

DOI:10.19353/j.cnki.dzsj.2019.09.078

暴力行为检测是计算机视觉领域的一个热点问题。然而,近年来的研究者主要使用传统的特征提取方式来进行检测,而没有充分利用深度学习在计算机视觉领域的优势。本文提出了基于深度学习的监控场景中暴力行为检测框架,并在公开数据集(Hockey fights, Movies, VID)和本文录制的数据集上进行了验证,证明了本文方法的有效性和可靠性。

引言:近年来,真实视频场景中的行为自动检测和识别在人机交互、视频监控和基于内容的视频检索领域变得越来越重要。暴力行为检测是行为识别的一个特殊子问题。随着计算机视觉的发展,暴力行为检测领域的研究取得了很大的进展。然而,因为监控视频的质量太低,视频中存在遮挡、运动变化和照明的变化,暴力行为检测仍然是一项具有挑战性的任务。

暴力行为检测主要是识别视频中是否发生了暴力行为,例如打架。现有公开的数据集使用人工修剪的、仅包含动作的视频片段,并将这些片段标记为“暴力行为”和“非暴力行为”。早些年,Datta(Ankur Datta,Mubarak Shah,Niels Da,and Vitoria Lobo·Person-on-person violence detection in video data·In International Conference on Pattern Recognition,2002·Proceedings,volume 1,pages 433-438,2002)等人提出使用人体骨骼的轨迹信息和方向信息来检测暴力行为。另外一些研究者(Marco Cristani,Manuele Bicego,and Vittorio Murino·Audio-visual event recognition in surveillance video sequences·IEEE Transactions on Multimedia,9(2):257-267,2007; Jian Lin and Weiqiang Wang·Weakly-supervised violence detection in movies with audio and video-based co-training·In Pacific Rim Conference on Multimedia:Advances in Multimedia Information Processing,pages 930-935,2009)尝试使用声音特征来检测暴力行为。然而,对于相关的数据集和真实的监控视频中,包含高质量的声音信息是十分困难的,因此本文主要基于视觉特征来进行暴力行为检测。

事实上,先前的暴力行为检测方法检测性能比较低并且误报率很高。如果对于暴力行为的特征表示提取能够达到令人满意的效果,检测性能将会更好。较为准确的基于手工提取的时空特征,例如:MoSIF(Enrique Bermejo Nievas,Oscar Deniz Suarez,Gloria Bueno Garc'ia,and Rahul Sukthankar·Violence detection in video using computer vision techniques·In International Conference on Computer Analysis of Images and Patterns,pages 332-339,2011),STIP(Enrique Bermejo Nievas,Oscar Deniz Suarez,Gloria Bueno Garc'ia,and Rahul Sukthankar·Violence detection in video using computer vision techniques·In Computer Analysis of Images and Patterns,pages 332-339, Berlin,Heidelberg,2011·Springer Berlin Heidelberg)和ViF(Tal Hassner,Yossi Itcher,and Orit Kliper-Gross·Violent flows: Real-time detection of violent crowd behavior·In Computer Vision and Pattern Recognition Workshops, pages 1-6,2012),确实一定程度上提高了暴力行为检测(Fillipe D·M·De Souza,Guillermo Camara Ch'avez,Eduardo Alves Do Valle Jr,and Arnaldo De Albuquerque Araujo·Violence detection in video using spatio-temporal features·In Graphics,Patterns and Images,pages 224-230,2011)的性能。基于光流的方法(Amira Ben Mabrouk and Ezzeddine Zagrouba·Spatio-temporal feature using optical flow based distribution for violence detection·Pattern Recognition Letters,92(C):62-67,2017; Tao Zhang, Zhijie Yang, Wenjing Jia, Baoqing Yang,Jie Yang,and Xiangjian He·A new method for

基于深度学习的暴力行为检测  
西安特种设备检验检测院 李红昌 王晶 韩建军 张金民 杨玉山 赵岳

violence detection in surveillance scenes·Multimedia Tools and Applications,75(12):7327-7349,2016) 能够比较好的建模时序信息对于检测任务也有很大的性能提升。然而,它并不能应用在实时场景下的检测任务。

近几年来,随着深度学习的发展,在暴力行为检测领域也出现了一些使用深度学习的方法(Zihan Meng, Jiabin Yuan, and Zhen Li·Trajectory-pooled deep convolutional networks for violence detection in videos·In Computer Vision Systems, pages 437-447, Cham, 2017·Springer International Publishing; Chunhui Ding, Shouke Fan, Ming Zhu, Weiguo Feng, and Baozhi Jia·Violence detection in video by using 3d convolutional neural networks·In Advances in Visual Computing, pages 551-558, Cham, 2014·Springer International Publishing), 并获得比较好的检测性能。然而,这种方法仅仅考虑全局的时空信息,忽略了局部信息的影响。

