

# 基于卷积神经网络的微表情识别研究

于赫鑫 钟李明

(泰国格乐大学, 泰国 曼谷 10700)

**摘要:** 人的脸部可以做出不同表情, 这是通过特殊肌肉群组织组合实现的。我们观察人们的微表情, 可以了解到人的真实情感与心理活动, 所以识别微表情开始被应用于公共安全、心理治疗、商业谈判等领域。因为微表情有瞬时性与细微性的特点, 肌肉群图像比例很小, 常规的表情处理方式无法识别, 因而就需要对相关系统进行研发。本文以卷积神经网络技术为前提, 对微表情识别系统的设计进行了研究, 希望可以更加精确的识别出人的微表情。

**关键词:** 卷积神经网络; 微表情; 识别; 困境

微表情是人的自身情感被压抑时, 人体防御机制为表达内心情感而发出的自发式表情。因为微表情具有不能伪造与抑制的特点, 因此在很多领域都有必要应用, 如审讯中的测谎仪就是采用了此原理, 医学领域则能够了解患者的精神状态等。鉴于微表情识别在生活中的应用越来越重要, 因此我们本次将目光聚焦在了卷积神经网络的微表情识别上, 希望可以为后续的微表情识别发展做出贡献。

## 一、卷积神经网络的困境

### (一) 数据集匮乏问题

为了实现对不同类型的人的微表情进行精确的分类, 必须明确提取不同的特征。然而, 由于微表情的动态信息少、持续时间短等特点, 使得其难以捕捉, 因此需要采集高速相机、光线充足、情感诱导合理、样本分布均匀、覆盖不同种族的样本, 以及经过心理专家的鉴定后, 将其输入到数据库中。在此基础上, 构建完整、可信的微表情库难度很大, 导致目前可用的微表情库匮乏。

### (二) 特征提取问题

研究表明, 面部微表情是一种精细的、易受外界因素影响的面部微表情, 其特征的抽取可能与头部转动和室内光线变化有关。随着人们对面部特征的认识, 人们对其进行了深入研究。由于图像中的图像具有不同的属性, 因此在进行图像分割时, 首先要解决的问题就是如何进行图像的分割, 以及如何对图像进行归一化处理以及如何选择合适的特征。

### (三) 样本量过小与深度学习不匹配问题

深度神经网络具有从底层向高层抽取的优点, 利用其微表情特点, 可以实现与其良好的匹配。然而, 当前建立的微表情库中样本数量小、分布不均, 在学习时难以收敛。为此, 本项目提出了新的、适用于微表情识别的深度神经网络方法。

## 二、卷积神经网络概述

### (一) 卷积神经网络的结构

卷积运算是常用的基于卷积结构的神经网络运算, 在图像的降噪、关键信息的提取、标识亮度的改变以及对类别图像的特征的辨识等方面有着广泛的应用。卷积核心是卷积运算的中心, 它是具有一定规模的、以一定间距在输入数值上运动的矩阵, 然后将卷积核的各个参量与所涵盖的区域的参量相乘, 再将所得的积相加, 所得的结果即为输出量, 依次进行类推。一般来说, 一个卷积核对应图像当中的某个特征, 二者一一对应之后就与图像中某一种特性相对应的卷积函数进行抽取运算, 比如索贝尔边界检测器。该算法使用边界卷积核算子来获取图像的边缘信息, 是按照像素的尺寸来确定的, 数值大则是边缘, 数值趋近于 0 则不是。层层相连的网络结构使得卷积神经网络能够有效抽取各种类型的特性, 仅需设定卷积核的个数。

### (二) 卷积神经网络的优点

利用卷积网络对卷积网络中的各个结点进行完全连通, 从而产生了局部连通。局域连接技术通过切断各层级的神经元结点, 通过对其进行激励, 仅对上层的特定结点进行链接, 从而减少了各层级间的计算复杂性。加权分享结构充分发挥了单一卷积核对单一特性的探测能力, 对不同区域使用统一的连接权和偏倚, 使得卷积计算核层以上的加权系数与空域值共享, 提高网络性能。

针对神经网络学习过程中存在的“过度学习”“过度学习”“局部优化”等问题, 本项目提出在卷积神经网络中引入“局部连通”与“权重分享”两种结构。这两种方法的架构见图 1, 左侧为本地连通, 表示各层各结点仅与上层三个结点相连, 并未将上层各结点与各层的结点相连。在此基础上, 本文提出了一种基于遗传算法的改进算法, 该算法能够有效地降低网络中的权重, 从而有效降低了网络的学习成本。图 1 右侧是权重分享, 可以看出在权重分享的过程中, 增加了权重分配后, 网络中的部分权重系数被缩减了 9 个, 从而使网络的学习效率得到了进一步提升。

