

## Classroom Analysis System based on Expression Recognition

Mingjie WU    Zhengde BAO

School of Computer and Software, Jincheng College, Sichuan University, Chengdu, Sichuan, 611731

### Abstract

New methods are adding to the management of college students with technological innovation of Internet, artificial intelligence and so on. Classroom learning situation analysis, as an essential part of detecting students' learning situation and evaluating the quality of classroom teaching, has introduced expression recognition technology. Giving school a better understanding of classroom management in order to facilitate the targeted adjustment of the way of education and provide the education which follows the principle of teaching students in accordance with their aptitude. In this system, a convolution neural network (CNN), is built by using Keras framework to train the facial expression data in fer2013 database to obtain the expression recognition model and detect the face by OpenCV. Put the detected face into the expression recognition model and according to the picture content, the model gives the probability of each expression, and selects the expression with the highest probability as the result expression. Finally, through the analysis of the relationship between the number of each expression and the total number of people to analysis the situation of classroom learning.

### Key Words

Convolutional Neural Network (CNN), Expression Recognition, Flask OpenCV

DOI:10.18686/jsjxt.v1i2.661

## 基于表情识别的课堂学情分析系统

吴明杰  鲍正德

四川大学锦城学院计算机与软件学院, 四川成都, 611731

### 摘要

高校学生管理顺应互联网、人工智能等技术变革也注入了新的方法, 课堂学情分析作为检测学生学习情况、评估课堂授课质量的重要组成部分, 引入了表情识别技术, 更直观的让学校了解课堂管理情况以便于针对性的调整教育方式, 更好的做到因材施教。本系统利用 Keras 框架搭建了一个卷积神经网络(CNN), 经由该网络对 fer2013 数据库中的人脸表情数据进行训练获得表情识别模型, 并使用 OpenCV 检测人脸, 将检测到的人脸放入表情识别模型中, 模型根据图片内容给出每种表情的概率, 选择概率最大的表情为结果表情, 最后, 通过分析各个表情的人数与总人数的关系得出课堂学情情况。

### 关键词

卷积神经网络 表情识别 Flask OpenCV

### 1. 引言

人脸表情识别以其准确性及实用性在心理学、智能监控、虚拟现实等领域存在着极大的潜在应用价值, 将表情识别融入学情分析, 不仅为学校提供了学生学习情况数据便于督促老师改进教学方式, 同时学生也可了解自身学习情况并做出学习计划的调整。本系统通过研究卷积神经网络(CNN)用于人脸面部表情识别的可行

性和准确性, 将面部表情识别系统投入到课堂辅助教学中。利用视频或摄像头, 结合 OpenCV 计算机视觉库, 对学生进行面部表情的捕捉与分析, 实时分析出学生上课时的倾听、疑惑、抗拒和不屑等各种心理状态, 从而为分析课堂中学生的关注度、参与度和活跃度提供了依据。

## 2. 主要技术

### 2.1 Python 语言

Python 语言目前被广泛的运用在 Web、人工智能、网络爬虫、机器学习等方面,是时下最流行的计算机编程语言之一。

Python 语言优点:

(1) 该语言属于解释型语言,所以代码在运行时不用进行编译。

(2) 代码开源,面向对象。

(3) 语法简洁,易上手。

(4) 拥有大量的标准库,第三方库。

缺点:

(1) 相较于 C/C++ 来说,Python 代码的运行速度较慢。

(2) 采用缩进来代替 {},这种方式在实际使用中,要严格对齐,给开发人员带来不便。

### 2.2 Flask 框架

Flask 是现在 Python 在小型 Web 应用开发上使用得最多的微型框架之一。它的 WSGI 工具箱使用的是 Werkzeug,模板引擎是 Jinja2,[1]使用 BSD 授权。<sup>[2]</sup>Flask 本身自带功能相对较少,但是可以用 extension 增加其他功能。<sup>[3]</sup>

Flask 的优点:

学习简单,入门门槛低。

语法简洁,易于理解。

兼容性好,对多种数据库、模板兼容。

极好的扩展性,可以通过 Flask-extension 增加项目所需要的功能模块。

轻巧。

Flask 的缺点:

