

基于数据增强的人脸表情识别方法研究

□ 杨兰兰¹ 高铭宇¹ 王晨宁² 冯冬杰³ 吕欣芮³

摘要: 深度学习在表情识别中的应用, 大多基于 VGGNet、GoogLeNet 与 ResNet 网络模型, 其核心结构均为深度卷积神经网络 (Deep convolution neural networks, DCNN)。本文针对现有的传统静态人脸表情识别, 对模型训练参数调节以及内部架构有很强的的依赖性这一问题, 拟采用数据增强方法, 在数据预处理层面实现对面脸表情识别的效果的提升以及模型运算效率的提升; 从而降低对于硬件设备的要求, 提升模型的适用性。

关键词: 人脸识别; 深度卷积神经网络; 数据增强

随着机器学习和深度神经网络的迅速发展以及智能设备的普及, 人脸识别技术正在经历前所未有的发展, 目前人脸识别精度已经超过人眼。人脸表情识别 (facial expression recognition, FER) 作为人脸识别技术中的一个重要组成部分, 近年来在人机交互、安全、机器人制造、自动化、医疗、通信和驾驶领域得到了广泛的关注, 成为学术界和工业界的研究热点。

人脸表情是传播人类情感信息与协调人际关系的重要方式, 据心理学家 A. Mehrabia 的研究表明, 在人类的日常交流中, 通过语言传递的信息仅占信息总量的 7%, 而通过人脸表情传递的信息却达到信息总量的 55%。面部表情是面部肌肉的一个或多个动作或状态的结果, 表达了个体对观察者的情绪状态。除了常见的高兴、吃惊、悲伤、愤怒、厌恶和恐惧 6 种表情, 还有惊喜 (高兴+吃惊)、悲愤 (悲伤+愤怒) 等 15 种可被区分的复合表情。一些表情可以准确解释, 一些表情则难以解释^[1]。

一、研究的现状与分析

面部表情的研究始于 19 世纪, 1872 年, 达尔文在他著名的论著《人类和动物的表情 (The Expression of the Emotions in Animals and Man, 1872)》中就阐述了人的面部表情和动物的面部表情之间的联系和区别。

深度学习在图像识别领域取得的巨大成绩为表情识别提供了新的思路。许多研究者开始将深度神经网络 (Deep neural network, DNN) 用于表情识别。深度学习的出现打破了表情识别中传统的先特征提取后模式识别的固定模式, 可以同时进行特征提取与表情分类, 而且深度学习对特征的提取是通过反向传播算法与误差优化对权值进行迭代更新, 所以能够提取出人类预想不到的关键点和特征。深度学习在表情识别中的应用, 大多基于 VGGNet、GoogLeNet 与 ResNet 网络模型, 其核心结构均为深度卷积神经网络 (Deep convolution neural networks, DCNN)。

在此基础上, Mollahosseini 等将 AlexNet 与 GoogleNet 模型结合, 构建了一个 7 层 CNN 用于人脸表情识别, 得到了较好的识别效果。Jung 等提出利用彩色图像与人脸特征点联合训练 CNN 网络, 其网络包括 3 个卷积层与 2 个隐藏层, 但这使得其表情特征学习不够精确。2014 年 Lopes 等将部分特定特征提取方法与卷积神经网络进行组合, 先采用预处理技术提取部分特定表情特征, 但这会导致未知环境下正确率不高、鲁棒性不足。2015 年 Verma 等提出了一个具有视觉和面部标识分支的网络, 其视觉分支负责图像序列的输入, 并引入从低层到高层的跳转连接, 以此考虑到底层特征的重要性; 面部标识分支研究面部标记轨迹, 考虑到因面部大动作造成的面部特征的变化, 如眼睛、鼻子、嘴唇的动作变化等, 该方法在 CK+ 数据集

上取得了较高的识别率。除了对基础模型及网络结构进行改进外, 有些研究者还对损失函数进行了研究与改进。2019 年, 吴慧华等对 AM-softmax 损失函数与 island loss 损失函数进行结合, 提出了一种基于余弦距离的损失函数 co-sine distance loss, 用于对 CNN 学习强区分度表情特征进行监督, 取得了较高的识别准确率。2018 年, Qiang 等提出了一种全新的损失函数 Knot magnify (KM) loss, 通过自适应的方式对训练过程中的每个样本添加一个权重, 从而提升强样本在 DCNNs 中训练的效果。

面部表情识别技术是近几十年来才逐渐发展起来的, 由于面部表情的多样性和复杂性, 并且涉及生理学及心理学, 表情识别具有较大的难度, 因此, 与其它生物识别技术如指纹识别、虹膜识别、人脸识别等相比, 发展相对较慢, 应用还不广泛。但是表情识别对于人机交互却有重要的价值, 因此国内外很多研究机构及学者致力于这方面的研究, 并已经取得了一定的成果。表情识别方法分类大致分为 4 种情况: 基于模板的匹配方法; 基于神经网络的方法; 基于概率模型的方法; 基于支持向量机的方法。

