

文章编号: 2095-2163(2021)09-0094-05

中图分类号: TP391.4

文献标志码: A

# 基于 MTCNN 的多特征融合学生疲劳检测算法研究

陈 藩, 施一萍, 胡佳玲, 谢思雅, 刘 瑾

(上海工程技术大学 电子电气工程学院, 上海 201620)

**摘要:** 目前疲劳检测主要是通过眼部 PERCLOS 值来判断, 但是这种方法检测疲劳特征较为单一, 影响了检测的准确率。本文提出一种基于卷积神经网络的多特征融合学生疲劳检测算法, 首先用 MTCNN 对人脸进行关键点定位, 在此基础上用人脸归一化的方式精准提取眼睛和嘴部的特征图像并进行眼部定位; 其次, 构建眼、嘴数据集, 完成眼、嘴部状态分类模型训练; 最后, 用训练好的模型将眼、嘴部的疲劳特征相融合并根据改良的 MAR 值判断方法进行疲劳检测。实验结果表明该方法的准确率达到了 96.2%, 实时性也得到了极大的改善。

**关键词:** MTCNN; 疲劳检测; 眼部定位; 归一化; 特征融合; MAR 值

## Research on multi feature fusion algorithm for student fatigue detection based on MTCNN

CHEN Fan, SHI Yiping, HU Jialing, XIE Siya, LIU Jin

(School of Electronic and Electrical engineering, Shanghai University of Engineering and Technology, Shanghai 201620, China)

**[Abstract]** The fatigue detection is mainly judged by the PERCLOS value of the eye, but this kind of characteristics is relatively single, which affects the accuracy of the detection. In this paper, a multi feature fusion student fatigue detection algorithm based on convolutional neural network is proposed. Firstly, MTCNN is used to locate the key points of the face. On this basis, a face normalization method is proposed to accurately extract the feature images of the eyes and mouth and locate the eyes. Then the eye and mouth dataset is constructed to complete the training of eye and mouth state classification model. Finally, the trained model is used to integrate the fatigue characteristics of eyes and mouth, and the fatigue detection is carried out according to the improved MAR value judgment method. The experimental results show that the accuracy of this method is 96.2%, and the real-time performance is greatly improved.

**[Key words]** MTCNN; fatigue detection; eye location; normalization; feature fusion; MAR value

## 0 引言

在课堂教学中, 学生会出现长时间精神不集中的疲劳状态。如果能够及时检测到学生课堂疲劳状态, 并进行有效的处理, 就能够帮助学生提高课堂的学习效率, 提高课堂教学质量, 左国才等人将 CNN 疲劳检测方法用于大学课堂之中<sup>[1]</sup>。疲劳检测为智慧课堂的实现提供了重要的保证。传统的疲劳检测只是通过眼部的 PERCLOS 值来判断, 判断标准较为单一(特征), 结果不够准确<sup>[2-3]</sup>。也有一些学者如陈建伟提出了视觉特征多通道融合的方式实现疲劳检测, 但是鲁棒性不强<sup>[4]</sup>; 传统的人脸识别方法存在容易被光照阴影等因素干扰的问题, 江伟坚等人提出新 Haar-like 特征的 Adaboost 人脸检测, 但有较高的时间开销<sup>[5]</sup>; 如今这些疲劳检测方法主要

应用于疲劳驾驶领域, 如褚晶辉、耿磊等人均提出基于驾驶员行为分析疲劳检测方法, 然而在教育领域却较少被应用<sup>[6-7]</sup>。本文提出一种基于卷积神经网络的多特征融合的疲劳检测算法, 实现学生疲劳检测。首先, 通过 MTCNN<sup>[8]</sup>实现人脸检测和特征点定位, 相对于基于多尺度卷积网络, MTCNN 有着更强的关键点定位的能力; 其次, 通过人脸归一化的方式精确提取眼、嘴部特征区域, 用神经网络对特征区域的图像分类; 最后, 将眼部 PERCLOS 值与嘴部 MAR 值相结合, 实现疲劳检测, 并在实际教室环境中取得了理想的效果。

