

2018 年英特尔杯大学生电子设计竞赛嵌入式系统专题邀请赛

2018 Intel Cup Undergraduate Electronic Design Contest

- Embedded System Design Invitational Contest

作品设计报告

Final Report



Intel Cup Embedded System Design Contest

报告题目：多对象情绪自动识别分析系统

学生姓名：王鑫辰 尤晓雯 张家恺

指导教师：刘闯

参赛学校：上海科技大学

2018 年英特尔杯大学生电子设计竞赛嵌入式系统专题邀请赛

参赛作品原创性声明

本人郑重声明：所呈交的参赛作品报告，是本人和队友独立进行研究工作所取得的成果。除文中已经注明引用的内容外，本论文不包含任何其他个人或集体已经发表或撰写过的作品成果，不侵犯任何第三方的知识产权或其他权利。本人完全意识到本声明的法律结果由本人承担。

参赛队员签名：

日期：年月日

多对象情绪自动识别分析系统

摘要

多对象情绪自动识别分析系统是一个能够自动识别人脸、区别不同人物身份、由脸部信息进行实时情绪评估，并且将情绪数据进行记录与分析的系统。从结构上讲，它由感知系统，信息实时处理系统、预训练系统以及信息存储系统组成。目前在计算机视觉领域，人脸识别算法已经相当成熟，而人脸表情识别算法(FER)也已有较多人涉足，但后者的推广还不广泛。我们结合了这两种算法，使得 UpSquare 板能像人一样记录不同对象的感情。这其中的技术难点在于对人脸信息的识别和归类，对此我们采用了卷积神经网络的算法训练表情识别模型，达到较理想的识别效果。最后，我们还开发了一套情绪监控的互动界面。

本作品将当下热门的机器学习应用于人类行为学领域，意在通过技术性手段，帮助人们更好的理解与研究人与人行为及其内在含义，有效地预防负面情绪以及心理疾病所带来的危害。同时，在人工智能的人机互动领域，有效地理解指定对象的感情并作出反应有利于机器人表现得更加有灵性。因此，这个项目有相当广泛的应用前景。

关键词：情绪自动识别分析系统，人脸识别，情绪分析，卷积神经网络

IDENTITY-AWARE AUTOMATIC EMOTION DETECTION AND ANALYSIS SYSTEM

ABSTRACT

An Identity-Aware Automatic Emotion Detection and Analysis System (IAAEDAS) is a system which can automatically distinguish between different person's faces, assess certain emotions on the faces and finally record or analyze the corresponding emotion data. In terms of its structure, IAAEDAS consists of a sensory system, a real-time information processing system, a pre-training system and an information storage system. Recently, in the field of computer vision, the algorithms for Facial Recognition have already been mature, while Facial Expression Recognition is a problem explored by many. Nonetheless, algorithms for the latter have not been widely promoted. We combine those two kinds of algorithms, which enabled the board to recognize and memorize people's emotions like a human. The difficulty in the procedure is how to recognize and classify the information delivered by a face. To solve this, we resorted to the CNN (Convolution Neural Network) to train our emotional recognition model. Finally, the results were presented by an interactive emotion management interface.

Our design has applied machine learning, a recently hot topic, to the realm of human behavioral science. It aims to help people comprehend interpersonal behaviors and their implicit meanings through technological methods, while at the same time prevent the detrimental effects of negative emotions and mental illness. Furthermore, understanding and responding to people's emotions would be the next milestone for AI to communicate with people. Thus, the importance of emotion recognition and analysis is self-evident, and the prospect of this topic is brilliant.

Key words: automatic emotion detection and analysis, face recognition, emotion analysis, convolution neural network

目 录

| | |
|--|----|
| 第一章 绪论 | 1 |
| 1.1 表情识别算法综述 | 1 |
| 1.2 人类的面部表情及情绪 | 2 |
| 1.2.1 FACS 及表情分类 | 2 |
| 1.2.2 表情与情绪 | 4 |
| 1.2.3 情绪与效率 | 4 |
| 1.2.4 情绪与大数据分析 | 4 |
| 1.3 卷积神经网络 | 5 |
| 1.3.1 全连接层 | 5 |
| 1.3.2 全局平均池化层 | 5 |
| 1.4 本章小结 | 5 |
| 第二章 系统介绍 | 6 |
| 2.1 硬件组成 | 6 |
| 2.2 功能实现流程 | 6 |
| 2.3 系统构架 | 7 |
| 2.3.1 感知系统 | 7 |
| 2.3.2 预训练系统 | 8 |
| 2.3.3 信息实时处理系统 | 8 |
| 2.3.4 信息存储系统 | 8 |
| 2.4 本章小结 | 8 |
| 第三章 模型的建立与训练过程 | 9 |
| 3.1 卷积神经网络概述 | 9 |
| 3.2 图像处理 | 9 |
| 3.2.1 图像预处理 | 9 |
| 3.2.2 图像进入卷积层(Convolutional Layer) | 10 |
| 3.2.3 特征图进入批归一化层(Batch Normalization Layer) | 10 |
| 3.2.4 特征图进入激活层(Activation Layer) | 11 |
| 3.2.5 特征图进入池化层(Pooling Layer) | 12 |
| 3.2.6 经过以上处理多层后进入全局平均池化层(Global Average Pooling Layer) | 12 |
| 3.2.7 对最终输出的分类向量再次进行 SoftMax 激活 | 13 |
| 3.2.8 基本表情识别模型的建立 | 13 |
| 3.3 训练 | 14 |
| 3.3.1 Loss 函数的计算 | 14 |
| 3.3.2 优化函数的选择 | 15 |
| 3.3.3 数据集的选取和训练 | 15 |
| 第四章 数据库建立与数据分析 | 16 |

| | |
|----------------------|----|
| 4.1 数据存储形式----- | 16 |
| 4.2 个体情绪跟踪的实现----- | 16 |
| 4.3 群体情绪检测----- | 16 |
| 4.4 样例分析----- | 17 |
| 4.4.1 样例介绍----- | 17 |
| 4.4.2 数据分析与误差处理----- | 17 |
| 4.5 本章小结----- | 18 |
| 第五章结论----- | 19 |
| 5.1 预测结果----- | 19 |
| 5.2 潜在问题----- | 20 |
| 5.3 项目应用背景----- | 20 |
| 5.4 项目优化方向----- | 21 |
| 参考文献----- | 22 |

第一章 绪论

随着计算机深度学习技术的不断成熟,计算机在人脸识别、模式识别方面准确率不断提升,采集与识别人脸、人脸表情已经不是一件非常困难的事情。美国心理学家 Mehrabian 认为人类情感信息表达等于 7%语言加 38%声音加 55%面部表情。可见,人脸表情识别技术的实现,将使通过机器分析人的情感变化更加准确。

我们设计的系统使得 IntelUpSquared 板能够通过摄像头实时地识别不同对象的表情所传达的情绪,并将情绪随时间的变化记录下来加以后续分析。要实现这些功能,重中之重是选择有效的的人脸识别特别是表情识别算法,而相关算法层出不穷,有必要作系统性概述。

同时,由于我们最终选择使用卷积神经网络作为表情识别模型的框架,采用一些内置库进行人脸的身份的识别,并将两者结合。在板上运行系统前,先要在其它电脑或服务器上使用标记过情绪标签的图片集训练神经网络,将训练好的网络导入到板上,所以这是一个监督学习问题。而在进行图片所传达信息的标定时,就需要一个权威的的评定标准,FACS 面部表情分类系统是一个很好的借鉴,而 Fer2013 图片训练集为我们节省了不少时间。

1.1 表情识别算法综述

人脸表情识别是模式识别中的一个分支,之前已经有人通过许多不同的方法实现了表情识别,这些算法不全应用了机器学习,但总体而言使用了机器学习算法准确度会相对高一些。这些算法在一般都为图像预处理、特征提取、分类三个步骤。

图像预处理阶段对于不同算法来说大同小异,主要是进行图像去噪、亮度与灰度归一化、尺寸像素调整等,目的是提高图像的质量,更好地进行后续处理。一般会截取脸部部分的图像。Lucy Nwosu等仅截取嘴巴和眼睛的图片用于训练神经网络,结果识别的准确率较其它非机器学习类算法都有所提高^[1];Shikhar Sharma,等更进一步,仅截取嘴巴的图片^[7]。这样做的确能够有效地去除干扰,提高准确度,但有些有价值的信息可能会丧失。

