

一种基于卷积神经网络的哈欠检测算法

马素刚^{1,2} 赵琛² 孙韩林² 韩俊岗²

(长安大学信息工程学院 西安 710064)¹ (西安邮电大学计算机学院 西安 710121)²

摘要 哈欠检测可以用于对驾驶员的疲劳驾驶行为发出警告,从而减少交通事故的发生。提出了一种基于卷积神经网络的哈欠检测算法,可以把驾驶员的面部图片直接作为神经网络的输入,避免对面部图片进行复杂的显式特征提取。利用 Softmax 分类器对神经网络提取的特征进行分类,判断是否为打哈欠行为。该算法在 YawDD 数据集上取得了 92.4% 的哈欠检测准确率。与现有多个算法相比,所提算法具有检测准确率高、实现简单等优点。

关键词 哈欠检测,卷积神经网络,权值共享,Softmax 分类器,YawDD 数据集

中图分类号 TP183 文献标识码 A

Yawning Detection Algorithm Based on Convolutional Neural Network

MA Su-gang^{1,2} ZHAO Chen² SUN Han-lin² HAN Jun-gang²

(School of Information Engineering, Chang'an University, Xi'an 710064, China)¹

(School of Computer Science and Technology, Xi'an University of Posts and Telecommunications, Xi'an 710121, China)²

Abstract Yawning detection can be used to warn drivers of fatigue driving behavior, thereby reducing traffic accidents. A yawning detection algorithm based on convolutional neural network was proposed. The driver's facial image can be directly used as input for neural network, so as to avoid the complex explicit feature extraction of the facial image. The Softmax classifier is used to classify the features extracted from the neural network to determine whether the behavior is yawning or not. This algorithm achieves 92.4% accuracy in the YawDD dataset. Compared with other existing algorithms, the proposed method has the advantages of high detection accuracy and simple implementation.

Keywords Yawning detection, Convolutional neural network, Weight sharing, Softmax classifier, Yawning detection dataset

1 引言

驾驶员在疲劳驾驶时会表现出注意力分散、反应迟钝、视力变弱等现象,极易导致交通事故的发生。研究表明,随着疲劳程度的加重,驾驶员的驾驶能力会不断下降。美国国家公路交通安全管理局(NHTSA)估计,每年由于司机疲劳或睡眠不足,造成了约 10 万次交通事故^[1]。对 100 辆车自然驾驶数据的研究发现,疲劳驾驶导致了 22%~24% 的交通事故。在没有任何警示的情况下,疲劳驾驶发生碰撞或接近碰撞的风险高出 4~6 倍^[2]。鉴于此,疲劳驾驶检测近年来一直是国内外研究的热点问题。

2 相关工作

驾驶员在疲劳驾驶时表现出来的行为特征有眨眼、点头、打哈欠等,其中打哈欠是一种主要表现形式^[3]。国内外学者已经对哈欠检测做了大量研究。文献[4]基于机器视觉的方法,采用改进后的 Viola-Jones 算法分别对人脸和嘴部进行检测,接着用反向投影原理(back-projection)对嘴部变化的频率和尺度进行测量,从而判断驾驶员是否出现打哈欠行为。谢

国波等^[5]提出了一种基于嘴巴内轮廓角点检测与曲线拟合的哈欠检测方法,利用角点检测获取嘴巴内轮廓上的若干点,对这些点进行曲线拟合,并建立嘴唇内轮廓数学模型,通过对张口度曲线进行时间维度的分析来对哈欠进行判决。文献[6]采用肤色检测算法获得人脸区域,为了解决人脸区域出现的多斑点问题,进一步标识了嘴的位置并检测其闭合度,然后计算脸的下部区域的垂直投影的直方图值,并把该值作为特征以用于哈欠检测。文献[7]提出的算法首先基于驾驶员脸部检测的结果,通过灰度投影检测出驾驶员的两侧嘴角,然后利用 Gabor 小波提取驾驶员嘴角的纹理特征,并基于这一特征利用 LDA 算法进行分类,从而判别是否出现哈欠行为。Li 等^[8]在获取人脸图像后,采用 Haar-like 特征检测驾驶员的嘴部并根据历史位置进行跟踪,然后基于嘴部高宽比对哈欠行为进行检测。Lu 等^[9]利用差值图像进行人脸分割,基于定向积分投影方法获得鼻孔的中心位置,把鼻孔中心到下巴的垂直距离作为哈欠检测的依据。文献[10]提出的算法使用局部二值模式(LBP)特征来标识嘴部变化,能够在嘴部被遮挡或变形时,有效地进行哈欠检测。

