

基于神经网络的交互式异常行为识别研究

关百胜^{1,2}, 卞春江¹, 冯水春¹, 陈实¹

(1. 中国科学院国家空间科学中心 北京 100190; 2. 中国科学院大学 北京 100094)

摘要: 在智能安防领域中, 异常行为识别一直是该领域的一个研究热点。但目前的研究技术仅能够识别简单的行为特征, 对于复杂行为的识别还存在诸多问题。随着人工智能技术的不断发展以及在行为识别上的应用, 可以尝试通过深度学习技术对监控视频中的异常行为进行识别。因此本文针对监控视频中的交互式异常行为进行分析研究, 首先通过背景减除算法对视频数据进行预处理, 然后采用深度学习技术在训练数据集上进行算法模型的训练, 最后在测试数据集上进行算法模型的验证。该算法模型能够有效区分异常行为和正常行为, 识别准确率为73%。

关键词: 交互式; 异常; 行为识别; 神经网络; 深度学习

中图分类号: TN919.8

文献标识码: A

文章编号: 1674-6236(2018)20-0001-05

Abnormal action recognition of interaction research based on neural network

GUAN Bai-sheng^{1,2}, BIAN Chun-jiang¹, FENG Shui-chun¹, CHEN Shi¹

(1. National Space Science Center, Chinese Academy Sciences, Beijing 100190, China; 2. University of Chinese Academy Sciences, Beijing 100094, China)

Abstract: Abnormal action recognition has been a hot topic in the field of intelligent security. However, current research techniques can only recognize simple action, and there are still many problems in complex action recognition. With the continuous development and application in action recognition of artificial intelligence, we can try to recognize abnormal action of interaction in the surveillance video by deep learning. Therefore, the paper analyzes abnormal action of interaction in the surveillance video. Firstly, background subtraction is used to preprocess the video data, and then deep learning is used to train the algorithm model in train dataset. Finally, the algorithm model is validated on the test dataset. The model can effectively distinguish between abnormal action and normal action, and the recognition accuracy is 73%.

Key words: interaction; abnormal; action recognition; neural network; deep learning

人的行为识别一直是计算机视觉领域的一个重要研究方向。近年来随着深度学习技术的不断发展, 行为识别已经成为了计算机视觉方向的一个研究热点^[1]。在智能安防领域中, 人的行为识别研究范围主要包括越界检测、遗留物力检测、徘徊检测和打架斗殴等异常行为。

目前, 行为识别研究方面的一些综述性文章^[2-7]对最近几年行为识别的发展现状进行了分析和总结。行为识别特征提取的方法分为传统的特征提取方法和基于深度学习的特征提取方法。针对人的行为分析主要包括人的姿态分析^[8]、单人行为分析^[9]、

交互式行为分析^[10]和人群行为分析^[11]。

交互式行为识别的研究工作主要开始于2010^[12]年。在2014年的Computer Vision and Pattern Recognition(CVPR)会议上, 相关专家对交互式行为识别的研究现状以及存在的问题进行了讨论分析。之前的研究方法主要是通过传统的图像特征提取方法和分类算法等手段进行交互式行为的识别, 因此无法对复杂的交互式异常行为进行识别, 算法的鲁棒性较差。考虑到目前深度学习在行为识别上已经取得初步的成果, 故本文提出基于神经网络的交互式异常行为识别方法, 对异常行为进行识别。神经网络技术能够自动提取视频中的交互式行为特征,

收稿日期: 2018-01-24 **稿件编号:** 201801133

作者简介: 关百胜(1991—), 男, 山西运城人, 硕士研究生。研究方向: 图像处理与计算机视觉。

避免了设计复杂的行为特征提取算法,减化了算法的设计流程,同时也提高了系统的鲁棒性。

1 交互式异常行为识别概述

交互式异常行为识别的主要目的是自动识别出监控视频中的异常行为,保障公共安全,减轻人力资源。文献[14]提出了基于 Hough-Voting 的行为识别方法,文献[15]提出了基于 Spatio-Temporal 的行为识别方法,文献[16]提出了基于 Poselet Key-framing 的行为识别方法。以上研究都是通过图像处理技术提取行为特征,而本文通过深度学习技术对两个人的交互式行为进行识别,研究的交互式行为种类有握手、拥抱、踢腿、挑逗、推动和出拳6种,其中握手和拥抱属于正常行为,踢腿,挑逗,推动和出拳属于异常行为。交互式行为示例如图1所示。