特征提取阶段则在图像中找出更加高级抽象的特点,降级了只够进行分类时的计算量,也能有效地去除没用的干扰。不过这些特征往往都是人根据常识设定的,有些人无法直观理解的有用信息可能会丢失。已有的方法包括:(1)提取脸部几何特征,如五官轮廓等,如 Shirkhar采用MCD(MinimumCovariance Determinant)算法标定出嘴唇上的关键点,再求欧几里德距离^[2];Anagha等使用AAM(Active Appearance Model)算法标定整张脸上的关键点与轮廓^[3];(2)提取脸部图片的纹理,相关的算法有LBP(Local Binary Pattern)算法、Gabor Filter算法等;(3)提取整体统计特征,这种方法较多地保有了原图的所有信息,但准确率容易受光照、背景的影响,主要算法是PCA与ICA(Independent Principal Component Analysis);(4)提取脸部运动特征,常见算法是光流法。

分类是在前一步的基础上完成的,常见算法有 K 临近算法、支持向量机(SVM)、隐马尔科夫模型、直方图匹配(Histogram Matching)等,当然还有各种神经网络算法。

上述不同的表情识别方法各有长处与短处,因而有文章使用两个不同的算法模型,将输出的结果作综合,达到了模型互补的效果。如 Heechul Jung 等同时训练了两个神经网络,一个针对整张脸部图片,另一个针对脸部几何特征,最后将两个网络的输出结果进行“联

合微调”(Joint Fine Tuning), 准确率较其它只使用了一个模型的算法有显著提高^[4]。不过实际运行时计算量会翻倍, 如果软硬件不允许运行时可能会卡。

本文所采用的卷积神经网络算法, 特征提取与分类两个步骤的训练是同时完成的, 因为训练与实际使用时都是以整张图片作为网络的输入, 学习过程中自行判断出有价值的特征以及分类标准。这样做有利也有弊, 好处是神经网络往往能提炼出一些人类难以想象的标准, 坏处是如果参数设置不当或噪音、干扰过多, 网络可能会将一些没有价值特征作为归类标准, 发生误判。除此之外, 当训练样本太小时, 容易过拟合。其表现为用以欧洲人为主的训练集训练出的网络在识别亚洲人脸时准确度会降低、光线角度不同时同一表情对象的识别结果不同等。

1.2 人类的面部表情及情绪

自然语言和形体语言是人体语言的主要部分。人类通过面部肌肉的不同变化来表达是一种典型的形体语言, 我们将这种方式成为面部表情。这是我们进行沟通和交往的一种极为重要的方式。我们在与他人沟通时, 也常常会通过观察对方表情来感知对方的喜怒哀乐^[5]。

人类的面部表情蕴含了丰富的信息, 也是人类情绪的主要载体之一。对其进行深入研究, 可以进一步了解到人类对应的心理状态。Mauss 等 (2005)^[6] 发现, 情绪状态和面部表情相关很大 ($r_s > .80$)。因此理解人类面部表情是进一步理解人类情绪的基础。

1.2.1 FACS 及表情分类

美国心理学家 Paul Ekman 和 Friesen 较早地对脸部肌肉群的运动及其对表情控制作用进行了深入研究, 开发了面部动作编码系统 (Facial Action Coding System, FACS) 来描述面部表情 [P. Ekman 1978]。他们根据人脸的解剖学特点, 将其划分为若干既相互独立又相互联系的运动单元 (AU) (见表1-1), 并分析了这些运动单元的运动特征及其所控制主要区域以及与之相关的表情 (见表1-2)^[7]。

表1-1 面部基本活动单元描述

| 序号 | 描述 | 序号 | 描述 |
|----|----------|----|----------|
| 1 | 内眼角上升 | 14 | 酒窝出现 |
| 2 | 外眼眉上升 | 15 | 嘴角向斜下方运动 |
| 4 | 眼眉下降 | 16 | 下嘴唇向下运动 |
| 5 | 上眼帘上升 | 17 | 下巴抬起 |
| 6 | 面颊上升 | 18 | 嘴收拢撅起不张开 |
| 7 | 眼帘绷紧 | 19 | 嘴唇紧抿 |
| 8 | 上下嘴唇互相靠拢 | 20 | 嘴唇张开 |
| 9 | 鼻子皱起 | 21 | 下巴下移 |
| 10 | 上嘴唇上升 | 22 | 嘴张大纵向拉长 |
| 11 | 鼻翼两侧皱纹加深 | 23 | 嘴唇内翻嘴抿紧 |
| 12 | 嘴角向斜上方运动 | 24 | 下颚向前伸 |
| 13 | 面颊鼓起 | 25 | 下颚向两侧移动 |

续表 1-1

| 序号 | 描述 | 序号 | 描述 |
|----|----------|----|----------|
| 26 | 下颚收紧 | 37 | 嘴唇来回摩擦 |
| 27 | 嘴唇抿紧 | 38 | 鼻孔膨胀 |
| 28 | 嘴唇张开 | 39 | 鼻孔收缩 |
| 29 | 下巴下移 | 48 | 上眼帘合起 |
| 30 | 嘴张大纵向拉长 | 42 | 咬住嘴唇 |
| 31 | 嘴唇内翻嘴抿紧 | 43 | 面颊鼓起 |
| 32 | 咬住嘴唇 | 44 | 面颊瘪下 |
| 33 | 面颊鼓起 | 45 | 面颊如吸吮状收缩 |
| 34 | 面颊瘪下 | 46 | 舌头鼓出 |
| 35 | 面颊如吸吮状收缩 | | |
| 36 | 舌头鼓出 | | |

表1-2 六种基本表情的面部单元的运动特征描述

| 表情 | 额头、眉毛 | 眼睛 | 脸的下半部 |
|----|-------------------------|--------------------------|---|
| 惊奇 | 1. 眉毛被抬起，以至于变高变弯 | 1. 眼睛睁大了，上眼皮被抬高，下眼皮下落 | 下颌下落。嘴张开，以至于唇和齿分开 |
| | 2. 眉毛下的皮肤被拉伸 | 2. 眼白可能在瞳孔的上边露出来 | 但嘴部并不紧张，也不分开 |
| | 3. 皱纹可能横跨额头 | | |
| 恐惧 | 1. 眉毛抬起来并皱在一起被拉上来 | 上眼睑抬起来，下眼皮非常紧张 | 嘴张开了，嘴唇或者轻微紧张，向后拉；或拉长，同时向后拉 |
| | 2. 额头的皱纹只集中在中部，而不横跨整个额头 | | |
| 厌恶 | 眉毛压低了，并压低了上眼睑 | 在下眼皮下部出现了横纹，脸颊推动其向上，并不紧张 | 1. 上唇被抬起来 2. 下唇与上唇紧闭，推动上唇向上，嘴角下拉，唇轻微凸起 |
| | | | 嘴角下拉 |
| 悲伤 | 眉毛内角皱在一起 | 眼内角的上眼皮被抬高 | |

续表 1-2

| 表情 | 额头、眉毛 | 眼睛 | 脸的下半部 |
|----|------------------------------------|--|---|
| 愤怒 | 1. 眉毛皱在一起, 并且被压低 2. 在眉宇间出现了竖直皱纹 | 1. 下眼皮非常紧张, 可被抬起或不被抬起 2. 上眼皮是紧张的, 在眉的动作下可能被压低 3. 眼睛愤怒的瞪着, 可能鼓起 | 1. 嘴角向后拉并太高 2. 嘴角可能被张大, 牙齿露出来 3. 一道皱纹从鼻子一直延伸到嘴角外部 4. 脸颊被抬起 |
| 高兴 | 眉毛稍微下弯 | 1. 下眼睑下边可能有皱纹, 可能鼓起, 但并不紧张 2. 鱼尾纹从外眼角向外扩张 | 1. 唇有两种基本位置: 紧闭, 唇角拉直或向下; 张开, 仿佛要呐喊 2. 鼻孔可能是张大的 (非必要) |

1.2.2 表情与情绪

将面部表情作为情绪的研究手段在现如今是具有相当充分的理论依据的。对于情绪的研究涉及了它自身的很多性质和组成成分, 而面对这些相对复杂的问题, 情绪心理学家积累了多年的研究成果, 逐渐地理出了一个头绪, 把情绪归纳为三个不可分割的成分。对这三个成分, 伊扎德 (C.Izard) 称之为生理的, 体验的和表情的三方面。刘易斯 (M.Lewis) 称之为状态, 体验和表情。他们的理论概括已经成为当代强有力的情绪心理学说。研究实践也表明, 通过运用面部肌肉运动模式测量情绪的外部表现被日益认为是重要的指标^[8]。

1.2.3 情绪与效率

经过研究发现, 情绪对于生产效率具有显著的影响。具体表现为:

- 1) 正性情绪对生产效率无显著影响;
- 2) 负性情绪对生产效率有显著影响;
 - ① “紧张的”负性情绪子变量可提高员工的生产效率;
 - ② “痛苦的”负性情绪子变量降低了员工的生产效率;