## 1 算法原理

本文提出的疲劳检测算法的总体流程如图 1 所示。

**基金项目:** 国家自然科学基金(61701296)。

**作者简介:** 陈 藩(1994-), 男, 硕士研究生, 主要研究方向: 深度学习和图像处理; 施一萍(1964-), 女, 硕士, 副教授, 硕士生导师, 主要研究方向: 深度学习和智能控制。

**通讯作者:** 施一萍 Email: syp@sues.edu.cn

**收稿日期:** 2021-01-30

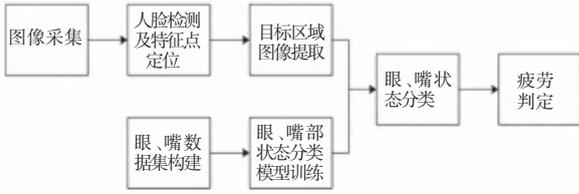


图 1 疲劳检测流程

Fig. 1 Fatigue detection process

1.1 人脸检测及关键点定位

疲劳检测过程中, 第一点便是要对学生的面部进行准确的检测和关键点的定位。由于人的面部容易受到光照、发型、佩戴眼镜的影响, 在这些复杂的环境下, 使用传统卷积神经网络检测效果较差。因

此, 本文采用多任务级联卷积神经网络 (Multi-task convolutional neural networks) 进行人脸检测及关键点定位, 其基本原理是通过级联的三级卷积神经网络 P-Net、R-Net 和 O-Net 组成漏斗状的检测器, 形成宽进严出的检测模式。

本文设计的 MTCNN 的网络结构模型如下:

P-Net: 第一个卷积层采用 3×3, 步长为 1 的卷积核, 池化层采用 2×2, 步长为 2 的最大值池化; 第二个卷积层采用 3×3, 步长为 1 大小的卷积核; 第三个卷积层采用了 3×3, 步长为 1 大小的过滤器; 最后, 用 1×1 大小的过滤器进行面部分类、边界框回归和特征点定位。其网络结构如图 2 所示。

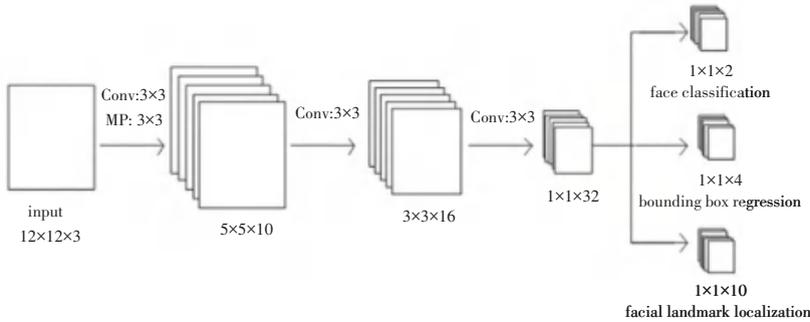


图 2 P-Net 层

Fig. 2 P-Net Layer

Q-Net: 第一个卷积层采用 3×3 的卷积核, 池化层采用 3×3 最大值池化; 第二个卷积层采用 3×3 大小, 池化层同样采用 3×3 的最大值池化; 最后一个

