

文章编号: 2095-2163(2021)06-0043-08

中图分类号: TP37

文献标志码: A

人脸表情识别研究进展

姜月武, 路东生, 党良慧, 杨永兆, 施建新

(上海工程技术大学 电子电气工程学院, 上海 201620)

摘要: 随着计算机视觉的发展和人工智能产业的兴起,人脸表情识别技术在人工智能产业有着广泛的应用需求。人脸表情识别在传统机器学习算法下对环境及姿态的改变不具备良好的鲁棒性,而且识别精度也达不到实际应用的要求。计算机和图像处理等硬件性能的提升,以大数据为核心的深度学习算法得到快速发展,人脸表情识别技术开始趋于在深度学习算法上研究。本文分别对人脸表情图像预处理、特征提取、特征分类3个关键技术进行介绍,具体叙述了从传统的机器学习到基于深度学习的人脸表情识别技术的研究进展,分析了人脸表情识别技术目前面临的挑战和发展趋势。

关键词: 人脸表情识别;深度学习;特征提取;特征分类

Research progress of facial expression recognition

JIANG Yuewu, LU Dongsheng, DANG Lianghui, YANG Yongzhao, SHI Jianxin

(School of Electronic and Electrical Engineering, Shanghai University of Engineering Science, Shanghai 201620, China)

[Abstract] With the development of computer vision and the rise of artificial intelligence, facial expression recognition (FER) technology has a wide range of applications in the artificial intelligence industry. FER under the traditional machine learning algorithm does not have good robustness to the change of environment and posture. Moreover, the accuracy of identification is not up to the requirement of practical application. With the improvement of hardware performance such as GPU, deep learning algorithm with big data as its core has been developed rapidly. FER technology tends to be studied in deep learning algorithm. This paper introduces three key technologies of facial expression image preprocessing, feature extraction and feature classification. The research progress from traditional machine learning to FER based on deep learning is described, and the current challenges and development trend of facial expression recognition technology are analyzed.

[Key words] facial expression recognition; deep learning; feature extraction; feature classification

0 引言

人脸表情是肢体语言的一种,肢体语言在人际交流中具有表达情感的作用。人脸表情识别是一个多学科交叉研究的热点,主要有心理学、生物学、计算机科学等。早在20世纪70年代,Ekman和Friesen就在跨文化研究结论的基础上定义了6种基本情绪:愤怒、厌恶、恐惧、高兴、悲伤和惊讶,而这项研究表明人类对某些基本情绪的感知方式是相同的^[1]。还有其它情感描述模型如面部动作编码系统(FACS),其可以进行更宽泛和多种类的人脸表情刻画和分类。人脸识别技术的早期探索始于SUWA M等通过在连续的人脸图像序列上进行特征点的标记,再去跟踪识别这些特征点的变化与原模型进行比较得到表情信息^[2]。基于传统机器学习的人脸表情识别技术依赖于人工特征。局部二值算法(LBP)、Gabor特征提取、主动形状模型(ASM)等是机器学习算法常用提取特征的方法,这些方法常常用于

小数据集。随着计算机硬件的提升和高性能GPU的出现,深度学习网络重新成为了热点,2012年基于CNN搭建的AlexNet模型的出现,表情识别的准确率实现了大幅提升,远超传统机器学习的准确率^[3]。

本文对人脸表情识别在机器学习和深度学习上国内外的研究进展进行综述,首先介绍常用的人脸表情数据集;而后分别就机器学习与深度学习在人脸表情识别的主要算法分别进行了总结;最后展望未来表情识别技术的挑战与机遇,对表情识别技术的未来发展趋势进行了分析。

1 面部表情数据集

常用的人脸表情数据集有CK+数据集、MMI数据集、JAFPE数据集、BHU数据集、FER2013数据集、AFEW数据集等。FER2013包含的7种表情,如图1所示。本文根据数据集的属性,从数据集数据样本、收集环境及方式和表情分布进行了总结,见表1。

基金项目: 上海市自然科学基金(17ZR1411900);上海市科委重点项目(18511101600)。

作者简介: 姜月武(1995-),男,硕士研究生,主要研究方向:人脸表情识别、深度学习、计算机视觉。

收稿日期: 2020-12-02



图 1 FER2013 数据集包含的 7 种基本表情

Fig. 1 The seven basic emoticons of FER2013 dataset

表 1 表情数据集及简介

Tab. 1 Expression data set and introduction

Datasets	Data Sample	Collect Environmental	Expression Distribution
<i>CK +</i>	593 image sequence	Lab (front+induce+spontaneous)	7 kinds of expressions (6 basic+1 neutral)
<i>MMI</i>	740 images and 2 900 videos	Lab (induce)	7 kinds of expressions (6 basic+1 neutral)
<i>JAFFE</i>	213 images	Lab (induce)	7 kinds of expressions (6 basic+1 neutral)
<i>FER2013</i>	35 887 images	web (front+induce+spontaneous)	7 kinds of expressions (6 basic+1 neutral)
<i>AFEW</i>	1 809 videos	movie (induce+spontaneous)	7 kinds of expressions (6 basic+1 neutral)
<i>SFEW</i>	1 766 images	movie (induce+spontaneous)	7 kinds of expressions (6 basic+1 neutral)
<i>RaFD</i>	1 608 images	Lab (induce)	8 kinds of expressions (6 basic +1 neutral +1 contempt)
<i>RAF - DB</i>	29 672 images	web (induce+spontaneous)	7 kinds of expressions (6 basic+1 neutral)+12 mixed expressions

