

# 基于面部表情识别的在线教学质量评价

周林好

## 摘要

本文复现了 Changbo Hou 等人<sup>[1]</sup>设计的基于表情识别的学生课堂集中度分析系统。在该系统中,使用 OpenCV 调用摄像头实时采集学生的学习状态,并使用 MTCNN 算法检测视频的人脸,以识别学生面部图像的位置。最后,利用 VGG16 网络对所得人脸图像进行实时表情识别,获取学生在课堂上的情绪。本实验成功复现了原论文设计的系统,能够实时识别学生的学习状态,并评估学生的学习质量。但是实验结果与原论文存在差距,未能成功复现添加 ECA 的 VGG16 网络,并且 VGG16 网络的训练结果与原论文差距大。

**关键词:** 在线教育; 人脸检测; 表情识别; 学习质量

## 1 引言

随着信息技术的飞速发展,各行各业的科技力量不断增强,教育领域也开始逐步引进高新技术。受疫情影响,线上教学在全国范围内落地,形成了线上线下教学“双融合”的教育模式。但是,在线教学的缺点也非常明显;也就是说,教师无法实时了解学生的学习状态。因此,这就需要一种能够实时采集学生学习状态并及时反馈给教师的技术,让教师通过分析学生的学习状态数据来定制个性化教学,激发学生的学习和思考热情,回归“因材施教”的教育本质。

当今一直采用的教学评价大多是课后调查的方式,不仅繁琐,而且收集到的信息也不全面准确。此外,人们缺乏一个完整和统一的评估方案来衡量学校和教师的教学水平。近年来,随着人工智能的发展,人们开始将人脸检测和表情分析应用于课堂教学评估。但是,由于精度和准确性等原因,它们尚未在课堂上广泛使用。因此,如何提高表达分析的准确性是一个亟待解决的问题。

Changbo Hou 等人<sup>[1]</sup>设计了基于表情识别的学生课堂集中度分析系统,该系统能够准确地识别学生在课堂上的情绪,进行教学效果评价,在智慧课堂、远程学习等智慧教育领域具有一定的应用价值。因此,本人对此系统进行了复现。

## 2 相关工作

### 2.1 数据集

本文使用复现论文提供的 FER2013 数据集,这是在 Kaggle 中公开的人脸数据集。FER2013 包含 35886 张面部表情图片,其中 28709 张在训练集中,3589 张在公共验证集中,3589 张在私有验证集中。共有 7 种表达方式,即愤怒,厌恶,恐惧,快乐,悲伤,惊讶和中性。

## 3 本文方法

### 3.1 本文方法概述

系统由视频采集模块、人脸检测模块、表情分析模块、教学评价模块四部分组成。系统组成如图 1 所示。

视频采集模块使用 OpenCV 调用摄像头，摄像头负责实时采集学生的课堂视频，对采集到的视频进行测试和预处理。人脸检测模块运用 MTCNN 算法标定图像人脸位置，提取人脸特征区域。表情分析模块运用 VGG16 网络对采集到的学生面部图像进行表情分析，获取学生的课堂情绪。教学评价模块运用情感得分公式获取学生的课堂情况和教师的教学情况。

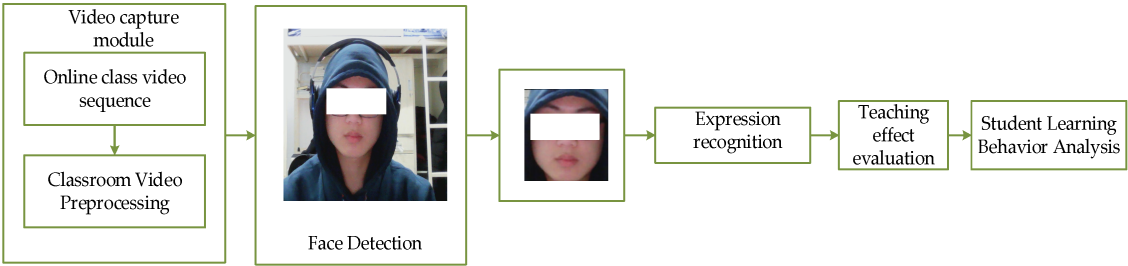


图 1: 系统组成图

3.2 视频采集模块

视频采集模块利用 OpenCV 调用摄像头实现。具体方法为：引用 cv2 包，主要使用 cv2 中的 VideoCapture() 和 imshow() 方法。

3.3 人脸检测模块

人脸检测过程使用 MTCNN 算法实现。MTCNN 算法将人脸区域检测与人脸关键点检测放在一起。总体可分为 P-Net、R-Net 和 O-Net 三层网络结构。MTCNN 算法实现流程如图 2所示。

