

DOI: 10.19650/j.cnki.cjsi.J1905426

基于深度分离卷积的情绪识别机器人即时交互研究*

徐桂芝^{1,2}, 赵阳^{1,2}, 郭苗苗^{1,2}, 金铭^{1,2}

(1. 河北工业大学 省部共建电工装备可靠性与智能化国家重点实验室 天津 300130;

2. 河北工业大学 河北省电磁场与电气可靠性重点实验室 天津 300130)

摘要:情绪识别是人工智能领域的研究热点,人机交互系统若能感知人类的情感行为并能表达情感,将会使机器人与人类的交互更加自然。人类主要通过面部表情、语义语调、肢体语言等几个方面获取情感信息。以拥有高自由度的 NAO 机器人为应用平台,设计了机器人面部情绪识别和肢体情感表达的人机交互系统。首先,引入深度分离卷积算法对人脸表情(生气、恐惧、伤心、高兴、惊讶和中性)进行特征提取和分类,结果表明通过训练得到的网络模型对 FER2013 人脸表情测试集的预测正确率可以达到 0.711;其次,设计 NAO 机器人的肢体动作,对 6 种面部情感做出了分类;最后,对机器人实时表达使用者的情绪状态进行了测试,反馈时间均在 2 s 内,并对连续 10 帧预测结果进行了统计分析。

关键词:深度分离卷积;人形机器人;情感交互;情绪识别

中图分类号: TH7 **文献标识码:** A **国家标准学科分类代码:** 520.20

Research on real-time interaction for the emotion recognition robot based on depthwise separable convolution

Xu Guizhi^{1,2}, Zhao Yang^{1,2}, Guo Miaomiao^{1,2}, Jin Ming^{1,2}

(1. State Key Laboratory of Reliability and Intelligence of Electrical Equipment, Hebei University of Technology, Tianjin 300130, China;

2. Key Laboratory of Electromagnetic Field and Electrical Apparatus Reliability of Hebei Province,

Hebei University of Technology, Tianjin 300130, China)

Abstract: Emotion recognition is a research hotspot in the field of artificial intelligence. If the human-robot interaction system can perceive human emotional behavior and express emotion, it will make the interaction between robot and human more natural. Humans acquire emotional information mainly through facial expression, semantic intonation and body language. Taking the NAO robot with high degree of freedom as an application platform, a human-robot interaction system is designed for facial emotion recognition and body emotion expression. Firstly, the depthwise separable convolution algorithm is introduced to extract and classify features of facial expressions (e.g., angry, fear, sad, happy, surprise and neutral). Results showed that the prediction accuracy of FER2013 facial expression test set could reach 0.711 by the trained network model. Secondly, the body movement of NAO robot are designed and classified according to six facial emotions. Finally, the real-time expression of the user's emotional state by the robot is tested, and the feedback time is within 2 s. The statistical analysis of the prediction results of 10 consecutive frames is carried out.

Keywords: depthwise separable convolution; humanoid robots; emotional interaction; emotion recognition

0 引言

机器人能够识别人类的面部情绪,理解人类情感并做出适当的反应,这是人们对人机交互提出的新要求^[1]。

美国心理学家 Albert 提出情绪表达 = 7% 语言 + 38% 声音 + 55% 面部表情,面部表情则是情绪表达最主要的表现形式,且在不同国家、不同民族、不同文化中面部表情传递的情绪信息具有共通性,基于面部表情识别的情绪认知系统实验结果表明,根据面部表情推测情绪状态的准确

率达到90%以上。在此基础上拥有情感能力的机器人越来越引起人们的关注^[2]。情感机器人的研究涉及到计算机、心理学等交叉领域,在医疗康复中有很大的潜力^[3-4],但目前还处于初级阶段。文献[5]中引入了一种名为KAPPA的人脸机器人,它可以通过面部表情识别情绪,并产生6种基本的情绪。文献[6]中引入了Minotaurus机器人系统,构建了一个智能人机交互环境,通过手势、语音和面部表情与用户进行交互。文献[7]提出了一种基于表情情感识别的人机交互系统,将情感控制信号传递给机器人,机器人通过表达卡通符号组成的面部表情来回应用户。虽然很多人机交互系统都涉及到机器人的情感,但只有少数研究人员同时研究机器人的情感识别和情感表达,以促进人与机器人之间的顺畅沟通。以上人机交互中表情情绪识别研究都是建立在Cohn等^[8]提出的6种基本情绪状态,其中对每种表情进行了面部动作的编码形成了一系列面部肌肉的孤立运动,例如喜悦的表达可以被解释为眼轮匝肌眶部与颧大肌的联合收缩,这在目前可用的技术是可以被检测到的。人机交互中的面部表情识别大多采用小样本人为的提取特征,如Gabor小波、局部二值模式、局部Gabor二值模式、梯度直方图等^[9],对新的人脸图片泛化能力不足,容易受到环境因素影响,并受制于算法的设计。近年来随着深度学习的发展,卷积神经网络广泛应用于模式识别、图像处理等领域^[10-12],它是一种结构简单、训练参数少和适应性强的高效识别算法,但同样也存在问题,比如浅层网络提取特征不足鲁棒性较差;深层网络容易过拟合,计算参数较多,计算量较大,不适用于对智能终端的移植和实际应用等。

