

摘要：随着公交车和出租车的数量不断增加，司机在开车时的不规范驾驶动作造成很多交通事故。为了有效降低这种交通事故发生的可能性，设计了一个基于深度学习的规范驾驶行为智能识别系统。本系统将摄像头安装在车内并实时采集司机驾驶状态的图像，控制器与后台电脑服务器远程连接，服务器通过网络通信接收司机图像并通过深度学习对采集的司机图像进行识别并处理，再把处理的结果传给控制器。这样司机在开车时发生喝水、打电话、玩手机等7种危险动作行为时，就可以通过控制器及时对司机进行智能安全语音提示，后台也可以实时监督和警示司机。

关键词：不规范驾驶动作；网络通信；深度学习；图像处理；智能安全语音提示

中图分类号：TP391.41

文献标识码：A

文章编号：1006-883X(2020)02-0007-05

收稿日期：2019-12-04

基于深度学习的规范驾驶行为智能识别系统

李光东 程进 周子肖 刘宝宇 王骏琦 史立根 姚宇军 常创创 李文月 邹小平

北京信息科技大学北京市传感器重点实验室，北京 100101

一、引言

随着越来越多的人乘坐公共交通出行，司机的驾驶行为也密切关乎着乘客的安全。近年来出现的交通事故越来越多，有很多交通事故主要是由于司机在开车中分心造成的，司机在车里的一点点不规范动作和分心都有可能产生不良的后果^[1]。

为了更好地警示和监督司机，降低这类事故发生的可能性，给乘客提供安全服务，本文针对公共交通系统，设计了基于深度学习的规范驾驶行为智能识别系统，通过深度学习识别司机在开车时的危险动作行为并及时对司机进行智能安全语音提示，并且后台公共交通公司也可以远程监督司机，可以对司机起到警示作用。

深度学习技术在图像识别中达到非常好的效果^[2-3]，作为深度学习中重要的卷积神经网络（Convolution Neural Network, CNN）则可以追溯到20世纪60年代早期，当时是由两位著名的生物学家—Hubel和Wiesel在文献中提出了感受野的概念^[4]。到20世纪80年代

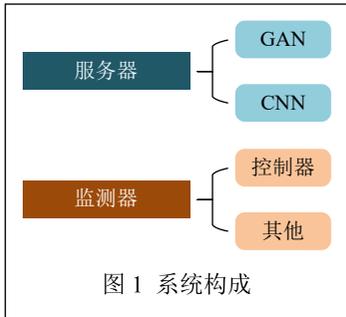
中期，Fukushima在文献中提出了基于感受野概念的神经认知机^[5]，可以说是第一次真正实现了卷积神经网络。

对司机不规范驾驶行为状态是通过卷积神经网络来检测^[6]。本系统结构相对简单而且可以识别更多的不规范驾驶状态，只要把带有摄像头的控制器系统安装在副驾驶车窗上方来拍摄司机，主要针对司机开车中像玩手机、打电话、操作音响、喝水等7种危险动作状态，后台通过人工智能技术对司机驾驶行为状态进行实时监测，并通过卷积神经网络对采集的司机图像进行识别处理，及时发出语音警示，提醒司机安全驾驶，后台也可以保存这些图像和数据集。将来该系统可以安装在更多的车辆中进行使用。