以上算法主要利用驾驶员的嘴部特征作为判断依据,不

本文受国家自然科学基金(61373116),陕西省自然科学基金(2016JM6048),陕西省教育厅专项科研项目(17JK0696)资助。

马素刚(1982—),男,博士生,高级工程师,CCF 会员,主要研究方向为机器学习与数据挖掘,E-mail:msg@xupt.edu.cn;赵琛(1993—),男,硕士生,主要研究方向为机器学习与数据挖掘,E-mail:zhaochen@126.com;孙韩林(1980—),男,博士,副教授,主要研究方向为复杂网络分析,E-mail:sunhanlin@xupt.edu.cn;韩俊岗(1943—),男,教授,CCF 会员,博士生导师,主要研究方向为计算机体系结构,E-mail:hjg@xupt.edu.cn。

断提升哈欠检测的效果。然而,打哈欠行为不仅带来了嘴部特征的变化,同时也导致眼睛和鼻子特征发生变化,因此应该综合考虑驾驶员的面部表情。另一方面,这些算法中用来作为检测依据的各种特征,均是通过人工方式显式提取,无法充分挖掘面部变化与哈欠行为之间深层次的关联。本文提出了一种基于卷积神经网络(CNN)的哈欠检测算法,对驾驶员的脸部原始图像进行简单的预处理后,将其作为神经网络输入,利用机器学习方法提取面部特征,借助 Softmax 分类器^[11]对这些特征进行分类,从而实现哈欠行为的检测。

3 CNN 的工作原理

Lecun 等^[12]在研究手写数字识别问题时,首先提出了使用梯度反向传播算法训练的卷积神经网络模型,并取得了很好的性能。卷积神经网络是一种典型的深度学习网络模型,已经在图像处理、语音识别等领域取得了很大成功。它可以直接输入原始图像,不需要进行人工特征提取,避免了对图像进行复杂的预处理。CNN 主要有卷积层和子采样层两种类型的网络层,其中卷积层的作用是提取图像的各种特征,子采样层是对原始特征信息进行抽象,从而减少训练参数,减轻模型过拟合的程度。CNN 具有两大特点,即局部连接和权值共享,这种结构对平移、比例缩放、倾斜或其他形式的变形具有高度不变性。