图1 6种交互式行为

交互式异常行为识别分为视频数据预处理、特征提取和行为特征判断3步,整体框架如图2所示。首先需要对视频数据进行预处理,例如滤波去噪、图像增强和图像背景减除等处理;其次需要提取视频中的交互式行为特征,例如通过提取 Hog 特征、轮廓模板特征和神经网络自提取特征等方法提取行为特征;然后通过已经训练好的分类器对交互式异常行为进行判断,区分正常行为和异常行为。

2 系统方案设计

交互式异常行为识别系统方案设计如图3所示。系统方案中包括3部分,分别是背景减除、

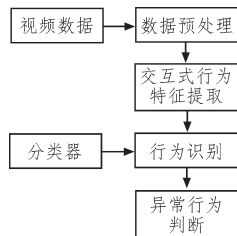


图2 交互式异常行为识别框架

AlexNet 网络和双隐层网络。

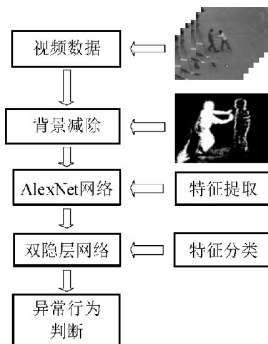


图3 系统方案设计

背景减除的作用是去除视频图像背景,保留前景目标图像,并将前景图像作为 AlexNet 网络的输入; AlexNet 网络的作用是根据输入的目标图像提取人体行为特征,作为特征分类的依据;双隐层网络的作用是根据已标注的标签数据和 AlexNet 网络已提取的行为特征训练分类模型,并根据训练后的模型判断异常行为。

2.1 视频图像背景减除

背景减除是对视频图像进行预处理,提取运动目标,为 AlexNet 网络提供输入。基于运动的目标检测算法是指从图像序列中将运动的目标区域检测出来。目前常用的基于运动的目标检测方法有帧差法、光流法和背景差(背景减除)法^[17]。考虑到基于高斯混合模型(Gaussian mixture model)的背景差法可以在线更新背景模型,很好地解决了光照变化、背景混乱变化的干扰,并且可以检测视频中停止运动的物体,因此本文选择基于高斯混合模型的背景差法来检测视频中运动的人体,去除图像背景并保存提取到的前景人体行为图像。高斯混合模型是指具有如下形式的概率分布模型^[19]:

$$P(y|\theta) = \sum_{k=1}^K \alpha_k \phi(y|\theta_k) \quad (1)$$

其中 α_k 是系数, $\alpha_k \geq 0$, $\sum_{k=1}^K \alpha_k = 1$; $\phi(y|\theta_k)$ 是高斯分布密度, $\theta_k = (\mu_k, \sigma_k^2)$,

$$\phi(y|\theta_k) = \frac{1}{\sqrt{2\pi}\sigma_k} \exp\left(-\frac{(y-\mu_k)^2}{2\sigma_k^2}\right) \quad (2)$$

称为第 k 个分模型。

2.2 AlexNet 网络模型

ALexNet 是一个经典的卷积神经网络 (Convolutional Neural Networks, CNN), 包含 5 个卷积层和 3 个全连接层, 最后一个全连接层的输出被送到 softmax 层中。文中通过 ALexNet 提取视频中的交互式行为特征, 并生成概率模型。AlexNet 网络的结构