如果能适时地检测出员工相应的负情绪, 并且采取一些措施来缓解负情绪的进一步加深, 那么工作效率提升便能带来更多的收益^[9]。

1.2.4 情绪与大数据分析

目前对于生产效率的大量研究主要强调资源分配、劳动分工、技术进步、工业组织有效

性等外部因素对生产效率的影响但却忽略了人的因素在生产效率中的重要作用；X 效率理论虽然意识到人为因素，因此将个体行为、习惯纳入考虑因素，但依然对于个人情绪、情感等心理因素对生产效率的影响关注甚少。而对于此方面的研究，研究方式相对比较传统，从少量对照试验的数据分析推而广之。也受限与此前的技术原因，很难大量分析总体趋势。本项目希望能通过对于表情数据的记录，来建立数据库，以供相关研究人员进一步分析。

1.3 卷积神经网络

机器学习和计算机视觉领域的利用构建卷积神经网络的模型对人脸进行识别和表情检测。时下较为热门的话题有无人驾驶，人工智能，智能家居机器人等。而这些问题都离不开对人脸部和面部表情的精准识别，我们需要一种高效率的，准确率高的方法对人的面部表情进行识别。

目前在图像识别相关的领域，卷积神经网络是最为先进的模型，它通过调整上万个参数，实现对图像特征的学习与识别。本文使用了一种较为快速准确的卷积神经网络的模型，可以实时的检测出人脸的表情。目前的表情识别成功率已经基本达到了人类水平。

1.3.1 全连接层

经典的卷积神经网络模型会使用全连接层（Fully Connected Layer）作为最后输出结果前的准备。实际上，全连接层中会含有绝大部分的参数用于调整和计算，这样就使得结果的计算和优化进行的较为缓慢，同时增加了神经网络的黑盒效应以及过拟合现象。如经典的 VGG16 卷积神经网络模型，在最后的全连接层中的参数，占据了总参数量的 90%。

1.3.2 全局平均池化层

针对大量参数用于调整和计算的问题，目前机器学习领域的学者们提出了全局平均池化层（Global Average Pooling Layer）的方法来替代了全连接层，平均的操作使得网络可以从整张特征图中提取特征。最新的卷积神经网络模型的组成部分 Xception 利用了机器学习中目前最成功的两大试验假设的组合，一是残差网络，二是深度可分离卷积结构，实现了在参数量不变的情况下，大大提高了网络的性能。

1.4 本章小结

从人脸识别出发，计算机已经能够从相对复杂的环境中识别出人类脸部轮廓。进而，人脸上的表情能够传达出更多的信息。通过适当的分类方式，可以将采集的到的面部表情分类进而推测出人物的情绪状态。而人物的情绪状态和生产效率又是息息相关的，无论从提高生产效率的角度来说还是进一步研究生产效率和人物情绪状态来说，建立表情数据库的重要性不言而喻。技术上来说，目前机器学习和计算机视觉领域所构建出的利用卷积神经网络模型对人脸进行识别和表情检测的技术逐渐走向成熟。本项目将融合这两个主要领域的研究成果，将其应用在 Intel Up Squared 嵌入式开发板上。

第二章 系统介绍

本系统的硬件构成较简单，其核心为 IntelUp2 板卡，外接摄像头、显示器与键鼠。使用 Linux 系统作为 Up2 的软件架构，目前主要通过命令行对系统进行操作，之后可能还会搭建可视化的操作界面。由于此板卡十分轻便，这套系统可以在很多场合使用，如教室，医院，超市等等。本版卡也有十分出众的 I/O 扩展能力，未来或许可以加如更多的感知系统用以增添功能。

接下来将对系统的硬件参数、系统结构以及运行流程作详述。

2.1 硬件组成

- (1) Intel Up2 板卡
- (2) 200 万像素 OV2710 无畸变摄像头
- (3) 任意外接键鼠与显示器
- (4) 远程服务器

2.2 功能实现流程

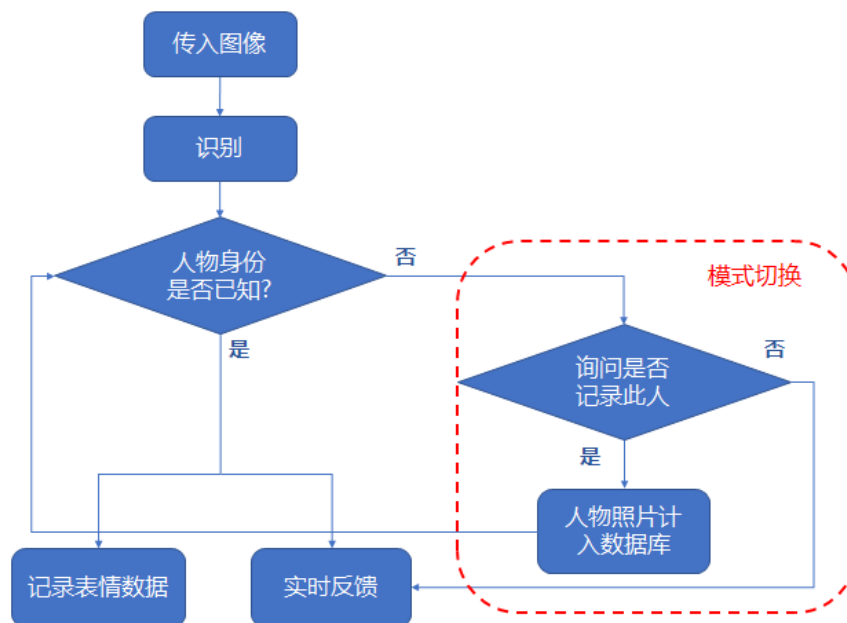


图 2-1 系统功能实现流程

系统的功能实现流程参考图 2-1，本系统在运行时有两种模式，区别是是否运行红色方框图中所示的功能。

在第一种模式下所有功能都会运行，开机后实时读取摄像头的图像显示在外接屏幕上，同时屏幕中会用方框框出所识别的人脸，在方框图上方标出识别出的情感；若系统认识某个人（技术上是指已经将该人的照片存入数据库），方框下方会标出该人的名字，反之则标

为 unknown；同时根据所识别的情绪不同，方框和字体的颜色也不同。除了显示屏上的实施反馈，系统还会纪录每一时刻所读出的表情数据，但是这种数据

记录是有选择的，仅会记录系统已经认识的人物的表情数据，未知人物的表情数据不会作记录。尽管可以通过给未知人物排序标号来实现对于未知人物表情数据的记录，但实际操作起来有一定困难：不仅系统有时会在短时间内将已知人物或非人对象认作未知人物，而且若应用于人流量很大的地方必然会出现很多未知人物，如果把所有识别到的未知人物都记录下来，会导致数据量过大存储不下或产生过多冗余信息。除此以外，对于同一个未知的人，机器可能会认作多个不同的人，将这些数据合成一个人又要大费周折。这固然可以通过减小误判、增加人物识别准确率、引入联网的大数据平台、开发甄别干扰数据与数据合成的算法来解决，但时下我们的设计只是一个雏形，仅针对一个板卡，上述改善方法的实施还是后话。

在系统识别出画面中存在未知人物时，系统会在命令行中询问用户是否要将该未知人物加入数据库，如果选择加入的话用户要给该人物起名。这时系统会创建文件夹把该人无照判存储起来。如果用户也可以实现在数据库中添加默认的人物及照片用于身份识别。已经询问过的未知人物不会再次询问。

在第二种模式下，除了图示红框内的流程不运行以外，其它都与第一种模式相同，命令行内不会弹出询问内容。

2.3 系统构架

本系统的结构如下图 2-1 所示，主要由感知系统、信息实时处理系统、预训练系统和存储系统组成。

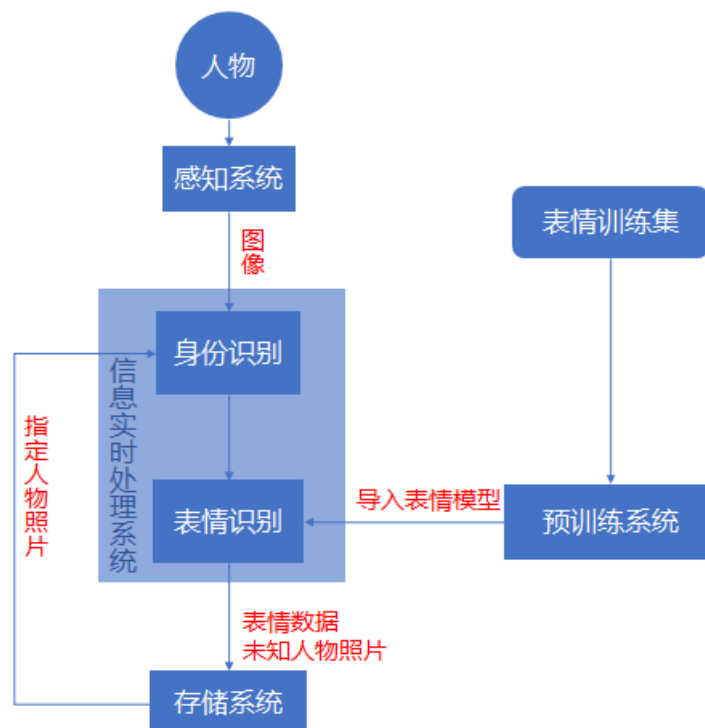


图 2-2 系统基本架构

2.3.1 感知系统

感知系统是指与板卡相连的用以采集实时数据的传感系统，是板卡 and 外界交互的主要方式。目前感知系统中只有用于读取图像的摄像头，将来可以加入麦克风、EMG、心电、体

