

基于 DenseNet 的人脸表情识别方法研究*

顾状状¹ 许学斌¹ 路龙宾¹ 豆阳光²

(1. 西安邮电大学计算机学院 西安 710000)(2. 长安大学地质工程与测绘学院 西安 710054)

摘 要 在人脸识别方向,文章提出了一种基于稠密卷积神经网络的人脸表情识别方法。该方法通过特征复用与旁路连接的策略,实现低计算资源的高效特征表达,从而达到提高表情识别系统准确率的目的。文中提出的人脸表情识别方法,使用 GPU 运算优化模型,最终在拓展的 KDEF 人脸表情数据集上取得了 96.88% 的准确率,优于目前已公开的大部分人脸表情识别方法。文章利用预先训练好的人脸表情识别模型,设计出表情识别原型软件,通过使用摄像头实时捕捉人脸表情进行识别,能够取得很好的识别效果。

关键词 稠密卷积神经网络;人脸表情识别;深度学习;图像分类

中图分类号 TP181 **DOI**:10.3969/j.issn.1672-9722.2023.10.039

Research on Human Facial Expression Recognition Method Based on DenseNet

GU Zhuangzhuang¹ XU Xuebin¹ LU Longbin¹ DOU Yangguang²

(1. School of Computer Science & Technology, Xi'an University of Posts & Telecommunications, Xi'an 710000)

(2. College of Geological Engineering and Geomatics, Chang'an University, Xi'an 710054)

Abstract In the face recognition field, this paper proposes a human facial expression recognition method based on dense convolutional neural network. The method achieves efficient feature expression with low computing resources through feature reuse and bypass connection strategy, so as to improve the accuracy of facial expression recognition system. The human facial expression recognition method proposed in this paper uses GPU computing to continuously research and optimize the model, and finally it achieves 96.88% accuracy on the human facial expression database which is composed with KDEF dataset and the human facial expression dataset produced in the article, which is better than the most of the current published human facial expression recognition methods. This paper uses the pre-trained facial expression recognition model to design the prototype software of facial expression recognition, which can achieve good recognition accuracy by using the camera to capture facial expression in real time.

Key Words DenseNet, facial expression recognition, deep learning, image classification

Class Number TP181

1 引言

近年来,随着计算机科学技术的迅猛发展,机器学习以及深度学习技术伴随着智能终端的推广成为了一个极其热门的研究方向,推动了人脸识别技术的飞速发展。目前人脸识别^[1]技术的识别精度已经远超人类所能识别的范围,使得其在多个领域占据着不可或缺的一席之地。人脸表情识别作为人脸识别技术的一个重要分支:其在水机交互、

安全、机器人制造、自动化、医疗、通信和驾驶等领域备受关注^[2],且在近十年里也取得了由量到质的突破,成为国际前沿研究领域的热门领域之一。

人脸表情的研究,起源于心理学。1972 年达尔文就从心理实验中得出结论,并在他的著名论著《The Expression of the Emotions in Animals and Man》中阐述了人脸表情和动物面部表情的联系和区别^[3]。在 1978 年,Ekman 和 Friesen 制定了 FACS (Facial Action Coding System) 系统,该系统主要用

* 收稿日期:2023 年 4 月 10 日,修回日期:2023 年 5 月 16 日

作者简介:顾状状,男,研究方向:计算机视觉,机器学习。许学斌,男,博士,副研究员,研究方向:人工智能,生物信息学等。路龙宾,男,博士,讲师,研究方向:人工智能,机器学习等。豆阳光,男,硕士研究生,研究方向:信息提取,深度学习。