本文的应用平台为软银Aldebaran Robotics公司研制的人工智能机器人NAO,全身拥有25个自由度,两个2D摄像头,可以做出丰富的肢体动作和获取清晰的图片和视频流,已广泛应用于教育、辅助医疗等领域^[13-14]。本文拟以机器人NAO为应用平台,将深度学习和机器人结合,构建一种基于深度分离卷积的能够根据面部表情识别实时进行肢体情感反馈的人机交互系统,在开发具有情感和社交智能的个性化机器人方面,展示了巨大的潜力。

1 基于深度分离卷积的面部情绪识别

1.1 模型设计

情感交互的系统架构如图1所示。将表情识别应用在实时的人机交互中需要计算低延迟,识别率要接近于人的水平,深度分离卷积在满足一定的深度下,整体的参数量和计算量都大大降低,方便了卷积神经网络

在移动端和嵌入式设备中的应用,保证了模型的可移植性。

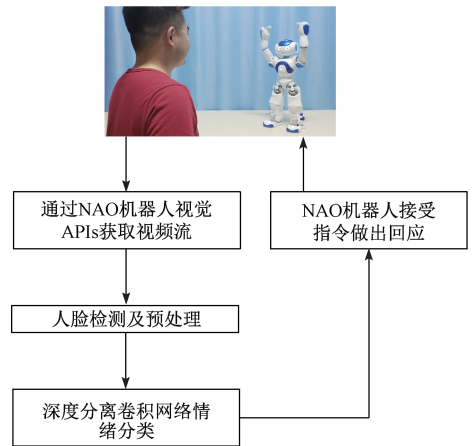


图1 人机情感交互系统

Fig.1 Diagram of human-robot emotional interaction system

Szegedy等^[15]提出了Inception结构,后续提出改进的V2^[16]、V3^[17]、V4^[18]版本。主要思想是先使用 1×1 的卷积核将特征图的各个通道映射到一个新的空间,在这一过程中学习通道间的相关性,再通过常规的 3×3 或 5×5 的卷积核进行卷积,以同时学习空间上的相关性和通道间的相关性。Francois^[19]将这种想法“极致”化,用二维的深度分离卷积(depthwise separable convolution, Separableconv2D)的方法,将通道的相关性和空间上的相关性达到了完全分离的效果,增加了网络的宽度,对分类提升正确率有很大的作用。

深度分离卷积^[20-21]分为逐层卷积和逐像素卷积两个过程,图2所示为普通卷积和深度分离卷积过程的对比。其中图(a)是普通二维卷积(Conv2D)的过程,图(b)是Separableconv2D的过程, C 为通道数, H 、 W 分别为每个通道的长和宽,两种卷积方式都以 $C \times H \times W$ 为输入,假设Conv2D有 k 个 3×3 的卷积, pad, stride都设为1, b 为偏置项,可得Conv2D的计算量为 $C \times H \times W \times k \times 3 \times 3$,对于Separableconv2D按通道数把输入分为 C 组,对每一组进行 3×3 卷积,得到逐层卷积特征图,接着进行 k 个 1×1 卷积得到逐像素卷积特征图,对通道和区域的完全分离得到对Conv2D计算量的压缩比为:

$$\frac{C \times H \times W \times 3 \times 3 + C \times H \times W \times k}{C \times H \times W \times k \times 3 \times 3} = \frac{1}{k} + \frac{1}{3 \times 3} \quad (1)$$

式(1)实现了对卷积计算的加速。根据Francois提出的网络模型,在每一次Separableconv2D后加入了批量标准化(batch normalization, BN)和激活函数,减少了原模型的层数,得到如图3所示的模型。