3.1 卷积神经网络结构

CNN 结构如图 1 所示,包含输入层,卷积层 C1,C3,子采样层 S2,S4,以及全连接层 F5,F6。

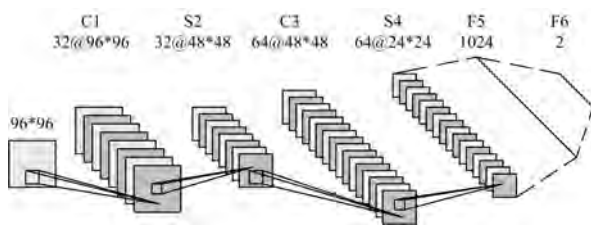


图 1 卷积神经网络结构

1)输入层:输入经过简单预处理的面部图像,大小为 96×96 。

2)卷积层 C1:卷积核大小为 5×5 ,共有 32 个卷积核,得到 32 组 96×96 的特征映射。

3)子采样层 S2:采用最大值子采样法,将 C1 层的每组特征映射中的 2×2 领域点采样为 1 个点,得到 32 组 48×48 的特征映射。

4)卷积层 C3:采用 64 个大小为 5×5 的卷积核,得到 64 组 48×48 的特征映射。

5)子采样层 S4:采用最大值子采样法,把 C3 层每组特征映射中的 2×2 领域点采样为 1 个点,得到 64 组 24×24 的特征映射。

6)全连接层 F5:拥有 1024 个神经元,与 S4 层神经元全连接。

7)全连接层 F6:与 F5 层神经元全连接,用 Softmax 回归实现二分类。

3.2 局部连接

CNN 在网络的相邻两层之间使用局部连接来获取图像的局部特性,具体来说,本隐藏层的神经元只与上一层的局部区域有连接,上一层的这些局部区域被称为空间连续的接受域。

3.3 权值共享

在 CNN 中对整幅图像进行一次卷积操作,不同局部区域使用同一个卷积核(权值参数相同),由此生成了一个特征映射。通过不同的卷积核对整幅图像进行卷积可生成不同的特征映射。权值共享就是同一个特征映射中的所有神经元使用同一个权值,因此权值参数的个数与神经元的个数无关,只与卷积核的大小及特征映射的个数有关。但是,连接总数与神经元的个数有关,神经元的个数由特征映射的大小决定。

4 数据集

4.1 YawDD 数据集

Abtahi 等研究整理的数据集 YawDD^[13]可以用于对人脸检测、人脸特征提取、哈欠检测等算法进行验证,在非商业或学术领域可以免费使用。YawDD 包含 2 个视频集,采录了多个具有不同面部特征的驾驶员行为。视频录制的分辨率为 640×480 ,采用 24bit 真彩色,每秒录制 30 帧。第一个视频集中包含了 322 个视频,拍摄第一个视频集的摄像头安装在车内后视镜的下方,其对每个驾驶员录制 3~4 段视频,包含了正常驾驶、说话/唱歌、打哈欠等多种不同的面部行为;第二个视频集中包含了 29 个视频,拍摄第二个视频集的摄像头安装在仪表盘上方,每个驾驶员只有一段视频,包含了正常驾驶、驾驶时说话和驾驶时打哈欠 3 种行为。

4.2 数据集构建

YawDD 数据集并不能直接用来对本文所提算法进行验证,因此本文对其进行进一步处理并构建了数据集 YawDD_CNN。YawDD_CNN 分为训练集和测试集两个部分,其中训练集包含 10400 张图片,测试集包含 1400 张图片。YawDD_CNN 中的每张图片均有一个人工标注的标签,即 0 表示非打哈欠状态,1 表示打哈欠行为。

数据集 YawDD_CNN 是在 YawDD 的基础上,根据图 2 所示的流程构建而成。构建流程包括图片获取、人工标记、人脸检测、尺寸变换和灰度转换等 5 个主要步骤。



图 2 数据集 YawDD_CNN 的构建流程

1)图片获取:以每秒 30 帧的速率对 YawDD 数据集中的视频进行采样,获得一组图片。这些图片包含了不同驾驶员的不同驾驶行为,如正常驾驶、说话/唱歌或者打哈欠。

2)人工标记:以人工方式处理,对每一张图片进行识别,标记出是否为打哈欠行为,把打哈欠行为标记为 1,否则标记为 0。

3)人脸检测:使用 Adaboost 人脸检测算法^[14]对每一张图片进行检测,保留包含人脸部分的矩形区域。

4)尺寸变换:在上一步人脸检测过程中得到的图片大小

并不相同,接下来利用双线性插值方法^[15]对图片的大小进行归一化。

5)灰度转换:把彩色图片转换为黑白图片,即对图片进行灰度化处理^[16]。

5 实验结果与分析

5.1 实验环境

本文借助 TensorFlow 平台^[17]搭建 CNN 模型,并使用数据集 YawDD_CNN 对该模型进行测试。TensorFlow 是一个采用数据流图描述计算过程和用于数值计算的开源软件库,由 Google 公司开发,可以用于机器学习和深度神经网络方面的研究工作,其灵活的架构适用于 CPU、GPU 等多种硬件环境。运行 TensorFlow 的服务器配置为:CPU 型号为 Intel(R) Xeon(R) CPU E5-2620 v3,主频为 2.40GHz,内存为 128GB,硬盘为 4.8TB;同时配有一块 NVIDIA Tesla K80 GPU,安装 CentOS Linux release 7.2 操作系统。