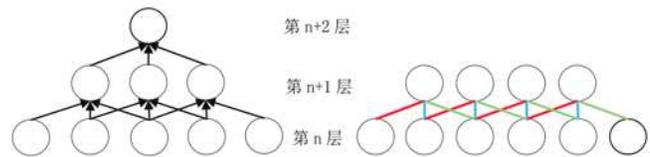


图 1 局部连接、权值共享示例图

### (三) 卷积神经网络的训练

**向前传播阶段:** 首先随机选择一个数据样本, 将其传至网络系统中, 然后从上到下以逐层的方式映射操作, 最后就可以算出误差。

**反向传播阶段:** 研究人员应用极小化误差法能够对流程加以优化, 从而使调节系统中不同的权值进行调整, 最后就可以实现网络参数优化。

通过以上两方面的研究, 我们可以归纳出具体的学习过程: 第一步, 将样本资料集合分成不同的类别, 并根据比例将其分割成不同的类别。根据实际情况, 把它分成了一个训练组和一个试验组, 如果有必要, 还可以把它分成一个检验组。其次, 网络的权重, 门限, 参数的初始分配要尽量少。然后, 对所分割出来的训练集合进行输入, 并对其进行运算, 并对其进行误差分析。同时, 依据这些偏差, 调整并设定每一层的参量及门限值, 并且要看每一个参量能否满足预期的网络准确度, 如果符合, 就不必再进行回路运算。如果不符合要求, 再次将所得到的训练集合重新投入, 如此反复, 直至精确程度符合要求。在此基础上, 将目前最适合的权重、阈值及学习参数存储起来, 得到调整后的网络模型。最后, 通过对试验样品进行统计分析, 验证该方法的精度。

### 三、微表情识别设定

拟采用稠密光流和卷积神经网络,实现对面部微表情的最大程度提取,同时兼顾时空两个维度,同时减小无关信息对面部微表情的影响。将两者相结合,构建面向用户的用户微表情识别算法。针对微表情的时间特征,拟采用稠密光流法,将微表情视频的运动信息进行捕获融合到一张图像中,通过卷积神经网络对其时间特征进行抽取,以此进一步改进网络深浅和优化空间步长,提高空间特征的提取能力。在数据预处理环节,提出了新的基于稠密光流法和卷积神经网络微表情识别方法——OFCNN。

### (一) 微表情数据集

微表情处理的发展离不开成熟完备的数据集。微表情因其非自主、瞬时性和微小化等固有特性,难以实现在自然情况下对其进行有效的捕获与收集,即使是在受控的条件下也难以实现。已有研究利用高风险情境诱发人的面部微表情。比如,让被试说谎,掩盖令人不快的影片带来的消极效果,模仿愉悦感,或是模仿犯罪现场的偷窃和撒谎之类的。从2011年起,国内外学者先后建立了8种典型的微表情数据库:USF-HD、Polikovskys' sdataset、York DDT、SMIC、CASME、CASME II、SAMM和CAS(ME)。其中,USF-HD和Polikovskys的数据集含有摆姿势的微表情,这些动作需要被试模拟微表情运动,所以这些都是不自然的,带有很大的主观色彩。很难表现出真正的人类情绪。另外,York DDT库中包含了谎言诱导的微表情,其生态效度也很高。但是这份资料中包含了一些与讲话有关的无情感的脸部动作。另外,由于这三组资料均未对公众开放,所以这三组资料在当前的研究中几乎都处于闲置状态。然而,由于诱导被试进行谎言诱导的微表情具有两大缺点:容易受到不相关人脸动作干扰以及微表情种类的局限,因此,未来研究人员一般会假设,在情感类影片中,尽量维持中立或抑制情感表达。这是最好的触发方式。在该模型基础上,构建了5个不同的微表情库:SMIC,CASME,CASME II,SAMM,CAS(ME)5个。

我们将利用奥卢大学的微表情研究课题组创建的三种类型的摄像机采集的数据子集:高速摄像机、NIR相机。由于微表情短暂的特性,HS可以用来研究其迅速改变的特性。另外,VIS和NIR子集可用于增加数据集的多样性,从而发展多种方法来实现对面部微表情的分析。但SMIC数据集的标注方式并不全面,仅包含积极、消极和惊讶三种情绪标签,并且没有包含标签的行动单元。

CASME系列的数据集的建立和开发均来源于中国最先进的微表情研究小组傅小兰课题组。他们都采取了同样的试验方法,选择四盏灯提供稳定且高强度的照明,保证了光源的稳定性和亮度(防止了由于交变电流而产生的闪光)。研究人员让受试者在看到具有强烈情绪冲击力的录像剪辑时,脸上的表情不会有任何变化。这三个数据集都有AU标注和更加丰富的情感标注。CASME有开心、悲伤、讨厌、震惊、压抑、害怕、焦虑和轻蔑等8种情感标签,CASME II基于首个数据集,利用更快、更高级的摄像头捕捉的微表情,将采集速度从60帧提升到200帧,包含7个情感标签,分别是开心、悲伤、震惊、恐惧、压抑、讨厌和其他,每一种情感标记下,都有更多的样本,而且标记的质量也更好。

