

基于多尺度卷积网络的司机嘴部异常检测^①



寿博, 朱明

(中国科学技术大学 网络空间安全学院, 合肥 230026)

通讯作者: 寿博, E-mail: vnkok@mail.ustc.edu.cn

摘要: 司机的异常驾驶行为会引发交通事故从而威胁司机、乘客及他人的生命安全, 因此检测司机的异常驾驶行为可以有效保障人们的出行安全. 在实际驾驶过程中, 发生在司机嘴部区域的异常行为复杂多样, 因此本文提出一种无监督检测算法来检测嘴部区域的异常行为. 算法首先使用人脸关键点检测网络获取嘴部异常高发区域; 之后通过改进后的 Convolutional Auto-Encoder (CAE) 算法重构嘴部区域图片并计算重构误差以此来判断是否发生异常. 包括以下 3 点改进: 加入 skip connect 结构用来更好地重构输入图片; 加入 Inception 结构并调整优化分支通道比例, 用来更好的拟合输入图片特征; 在训练时加入高斯白噪声进一步提高模型检测的鲁棒性. 实验结果表明: 本文提出的算法框架与传统 CAE 算法相比, AUC 从 0.682 提高到 0.938, 并且能够在嵌入式系统上运行.

关键词: 异常检测; 分心驾驶; 无监督学习; 卷积自编码器 (CAE); CNN

引用格式: 寿博, 朱明. 基于多尺度卷积网络的司机嘴部异常检测. 计算机系统应用, 2021, 30(11): 224-230. <http://www.c-s-a.org.cn/1003-3254/8192.html>

Driver's Mouth Anomaly Detection Based on Multiscale Convolutional Auto-Encoder

SHOU Bo, ZHU Ming

(Cyberspace Security Academy, University of Science and Technology of China, Hefei 230026, China)

Abstract: Abnormal driving behaviors of drivers pose a high risk of traffic accidents, threatening the life safety of drivers, passengers, and others. Thus, detecting the abnormal driving behaviors of drivers is of great significance for ensuring people's travel safety. In an actual driving process, abnormal behaviors in the driver's mouth region are complex and diverse. In view of this, this study proposes an unsupervised detection algorithm for the abnormal behaviors in the mouth region. The algorithm first uses the facial landmark detection network to obtain the mouth area with a high probability of anomaly. Then, the mouth area image is rebuilt with the improved Convolutional Auto-Encoder (CAE) algorithm. Abnormal behaviors are determined by the computation of the reconstruction error. The proposed algorithm is improved in three aspects: (1) the introduction of the skip connection structure to better reconstruct the input image; (2) the introduction of the Inception structure and the optimization of the proportions of branch channels to better fit the features of the input image; (3) the addition of Gaussian white noise in the training process to improve the robustness of model detection. The experimental results show that the AUC of the proposed algorithm framework is improved from 0.682 to 0.938 as compared with the traditional CAE algorithm and it can run on embedded systems.

Key words: anomaly detection; distracted driving; unsupervised learning; Convolutional Auto-Encoder (CAE); CNN

① 基金项目: 安徽省 2019 年重点研究与开发计划 (201904a05020035)

Foundation item: Key Research Plan Project of Anhui Province in 2019 (201904a05020035)

收稿时间: 2021-01-30; 修改时间: 2021-02-26; 采用时间: 2021-03-24; csa 在线出版时间: 2021-10-22

1 引言

随着我国经济的平稳发展,交通基础设施的不断完善,人民的物质生活水平的不断提高,自驾出行成了越来越多人选择的交通方式.汽车行业得以迅猛发展的同时也带来了交通事故频发的安全问题.

异常驾驶,是指一切与正常驾驶行为不同的驾驶行为,包括所有分散正常驾驶注意力的所有行为,通常包括打电话,喝水,抽烟等行为.由于打电话,喝水,抽烟等行为导致驾驶员在驾驶时注意力难以集中,此时遇到突发状况,驾驶员会出现反应不及时甚至陷入呆滞与慌乱,以至于出现判断失误以及操作失误,造成交通事故.然而驾驶员的异常驾驶行为是可控的,因此准确并且快速地检测出驾驶员的异常驾驶行为,可以有效地保障司机的出行安全,长期还能让司机养成规范驾驶的好习惯.

