

# 基于 FER2013 数据集的人脸表情识别

姜洋 李丹

四川大学锦城学院计算机与软件学院 四川 成都 610000

**【摘要】**面部情绪是表明心中想法的一种方法，同样是了解他人情绪的有效路径。现如今科技发展迅速，计算机也能够通过卷积神经网络、深度学习等方式对人脸面部表情进行识别，并且对结果进行分类。在整个实验过程中，我们选择了 FER2013 数据来作为模型的训练集，最终能够得到 62% 左右的准确度，并且还将其与 SFEW 数据集做了对比。表情识别出现的地方在将来会与之增加，它在教学监督上的应用是我们在此所探讨的，其主要作用可以用来进行监考、考勤、查看上课状况等等。

**【关键词】**人脸表情识别；FER2013 数据集；卷积神经网络；协方差；教学监督

## 1 表情识别的意义

### 1.1 意义

如今人机交互系统在日常中随处可见，也正是这个原因，人机交互系统的智能性和可靠性愈发重要。而在人机交互系统中使用面部表情识别则可大大增加其智能性，通过识别用户的面部表情，判别其情绪变化，能够使得产品与用户的交流更加地自然。当然，不仅仅只有人机交互系统才能用到人脸表情识别，因为科技的进步，人脸表情识别也趋之优化，它的应用范围更是在不停的扩充，比如安全驾驶、医疗疼痛检测、教学监督等等。

### 1.2 表情识别的现状

面部表情识别技术发展多年，到现在它的主要核心环节主要是人脸检测、提取面部特征、表情分类三个部分。对于人脸检测这个部分，在现在已经不算一个难题，目前主要是如何提取面部特征这一部分是研究重点。关于表情分类，现在把表情分为基本表情和复杂表情，基本表情分为 7 种：开心、难过、惊讶、害怕、发怒、讨厌、平静。基本表情无法将日常生活中人类情感的复杂呈现出来，而面部运动编码系统 FACS 则可以用来表示更复杂的人脸表情。

因为在日常生活中，我们最常表现的是基本表情，所以目前关于复杂表情的研究很少，许多研究都是专门关于 7 种基本表情的。除此以外，数据集的获取较为困难，因为进行表情识别的研究需要大量的面部表情数据用于训练，而目前的数据集有限。SFEW、CK+、FER2013 等数据集是我们经常使用的一些数据集。

人脸表情识别技术最先开始于国外，在 21 世纪得到快速发展，并将人脸表情识别技术分为关于图片或者照片的静态识别以及关于视频、录像的动态识别。一开

始人们是用传统方式来提取人脸特征，即 LBP 特征、主成分分析 PCA 以及 Gabor 小波等。而之后随着深度学习的出现，利用深度学习来提取特征便成为了最常用的方法。我们国家进行人脸表情分析的研究最早是在 1997 年，由哈尔滨工业大学将此技术引入我国，之后，经过不断的研究发展，我们已经在人脸识别上已经达到了不错的高度。

## 2 技术核心

### 2.1 卷积神经网络

卷积神经网络 (Convolutional Neural Network, CNN) 是一种前馈神经网络，它的组成如图 1 所示，其中除了输入输出层以外，其他层在 CNN 中的地位类似于心脏关于人的重要性。卷积神经网络除了可以处理此类文件以外，还能够处理音频、文本等这样的信息，只要是数据能够变成图像格式，那么它就能够被卷积神经网络处理。

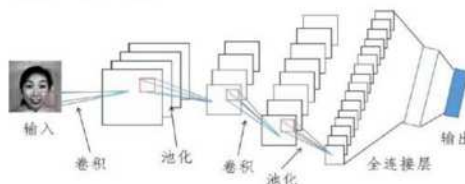


图 1 卷积神经网络基本结构

#### 2.1.1 卷积层

使用图像的某一部分去进行匹配从而避免使用整体被称为卷积。卷积运算就是比较两个图像中的各个局部，这些局部就是卷积核，也被成为特征。而卷积核一开始

是需要人为设定的，在经过大量的运算后会得到一个最优的卷积核。运算过程就如图 2<sup>1</sup> 所示，左侧为输入矩阵，中间是卷积核，它会以一定步长在左侧上进行点积运算，得到右侧的输出矩阵，这就是卷层最基础的运算。