2 基于机器学习的人脸表情识别

人脸表情识别是图像识别的一种,机器学习算法一直是图像识别任务的常用算法,具有较高的识别精度。基于机器学习的人脸表情识别算法的主要处理步骤包含:图像预处理、特征提取和表情分类。

2.1 图像预处理

图像预处理的 2 个主要作用:一是去除与人脸表情无关的成分;二是削弱对特征有影响的因素。去除与人脸无关的成分,包含去除背景、截取只含有人脸的图片。一幅图像除图像信息外还包含噪声信息,这些噪声信息会对特征提取造成影响。由于光照的波动影响特征的提取率,人脸表情特征提取的纹理特征比较容易受到不平衡光照强度的影响。通过光照归一化可有效的去除了光照的影响^[4]。还有一些处理涉及到图像的大小问题,只需要考虑关键特征不丢失的情况,适当的减小图像尺寸。

2.2 图像的特征提取

特征提取在分类任务是核心问题,提取到能代表分类信息的有用特征关乎到人脸表情识别的准确度。下面列举一些常用的人脸表情特征提取算法:

局部二值算法(LBP)是一种描述图像局部纹理的算子,由于 LBP 具有良好的灰度不变性与旋转不

变性,根据这种特性 LBP 常被用来应用于纹理分类、人脸识别、手势识别等领域^[5]。王宪等根据基本 LBP 算子提取的特征缺乏完整性的特点,提出了分块的完备局部二值模式 (CLBP),先进行不同的算子分块的直方图统计特征的提取,然后将所有分块的 CLBP 直方图序列连接起来得到人脸的 CLBP 特征^[6];Jabid 提出基于局部方向模式 (LDP),通过计算边缘反应来增强算法的鲁棒性^[7]。

Gabor 小波的响应与人类视觉系统中简单细胞的视觉刺激响应非常相似,Gabor 小波对图像边缘具有良好的敏感性,同时对光照具有鲁棒性,Gabor 小波被广泛的应用于视觉处理领域^[8]。针对传统的 Gabor 滤波器特征冗余的缺点,改进了一种新颖的局部 Gabor 滤波器组,在人脸表情特征提取上有了很大的提升^[9];刘帅师等提出采用 Gabor 多方向特征融合与分块直方统计图相结合的方法来提取表情特征,能够有效地表征图像的全局特征^[10]。

Haar-like 特征定义了 4 种由黑色区域和白色区域组成的基本窗口,这些窗口在图像上滑动,遍历整个图像,黑色区域像素之和减去白色区域像素之和的差就是特征的一个维度;Lienhart R 等在 4 个特征基础上进行了扩展,扩展为 14 个,增加了旋转性,能够提取到更为丰富的边缘信息^[11];胡敏等提出 Haar-like 特征与直方图加权的方法,使得局部特征

描述的更加充分^[12]。

光流法是基于光流的运动目标检测算法,应用于行人识别、目标检测。光流法在人脸表情识别中被应用于基于序列的表情识别。吴新根等提出了一种改进的瞬时位置速度估计方法,克服了光流算法的漂移问题,也加快了收敛速度^[13]。

人脸表情特征提取是表情识别的关键问题,符合任务场景的特征有利于表情识别的完成,上述提取特征算法是人脸表情识别常用的算法,另外还有颜色特征、形状特征、HOG 特征等也被用来刻画人脸表情。

2.3 表情分类算法

特征提取后,通过分类器进行表情所属类别的判断。合适的分类器对表情分类起着重要作用,如何设计高性能的分类器也是表情识别任务的一个重要研究方向,人脸表情识别几种分类器介绍:

支持向量机(SVM)是建立在统计学 VC 理论和结构风险最小化原理基础上的机器学习方法,其对于小样本、非线性和高位模式识别问题的分类问题具有良好的解决能力。标准的 SVM 问题的本质是

求解一个受约束的二次型规划。通过特征融合,再利用 SVM 进行人脸表情的识别,在 CK 数据集上取得了 95% 的准确率^[14];结合 Gabor 小波与 SVM,减少了数据量的处理,取得了很好地效果^[15]。