目前的基于视频监控的异常驾驶检测技术主要分为以下3类:第1类是使用图像特征分析的方法,这类方法往往使用图像处理技术直接分析图片本身特征,比如肤色,边缘,纹理等特征来判断驾驶员是否存在异常行为^[1-3].这类方法往往着眼于像素单位,对采集到的图片清晰度要求很高,优点在于计算简单,检测速度快,但是容易受到光线的干扰,鲁棒性不高,适合作为异常检测的辅助手段.第2类是基于有监督学习的方法,这类方法往往都有分类具体、标签明确的异常动作图片来做训练集,研究的主要目标也都局限在检测训练过的异常动作上^[4-6].这种方法的缺点在于鲁棒性较差,不能检测出训练集之外的异常驾驶行为,往往只能检测某种单一的异常驾驶行为,而且缺乏足够的训练集来训练.第3类是基于无监督学习的方法,这类方法一般只需要正常驾驶的图片来训练,模型的目标是筛选出所有与正常样本不同的数据样本点,也可以看作是一种单分类算法^[7].这类方法的优点在于对未知的异常行为检测鲁棒性较高,但是检测效果没有有监督的分类检测方法好.基于以上问题,本文提出了一种基于深度多尺度卷积网络的驾驶员嘴部异常检测算法,该算法先通过人脸关键点检测网络多任务级联卷积网络(Multi-Task cascaded Convolutional Neural Network, MTCNN)^[8]来提取驾驶员人脸附近的异常高发区域,然后针对这一区域设计了改进后的卷积自编码器(Convolutional Auto-Encoder, CAE)网络结构来训练并检测嘴部区域发生的任何异常行为.

本文贡献如下:(1)在传统的CAE算法上做出3点改进:1)编码器端加入多尺度卷积结构,并且实验调优了多尺度卷积分支通道比例,使得模型能够更好地拟合输入图片特征;2)加入skip connect结构,使模型能够更好地重构图片;3)训练时加入噪声,使得模型检测的鲁棒性大大提高.(2)使用RGB摄像头拍摄并建立了司机异常驾驶数据集.

2 相关工作

2.1 人脸关键点检测

因为所检测的异常发生区域都与人脸呈高度相关,所以我们首先需要对驾驶员面部区域做实时定位,再根据人脸关键点位置,得到相对原图较小的嘴部区域图片(如图1所示),通过这种方式可以减少大部分背景的干扰,有利于提高检测异常准确度.



图1 嘴部区域提取

对于人脸关键点检测,我们使用了MTCNN模型进行识别^[8].MTCNN是2016年中国科学院深圳研究院提出的用于人脸检测任务的多任务神经网络模型,该模型主要采用了3个级联的网络,采用候选框加分类器的思想,进行快速高效的人脸检测.对于给定的一张包含人脸的原始图片,经过MTCNN网络处理,最终输出人脸框和人脸的5个特征点,包括左眼、右眼、鼻子、嘴唇和左右嘴角的坐标信息,通过这些信息能得到待检测的嘴部区域.

2.2 Autoencoder

Autoencoder,自编码器是传统的无监督学习算法,它尝试学习一个重构输入的函数.Autoencoder接收一个输入向量 $x \in R^d$ 然后先将这个输入向量通过非线性函数式(1)映射到一个向量 $h \in R^{d'}$,通常 $d < d'$, h 也被称为隐藏层表示.函数 σ 被称为编码器,其参数 θ 包括权重 W ,偏置 b .

$$h = f_{\theta} = \sigma(Wx + b) \quad (1)$$

$$\theta = \{W, b\} \quad (2)$$

解码器将隐藏层表示 h 作为输入,通过反向映射函数式(3)得到重构向量 y ,其中 y 与 x 维度一致.

$$y = f_{\theta'}(h) = \sigma(W'h + b') \quad (3)$$

$$\theta' = \{W', b'\} \tag{4}$$

编码器与解码器的参数 θ, θ' 通过优化目标函数使得重构误差最小而得到, 训练集为 D_n , 其中 x_n 表示输入信号, y_n 表示重构信号, 如式(5)、式(6)所示. 损失函数 L 可以选择均方误差函数、联合熵等常见损失函数. 通过训练后的自编码器能够通过计算输入与输出的重构误差来检测与训练数据不同的数据, 即能检测异常数据.

