

面向面部表情识别的双通道卷积神经网络

曹金梦¹,倪蓉蓉²,杨彪¹

(1.常州大学信息科学与工程学院,江苏 常州 213164)

(2.常州纺织服装职业技术学院机电系,江苏 常州 213164)

[摘要] 面部表情识别是机器感知人类情绪变化的重要途径,但表情识别受不同个体及情绪强弱差异影响较大,难以手动设计准确的特征.提出一种基于双通道卷积神经网络的面部表情识别方法,首先对采集得到的人脸图像进行预处理以限制分析范围,同时分析人脸灰度图像与对应的 LBP 图像以兼顾全局与细节特征;针对双通道输入数据,利用不同参数的卷积神经网络自动提取面部特征,通过加权融合分类网络进行特征融合,并利用 softmax 分类不同表情.实验结果表明,该算法能够以较高的准确率识别 6 种基本面部表情(高兴、悲伤、愤怒、沮丧、恐惧及惊讶).该方法性能优于基于手动设计特征的面部表情识别方法及单通道 CNN 方法,相比于其他双通道 CNN 方法,能通过更简单的处理得到近似的识别结果.

[关键词] 人脸检测,面部表情识别,双通道,卷积神经网络,LBP 图像

[中图分类号] TP391.41 **[文献标志码]** A **[文章编号]** 1672-1292(2018)03-0001-09

Binary-Channel Convolutional Neural Network for Facial Expression Recognition

Cao Jinmeng¹, Ni Rongrong², Yang Biao¹

(1.School of Information Science and Engineering, Changzhou University, Changzhou 213164, China)

(2.Department of Mechanical and Electrical, Changzhou Vocational Institute of Textile and Garment, Changzhou 213164, China)

Abstract: Facial expression recognition is an important way for machines to understand emotional changes of human beings. However, accurate hand-crafted features are hard to extract due to the fact that emotion recognition may be severely influenced by individual differences and emotional intensity differences. A facial expression recognition approach based on binary-channel convolutional neural network (BC-CNN) is proposed in this paper. Initially, the sampled facial images are pre-processed for region limitation of further analysis. Global and detailed features are both considered by studying facial gray-scale images and corresponding LBP images simultaneously. For binary-channel input data, two CNNs with different parameters are employed to extract facial features automatically and further, a weighted merge classify network (WMCN) is used for feature fusion and the softmax is finally used to recognize different facial expressions. Experimental results indicate that our algorithm can recognize six fundamental facial expressions (happy, sadness, angry, disgust, fear and surprise) with high accuracy. Its performance is better than that of those methods based on hand-crafted features or single channel CNN. Compared with other binary-channel CNN approach, our approach is easier to implement with similar recognition accuracy.

Key words: face detection, facial expression recognition, binary-channel, convolutional neural network, LBP images

面部表情识别是指利用计算机视觉技术从人脸图像中预测表情,它在揭示人的意图、情感及其他内在状态方面发挥着极大的作用,是机器感知人类情绪变化并与人类进行沟通的重要手段,在人机交互、健康监控、辅助驾驶等方面获得了广泛应用.

面部表情识别流程包括图像预处理、面部特征提取以及表情分类^[1].人脸检测通常利用级联分类器

收稿日期:2018-04-18.

基金项目:国家自然科学基金(61501060)、江苏省科技厅青年基金(BK20150271)、江苏省道路载运工具新技术应用重点实验室开放课题(ZMF15020068).

通讯联系人:杨彪,博士,讲师,研究方向:机器视觉、模式识别. E-mail: yb6864171@cczu.edu.cn

实现^[2],例如目前流行的 Viola-Jones 人脸检测框架^[3].检测到人脸后,可以利用眼睛、嘴角等特征点进行旋转校正^[4],并从校正后的脸部区域中提取能够反映表情变化的面部特征,包括嘴、眼睛、眉毛、鼻子等典型区域的位置与形状^[5]或脸部的全局/局部灰度特征^[6].最后,利用分类器(支持向量机 SVM^[7]及随机森林分类器 RF^[8]等)识别不同面部表情.

近年来,面部表情识别取得了较大进展.但是,光照变化、局部遮挡、面部旋转等复杂因素会影响人脸检测结果,进而降低表情识别的准确率.即使准确检测到人脸,面部表情识别也是一项极具挑战的任务,其难点在于:(1)相同表情的面部图像可能因人而异;(2)相同被试者不同表情的差异可能不明显;(3)相同表情的强弱程度可能导致面部图像存在差异.因此,面部表情识别不仅需要准确检测人脸,更需要能够有效描述表情变化的特征以及准确区分不同表情的分类器.