温等不同传感器，使得其能够更加全面地感知人类的情感，做成诸如聊天机器人、测谎仪之类的成品。本设计也可以从外部将视频直接导入系统当中，将来可以考虑将多个摄像头的的数据通过拓展槽或无线途径传入板卡，进行多场景或同一场景多角度的识别。

2.3.2 预训练系统

预训练系统用于事先训练表情识别所用的卷积神经网络，不位于板卡上，因为训练神经网络计算量大，无法在没有计算棒的情况下完成。本设计使用远程主机进行神经网络的训练及优化。未来可以考虑给板卡外加神经计算棒，将模型训练与信息实时处理集成于板卡一身，增加 Up2 的计算速度。这样既可以将板卡作为多个信号输入的处理主机，也可以读取数据的同时增加一个实时受训练的网络系统，实时给予网络识别结果的反馈，这样训练的准确度就不会再受限于训练集的选取。这样的话，可能的人工智能机器人在与人互动的过程中有了短期记忆与逻辑判断，不会像时下的聊天机器人那么呆板了。

2.3.3 信息实时处理系统

信息实时处理系统和存储系统都位于 Intel Up2 板卡上，是实时的系统，由人物身份识别部分和表情判别部分组成，同时也隐含了从画面中辨别出人脸，用方框标出人脸位置等基本程序。该系统一方面实施地将处理结果反馈在显示器上，一方面将输出的表情数据喂给存储系统，另一方面通过命令行实时与用户互动。我们在 Up2 上安装了 Pytorch、OpenCV 和 Python 人脸识别库。人物身份识别部分调用了 python 的库，身份识别后把所需的人脸图片传给表情识别程序或是否加入内存的判别程序；表情识别程序则需要调用已经训练好的神经网络模型，最后输出之前所有结果。

2.2.4 信息存储系统

读取的数据存取在板卡内存里，以后可以考虑通过无线网卡把数据传到服务器上，这样可以进行数据库与大数据的管理和分析。

2.4 本章小结

本章介绍了我们设计的硬件与软件结构、设备的详细参数，理清了系统运行与准备的流程与设计与用户互动的方式，也提到了我们所使用的操作系统与算法所使用的程序库。

第三章 模型的建立与训练过程

3.1 卷积神经网络概述

卷积神经网络（Convolutional Neural Network）是一种深度前馈式人工神经网络，它利用神经网络的一个个单元，通过卷积的方法来模拟生物的神经网络构造，在卷积神经网络中，其单元，又称卷积核（Convolution Kernel），功能如下图 3-1：

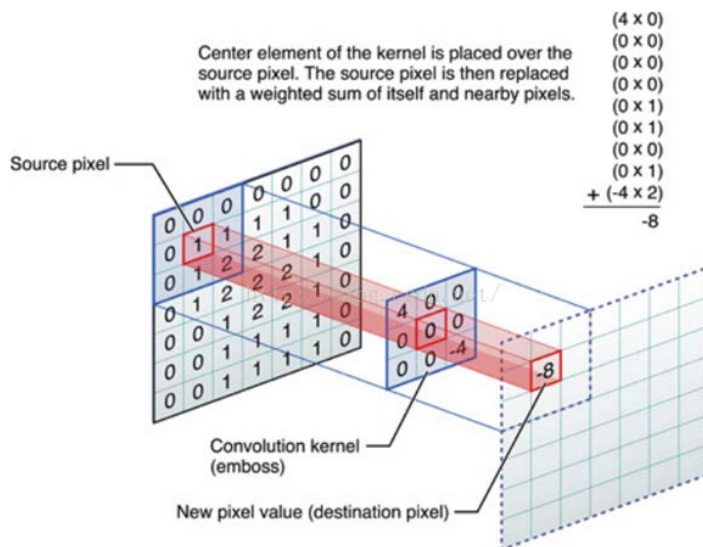


图 3-1 卷积核的基本功能

如图所示是一个感受野为 3*3 的卷积核对图像进行操作，从而在目标像素处得到新的图像值的过程，这一步的目的是，提取前一层图像的一部分特征，而卷积核的参数就是我们深度学习过程中主要需要优化的部分。我们将得到的新的图像叫做特征图（Feature Map），这一部分可以用一个公式来进行总结：

$$FeatureMap(i, j) = \sum_{m=1} \sum_{n=1} K(m, n) * Input(i - m + 1, j - n + 1) + b \quad (3-1)$$

其中 K 为卷积核，Input 为输入图像，Feature Map 为输出图像，

m, n 用来确定卷积核上的一点的值与 Input 进行卷积, b 为偏移项。

在明白了卷积的基本原理后，我们通过图像的预处理和对卷积神经网络各层作用的解释来介绍我们模型的建立过程。

1.2 图像处理

3.2.1 图像预处理

我们采用的是 FER-2013 的数据集进行训练，通过简单的抽查，我们发现这些图片都是灰度的，并且大小为 76*76，我们将图片 Resize 为 48*48 的图像后，对图像进行了去均值化（Zero-Centered），归一化（Normalized）。即

$$x'_i = \frac{x_i - \mu}{\sigma} \quad (3-2)$$

μ 为 x_i 的平均值

σ^2 为 x_i 的方差

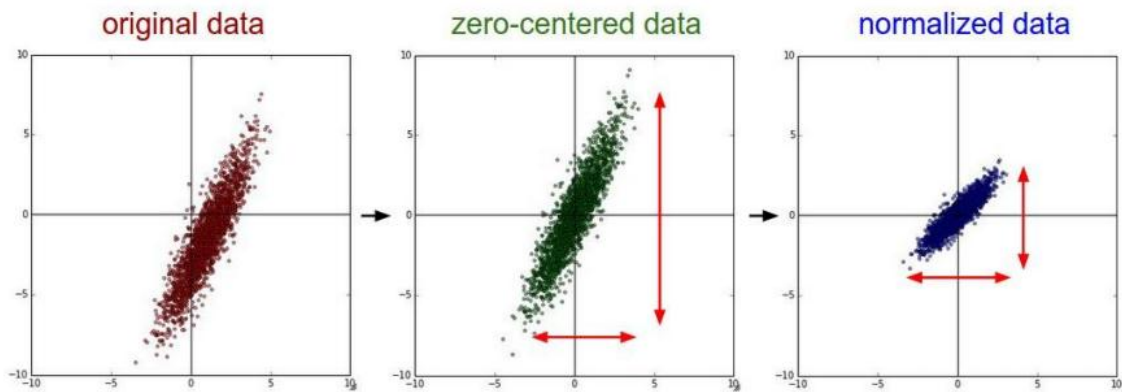


图 3-2 像素点处理的可视化流程

并且在将图像输入模型前进行图像的随机旋转和对称。

接下来就可以将图像输入网络进行训练了。

3.2.2 图像进入卷积层（Convolutional Layer）

之前我们对卷积的概念进行了解释，卷积层的作用就是用多个卷积核对图像进行感知得到多张特征图，通过不断的学习来从图像中感知出不同的特征，从而使得模仿人的学习过程。

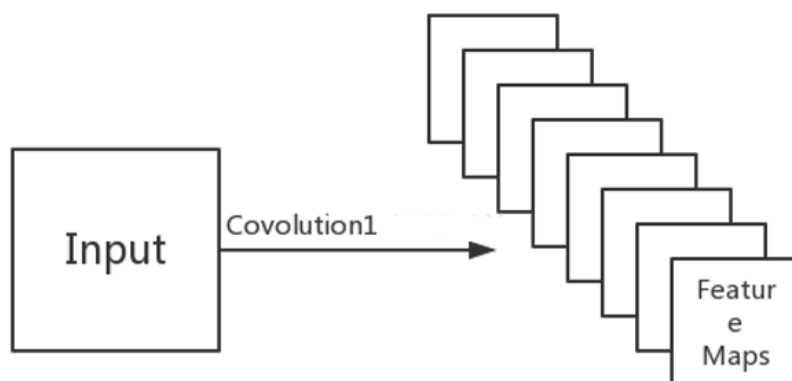


图 3-3 卷积层作用的可视化表示

3.2.3 特征图进入批归一化层（Batch Normalization Layer）

批归一化层也是近几年来才提出的一个神经网络方面的新结构，其作用是解决网络的梯度弥散和梯度爆炸效应，随着神经网络层数的增加，深层的参数变化受前层的参数变化影响较大会发生梯度爆炸，而太小则会发生梯度弥散，而批归一化的作用可以消除这种效应