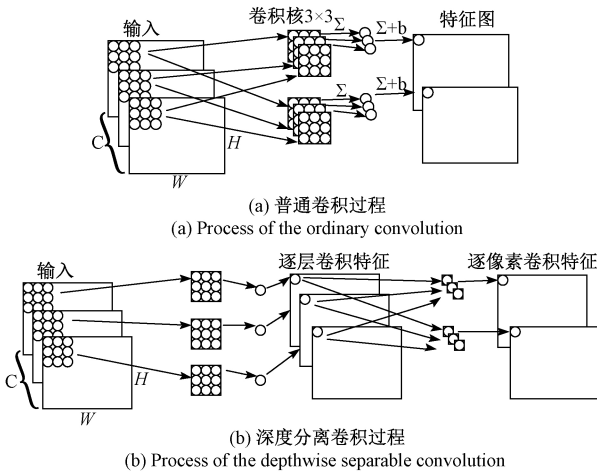


图 2 普通卷积过程和深度分离卷积过程
Fig.2 Process of the ordinary convolution and the depthwise separable convolution

图 3 中小括号标注部分为 5 个主要模块, *filters* 对应模块中每个卷积层的卷积核数量(图中括号 3、4、5 结构省略),其他结构都一样,每层卷积后进行 BN,公式如下:

$$y^{(k)} = \gamma^{(k)} \hat{x}^{(k)} + \beta^{(k)} \quad (2)$$

式中: γ 和 β 为参数, \hat{x} 为同一批次的归一化项, BN 可以提高网络的泛化能力,加速网络的收敛,有正则化的效果。模块中第一个 BN 后跟激活函数 ReLU,第 2 个 BN 后跟最大池化减少特征及参数,保留更多的纹理信息。模块左侧通过一个 1×1 卷积加入了类似 ResNet^[22] 的残差连接,一定程度上防止了网络稀疏后特征过多的丢失,也可以加快网络的收敛过程并获得更高的准确率。用全局平均池化代替全连接层又是对模型的压缩,减少了参数的数量,减轻了过拟合的发生。最后经过 softmax 将结果以总和为 1 的概率分布形式分类。

$$S_i = \frac{e^{z_i}}{\sum_{j=1}^N e^{z_j}} \quad (3)$$

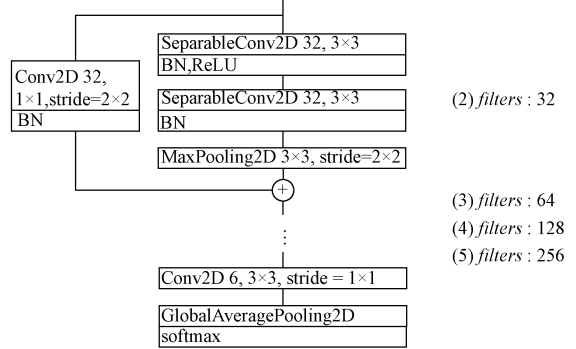
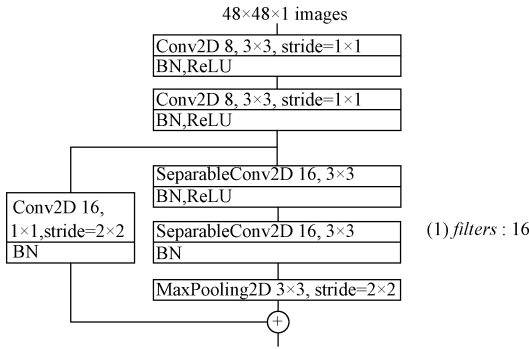


图 3 基于深度分离卷积的表情分类网络模型
Fig.3 Expression classification network model based on depthwise separable convolution

全局平均池化后的数组中第 i 个元素 z_i 被映射到 $(0,1)$ 区间内,其中 N 为数组中元素的总个数。整个网络不仅实现了稀疏化,大大减少了参数量和计算量,还在 13 层卷积的深度下保证了模型分类的准确率。

1.2 模型训练

在人脸情绪的识别中常用的数据集有 CK^[23]、JAFFE^[24] 等数据集,但这些数据集的图片数量较少,不满足深度神经网络的需求,容易造成模型的过拟合,使泛化能力不足。所以本文选用 FER2013^[25] 人脸表情数据集,由 28 709 张训练图,3 589 张公开测试图和 3 589 张私有测试图组成,每一张图片都是像素为 48×48 的灰度图,数据集中一共有生气、恶心、恐惧、高兴、悲伤、惊讶和中性 7 种表情,数据集人为测试的准确率是 $65\% \pm 5\%$ ^[26]。后续的机器人肢体表达中没有涉及“恶心”,所以模型训练的数据集将“恶心”去除,对剩下的 6 种表情进行训练,表情标签采用 one-hot 的编码方式,将 48×48 的灰度图的每个像素数据类型转化为浮点型,并归一化到 $[-1,$