卷积神经网络(convolutional neural network, CNN)在模式识别领域也获得了飞速发展,尤其在脸部识别领域表现出色^[9].CNN 可以同时完成图像的特征提取与分类,随着网络层数的增加,它能更好地理解并学习对象的抽象特征^[10].因此,相比于手动设计得到的特征,利用深度网络,尤其是 CNN 自动提取到的脸部特征,能够更好地反映出人脸的表情变化,可以有效提高人脸表情识别的准确率.

本文主要考虑面部表情特征的提取与表情识别问题,搭建了一个双通道卷积神经网络,同时从人脸灰度图像与对应的 LBP 图像中自动提取特征,灰度图像倾向于面部全局信息的获取,LBP 图像倾向于面部细节信息的捕捉,针对不同图像的特点调整网络参数;设计了基于加权的融合策略,通过实验验证了不同权值的影响,并利用 softmax 直接输出表情分类;在多个基准数据库(CK+、JAFPE、Oulu-CASIA)上与主流面部表情识别算法进行对比,验证了本文算法的性能.

1 相关工作

近年来,大量研究者从基于手动设计特征与基于深度学习两个方面对面部表情识别展开了研究.

1.1 基于手动设计特征的面部表情识别

Zhong 等人^[11]提出利用双阶段多任务稀疏学习框架进行面部表情识别,通过多任务学习有效定位脸部重要区域,并通过稀疏编码提取特征,最后利用 SVM 进行表情分类.Zhang 等人^[12]对具有互补性的纹理特征与标记点特征进行多模态学习,并通过结构化约束增强不同模态数据的稀疏性,从而实现了对面部细微表情变化的识别.上述两项工作的研究对象是静止的面部图像,也有学者对动态的面部图像序列进行了研究.Walecki 等人^[13]认为面部表情的时序变化信息可以帮助算法提高表情识别的准确率,他们采用条件随机场算法识别面部表情以及运动单元.Jain 等人^[5]利用线性链条 CRF 模型、隐 CRF 模型以及额外的隐含层变量来建立人脸表情变化的动态模型,并通过相似度分析区别不同的面部表情.

1.2 基于深度学习的面部表情识别

随着深度学习在模式识别领域获得巨大的成功^[14],特别是多届 EmotiW 竞赛(Emotion Recognition in the Wild Challenge)的顺利举办,大量学者开始利用深度学习研究面部表情识别问题.何俊等人^[15]提出一种新的基于深度置信网络(deep belief net, DBN)模型无监督的表情特征提取和识别的方法.Lopes 等人^[16]利用 CNN 识别面部表情,针对 CNN 训练中需要大量样本的问题,通过数据增益扩大了样本量,并通过预处理保存了面部图像中与表情相关的特征,最终在 CK+数据库取得了较好的表现.Yu 等人^[17]则采用多个 CNN 来识别面部表情,并通过学习网络响应的集合权重实现不同 CNN 模型的融合,具有一定跨数据库表情识别能力.

2 双通道卷积神经网络搭建

图 1 所示为本文提出的面部表情识别算法流程,包括图像预处理模块与双通道卷积神经网络(binary channel-convolution neural network, BC-CNN),后者又由双通道特征提取网络(binary channel-feature extraction network, BC-FEN)与加权融合分类网络(weighted merge classify network, WMCN)构成,可以同时完成人脸图像的特征提取与表情分类.由于本文同时对人脸两个通道的图像进行处理,因此需要两个特征提取网络 BC-FEN_1 与 BC-FEN_2.对输入的人脸图像进行预处理,包括人脸检测、旋转校正、降采样以及数据样本扩充,若输入 RGB 图像,还需要进行灰度化以降低计算复杂度.对于样本扩充后的灰度图像,

计算对应的 LBP 图像,从而构成双通道样本集,用于后续模型训练与测试。

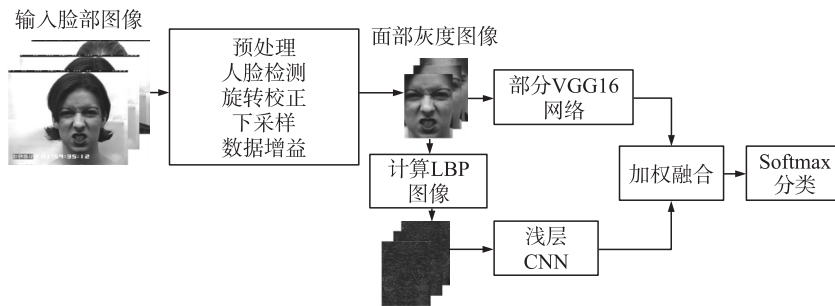


图1 本文提出的面部表情识别算法流程

Fig. 1 Pipeline of the facial expression recognition method proposed in this paper