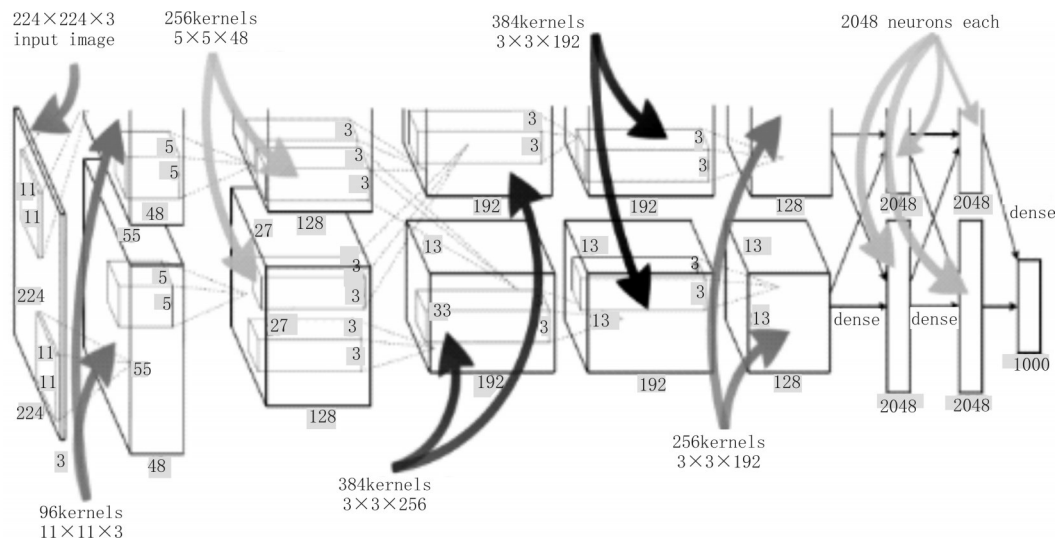


图4 ALEXNet网络结构图

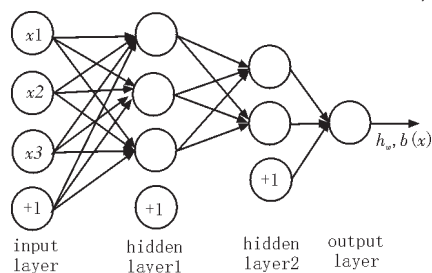


图5 双隐层网络示例

图5中的双隐层网络有3个输入单元,一个输出单元,隐藏层1有3个输入单元,隐藏层2有两个输入单元,图中的“+1”表示偏置节点。

3 算法具体实现

3.1 实验平台搭建

本实验在 Ubuntu16.04 环境下搭建深度学习开发平台,系统开发环境如表 1 所示。

其中 OpenCV3.3 是一个开源的计算机视觉库, 包含许多图像处理和计算机视觉的通用算法, 用户可以通过 OpenCV 快速实现相关功能, 不需要关注底

如图4所示^[20],输入图像大小为 $224 \times 224 \times 3$ (RGB),输出为1 000个float类型的值,即提取出的行为特征概率值。

2.3 双隐层网络模型

文中设计的双隐层网络是对已标注的行为和ALexNet网络生成的概率模型进行训练,最终生成一个行为分类器,对异常行为进行判断。双隐层网络包含一个输入层,一个输出层和两个隐藏层。双隐层网络的结构如图5所示。

表1 开发环境介绍

系统平台	Ubuntu16.04
开发语言	Python3.5
编程环境	PyCharm
依赖库	OpenCV3.3 TensorFlow1.4

层算法的具体实现过程;TensorFlow1.4是Google的第二代人工智能学习系统,可用于机器学习或图像识别等多项机器学习和深度学习领域,借助TensorFlow开发平台可以方便设计自己的神经网络。

3.2 背景减除算法实现

在算法的具体实现中,考虑到 OpenCV 视觉库中已有基于混合高斯模型的背景减除算法,并通过 BackgroundSubtractorMOG2 函数实现,因此本文对该函数进行研究分析并利用 BackgroundSubtractorMOG2 函数实现视频图像的背景减除。该算法的一个重要特点是它为每个像素选择一个合适的高斯分布。原图与背景减除后的图像如图 6 所示。