隐马尔可夫模型(HMM)是一种基于数理统计模型,描述一个含有隐含未知数的马尔科夫过程,其在语音识别、模式识别被广泛应用,一般用于解决基于序列的问题。在表情识别领域,隐马尔可夫模型一般用于处理基于动态序列的人脸表情识别问题。

贝叶斯分类器是指以贝叶斯定理为核心的算法总称,贝叶斯决策通过相关概率已知的情况利用误判的损失来选择最优的类别分类。改进贝叶斯分类器并结合传统的人脸特征,提高了识别率并对光照和表情变化有较好的鲁棒性^[16]。

传统的机器学习分类方法对大型数据处理往往面临着各种问题,传统机器学习算法大都只学习出一种分类边界,不具有多维分类的特性。人脸表情识别也逐渐向深度学习领域发展,但传统机器学习在表情识别上的应用依然值得探索。机器学习算法在常用的数据上的识别结果见表 2。

表 2 传统机器学习算法在常用表情数据集上的比较

Tab. 2 Comparison of traditional machine learning algorithms on commonly used expression datasets

Datasets	Researcher	Feature	Classification method	Accuracy rate / %
CK	Xu Chao ^[17]	Haar-like	SVM	98.10
	Zhan Yong-zhao ^[18]	Gabor wavelet	FBMM	94.90
	Li Wen-shu ^[19]	Gabor wavelet	GPC+AdaBoost	91.50
	Hu min ^[12]	HCBP	SVM	97.7
JAFFE	Li Wen-shu ^[19]	Gabor wavelet	GPC+AdaBoost	90.90
	Zhou Shu-Ren ^[20]	CA	HMM	95.80
	Li Dong ^[21]	Gabor+LBP	RBF	93.42
	Hu min ^[12]	HCBP	SVM	95.3

3 基于深度学习的人脸表情识别

深度学习的相关研究在图像识别领域取得很大进展,在 ImageNet 数据集的识别率已经接近人类水平。在表情识别领域,深度学习相关算法已经居于主要地位,被广泛应用于基于静态图像和动态序列的表情识别问题。研究者们从各个角度出发提高表情识别的准确率,从特征的优化到数据的增强,从网络结构的变化到新的分类器的提出。

3.1 数据预处理

深度学习同样也需要对数据进行预处理,去除

与面部表情无关的背景和图像在光照、头部姿态上的差异。这些因素会影响到深度网络的特征学习,在训练之前对数据进行预处理是很有必要的。

3.1.1 人脸对齐

人脸对齐是表情识别的一个关键步骤。给定一系列的数据,首先要检测出人脸,之后再行人脸图像的裁剪以及其它变化。Viola-Jones 检测器是最广泛和经典的检测器,其输入特征为 haar-like 特征,对矩形图像区域像素及和差的阈值化,具有良好的鲁棒性,用于正面人脸识别^[22];监督下降法(supervised descent method, SDM)通过级联回归函数,将图像外

观映射到关键点,取得比较好的效果,深度网络被广泛用于人脸识别^[23];级联 CNN 是早期以级联方式预测关键点,级联回归在人脸对齐任务上具有高速度和准确性,除此之外还考虑多个人脸检测器组合来提高性能^[24]。将不同的面部关键点探测器串联起来,通过不同关键点的几何位置进行人脸对齐^[25]。

3.1.2 人脸标准化

在表情识别中光照和姿态的变化往往对特征提取造成影响,导致提取不到人脸表情特征,光照归一化和姿态归一化是比较常用的归一化方法。基于各向同性扩散(IS)的归一化方法、局域离散余弦变换归一化与高斯分布的差分(DOG)在光照归一化上取得了很好地效果,有效的去除了光照不均衡对特征的影响;直方图均衡化于光照归一化在人脸识别取得不错的性能^[26]。姿态归一化问题一直以来是一个难题,野外数据集的姿态复杂,不同于实验室数据集,对表情识别任务造成严重影响,最新的研究方法是基于 GAN 网络进行正面表情的合成。

3.2 深度网络结构

深度学习网络利用组合低层特征去形成更加抽

象的高维表示特征,应用这些高维特征来发现数据的分布,从而实现数据的分类。深度学习网络有着丰富的结构,实现端到端的分类任务。表情识别常用的网络结构:

3.2.1 卷积神经网络(CNN)

由于 CNN 对人脸位置变化和尺度变化具有很强的鲁棒性,CNN 网络被广泛的应用。使用 CNN 对人脸表情识别中主体独立性、平移、旋转、尺度不变性问题有了很好地解决^[27]。CNN 由卷积层、池化层、全连接层组成,其结构如图 2 所示。卷积层主要利用卷积核计算,卷积核是指对于输入图像的一部分区域,进行加权平均的处理,这个过程的权重由一个函数定义,这个函数就是卷积核。通过核函数对输入的图像进行卷积操作,经过激活函数线性化后得到特征图;池化层连接在卷积层之后,其作用相当于对特征图进行特征提取,减小输入特征图尺寸和数据量,同时获取更抽象的特征,防止过拟合,提高泛化性;全连接层将高维的特征图转化为一维特征图,实现分类的任务。常见的 CNN 网络架构有 AlexNet、VGGNet、GoogleNet、ResNet 等。