$$\theta = \theta' = \arg \min L(x, f_{\theta'}(f_{\theta}(x))) \tag{5}$$

$$D_n = \{(x_0, y_0), \dots, (x_n, y_n)\} \tag{6}$$

传统的神经网络自编码器由全连接层组成, 忽略了二维图像结构, 而 CAE 使用卷积神经网络代替全连接层, 使得网络的权重在输入的所有位置之间共享, 保留了空间信息. CAE 的编码器映射函数如式(7)所示, h^k 表示第 k 个特征图的隐藏层输出, $*$ 表示 2D 卷积操作, σ 表示非线性激活函数, w^k 表示卷积网络的权重, 单一偏置 b^k 被整个特征图共同使用.

$$h^k = \sigma(x * W^k + b^k) \tag{7}$$

解码器的映射函数如式(8)所示, 输入是编码器的输出 h^k , 权重为 \widetilde{W}^k , 对于每一个输入通道共享同一个偏置 c , y 表示重构后的输出. 卷积自编码器更适合完成图片异常检测任务^[9].

$$y = \sigma\left(\sum_{k \in H} h^k * \widetilde{W}^k + c\right) \tag{8}$$

传统的 Autoencoder 模型分类中有一类编码器被称为降噪自编码器 (Denoising Auto-Encoder, DAE)^[10]. 和自编码器不同的是, 降噪自编码在训练过程中, 输入数

据是训练数据和噪声的叠加, 即相当于输入数据被人为地“损坏”, 如式(9)所示降噪自编码器的编码器与式(1)不同, 输入为 $x+x'$, 解码器与 loss function 均与式(3)、式(5)相同. x' 即为噪声, 通常使用服从标准正太分布的随机噪声, 其一维概率密度如式(10)所示. 降噪编码器的核心思路是一个能够将被“损坏”的数据恢复正常自编码器比一个只能从正常数据恢复出正常数据的自编码器优秀, 即降噪自编码器能够找到训练数据更好地表示特征. 从信息论的角度解释即有噪声的数据转变成无噪声数据, 训练过程中信息量增大, 即降噪编码器隐藏层中保留了更多信息.

$$h = f_{\theta} = \sigma(W(x+x') + b) \tag{9}$$

$$p(x) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right) \tag{10}$$

3 算法

3.1 算法框架

本算法由人脸关键点识别网络 MTCNN 和改进后的 CAE 组成, 如图 2 所示, 由 MTCNN 检测人脸框和人脸关键点信息进而得到嘴部区域图片, 然后将嘴部区域图片作为输入, 改进后的 CAE 模型重构嘴部区域图片, 通过计算重构误差, 最终输出嘴部区域是否存在异常. 改进后的 CAE 框架结构如图 3 所示, 模型的输入为 $1 \times 128 \times 128$ 的灰度图片 X_{input} , 经过 7 次多尺度卷积下采样 (灰色背景框内的网络结构重复 3 次), 得到 $512 \times 1 \times 1$ 的隐藏层特征表示, 再经过 7 次反卷积上采样得到 $1 \times 128 \times 128$ 的重构图片, 模型的结构细节如图 1 所示.

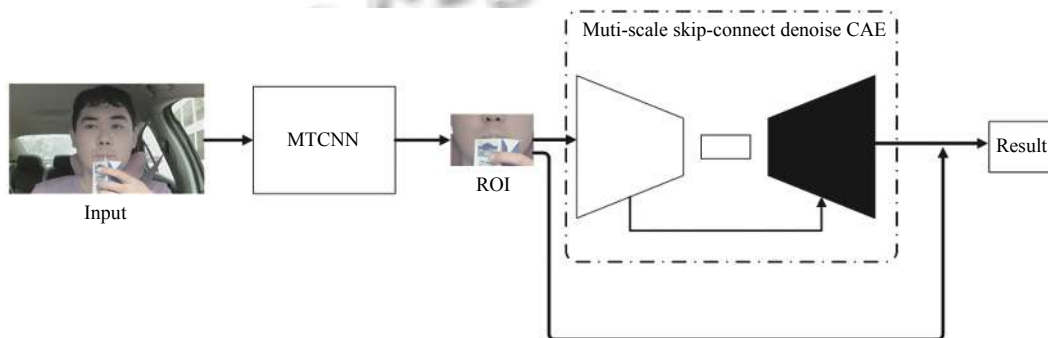


图 2 算法整体框架

3.2 多尺度卷积和池化层

为了提高对不同尺度的异常物体 (如图 4 所示) 的检测成功率, 受到 GoogLeNet 网络^[11]的启发, 本文对

编码器的单层卷积网络采用了与 Inception 相似的结构, 融合了 4 种不同尺度的卷积核, 能够找到更优化的局部特征与全局特征.

