基于级联卷积神经网络的视频动态烟雾检测

陈俊周，汪子杰，陈洪瀚，左林翼

(西南交通大学信息科学与技术学院 成都 610031)

【摘要】复杂场景中烟雾特性的提取是目前视频烟雾检测领域的主要挑战。针对该问题，提出一种静态和动态特征结合的卷积神经网络视频烟雾检测框架。在静态单帧图像特征检测的基础上，进一步分析其时空域上的动态纹理信息以期克服复杂的环境干扰。实验结果显示，该级联卷积神经网络模型可有效应用于复杂视频场景中烟雾事件的实时检测。

关 键 词 卷积神经网络; 深度学习; 纹理特征; 视频烟雾检测

中图分类号 TP391 **文献标志码** A doi:10.3969/j.issn.1001-0548.2016.06.020

Dynamic Smoke Detection Using Cascaded Convolutional Neural Network for Surveillance Videos

CHEN Jun-zhou, WANG Zi-jie, CHEN Hong-han, and ZUO Lin-yi

(College of Information Science & Technology, Southwest Jiaotong University Chengdu 610031)

Abstract The extraction of stable smoke features in complex scenes is a challenging task for video based smoke detection. For this issue, a convolutional neural network (CNN) framework which employs both static and dynamic features of the smoke is proposed. On the basis of analyzing the static features of individual frame, we further explore the dynamic features in spatial-temporal domain to reduce the influence of the noise from environment. Experimental results show that the proposed cascaded convolutional neural network framework performs well in real-time video based smoke detection for complex scenes.

Key words convolutional neural networks; deep learning; texture features; video smoke detection

烟雾检测作为消防探测重要手段，已广泛应用于火灾、爆炸的探测与预警。传统基于烟雾传感器的探测技术监测范围小，在工厂、仓库、森林等较大的区域铺设成本高，且此类传感器易老化而灵敏度降低。近年来，视频烟雾探测技术因其响应时间短、灵敏度高、覆盖面积大等优势备受国内外研究者关注。

现有视频烟雾检测方法主要依靠运动、颜色、形状、透明度、纹理等视觉特征。文献[1]提出一种利用烟雾颜色和运动特征的检测方法，首先采用背景提取和颜色过滤获取候选烟雾区域，然后计算光流将其速度及方向的均值和方差作为特征，最后采用BP神经网络完成分类识别，其所获特征向量维度偏低难以有效描述烟雾在复杂环境下的不同表现形式。文献[2]提出积累运动模型并利用积分图快速估计烟雾运动方向，该方法假设烟雾做向上运动其适用范围较有限。随后，文献[3]又提出一种双映射框架特征与AdaBoost结合的烟雾检测方法。第一层映射将每帧图像分块，提取各图像块的边缘方向直方图、边缘强度直方图、LBP直方图、边缘强度密度以及颜色和饱和度密度等特征。第二层映射将图像分区，统计各区块特征的均值、方差、峰态、偏度等。这些统计量最终被用于AdaBoost模型的训练和分类。文献[4]针对固定摄像头视频，提出一种基于轮廓和小波变换的烟雾探测方法，隐马尔科夫模型(HMM)被用于分析烟雾轮廓时域上周期性的变化。烟雾通常具有一定的透明度，其视觉特征受到背景影响，若能克服背景干扰则可有效降低烟雾识别难度。针对这一问题，文献[5]分析了烟雾与背景的混合机制，构建了一套烟雾前景提取方法，利用稀疏表达、局部平滑等约束求解混合系数。该方法可在一定程度上降低背景干扰，提高烟雾识别准确率。

收稿日期：2015 – 12 – 17；修回日期：2016 – 06 – 17
基金项目：国家自然科学基金(61003143, 61202191)
作者简介：陈俊周(1979 –)，男，博士，副教授，主要从事计算机视觉、模式识别、机器学习方面研究.

在烟雾纹理特征提取方面，GLCM、LBP、Wavelet等应用最为广泛。文献[6]基于GLCM分析烟雾纹理实现了一套火焰、烟雾实时检测系统。文献[7]引入LBP提取烟雾纹理特征。文献[8]提出一种基于金字塔直方图序列烟雾检测方法。首先金字塔采样为三层多尺度结构，对每一层图像提取不同模式的LBP及LBPV特征，最后将LBP和LBPV特征序列拼接作为烟雾纹理特征，并由BP神经网络进行分类。然而，实际应用中现有方法均存在较多误检，主要原因在于：1) 烟雾在不同环境下呈现出多样的状态，现有文献选用数据集较小，难以训练出稳定、可靠的分类器以拟合其复杂表现形式。2) 烟雾视觉特征提取一直是视频烟雾检测的难点，仅依赖静态特征不足于将烟雾与一些似烟对象区分(如：云、喷泉等)。如何构建稳定、高效的特征提取算法，融合视频中静态与动态信息，成为降低烟雾误检的关键。