二、系统设计

1、系统构成

本系统主要由2个部分组成（图1）：后台电脑服务器和监测器，其中服务器主要用来后台监控司



机驾驶状态并识别司机危险行为的，包括用到去运动模糊原理和CNN卷积

神经网络原理；监测器主要是拍摄司机状态并传输给服务器的，包括带有摄像头的控制器或者其它无线网络摄像头。服务器和监测器主要是通过网络通信来相互传输信息的。

2、去运动模糊模型

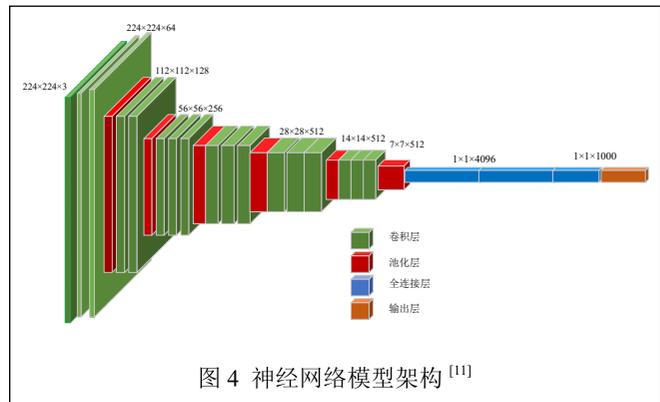
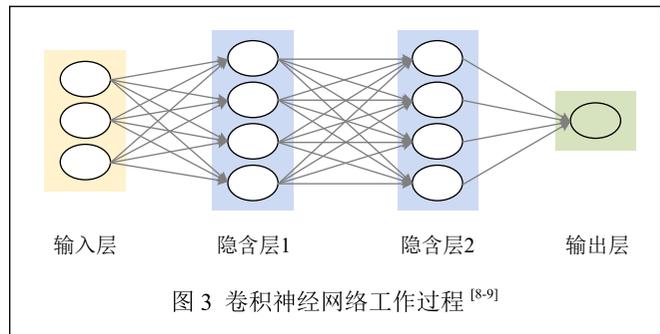
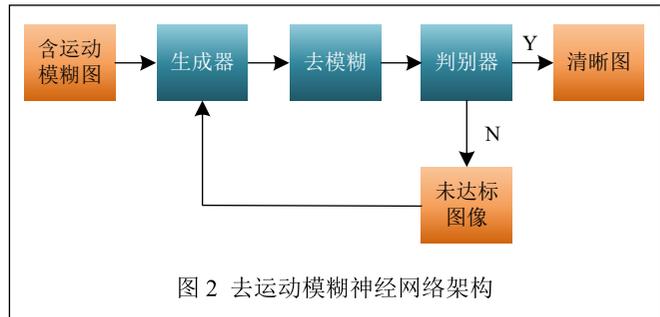
由于车在行驶中，司机一直是运动状态，摄像头采集司机的图像由于司机的运动会产生运动模糊，所以需要去模糊处理。运用了生成式对抗网络（Generative Adversarial Networks, GAN）博弈模型来实现去运动模糊。GAN是一种深度学习模型，它由生成器和判别器网络相互博弈来达到最好的生成结果^[7]。如图2所示，首先，含有运动模糊的图像通过生成器进行去模糊处理，再通过判别器进行运动模糊判别，如果图像去运动模糊没有达标，则再次返回生成器，进行去模糊处理，直到生成清晰的图像。分别对司机这7种不同驾驶状态的运动模糊图像模型进行去运动模糊训练50次，一共训练了350次，其中的网络训练的数据来源于kaggle平台里检测司机状态的数据集。最后对司机其中的100张不同驾驶状态的模糊图像进行实验验证，结果生成95张清晰的图像，精确度达到了95%左右。

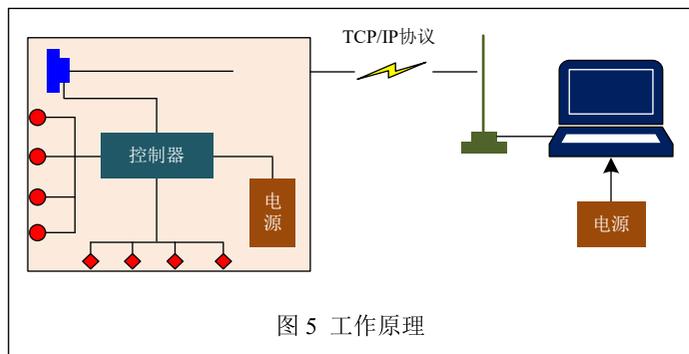
3、CNN 卷积神经网络模型

采集的司机图像通过卷积神经网络进行识别和处理，如图3从左向右可以看出，对司机的真实图像，首先通过输入层方块中的滤波器进行滤波，将滤波器分别与图像的同大小区域进行点乘，每次滤波器依次从左往右从上往下滑过该图像所有的区域，点乘后各个乘积求

和得到新的过滤后的图像，传递到下一层，隐含层方块是下采样的块，通过下采样来压缩过滤后的图像，压缩后的图像再通过滤波器滤波产生新的图像，再通过下采样进行压缩，最后输出结果。

神经网络架构由卷积层、池化层、全连接层以及输出层构成^[10]。如图4所示是神经网络模型架构图，探索出该模型是最适合本系统。对采集的司机图像进行识别处理，首先，要将司机图像分解为一系列有重叠的 3×3 的像素块。在这之后，将每个像素块输入一个简单的、单层的神经网络，保持权重不变。这一步将像素块集转化成一个矩阵。只要保持每个像素块都比较小（这里是 3×3 ），处理它们所需的网络也可以保持可控。