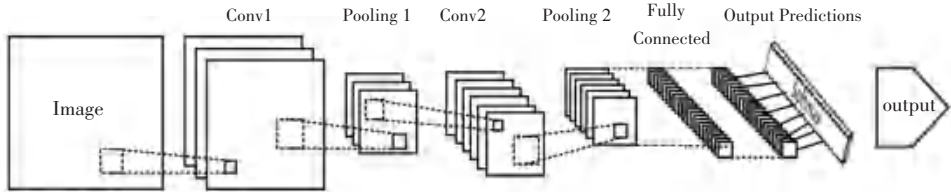


图 2 CNN 结构图

Fig. 2 CNN structure diagram

3.2.2 深度信念网络(DBN)

深度信念网络(DBN)由 Geoffrey Hinton 提出,其是一种生成模型,通过训练不断优化其神经元间的权重,可以让整个神经网络按照最大概率来生成训练数据。DBN 由多层神经元构成,这些神经元分为显性神经元(显元)和隐性神经元(隐元),显元接收输入,隐元提取特征,如图 3 所示。最上面的两层间的连接是无向的,组成联合内存;较低的其它层之间有连接上下的有向连接;最底层代表了数据向量,每一个神经元代表数据向量的一维。DBN 的组成元件是受限玻尔兹曼机(RBM),其训练过程一般包含 2 个阶段,第一阶段被称为预训练阶段,通过无监督地训练每一层 RBM 网络,使特征向量尽可能的映射到不同特征空间,从而尽可能多地保留特征信息;第二阶段叫做微调,通过反向传播网络,将错误信息自顶向下传播至每一层 RBM,微调 DBN 网络。

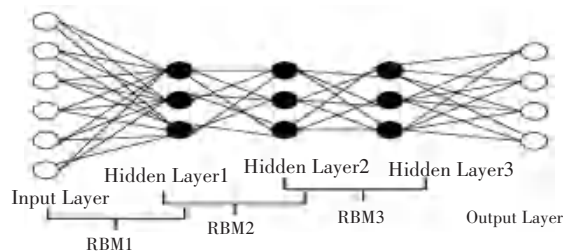


图 3 DBN 网络结构图

Fig. 3 DBN network structure diagram

3.2.3 循环神经网络(RNN)

循环神经网络(RNN)常用于挖掘数据中的时序信息以及语义信息的深度表达,在语音识别、语言模型、机器翻译以及时序分析等方面实现了突破。网络架构上,循环神经网络会记忆之前的信息,并对后面的信息造成影响。长短时记忆网络(LSTM)解决了 RNN 在信息间隔比较大时丧失学习能力,或者在信息间隔大小不同、长短不一时 RNN 的性能有所

降低^[28]。LSTM 通过一些“门”结构让信息有选择的影响循环神经网络每个时刻的状态,LSTM 的门一般包括遗忘门、输入门和输出门 3 种,如图 4 所示。遗忘门在 LSTM 中即以一定的概率控制是否遗

忘上一层的隐藏细胞状态。一种嵌套 LSTM,内部进行时间动态建模,外部进行输出,集成获取多层次特征^[29]。

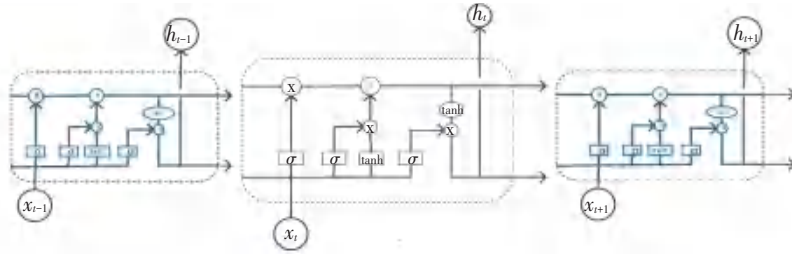


图 4 LSTM 的门运算结构

Fig. 4 Gate operation structure of LSTM

3.3 基于深度网络的表情识别

随着深度学习在人脸表情识别领域的发展,基于深度网络的表情识别算法大量出现,根据处理数据类型可分为静态图像的表情识别和动态序列的表情识别算法。

3.3.1 静态图像

静态图像表情识别算法通过预训练和微调,增加功能模块、构建级联网络等策略来提升表情识别的准确率。Yao 等提出了一个新颖的 CNN 架构 HoloNet,其将 CReLU 激活函数与残差网络相结合,在不降低效率的情况下增加网络深度,并设计了一个残差块,使得人脸表情识别系统通过学习多尺度