传统的分类器如SVM、决策树等在小数据集中表现良好，但在数据量较大时却难以更好地提高分类精度。近年来，深度神经网络(deep neural network, DNN)被成功地应用于计算机视觉领域。DNN通过建立类似于人脑的分层网络模型结构，对输入数据逐级提取从底层到高层的特征，以便更好地获得从底层信号到高层语义的映射关系。卷积神经网络(convolutional neural networks, CNN)作为其中最重要的网络模型之一，伴随大数据和高性能计算的驱动，在人脸识别、图像分类等方面取得突破性进展。文献[9]首次将CNN引入手写数字识别，其提出的LeNet网络结构被美国银行业广泛用于支票识别，并成为小尺度图像识别的基础模型。2012年，文献[10]在著名的ImageNet图像数据集上用更深的CNN取得当年世界最好结果，将识别错误率从26%降到15%，大幅度提升了大规模图像识别的精度。此后，更多的基于深度卷积神经网络模型和方法[11-12]被提出，并向人脸识别[13]、行人检测、行为识别[14]等分支发展。深度的卷积神经网络能以原始图像作为输入，学习到从底层像素级到高层表示级的特征，将人工提取特征的模式向从数据中自动学习特征的模式转变。并且，该模型在大数据上效果更为显著。本文将卷积神经网络引入烟雾纹理特征提取，提出一种级联的卷积神经网络烟雾纹理识别框架融合静态和动态纹理信息，在静态纹理上将原始图像作为输入，在动态纹理上将原始图像的光流序列作为输入，最终实验结果显示，本文方法在烟雾识别准确率和误检率上均取得更好表现。

1　方法

1.1　视频预处理

监控系统中摄像机通常处于静止状态，在视频场景中包含大量静止背景，而烟雾属视频中运动前景。为了达到系统实时性的要求，对视频做预处理，过滤掉静止区域至关重要。由于烟雾在视频图像中呈现不规则的形态，本文采用分块检测的方法，将视频图像划分为固定大小的块用于CNN的输入。具体实现中，将每帧图像划分为24×24不重叠的小块，采用帧间差分法滤除其中静止块，将剩余运动块作为候选烟雾区域。这些候选块可能为烟雾或非烟(如：运动的行人、车辆以及植物等)运动区域。

为此，本文提出一种级联的卷积神经网络框架以检测视频中的烟雾事件。

1.2　烟雾识别系统框架

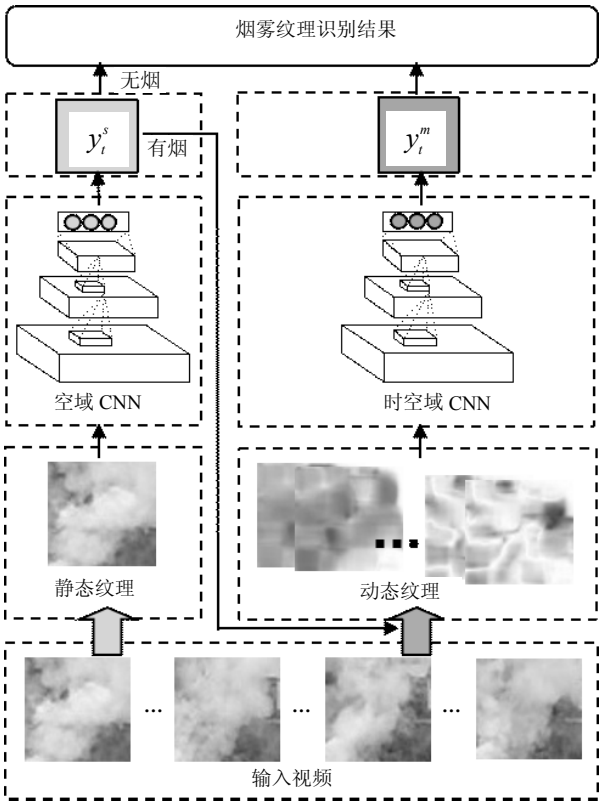

图1　烟雾纹理识别框架

本文提出的视频烟雾识别系统由两部分级联组成：静态烟雾纹理识别网络(对应空域CNN)和动态烟雾纹理识别网络(对应时空域CNN)，系统整体框架如图1所示。被检视频经预处理后，提取当前帧候选烟雾块区域输入空域CNN提取其静态特征进行第一步判别。在实际检测过程中，由于可能存在某些与烟雾外观相似的运动干扰(如：喷泉、云等)，会有