批归一化层一般放于激活函数后，而在激活函数比较特殊的时候，有时放在激活函数前比较好，对于这次我们所使用的 ReLU 激活函数来说，BN 层放在激活函数前更好。

批归一化层主要做的工作可以用以下四个公式和一张图概括：

$$\mu_B = \frac{1}{m} \sum_{i=1}^m x_i \quad \text{mean of mini-batch}$$

$$\sigma_B^2 = \frac{1}{m} \sum_{i=1}^m (x_i - \mu_B)^2 \quad \text{variance of mini-batch}$$

$$\hat{x}_i = \frac{x_i - \mu_B}{\sqrt{\sigma_B^2 + \epsilon}} \quad \text{normalize}$$

$$y_i = \gamma \hat{x}_i + \beta \stackrel{\text{def}}{=} BN_{\gamma, \beta}(x_i) \quad \text{scale and shift} \quad (3-3)$$

y_i 为输出的特征图上的像素

γ , β 用于还原卷积后的特征图要学习的分布^[10]

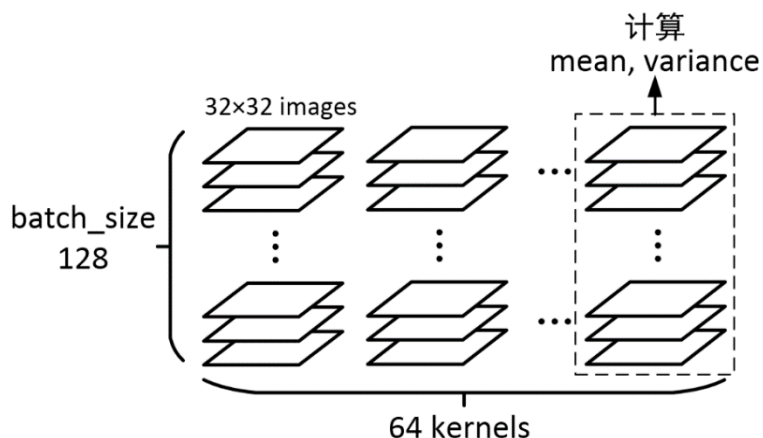


图 3-4 批处理化的可视化操作

3.2.4 特征图进入激活层（Activation Layer）

和图像预处理所相似，我们在学习过程中也要不断的处理特征图来使得学习更加的准确和有效率，因此我们要使用激活函数对特征图来进行进一步的处理，如对信号的非线性话，稀疏化等。常见的激活函数有：

$$\text{Sigmoid}(x) = \frac{1}{1 + e^{-x}} \& \text{tanh}(x) = 2\text{sigmoid}(2x) - 1$$

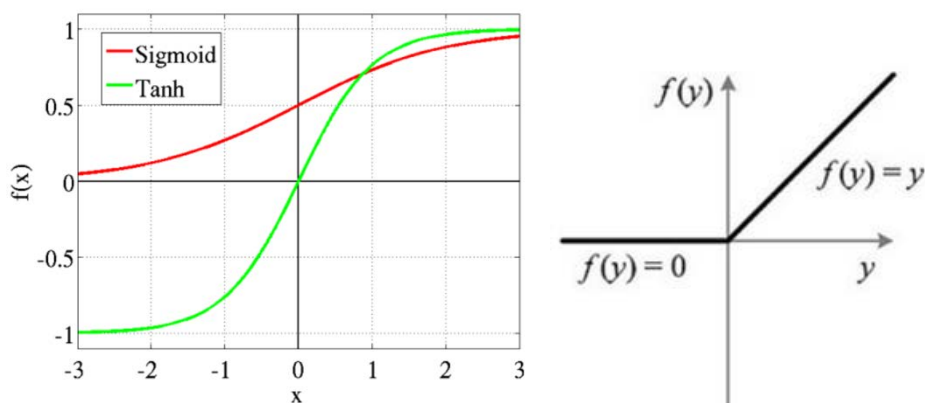


图 3-5 Sigmoid, Tanh 和 ReLU 激活函数的图像

而这次我们使用的是之前所提到的 ReLU 函数：

$$\text{ReLU}(x) = \begin{cases} 0 & x < 0 \\ x & x \geq 0 \end{cases} \quad (3-3)$$

不同的激活函数的作用和效果都是不同的，最早人们常常使用的 Tanh 和 Sigmoid 激活函数因为较为致命的缺点被人们用的已经越来越少了，具体原因是这两个函数都是把输入压缩到了一个长度为 2 的区间内，起到了归一化的作用，可当输入的数据非常大或者非常小时，这些数据的特征会被隐藏，也就是说，如果输入的数据落到了激活函数的非线性区间内，那么对于参数的调整可能使得网络实际的运行并没有实际上的改变，因为数据进入了激活函数的饱和区，这也会导致较为严重的梯度弥散现象。

ReLU 激活函数是学者们通过模拟生物的神经元在特定事件下总是部分激活的而构建出的函数，目前 ReLU 函数和其变种在深度学习领域运用的十分广泛。其优点在于当输入大于零时都是线性区间，因此可以避免梯度弥散效应，同时可以讲输出数据稀疏化来优化学习效率 and 准确度。

3.2.5 特征图进入池化层（Pooling Layer）

池化层的作用是将图像进行压缩，其作用是防止过拟合和减少特征的冗余，但与此同时也会带来信息的缺失，因此 Pooling 层的使用需要格外的注意，目前网络中比较常见的两种池化方式，最大池化（Max Pooling）和平均池化（Average Pooling）。

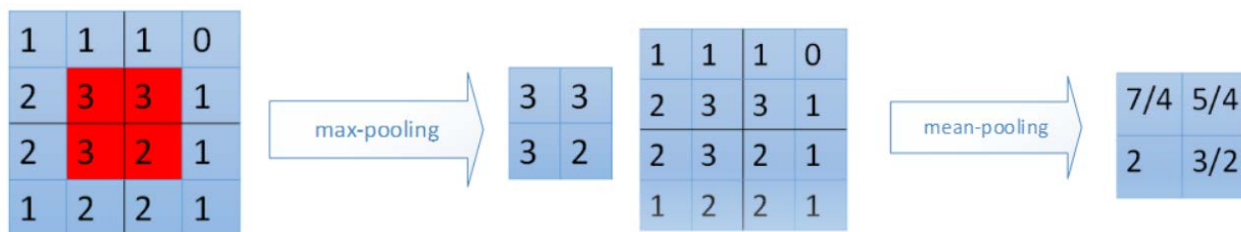


图 3-6 最大池化和平均池化的可视化表示

3.2.6 经过以上处理多层后进入全局平均池化层（Global Average Pooling Layer）

在此前已经提到了全局平均池化的优点，这里用一张图简单的说明平均池化的作用

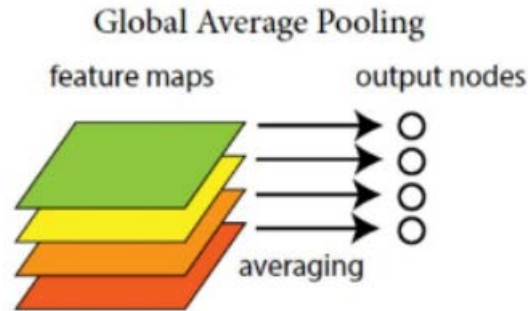


图 3-7 NIN^[11]中对 GAP 的可视化表示

3.2.7 对最终输出的分类向量再次进行 SoftMax 激活

softmax 函数的作用是将多个神经元的输出，映射到 0, 1 区间内，将其看成概率来理解，多用于类似表情识别的分类问题中，其表达式如下：

$$S_i = \frac{e^i}{\sum_{j=0} e^j} \quad (3-4)$$

3.2.8 基本表情识别模型的建立

基于以上神经网络的理解，我们构建了更为复杂的 CNN 网络，其中参考了 Xception 方法对学习的参数和网络的性能进行了优化。Xception 方法是利用了 Depthwise Separable Convolution 的方法对参数进行优化的，简单而言，对于输入为 M 张特征图，输出为 N 张特征图的 3*3 卷积层，本来一共需要 M*N 个卷积核进行卷积，而这里我们先利用 M 个 3*3 卷积核对每张输入的特征图进行一对一卷积，再用 N 个 1*1 卷积核对 M 个卷积核输出的 M 张新的特征图进行卷积，从而得到 N 张用于输出的特征图。经研究表明，这种方法不会损失训练的准确度，而且可以大幅度的提高训练的效率。应用这种方式，我们构建出的基本神经网络模型如图 3-8：