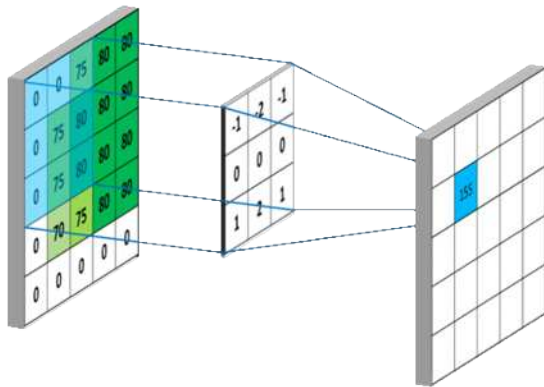


图 2 卷积运算过程

### 2.1.2 池化层

池化就是根据卷积核将卷积层的输出缩小，从而达到减少计算量的目的，虽然会牺牲掉一部分信息，但只是可以接受的。池化一般包含最大池化和平均池化，根据图 3 所示，最大池化就是将指定范围也就是 2\*2 的方框内最大的数提取出来，也就是 22，而平均池化就是将那四个数求平均值后再提取出来，最后结果为 9，一般使用的是最大池化。

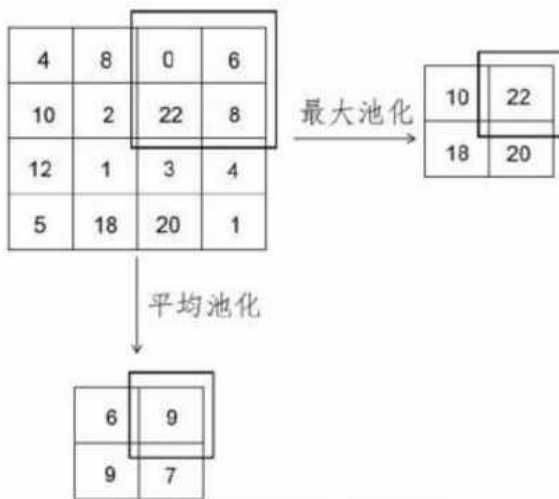


图 3 最大池化和平均池化

### 2.1.3 全连接层

全连接层与分类器相似，在卷积神经网络中它能够根据提取到的特征对他们进行分类。全连接层的任务是集合高层中所筛选过的图像，将这些特征资讯转化为票数，也就是权重，每一个矩阵的值都需要和神经元相

连，并且每一个神经元都需要去乘以自身的权重，最后加起来就是所获得的票数，最后所输出的类型即为最高的。全连接层可以不止有一层，每新加一层全连接层，卷积神经网络所做出的判断也就会更加的准确。

### 2.1.4 ReLu 函数

ReLu 函数在卷积神经网络中也起到了很大的作用，它是一种激活函数，其主要作用就是在计算中引入非线性。它的数学原理是将特征中那些小于 0 的负值转化为 0，因为这样在矩阵运算中，0 与其他值相乘可以方便计算。

### 2.2 协方差池

协方差池就是想池化层中池化的步骤进行改变，将选取最大值的那一步改为提取协方差，由协方差提取出来的特征相较于其他方式提取的特征更为精准。

#### 2.2.1 协方差

若是想要对两个随机变量进行度量，就需要用到协方差。其公式如图 4 所示，若算出的结果为正值，那么说明这两个随机变量为正相关关系；若结果为负值，说明二者关系为负相关关系；若是零值的话，则说明二者不相关，也就是相关统计学中所谓的相互独立。

$$\begin{aligned} Cov(X, Y) &= E[(X - E[X])(Y - E[Y])] \\ &= E[XY] - 2E[Y]E[X] + E[X]E[Y] \\ &= E[XY] - E[X]E[Y] \end{aligned}$$