数据报组成帧，并通过比特流的形式传到物理层，最后数据就由源主机传输到了目的主机。两者反过来也是一样，就实现了网络通信。如果两者不在同一局域网中，可以使用监测器和服务器都可以访问的公共服务器，两者就可以通过共同访问公共服务器来实现网络通信。如图 5 所示。

三、系统安装与工作流程

监测器模块采用带有摄像头的控制器，安装在如图 6 (a) 示车内副驾驶位置，通过摄像头实时采集司机的视频帧，所采集的司机图像如图 6 (b) 所示。

系统的工作流程及程序流程图如图 7 所示。运行后台电脑服务器程序，带有摄像头的控制器通过网络通信接入电脑服务器，把采集的图像传到后台服务器，服务器先通过去运动模糊原理对模糊的图像进行去模糊处理，再通过 CNN 卷积神经网络对清晰的图像进行识别^[14-15]，最后把识别的结果传回到控制器，控制器对不规范驾驶行为及时在车内对司机进行智能语音安全提示，例如“请勿操作音响，注意行车安全”。同时，后台公共交通公司也可以对保存在后台服务器的司机驾驶行为数据进行统计分析，远程对司机实时监控。系统可以主动学习，不断优化自身的识别判断能力。后台系统也将记录和保存所有触发预警的行为数据，同时提醒管理人员及时进行处理。



下一步，输出值会被排成矩阵，以数据形式表示照片中每个区域的内容，不同轴分别代表颜色、宽度和高度通道。对于每一个图像块都有一个 $3 \times 3 \times 3$ 的表示。接下来是池化层。池化层对这些 3 或 4 维的矩阵在空间维度上进行下采样。处理结果是池化阵列，其中只包含重要部分图像，并且丢弃了其他部分，这样一来最小化了计算成本，同时也能避免过拟合问题。经过下采样的矩阵作为全连接层的输入。由于经过了池化和卷积操作，输入的尺寸被大幅减小，现在有了正常网络能处理的，同时能保持数据最重要特性的东西。这样通过每一层的工作，最后的输出结果则代表了系统对采集的司机图像这一判断的精确度。

4、网络通信

监测器和服务器是通过网络通信来相互运行工作，如果两者在同一局域网内，则通过 TCP/IP 协议来实现监测端和服务器的数据和信息传输^[12-13]。把带有摄像头的控制器作为源主机，把后台电脑服务器作为目的主机，通过源主机发送数据，数据先通过应用层和控制信息形成报文流，传递给传输层；通过传输层将数据和控制信息形成报文段、数据报，再传到网际层；通过网际层和控制信息形成 IP 数据报，再传到网络接口层；网络接口层将 IP

四、训练结果

通过搭载深度学习模型对司机不规范驾驶状态图像模型进行了训练，对司机打电话的 100 张不同的图像分别训练 10 次，对司机玩手机的 100 张不同的图像分别训练 10 次，以此类推，对司机 7 种不规范驾驶状态则总共训练了 7000 次，实验训练结