2.1 数据预处理模块

2.1.1 人脸检测

本文采用 Viola-Jones 算法^[3]进行人脸检测,从而减少人脸图像中与表情分析无关的背景信息,增强特征描述子对不同面部表情的表达能力。

2.1.2 旋转校正

针对检测到的人脸图像中可能存在的角度差异,需要进行旋转校正。本文利用旋转变换矩阵进行校正,定义如下:

$$(Lx', Ly', 1) = [Lx, Ly, 1] \begin{bmatrix} \cos\theta & \sin\theta & 0 \\ -\sin\theta & \cos\theta & 0 \\ 0 & 0 & 1 \end{bmatrix}, \quad (1)$$

式中, Lx 、 Ly 为原始坐标; Lx' 、 Ly' 为变换后的坐标; θ 为旋转角度(通过测量两眼中心向量与水平方向的夹角得到)。

2.1.3 降采样

降采样是为了减小待处理图片尺寸,降低算法复杂度,并保证输入数据尺度一致。本文采用线性插值方法进行降采样,同时,为了避免过分降采样导致人脸 LBP 图像信息丢失,将降采样后的图像尺寸定为 64×64 像素。

2.1.4 数据样本扩充

BC-CNN 训练需要足够多的训练样本,而目前人脸表情基准数据库中的样本量无法满足要求,因此本文对原始数据进行旋转与平移产生合成数据^[18],并添加高斯白噪声,从而人为地扩充数据样本。

2.1.5 人脸 LBP 图像计算

局部二值模式(local binary pattern, LBP)是一种常用的图像局部纹理描述方法,它通过对比局部范围内的中心点与邻域点的像素值,得到关于该区域的二进制编码^[19]。图2所示为 LBP 编码示意图,其中左图表示局部区域中各点的像素值,右图为中心像素 5 对应的 LBP 编码。

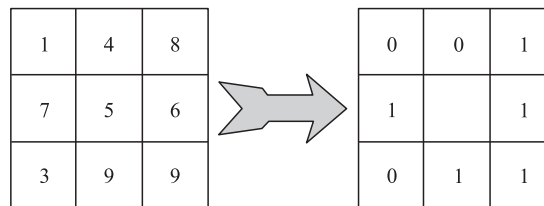


图2 LBP 编码示意图

Fig. 2 Illustration of LBP coding

得到某像素对应的 LBP 编码后,可以计算该像素的 LBP 值:

$$\text{LBP} = \sum_{n=1}^N S(g_n - g_c) \times 2^n, \quad (2)$$

式中, $S(\cdot)$ 为符号函数; N 为邻域点的数目; g_c 与 g_n 分别表示中心点与邻域点的像素值。通过计算图像每一点的 LBP 值,可以得到对应的 LBP 图像,图3所示即为惊讶表情下人脸及对应的 LBP 图像。从图3可以看出,通过提取人脸的 LBP 特征,使得与表情关联较大的嘴巴、眼角、眉毛等区域更加显著,使得深度网络更倾向于从这些区域中提取与表情分类相关的特征。



图 3 人脸 LBP 图像提取

Fig. 3 Calculation of facial LBP image

2.2 双通道特征提取网络设计

本文借鉴 DeepID 结构^[20]搭建了双通道特征提取网络(BC-FEN),以获取不同通道人脸图像与表情变化相关的特征. 图 4 所示为单通道特征提取网络的结构,包括一组输入层、两组卷积-下采样层(convolution-down sampling, C-S)以及一组特征向量层(feature vector, fv). 对于不同通道的输入图像,采用相同的网络结构,但针对不同通道图像的特点调整网络参数. 对于人脸灰度图像(描述全局信息),输入层大小为 32×32 像素,两组隐含层中卷积层与下采样层 C_1, S_1, C_2, S_2 大小分别为 $28 \times 28, 14 \times 14, 12 \times 12, 6 \times 6$ 像素,输出特征向量维度为 500. 对于对应的 LBP 图像(描述细节信息),输入层大小为 32×32 像素,两组隐含层的卷积层和下采样层 C_1, S_1, C_2, S_2 大小分别为 $30 \times 30, 16 \times 16, 14 \times 14, 7 \times 7$ 像素,输出特征向量维度同样为 500. 对于任意通道,特征向量层与两个下采样层 S_2 及 S_1 同时相连,保证所提取特征具有多尺度特性. 在此过程中,为了防止过拟合,在下采样层-特征向量层之间进行“dropout”^[21]操作(参数为 0.5).