首先构建图像金字塔，即将图像进行不同尺度的变换，以适应不同大小的人脸的进行检测。构建方式是通过不同的缩放系数 factor 对图片进行缩放，每次缩小为原来的 factor 大小。其次，在 P-Net 中进行初步特征提取与边框标定。第三，在 R-Net 中滤除大量效果比较差的候选框。最后，通过 O-Net，O-Net 的效果与 R-Net 的区别在于这一层结构会通过更多的监督来识别面部的区域，而且会对人的面部特征点进行回归，最终输出五个人脸面部特征点。

人脸检测实现流程：利用视频采集模块调用摄像头；创建 MTCNN 模型检测人脸；根据模型返回值 rectangles 画框；使框实时显示在实时视频中。

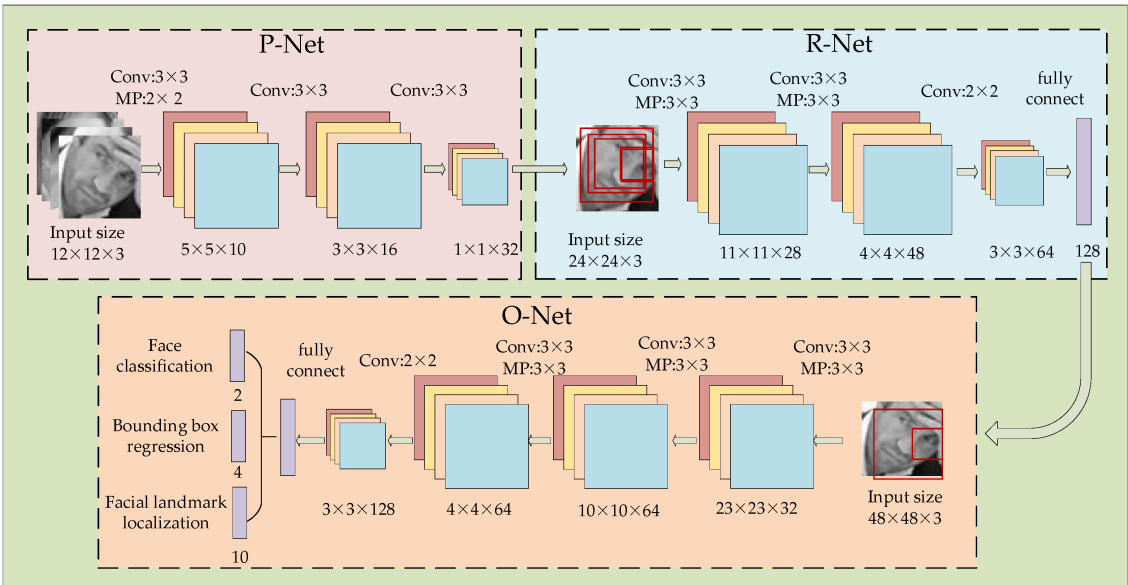


图 2: MTCNN 网络结构图

### 3.4 表情分析模块

表情分析模块使用预训练好的 VGG16 模型，该模型的网络结构如图 3 所示。VGG16 共有 16 层，包括 13 个卷积层和 3 个全连接层。首先输入尺寸为 224x224x3 的图像；经过两次 64 个卷积核卷积后，采用 pooling；经过两次 128 个卷积核卷积后，采用 pooling；经过三次 256 个卷积核卷积后，采用 pooling；再重复两次三个 512 个卷积核卷积后，再 pooling；最后经过三次全连接。