对于大型项目来说,需要自己设计代码的架构,对于经验教少的开发人员来说,容易造成代码混乱。

### 2.3 前后端通信技术

#### 2.3.1 WebSocket 协议

WebSocket 协议实现了浏览器和服务器之间异步、全双工通信(可以互相发送消息),于 2011 年被 IETF 定为标准 RFC 6455,并由 RFC7936 补充规范。<sup>[4]</sup>它是一个长连接协议。即在通信时浏览器端与服务器端一直

保持连接,这种方式可在前后端需要高频率通信的情况下,有效减少因为前后端反复建立连接而造成的计算机资源浪费,从而使系统的整体性能得到一定的提升。

#### 2.3.2 Socket.io

Socket.io 将 WebScket、Ajax 等一些常见通信方法整个封装成了一个接口,其特点是:对所有浏览器兼容、自动选择通信方式、开发难度低。使用时,它会根据浏览器对协议的支持与否自动选择一个最佳的通信方式。如果客户端浏览器支持 WebSocket 协议,通信方式便采用 WebSocket,如果不支持,它会选择其他的通信方式来模拟 WebSocket 通信,这其中用的比较多的替代方式是 Ajax 长轮询。

### 2.4 OpenCV

OpenCV 是一个开源的计算机视觉库<sup>[5]</sup>,它对 python 也供给了接口,因而 python 可以非常便利的利用 OpenCV 进行应用开发。OpenCV 在人脸检测中提供了包括了人脸,眼睛,嘴巴,鼻子,耳朵等一系列特征分类器,利用这些分类器能够快速高效的在图片中找到需要的特征。开发者还可以利用该视觉库训练自己的物体分类器。训练分类器,需准备正样本数据与负样本数据,正样本即只包含了找寻目标的图片,大小必须一致(OpenCV 推荐大小 20x20 像素),负样本即不包含找寻目标的图片,大小可以不一致但必须不小于正样本数据,内容最好是目标物体附近常出现的東西,可根据实际应用的场合而定。

## 3. 训练数据

### 3.1 Fer2013

Fer2013 人脸表情数据库,该数据库以 csv 表格的形式搜集了 35886 张灰度表情图片的数据,图片大小为 48x48 像素,其中囊括了训练集(Training)28,709 张,公共测试集(Public Test)3589 张、私有测试集(Private Test)3589 张。Fer2013 数据库中把人脸的表情分成了 7 种,分别是 0-angry(气愤)、1-disgust(厌恶)、2-fear(害怕)、3-happy(开心)、4-sad(悲伤)、5-surprised(惊喜)、6-neutral(中立)。该数据集还包含了部分卡通人物的表情数据,它们会对模型训练造成干扰,影响准确率。本实验为了得到更好的训练精度,在数据处

理阶段手动清除了数据集中的无效图片数据。

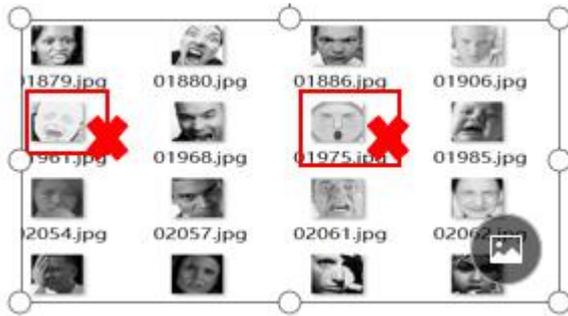


图 1 fer2013 人脸数据集中的卡通

## 4.前期准备

### 4.1 训练模型

#### 4.1.1 搭建神经网络

本系统的 CNN 网络由 4 个卷积层，3 个池化层和 2 个全连接层组成，结构如下：

Height	Width	Depth	filter Height	filter Width
48	48	1	1	1
48	48	32	5	5
24	24	32	3	3
12	12	64	5	5
6	6	64	0	0