本文设计了一个多流的融合网络来处理上述的问题,融合基于注意力的空间RGB流、时序流(通过optical flow实现)和局部空间流(基于分块空间RGB实现)的检测结果,考虑视觉注意力机制对外观特征提取的影响、光流场对运动信息的获取以及局部特征对检测结果的综合影响。在实验部分,从不同角度,包括等,分别分析本文所提方法的优劣,并与现有流行的方法进行对比,深入分析本文方法的准确性和稳定性。

## 1 多流深度学习算法

本文基于多流深度学习提出了一个新的暴力行为检测模型。包括基于注意力的空间RGB流,时序流,局部空间流三个模式,最后融合这三个流的结果来获得最终的识别结果。

首先将每个修剪的视频分成若干个片段,然后随机的从每个片段中采样一帧。因此,对于每个输入视频可以得到若干的输入帧,每帧使用BN-Inception网络来提取空间特征。基于提取的空间特征,我们使用注意力机制来使得该流更加的关注于发生动作的主体的区域,来提高该流的性能。

注意力机制可以看做是有特征图生成的一系列权重。如果一个区域在原始的特征图上属于包含暴力行为的区域,该区域的权重将趋向于1。注意力机制关注于包含暴力行为的区域,并且使得模型在预测时忽略掉一些背景噪音的影响

从而提高预测的准确度。这种方式提高了RGB帧提取的特征表示的质量。

本文在暴力行为检测中使用RGB帧的分块来提取人体的局部信息。为了使得不同分块的区域信息能够交互,本文将从这三个子块获得的局部信息进行拼接,并使用全连接层用于局部信息间的交互。通过使用上面提出的分块方法,RGB帧的三个子块被输入到局部空间流,并且我们可以获得他们的一维特征表示。最后,使用BN-Inception在原始图像上提取全局特征表示用于和融合后的局部特征表示相互补充,来获得能够提升性能的更好的特征表示。

为了使这三个流融合以获得更好的识别性能,我们使用晚融合的方式来融合三个流输出的类概率。

对于损失函数,本文使用交叉熵损失函数和SGD优化器。损失函数可以表示为:

$$L = -\sum_{c=1}^C y_{i,c} \log p_{i,c}$$

其中,  $C$  是行为类别的概率,  $y_{i,c}$  是样本  $i$  对应的真实标签,  $p_{i,c}$  是样本  $i$  的预测类别  $c$ 。

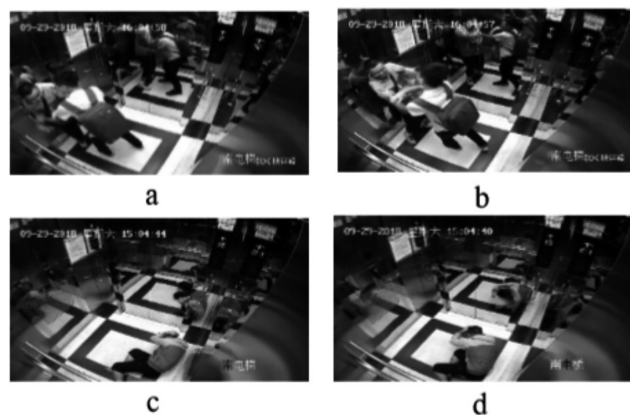


图1 本文提出的数据集的样例图片

## 2 实验结果与分析

本文在四个相关的数据集: Hockey fight dataset, hockey fight, movies (Enrique Bermejo Nieves, Oscar Deniz Suarez, Gloria Bueno Garc'ia, and Rahul Sukthankar·Violence detection in video using computer vision techniques·In Computer Analysis of Images and Patterns, pages 332-339, Berlin, Heidelberg, 2011·Springer

(下转第142页)

输出电压的波形, 并使用电压表XMM1测量电路输出电压值。

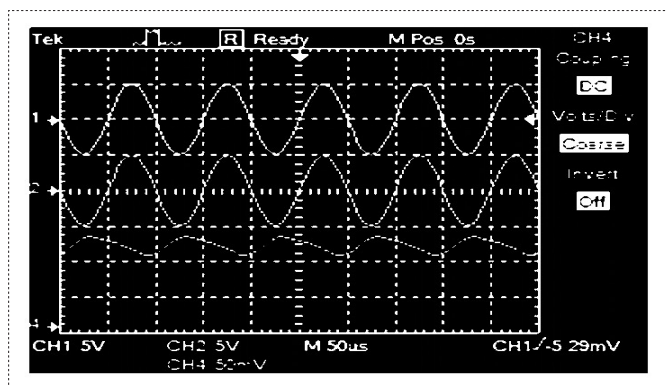


图5 电路仿真输出波形图

设置传感器最大电容值 $C_0=100\text{pF}$ , 当移动极板A的位移为0时, 即位于两个固定极板的中间时, 取值 $C_1=C_2=C_0/2$ 。当移动极板A向一侧运动时, 差动可变电容式传感器的容量变化随位移量线性变化。所以假定两差动电容容量每次增量为 $\pm 1\%$ , 分别完成正反行程两次测量, 得到如下数据。