来检测面部的变化,并定义了六种基本表情,即:伤心、开心、厌恶、害怕、愤怒和惊讶^[4]。计算机领域对人脸表情的研究最早是在1978年,Suwa和Sugie等人对人脸表情图像中面部的20个点进行追踪进行表情识别,提出了在图片流中对人脸表情进行自动分析^[5]。1981年,通过仿生学方法建立了人脸表情模型^[6]。20世纪90年代初,K.Mase和A.Pentland提出了光流法,该方法很快被应用到人脸表情识别研究中^[7]。Rashid T A提出了一种使用不同的面部表情不平衡数据集来识别面部表情的技术^[8]。Fathallah等提出基于CNN的用于面部表情识别的新架构网络,并使用视觉几何组模型(VGG)微调了架构以改进结果,精度达到了71.04%^[9]。Cheng S等提出一种改进的视觉几何组(VGG)深度卷积神经网络(CNN)的表情识别模型,改进后的VGG-19网络模型在人脸表情识别上可以达到96%的准确率^[10]。目前为止人脸表情识别研究已历经了半个世纪,然而它的热度不降反升,近年来更是伴随着神经网络技术的发展成为了一个热门的研究课题。

本文提出了一种基于稠密卷积神经网络的人脸表情识别方法,通过特征复用与旁路连接的方式,实现人脸表情特征提取,在扩展的KDEF数据集数据集上训练卷积神经网络,对面部表情进行分类,精度达到了96.88%,并利用预先训练的网络模型开发出一个可以实时检测人脸表情的操作软件。本文不仅在人脸表情识别网络的学术领域取得了一定的成果,也对社会生产实践有一定的指导意义。

2 人脸表情识别方法

2.1 人脸表情识别过程

人脸表情^[11]是传递人类情绪信息和协调人际关系的有效手段。人脸表情是由面部多处肌肉(如,眼部,口部等)的变化组合来表达人类自身的各种心理状态。人类的面部表情至少包含21种,1971年,心理学家Ekman与Friesen首次提出人类的6种主要情绪状态,分别是:生气(Anger)、开心(Happiness)、伤心(Sadness)、惊讶(Surprise)、厌恶(Disgust)和害怕(Afraid)^[4]。本文所研究的人脸表情即是在此六种基本表情基础上加入了中立(Neutrality)表情。

人脸表情识别主要基于三个步骤组成:人脸表情图像获取、图像特征提取、特征分类^[12],如图1所示。

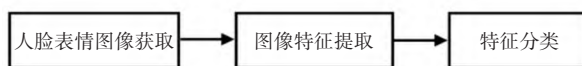


图1 人脸识别过程

2.2 DenseNet网络模型

从CNN开始不断地加深网络,梯度消失和模型退化问题就一直存在,其中批量规范化使得梯度消失问题得到一定程度的缓解,ResNet^[11]的出现进一步减少梯度消失和模型退化问题的发生。DenseNet^[13]的优势在于其通过特征重用和旁路设置,使得网络的性能得到大幅度提升。相比于ResNet,减少了可训练参数量,加强了特征的传递,更加有效地利用特征,减轻了梯度消失和模型退化问题。

ResNet将每一层和前面的某一层短路连接起来,在一定范围内减轻了该问题的发生,而DenseNet做了一个更加激进的紧密连接机制:将每一层网络都和前面所有的层连接起来(如图2),即在DenseNet网络的第 n 层有 $(n-1)$ 个连接与前面的所有层。这种紧密连接机制,避免了特征信息在层之间传递导致的梯度消失问题。