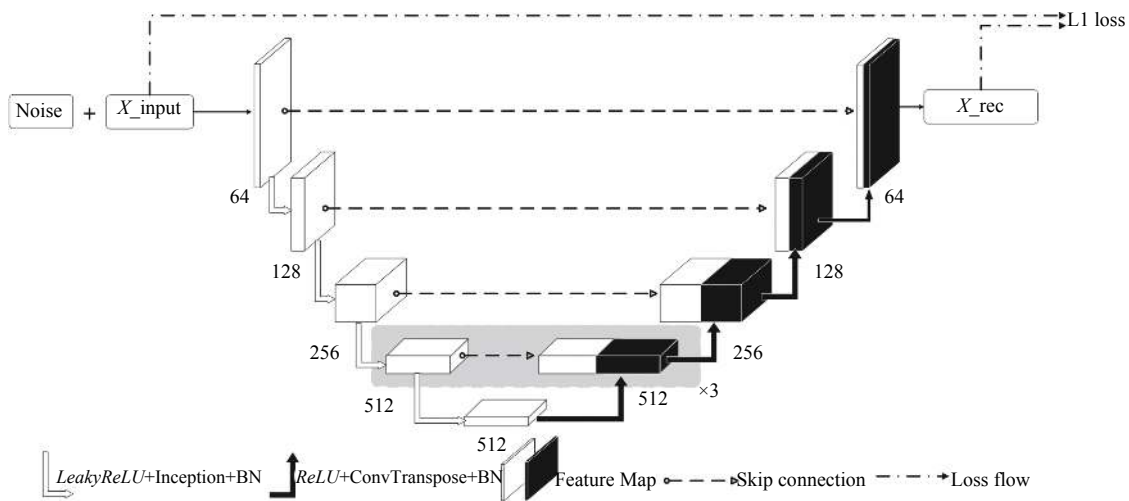


图3 Multi-Scale Skipconnect Denoising CAE

表1 模型结构细节

网络组成	结构细节
Encoder	Inception(1,64)
	LeakyReLU(0.2)+Inception(64,128)+BN
	LeakyReLU(0.2)+Inception(128,256)+BN
	LeakyReLU(0.2)+Inception(256,512)+BN
	LeakyReLU(0.2)+Inception(512,512)+BN
Decoder	LeakyReLU(0.2)+Inception(512,512)
	ReLU+ConTranspose(512,512)+BN
	ReLU+ConTranspose(1024,512)+BN
	ReLU+ConTranspose(1024,512)+BN
	ReLU+ConTranspose(1024,512)+BN
	ReLU+ConTranspose(512,128)+BN
	ReLU+ConTranspose(256,64)+BN
ReLU+ConTranspose(128,1)+tanh	

3个分支再接2×2 MaxPooling层. 两个3×3卷积的堆叠与一个5×5卷积的感受野相同, 3个3×3卷积的堆叠与一个7×7卷积的感受野相同, 利用这种方式既可以减少网络的权重参数, 又可以增加网络的深度, 从而更好地提取特征^[12,13].

如图3所示, 在编码器的7次下采样中, 每一层的Inception结构都需要确定图5中从左到右4种分支的通道比例. 经过笔者的不断调参实验, 最终确定的Inception分支通道比例如表2所示. 前两层分支通道比例为4:2:1:1, 后5层分支通道比例为1:1:1:1, 实验结果表明在较底层下采样时适当增加1×1分支和3×3分支的比例能够更好提高的异常检测效果, 原因在于这种通道比例能够更细致的刻画输入图片的局部特征.



图4 嘴部区域异常示例

单层卷积和池化结构如图5所示, 由4种卷积和池化层concat而成, 分别为2×2 MaxPooling层接1×1卷积层; 1×1卷积层接3×3卷积层; 1×1卷积层接两个3×3卷积层; 1×1卷积层接3个3×3卷积层, 后面

表2 Inception分支通道比例

层数	第1分支	第2分支	第3分支	第4分支	通道总数
1	32	16	8	8	64
2	64	32	16	16	128
3	64	64	64	64	256
4	128	128	128	128	512
5-7	128	128	128	128	512