这个模型的优点不言而喻，一是使用了 BN 层来防止梯度弥散和梯度爆炸效应，二是使用了 GAP 层和 Xception 的方法实现了对训练参数的大幅缩减，这样构建出的模型的参数数量比使用全连接层的序贯式模型（Sequential Model）的参数数量少了约 90%，其次在训练结果上来看，并没有出现任何的过拟合现象，测试集的准确率甚至高于训练集，可见模型对数据的泛化能力非常的高。

同时本模型由于参数较少，家用电脑可在 20ms 左右就实现对表情的识别，这意味着我们可以直接将此模型用于实时的预测人的面部表情和实时的刷新，这是十分振奋人心的，因为经典的模型由于参数量过大，必须需要大量的计算才能得到结果。

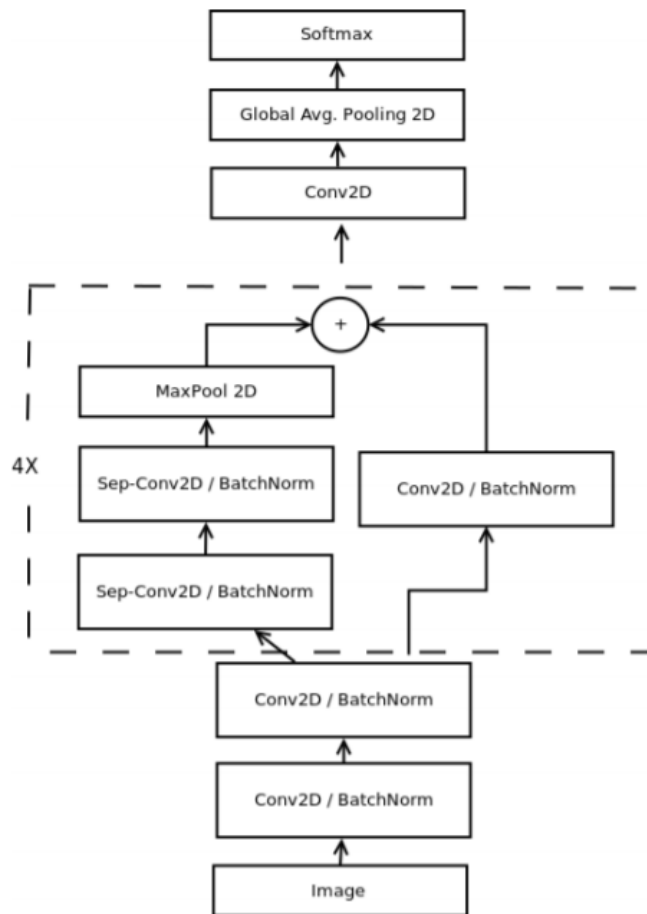


图 3-8 我们所构建的现代神经网络模型^[12]

1.3 训练

训练的最终目标是实现 Loss 函数不断的减小，直到收敛到某一个值为止，如果网络构建的好的话，训练集和测试集的准确率将会十分相近。因此 Loss 函数和优化函数的选择在这里显得十分重要

3.3.1 Loss 函数的计算

与 Softmax 函数相对应的 Loss 函数是交叉熵 loss 函数，其表达式如下：

$$\text{Loss} = -\sum_x p(x) \ln(q(x)) \quad (3-5)$$

其中 p 为期望输出， q 为实际输出， x 为向量的维度。

举个例子就是假如期望的输出为 $(1, 0, 0)$ ，而实际的输出为 $(0.5, 0.4, 0.1)$ 那么

$$\text{Loss} = -1 * \ln(0.5) - 0 * \ln(0.4) - 0 * \ln(0.1) \quad (3-6)$$

当实际情况与期望的输出越接近的时候 Loss 就会不断的减小。

3.3.2 优化函数的选择

这里我们选择的是 Adam 优化算法对网络权值进行优化, 这种算法的好处在于可以自动调整学习效率, 适用于大数据的处理, 这里不再对 Adam 算法的具体操作进行讲解, 下图为算法的具体实现过程。

Algorithm 1: *Adam*, our proposed algorithm for stochastic optimization. See section 2 for details, and for a slightly more efficient (but less clear) order of computation. g_t^2 indicates the elementwise square $g_t \odot g_t$. Good default settings for the tested machine learning problems are $\alpha = 0.001$, $\beta_1 = 0.9$, $\beta_2 = 0.999$ and $\epsilon = 10^{-8}$. All operations on vectors are element-wise. With β_1^t and β_2^t we denote β_1 and β_2 to the power t .

Require: α : Stepsize

Require: $\beta_1, \beta_2 \in [0, 1)$: Exponential decay rates for the moment estimates

Require: $f(\theta)$: Stochastic objective function with parameters θ

Require: θ_0 : Initial parameter vector

$m_0 \leftarrow 0$ (Initialize 1st moment vector)

$v_0 \leftarrow 0$ (Initialize 2nd moment vector)

$t \leftarrow 0$ (Initialize timestep)

while θ_t not converged **do**

$t \leftarrow t + 1$

$g_t \leftarrow \nabla_{\theta} f_t(\theta_{t-1})$ (Get gradients w.r.t. stochastic objective at timestep t)

$m_t \leftarrow \beta_1 \cdot m_{t-1} + (1 - \beta_1) \cdot g_t$ (Update biased first moment estimate)

$v_t \leftarrow \beta_2 \cdot v_{t-1} + (1 - \beta_2) \cdot g_t^2$ (Update biased second raw moment estimate)

$\hat{m}_t \leftarrow m_t / (1 - \beta_1^t)$ (Compute bias-corrected first moment estimate)

$\hat{v}_t \leftarrow v_t / (1 - \beta_2^t)$ (Compute bias-corrected second raw moment estimate)

$\theta_t \leftarrow \theta_{t-1} - \alpha \cdot \hat{m}_t / (\sqrt{\hat{v}_t} + \epsilon)$ (Update parameters)

end while

return θ_t (Resulting parameters)

图 3-9 Adam 算法的具体实现 ^[13]

3.3.3 数据集的选取和训练

这次的训练我们使用的是 FER-2013 数据集进行训练, 训练中我们使用了交叉验证 (Cross-validation) 的方式对数据进行分类, 分为了训练集、验证集和测试集。

FER-2013 数据集含有“生气、恶心、害怕、开心、难过、惊讶、平和”七种表情, 其中三万八千多张人类的表情图片, 这个数据集的训练比较困难和缓慢, 因此我们最终取得的正确率看上去并不是很高, 只达到了 65%, 但实际上的效果其实已经达到了人类水平的表情识别准确程度。

第四章 数据库建立与数据分析

人脸表情识别模型最终输出所读取画面中所有个体或所有指定个体情绪随时间的变化,存入数据库中,并进行分析。在分析的过程中要结合心理学和统计学的方法,可以应用于许多不同的领域,这可能需要进一步的开发研究。本文先尝试了几种就目前而言最方便实施的分析方法,接下来将做介绍。

将画面中的某个个体用函数 $f_i(t)$ 表示,其中 i 是索引,用于界定个体, t 表示时间。在每一个时间点的输出用六维向量 (a,b,c,d,e,f) 表示,其中字母按照顺序为六种所界定的情感:惊奇(surprise),恐惧(fear),悲伤(sad),厌恶(disgust),高兴(happy),愤怒(angry)。在某一个时间点六个维度中只可能有一个维度的值为 1,其它维度的值都为 0。当所有维度的值都为零时表示情感为中性(neutral),故任意时刻函数有七种可能的输出结果。

针对情绪数据的分析可分为对于指定个体的情绪跟踪以及对于群体的情绪检测。对于个体的情绪跟踪可用于心理疾病的检测与预防,或者推断是谈话中某一个体对于另一个体的情感态度;对于群体的情绪检测则主要用于群体对于公共演讲者的情感态度。

4.1 数据存储形式

当被采集人加入数据库后,每次读取表情数据,都会被写入.csv 文件中。文件中每一条数据的形式为:“name + time + emotion + emotion_vector”name 为被采集人自己输入的姓名,time 为实际时间,emotion 为检测出的表情,emotion_vector 为表示六种表情的六维向量(分别对应 surprise,fear,sad,disgust,happy 和 angry),这个向量的作用主要是为了简化后续计算群体情绪。

4.2 个体情绪跟踪的实现

对于个体的情绪跟踪,可分为分析个体在某一时间段内的情感倾向以及个体情绪随时间的变化。可将已知的外部刺激因素结合起来,分析情感倾向与变化可能的诱因,从而推知个体的喜好;或者根据情感变化模式预测心理疾病或个体行为。