一部分无烟的图像块被误检为有烟，造成虚警。为此，当候选烟雾块被空域CNN识别为有烟时，提取该块对应区域过去连续多帧图像块序列，计算其对应的光流序列作为第二级时空域CNN的输入，分析该候选块区域的时空域动态特性，以进一步降低误检。

1.3 静态纹理特征

识别烟雾静态纹理特征的空域CNN模型包括6层，网络构建过程如图2所示。网络的输入层为24×24的RGB图像，灰度级范围为[0, 255]，将其归一化至[0, 1]以适应神经网络训练的需要。输入层连接第一层卷积层，卷积的滤波器参数设置20×3×5×5，得到C_1的通道数为20，大小为20×20。接下来连接第一层下采样层，采用Max pooling的下采样方法，核大小为2×2，得到S_1的通道数为20大小为10×10。第二层卷积层滤波器参数设置为50×20×5×5，得到C_2的通道数为50，大小为6×6。再次连接下采样层，和第一次下采样一样的方法，得到S_2的通道数为50，大小为3×3。最后两层全连接层，F_1的神经元个数为100，最后输出层得到识别结果。

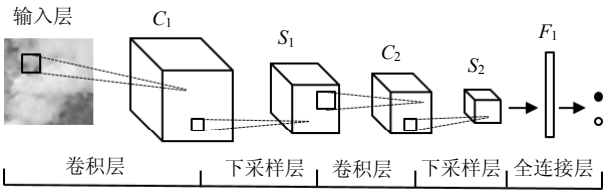

图2 本文CNN的网络结构

1.4 动态纹理特征

在每一帧静态纹理识别的基础上，将候选烟雾块(包含有烟和无烟误检为有烟)以及历史帧中相应位置的图像块作为动态部分检测数据。光流是一种图像中像素级运动的表示方法，文献[14]将光流图像作为CNN的输入来识别视频中的动作。受此启发，本文将时间连续的烟雾块的光流序列作为识别烟雾动态纹理特征的时空域CNN的输入。稠密的光流可以看成是连续两帧t，t+1时刻的位移向量场d_t，$d_t(u,v)$表示点(u,v)从t时刻到t+1时刻的位移矢量，d_t^x和d_t^y分别表示水平和垂直方向上的分量。为了表示一个帧序列的运动，将连续L帧$d_t^{x,y}$重叠起来形成通道为$2L$的光流序列作为输入。将w、h表示成输入的宽度和高度，那么时空域CNN的输入$I_\tau \in R^{w\times h\times 2L}$为：

$$I_\tau(u,v,2k-1)=d^x_{\tau+k-1}(u,v)$$
$$I_\tau(u,v,2k)=d^y_{\tau+k-1}(u,v)$$
$$u\in[1,w],v\in[1,h],k\in[1,L]$$

对任意点(u,v)，通道$I_\tau(u,v,c)$，$c\in[1,2L]$表示L帧序列的运动编码。

时空域CNN通过预先训练好的模型判断输入的图像块是否属于真实烟雾块。在时空域CNN的网络结构中，输入图像大小不变，但通道数由RGB的3通道变成$2L$。CNN中卷积的一个滤波器可以理解为提取图像中一种特征，由于输入层通道数增多，提取的特征数也增多了。本文在训练模型时发现，适度增加时空域CNN各层特征图数量，能够提高模型准确率，但时间复杂度也相应增加了。综合考虑模型精度与时间成本，确定了时空域CNN第一次卷积下采样后通道数为40，第二次卷积下采样后为80。全连接层的神经元个数不变。具体实验中，L的大小取5，此时输入通道数为10。

2 实验结果与分析

本文的CNN模型采用Caffe[15]训练，网络的模型结构和训练参数主要参考LeNet[9]。Caffe是采用C++与CUDA实现的深度学习框架，具有模型描述简单、代码易扩展、速度快等优点，被学术界与工业界广泛使用。LeNet是一个6层的CNN结构，包含两层卷积层、两层下采样层以及两层全连接层，其在手写字符识别经典数据集MNIST上达到99%以上的识别准确率。本文CNN输入图像大小为24×24，与LeNet的输入图像大小28×28相似，故本文在LeNet的基础上，更改了输入和输出的结构，在时空域CNN上更改了每一层通道数。