3.3 Skip connect 结构

为了更好的重构图片 X_{rec} , 受到 U-net 网络^[14] 的启发, 我们在多尺度卷积自编码器的基础上, 加入 skip connect 结构. 如图3所示, 在编码器网络中, 每一次下采样得到的 Feature Map 都传递到相对应解码器网络层, 解码器网络每一次上采样时都在原有的输入中叠加由编码器传递的 Feature Map. 例如, 第1层下采样输出 Feature Map 为 $64 \times 64 \times 64$, 与之对应的第7层上

采样的输入 Feature Map 大小为 $64 \times 64 \times 64$, 两者叠加成大小为 $128 \times 64 \times 64$ 的 Feature Map 进行第 7 层上采样操作. 由于从低层到高层都加入了 skip connect 结构, 输入图片的全局特征和局部特征都得到了更好的保留.

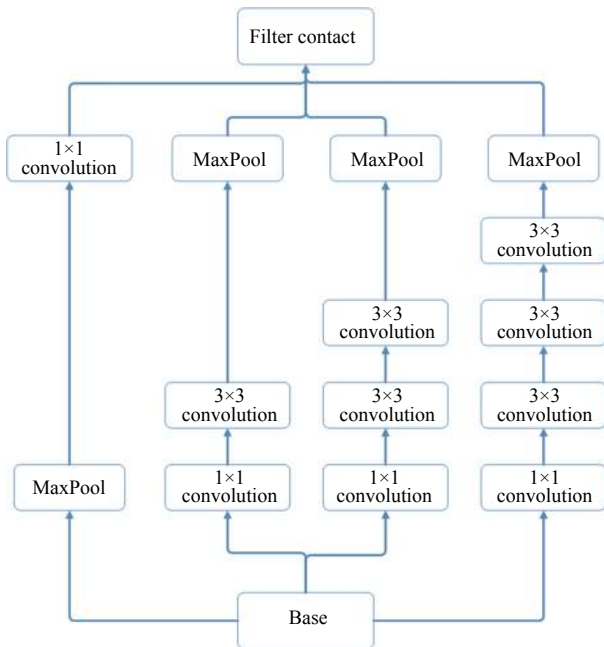


图 5 编码器单层卷积池化结构

3.4 激活函数和目标函数

解码器与编码器每层卷积与反卷积之后都有激活函数, 编码器采用的激活函数为 *LeakyReLU*, 解码器的最后一层的激活函数为 *tanh*, 其他层的激活函数为 *ReLU*.

$$ReLU(x) = \max(0, x) \quad (11)$$

ReLU 数学表达式如式 (11) 所示, 当输入值为负时输出 0, *LeakyReLU* 数学表达式如式 (12) 所示, 输入值为负时输出 ax , 系数 a 选定为 0.2, *tanh* 数学表达式如式 (13) 所示, 将网络的输出映射到 $[-1, 1]$. 解码器输出与输入图像维度相同的单通道矩阵. 如式 (14) 所示, 目标函数 L 为 L1 Loss, label 值 $x_{i,j}$ 与输入矩阵相同, $x'_{i,j}$ 表示重构矩阵的值, m 和 n 表示图像矩阵的行和列.

$$LeakyReLU(x) = \begin{cases} x, & x \geq 0 \\ ax, & x < 0 \end{cases} \quad (12)$$

$$\tanh(x) = \frac{1 - e^{-2x}}{1 + e^{-2x}} \quad (13)$$

$$L = \frac{1}{m} \frac{1}{n} \sum_{i=1}^m \sum_{j=1}^n |x_{i,j} - x'_{i,j}| \quad (14)$$

4 实验分析

4.1 数据集

本文采用的数据集是笔者自己采集的数据集. 采集装置在车辆驾驶室正前方如图 6 所示. 笔者使用 1280×720 分辨率的 RGB 摄像头拍摄了一个数据集, 整个数据集采集了 10 个人的行车视频录像, 每人采集两段视频, 第一段正常驾驶视频, 时长 20 min, 从中分别抽取了 4000 帧图片, 再使用 MTCNN 检测人脸区域得到嘴部区域, 剔除成像质量过差和人脸检测失败的图片, 最终获得 3600 帧嘴部区域图片, 最后 *resize* 到 128×128 大小, 用于模型的训练. 第二段视频每隔 30 s 做出抽烟, 喝水, 打哈欠, 吃食物等任意异常行为, 时长 10 min, 用同样的方法得到了 600 帧正常驾驶的嘴部区域图片和 600 帧嘴部区域异常图片, 用于测试模型的检测效果. 最终从所有视频中获取了 48 000 张图片, 其中 75% 用于训练, 25% 用于测试.