5.2 实验结果分析

假设图片的大小为 $L \times L$ (长度和宽带均为 L),训练次数为 N ,批量样本数为 S ,学习率为 l ,训练准确率为 P 。

当 $L=32, N=1000, l=0.0001$ 时,准确率 P 与批量样本数 S 之间的变化关系如图 3 所示,这里准确率 P 取重复 10 次训练的平均值。

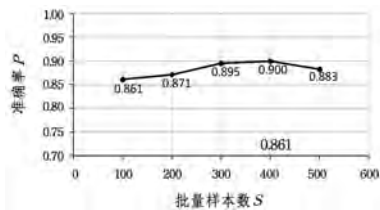


图 3 准确率 P 随批量样本数 S 的变化($L=32$)

图 3 显示,当每次训练输入的样本数 S 从 100 逐渐变化至 400 时,训练准确率逐步增大至 90%,然而 S 取值为 500 时,准确率降至 88.3%。

当 $L=64, N=1000, l=0.0001$ 时,准确率 P 与批量样本数 S 之间的变化关系如图 4 所示,准确率 P 仍然取重复 10 次训练的平均值。

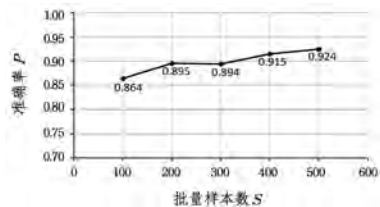


图 4 准确率 P 随批量样本数 S 的变化($L=64$)

图 4 显示,当每次训练输入的样本数 S 从 100 逐渐变化至 500 时,训练准确率总体来看呈逐步增大的趋势,最大值达到了 92.4%。

当 $L=96$ 或 $L=128, N=1000, l=0.0001$ 时,实验结果显示,随着批量样本数 S 从 100 不断增大,准确率 P 的值均在 80% 以下。

综合以上分析,在 $L=64, N=1000, l=0.0001, S=500$ 的情况下,可以获得最佳的训练准确率,即 $P=92.4\%$ 。

将本文算法与不同文献提出的几种哈欠检测算法进行对

比,检测准确率如表 1 所列。所有算法均在数据集 YawDD^[13] 上进行测试。

表 1 不同哈欠检测算法的准确率

算法	检测准确率
基于反向投影原理的算法 ^[4]	75
基于双代理专家系统的算法 ^[6]	94
基于 CNN 和鲁棒跟踪的算法 ^[18]	92
本文所提基于 CNN 的算法	92.4

由表 1 可知,本文提出的算法的准确率高于文献^[4]中提出算法。与文献^[6]、文献^[18]的算法相比,本文所提算法的准确率差别不大,但是本文所提算法不需要对驾驶员面部图像中的眼睛、鼻子、嘴等部位特征进行进一步处理,复杂度较低、容易实现。

结束语 本文提出了一种基于卷积神经网络的哈欠检测算法,以驾驶员的整个面部图像作为输入,充分考虑了眼睛、鼻子、嘴巴等面部特征对哈欠行为的影响。面部图像特征通过神经网络逐层自动提取,避免了复杂的人工特征提取工作,从而简化了算法。该算法具有复杂度低、准确率高等优点。如果能对神经网络层数、分类器类型、学习率等参数进行进一步优化,算法的准确率有望得到进一步提高。利用本文提出的算法检测打哈欠行为时,没有考虑连续多帧图像之间的关联性,下一步工作可以通过累计检测相邻多帧数据来提高检测准确率。综合考虑驾驶员的生物特征(如脑电、心电数据)、车辆行驶状态等多模态数据进行融合判断,也是未来工作的重要方向。