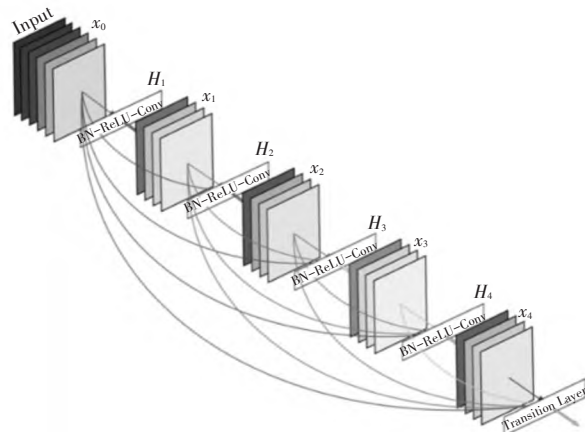


图2 DenseNet短路连接机制

2.3 基于DenseNet预训练网络的人脸表情识别方法

2.3.1 人脸表情识别数据集

本文所采用的数据集由KDEF和手动制作的人脸数据集组成。

KDEF^[14]数据集包含了70个年龄位于20岁~30岁之间的成年人,由35个男性,35个女性从5个不同的角度:正前方,左前方,左侧,右前方,右侧,拍摄了7种表情:生气、开心、伤心、惊讶、厌恶、害怕和中立,每个表情拍摄两次,总共4900张人脸表情图片。图片为562*762像素的RGB图像,图3所示为KDEF中的部分表情图片。



图3 KDEF 部分图片示例

手工制作的人脸表情数据集同样包含 7 种表情,每种表情包含 10 张不同的图片,共 70 张,图片为 420*620 像素的 RGB 彩色图像,部分图片如图 4 所示。

训练集包含 KDEF 数据集中的 3780 张图片,以及手工制作数据集中的 56 张图片,验证集包括 KDEF 数据集中的 1120 张图片,手工制作数据集中的 14 张图片。



图4 手工制作数据集部分示例

2.3.2 数据处理

归一化处理^[15]可以提升训练模型的收敛速度、训练精度以及防止梯度爆炸。文章采用了最大-最小标准化方法对输入的图像数据进行归一化处理,数学定义式为

$$x^* = \frac{x - \min}{\max - \min} \tag{1}$$

式中: x 为数据灰度值, $\max$ 为图像灰度最大值, $\min$ 为图像灰度最小值。

文中采用了随机上下、左右翻转和随机裁剪对图像进行数据增强,对图像进行随机上下左右翻转和随机裁剪,大大扩充了样本量,训练出来的模型泛化能力更强^[16]。

2.3.3 网络模型

本文采取的人脸表情识别方法是基于 DenseNet 预训练网络搭建的。为了减轻梯度消失和梯度爆炸问题,在模型中用添加 BatchNormalization 层做批量正则化处理,选择 ReLU 作为激活函数,可以减轻梯度消失和梯度爆炸问题,而且计算量小,收敛迅速。添加一层 GlobalAveragePooling,对特征信息的高度和宽度两个维度进行平均化求值。

两层 Dense 做分类器,第一层 Dense 有 1024 个

隐藏单元,使用 ReLU 作为激活函数,最后一层输出层 Dense 隐藏单元数为 7,使用 softmax 作为激活函数。

网络模型如图 5 所示,该模型共有 8098375 个参数,其中可训练参数为 8012679 个。

2.4 编写实时人脸表情识别系统界面

文章编写了一个系统,来实时检测人脸表情。将训练好的模型加载进来,调用计算机摄像头,在视频流中获取到每一帧的图片,然后调用 OpenCV 提供的人脸识别分类器,找出人脸所在的坐标,并记录下来,按下空格键,系统则会先保存下当前图像,将图像按照人脸识别出来的坐标裁剪出合适的图片,用于人脸表情识别。此时载入人脸表情识别模型,对图片内容进行识别,并识别内容返回,使用 OpenCV 在显示图像的每一帧时将上一次的识别结果放在图片左上角,按下 Q 键则退出系统,软件操作界面如图 6 所示。

Model: "sequential"

Layer (type)	Output Shape	Param #
densenet121 (Model)	(None, None, None, 1024)	7037504
batch_normalization (Batch Normalization)	(None, None, None, 1024)	4096
activation (Activation)	(None, None, None, 1024)	0
global_average_pooling2d (Global Average Pooling)	(None, 1024)	0
dense (Dense)	(None, 1024)	1049600
dense_1 (Dense)	(None, 7)	7175