1] 区间内,归一化公式如下:

$$J = (I/255.0 - 0.5) \times 2 \quad (4)$$

式中: J 为归一化结果; I 为输入图像的像素矩阵。其作用是减弱光线和光照强度等影响,使图像的细节更加清楚,还能加快梯度下降求最优解的速度。为了防止网络过快拟合,还对图像进行了随机水平偏移和随机竖直偏移、随机缩放和水平翻转、随机转动角度等数据增强,扩大数据集的数据量,增强训练网络的鲁棒性。损失函数选择交叉熵损失函数:

$$C = -\frac{1}{n} \sum_x [y \ln a + (1 - y) \ln(1 - a)] \quad (5)$$

式中: a 为真实输出; y 为预测输出; n 为类别数。优化函数选用 Adam,在训练中可适当调整学习率,公式如下:

$$\theta_{i+1} = \theta_i - \frac{\eta}{\sqrt{\hat{v}_i}} \tilde{m}_i \quad (6)$$

式中: θ 为更新值; \hat{v}_i 为第二时刻方差; \tilde{m}_i 为第一时刻平均值; η 为学习率初始值设为 0.001。数据集按随机抽取

10%的部分作为验证集,采用 Tensorflow 为后端的 keras 深度学习框架在 Windows10 Titan Xp 上训练,保存训练过程中最优的网络模型。

2 基于 NAO 机器人平台的情绪感知与表达

2.1 视频获取及预处理

NAO 机器人的两个摄像头安装在面部的不同位置,它们具有相同的参数,图 4 所示为摄像头的位置分布及可视范围,摄像头的最高分辨率为 $1\ 280 \times 960$,上下摄像头的垂直视角范围均为 47.64° ,水平视觉范围是 60.97° ,最高每秒可传输 30 帧的图像,聚焦范围从 30 cm 到无穷。通过 WiFi 创建“ALVideoDevice”代理和机器人进行连接,实时地从机器人上获取视频流。以 NAO 机器人水平观察人类的角度为准,选择头部摄像头,颜色空间选 kBGRColorSpace,为了不影响传输速率又能达到表情分类效果,设置分辨率为 160×120 ,速度为 30 帧/s。对每帧的图片用计算机视觉库 opencv 将图片进行灰度化,根据 Viola 等^[27]提出的方法对图片中的人脸进行检测,从图片中得到本文需要的人脸图片。每张人脸图片统一分辨率为 48×48 ,同样用式(4)进行归一化,再将图片形状设置为 $(1, 48, 48, 1)$,预处理的图像被送入到训练好的网络模型中,对图像中人脸表情进行预测。

2.2 机器人互动模型设计

机器人互动模型的输入为通过摄像头采集的人脸图像,再根据人脸特征做出情感判断,最后输出为机器人自

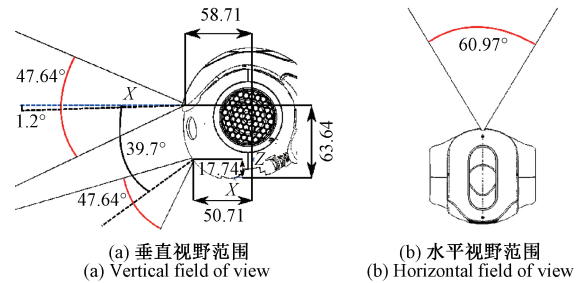


图4 NAO 机器人摄像头位置及可视范围

Fig.4 NAO robot camera location and visual range

身对应情感的表达。但由于 NAO 机器人没有面部特征,本文初步根据文献[28]设计了 6 种机器人情感表达的肢体动作,例如表达情感为高兴时,机器人举起双手,略微抬起头部双眼变为淡粉色,文献[28]中 14 个 8~9 岁的儿童参与了 NAO 机器人情感动作的识别实验,经实验验证身体姿势传达的情感状况与人的面部表情表达一致,而且发现机器人眼睛的颜色可以提高情绪表达的准确程度,在 NAO 的姿势上添加适当的眼睛颜色会使姿势更有效,机器人的身体自由度角度值和眼睛的颜色范围的具体数值如表 1 所示。表中是对应每种情感表达的最优关节自由度,由文献[28]识别实验得分最高一组关节自由度值对应现有版本机器人改进得到,每种情感的身体表达都由机器人唤醒站立姿势开始过渡,表中用横线代替数值的地方表示当前关节角度或颜色通道的数值相对于唤醒站立姿势不变,底部 3 行表示用于点亮和着色眼睛 LED 的参数。