卷积层采用的卷积核大小为 2×2, 并且全连接层与最后一层卷积层连接。最后进行面部分类、边界框回归和关键点定位。其网络结构如图 3 所示。

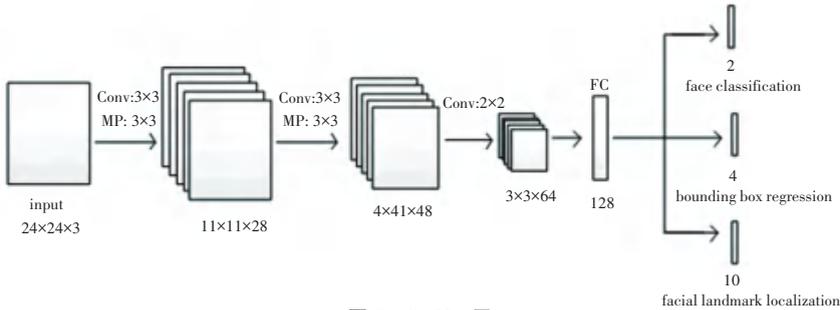


图 3 Q-Net 层

Fig. 3 Q-Net Layer

O-Net: 第一个卷积层采用 3×3 卷积核, 池化层采用 3×3 的最大值池化; 第二个卷积层采用 3×3 卷积核, 池化层采用 3×3 的最大值池化; 第三个卷积层采用 3×3 卷积核, 池化层采用 2×2 的最大值池化; 第四层卷积层采用 2×2 卷积核, 全连接层连接着最后一层卷积层。最后, 进行面部分类、边界框回归和关键点定位这 3 项工作。以上卷积的步长为 1, 池化的步长为 2, 其网络结构如图 4 所示。

为“valid”。而池化层的滑动步长为 2, 并且填充方式为“same”, 在所有的卷积层和全链接层之后为激励层。

对于 MTCNN 而言, 主要完成 3 个任务:

(1) 面部分类: 此任务通过交叉熵损失函数来实现, 如公式(1)所示。

$$L_i^{det} = -(y_i^{det} \log(p_i) + (1 - y_i^{det})(1 - \log(p_i))) \tag{1}$$

其中  $P_i$  表示人脸的概率,  $y_i$  为真实类别的标

上述 MTCNN 网络中, 卷积步长为 1, 填充方式

签,若为人脸,则值为 1,若不是则值为 0。

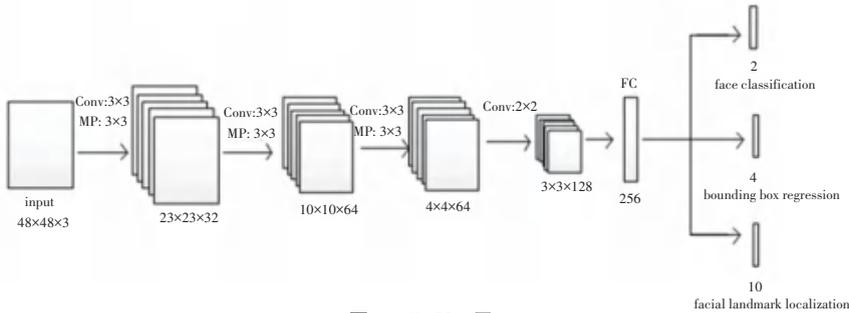


图 4 Q-Net 层

Fig. 4 Q-Net Layer

(2)边界框回归:对于候选窗  $\hat{y}_i^{box}$  而言,由于预测的真实值  $y_i^{box}$  之间的偏移,是回归问题,所以可以采用欧几里得损失,如公式(2)所示。

$$L_i^{box} = \|\hat{y}_i^{box} - y_i^{box}\|_2^2 \quad (2)$$

(3)面部关键点定位:由于人脸关键点检测类似于边界框回归任务,可通过最小化欧氏距离回归获得人脸关键点坐标,如公式(3)所示。

$$L_i^{landmark} = \|\hat{y}_i^{landmark} - y_i^{landmark}\|_2^2 \quad (3)$$

其中,  $\hat{y}_i^{landmark}$  和  $y_i^{landmark}$  分别为预测和真实的面部关键点。

### 1.2 目标区域提取

本文是通过 MTCNN 确定关键点,所以关键点较少,因此提出一种基于稀少特征点提取目标区域图像的方法,即通过仿射变换将不同位置、不同姿势以及不同大小的图像归一化到同一位置尺度。