图 4 协方差公式

#### 2.2.2 协方差矩阵

在进行图像表情识别时，协方差矩阵是通过将卷积层的输出扁平化后进行向量运算得到的；而在视频表情识别时，是通过将全连接层的输出作为图像集特征，然后从中计算得来。

## 3 数据集

数据集在每一个模型中都不能少，它会直接影响到整个模型的最终结果。在模型算法中，我们通常会将数据集分为训练集、验证集和测试集。顾名思义，训练集是用来训练模型；验证集是用来调整学习模型；而测试集则是用来评估整个模性能。

### 3.1 常用数据集

在人脸识别领域，已经有许多相关的数据集可以用来进行试验与比较，比如 CK+ 人脸表情数据库、JAFFE 表情数据集、FER2013 表情数据集等。这些数据集在网上都能找到，可以通过这些数据集来帮助我们队模型获得更高的准确度。

<sup>1</sup> 此图出自 Convolutional Neural Networks - Basics

### 3.2 FER2013 数据集

数据集总共包含有三万五千多张灰度图像，共分为 7 中不同的面部表情，7 种表情用 0 到 7 的数字标签来替代，其中，0 代表生气，1 代表厌恶，2 代表害怕，3 代表开心，4 代表伤心，5 代表惊讶，6 代表自然。每一张图像的大小都是 48\*48 个像素，但是数据集是将其保存在 csv 文件中的，并没有直接给出图像。如图 5 所示，其中 emotion、pixels、usage 代表了数据的属性，emotion 列为面部表情的类别，pixels 列为图像的具体数据，而 usage 列是属于哪一个数据集。比如 0 代表的就是生气，结尾的英文则代表用于训练集。

emotion	pixels	Usage
0	70 80 82 72 58 58 60 63 54	
01	58 57 56 69 75 70 65 56 54	
50	43 54 64 63 71 68 64 52 66	
47	38 44 63 55 46 52 54 55 83	
47	45 37 35 36 30 41 47 59 94	
43	56 54 44 24 29 31 45 61 72	
53	47 41 40 51 43 24 35 52 63	
54	48 54 73 100 73 36 44 31 37 53 51	
8	61 63 91 65 42 37 22 28 39 44 57 68	
72	56 43 77 82 79 70 56 28 20 25 36 56	
101	105 70 46 77 72 84 87 87 81 64 37	
116	95 106 109 82	Training

图 5 FER2013 数据集

### 3.3 SFEW2.0 数据集

SFEW2.0 数据集是国外一个面部表情识别比赛中所使用的数据集，它经常用于静态面部表情识别，也就是基于图像的表情识别。该数据集的图像绝大多数都是来自于电影中的人物，将不同的表情裁剪下来进行分类从而创造了此数据集。SFEW2.0 数据集同样也分为了 7 种表情。

### 3.4 结果比较

在测试时，原数据集，也就是 SFEW2.0 数据集在测试时只有 58% 左右的准确度，而经过更换后，模型的准确率在 62% 附近摆动。关于原因，主要是因为 SFEW2.0 数据集来源于电影截图，图像内的无关信息较多，对检测有一定的影响，且表情有很多都来自于夜晚，让人脸检测有一点难度从而导致准确度较低。而 FER2013 数据集就没有那么多的无关信息，且都是灰度图像，不存在光线的影响，但是在 FER2013 数据集里面也存在某些错误的标签，所以它的准确度也没有达到一个太高的水准。通过实验比较发现，协方差池的确能够提高表情识别的准确度，除此以外，通过调整迭代次数、全连接层层数以及更改模型框架等，这些方法都可以帮助我们提高实