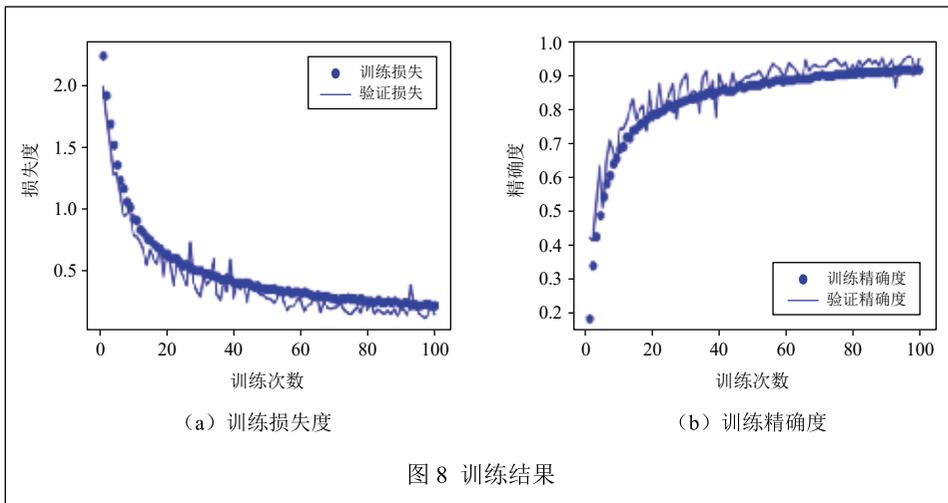
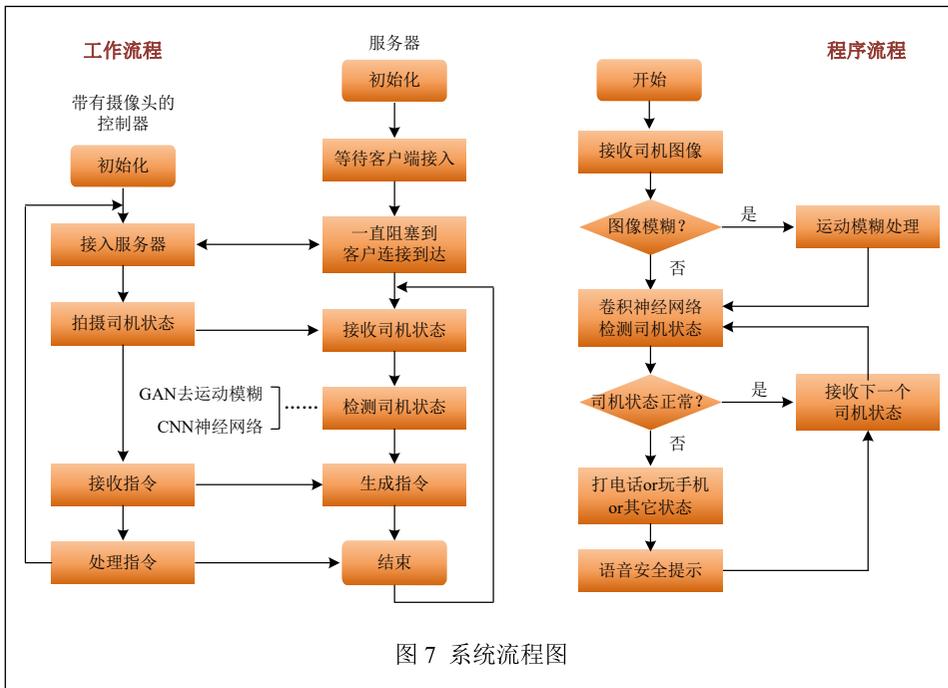


表1 系统的测试结果

司机行为	训练次数	训练精度 (%)	测试次数	语音正确提示次数	测试精度 (%)	语音提示
打电话	1000	约 96	50	47	94	请勿打电话, 注意行车安全
玩手机	1000	约 95	50	46	92	请勿玩手机, 注意行车安全
操作音响	1000	约 92	50	43	86	请勿操作音响, 注意行车安全
与乘客聊天	1000	约 92	50	42	84	请勿聊天, 注意行车安全
喝水	1000	约 80	50	38	76	喝水时, 请注意行车安全
弄头发和化妆	1000	约 90	50	43	86	请勿弄头发和化妆, 注意行车安全
向后伸手	1000	约 85	50	39	78	请勿向后伸手, 注意行车安全
其他情况	无	无	无	无	无	无

果如图8所示。随着训练次数的增加, 损失度不断降低, 精确度逐渐提高, 最后平均达到90%。又在车内进行了测试, 每种动作行为分别测试了50次, 由于车里面现场环境等各种因素的影响, 车内测试精确度稍低于实验训练的精确度。如表1所示是系统的测试结果, 当司机在开车中有危险动作行为时, 对司机的语音提示如表格中右侧所示。

五、结束语

本装置结构简单, 操作简便, 成本低, 易安装, 功能强。在任何车中都可以安装, 有助于提高司机安全意识, 减少交通事故的发生。通过应用到公共交通系统中, 出租车公司或公交车公司可以后台实时监测司机驾驶状态, 可以对司机起到警示作用, 并且后台可以保存图片和数据, 可以统计不同司机驾驶状态数据集。

参考文献

[1] 汽车与安全编辑部. 拒绝分心驾驶 为生命安全负责——视觉分心: 危险在不经意间来袭 [J]. 汽车与安全, 2016, (10): 42-46.

[2] 张曰花, 王红, 马广明. 基于深度学习的图像识别研究 [J]. 现代信息科技, 2019, 3(11): 111-112, 114.