### 1.3 人脸归一化

设  $(x_i, y_i)^T$  是脸上的第  $i$  个特征点坐标,  $(x'_i, y'_i)^T$  是通过人脸定位出的第  $i$  个特征点的坐标,矩阵如公式(4)所示:

$$\begin{bmatrix} x'_1 & y'_1 & 0 & 0 \\ x'_2 & y'_2 & 0 & 0 \\ x'_3 & y'_3 & 0 & 0 \\ x'_4 & y'_4 & 0 & 0 \\ x'_5 & y'_5 & 0 & 0 \end{bmatrix} = \begin{bmatrix} c & d \\ e & f \\ 0 & 1 \end{bmatrix} \begin{bmatrix} x_1 & y_1 & 0 & 0 \\ x_2 & y_2 & 0 & 0 \\ x_3 & y_3 & 0 & 0 \\ x_4 & y_4 & 0 & 0 \\ x_5 & y_5 & 0 & 0 \end{bmatrix} = M \begin{bmatrix} x_1 & y_1 & 0 & 0 \\ x_2 & y_2 & 0 & 0 \\ x_3 & y_3 & 0 & 0 \\ x_4 & y_4 & 0 & 0 \\ x_5 & y_5 & 0 & 0 \end{bmatrix} \quad (4)$$

将矩阵简写为  $Z = Nk$ ,  $Z$  是平均人脸的 5 点坐标值矩阵;  $N$  是通过人脸定位后的 5 点坐标;  $k$  为仿射变换矩阵。通过最小二乘法求出方程的解  $k = (N^T N)^{-1} N^T Z$ 。检测定位出的不同姿势、不同位置以及不同大小的图像归一化到同一位置尺度的图像。本文选取学生上课最有可能出现的姿势,分别是歪着头和手托着头听课,通过归一化操作,效果如图 5 所示。

经过归一化操作之后,需要对目标区域的图像进行快速提取。由于特征点较少,所以对目标区域

外接一个最小面积的矩形,来获取目标区域,效果如图 6 所示。

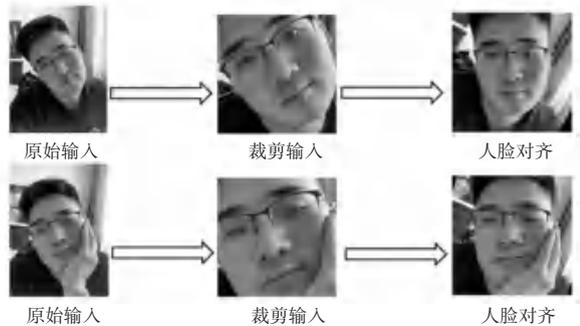


图 5 归一化操作

Fig. 5 Normalization operation

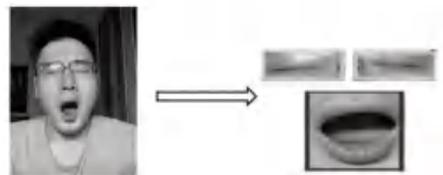


图 6 眼、口提取效果图

Fig. 6 Eye and mouth extraction effect

### 1.4 眼、口状态识别

在对眼睛、嘴部进行状态分类时,考虑到待分类的图片尺寸较小,所提取的特征图像排除了眉毛和鼻子的干扰,使得图像特征得到了简化。考虑到教室空间有限而且检测又要有较高的准确性和实时性要求,因此使用网络结构较为简单的网络进行模型训练,模型结构如下图 7 所示。

网络一共包含 3 个卷积层,第一层和第二层的卷积核为 32,第三层卷积层为 64,卷积核的大小均为 5。除了第一个池化层采用最大池化方式,其余两个均为平均池化,共计 3 个池化层,共有两个全连接层,神经元的个数分别为 64 和 2,识别效果如图 8 所示。