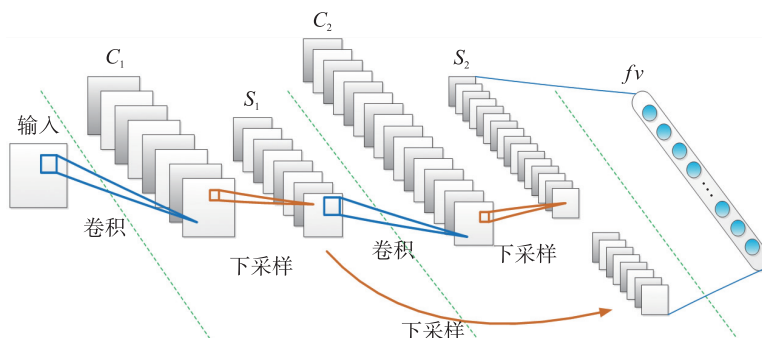


图 4 单通道特征提取网络结构图

Fig. 4 Structure of the single channel used to extract features

2.3 加权融合分类网络设计

图 5 所示为加权融合分类网络结构,其中 $fv_1 = \{s_1, s_2, \dots, s_m\}$, $fv_2 = \{l_1, l_2, \dots, l_m\}$ 分别表示从人脸灰度图像与对应的 LBP 图像中提取的特征向量, m 表示特征向量的维度. 对 fv_1, fv_2 进行加权融合得到向量 $fw = \{p_1, p_2, \dots, p_m\}$, 其中第 i 个成分 p_i 通过如下方式计算得到:

$$p_i = \alpha \cdot s_i + (1 - \alpha) \cdot l_i, \quad (3)$$

式中,权重 α 用以衡量人脸灰度图像与对应的 LBP 图像对最终融合特征的贡献, α 的取值通过实验方式确定. 对于融合后的特征向量 fw , 将其输出到两个级联的全连接层 fc_1 与 fc_2 , 其中 fc_1 维度为 300, fc_2 维度为 100, 而后利用 softmax 函数解决多分类问题, 从而识别不同的人脸表情.

softmax 函数可用于计算多类别分类问题的类别概率分布, 假设对于给定输入 x :

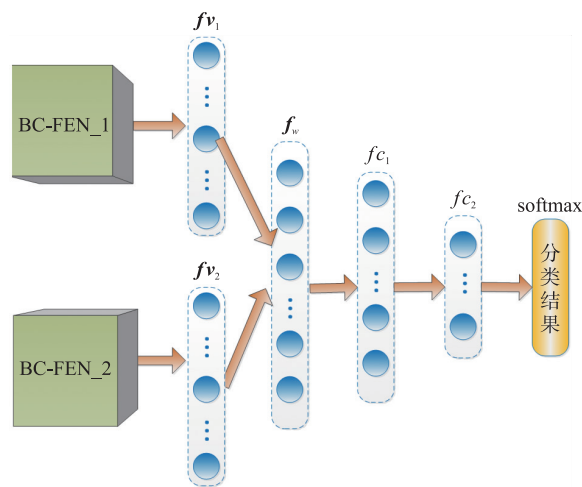


图 5 加权融合分类网络结构图

Fig. 5 Structure of the weighted fusion network

$$y_i = \frac{e^{x_i}}{\sum_{j=1}^K e^{x_j}} \text{ for } i=1, \dots, K, \quad (4)$$

式中,输入 \mathbf{x} 与输出 \mathbf{y} 都是 K 维向量, $\mathbf{x} = \{x_i\}_{i=1}^K$, $\mathbf{y} = \{y_i\}_{i=1}^K$. 式(4)中的分母起到正则化作用,保证输出向量的每个成分在 $[0, 1]$ 范围内. 作为神经网络的输出, softmax 函数中的值可用 K 个神经元来表示,即对给定输入 \mathbf{x} , 每种分类的概率 \mathbf{y} 可表示为

$$\begin{bmatrix} P(\mathbf{y}=1|\mathbf{x}) \\ \vdots \\ P(\mathbf{y}=K|\mathbf{x}) \end{bmatrix} = \begin{bmatrix} \zeta(\mathbf{x})_1 \\ \vdots \\ \zeta(\mathbf{x})_K \end{bmatrix} = \frac{1}{\sum_{j=1}^K e^{x_j}} \begin{bmatrix} e^{x_1} \\ \vdots \\ e^{x_K} \end{bmatrix}, \quad (5)$$

式中, $P(\mathbf{y}=k|\mathbf{x})$ 表示输入 \mathbf{x} 属于第 k 类的概率. 对于 softmax 函数的求解, 本文采用交叉熵作为损失函数, 定义如下:

$$\text{Loss}(\mathbf{y}, \mathbf{z}) = - \sum_{i=1}^K z_i \cdot \log(y_i), \quad (6)$$

式中, z_i 表示输入样本的真实值, y_i 表示 softmax 输出值. 对于式(6), 可采用基于梯度下降优化的反推算法进行最小化^[22], 从而得到网络输出.