图 6 数据集采集装置

4.2 实验过程

笔者在只含有正常驾驶图片的训练集上训练模型, 在既含有异常驾驶图片又含有正常驾驶图片的测试集上评价模型. 评价的标准为 AUC, AUC 被定义为 ROC (Receiver Operating characteristic Curve) 曲线下与坐标轴围成的面积, 其值越大表示分类效果越好. 其中 ROC 曲线是根据一系列不同的阈值, 以真阳性率 (True Positive Ratio, TPR) 为纵坐标, 假阳性率 (False Positive Ratio, FPR) 为横坐标绘制的曲线. 本模型的阈值即为图片的重构误差, 计算方式为 L1 Loss (式 (12)), 值越大代表异常的可能性越高. 如图 3 所示, 训练时输入叠加了一个与输入矩阵相同大小的服从标准正态分布的随机噪声, 用以提高模型检测的鲁棒性. 学习率设置为

0.002, 训练的 epoch 为 20, 为了防止过拟合的发生, 当训练过程中 Loss 不再下降, 立刻停止训练, 记录此前的 best AUC 作为实验结果。

4.3 实验结果

为了验证提出的改进方法的效果, 笔者做了模型自身的纵向对比实验, 实验结果如表 3 所示, 可以看到每处改进对实验结果的提升。表中 CAE 为与模型结构相似的卷积自编码器, 其中解码器完全一致, 编码器用 kernel_size=4, stride=2, padding=1 卷积代替原模型的 Inception 操作。表中 Inception(Mix ratio) 即为表 2 所示的最佳分支通道比例。

表 3 纵向对比实验结果

网络优化结构	AUC
CAE	0.693
CAE+skip-connect	0.711
CAE+skip-connect+Inception(1:1:1:1)	0.776
CAE+skip-connect+Inception(Mix ratio)	0.798
CAE+skip-connect+Inception(Mix ratio)+Denoise	0.938

为了验证此模型的有效性, 笔者也与其他无监督检测方法做了横向对比实验, 结果如表 4 所示, 可以看到模型的异常检测结果优秀。

表 4 横向对比实验结果

方法	AUC
CAE	0.693
HOG+OCSVM	0.762
Skip-GANomaly ^[15]	0.921
Proposed Methods	0.938

以约登指数寻找最佳 ROC 曲线阈值, 既寻找 ROC 曲线上横坐标与纵坐标差异最大的点, 以这一点的阈值 t 作为分类标准, 重构误差小于 t 视为正常驾驶, 重构误差大于 t 视为异常驾驶。以检测异常驾驶作为 True Positives, 在 12 000 张测试图片上实验检测效果, 给出实验结果的混淆矩阵如表 5 所示。

表 5 实验结果的混淆矩阵

	True	False
Positive	5008	992
Negative	5273	727

4.4 算法时间复杂度分析

模型使用如表 6 所示的配置机器上训练, 为了证明该算法的实际应用效果, 笔者将训练好的模型布置在嵌入式开发板 NVIDIA Jetson TX2 (配置见表 7) 上

测试检测异常所需时间, 训练用时和测试用时见表 8、表 9 所示, 在嵌入式系统上检测异常的帧率达到了 3.38 fps, 基本满足检测需求。

表 6 训练所用机器配置表

类型	型号	参数
系统	Ubuntu	16.04.12LTS
CPU	Intel(R) Core(TM) i7-7800X 1080 Ti	6核
显卡	NVIDIA GeForce GTX 1080 Ti	8 GB
内存	DDR4	32 GB

表 7 NVIDIA Jetson TX2 配置

类型	型号	参数
系统	Ubuntu	18.04LTS
CPU	NVIDIA Denser and ARM Cortex-A57	6核
显卡	NVIDIA Tegra X2	8 GB
内存	LPDDR4	8 GB

表 8 训练模型时间复杂度分析 (Frame=36 000)

训练/检测	时间 (s)	帧率 (fps)
Face/Mouth Detection	3420.54	10.52
Model Training	176.79	203.62
Total	3596.76	10.00

表 9 模型检测时间复杂度分析 (Frame=12 000)