表 1 NAO 肢体情绪最优表达时关节值的参数

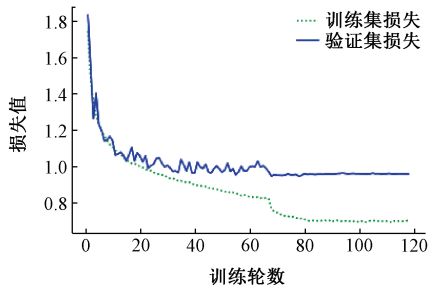
Table 1 Parameters of joint value when NAO body emotion is expressed optimally

关节名称	高兴	生气	恐惧	悲伤	惊讶	中性
LShoulderRoll	9.226	35.330	1.688	13.885	37.176	10.2
LShoulderPitch	-62.054	79.100	37.879	91.229	8.523	-10.2
RShoulderRoll	-8.176	-45.882	-1.672	-8.704	-47.903	84.1
RShoulderPitch	-63.983	84.642	6.594	84.203	25.931	84.1
LElbowYaw	-71.195	-7.737	-24.964	-58.978	-119.496	-24.0
LElbowRoll	-36.736	-81.649	-39.197	-24.959	-55.369	-68.7
RElbowYaw	55.809	6.589	50.535	52.733	119.496	68.7
RElbowRoll	29.007	86.048	88.334	5.979	49.310	24.0
HeadYaw	-	21.355	-38.059	24.871	-24.261	0
HeadPitch	-	-15.384	1.843	28.035	-15.823	-9.6
LHipYawPitch	-	-	-	-37.051	-10.544	-
LHipRoll	-	-	-	-	12.809	-
R	255	209	255	6	255	255
G	170	0	170	45	255	255
B	255	3	0	117	127	255

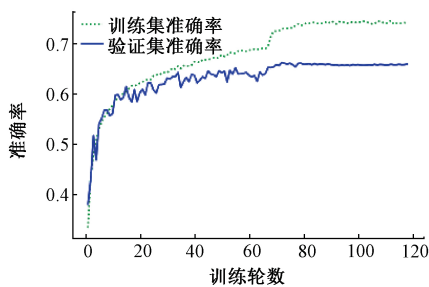
3 结果分析及实际交互

3.1 模型训练结果

图 5 所示为前 120 轮的训练结果,分别对比了训练集和验证集的损失值和精确率,以验证集为准,从图中可以看到,在 68 轮左右损失值停止下降,网络达到收敛,模型达到了最高的准确率。训练集在 68 轮后出现了过拟合的趋势,所以本文最后保存的是第 68 轮的训练网络模型。图 6 所示为最优模型对所有测试集的混淆矩阵分析,对于公共和私有测试集整体测试的准确率为 0.711,模型的识别率显然接近人类水平,但同时也发现快乐、中性的准确率明显高于其他表情,恐惧的识别率很低,对于此问题,经分析有以下两点原因:1)因为数据集不同表情的数量不均衡会使分类出现误差;2)因为恐惧和惊讶有一定的相似性,这些类易于混淆。为了满足人机交互的实时性,对不同数量的图片分别统计了网络模型对其的预测时间如表 2 所示(以人机交互 PC 端测试为准)。总体来看设计的深度学习网络模型在准确率和实时性方面满足了基本的人机交互要求。



(a) 损失值随训练轮数的变化
(a) Loss vary with the number of epochs



(b) 准确率随训练轮数的变化
(b) Accuracy vary with the number of epochs

图 5 训练集和测试集的损失值和准确率值随训练轮数的变化

Fig.5 Loss and accuracy in training set and test set vary with the number of epochs