特征以捕获表情的变化,在人脸表情识别上取得很好效果^[30];基于单个深度网络能取得了良好的结果,研究人员又把方向转向多网络集成的方向,一个良好的集成网络应考虑各个网络的集成效果和网络的互补性^[31]。S.A.Bargal 等提出了集合最优的表情识别网络如图 5 所示,通过串联特征一维输出层再进行分类任务^[32]。级联网络通过对不同任务的模块进行组合构成一个更深层次的网络,Lv.Y 等利用级联结构设计了新的人脸表情识别结构,使用 DBN 检测人脸和与表情相关的区域,用堆叠式自动编码器对这些解析过的面元进行分类^[33]。

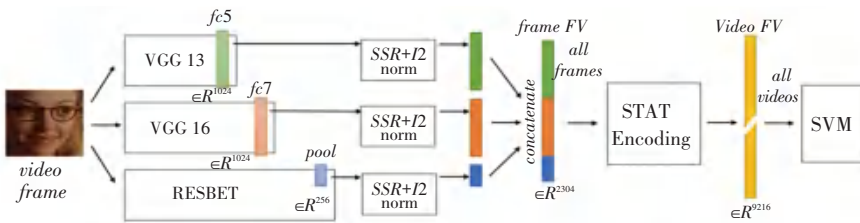


图 5 网络集成结构网络

Fig. 5 The network integration structure

3.3.2 动态序列

基于动态序列的表情识别通过从连续帧的时间相关性来获取信息,常用的策略有帧聚合、表情强度网络、深度时空网络。帧聚合策略利用视频片段中,帧在表达强度上不同的特点进行特征的提取,通过各种算法来聚合帧的输出^[34]。通过将所有帧的特征的均值、方差、最小值和最大值串联起来实现帧聚合^[35];对于大多基于序列的表情识别算法往往忽略细微的表情,表情强度以不同强度的训练样本作为输入,利用表情强度不同序列中表达之间的内在相

关性进行特征的优化提取^[35];Zhao 等提出峰值导数深度网络,网络结构如图 6 所示,将同一表情、同一对象的一对峰值和非峰值图像作为输入,利用 L2 范数损失来最小化两幅图像之间的距离^[36]。在反向传播过程中,提出了一种峰值梯度抑制(PGS)方法,在避免反向传播的同时,将非峰值表达式的学习特征推向峰值表达式的学习特征,从而提高了对低强度表达式的网络识别能力。虽然帧聚合技术可以在视频序列中集成帧,但时间依赖性没有被利用。利用深度时空网络在未知表情强度的前提下,将时间

窗口中的一系列帧作为单个输入,利用纹理和时间信息对更细微的表达进行编码^[37]。在真实场景中,人的面部表情是一个动态的过程,从细微到明显,通过序列进行人脸表情的分析是未来研究的重要方向。

构建良好的深度表情识别系统主要存在以下问题:缺乏丰富多样的训练数据和表情无关的变化因素,如光照、头部姿势和入种。为了解决训练数据不足和过度拟合的问题,预训练和微调、网络级联、网

络集合等成为人脸表情识别的主流。图像预处理阶段进行光照归一化、姿态归一化和图像裁剪操作,都能很好地去除无关因素的影响。面对头部姿态的变化,大多数深度表情识别网络还是没能很好地处理这个问题。本文对深度学习在人脸识别任务上的算法进行了总结,以不同的表情库为依据,分别对研究者、深度网络结构、网络层数、训练数据的选择和准确率进行了统计,见表3。