参考文献

- [1] DWIVEDI K, BISWARANJAN K, SETHI A. Drowsy driver detection using representation learning[C]// IEEE International Advance Computing Conference. IEEE, 2014: 995-999.
- [2] ABTAHI S, SHIRMOHAMMADI S, HARIRI B, et al. A yawning measurement method using embedded smart cameras[C]// Instrumentation and Measurement Technology Conference. IEEE, 2013: 1605-1608.
- [3] GALLUP A C, ELDAKAR O T. The thermoregulatory theory of yawning: what we know from over 5 years of research[J]. Frontiers in Neuroscience, 2013, 6: 188.
- [4] OMIDYEGANEH M, SHIRMOHAMMADI S, ABTAHI S, et al. Yawning Detection Using Embedded Smart Cameras[J]. IEEE Transactions on Instrumentation & Measurement, 2016, 65(3): 1-13.
- [5] 谢国波, 陈云华, 张灵, 等. 基于嘴巴特征点曲线拟合的哈欠检测[J]. 计算机工程与科学, 2014, 36(4): 731-736.
- [6] ANITHA C, VENKATESHA M K, ADIGA B S. A Two Fold Expert System for Yawning Detection[J]. Procedia Computer Science, 2016, 92: 63-71.
- [7] FAN X, YIN B C, SUN Y F. Yawning detection based on Gabor wavelets and LDA[J]. Beijing Gongye Daxue Xuebao, 2009, 35(3): 409-413, 432.
- [8] LI L, CHEN Y, LI Z. Yawning detection for monitoring driver fatigue based on two cameras[C]// International IEEE Conference on Intelligent Transportation Systems. IEEE Xplore, 2009: 1-6.

灰度差异就可能变大,从而造成误检。

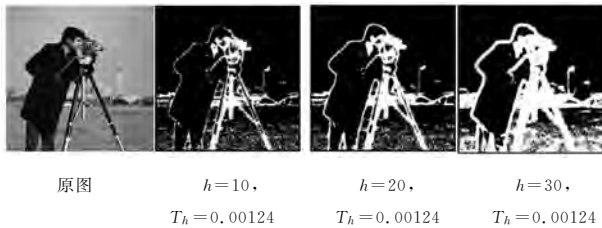


图 4 T_h 不变, h 改变时的检测效果图(7×7 窗口)

最后将图像加入高斯噪声进行分析。图 5 是 0.01 的高斯噪声图,可发现本文方法易受噪声影响,因为它是基于邻域像素间差异的方法,当出现噪声时,会使所计算出的核密度估计值偏小,从而使该像素点被误认为边缘点,因此与 Laplace 算子一样,需要事先进行滤波。



(a) 0.01 的高斯噪声图 (b) 本法检测图

图 5 椒盐噪声检测效果图

根据以上分析,使用经典 Sobel 算子法及本文算法进行了边缘检测对比,检测效果如图 6 和图 7(左图为原图,右图为检测图)所示。Sobel 算子的检测比较细腻,检测出了图像显著的边缘,也会造成漏检。本文方法检测出的边缘虽不够细腻,但可以检测出几乎所有的边缘,且无水平、垂直的方向限制,对不复杂图像的检测很有成效,如图 6 中的电路图。



图 6 Sobel 算子法的检测效果图

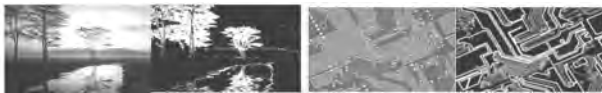


图 7 本文方法的检测效果图

结束语 在检测之前,可以先对图像进行滤波处理,然后

根据检测图像的特点选择合适的窗口以及合适的核函数,选择好阈值后即可进行检测。本文的方法在边缘检测上是有效的,能够根据需求通过选择合适的带宽或者阈值进行边缘检测,灵活方便;且其计算简单,不需要图像训练,运行速度较快,能够快速进行检测。另外,为了追求更快的检测速度,也可以使用与高斯核函数类似的 Epanechnikov 核函数^[8]进行计算。进一步,我们也可通过设定带宽或者阈值来进行不同程度的检测,可分析图像的核密度来检测出感兴趣区域,而忽略不感兴趣区域。对于线条简单的图像,本文算法的检测是准确且快速的,这对于大型、大量的图片检测而言很有效。