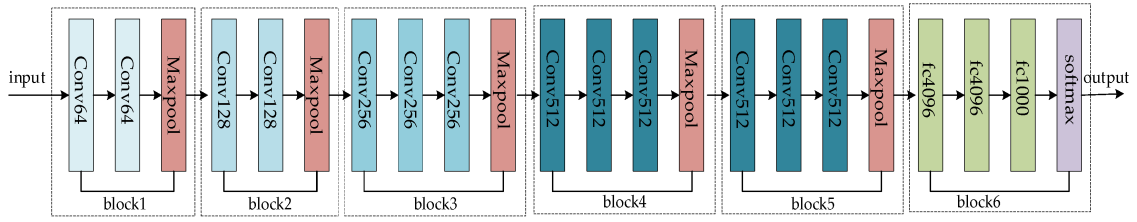


图 3: VGG16 网络结构图

### 3.5 教学评价模块

根据 FER2013 数据集，面部表情分为七种类型：即愤怒、厌恶、恐惧、快乐、悲伤、惊讶和中性。其中，愤怒、厌恶、恐惧和悲伤是消极的学习行为，而快乐、惊讶和中立是积极和专注的学习行为。将面部表情识别模块的识别结果发送到教学评价模块后，根据情绪的分类获取学生的学习情况。

7 种情绪的置信度可以通过面部表情识别获得。置信水平反映了每个状态的可能性，可用作评分标准。把恐惧和悲伤作为学生对课堂不满意的情绪，将他们的权重设置为-1; 将厌恶和愤怒视为更不满意的情绪，并将其权重设置为-2; 以中立作为学生对课堂的满意度，将其权重定为 1; 把快乐作为对课堂比较满意的学生的感情，把权重定为 2; 把惊喜作为对课堂情况最满意的学生的感情，并将其权重设置为 3。加权求和可以得到样本的情感得分，如公式 (1)，其中， $p_i (i = 0, 1, \dots, 6)$  表示 7 中情绪的置信度。所有分数都归一化为介于  $[0, 1]$  之间以获得最终分数。分数归一化的计算方法如公式 (2) 所示。

$$SCORE = p_0 + 2p_1 + 3p_2 - (p_3 + p_4 + 2p_5 + 2p_6) \quad (1)$$

$$norm(score) = \frac{score_{min}(X)}{max(X) - min(X)}, score \in X \quad (2)$$

分数越高，学生此刻的专注度就越高，分数是相对于全班的。当分数高于 0.5 分时，可以认为课堂情绪高于全班平均水平。对一定时间段内获得的评分结果进行统计平均，得到目标学生在一定时间段内的学习情况。最后，取全班所有学生分数的平均值，可以得出全班学生的整体情绪分数，从而评价课堂教学的质量。