3 实验结果分析

3.1 实验平台配置及基准数据库

本实验在 Linux 平台下基于 Keras 框架搭建了 BC-CNN 对所提出的算法进行实验分析, 计算机配置为 Intel i7 处理器、32G 内存、英伟达 1080 显卡、8G 显存, 并安装有 Python 3.0、NVIDIA CUDA Framework 6.5 以及 cuDNN 库.

为了使实验结果具有说服力, 本文采用以下 3 个人脸表情基准数据库进行分析:

(1) CK+^[23]: 该数据库包含 6 种基本表情(高兴、伤心、惊讶、沮丧、恐惧、愤怒)以及中性表情, 数据来自 70 个被试者, 总共 593 个样本, 每种基本表情约有 80 个样本, 将其按照数据样本扩充策略扩大 50 倍, 则每种表情可以得到约 4 000 个样本, 并采用 10-折交叉验证进行实验;

(2) JAFFE^[24]: 该数据库来自 10 个被试者, 总共 213 个样本, 除去 30 个中性表情其余 6 种基本表情每种包含约 30 个样本, 将其按照数据样本扩充策略扩大 100 倍, 则每种表情可以得到约 3 000 个样本, 并采用 10-折交叉验证进行实验;

(3) Oulu-CASIA^[25]: 该数据库来自 80 个被试者, 包含 6 种基本表情共 10 880 个样本, 每种基本表情约有 1 800 个样本, 将其按照数据样本扩充策略扩大 3 倍, 则每种表情可以得到约 5 400 个样本, 并采用 10-折交叉验证进行实验.

需要注意的是, 对标准数据库的原始数据样本进行扩充可能会导致过拟合, 因此在搭建的网络中引入了“dropout”操作, 从而尽可能减轻过拟合的影响.

3.2 网络参数设置及收敛性能分析

BC-CNN 的网络参数设置如表 1 所示.

本文通过准确率判断表情识别精度, 计算公式为:

$$\text{Accuracy} = (TP + TN) / (P + N), \quad (7)$$

式中, P 为正样本数; N 为负样本数; TP 表示被正确地划分为该类表情的样本数; TN 表示被正确地划分为不是该类表情的样本数.

图 6 所示为训练过程中每个 Epoch 对应的损失 (Loss) 及预测准确度 (Accuracy). 不难看出, 任意数据库都能在 15 ~ 20 个 Epochs 后趋于稳定 (Loss 与 Accuracy 基本不变).

表 1 BC-CNN 参数设置

Table 1 Parameter settings of the BC-CNN

参数	参数值
学习率 (Learning rate)	0.01
动量 (Momentum)	0.0
Epoch 数目	30
性能评估 (Performance Evaluation)	Accuracy

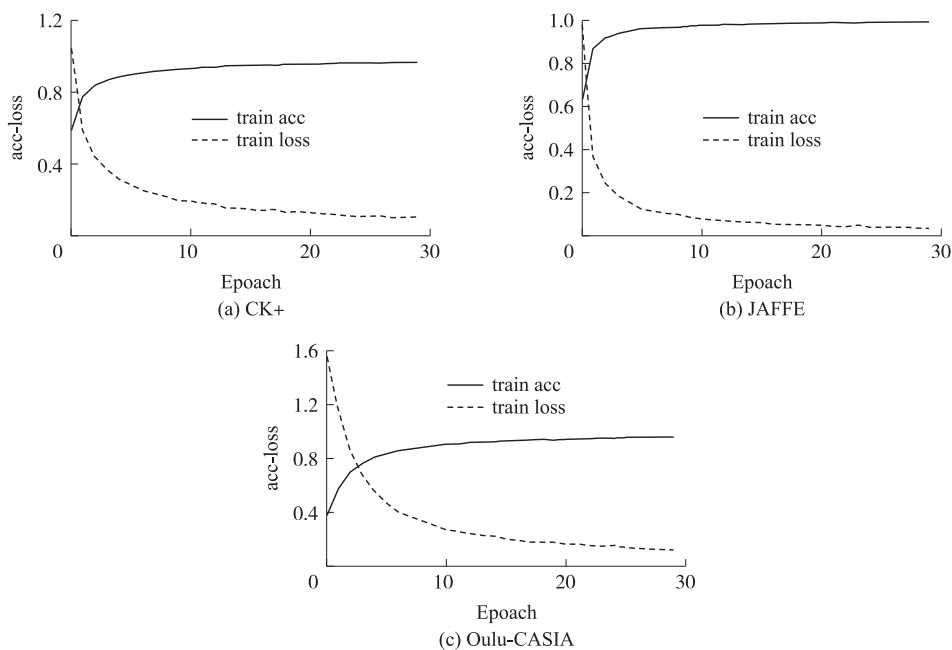


图 6 损失与准确度变化曲线
Fig. 6 Curves of “loss” and “accuracy”

3.3 算法有效性分析

为了验证双通道方法的有效性,分别测试了“灰度”CNN 与“LBP”CNN 在基准数据库 CK+、JAFFE 及 Oulu-CASIA 下的平均识别率,并与双通道方法的识别结果进行对比.如图 7 所示,双通道方法的平均识别率要高于任意单通道方法.