### 1.5 疲劳判断及多特征融合

学生在疲劳时会出现闭眼时间较长和打哈欠频

率较高的情况,为了更为准确的检测疲劳状态,本文

将眼部和嘴部的情况相结合来综合判断。

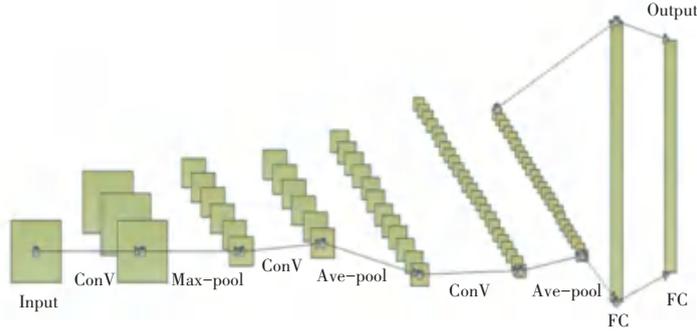


图 7 眼、口状态识别网络

Fig. 7 Eye and mouth state recognition network



图 8 眼、口状态识别效果

Fig. 8 Eye and mouth state recognition effect

根据人的面部特征判断疲劳状态时,卡内基梅隆研究所提出的 PERCLOS 方法是一种较为常用的检测方法,在所选取的时间范围内,人眼处于闭合状态所占时间的百分比,计算公式如公式(5)所示:

$$P = \frac{N_c}{N_t} \times 100\% \quad (5)$$

其中,  $P$  表示一定时间内,眼睛处于闭合状态的时间占据比;  $N_c$  表示某段时间内眼睛闭合状态的帧数量;  $N_t$  表示该时间段内的总帧数量。根据经验,当计算值  $\geq 0.4$  时,就表示学生处于疲劳状态,否则为正常状态。

当学生处于疲劳状态时另一个特征就是打哈欠,因此可以通过嘴部开合状态和张开的时间来判断学生是否疲劳。但是在检测过程中,首先要排除说话这一重要干扰因素。据研究,人打一次哈欠的时间大约在 3 s,而说话时嘴巴处于不停的开闭合状态,因此可以通过改进的 MAR 机理来进行哈欠判定。由于嘴部在运动时是处于整体运动的,而且嘴部区域共有 8 个特征点,如图 9 所示。所以使用嘴唇内部坐标进行计算,通过嘴唇纵横比的值  $MAR$  来判断是否疲劳,如公式(6)所示。

$$MAR = \frac{\|p_2 - p_8\| + \|p_3 - p_7\| + \|p_4 - p_6\|}{3 \|p_1 - p_5\|} \quad (6)$$

权威研究发现,当  $MAR$  值  $> 0.9$  时,可判定为学生处于疲劳状态。

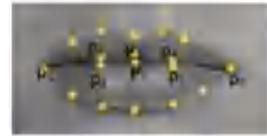


图 9 嘴唇区域坐标

Fig. 9 Lip area coordinates

为了更为准确地检测判定疲劳状态,本文将眼部疲劳和嘴部疲劳特征相融合,若  $P \geq 0.4$ , 或  $MAR \geq 0.9$  时学生处于疲劳状态。本文提出新的综合判断法,如公式(7)所示:

$$T = \alpha_1 P + \alpha_2 Q \quad (7)$$

其中,  $T$  为综合疲劳度;  $P$  为 PERCLOS 值;  $Q$  为  $MAR$  值;  $\alpha_1$  和  $\alpha_2$  为权重,并且  $\alpha_1 + \alpha_2 = 1$ 。经过多组实验可得到  $P$  的阈值为 0.5,即  $P \geq 0.5$ , 学生处于疲劳状态。 $\alpha_1$  为 0.8,  $\alpha_2$  为 0.2。

## 2 实验及结果分析

### 2.1 数据集准备

为了解决睁闭眼、张闭嘴样本不均衡的问题,本文通过对 ZJU 眨眼视频数据集采集包括睁闭眼两种状态共计约 9 000 张图片,其中 1 000 张图片 ( $\pm$  样本各 500 张)用于后续的模型测试,其余的样本则用于训练。

嘴部数据集则是通过 YawDD 疲劳驾驶视频数据集和其它视频数据集所采集的嘴部样本。为了提升嘴部样本数量,本文对样本进行垂直旋转和水平翻转得到样本数量共计 2 000 张,其中 200 张 ( $\pm$  样本各 100 张)用于模型的测试,其余的样本则用于模型训练,部分样本如图 10 所示。