## 4 复现细节

### 4.1 与已有开源代码对比

无开源代码

### 4.2 实验环境搭建

本实验使用语言为 python，具体版本如下：

python: 3.7

pytorch: 1.12.0

cuda: 11.6

### 4.3 界面分析与使用说明

本实验直接运行 main.py 即可实时对学生学习状态进行评估, 如图 4所示。红框圈出人脸, 白色文字显示识别的学生情绪以及学习状态得分。

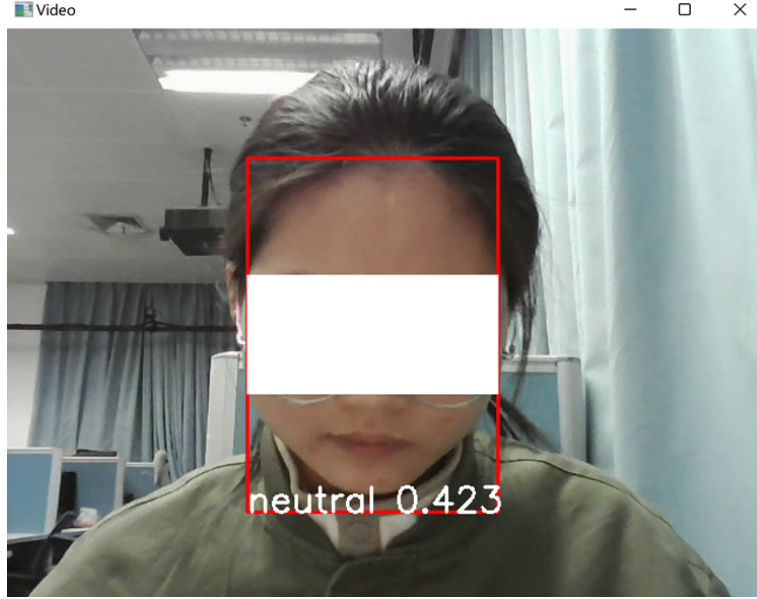


图 4: 系统界面图

### 4.4 创新点

无

## 5 实验结果分析

在 FER2013 数据集上训练时, 网络超参数如下: 实验迭代 300 次。训练集的批量大小设置为 128, 验证集的批量大小设置为 128, 初始学习率设置为 0.0001。通过公共验证集优化网络的权重, 最后在私有验证集上测试模型以评估模型性能。公式 (3) 中显示的随机梯度下降 (SGD) 用于优化, 公式 (4) 中显示的分类交叉熵损失函数用于模型训练。

$$g_t = \nabla_{\theta} f(\theta; x_i; y_i) + \nabla_{\theta} \phi(\theta) \quad (3)$$

$$Loss = - \sum_{i=1}^n \widehat{y}_{i1} \log y_{i1} + \widehat{y}_{i2} \log y_{i2} + \dots + \widehat{y}_{im} \log y_{im} \quad (4)$$

图 5显示了使用 VGG16 方法的精度曲线, 这两个函数会随着迭代次数的增加而逐渐降低, 最后趋于稳定。使用 VGG16 进行模型训练时, 训练集的准确率趋于 0.52, 验证集的准确率趋于 0.48, 与原论文训练集的准确率趋于 0.94, 验证集的准确率趋于 0.64 的结果相差很大。

图 4为系统运行结果, 能够实时识别学生的学习状态, 并对学生的学习质量进行评估。

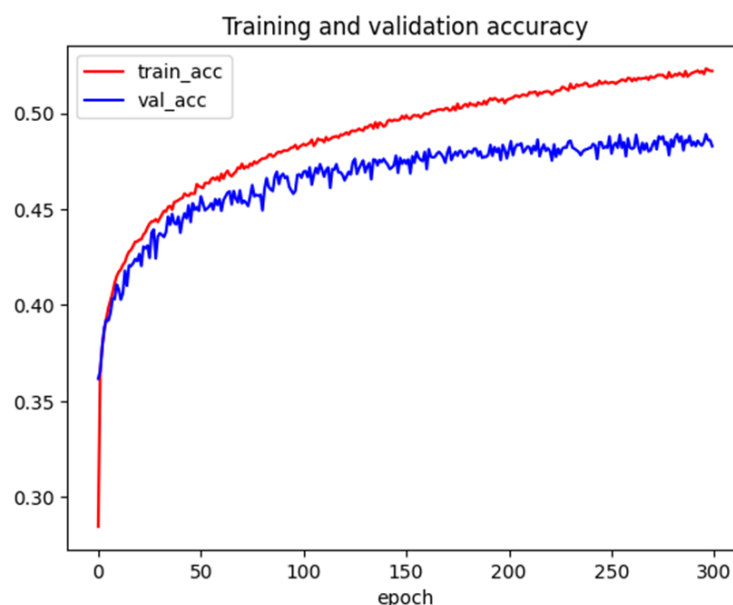


图 5: VGG16 训练累积曲线图

## 6 总结与展望

针对在线教学过程中教师无法与学生实时面对面交流以了解学生学习状态的问题，本实验利用 VGG16 网络对在线课堂学生进行实时面部表情识别，获取学生的学习状态并及时反馈给教师。虽然成功复现了原论文的系统，但存在着诸多不足。首先是在表情识别模块中，未能成功复现添加 ECA 的 VGG16 网络，使得系统不够完整。其次，使用 VGG16 进行模型训练的结果差强人意，与原论文结果存在很大差距。第三，未能根据原论文进行改进。

在后续研究中，可以创建学生表达数据集以提高模型的准确性。此外，通过手臂动作和抬头率可以反映学生的学习状态，可以作为后续研究中学生注意力评价标准的一部分，使课堂教学评价更加客观全面。

## 参考文献

- [1] HOU C, AI J, LIN Y, et al. Evaluation of Online Teaching Quality Based on Facial Expression Recognition[J/OL]. Future Internet, 2022, 14(6). <https://www.mdpi.com/1999-5903/14/6/177>. DOI: 10.3390/fi14060177.