3.3 AlexNet 网络实现

利用 TensorFlow 深度学习平台实现 AlexNet 网

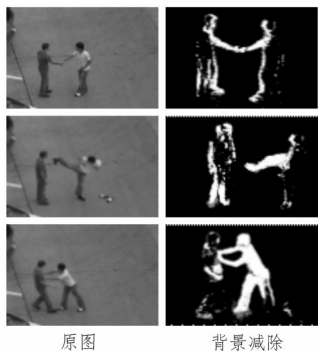


图6 背景去除前后

络,并根据实际需求对相关参数进行重新设置。AlexNet网络实现中用到的函数主要有7个,分别是`fully_connected()`、`tf.reshape()`、`local_response_normalization()`、`conv_2d()`、`max_pool_2d()`、`dropout()`和`regression()`。其中`conv_2d()`用于神经网络卷积操作;`local_response_normalization()`用于数据标准化;`tf.reshape()`用于格式化输入数据;`max_pool_2d()`用于池化操作;`dropout()`用于防止数据过拟合;`fully_connected()`用于全连接操作;`regression()`用于回归分析。AlexNet网络实现流程如图7所示。

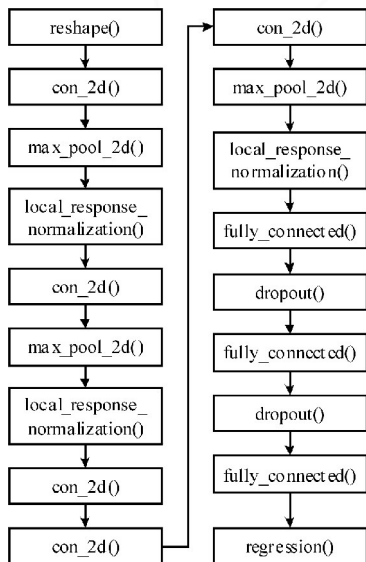


图7 AlexNet网络实现流程

完成 AlexNet 网络的实现并通过 AlexNet 网络提取交互式行为特征,提取过程中损失函数的变化如图8所示,最终的损失值为0.127。

3.4 双隐层网络实现

利用 TensorFlow 深度学习平台设计并实现双隐层网络,其设计结构简图如图9所示。双隐层网络可以将交互式行为分为正常行为和异常行为,其中输入层有20个输入单元,隐藏层1有16个输入单

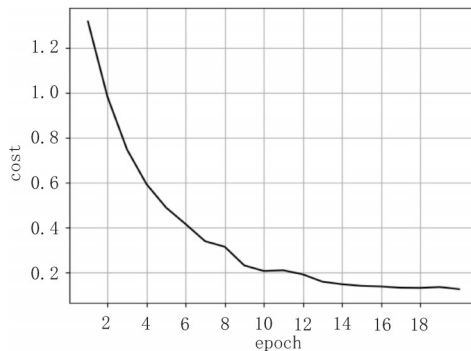


图8 AlexNet网络损失值

元,隐藏层2有8个输入单元,输出层有1个单元。关于双隐层网络的详细设计请参考图5。

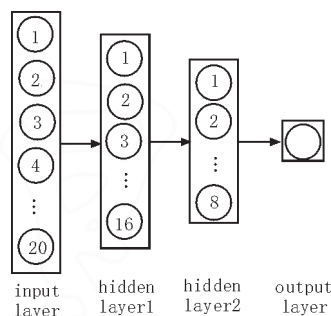


图9 双隐层网络结构简图

完成双隐层网络的实现并训练模型,训练过程中损失函数的变化如图10所示,最终的损失值为0.112。

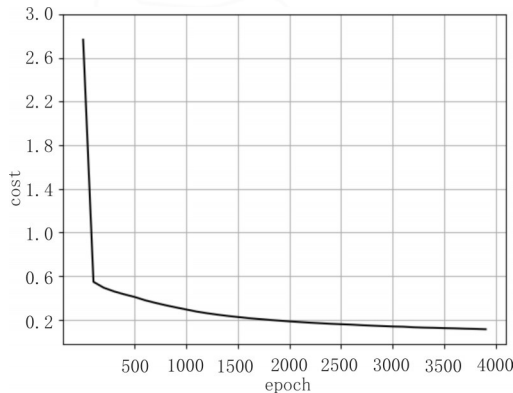


图10 双隐层网络损失值

4 实验结果及分析

经过对6种交互式行为进行测试分析,计算出行为分类的准确率为85%,识别的准确率为73%;文献[21]对目前交互式行为识别的准确率进行了总结,分类准确率范围基本在70%~90%之间。

