟火；

假阳性：负样本被错误判别为有烟火；

真阴性：负样本被正确判别为无烟火；

假阴性：正样本被错误判别为无烟火；

序列二分类模型性能可用真阳性率 (True Positive Rate, TPR)，真 阴 性 率 (True Negative Rate, TNR) 和准确率 (Accuracy Rate, AR) 几个指标进行评估。TPR 主要用于评估对于森林火灾的捕获率，TNR 主要用于评估对于正常非火灾样本的识别能力，AR 表示模型的综合识别能力。3 个评价指标的计算公式如下：

$$TPR=\frac{TP}{TP+FN} \dots\dots\dots(7)$$

$$TNR=\frac{TN}{TN+FP} \dots\dots\dots(8)$$

$$AR=\frac{TP+TN}{P+N} \dots\dots\dots(9)$$



注：第一行到第四行展示了 4 次早期森林火灾样本，每列图像的采样时间间隔是 200 ms。
Note: row 1 to row 4 show samples of four early forest fires, each with a sampling interval of 200 milliseconds

图 6 林火识别任务中的序列样本

Fig.6 Sequence samples in forest fire recognition task

表 1 序列数据集构成
Tab.1 Composition of sequence dataset

类型 Type	序列数目 Quantity	帧数 Frames	用途 Purpose
正类（有烟）	800	16 000	训练集
负类（无烟）	800	16 000	训练集
正类（有烟）	200	4 000	测试集
负类（无烟）	200	4 000	测试集

2.2 模型训练与优化

本文实验在 Intel Core i5-6500 和 NVIDIA GTX1080 的工作站上进行。本文所提出的模型基于 Keras 实现。在大多数文献中,研究人员通常是基于单帧图像计算模型精度。在我们的工作中,模型的识别精度是基于序列进行统计的,因为本文所提方法是针对序列图像进行识别。

本文所提出的识别模型包括静态特征提取网络和时序网络两部分。在第一阶段,我们使用 Adam 算法对 Inception V3 模型进行训练优化,使用 0.000 01 的学习率和 32 的 batchsize。我们将 Inception V3 模型最后的分类器修改为二分类的全连接层后,对预训练模型进行迁移学习。在第二个阶段,将 avg_pool 层的输出作为每帧图像的空间特征输入到循环神经网络中进行训练。我们使用 RMSprop 算法对时序网络部分进行训练优化,同样使用 0.000 01 的学习率和 32 的 batchsize。

3 结果与分析

本节对比实验前,首先使用测试集挑选了最合适的参数,包括学习率、节点数等。为了验证本文模型的有效性,我们将本文的模型与多种方法进行了对比。其中包括单帧的 Inception V3^[18] 深度学习方法,3DCNN^[15] 多帧深度学习方法和传统烟火识别方法^[20]。

从表 2 可知,我们提出的基于循环神经网络的识别方法实现了 96.3% 的识别精度,97.0% 的正类识别率和 95.5% 的负类识别率。本文所提的方法在测试集上整体优于单帧深度模型、3DCNN 和^[20]中的方法。实验结果表明,通过循环神经网络提取动态特征后,相比 Inception V3^[18] 模型识别精度提升了 2.2 个百分点。算法在保证较高的烟雾捕获率同时,也能有效过滤非烟雾的图像序列。该实验结果验证了本文的设想,通过循环神经网络确实提高了卷积神经网络模型对于烟雾动态特

征的识别能力。

4 结论与讨论

林火监控视频的场景复杂,图像中的干扰因素众多。传统的烟火识别算法通常对疑似的烟雾区域提取特定的特征后,对特征进行分类识别。基于规则的识别方法为手工挑选的特征选取阈值和规则,对目标进行判别;基于学习的方法则通过分类器对特征进行识别。烟雾作为一种随机的湍流运动现象,受林区天气、地形等因素影响,外观差异较大且存在一定的透明度,人工挑选的特征常常鉴别力不足,泛化能力较低。

近年来以卷积神经网络为代表的深度学习技术得到快速发展,本文尝试通过卷积神经网络对单帧局部图像进行特征提取。基于 Inception V3 的深度学习模型,通过在 ImageNet 数据集上的预训练,拥有比传统特征提取算法更鲁棒的特征提取能力。为了对算法模型进行训练与测试,本文建立了一个大规模的森林火灾烟雾视频数据集来对算法进行评估。

从单帧图像中识别烟火目标面临的另一个挑战是单帧图像缺乏烟火的运动信息,然而运动特征是烟火的本质特征之一。因此本文进一步探究了通过循环设计网络学习动态特征,设计了一种针对序列图像的森林火灾识别方法。识别模型包括空间特征提取网络和时序网络两个部分。其中以 Inception V3 为主干网络用来对每帧疑似图像提取静态特征,以循环神经单元为基础构建的时序网络对静态特征融合与分析,学习序列图像中的动态特征。时序网络通过循环神经网络结构中的有向环设计,赋予了模型学习与表示时间序列的能力。通过循环神经网络提取烟火动态特征后,本文提出的多帧识别模型相比于 Inception V3 的单帧识别模型,精度提升了 2.2 个百分点。实验结果证明了本文模型设计的合理性。