为了验证加权融合分类网络中权重 α 对算法精度的影响,在不同标准数据库下验证了不同 α 值对算法精度的影响.如图 8 所示,对于任意数据库,计算 10-折交叉验证下的准确率, α 的步进设置为 0.1, α 为 0 表示仅使用 LBP 人脸图像的情况, α 为 1 表示仅使用灰度人脸图像的情况.图中实线表示 CK+的结果,点划线表示 JAFFE 的结果,虚线表示 Oulu-CASIA 的结果.不难看出,仅使用灰度人脸图像算法准确率高,对于加权融合,当 α 为 0.6、0.7 时算法分类准确率高,表明灰度人脸图像对于识别不同表情的贡献大于 LBP 人脸图像.根据实验测试结果,本文令 $\alpha = 0.6$.

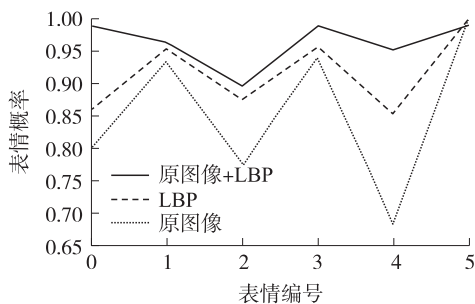


图 7 CNN 表情识别结果对比

Fig. 7 Comparisons with CNN facial expression recognition

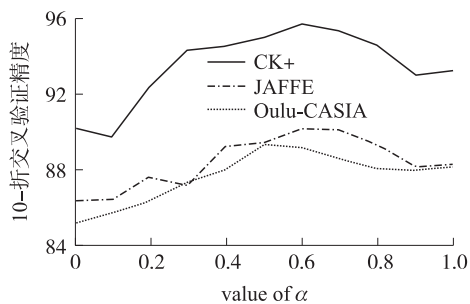


图 8 融合权重对算法精度的影响

Fig. 8 Evaluation of fusion weight on algorithm accuracy

3.4 面部表情识别准确率分析

图 9 分别给出了本文算法在 3 个数据库上对 6 种基本面部表情的识别结果.对于任意数据库,采用 10-折交叉验证,结果以混淆矩阵的形式给出.从图 9 可见,本文算法在 CK+数据库上有不错的表现,识别准确率基本在 96%以上;对于 JAFFE 数据库,算法在识别“fear”与“sadness”表情时表现较差,其原因可能是真实样本量过少导致较难准确识别这两种表情;对于 Oulu-CASIA 数据库,算法整体表现较好,但在识别“sadness”表情时易将其错误地识别为“anger”或“surprise”,究其原因可能是该数据库中少量被试者的不同表情比较接近.