[3] 王禹嫣. 深度学习在图像识别中的应用与发展 [J]. 中国战略新兴产业, 2017, 132(48): 127-128.

[4] HUBEL D H, WIESEL T N. Receptive fields, binocular

interaction and functional architecture in the cat's visual cortex[J]. *The Journal of Physiology*, 1962, 160(1): 106-154.

[5] FUKUSHIMA K, MIYAKE S. Neocognitron: A new algorithm for pattern recognition tolerant of deformations and shifts in position[J]. *Pattern Recognition*, 1982, 15(6): 455-469.

[6] 田文洪, 曾柯铭, 莫中勤, 等. 基于卷积神经网络的驾驶员不安全行为识别[J]. *电子科技大学学报*, 2019, 48(03): 63-69.

[7] KUPYN O, BUDZAN V, MYKHAILYCH M, et al. DeblurGAN: Blind Motion Deblurring Using Conditional Adversarial Networks[C]// *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2018: 8183-8192.

[8] 高华照. 基于深度卷积神经网络的图像识别算法研究[D]. 长春: 吉林大学, 2018.

[9] 钱程. 基于深度学习的人脸识别技术研究[D]. 成都: 西南交通大学, 2017.

[10] 李俊俊, 杨华民, 张澍裕, 等. 基于神经网络融合的司机违规行为识别[J]. *计算机应用与软件*, 2018, 35(12): 228-233, 325.

[11] 杨林川. 基于深度神经网络的司机行为识别技术与实现[D]. 成都: 电子科技大学, 2018.

[12] 吕焦盛. 基于TCP/IP协议的网络通信服务器设计与实现[J]. *赤峰学院学报(自然科学版)*, 2019, (09): 50-52.

[13] 张彦青. 基于TCP/IP协议的网络通信服务器设计[J]. *中国战略新兴产业*, 2018, (36): 99.

[14] 卓胜华. 基于机器视觉的违规驾驶行为检测研究[D]. 广州: 华南理工大学, 2012.

[15] 王瑞. 基于卷积神经网络的图像识别[D]. 开封: 河南大学, 2015.

Irregular Driving Behavior Intelligent Recognition System Based on Deep Learning

LI Guang-dong, CHENG Jin, ZHOU Zi-xiao, LIU Bao-yu, WANG Jun-qi, SHI Li-gen, YAO Yu-jun, CHANG

Chuang-chuang, LI Wen-yue, ZOU Xiao-ping

(*Beijing Key Laboratory of Sensors, Beijing Information Science and Technology University, Beijing 100101, China*)

Abstract: With the increasing number of buses and taxis, drivers' irregular driving behaviors while driving cause many traffic accidents. In order to effectively reduce the possibility of such traffic accidents, an irregular driving behavior intelligent recognition system is designed based on deep learning. A camera with a controller is installed in the car to collect images of the driver's driving status in real time, and the controller is remotely connected to the background computer server. The server receives the collected images through network communication, identifies and processes these images based on deep learning, and then transmits the processed results to the controller. In this way, 7 kinds of dangerous actions while driving, such as drinking water, talking or playing on the phone and so on, can be promptly and intelligently given voice prompts to the driver under the control of the controller in the car, and the background server can also monitor and alert the driver in real time.

Key words: irregular driving actions; network communication; deep learning; image processing; intelligent safety voice prompts

作者简介

李光东: 北京信息科技大学, 硕士研究生, 研究方向为物联网技术。
通信地址: 北京北四环 35 号教 2 楼 117 实验室

邮编: 100101

邮箱: 1511455720@qq.com

程进: 北京信息科技大学, 北京市传感器重点实验室, 副研究员, 主要从事激光光纤声学传感器方面的研究。

周子肖: 北京信息科技大学, 硕士研究生, 研究方向为物联网技术。

刘宝宇: 北京信息科技大学, 硕士研究生, 研究方向为物联网技术。

王骏琦: 北京信息科技大学, 硕士研究生, 研究方向为物联网技术。

史立根: 北京信息科技大学, 硕士研究生, 研究方向为物联网技术。

姚宇军: 北京信息科技大学, 硕士研究生, 研究方向为物联网技术。

常创创: 北京信息科技大学, 硕士研究生, 研究方向为物联网技术。

李文月: 北京信息科技大学, 硕士研究生, 研究方向为物联网技术。

邹小平: 北京信息科技大学硕士生导师, 研究员, 主要从事物联网方面的研究。