2.1 数据集

静态纹理数据集包含正负样本各30 000的24×24的单幅图像，其中有烟雾部分图片从有烟视频(不与测试视频重复)中获取，无烟部分图片从Caltech101[16]数据集中选择无烟雾背景图片里截取。随机选择80%作为训练集，余下20%作为测试集。

动态纹理数据集包含正负样本各30 000的6×24×24帧序列(一个序列连续6帧，即网络的输入是10个通道的光流序列)，其中有烟部分从有烟视频中截取，无烟部分从UCF-101[17]数据集和部分无烟雾视频里选择无烟雾运动部分截取(均不与测试视频重复)。随机选择80%作为训练集，20%作为测试集。

完整视频数据集包含视频样本20个。其中有烟视频10个，无烟视频10个。

2.2 评价指标

为验证算法的有效性，实验的指标为：

$$\mathrm{ACC}=(\mathrm{TP}+\mathrm{TN})/N \tag{1}$$

$$TPR = TP/(TP + FN) \quad (2)$$

$$TNR = TN/(TN + FP) \quad (3)$$

式中，ACC为准确率；TPR为真正率；TNR为真负率；N为总样本数；TP为真正样本数；TN为真负样本数；FP为假正样本数；FN为假负样本数。

此外，对于完整视频烟雾检测的评价指标还包含两个指标：针对有烟的视频，第一次发出烟雾警报的帧号FAFSV(first alarm for smoke video)，该值越小说明越早报警；针对无烟的视频，整个视频中误检的帧数FAFNSV(false alarms for non-smoke video)越小说明鲁棒性越好。本文分别比较了10个非烟视频和10个有烟视频，部分视频如图3所示。

图3 部分实验视频

2.3 实验结果

单独使用静态纹理识别实验：静态纹理识别部分，将本文方法与LBP+SVM方法进行了实验对比。LBP[18]是一种纹理特征描述方式。LBP统计图像中每个像素与其邻域像素的亮度关系，并将其统计成直方图，从而能有效的描述一副图像的纹理特征。本文的静态纹理识别对比LBP+SVM的结果如表1所示，结果表明，采用卷积神经网络对烟雾静态纹理具有更好的识别效果，准确率从93.43%提高到99.0%。然而发现通过静态纹理识别后烟雾的误检率较高，分别为7.09%(LBP+SVM)和1.78%(本文方法)。静态纹理误检的原因是：有许多在静态纹理上类似于烟雾的图像(如：云、喷泉、颜色灰暗的区域等)，而非烟雾的情况远多于有烟雾的情况。

表1　单独使用静态纹理的识别结果

方法	ACC	TPR	TNR
LBP+SVM	93.43%	93.95%	92.91%
本文静态纹理	99.0%	99.76%	98.22%

按照本文测试视频的大小，每一帧将会划分成130个小块，这样最终视频对非烟雾的误检比较高。因此，进一步的动态纹理检测至关重要。

单独使用动态纹理识别实验：动态纹理识别部分，将本文方法与LBP-TOP+SVM方法进行了实验对比。LBP-TOP[19]是一种动态纹理特征提取方法，它是将LBP特征扩展到3维空间上，具有良好的动态纹理表示特性。基于动态纹理数据集的实验结果如表2所示，相对于LBP-TOP+SVM的识别方法，卷积神经网络对烟雾动态纹理的识别具有更好的效果，在准确率上提高了0.82%，并且在真负率上提高了1.54%，这表明本文方法在保证正检率的同时减少了误检率。同时在后续完整视频数据集的测试上，发现将动态纹理与静态纹理相结合的误检率相对于仅使用静态纹理的误检率大大降低，可见烟雾的动态特征可有效地作为静态特征的补充。

表2　动态纹理的识别结果

方法	ACC/%	TPR/%	TNR/%
LBP-TOP+SVM	97.16	97.17	97.15
本文动态纹理	97.98	97.28	98.69

完整视频的检测实验：对于完整的监控视频，将本文方法(静态纹理与动态纹理相结合)与文献[3]的方法进行了实验对比。非烟视频误检帧数实验结果见表3所示，本文方法的FAFNSV低于文献[3]的方法，说明本文方法有效减少了非烟雾区域的误检率。特别是对于视频2、3、5、9、10，本文方法有效避免了虚警发生。有烟视频首次报警帧号实验结果见表4，本文方法的FAFSV均低于文献[3]方法，说明本文方法能够更早地发现视频中的烟雾事件及时预警、降低火灾带来的危害。(本文动态纹理识别需要连续6帧视频以计算光流序列作为输入，故本文烟雾报警最早从第6帧开始。)