另外,阈值的选择是一个关键问题,因为最大类间方差法也无法完全胜任,所以需要进行调参,且噪声也是一个影响检测质量的因素。本文是一种探索尝试,希望今后能对算法做进一步改进,以解决上述问题,使检测效果更加高效、细腻、完美。

参 考 文 献

(上接第 229 页)

- [9] LU Y, WANG Z. Detecting Driver Yawning in Successive Images[C]// International Conference on Bioinformatics and Biomedical Engineering. IEEE Xplore, 2007: 581-583.
- [10] IBRAHIM M M, SORAGHAN J J, PETROPOULAKIS L, et al. Yawn analysis with mouth occlusion detection[J]. Biomedical Signal Processing & Control, 2015, 18: 360-369.
- [11] 王忠民, 曹洪江, 范琳. 一种基于卷积神经网络深度学习的人体行为识别方法[J]. 计算机科学, 2016, 43(Z11): 56-58.
- [12] CUN Y L, BOSER B, DENKER J, et al. Handwritten digit recognition with a backpropagation network[J]. World of Computer Science & Information Technology Journal, 1990, 2(2): 396-404.
- [13] ABTAHI S, OMIDYEGANEH M, SHIRMOHAMMADI S, et al. YawDD: A Yawning Detection DataSet[C]// Proceedings of the 5th ACM Multimedia Systems Conference. New York: ACM Press, 2014: 24-28.
- [14] 曹莹, 苗启广, 刘家辰, 等. AdaBoost 算法研究进展与展望[J]. 自动化学报, 2013, 39(6): 745-758.
- [15] 王平, 全吉成, 赵柏宇. 基于双线性插值的图像缩放在 GPU 上的实现[J]. 微电子学与计算机, 2016, 33(11): 129-132.
- [16] 刘春晓, 朱臻阳, 伍敏, 等. 基于主色检测与灰度传播的彩色图像灰度化方法[J]. 计算机辅助设计与图形学报, 2016, 28(3): 433-442.
- [17] VOLOSHYN D. Application of deep learning and computer vision frameworks for solving video context prediction problem [C]// Proceedings of the 10th International Conference of Programming. 2016: 164-169.
- [18] ZHANG W, YI L M, WANG T, et al. Driver yawning detection based on deep convolutional neural learning and robust nose tracking[C]// International Joint Conference on Neural Networks. IEEE, 2015: 1-8.
- [1] 贾云得. 机器视觉[M]. 北京: 科学出版社, 2000.
- [2] 袁春兰, 熊宗龙, 周雪花, 等. 基于 Sobel 算子的图像边缘检测研究[J]. 激光与红外, 2009, 39(1): 85-87.
- [3] 解欢庆. 改进的蚁群算法在图像边缘检测中的应用研究[D]. 兰州: 兰州大学, 2011.
- [4] AJENAGHUGHURE I B, OGINI NICHOLAS O, ONYEKE-LI C O. Optimum Fuzzy based Image Edge Detection Algorithm [J]. I. J. Image, Graphics and Signals Processing, 2017(4): 44-45.
- [5] XIE S N, TU Z W. Holistically-Nested Edge Detection[C]// The IEEE International Conference on Computer Vision (ICCV). 2015: 1395-1403.
- [6] COMANICIU D, MEER P. Mean Shift: A Robust Approach Toward Feature Space Analysis [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence (S0162-8828), 2002, 24(5): 603-619.
- [7] 李大威, 徐立鸿, 吴阳, 等. 一种迭代的核密度估计视觉目标检测算法[J]. 系统仿真学报, 2013, 25(3): 558-564.
- [8] CHU C Y, HENDERSON D J, PARMETER C F. On discrete Epanechnikov kernel functions[J]. Computational Statistics and Data Analysis, 2017, 116.