4.3 群体情绪检测

在面对群体对于相同对象的情绪变化时,可以通过归一化的方式,定量求出某一时刻某一时刻某种情感的度。以“惊奇”情感为例,假设群体中一共有 N 个个体,每个个体的惊奇值为 a_i ($i = 1,2,3,\dots,N$),那么群体的惊奇度为:

在

$$A(t) = \frac{\sum_{i=1}^N a_i(t)}{N} \quad (4-1)$$

由此可求出 A,B,C,D,E,F 六个函数,值都在 0 到 1 间。

4.4 样例分析

4.4.1 样例介绍

为了测试我们的系统功能，我们选取了电影《楚门的世界》的片段来进行系统的测试，在影片中，Truman 对女演员 Lora 一见钟情，他羞涩的询问 Lora 是否单身，再被 Lora 指出不能和他说话也不能和他下周出去约会的时候，十分沮丧，而最后 Lora 表示如果要约会的话，现在可以，Truman 的心情又由悲转喜，这个片段中 Truman 的表情有着丰富的变化，很适合作为系统检测的研究对象。

4.4.2 数据分析与误差处理

经过了系统的分析以后，我们得到了 Truman 在整个片段中，每一帧上的检测出的表情，然而由于影片中镜头转换，摄像角度和系统本身的不稳定性等原因，某些帧中 Truman 的表情识别会出现突变，如从快乐变为愤怒，一帧后又变回快乐。

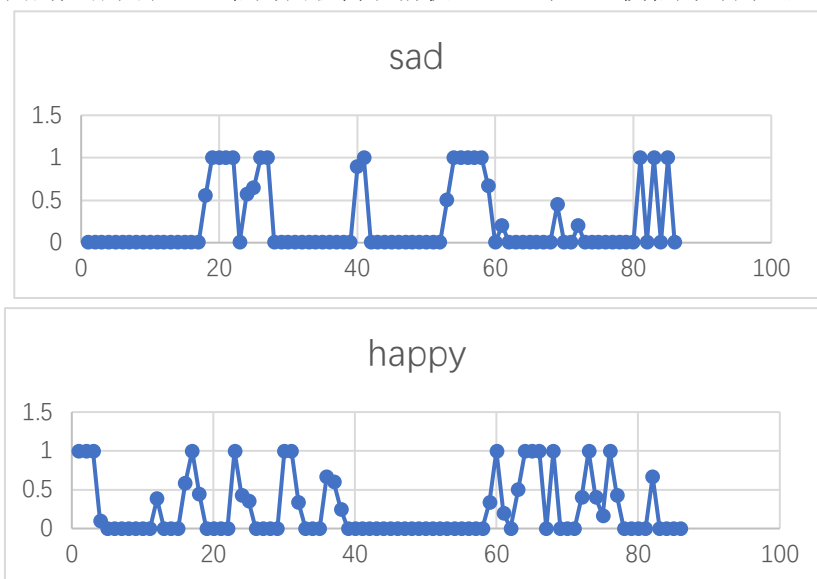
我们采取将两秒内的帧合并为一组数据，取他们的情绪向量之和再除以数据的个数，由此将产生的错误表情造成的影响降低。如我们在两秒内得到了 10 帧，他们的情绪向量分别为

$$\begin{aligned}
 V_1 &= [0,0,0,0,0,1] \\
 V_2 &= [0,0,0,0,0,1] \\
 V_3 &= [0,0,0,0,0,1] \\
 V_4 &= [0,0,0,0,0,1] \\
 V_5 &= [0,0,0,0,0,1] \\
 V_6 &= [0,0,0,0,0,1] \\
 V_7 &= [0,1,0,0,0,0] \\
 V_8 &= [0,0,0,0,0,1] \\
 V_9 &= [0,0,0,0,0,1] \\
 V_{10} &= [0,0,0,0,0,1]
 \end{aligned}$$

显然 V_7 这里情绪的识别出现了问题。经过我们的平均化处理以后：

$$V_0 = [0,0.1,0,0,0,0.9]$$

这样就实现了对于错误表情的纠正功能，对整个片段 Truman 的表情经过处理以后，我们得到了下面的几张图表（竖轴为两秒内表情权重（0 到 1），横轴为时间（/2s））：



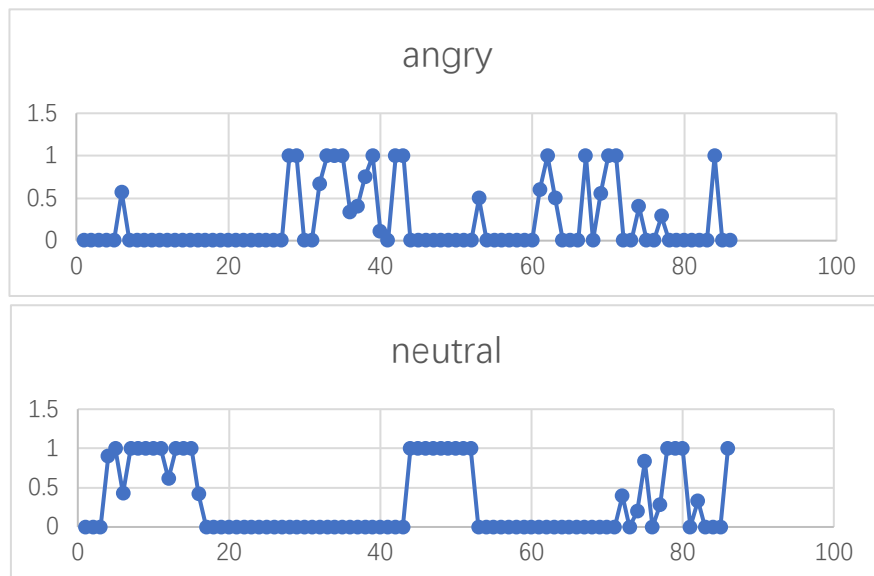


表 4-1: Truman 的四种表情检测折线图

4.5 本章小结

本章我们实现了对于数据库的建立，数据的存储和简单的分析，同时对于一个简单的例子进行了测试。今后我们还可以对大量的数据进行宏观的分析，或者对于一个人的数据进行长时间的分析，得到我们需要的分析结果。本文中我们只是对数据进行了简单的分析，希望可以起到抛砖引玉的效果。

第五章 结论

本项目将当下大热的计算机科学领域的深度学习与卷积神经网络技术应用于心理学,人类科学等社会科学层面,意在通过技术性手段,帮助人类更好的理解不同个体间的行为及其内在表达的意思。

事实上,人类的情绪远比心理学定义的这六种要复杂,根据实际表情所作出的判断是人们可控的。或许计算机还无法判别他们是否是“刻意”流露出的某种情绪,不过无论怎么样,这是我们向计算机科学、心理学及人类行为科学交叉学科迈出的第一步,相信随着更多技术的迭代与更新之后,如更多的微传感器、更多的高效算法以及更大的训练集被开发、组建起来后,在解读人类情绪方面,计算机会有更加出众的表现

5.1 预测结果

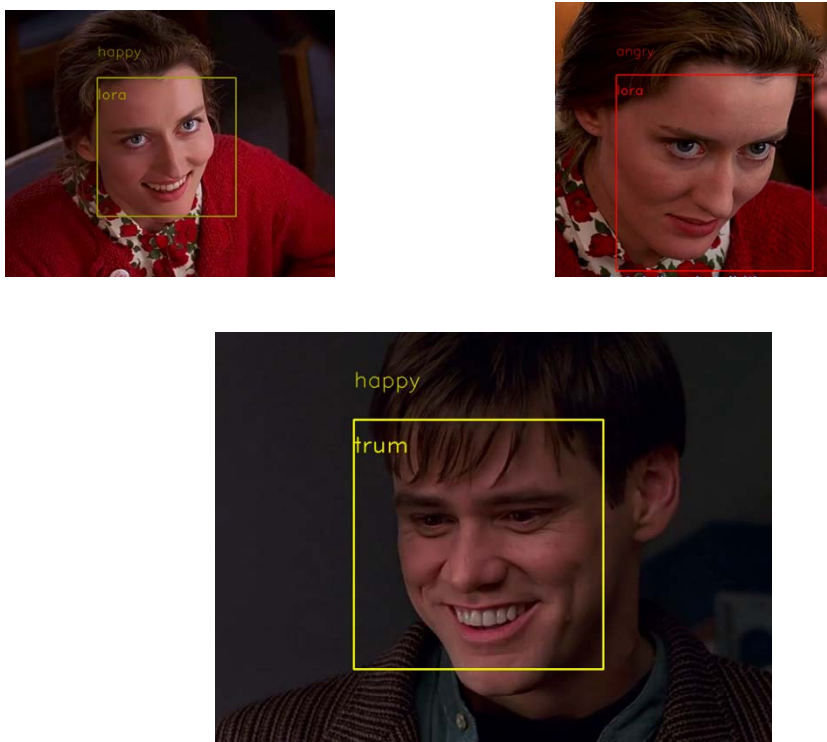


图 5-1 《楚门的世界》片段的人脸表情识别结果



图 5-2 数据库中的 Trum 和 Lora

5.2 潜在问题

当然，当科技的发展到达一定程度的时候，由此也会牵扯出一些如社会伦理的问题：我们对于自身表情及其他肢体所传达的意思是否享有隐私权，他人是否需要经过授权才能来解读这些动作表情？不过基于目前的技术来说，大可不必担心这样的问题。因为在你刻意掩盖表情的时候，计算机还是无法分辨的。