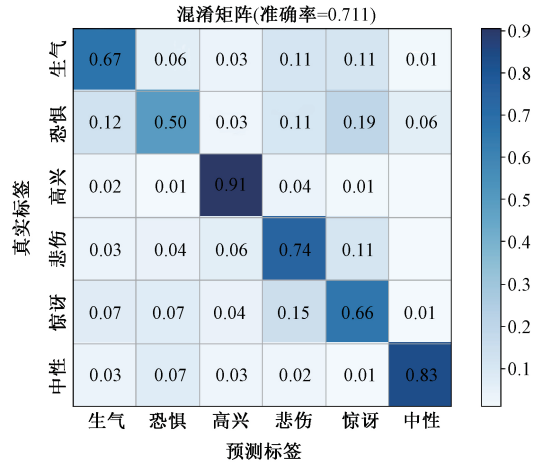


图 6 测试集混淆矩阵分析

Fig.6 Confusion matrix analysis of test sets

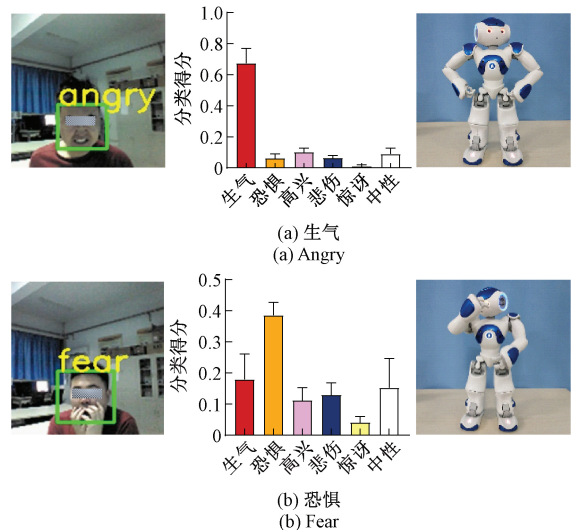
表 2 单张和多张图片的分类预测时间

Table 2 Single-picture and multi-picture classification prediction time

图片数量/张	分类预测时间/s
1	0.016
5	0.074
20	0.305
100	1.375

3.2 情感交互实际效果

由于机器人内部嵌入式芯片算力有限,本文通过 WiFi 创建代理和机器人进行连接,在 PC 端上首先搭建深度学习网络框架,PC 端系统为 Windows 8.1,深度学习框架后端用的是 Theano,用高级神经网络 API keras 载入我们训练好的最优人脸情绪识别模型,进行预测并将预测结果实时传输给机器人,机器人根据人的情绪状态做出肢体回应,图 7 所示为实时的测试结果。



(a) 生气
(a) Angry

(b) 恐惧
(b) Fear

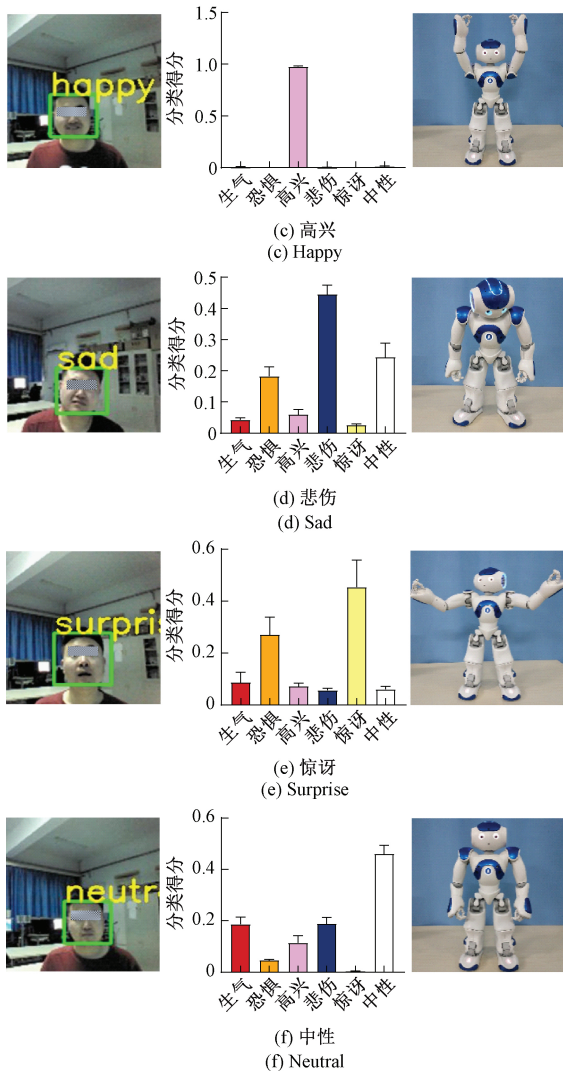


图7 人机情感交互实测结果

Fig.7 The real-time results of human-robot emotional interaction

实验员与机器人保持水平并模仿6种表情,根据测试表情的类别总共分为6组,其中每组表情结果的最左侧是由机器人实时采集的视频画面,画面中对识别的人脸和表情类别进行了标记,中间的柱状图是当前时间内实验员人脸表情的得分情况,即当前表情状态的可能性所占的概率大小,概率值是截取当前时间内连续10帧画面的预测结果的平均值,并计算了连续10帧分类得分的标准差,右侧是机器人根据推测结果实时做出的肢体回应。从图7中可以看出,视频画面清晰流畅实验员情感表达可以被机器人正确识别,生气、恐惧、高兴、悲伤、惊讶和中性6组的主情感得分(平均值±标准差)依次为 0.674 ± 0.095 、 0.385 ± 0.041 、 0.976 ± 0.008 、 0.445 ± 0.028 、 0.454 ± 0.032 和 0.460 ± 0.034 ,每种表情的实时识别率均超过其他类别的随机得分,即连续测试的表情得