表3　非烟视频误检帧数

视频ID	类型	文献[3]的方法	本文方法
1	非烟	18	14
2	非烟	114	0
3	非烟	4	0
4	非烟	23	1
5	非烟	1	0
6	非烟	0	0
7	非烟	0	0
8	非烟	67	10
9	非烟	6	0
10	非烟	2	0

表4 有烟视频首次报警帧号

视频ID	类型	文献[3]的方法	本文方法
11	有烟	8	6
12	有烟	93	6
13	有烟	7	6
14	有烟	16	27
15	有烟	7	6
16	有烟	110	20
17	有烟	7	6
18	有烟	7	6
19	有烟	7	6
20	有烟	8	6

3 结 束 语

本文提出一种基于级联CNN烟雾纹理识别框架视频烟雾检测方法，与传统方法相比，该方法在有效降低了对非烟视频误检的同时，可确保对有烟视频的中的烟雾事件及时检测和报警。本文系统采用C++编写，基于Caffe[15]对CNN网络进行训练和测试，并利用GPU加速，其运行速度可达到实时烟雾检测的需要。

参 考 文 献

[1] YU Chun-yu, FANG Jun, WANG Jin-jun, et al. Video fire smoke detection using motion and color features[J]. Fire Technology, 2010, 46(3): 651-663.

[2] YUAN F. A fast accumulative motion orientation model based on integral image for video smoke detection[J]. Pattern Recognition Letters, 2008, 29(7): 925-932.

[3] 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.

[4] TOREYIN B U, DEDEOGLU Y. Contour based smoke detection in video using wavelets[C]//14th European Signal Processing Conference. [S.l.]: IEEE, 2006: 1-5.

[5] TIAN H, LI W, WANG L, et al. Smoke detection in video: an image separation approach[J]. International journal of computer vision, 2014, 106(2): 192-209.

[6] YU Chun-yu, ZHANG Yong-ming, FANG Jun, et al. Texture analysis of smoke for real-time fire detection[C]// Second International Workshop on Computer Science and Engineering, WCSE'09. [S.l.]: IEEE, 2009, 2: 511-515.

[7] TIAN H, LI W, OGUNBONA P, et al. Smoke detection in videos using non-redundant local binary pattern-based features[C]//2011 13th IEEE International Workshop on Multimedia Signal Processing (MMSP). [S.l.]: IEEE, 2011: 1-4.

[8] YUAN F. Video-based smoke detection with histogram sequence of LBP and LBPV pyramids[J]. Fire Safety Journal, 2011, 46(3): 132-139.

[9] LÉCUN Y, BOTTOU L, BENGIO Y, et al. Gradient-based learning applied to document recognition[J]. Proceedings of the IEEE, 1998, 86(11): 2278-2324.

[10] KRIZHEVSKY A, SUTSKEVER I, HINTON G E. Imagenet classification with deep convolutional neural networks[C]//Advances in Neural Information Processing Systems. Lake Tahoe, USA: [s.n.], 2012: 1097-1105.

[11] SZEGEDY C, LIU W, JIA Y, et al. Going deeper with convolutions[C]//2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). [S.l.]: IEEE Computer Society, 2014: 1-9.

[12] HE K, ZHANG X, REN S, et al. Deep residual learning for image recognition[EB/OL]. (2015-12-10). http://arxiv.org/abs/1512.03385.

[13] YIN Q, CAO Z, JIANG Y, et al. Learning deep face representation: U.S, Patent 20,150,347,820[P]. 2015-12-03.

[14] ANNANE D, CHEVROLET J C, CHEVRET S, et al. Two-stream convolutional networks for action recognition in videos[J]. Advances in Neural Information Processing Systems, 2014, 1(4): 568-576.

[15] JIA Y, SHELHAMER E, DONAHUE J, et al. Caffe: Convolutional architecture for fast feature embedding [EB/OL]. (2014-06-20). http://arxiv.org/abs/1408.5093.

[16] LI Fei-fei, FERGUS R, PERONA P. Learning generative visual models from few training examples: an incremental Bayesian approach tested on 101 object categories[C]// Computer Vision and Image Understanding. [S.l.]: Elsevier, 2004, 106(1): 59-70.

[17] SOOMRO K, ZAMIR R A, SHAH M. UCF101: a dataset of 101 human action classes from videos in the wild [EB/OL]. (2012-12-03). http://arxiv.org/abs/1212.0402.

[18] OJALA T. Multiresolution gray-scale and rotation invariant texture classification with local binary patterns[J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2002, 24(7): 971-987.

[19] ZHAO G, PIETIKÄINEN M. Dynamic texture recognition using local binary patterns with an application to facial expressions[J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2007, 29(6): 915-928.

编 辑 蒋 晓