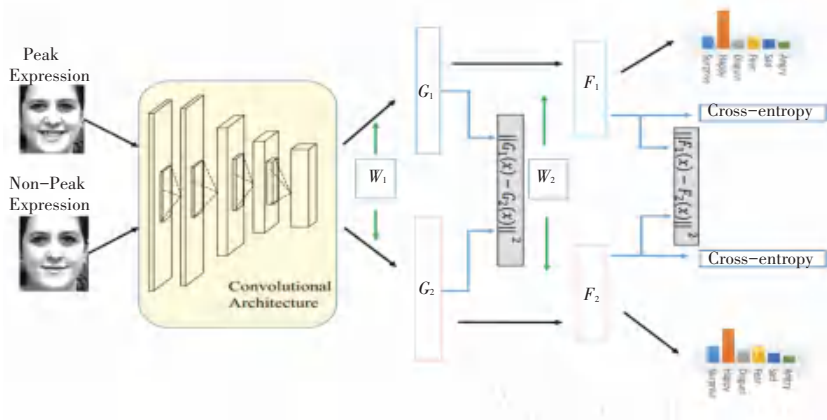


图6 PPDN级联网络

Fig. 6 PPDN cascade network

表3 基于深度网络的人脸表情识别算法

Tab. 3 Facial expression recognition algorithm based on deep network

Dataset	Researcher	Network Configuration	Number of Layers	The Selection of Training data	Accuracy/%	
CK+	S.Ouellet .2014 ^[38]	CNN AlexNet	8	last frame	6 classes (94.4)	
	Li 2015 ^[39]	RBM	4	last frame	6 classes: 96.8	
	Static	Liu 2014 ^[40]	Cascade connection DBN	5	first frame and last frame	6 classes: 96.7
		Liu 2015 ^[41]	Cascade connection DBN	5	first frame and last frame	7 classes: 93.70
	Ding 2017 ^[42]	CNN Fine tuning	8	first frame and last frame	6 classes: (98.6)	
	Zhang 2018 ^[31]	CNN Multitask network	—	the last three frames	8 classes: (96.8)	
	Zhao 2016 ^[36]	The intensity of expression	22	from frame 7 to the end	6 classes: 98.9	
	Dynamic	Yu 2017 ^[35]	The intensity of expression	42	from frame 7 to the end	6classes: 99.3
		kuo 2018 ^[43]	Frame Aggregation	6	Fixed length 9 frames	6 classes: 99.6
		Zhang 2017 ^[31]	Network Integration	7+5	space: the last frame time: All the frames	7 classes: 98.47
MMI	Liu 2015 ^[41]	CNN,RBM(Cascade Network)	5	the middle three frames and the first frame	7 classes: 98.50	
	Static	Yang 2018 ^[44]	GAN(GANc)	—	the middle three frames	7 classes: 75.85
		Dynamic	Sun 2017 ^[37]	Cascade Network	3*GoogleNet	space: network of expression data set time: neutral + expression
FER2013	Zhang 2017 ^[31]	CNN Multitask network	6	training:28709	6 classes: 91.46	
	Guo 2016 ^[45]	CNN + Fresh loss layer	10	validation:3589 test:3589	Test:75.10 Test:71.33	
SFEW2.0	Li 2017 ^[46]	CNN Cascade Network	VGG13+VGG16+ GoogleNet	Training:891 validation:431 test:372	Validation:51.75 Test:54.56	

4 表情识别的挑战和发展趋势

表情识别目前无论是基于图像还是基于序列的人脸表情识别都已经很成熟了,并实现了很高的识

别率,但是仍然存在一些问题和亟待突破的问题:

(1) 表情识别的数据库往往都是在实验室环境下获取的,数据集相对单一;

(2) 在 SFEW 数据集上的表情识别准确率比较

低,原因在于野外静态面部数据集的数据来自自然环境,包含很多影响因素如光照、姿态等,凸显了表情识别算法的鲁棒性不足;

(3)数据集的匮乏也是人脸识别问题的关键问题。面对人脸表情的差异性、肤色、外貌都有可能不同,数据对于深度学习方法有着重要作用,使用丰富度更高的数据集可训练出泛化性更好的表情识别网络;

(4)传统的表情分类相对较少。人的情绪比较复杂,传统的几类表情不能适应人机交互的复杂性,实用性可能大打折扣。

人脸表情识别基于图像与其它媒介进行结合研究也取得了较大进展。语音与图像的人脸表情识别算法、3D表情的人脸识别在解决遮挡问题上有了很大的突破;通过红外线获取面部的热数据,微调学习热特征来进行人脸表情识别,有效的避免了光照的影响。人脸表情识别在人脸遮挡上也是一个比较热门的研究方向,遮挡和非正面识别目前是人脸识别及表情识别的两大难题。