分比较稳定,机器人对人类的情感预测可信度较高。人机实时交互的平均交互总时间为 0.079 ± 0.010 s,其中情绪识别时间平均为 0.058 ± 0.006 s,分别包括单帧图片传输、图片预处理和分类计算的时间,控制响应时间平均为 0.020 ± 0.004 s,即控制指令发出到机器人动作启动时刻的时间间隔。此外高兴的预测得分是最高的,经分析认为高兴与其他情感相比区分度更高,在日常的社交中最普遍也是最容易认知的表情之一,机器人的情感姿态和眼睛的颜色合理的对机器人情感进行了表征,在交互过程中做到了实时反馈。

4 结 论

本文设计了结合逐层卷积和逐像素卷积的深度分离卷积外加残差连接的含有13层卷积层的网络模型,在满足精确率的模型深度下,对模型进行了稀疏化,大大减少了参数和计算量,同时还防止特征在每个网络模块间传播时过多的丢失,在低延迟预测的基础上满足了对人机交互的需求,并将训练的面部表情识别的卷积神经网络成功运用到NAO机器人中,为NAO机器人设计了情感肢体动作,赋予了机器人感知情感和表达情感的能力,本文的研究内容在辅助治疗、智能家居等方面具有广阔的应用前景。但在表情识别中发现,恐惧和惊讶表情很难区分,在后面会对网络的结构再做调整,并对数据集进行扩充。之后,还会为机器人设计适合各个表情的语音语调,使情感表达更完整。

参考文献

- [1] AWAIS M, HENRICH D. Human-robot Interaction in an Unknown Human Intention Scenario [C]. IEEE International Conference on Frontiers of Information Technology, 2014:89-94.
- [2] WANG T M, TAO Y, LIU H. Current researches and future development trend of intelligent robot: A review[J]. International Journal of Automation and Computing, 2018, 15(5):525-546.
- [3] 郑悦, 景晓蓓, 李光林. 人机智能协同在医疗康复机器人领域的应用[J]. 仪器仪表学报, 2017, 38(10): 25-32.
ZHENG Y, JING X B, LI G L. Application of human-machine intelligence synergy in the field of medical and rehabilitation robot [J]. Chinese Journal of Scientific Instrument, 2017, 38(10): 25-32.
- [4] 徐国政, 宋爱国, 高翔, 等. 基于焦虑情绪与混杂控制的机器人辅助临床康复实验[J]. 仪器仪表学报, 2017, 38(10): 16-24.