从表1和图6可看出, 当传感器电容 $C_1=C_2$ 时, 在零点调整后输出电压最小。理论计算此时输出应该为0V, 而实验中因为有误差存在, 所以有零位输出-0.4mV。在电容变化量大于 $\pm 7\%$ 时, 输出线性变差, 在反行程测量区间段尤为明显。当电容相对变化量在 $\pm 6\%$ 的区间内时, 无论是正行程测量还是反行程测量, 电路均具有良好的线性,

最大非线性误差小于1.2%, 测试结果与差动电容传感器的理论一致。

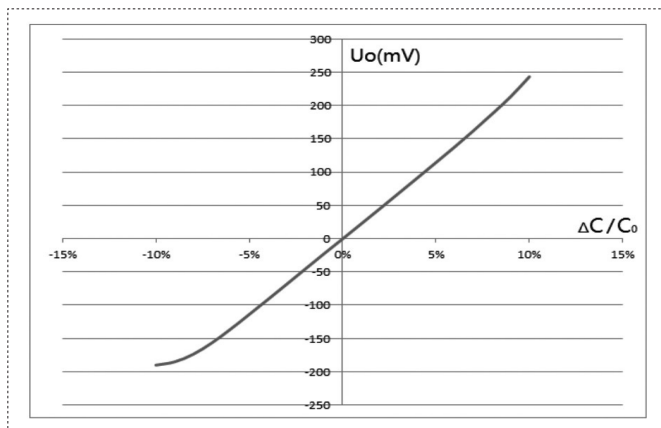


图6 差动电容传感器电路转换特性曲线

#### 4. 总结

仿真结果及数据分析表明, 设计的电路能较好的完成差动电容传感器输出电量的转换工作, 使用Multisim仿真达到了预期的效果。后期可利用该软件对多种传感器及其转换电路进行仿真分析, 对传感器研究有很大的帮助。

作者简介: 朱伟 (1982—), 男, 山东滕州人, 工程硕士学位, 枣庄科技职业学院讲师, 主要研究方向: 自动化技术。

(上接第139页)

Berlin Heidelberg), VID (Peipei Zhou, Qinghai Ding, Haibo Luo, and Xinglin Hou: Violent interaction detection in video based on deep learning. Journal of Physics: Conference Series, 844(1):012044, 2017) 和本文构建的数据集EsV上对本文提出的算法和其他方法进行了比较。此外, 本文录制了自己的数据集, 电梯监控视频数据集, 用于暴力行为检测 (图1)。

本文分析了不同模态在暴力行为检测任务中的性能表现。它主要包括光流模态 (对应于时序流), warped flow模态 (对应于TSN (Qing Xia, Ping Zhang, Jingjing Wang, Ming Tian, and Chun Fei: Real time violence detection based on deep spatio-temporal features. In Biometric Recognition, pages 157-165, Cham, 2018: Springer International Publishing) 中的warp flow), RGB模态 (对应于无attention的空间RGB流) 和基于注意力的空间RGB流。本文比较了不同模态融合的结果, 给出了暴力行为检测任务的最优的多模态融合的方法。

从数据集上实验结果可以看出, 本文提出的多流模型在四个相关的数据集上相比于其他暴力行为检测模型具有很大的性能提升。原因主要在于引入注意力机制, 能够提升算法自动去掉冗余; 局部

空间流从局部信息中有效的对局部特征进行建模, 因此单个流可以实现相比其他模型更好的检测性能。

#### 3 结论

本文提出了一种用于暴力行为检测的新的多流深度学习算法, 通过对全局信息以及局部信息的融合, 在验证数据集上取得了不错的效果。通过分析不同模态对于检测性能的影响, 本文找到了不同模态间的最佳融合的组合。该算法不仅可以用于暴力行为检测任务, 同时可以用于其他的行为识别相关的任务。

作者简介:

李红昌 (1971—), 男, 陕西宝鸡人, 大学本科, 副高级职称, 现供职于西安特种设备检验检测院, 主要从事机电类特种设备检验检测工作。

王晶 (通讯作者) (1986—), 女, 陕西西安人, 工学硕士, 工程师, 现供职于西安特种设备检验检测院, 主要从事机电类特种设备检验检测工作。