验的准确率。

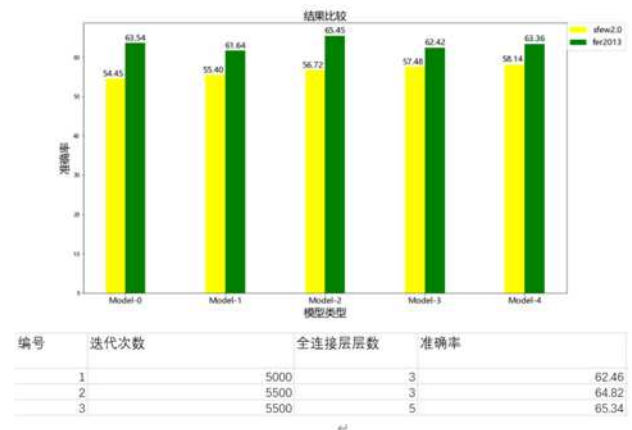


图 6 SFEW 数据集与 FER2013 数据集在不同模型下的准确率及不同参数下的结果

## 4. 应用于教学监督

因为科技的不断发展，面对面的上课方式也遭到了一点冲击，教学模式变得越来越多种多样，比如网络平台教学的出现。但是“新发展新问题”，新的教育方式的出现，也产生了许多问题。有些学生上课玩手机、睡觉，而老师却并不能立即发觉。但人脸表情识别就能帮助老师及时地发现这类情况。人脸识别系统还能发挥许多其他的作用，比如防止考试作弊，提高出勤率等等。

### 4.1 防止考试作弊

在中学阶段，考试时都会有两个监考老师，而一个考场也就四五十个考生而已，所以作弊情况并不严重。而在大学，一个考场往往有接近上百名考生，而监考老师依旧只有两名，这就给平时学习不用功的考生一个“机会”。很多人就找人替考，从而使自己得到不错的分数，而通过检查学生证、身份证等手段还是会存在一些漏洞。但是通过人脸表情识别，我们能够一一核对学生信息，将代考学生拒之门外，而通过考试时考生的表情，我们还可以及时发现考生是否存在作弊行为，极大保证了考试的公平性。

### 4.2 提高出勤率

不管是什么大学，总会出现学生找人帮忙代课的情况，这样的现象屡见不鲜。而使用人脸表情识别，不仅可以杜绝这样的现象，而且可以省去老师点名的时间。对同学们上课时的表情进行分析后，我们能够得知同学们有没有开小差等等，从而提高教学质量。

## 5 总结

如今，通过学习卷积神经网络以及自我的拓展积累，

我们能够很轻松地实现这一项目。它不仅仅只能运用于教学监督领域,目前,医疗、安全驾驶、人机交互等不同领域也开始广泛应用人脸识别技术。在这个用于教学监督的人脸表情识别中,我们通过更换数据集以及调整参数来获取更高的准确度,但最后的准确率还是较低。所以,后续打算通过更换数据来提高准确率,并且加入更多的表情特征,从而增加对于复杂表情识别的能力。

### 【参考文献】

- [1] 陈春燕. 基于深度学习的人脸表情识别 [D]. 浙江: 宁波大学, 2019.
- [2] 施宇楠, 吴自万, 孙文. 基于卷积神经网络的人脸表情识别研究 [J]. 科学技术创新, 2020(05):75-76.
- [3] 周本君. 基于卷积神经网络的人脸表情识别研究 [D]. 南京: 南京邮电大学, 2019.
- [4] 高文, 金辉. 面部表情图像的分析与识别 [J]. 计算机学报, 1997,20(9):782-789.
- [5] <https://blog.csdn.net/tjlakewalker/article/details/83275322>.
- [6] 吴恩达《Machine Learning Yearning》.
- [7] The Extended Cohn-Kanade Dataset(CK+).
- [8] Michael J. Lyons, Shigeru Akemastu, Miyuki Kamachi, Jiro Gyoba. Coding Facial Expressions with Gabor Wavelets, 3rd IEEE International Conference on Automatic Face and Gesture Recognition, pp. 200-205 (1998).
- [9] <https://www.kaggle.com/c/challenges-in-representation-learning-facial-expression-recognition-challenge/data>.