在行为识别的准确率上,本文提出的方法与当

前的方法没有明显差异,主要原因是样本数据较少,无法进行大规模的深度学习。如果能够增加样本数量,进行大规模数据的学习,会再次提高识别的准确率。在行为识别的方法上,本文提出的基于深度学习的方法可以自学习行为特征,不需要人工进行特征标注并提取行为特征,简化了行为特征的提取过程,更有利于算法的工程化实现。

5 结 论

文中通过神经网络方法提取行为特征进行交互式异常行为的识别,验证了神经网络技术可以用于交互式行为识别,并且取得了很好的效果,准确率也不低于传统的行为识别方法。但是本文目前只是针对单组的交互式行为进行研究,对于多组交互式行为的分类问题还在研究中,后续将在本文研究的基础上,对多组交互式行为进行分析研究,提高算法模型的适应性。

参考文献:

- [1] 张少博. 安防监控中人体行为模式分类算法研究[D]. 武汉: 武汉理工大学, 2015.
- [2] Aggarwal J K, Ryoo M S. Human activity analysis: A review[J]. ACM Computing Surveys (CSUR), 2011, 43(3): 16.
- [3] 黄凯奇, 陈晓棠, 康运锋, 等. 智能视频监控技术综述[J]. 计算机学报, 2015, 38(6): 1093-1118.
- [4] Popoola O P, Wang K. Video-based abnormal human behavior recognition—A review[J]. IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews), 2012, 42(6): 865-878.
- [5] 朱煜, 赵江坤, 王逸宁, 等. 基于深度学习的人体行为识别算法综述[J]. 自动化学报, 2016, 42(6): 848-857.
- [6] Ke S R, Thuc H L U, Lee Y J, et al. A review on video-based human activity recognition[J]. Computers, 2013, 2(2): 88-131.
- [7] 李瑞峰, 王亮亮, 王珂. 人体动作行为识别研究综述[J]. 模式识别与人工智能, 2014, 27(1): 35-48.
- [8] Cao Z, Simon T, Wei S E, et al. Realtime multi-person 2d pose estimation using part affinity fields[J]. arXiv preprint arXiv:1611.08050, 2016.
- [9] 王世刚, 孙爱朦, 赵文婷, 等. 基于时空兴趣点的单人行为及交互行为识别[J]. 吉林大学学报(工学版), 2015, 45(1): 304-308.
- [10] Raptis M, Sigal L. Poselet key-framing: A model for human activity recognition[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2013: 2650-2657.
- [11] 卢晓威. 面向视频监控应用的人群分析研究[D]. 上海: 上海交通大学, 2015.
- [12] Ryoo M S, Chen C C, Aggarwal J K, et al. An overview of contest on semantic description of human activities (SDHA) 2010[C]//Recognizing Patterns in Signals, Speech, Images and Videos. Springer, Berlin, Heidelberg, 2010: 270-285.
- [13] Waltisberg D, Yao A, Gall J, et al. Variations of a hough-voting action recognition system[C]//ICPR Contests, 2010: 306-312.
- [14] Yao A, Gall J, Van Gool L. A hough transform-based voting framework for action recognition[C]//Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on. IEEE, 2010: 2061-2068.
- [15] Ryoo M S, Aggarwal J K. Spatio-temporal relationship match: Video structure comparison for recognition of complex human activities[C]//Computer vision, 2009 IEEE 12th International Conference on. IEEE, 2009: 1593-1600.
- [16] Raptis M, Sigal L. Poselet key-framing: A model for human activity recognition[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2013: 2650-2657.
- [17] 徐光柱, 雷帮军. 实用性目标检测与跟踪算法原理及应用[M]. 北京: 国防工业出版社, 2015.
- [18] 周志华. 机器学习[M]. 北京: 清华大学出版社, 2016.
- [19] 李航. 统计学习方法[M]. 北京: 清华大学出版社, 2012.
- [20] Krizhevsky A, Sutskever I, Hinton G E. Imagenet classification with deep convolutional neural networks[C]//Advances in neural information processing systems, 2012: 1097-1105.
- [21] Raptis M, Sigal L. Poselet key-framing: A model for human activity recognition[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2013: 2650-2657.