现如今, “深度学习”在图像识别领域取得的巨大成绩为表情识别提供了新的思路。许多研究者开始将深度神经网络 (Deep neural network, DNN) 用于表情识别。深度学习的出现打破了表情识别中传统的先特征提取后模式识别的固定模式, 可以同时进行特征提取与表情分类, 而且深度学习对特征的提取是通过反向传播算法与误差优化对权值进行迭代更新, 所以能够提取出人类预想不到的关键点和特征。

二、人脸识别的理论基础

(一) 目标检测

在计算机领域, 目标检测是一个很关键的技术。它的工作是在图像或视频中识别感兴趣的对象, 同时将对象的位置检测出来, 所以广泛应用在各个领域。

目前目标检测领域的深度学习算法主要分为一阶段 one-stage 与两阶段 two-stage 两类^[2]。如果输入一张图片, 要求将图片中所包含的对象、每个对象的位置输出, 这可以看作两个任务: 首先要找到图片中存在对象的区域, 接着判断出该区域中具体是什么对象。近几年 CNN 卷积神经网络的各种方法在对象识别方面已经达到了很不错的成绩, 所以我们最想要解决的问题就是, 对象在哪里。最简单的办法就是遍历图片中所有可能的位置, 地毯式地搜索不同大小, 不同高宽比以及不同位置的每个区域, 逐一进行检测, 最后输出概率最大的结果作为结果。当然, 这种方法的效率并不高。

RCNN 开创性地提出了候选区 (Region Proposals) 的方法, 先从图片中搜寻一些可能存在对象的候选区 (Selective Search), 然后识别出每个候选区中的对象, 这大幅度提高了目标检测的效率。但是总的来说, RCNN 系列算法仍然是两个阶段的识别模式: 先提取候选区, 再去识别候选区中的对象。

YOLO (You Only Look Once) 是一种基于深度学习的随想识别和定位算法, 它创造性地将传统目标检测的两个阶段合二为一, 所以是在检测速度上有了很大的提升。目标检测的本质, 就是识别与回归。而在处理图像的过程中, 用到最多的就是卷积神经网络 CNN, 所以, YOLOv3 从本质上来说, 就是一个能够实现回归功能的深度卷积神经网络。这种方法不仅检测速度非常快, 而且在此基础上还提高了预测的精度, 特别是增强了对

于小物体目标的识别能力。

(二) 图像特征提取

图像特征提取指的是通过计算机对输入图像的信息进行提取, 决定每个图像的点是否属于一个图像特征。通过特征提取, 图像上的点被分成了不同的子集, 这些子集通常是孤立的点、连续的曲线以及连续的区域^[3]。YOLOv3 检测模型与目标检测模型对图像特征提取的方法不同, FAST R-CNN 利用候选区域对图像进行特征提取, 而 YOLOv3 可以对图片的全局区域进行特征提取, 这样不仅使训练出的模型进行目标检测的速度得到提升, 还能够更好的对图片中的目标和背景进行区分。但是在对于大物体进行目标检测的过程中, 由于背景情况是复杂多变的, 所以背景也有可能被算进大物体目标的一部分。因此 YOLOv3 算法对于中小物体的识别效果较大物体来说会好一些。

三、研究的内容

通过在 FER-2013 和 IMDB 这两种数据集分别随机选取 100 张人脸表情数据作为人脸表情模型的训练和测试的数据, 采用高斯滤波平滑和多尺寸缩放这两种方法对人脸表情数据进行数据增强处理, 同时对原有模型进行 finetune 处理。

(一) 训练及测试数据集

1. FER-2013

该数据集来源于 Kaggle, 由 48×48 像素的面部灰度图像组成。面部已自动注册, 因此面部或多或少居中, 并且在每个图像中占据大约相同的空间量。任务是根据面部表情中显示的情感将每个面孔分为以下七个类别之一 (0 = 愤怒, 1 = 厌恶, 2 = 恐惧, 3 = 快乐, 4 = 悲伤, 5 = 惊奇, 6 = 中性)。训练集包含 28,709 个示例, 用于排行榜的公共测试集包含 3,589 个示例。

2. IMDB

从 IMDB 网站上列出了最受欢迎的 100,000 名演员的名单, 并 (自动) 从他们的个人资料的出生日期, 姓名, 性别和与该人相关的所有图片中抓取了这些演员。总共从 IMDB 的 20,284 名名人和 Wikipedia 的 62,328 名名人中获得了 460,723 张脸部图像, 因此总共获得了 523,051 张。

(二) 对人脸表情数据进行预处理

(1) 降噪层面: 采用高斯滤波, 对人脸表情数据进行一定的平滑处理, 使分布在原始人脸表情数据的噪声点模糊化, 降低非表情像素的干扰, 以及背景噪声对神经网络模型的扰动。

(上接第 125 页)