Total params: 8,098,375
Trainable params: 8,012,679
Non-trainable params: 85,696

图5 人脸表情识别方法整体结构



图6 人脸表情识别系统软件界面

3 实验结果分析

3.1 优化模型

经过测试 512*512、256*256、224*224、128*128 不同大小尺寸的图像输入网络后的结果,如表 1 所示,可以看出当输入图像尺寸为 256*256 和 224*224 的时候,都取得了 93% 的较高准确率,但是大小为 224*224 像素时每次迭代训练速度更快,模型收敛速度也略快,所以最终采用图像的输入尺寸

为 224*224 像素。

表 1 不同尺寸的图像输入对准确率的影响

大小	准确率/%
562*762	14
512*512	90
256*256	93
224*224	93
128*128	90

3.2 结果分析以及模型评估

文章建立的数据集对模型进行 5000 次迭代的训练,模型的准确率随着训练进行产生的变化如图 7 所示,图中红色的线表示训练集上的准确率情况,蓝色的线表示验证集上的准确率情况,横轴表示迭代的次数,纵轴表示准确率。在进行第 2085 次迭代之后,模型的准确率达到 5000 次训练的最大值 96.88%。

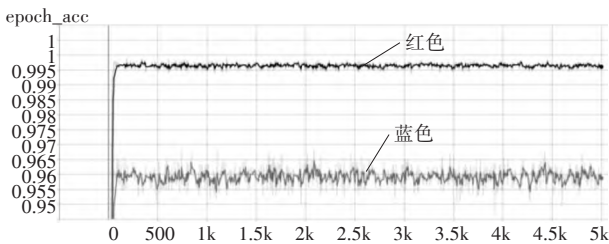


图 7 模型准确率变化情况

除了训练集和验证集的准确率和损失,本文还运用了其他评估方法对模型进行了评估。

1) 计算模型在不同人脸表情上的识别精度,如表 2 所示。

表 2 不同表情上的识别准确率精度

表情类别	识别准确率/%
生气 (Anger)	96.93
开心 (Happiness)	100
伤心 (Sadness)	95.71
厌恶 (Disgust)	91.41
害怕 (Afraid)	94.48
中立 (Neutrality)	97.55
惊讶 (Surprise)	98.16

2) 计算混淆矩阵。本文中的人脸表情识别模型的混淆矩阵^[17]如表 3 所示。引入了五个分类精度评价指标,分别为准确率 (Accuracy)、精确率 (Precision)、召回率 (Recall)、特异度 (Specificity)、F1-值 (F1-score)。其中,准确率是所有的预测正确的样本数占总预测样本个数的比重,数学公式为

$$ACC = \frac{TP}{TP + TN + FP + FN} \quad (2)$$

式中:TP 表示预测值为当前表情且预测正确,TN 则为预测值为非当前表情且预测正确,FP 表示真实值不是当前表情且预测错误,FN 则为真实值是当前表情且预测错误。

精确率表示预测值为当前表情且正确预测的样本数占预测正确的样本数的比重,精确率的数学公式:

$$PPV = \frac{TP}{TP + FP} \quad (3)$$

计算出每一个表情的精确率,如表 4。

召回率表示预测值为当前表情且正确预测的样本数占全部预测当前表情的样本数的比重,召回率的数学公式为

$$TPR = \frac{TP}{TP + FN} \quad (4)$$

每个表情的召回率如表 5。

特异度是错判的样本数占全部预测错误的样本数的比重,特异度的数学公式为

$$FPR = \frac{TN}{TN + FP} \quad (5)$$

计算结果如表 6 所示。

F1 Score 是 P 和 R 的调和平均数,F1 Score 的数学定义为

$$F1 \text{ Score} = \frac{2 * P * R}{P + R} \quad (6)$$

式中:P 代表精确率,R 代表召回率,所以每个表情的 F1 Score 值如表 7 所示。

表 3 混淆矩阵精度评估表

预测值	真实值						
	生气	开心	伤心	厌恶	害怕	中立	惊讶
生气	157	0	1	4	0	0	0
开心	0	163	0	0	0	0	0
伤心	0	0	157	6	0	3	0
厌恶	4	0	1	150	1	0	0
害怕	0	0	3	3	154	0	3
中立	2	0	1	0	0	159	0
惊讶	0	0	0	0	8	1	160