### 2.2 实验环境

Windows10 操作系统, Intel Core i5-8300H, 8GB

内存,python3.5 编程语言, Tensorflow1.7 框架,显卡型号 GeForce GTX1060。



图 10 部分数据样本

Fig. 10 Part of sample data

### 2.3 实验结果

为了进一步验证本文所提出方法的优越性,在相同实验条件下,分别与 MTCNN 结合随机森林<sup>[9]</sup>的方法以及 Adaboost 结合 CLM<sup>[10]</sup>的方法进行对比,结果见表 1。分别使用 3 种方法在 100 张人脸图像上进行测试。

表 1 眼、嘴部状态分类测试结果

Tab. 1 Eye and mouth state classification test results

方法	平均准确率/%	时间开销/s
MTCNN+HOG+random ferns <sup>[9]</sup>	94.1	14.20
MTCNN+Multihog+SVM <sup>[10]</sup>	94.4	48.67
本文方法	96.2	8.101

由实验结果可知,本文提出的方法在准确率上具备了一定的优越性,并且时间开销上能满足实时性要求。本文通过简化图像特征,使得利用简单的网络实现眼、嘴部状态分类成为可能。

### 2.4 运行效果展示

为了清晰的展示运行效果,依据 PERCLOS 疲劳判断标准,本文用实时视频进行测试,测试效果如图 11 所示。



图 11 疲劳检测系统示意

Fig. 11 Schematic diagram of fatigue detection system

本文提出疲劳检测方式的实时性也极高,能够及时的帮助老师发现疲劳学生,加以提醒,提高课堂教学效果。

### 3 结束语

本文提出了一种基于卷积神经网络的多特征融合学生课堂疲劳检测算法。在 MTCNN 的基础上通过归一化操作完成眼、嘴部区域的精确定位,并且将眼嘴部疲劳状态相结合实现多特征融合的疲劳检测方法。对比采用单一的特征进行检测,识别率有了明显的提升,说明本文所提方法能够提高系统判断的准确性。但是教室是个很复杂的环境,如何更加全面的检测到每个学生的状态,从而实现真正意义上的智慧课堂与智慧教育还需要进一步研究。

### 参考文献

- [1] 左国才,苏秀芝,王海东,等. 基于 CNN 深度学习模型的大学生课堂行为检测研究[J]. 智能计算机与应用,2020,10(2):158-160.
- [2] 曾鸽. 基于深度学习的驾驶员疲劳状态检测研究[D]. 长沙:湖南大学,2018.
- [3] 靳明浩. 基于人脸识别的驾驶人员疲劳检测系统设计[J]. 智能计算机与应用,2020,10(2):172-173,179.
- [4] 陈建伟,贺继林. 一种改进的多通道特征视觉跟踪算法[J]. 传感器与微系统,2020,39(3):120-123.
- [5] 江伟坚,郭躬德,赖智铭. 基于新 Haar-like 特征的 Adaboost 人脸检测算法[J]. 山东大学学报(工学版),2014,44(2):43-48.
- [6] 褚晶辉,张姗,吕卫. 基于卷积神经网络的驾驶行为分析算法研究[J]. 激光与光电子学进展,2019(12):1-18.
- [7] 耿磊,袁菲,肖志涛,等. 基于面部行为分析的驾驶员疲劳检测方法[J]. 计算机工程,2018,44(1):274-279.
- [8] ZHANG Kaipeng, ZHANG Zhanpeng, Li Zhifeng, et al. Joint face detection and alignment using multitask cascaded convolutional networks [J]. IEEE Signal Processing Letters, 2016, 23(10): 1499-1503.
- [9] DONG Yanchao, ZHANG Yan, YUE Jiguang, et al. Comparison of random forest, random ferns and support vector machine for eye state classification [J]. Multimedia Tools and Applications, 2016, 75(19): 11763-11783.
- [10] 刘朝涛,张雪佼. 基于图像处理的疲劳驾驶预警研究[J]. 电子技术应用, 2019, 45(8): 104-108.