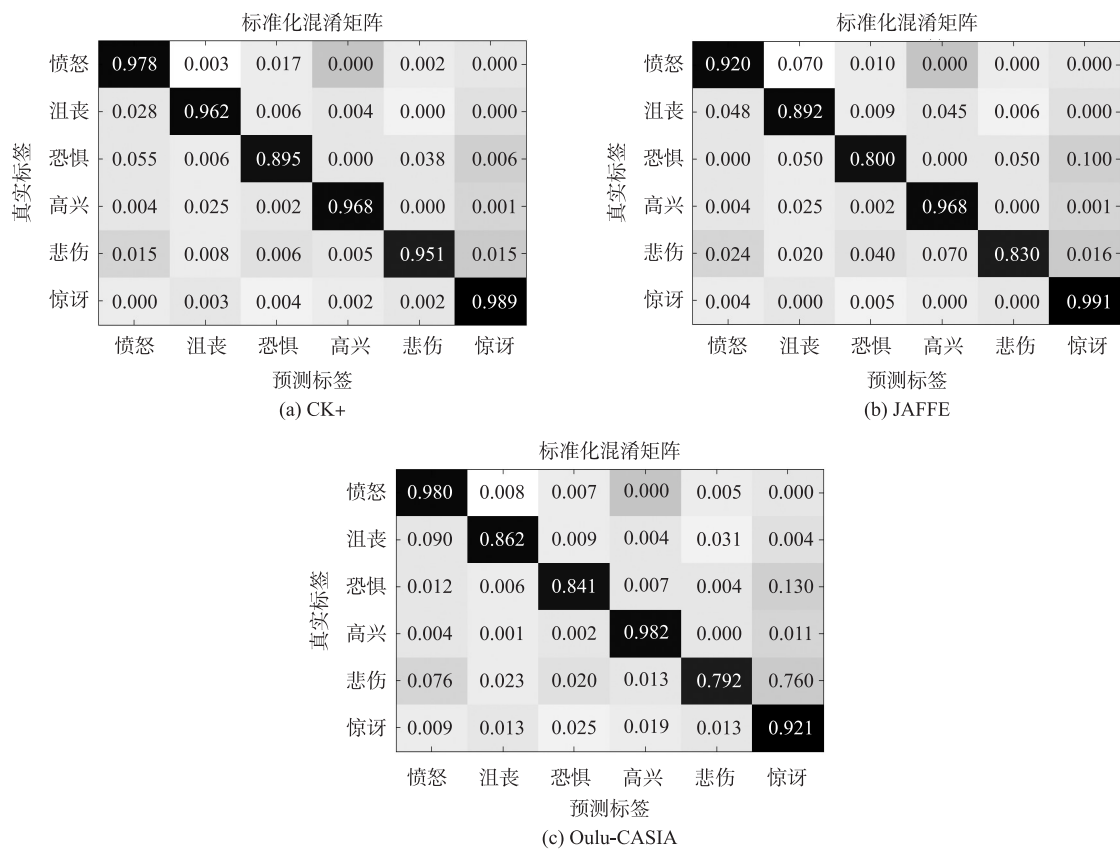


图9 面部表情识别结果

Fig. 9 Recognition results of facial expressions

同时,本文在基准数据库上对比了本文算法与主流表情识别算法的性能. 对于任意数据集,同样采用 10-折交叉验证方法,所采用对比算法的参数依照原文献进行设置. 表 2 所示为在同一实验框架下不同数据库上不同算法的平均识别准确率,其中 Aly^[26]与 Rivera^[27]通过手动设计特征进行面部表情识别,其识别精度低于基于深度学习的识别算法;Lopes^[17]使用单通道 CNN 进行面部表情识别,其精度略低于双通道 CNN 算法;Zhang^[12]使用脸部全局图像与局部区域图像(基于眼睛、鼻子、嘴等特征点检测的结果)作为 CNN 的输入,其识别性能与本文方法类似,但其网络结构更复杂,且需要预先检测特征点. 由此可知,本文算法可以较好地识别 6 种基本面部表情,且算法复杂度不高.

3.5 时间开销

本文算法需要 1.2 s 处理 72×72 像素的人脸图片. 其中,预处理(包括人脸检测、旋转校正及计算 LBP 图像)需要约 0.4 s,通过双通道网络识别面部表情需要约 0.8 s. 本文算法未经过任何优化,与 Zhang^[12]使用的方法(实测约 3.8 s)相比时间开销明显减少.

4 结语

本文搭建了一种双通道卷积神经网络来识别 6 种基本面部表情,同时考虑了人脸灰度图像与对应的 LBP 图像,兼顾了全局与细节信息. 该网络能够自动提取不同通道人脸图像的特征,并进行加权融合,从而给出较为准确的表情识别结果. 实验结果表明,基于深度学习获取的特征要优于传统手动设计的特征. 同时,本文提出的双通道 CNN 方法的表情识别性能要高于单通道 CNN 方法,且不需要预先检测脸部特征点,实施起来相对简单. 另一方面,由于训练需要足够多的样本,而目前人脸表情基准数据库中的小

表 2 主流面部表情识别算法平均准确率对比

Table 2 Comparison of average accuracy of the state-of-the-art facial expression recognition approaches

所采用 算法	平均识别准确率/%		
	CK+	JAFF	Oulu-CASIA
Aly	88.14	87.32	84.21
Rivera	91.51	88.75	85.18
Lopes	93.68	88.73	86.42
Zhang	95.12	91.48	87.88
本文方法	95.68	90.42	89.30

样本数据量过少,虽然对原始数据进行了数据增益,但数据样本仍然有限,此外,目前网络对小尺度样本普遍检测率不高,本文算法在处理小样本数据及表情差异不明显数据时存在不足. 如何调整样本扩充策略,进一步改善小样本训练问题,如何通过对网络结构进行调整,提取能够区分微弱表情变化的特征,是课题组的后续研究方向.