参考文献

- [1] EKMAN P, FRIESEN W V. Constants across cultures in the face and emotion [J]. *Journal of personality and social psychology*, 1971, 17(2):124-129.
- [2] SUWA M, SUGIE N, FUJIMORA K. A preliminary note on pattern recognition of human emotion expression [C]//Proceeding of the 4th International Joint Conference on Pattern Recognition, 1978:408-410.
- [3] LIU Dong, LI Su, CAO Zhidong. Overview of deep learning and its application in image object classification and detection [J]. *Computer science*, 2016, 43(12): 13-23.
- [4] ZHOU Wei, CHENG Yong, CAO Xuehong. Face recognition based on illumination normalization and segment-complete LBP feature [J]. *Computer engineering and application*, 2015, (11): 145-149.
- [5] OJALA T, PIETIKINEN M, HARWOOD D. A Comparative Study of Texture Measures with Classification Based on Feature Distributions [J]. *Pattern Recognition*, 1996, 29:51-59.
- [6] WANG Xian, ZHANG Yan, MU Xin, et al. Improved LBP face recognition algorithm [J]. *Optoelectronic engineering*, 2012, 39(7):12-13.
- [7] JABID T, KABIR M H, CHAE O. Robust facial expression recognition based on local directional pattern [J]. *ETRI journal*, 2010, 32(5): 784-794.
- [8] LEE T S. Image representation using 2D Gabor wavelets [J]. *IEEE transactions on pattern analysis and machine intelligence*, 1996, 18(10): 959-971.
- [9] LUO Fei, WANG Guoyin, YANG Yong. A new facial expression recognition method based on Gabor porter's sign [J]. *Computer science*, 2009, 36(1):181-183.
- [10] LIU Shuaishi, TIAN Yantao, WAN Chuan. Face expression recognition based on Gabor multi-directional feature fusion and block histogram [J]. *Chinese journal of automation*, 2011, 37(12):1455-1463.
- [11] LIENHART R, MAYDT J. An extended set of Haar-like features for rapid object detection [C]// Proceedings. International Conference on Image Processing. IEEE, 2002.
- [12] HU Min, LI Kun, WANG Xiaohua, et al. Facial expression recognition based on histogram weighted HCBP [J]. *Journal of electronic measurement and instrumentation*, 2015, 29(7):953-960.
- [13] WU X G, LUO L M. An improved optical flow field calculation method [J]. *Acta electronica sinica*, 2000(1):130-131.
- [14] LIEW C F, YAIRI T. Facial expression recognition and analysis: a comparison study of feature descriptors [J]. *IPSN transactions on computer vision and applications*, 2015, 7:104-120.
- [15] WANG Xiaohua, XIA Chen, Hu Min, et al. Video sequence facial expression recognition with spatio-temporal features [J]. *Journal of electronics and information*, 2018, 40(3):626-632.
- [16] FENG Xiaorong, HUI Kang-hua, LIU Zhendong. Face recognition based on convolution feature and Bayesian classifier [J]. *Journal of intelligent systems*, 2018, 13(5):769-775.
- [17] XU Chao, DONG Caichao, FENG Zhiyong, et al. Facial expression pervasive analysis based on Haar-like features and SVM [M]. Berlin Heidelberg; Springer, 2012:521-529.
- [18] ZHAN Yongzhao, CHENG Keyang, CHEN Yabi, et al. A new classifier for facial expression recognition: fuzzy buried Markov model [J]. *Journal of Computer Science and Technology*, 2010, 25(3): 641-650.
- [19] LI Wenshu, He Fangfang, QIAN Yuntao. For facial expression recognition based on Adaboost - Gaussian process classification [J]. *Journal of Zhejiang university*, 2012, 46(1):79-83.
- [20] ZHOU Shuren. Analysis and research on facial expression recognition algorithm [D]. Changsha: central south university, 2009.
- [21] LI Dong. Research on facial expression recognition based on local texture feature fusion [D]. Changchun: Jilin University, 2014.
- [22] VIOLA P A, JONES M J. Rapid Object Detection using a Boosted Cascade of Simple Features [C]// Computer Vision and Pattern Recognition, 2001. CVPR 2001. Proceedings of the 2001 IEEE Computer Society Conference on. IEEE, 2001.
- [23] XIONG X, FERNANDO DLT. Supervised Descent Method and Its Applications to Face Alignment [C]//Portland: 2013 IEEE Conference on Computer Vision and Pattern Recognition. IEEE Computer Society, 2013:532-539.
- [24] SUN Y, WANG X, TANG X. Deep convolutional network cascade for facial point detection [C]//Portland: Computer Vision and Pattern Recognition (CVPR), 2013 IEEE Conference on. IEEE, 2013:3476-3483.
- [25] LI K, JIN Y, AKRAM M W, et al. Facial expression recognition with convolutional neural networks via a new face cropping and rotation strategy [J]. *The Visual Computer*, 2019.
- [26] ÖZERDEM M S, POLAT H. Emotion recognition based on EEG features in movie clips with channel selection [J]. *Brain Inform*, 2017, 4(4):241-252.
- [27] XIA Tian, ZHANG Yifeng, LIU Yuan. Expression recognition based on feature points and multi-network training [J]. *Journal of computer aided design and graphics*, 2019, 31(4):552-559.
- [28] HOCHREITER S, SCHMIDHUBER J. Long short-term memory