检测	时间 (s)	帧率 (fps)
Face/Mouth Detection	3508.77	3.42
Model Decision	34.85	344.33
Total	3543.62	3.38

5 结束语

本文提出了一种基于多尺度卷积自编码器的驾驶员嘴部异常检测方法, 在检测之前先利用人脸关键点检测方法 MTCNN 筛选嘴部区域, 然后使用改进后的 CAE 进行异常检测, 实验表明加入多尺度卷积结构并优化分支通道比例、加入 skip connect 结构、训练时加入随机噪声等改进方法能够显著提高异常检测效果; 与其他无监督检测算法相比, 本文提出的算法表现优异; 在嵌入式开发平台 NVIDIA Jetson TX2 上的算法复杂度实验表明该算法具有实际应用价值。

参考文献

- 成波, 屈肖蕾, 张波. 基于肤色范围的驾驶员手持电话行为检测方法. CN: 20120004371.8. 2012-01-09.
- Omidyeganeh M, Javadtalab A, Shirmohammadi S.

- Intelligent driver drowsiness detection through fusion of yawning and eye closure. 2011 IEEE International Conference on Virtual Environments, Human-Computer Interfaces and Measurement Systems Proceedings. Ottawa: IEEE, 2011. 1–6.
- 3 Artan Y, Balci B, Elihoş A, *et al.* Vision based driver smoking behavior detection using surveillance camera images. In: Ricci E, Bulò SR, Snoek C, *et al.*, eds. Image Analysis and Processing. Cham: Springer, 2019. 468–476.
 - 4 Xing Y, Lyu C, Zhang ZZ, *et al.* Identification and analysis of driver postures for in-vehicle driving activities and secondary tasks recognition. IEEE Transactions on Computational Social Systems, 2018, 5(1): 95–108. [doi: [10.1109/TCSS.2017.2766884](https://doi.org/10.1109/TCSS.2017.2766884)]
 - 5 Huang W, Liu X, Luo MY, *et al.* Video-based abnormal driving behavior detection via deep learning fusions. IEEE Access, 7: 64571–64582
 - 6 姚巍巍, 张洁. 基于模型剪枝和半精度加速改进 YOLOv3-tiny 算法的实时司机违章行为检测. 计算机系统应用, 2020, 29(4): 41–47. [doi: [10.15888/j.cnki.csa.007348](https://doi.org/10.15888/j.cnki.csa.007348)]
 - 7 Chiou CY, Chung PC, Huang CR, *et al.* Abnormal driving behavior detection using sparse representation. 2016 International Computer Symposium (ICS). Chiayi: IEEE, 2016. 390–395.
 - 8 Zhang KP, Zhang ZP, Li ZF, *et al.* Joint face detection and alignment using multitask cascaded convolutional networks. IEEE Signal Processing Letters, 2016, 23(10): 1499–1503. [doi: [10.1109/LSP.2016.2603342](https://doi.org/10.1109/LSP.2016.2603342)]
 - 9 Masci J, Meier U, Cireşan D, *et al.* Stacked convolutional auto-encoders for hierarchical feature extraction. 21st International Conference on Artificial Neural Networks and Machine Learning. Espoo: Springer, 2011. 52–59.
 - 10 Vincent P, Larochelle H, Lajoie I, *et al.* Stacked denoising autoencoders: Learning useful representations in a deep network with a local denoising criterion. The Journal of Machine Learning Research, 2010, 11: 3371–3408.
 - 11 Szegedy C, Liu W, Jia YQ, *et al.* Going deeper with convolutions. 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Boston: IEEE, 2015. 1–9.
 - 12 Luo WJ, Li YJ, Urtaşun R, *et al.* Understanding the effective receptive field in deep convolutional neural networks. arXiv: 1701.04128, 2017.
 - 13 Szegedy C, Vanhoucke V, Ioffe S, *et al.* Rethinking the inception architecture for computer vision. 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Las Vegas: IEEE, 2016. 2818–2826.
 - 14 Ronneberger O, Fischer P, Brox T. U-net: Convolutional networks for biomedical image segmentation. 18th International Conference on Medical Image Computing and Computer-Assisted Intervention. Munich: Springer, 2015. 234–241.
 - 15 Akçay S, Atapour-Abarghouei A, Breckon TP. Skip-GANomaly: Skip connected and adversarially trained encoder-decoder anomaly detection. 2019 International Joint Conference on Neural Networks (IJCNN). Budapest: IEEE, 2019. 1–8.