表 2 数据集上的测试结果
Tab.2 Test results on dataset

方法 Methods	召回率 TPR	负正类率 TNR	精度 AR
Method in [20]	88.1	89.1	88.6
3DCNN [15]	95.4	95.0	95.2
Inception V3[18]	93.7	94.5	94.1
Our Method	97.0	95.5	96.3

未来我们会对算法的特征提取和时序融合部分进一步展开研究,如使用空间和时间注意力机制进一步提升模型对于烟火目标的关注度。

参考文献

- [1] 魏书精, 罗斯生, 罗碧珍, 等. 气候变化背景下森林火灾发生规律研究[J]. 林业与环境科学, 2020, 36(2): 133-143.
- [2] 王振师, 魏书精, 谢继红, 等. 林火燃烧环境对灭火效果的影响研究[J]. 林业与环境科学, 2019, 35(2): 84-88.
- [3] 孙红斌, 王佐霖, 朱子维, 等. 深圳市坝光湿地园银叶树群落优势树种与土壤生态化学计量特征分析[J]. 林业与环境科学, 2018, 34(6): 15-20.
- [4] 钟映霞, 周宇飞, 陈世清, 等. 基于无人机遥感的森林火烧迹地面积调查应用[J]. 林业与环境科学, 2019, 35(2): 89-93.
- [5] 周宇飞, 曾宇, 李小川, 等. 红外热成像仪林火识别性能测评技术研究[J]. 林业与环境科学, 2019, 35(6): 27-31.
- [6] 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.
- [7] TIAN H, LI W, OGUNBONA P O, et al. Detection and separation of smoke from single image frames[J]. IEEE Transactions on Image Processing, 2018, 27(3): 1164 - 1177.
- [8] YIN Z, WAN B, YUAN F, et al. A Deep normalization and convolutional neural network for image smoke detection[J]. IEEE Access, 2017(5): 18429 - 18438.
- [9] YUAN F, XIA X, SHI J, et al. Non-linear dimensionality reduction and gaussian process based classification method for smoke detection[J]. IEEE Access, 2017(5): 6833 - 6841.
- [10] DIMITROPOULOS K, BARMPOUTIS P, GRAMMALIDIS N. Higher order linear dynamical systems for smoke detection in video surveillance applications[J]. IEEE Transactions on Circuits & Systems for Video Technology, 2017, 27(5): 1143 - 1154.
- [11] LIN G, ZHANF Y, XU G, et al. Smoke detection on video sequences using 3d convolutional neural networks[J]. Fire Technology, 2019, 55(5): 1827 - 1847.
- [12] YUAN F. A double mapping framework for extraction of shape-invariant features based on multi-scale partitions with adaboost for video smoke detection[J]. Pattern Recognition, 2012, 45 (12): 4326 - 4336.
- [13] 李巨虎, 范睿先, 陈志泊. 基于颜色和纹理特征的森林火灾图像识别[J]. 华南理工大学学报(自然科学版), 2020, 48(1): 70-83.
- [14] REN S, HE K, 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, 2015, 39(6): 91 - 99.
- [15] TRAN D, BOURDEV L, FERGUS R, et al. Learning spatiotemporal features with 3D convolutional networks[C]// ICCV. 2015 IEEE International Conference on Computer Vision. Santiago: Institute of Electrical and Electronics Engineers Inc, 2015: 4489 - 4497.
- [16] SZEGEDY C, LIU W, JIA Y, et al. Going deeper with convolutions[C]// CVPR. 2015 IEEE Conference on Computer Vision and Pattern Recognition. Boston: IEEE, 2015: 1 - 9.
- [17] IOFFE S, SZEGEDY C. Batch Normalization: Accelerating deep network training by reducing internal covariate shift[J]. Proceedings of the 32nd International Conference on International Conference on Machine Learning, 2015, 37: 448-456.
- [18] SZEGEDY C, VANHOUCKE V, IOFFE S, et al. Rethinking the inception architecture for computer vision[C]// CVPR. 2016 IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas: IEEE, 2016: 2818-2826.
- [19] BARNICH O, DROOGENBROECK M V. ViBe: A universal background subtraction algorithm for video sequences[J]. IEEE Transactions on Image Processing, 2011, 20(6): 1709-1724.
- [20] YUAN F. Video-based smoke detection with histogram sequence of LBP and LBPV pyramids[J]. Fire Safety Journal, 2011, 46 (3): 132-139.