Vector Length
2048
1024
7

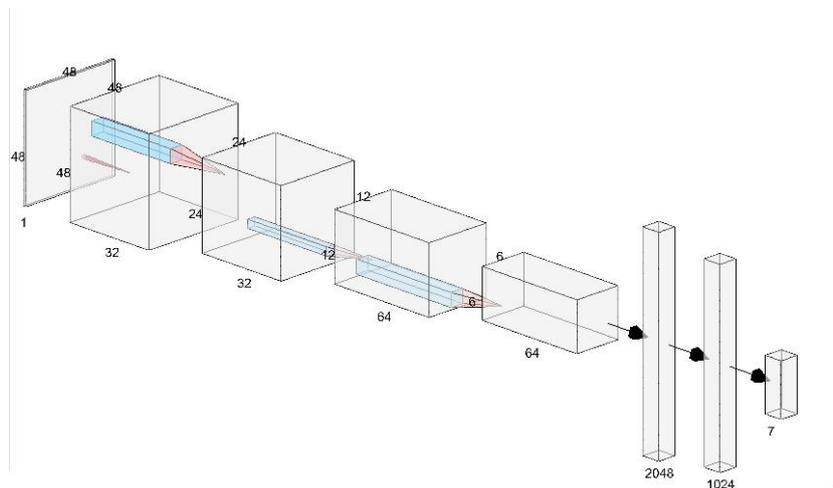


图 2 CNN 结构图

#### 4.1.2 训练环境

硬件环境: CPU:Core i5-7300U@2.6GHz  
内存:8GB, 硬盘 128GB, 操作系统 Windows10 Pro。  
软件环境: Anaconda3、Python3.6、keras

#### 4.1.3 训练结果

训练时长: 1 天 23 小时, 准确率: 74%。

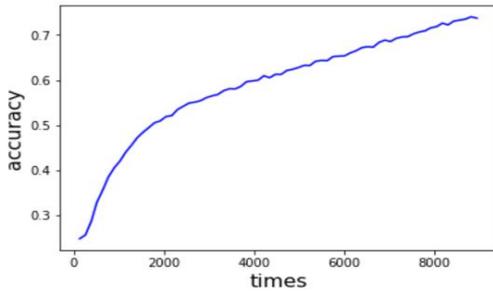


图3 在 fer2013 数据集上的测试正确率变化图

## 4.2 结果分析

### 4.2.1 训练时长

由于本实验使用 CPU 来进行模型训练，所以训练速度偏慢，花费时间较长。提高训练速度方法：

(1) 优化代码。让代码变得更加简洁高效，这是在不改变硬件条件的情况下最好方式，但这个提升的速度有限。

(2) 使用 GPU 训练。GPU 并行运算的特点使它的计算速度要远远快于 CPU。训练模型，就是通过大量的数据去不断调整参数的过程，这一过程的长短就取决于担任计算任务的 CPU 或 GPU 的计算速度。由于 GPU 的速率更快，以是利用 GPU 可以比 CPU 更快的完成训练。根据以往的数据显示，GPU 的速度一般会比 CPU 的速度快几十倍甚至上百倍，这取决于参与比较的 GPU 和 CPU 的性能。

(3) 使用性能更强的 CPU/GPU。

### 4.2.2 准确率

模型的准确率达到到了 74%，在同用 fer2013 数据集训练的模型中，处于中间位置，仍有提升的空间，但也达到了能够使用的标准。提高准确率方法：

(1) 增加训练的图片数量。通常情况下样本越多，参数就越精确，从而准确率也就更高了。

(2) 改进网络结构。网络结构会直接影响到模型准确率，比如，增加一个或减少一个卷积层会对模型的准确率造成较大的影响。

(3) 增加训练次数。如果发现准确率还有上升趋势，但训练已经结束了，就可以适当的增加一些训练次数，也可以提高准确率。

## 5.系统部分功能实现

### 5.1 系统工作流程：

- (1) 前端界面打开摄像头或视频显示课堂景象。
- (2) 从视频中截取图片。
- (3) 将图片传递到后端。
- (4) 后端接收图片。
- (5) 检测图片中所有人脸，获取人脸图片。
- (6) 将人脸图片依次送入训练好的模型中进行表情识别。
- (7) 根据识别结果进行学情分析。
- (8) 将学情分析结果返回前端。
- (9) 前端页面展示结果。

### 5.2 前端界面

前端 web 界面大致分为左右两部分，左边部分由监控区和控制区组成，监控区显示视频或摄像头中的课堂景象，控制区设置了一个按钮，该按钮用来控制视频与摄像头之间的切换。右边部分则实时展示当前的课堂质量，表情分布等信息。

### 5.3 截图功能

本系统利用 html5 的画布 (canvas) 元素实现视频截图，该元素可以通过前端代码在 Web 页面上绘制图像。截图实现：在 html 文件中添加 canvas 标签，通过 JavaScript 代码将视频 (video) 标签中的图像帧画在画布上，画上的图像便是视频的截图。