表 4 精确率的精度评估表

表情类别	精确率/%
生气 (Anger)	96.91
开心 (Happiness)	100
伤心 (Sadness)	94.58
厌恶 (Disgust)	96.15
害怕 (Afraid)	94.48
中立 (Neutrality)	98.15
惊讶 (Surprise)	94.67

表 5 召回率的精度评估表

表情类别	召回率/%
生气(Anger)	96.32
开心(Happiness)	100
伤心(Sadness)	96.32
厌恶(Disgust)	92.02
害怕(Afraid)	94.48
中立(Neutrality)	97.55
惊讶(Surprise)	98.16

表 6 特异度的精度评估表

表情类别	特异度/%
生气(Anger)	96.42
开心(Happiness)	95.81
伤心(Sadness)	96.42
厌恶(Disgust)	97.14
害怕(Afraid)	96.72
中立(Neutrality)	96.22
惊讶(Surprise)	96.11

表 7 F1 Score 值的精度评估表

表情类别	F1 Score/%
生气(Anger)	96.61
开心(Happiness)	100
伤心(Sadness)	95.44
厌恶(Disgust)	94.04
害怕(Afraid)	94.48
中立(Neutrality)	97.85
惊讶(Surprise)	96.38

3.3 对比其他人脸表情识别方法

本文提出的基于 DenseNet 预训练神经网络的人脸表情识别方法,相对于直接使用 DenseNet 有较大的优势,从表 8 中可以看出,DenseNet121 预训练网络模型在本文的数据集下最高准确率为 93.66%。

表 8 不同的网络模型的识别准确率

网络模型	识别准确率/%
本文提出的人脸表情识别方法	96.88
DenseNet121	93.66
MobileNet	90.54
ResNet121	85.62
PCA+KNN (K=1)	44.96
PCA+KNN (K=3)	44.79
PCA+KNN (K=5)	43.73
2DPCA	60.27
Gabor+2DPCA	82.51
NMF+SVM	83.17

使用文章中的数据集对 ResNet121、MobileNet 流行的预训练网络模型以及传统方法 PCA+KNN、2DPCA^[18]、Gabor+2DPCA、NFM+SVM 进行了对比试验。ResNet121 同样采取跳接的方式,取得了 85.62% 的准确率,但 DenseNet 更为激进的跳接方式在特征提取上取得了更好的效果。MobileNet 在体积上有着很大的优势,极大地减少了参数量,但与此同时损失了精度,准确率为 90.54%。传统机器学习方法:PCA+KNN、2DPCA 相对于卷积神经网络对图像特征提取能力较差,人脸图像识别准确率较低。Gabor+2DPCA 中 Gabor 类似于人体视觉细胞刺激响应方式,能够较好提取局部特征信息,取得了 82.51% 的准确率。NFM+SVM 能够快速学习到高阶线性组合图像特征,取得了较高的准确率 83.17%。综上,文章所提出的基于 DenseNet 的人脸表情识别方法在准确率上明显优于其他方法。

4 结语

本文提出的基于稠密卷积神经网络的人脸表情识别方法,最终取得了验证集上 96.88% 的准确率,与此同时设计出了人脸表情识别原型软件,利用已经训练好的模型实时检测人脸表情,取得了很好的效果。本文主要做出的贡献如下:

1) 提出基于稠密卷积神经网络的人脸表情识别方法,该模型能够有效提取人脸表情特征信息,识别率较高,收敛速度快。

2) 手工制作了人脸表情数据集并结合目前主流的公开数据集,测试了不同图像尺寸下模型的表现情况。

3) 设计出了人脸表情识别的原型软件,对人脸表情识别的普及推广具有重要的现实意义。

文中所提出的模型虽然已经取得了不错的识别效果,但是还需要继续增大训练数据样本,其普遍性无法推广到整个人类的表情数据,还需要经过更多的数据检验。其次,文中所提出的模型方法,需要提供拥有非常强大算力的计算机,训练参数达到了百万级,训练过程较长,在不降低模型性能的基础上,适当优化模型,减少模型训练参数,以减少计算量。

参考文献

- [1] Parkhi O. Deep Face Recognition[C]//BMVC 2015-Proceedings of the British Machine Vision Conference 2015, 2015.