- [J]. *Neural Computation*, 1997, 9(8):1735-1780.
- [29] YU Z, LIU G, LIU Q, et al. Deng, Spatio-temporal convolutional features with nested lstm for facial expression recognition [J]. *Neuro computing*, 2018, 317: 50-57.
- [30] YAO A, CAI D, HU P, et al. HoloNet: towards robust emotion recognition in the wild [C]//The 18th ACM International Conference. ACM, 2016:472-478.
- [31] ZHANG Z, LUO P, LOY C C, et al. From Facial Expression Recognition to Interpersonal Relation Prediction [J]. *International Journal of Computer Vision*, 2018, 126(5):1-20.
- [32] BARGAL S A, BARSOUM E, FERRER C C, et al. Emotion recognition in the wild from videos using images [C]//The 18th ACM International Conference on Multimodal Interaction. ACM, 2016:433-436.
- [33] LV Y, FENG Z, XU C. Facial expression recognition via deep learning [C]//2014 International Conference on Smart Computing (SMARTCOMP). IEEE Computer Society, 2014:303-308.
- [34] KAHOU S E, BOUTHILLIER X, LAMBLIN P, et al. Emonets: Multimodal deep learning approaches for emotion recognition in video [J]. *Journal on Multimodal User Interfaces*, 2016, 10(2): 99-111.
- [35] YU Z, LIU Q, LIU G. Deeper cascaded peak-piloted network for weak expression recognition [J]. *The Visual Computer*, 2017 (6-8):1-9.
- [36] ZHAO X, LIANG X, LIU L, et al. Peak-Piloted Deep Network for Facial Expression Recognition [C]//European conference on computer vision, Springer, 2016: 425-442.
- [37] SUN Ning, et al. Deep spatial-temporal feature fusion for facial expression recognition in static images [J]. *Pattern Recognition Letters*, 2017, (119): 49-61.
- [38] OUELLET S. Real-time emotion recognition for gaming using deep convolutional network features [J]. USA: arXiv preprint arXiv, 2014: 1408-1418.
- [39] LI J, LAM E Y. Facial expression recognition using deep neural networks [C]//Macau, China: 2015 IEEE International Conference on Imaging Systems and Techniques (IST). IEEE, 2015: 1-6.
- [40] LIU P, HAN S, MENG Z, et al. Facial Expression Recognition via a Boosted Deep Belief Network [C]//Columbus, Ohio: 2014 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). IEEE Computer Society, 2014: 1805-1812.
- [41] LIU Mengyi et al. AU-inspired Deep Networks for Facial Expression Feature Learning [J]. *Neuro computing* 2015, 159: 126-136.
- [42] DING H, ZHOU S K, CHELLAPPA R. FaceNet2ExpNet: Regularizing a Deep Face Recognition Net for Expression Recognition [C]//Gesture Recognition 12th IEEE International Conference on. IEEE, 2017: 118-126.
- [43] KUO C M, LAI S H, SARKIS M. A Compact Deep Learning Model for Robust Facial Expression Recognition [C]//Salt Lake City: 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW). IEEE, 2018:2121-2129.
- [44] YANG H, CIFTCI U, YIN L. Facial Expression Recognition by De-Expression Residue Learning [J]. *International Journal on Computer Science & Engineering*, 2018, 2(5):2220-2224.
- [45] GUO Y, TAO D, YU J, et al. Deep neural networks with relativity learning for facial expression recognition [C]//Curicó, Chile; in *Multimedia & Expo Workshops (ICMEW)*, 2016 IEEE International Conference on. IEEE, 2016: 1-6.
- [46] LI S, DENG W, DU J P. Reliable Crowdsourcing and Deep Locality-Preserving Learning for Expression Recognition in the Wild [C]//Honolulu, Hawaii Cite: 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). IEEE, 2017: 2584-2593.

(上接第42页)

3 结束语

本文针对颜色或亮度特征、纹理特征、梯度特征、深度特征等视觉表层特征无法有效检测出没有明显移动的伪装人体目标,提出了基于增强的微振动特征实现伪装人体检测的算法。在训练集上估计伪装人体目标的最优的呼吸率区间,并通过放大振动的相位信息来增强视频中伪装人体的微振动,根据增强的微振动视频和提出的检测模型实现伪装人体目标的定位检测,并通过图像形态学去噪后处理以降低检测结果的噪声。通过一系列的消融实验和对比实验,在提出的数据集上证明了本文检测算法的先进性。与其它检测特征的算法相比,本文算法可以在高光、深色、纹理不平坦等复杂背景中检测出没有明显移动的伪装人体目标。

参考文献

- [1] HUANG Z, JIANG Z. Tracking Camouflaged Objects with Weighted Region Consolidation [C]// *Digital Image Computing: Techniques and Applications*, 2005:24.
- [2] SENGOTTUVELAN P, WAHI A, SHANMUGAM A. Performance of Decamouflaging Through Exploratory Image Analysis [C]// *First International Conference on Emerging Trends in Engineering & Technology*. IEEE Computer Society, 2008:6-10.
- [3] PANY, CHEN Y, FU Q, et al. Study on the Camouflaged Target Detection Method Based on 3D Convexity [J]. *Modern Applied Science*, 2011, 5(4):152-157.
- [4] FANG Z, ZHANG X, DENG X, et al. Camouflage people detection via strong semantic dilation network [C]// *the ACM Turing Celebration Conference-China*. ACM, 2019:1-7.
- [5] WADHWA N, RUBINSTEIN M, FRÉDO DURAND, et al. Phase-based video motion processing [J]. *ACM Transactions on Graphics (TOG)*, 2013, 32(4CD):80.1-80.10.