### 5.4 图片编码

本系统前端向后台传递图片的信息格式采用 Base64 编码 (网络上最常见的用于传输 8Bit 字节代码的编码方式之一<sup>[6]</sup>) 后的字符序列 (字符串)。前端实现图片转换成 Base64 字符序列的方式：html 文件中添加 canvas 标签，把要编码的图片画在 canvas 上，再调用 canvas 的 toDataURL() 方法，该方法会对 canvas 上的图像进行 Base64 编码，并将编码后的字符序列返回。使用该编码的字符序列要注意的是：一、数据经过 Base64 编码后的字符序列如果直接放在 URL 里传输，字符序列中的“/”，“+”，“=”会分别转换成“%2F”，“%2B”，“%3D”。二、待解码的字符序列需满足长度能被 3 整除，因为该编码以三个字符为一组，如果不能被整除，则需要在该字符序列的末尾追加字符“=”来使序列长度为 3 的倍数。

### 5.5 人脸检测

输入训练模型的数据是 48X48 像素的单通道灰度图, 每张图只包含一张人脸。但在实验中, 后端接收到的图片大概率会是有多张人脸和复杂环境的高分辨率图片, 这样的图片不能直接送入模型进行表情识别。需要先检测出图片中的人脸, 再把检测到的人脸缩放到 48x48 像素大小后再送入模型中进行表情识别。

OpenCV 视觉库提供了方法可以准确地找到图片的人脸。本论文中使用 OpenCV 的 Haar 特征级联分类器来检测人脸, 该分类器会返回图片中的人脸位置信息, 得到人脸位置信息后便可将图片中所有人脸提取出。

### 5.6 人脸表情检测

将提取到的人脸图片处理成模型的输入格式

后, 放入模型中检测, 模型给出图片人脸属于各个表情的概率, 概率最高的表情作为检测的结果。本系统为让结果更加准确, 还对图片进行了水平翻转后再次放入模型中检测, 将两次的结果相加, 选择最大值的表情为最终识别的结果。

### 5.7 学情判定规则

- (1) 质量好: 超过 50%的学生是高兴的表情。
- (2) 质量一般: 超过 50%的学生是正常的表情。
- (3) 质量差: 不属于上面任何一种就判定为差。

### 6.最终效果

本系统基本实现了对小班教学的学情情况进行分析的功能, 并实时的在前端 web 界面显示当前时刻各个表情的人数以及课堂的质量。



图 4 系统最终效果图

### 7.结束语

本文通过 Keras 框架搭建的卷积神经网络对 Fer2013 数据库训练得到人脸表情识别模型, 再结合 OpenCV 实现了一个简单的课堂实时学情分析系统。该系统通过视频或摄像头获取当前课堂的实时景象, 检测课堂上各个学生的表情, 再根据各种表情人数与总人数的比例, 给出学情分析的结果, 即课堂氛围的实时情况。授课教师可通过本系统及时了解课堂氛围的变化, 及时调整授课方式、内容等, 同时学生也可了解自身学习情况并做出学习计划的调整, 以达到提高整体教学质量的目的。

### 参考文献

[1]张虹. 智能家居更新事务管理系统的研究与开发[D]. 东南大学,2017  
[2]:蔡敏.Python 语言的 Web 开发应用分析[J].无线互联科技,2019,16(04):27-28.

[3]:潘海珍,明小波,吴云.基于 Flask-RESTful 的定制化问卷系统设计与实现[J].自动化技术与应用,2019,38(02):37-41.  
[4]马志彬. 面向移动医疗的心电数据实时共享系统设计与实现[D].浙江大学,2018.  
[5]:李雪煜,杨松,秦炜茗,孙淑婷,宫明,曲洲.基于树梅派和 OpenCV 的车牌识别算法研究[J].电脑编程技巧与维护,2019(04):146-148.  
[6]张晋,李太君,冯思玲.局域网 Email 信息监控与过滤[J].电脑与信息技术,2009,17(02):67-70.

### 作者简介

第一作者: 吴明杰 (1998-), 男, 汉, 四川省资阳市, 本科, 四川大学锦城学院, 研究方向: J2EE、Python  
第二作者 (通讯作者): 鲍正德 (1989-), 男, 汉, 黑龙江哈尔滨, 研究生, 四川大学锦城学院, 研究方向: 电子商务。