5.3 项目应用背景

（1）企业的员工情绪管理：

现如今企业对于员工的考量已不仅仅覆盖于基本的问题，员工的生理心理状态也是和企业生产效率紧密联系的重要因素之一。基于之前的文献检索和搜查，我们发现员工的情绪和工作效率也有着密切的联系。如果企业的管理者能够发现员工的情绪异动并及时作出相应措施就能改善出于情绪消极而导致的工作效率低下问题；同时，企业的管理者还能够通过这一系统来观测员工针对企业颁布的某项措施的反应，来进一步考虑接下来如何操作。从数据分析的角度，我们的系统能够帮助企业在两种情形下进行分析：特异型和泛化型。

首先，是针对个人的特异性情绪数据分析。员工在每日打卡上下班时，我们系统会记录员工情绪，并对于每一个员工的情绪进行分析。如果发现情绪的异常波动，便会通知企业管理者，及时予以心理疏导，以提高企业员工的心里健康状况和工作效率。

其次，是针对群体性的泛化型情绪数据分析。企业可以采用我们的系统，在施行某一新措施后，对企业员工的情绪状态进行检测，通过采集到的数据与未实施措施的情绪数据作比较，通过和员工调查问卷相结合的方式，分析此项举措对于员工的影响。

（2）在心理学等社科领域辅助分析实验对象的情绪：

本系统可以实现对于实验者情感的初步分析，可以起到辅助心理医生、心理分析员的作用。依赖于专业知识和计算机分析，可以进一步提升心理诊断的判断力，对症下药，以此提升心理治疗的效果。同时，借由这一系统，心理医生也能通过远程更准确的为患者进行情绪及心理状态的评估，给出相应的建议，患者是否需要诉诸于专业机构或者自行调节，能够节省相当的人力物力。

如果能将这些记录下来的数据进行合理利用，这对于这些领域也是不可多得的宝藏。通过对大量数据结合案例的分析，可以得到在时代发展的过程中，某些心理疾病的微妙变化过程，进而进行数学建模，预测可能的变化；通过对个例的分析，也能得到某种疾病在特殊情形下的变体，这些情绪数据都能帮助到医生对于以后的治疗起到辅助作用。

对于特殊疾病的治疗，也有相当的作用。比如自闭症儿童、少儿孤独症等等，患者不愿和人交流，整体沉迷于自己的世界里。本系统可以应用于机器人或仿生人上，系统可以在治疗初期在医生和孩子之间打开突破口，进一步取得信任继而进行疏导和治疗。

同时，这些数据也有助于这个行业的门槛教育。在课堂上，教授能够通过曾经授权的案例数据来帮助学生理解并剖析情绪分析乃至心理分析的过程，以此可在某些阶段就极大地降低了实验的时间成本和人力成本。

（3）商场顾客消费情绪分析：

由于本系统的便携性和适配性相当高，可以通过放在商场、超市等场所对于顾客消费的情绪做出一定分析进而决定其他举措，作商业用途。事实上，类似于企业情绪管理又不同于

它，商家所做的各项措施促销、新品等等，甚至是商场的格局等等次要因此也都会影响到顾客的情绪。通过情绪数据的分析，可以得到整体顾客对于不同类型、不同规格商品的喜好程度；如果再对顾客做详细的定位，可以得到不同群体之间的差异，以此和本身的定位作比较，明确了目标群体后可以更加有针对性的改善潜在目标的购物体验，以此留住顾客的心，进而增加收入。

网络购物也是目前比较普遍的购物方式之一。如果在获得授权的情况下，商家可以通过分析消费者进入商店的情绪、消费者购买商品时的情绪，总结出自己商店的卖点和不足。掌握消费者的心理以此来提升销。

（4）教学反馈系统：

在教育业，学生情况反馈向来是个非常重要的评估教学质量和教学方式的标准。课堂上，本系统可以跟踪每个学生学习时的情绪变化，通过对个人分析，可以得到某一人课堂上的具体情况，以此来评估他是否有认真参与；通过对班级总体情况的分析，可以得到老师教学情况的反馈，以此完成对于老师的初步评估。通过对学生表情数据的进一步分析，可以跟踪获得不同时刻学生针对课堂特定内容的一手反馈，帮助老师提升教学质量，发现重难点并及时强调。当然，如果要做此方面的分析，需要对于学生状态的不同定义，以此来建立数据库并重新训练，此方面的应用仍需要一定的时间。

5.4 项目优化方向

本项目未来的优化方向主要有：感知系统优化，训练集优化，算法优化等等。受限于时间等因素，目前我们仅仅采用了摄像头采集图像，通过表情推测情绪，以此采集数据。事实上人的情绪还体现在很多其他方面，如脉搏变化、呼吸轻重缓急等。未来或许我们可以利用板卡的 IO 优势，加入这些传感器。通过多维度的数据来分析人的情绪变化，更有可信度。

目前我们使用的数据集 FER-2013 考虑到了不同年龄，不同性别之间的差异。但其在人种方面仍欠缺考量，因此在我们亚洲人情绪识别方面的确存在一定误差。为了能够更适用于此类环境，未来可以尝试建立以亚洲人为主导的，涵盖不同年龄，不同性别的情绪数据集。

算法方面，未来还能通过修改优化器、改善神经网络等方式进一步提升情绪识别的准确率。事实上，还有其他方式会将面部进行分割，先对局部区域的器官位置变化、运动等等先进性训练，再讲整体面部一起投入训练，不过因此也需要更强大的硬件新能，或许未来能够训练处更准确，更高效的模型。

参考文献

- [1] Lucy Nwosu, Hui Wang et al. Deep Convolutional Neural Network for Facial Expression Recognition using Facial Parts [R]. Houston, 2017.
- [2] Shikhar Sharma, Krishan Kumar, Member, and NavjotSingh.D-FES: Deep Facial Expression recognition System [R]. Srinagar Garhwal, India, 2017.
- [3] Anagha S. Dhavalikar and Dr. R. K. Kulkarni. Face Detection and Facial Expression Recognition System[R]. Mumbai, india, 2014.
- [4] Heechul Jung, Sihaeng Lee et al. Joint Fine-Tuning in Deep Neural Networks for Facial Expression Recognition[R]. Korea: Advanced Institute of Science and Technology, 2015.
- [5]王志良, 孟秀艳.人脸工程学[M]. 北京:机械工业出版社,2008:15-40.
- [6] Mauss, I.B., Levenson, R.W., McCarter, L., Wilhelm, F.H., &Gross, J.J. (2005) .The tie that binds Coherence among emotion experience, behavior, and physi-ology.Emotion, 5 (2) , 175-190.
- [7] 刘伟峰.人脸表情识别研究[D].安徽:中国科学技术大学自动化系模式识别与智能系统专业,2007:20-25.
- [8] 孟昭兰.为什么面部表情可以作为情绪研究的客观标准[N].心理学报,1987 年第二期.
- [9] 尚倩.基于心理负荷的生产效率研究[D].浙江:中国浙江大学管理学院管理科学与工程专专业,2013:173-180.
- [10] Ioffe S, Szegedy C. Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift[J]. 2015:448-456.
- [11]Lin M, Chen Q, Yan S. Network in network[J]. arXiv preprint arXiv:1312.4400, 2013.
- [12] Octavio Arriaga, Paul G. Plöger, Matias Valdenegro. Real-time Convolutional Neural Networks for Emotion and Gender Classification[J/OL], 2017, https://github.com/oarriaga/face_classification/blob/master/report.pdf
- [13]Kingma D, Ba J. Adam: A Method for Stochastic Optimization[J]. Computer Science, 2014.

附录：程序清单

详细代码见附件。

1. 训练部分

```
convert_csv2grey.py  
convert_fer2013.py  
face_train_use_keras.py  
load_face_dataset.py
```

2. 表情识别部分

```
Tensor_test.py  
emotions.py  
FaceAndEmotionDetection_v2.py
```

3. 数据格式转换

```
data.py
```

4. 环境搭建

需要在 Linux 系统下事先安装 python2，然后再安装一些常见的库包括 pytorch, pip, OpenCV2。还要安装一个叫作 ageitgey/face_recognition 的不常见的库，下载地址为：https://github.com/ageitgey/face_recognition