- [2] WANG Huanhuan. The Applications of Facial Expression Recognition in Human-Computer Interaction [C]//2018 IEEE International Conference on Advanced Manufacturing (ICAM), 2018:288-291.
- [3] Darwin C. The Expression of The Emotions in Man and Animals (1872)[J]. The Portable Darwin, 1993:364-393.
- [4] Ekman P, Friesen WV. Facial action coding system[J]. Environmental Psychology & Nonverbal Behavior, 1978.
- [5] Srivastava R, Roy S. Utilizing 3D Flow of Points for Facial Expression Recognition[J]. Multimedia Tools & Applications, 2014, 71:1953-1974.
- [6] Platt SM. Animating Facial Expressions[C]// Proceedings of The 8th Annual Conference on Computer Graphics and Interactive Techniques, 1981:245-252.
- [7] Paleari M. Evidence Theory-Based Multimodal Emotion Recognition[C]// International Conference on Multimedia Modeling, 2009:435-446.
- [8] Rashid TA. Convolutional Neural Networks Based Method for Improving Facial Expression Recognition[J]. Intelligent Systems Technologies and Applications, 2016, 2016: 73-84.
- [9] Fathallah A. Facial Expression Recognition via Deep Learning[C]// 2017 IEEE/ACS 14th International Conference on Computer Systems and Applications (AICCSA), 2017:745-750.
- [10] Cheng, S Facial Expression Recognition Method Based on Improved VGG Convolutional Neural Network[J]. International Journal of Pattern Recognition and Artificial Intelligence, 2020, 34(07):2056003.
- [11] Bettadapura V. Face Expression Recognition and Analysis: The State of The Art [J]. ArXiv Preprint ArXiv: 1203.6722, 2012.
- [12] 陈幼. 基于 Gabor 小波与粒子群优化算法的人脸表情识别研究[D]. 武汉: 武汉理工大学, 2014.
- CHEN You. Research on Facial Expression Recognition Based on Gabor and Particle Swarm Optimization Algorithm [D]. Wuhan: Wuhan University of Technology, 2014.
- [13] Huang G. Densely Connected Convolutional Networks [C]//Proceedings of the IEEE conference on computer vision and pattern recognition, 2017:4700-4708.
- [14] HE Kaiming. Deep Residual Learning for Image Recognition[C]// Proceedings of The IEEE Conference on Computer Vision and Pattern Recognition, 2016: 770-778.
- [15] 卢官明, 朱海锐, 郝强, 等. 基于深度残差网络的人脸表情识别[J]. 数据采集与处理, 2019, 34(1):50-57.
- LU Guanming, ZHU Hairui, HAO Qiang, et al. Facial Expression Recognition Based on Deep Residual Network [J]. Journal of Data Acquisition and Processing, 2019, 34(1):50-57.
- [16] Wood A. Fashioning The Face: Sensorimotor Simulation Contributes to Facial Expression Recognition[J]. Trends in Cognitive Sciences, 2016, 20(3):227-240.
- [17] Sokol K, Cécile C. On Multi-class Learning through The Minimization of The Confusion Matrix Norm [J]. ArXiv Preprint ArXiv:1303.4015, 2013.
- [18] YANG Wankou. Sequential Row--Column 2DPCA for Face Recognition [J]. Neural Computing and Applications, 2012, 21:1729-1735.

版权声明

本刊已许可万方数据库、中国学术期刊(光盘版)电子杂志社在中国知网及其系列数据库等产品中以数字化方式复制、汇编、发行、信息网络传播本刊全文。作者向本刊提交文章发表的行为即视为同意我编辑部上述声明。

《计算机与数字工程》编辑部