维持中立的面部表情本身就是一项挑战,尤其是在尝试同时引发微表情和宏表情时,难度更大。CASME(ME)数据集旨在收集包含微表情和宏表情的数据,以推动长视频微表情检测的研究。该数据集涵盖了快乐、悲伤、惊讶和中性等四种情绪类别。SAMM数据集与CASME系列类似,都使用了高速摄影机和优质的实验条件,并特别设计了照明设置以减少光线变化的干扰。在实验前,参与者需完成问卷,以确保情绪刺激视频能准确引发预定的微表情。SAMM数据集进一步细化了情绪分类,共定义了七种情绪,

比CASME数据集少了焦虑情绪标签。

### (二) 数据集预处理

因为采用了较高帧率的高速摄像机进行微表情的获取,使得视频中出现了很多类似的图像,帧间区分度较低,为解决该问题提出了基于尖峰帧先验知识的峰值帧识别方法,利用提取起始帧和峰值帧的方式,有效地消除了信号中的多余信息。

起始帧指的是微表情视频采样分解过程中的首张图像,而尖峰帧则是表示情感最丰富的单帧图像,峰值帧体现了样本序列情绪表达的峰值图像,展示了情绪的最高强度,因此在识别和分类后续情绪具有非常重要的作用。为此,本研究提出了新的基于光流法的峰值帧定位方法,并利用CASME和CASME II两个大样本的特点,实现数据集中峰值帧的准确获取。

利用照相机对移动目标进行解帧后,可以清楚地看到目标的运动幅度、方向以及移动的位置,并利用光流对其进行刻画。根据光流场的基本原理,提出了新的思路:光流刻画了一段时期的多幅影像间的相关关系,由此计算出单个影像间目标的移动状态,即像素点在影像中的瞬间灰度变化速率称为“光流矢”。同时,物体在同一幅影像上的运动信号往往被分解后的影像灰度分布反映出来,因此,在同一空间范围内,物体的运动特征也会被转化为光流,并被用RGB色域表达出来。在此基础上,根据光流图生成的原理,得出面部表情幅值愈大,则其颜色愈鲜明,从而使每一帧图像与第一帧图像构成光流图,从而构成颜色最显著的单个画面即为峰值帧图像。

### (三) OFCNN 微表情识别方法体系结构

卷积神经网络在对图像进行分析时具有很大的优越性,它能实现对图像的空间特性的自动抽取,同时具有计算速度快的特点。基于此,本文提出基于小波变换的人脸微表情识别算法。针对目前已建立的卷积神经网络无法有效地抽取面部微表情的特点,拟构造基于深度神经网络的面部微表情精细特征抽取方法。但是,单凭复杂的卷积神经网络无法实现对微表情时空特性的同步检测。由于神经网络中的输入数据往往是单张图片,因此只利用卷积神经网络不能提取出最关键的连续特征,而造成短时特征的损失,而时间特性的获取正是光流法的强项。

稠密光流法Farneback强调将微表情的时间特性用光流图形式表达出来。为此,采用了Farneback方法,在卷积神经网络进行特征抽取前,保证了时间维度和空间维度的一致性,保证两个维度中的动态信息不丢失,实现对多个尺度的连续微表情图像的融合。

## 四、结束语

微表情识别在很多领域应用广泛,其技术的更新和发展将会给人民群众带来便利。尽管微表情在观察与采集方面有一定的局限,不过相信在科学技术的助力下,加上相关设备的更新,微表情识别技术将会更加成熟。

### 参考文献:

- [1] 赖振意,陈人和,钱育蓉.结合空洞卷积的CNN实时微表情识别算法[J].计算机应用研究,2020,37(11):1-5.
- [2] 吴进,闵育,马思敏,张伟华.一种基于CNN与LSTM结合的微表情识别算法[J].电讯技术,2020,60(01):1-7.
- [3] 孔慧芳,钱世超,闫嘉鹏.基于不均衡数据与迁移学习的面部微表情识别[J].合肥工业大学学报(自然科学版),2020,43(07):895-900.
- [4] 钱泽锋,钱梦莹.基于改进特征融合的微表情识别方法[J].软件工程,2021,24(04):26-29+14.
- [5] 吴进,闵育,李聪,等.一种基于3D-CNN的微表情识别算法[J].电讯技术,2019,59(10):1115-1120.