等仓内环境进行智能监控。在不产生污染的时间段进行清仓处理。对污染垃圾填满, 可回收垃圾送到各地再加工, 对厨余垃圾利用堆肥技术进行预处理等。立体仓能有效节约垃圾存放空间, 并对垃圾分类起到积极作用。

(四) 提高公众垃圾分类意识的宣传教育措施——“区长制”

垃圾分类宣传教育对学生学校教育效果明显, 成年人长期形成的习惯很难改变。社区宣传和网络推送的宣传效果不明显。本研究提出“区长制”。通过社区, 党员, 保洁员和公众的力量, 将垃圾分类理念和实践落到实处。“区长制”分为三级, 一级“区长”由社区干部担任, 每人划格包干一个责任区, 对网格责任区内的垃圾分类工作实行总体的监督和指导; 二级“区长”由先进党员和热心居民组成, 以每人结对十至二十户家庭的方式指导结对家庭开展垃圾分类; 三级“区长”则以保洁员为主。每天轮流上岗, 在进行垃圾收集和清运工作时, 检查垃圾分类收集的情况, 免费发放对应不同垃圾种类的垃圾袋等等。将垃圾分类意识潜移默化到人们的日常生活中。

(五) 激励机制为垃圾分类增加动力

在“区长制”基础上, 增加奖励机制, 对入户问答考查合格的, 垃圾分类做的好的给予奖励, 可以建立垃圾分类信息管理系统, 对居民垃圾分类进行积分制度。表现好的, 加积分

高斯滤波平滑处理算法公式:

其中 G 为高斯核函数。

高斯滤波步骤:

- 移动相关核的中心元素, 使它位于输入图像待处理像素的正上方
- 将输入图像的像素值作为权重, 乘以相关核
- 将上面各步得到的结果相加做为输出

(2) 感受野层面: 大尺寸的人脸表情数据对模型的算力以及识别度都有一定的扰动, 缩小原始人脸表情数据尺寸, 可以减小感受野移动次数, 在模型识别阶段提升语义信息丰富度, 使模型对人脸表情特征信息的感知能力有所提升。尺寸缩放公式:

其中为缩放后的图像尺寸, 为原始图像尺寸, 为图像尺寸缩放系数。

(三) 模型的选取

在 Octavio Arriaga 等人的 Face Classification and Detection (FCD) 模型的基础上进行 finetune, 受 ResNet v2 启发, 在 shortcut 分支上加入一组 BN-ReLU 激活层。

四、实验结果

(1) 通过高斯滤波平滑处理, 在模型算力层面进行一定的优化处理, 使模型注意力更好的集中在人脸区域, 从而提升人脸表情识别的置信度。

(2) 通过对人脸表情数据的多尺寸的缩放, 使模型的感受野在其上面的滑动次数减少, 降低模型的算力, 从而提高表情的识别能力。

(3) 通过高斯滤波平滑处理和多尺寸缩放对 FER-2013 和 IMDB 数据集分别随机抽取的 100 张人脸表情数据进行数据预处理, 同时与原始数据进行对比实验, 从而验证上述两种数据增强方法对人脸表情数据的识别度较原始数据有一定的提升效果。

注释

- [1] 人脸表情识别研究技术综述 有三 AI 学院 知乎。
- [2] 陶磊, 洪韬, 钞旭. 基于 YOLOv3 的无人机识别与定位追踪 [J]. 工程科学学报, 2020, 42 (0): 463-468.
- [3] 周志华. 机器学习 [M]. 北京: 清华大学出版社, 2015: 114-115.

(作者单位: 1. 河南师范大学计算机与信息工程学院; 2. 河南师范大学电子与电气工程学院; 3. 河南师范大学商学院)

奖励, 做的不好的减积分。积分可以兑换生活用品或一些旅游景点减门票费等。激励措施对中老年人有效性大, 对垃圾分类起到推动力作用。

四、创新和价值

在市民素质和习惯不会在垃圾分类日常中变成自觉性行为的现状下, 本文提出了增强垃圾分类时效性的智能化手段。对智能化手段中的智能垃圾桶创新提出技术性和成本不高的个人创新方案。提出的“区长制”和激励措施对市民习惯养成起到促进作用, 对垃圾分类起到推动力作用。

参考文献

- [1] 龚文娟. 城市生活垃圾治理政策变迁 [J]. 学习与探讨, 2020 (2): 28-29.
- [2] 《习近平对垃圾分类工作作出重要批示》, 新华网, 2019 年 6 月 3 日, <http://www.xinhuanet.com/politics/leaders/2019-06/03/c-1124577181.htm>.
- [3] 陈子, 赵静, 马国强. 我国城市生活垃圾分类政策实施研究——以南京市为例 [J]. 地域研究与开发, 2016 (3): 64-68.

(作者单位: 漯河职业技术学院电气电子工程系)

作者简介: 陈亚丽 (1981 ~), 女, 硕士, 讲师, 研究方向为电气控制技术、电磁功能材料。

杨伟兵 (1983 ~), 男, 本科, 讲师, 研究方向为制冷技术、电气控制。