- XU G ZH, SONG AI G, GAO X, et al. Robot-assisted clinical rehabilitation experiment based on anxiety emotion and hybrid control [J]. Chinese Journal of Scientific Instrument, 2017, 38(10):16-24.
- [5] FUKUDA T, TACHIBANA D, ARAI F, et al. Human-robot mutual communication system [C]. IEEE International Workshop on Robot & Human Interactive Communication, 2001:14-19.
- [6] RÖNING J, HOLAPPA J, KELLOKUMPU V, et al. minotaurus: A system for affective human-robot interaction in smart environments [J]. Cognitive Computation, 2014, 6(4):940-953.
- [7] LIU Z, WU M, CAO W, et al. A facial expression emotion recognition based human-robot interaction system[J]. IEEE/CAA Journal of Automatica Sinica, 2017, 4(4):668-676.
- [8] COHN J F, AMBADAR Z, EKMAN P. Observer-based measurement of facial expression with the Facial Action Coding System. [J]. Neuroscience Letters, 2007, 394(3):203-221.
- [9] ZHUANG L, GUAN Y. Improvement of face recognition with Gabor, PCA, and SVM under complex illumination conditions [J]. Journal of Advanced Computational Intelligence and Intelligent Informatics, 2019, 23(3):465-473.
- [10] SUN X D, WU P CH, HOI S C H. Face detection using deep learning: An improved faster RCNN approach[J]. Neurocomputing, 2018, 299(1):42-50.
- [11] ZENG N Y, ZHANG H, SONG B Y, et al. Facial expression recognition via learning deep sparse autoencoders [J]. Neurocomputing, 2018, 273(9):643-649.
- [12] VOULODIMOS A, DOULAMIS N, DOULAMIS A, et al. Deep learning for computer vision: A brief review[J]. Computational Intelligence and Neuroscience, 2018;1-13.
- [13] PÖHNER N, HENNECKE M. Evaluation of a robotics course with the humanoid robot NAO in CS teacher education[C].ACM Proceedings of the 13th Workshop in Primary and Secondary Computing Education, 2018: 28.
- [14] LYTRIDIS C, VROCHIDOU E, CHATZISTAMATIS S, et al. Social engagement interaction games between children with Autism and humanoid robot NAO[C]. The 13th International Conference on Soft Computing Models in Industrial and Environmental Applications, 2018: 562-570.
- [15] SZEGEDY C, LIU W, JIA Y Q, et al. Going deeper with convolutions[C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2015:1-9.
- [16] IOFFE S, SZEGEDY C. Batch normalization: accelerating deep network training by reducing internal covariate shift [C]. International Conference on International Conference on Machine Learning, 2015: 448-456
- [17] SZEGEDY C, VANHOUCKE V, IOFFE S, et al. Rethinking the inception architecture for computer vision[C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016: 2818-2826.
- [18] SZEGEDY C, IOFFE S, VANHOUCKE V, et al. Inception-v4, inception-resnet and the impact of residual connections on learning [C]. Conference on Computer Vision and Pattern Recognition, 2016:4278-4284.
- [19] FRANÇOIS C. Xception: Deep Learning with Depthwise Separable Convolutions [C]. IEEE Conference on Computer Vision and Pattern Recognition, 2016: 1251-1258.
- [20] ZHANG X Y, ZHOU X Y, LIN M X, et al. Shufflenet: An extremely efficient convolutional neural network for mobile devices[C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018: 6848-6856.
- [21] SHENG T, FENG C, ZHUO SH J, et al. A quantization-friendly separable convolution for mobilenets [C]. Workshop on Energy Efficient Machine Learning and Cognitive Computing for Embedded Applications, 2018: 14-18.
- [22] HE K M, ZHANG X Y, REN SH Q, et al. Deep residual learning for image recognition [C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016: 770-778.
- [23] LUCEY P, COHN J F, KANADE T, et al. The extended cohn-kanade Dataset (CK+): A complete dataset for action unit and emotion-specified expression [C]. IEEE Computer Society Conference on Computer Vision and Pattern Recognition-Workshops, 2010:94-101.
- [24] LYONS M J, AKAMATSU S, KAMACHHI M G, et al. Coding facial expressions with gabor wavelets[C]. IEEE Conference on Automatic Face and Gesture Recognition, 1998, doi:10.1109/AFGR.1998.670949.
- [25] GOODFELLOW I J, ERHAN D, CARRIER P L, et al.

Challenges in representation learning: A report on three machine learning contests [J]. *Neural Networks*, 2015, 64(4):59-63.

- [26] LIU K, ZHANG M M, PAN ZH G. Facial expression recognition with CNN ensemble [C]. *IEEE International Conference on Cyberworlds*, 2016: 163-166.
- [27] VIOLA P, JONES M J. Robust real-time face detection [J]. *International Journal of Computer Vision*, 2004, 57(2):137-154.
- [28] COHEN I, LOOIJJE R, NEERINCX M A. Child's perception of robot's emotions: Effects of platform, context and experience [J]. *International Journal of Social Robotics*, 2014, 6(4):507-518.

作者简介



徐桂芝, 分别在 1983 年、1999 年和 2002 年于河北工业大学获得学士学位、硕士学位和博士学位, 现为河北工业大学教授, 主要研究方向为脑认知与神经工程、生物医学电磁技术。

E-mail: gzxu@hebut.edu.cn

Xu Guizhi received her B. Sc. , M. Sc. and Ph. D. degrees all from Hebei University of Technology in 1983, 1999, and 2002, respectively. She is currently a full professor at Hebei University of Technology. Her main research interests include brain cognition and neural engineering, and biomedical electromagnetic technology.



郭苗苗 (通信作者), 分别在 2009 年、2012 年和 2016 年于河北工业大学获得学士学位、硕士学位和博士学位, 现为河北工业大学讲师, 主要研究方向为脑认知与神经工程、人机交互。

E-mail: gmm@hebut.edu.cn

Guo Miaomiao (Corresponding author) received her B. Sc. , M. Sc. and Ph. D. degrees all from Hebei University of Technology in 2009, 2012 and 2016, respectively. She is currently a lecturer at Hebei University of Technology. Her main research interests include brain cognition and neural engineering, and human-machine interaction.