[参考文献](References)

- [1] 黄建,李文书,高玉娟. 人脸表情识别研究进展[J]. 计算机科学,2016,43(增刊2):123-126.
HUANG J,LI W S,GAO Y J. Research advance of facial expression recognition[J]. Computer science,2016,43(Suppl. 2):123-126. (in Chinese)
- [2] CHEN C R,WONG W S,CHIU C T. A 0.64 mm real-time cascade face detection design based on reduced two-field extraction[J]. IEEE transactions on very large scale integration system,2011,19(11):1937-1948.
- [3] BARNOUTI N H,ALDABBAGH S S M,MATTI W E,et al. Face detection and recognition using viola-jones with PCA-LDA and square euclidean distance[J]. International journal of advanced computer science and applications,2016,7(5):371-377.
- [4] DEMIRKUS M,PRECUP D,CLARK J,et al. Multi-layer temporal graphical model for head pose estimation in real-world videos[C]//2014 IEEE International Conference on Image Processing. Paris,France:IEEE,2014:3392-3396.
- [5] JAIN S,HU C,AGGARWAL J. Facial expression recognition with temporal modeling of shapes[C]//2011 International Conference on Computer Vision Workshops. Barcelona,Spain:IEEE,2011:1642-1649.
- [6] ALI G,IQBAL M A,CHOI T S. Boosted NNE collections for multicultural facial expression recognition[J]. Pattern recognition,2016,55:14-27.
- [7] 李宇成,刘昆. 一种基于人脸分割的 PCA 和 SVM 人脸识别方法[J]. 微型机与应用,2016,35(15):51-53.
LI Y C,LIU K. PCA and SVM face recognition method based on human face segmentation[J]. Microcomputer & its applications,2016,35(15):51-53. (in Chinese)
- [8] PU X,FAN K,CHEN X,et al. Facial expression recognition from image sequences using twofold random forest classifier[J]. Neurocomputing,2015,168:1173-1180.
- [9] LI H,LIN Z,SHEN X,et al. A convolutional neural network cascade for face detection[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Boston,USA:IEEE,2015:5325-5334.
- [10] LECUN Y,BENGIO Y,HINTON G. Deep learning[J]. Nature,2015,521(7553):436-444.
- [11] ZHONG L,LIU Q S,YANG P,et al. Learning multi-scale active facial patches for expression analysis[J]. IEEE transactions on cybernetics,2015,45(8):1499-1510.
- [12] ZHANG W,ZHANG Y,MA L,et al. Multimodal learning for facial expression recognition[J]. Pattern recognition,2015,48(10):3191-3202.
- [13] WALECKI R,RUDOVIC O,PAVLOVIC V,et al. Variable-state latent conditional random fields for facial expression recognition and action unit detection[C]//11th IEEE International Conference and Workshops on Automatic Face and Gesture Recognition. Ljubljana,Slovenia:IEEE,2015:1-8.
- [14] 郑胤,陈权崎,章毓晋. 深度学习及其在目标和行为识别中的新进展[J]. 中国图象图形学报,2014,19(2):175-184.
ZHENG Y,CHEN Q Q,ZHANG Y J. Deep learning and its new progress in object and behavior recognition[J]. Journal of image and graphics,2014,19(2):175-184. (in Chinese)
- [15] 何俊,蔡建峰,房灵芝,等. 基于 LBP/VAR 与 DBN 模型的人脸表情识别[J]. 计算机应用研究,2016,33(8):2509-2513.
HE J,CAI J F,FANG L Z,et al. Facial expression recognition based on LBP/VAR and DBN model[J]. Application research of computer,2016,33(8):2509-2513. (in Chinese)
- [16] LOPES A T,DE AGUIAR E,DE SOUZA A F,et al. Facial expression recognition with convolutional neural networks: coping with few data and the training sample order[J]. Pattern recognition,2017,61:610-628.
- [17] YU Z,ZHANG C. Image based static facial expression recognition with multiple deep network learning[C]//Proceedings of the 2015 ACM on International Conference on Multimodal Interaction. London,UK:ACM,2015:435-442.
- [18] SIMARD P,STEINKRAUS D,PLATT J C. Best practices for convolutional neural networks applied to visual document analysis [C]//Proceedings of the 7th International Conference on Document Analysis and Recognition. Edinburgh,Scotland:IEEE,2003:958-963.

- [19] GUO Z,ZHANG L,ZHANG D. A completed modeling of local binary pattern operator for texture classification[J]. IEEE transactions on image processing,2010,19(6):1657–1663.
- [20] SUN Y,LIANG D,WANG X,et al. Deepid3:Face recognition with very deep neural networks[DB/OL]. [2018–06–10]. <https://arxiv.org/abs/1502.00873>.
- [21] SRIVASTAVA N,HINTON G E,KRIZHEVSKY A,et al. Dropout:a simple way to prevent neural networks from overfitting[J]. Journal of machine learning research,2014,15(1):1929–1958.
- [22] THEODORIDIS S. Improving the way neural networks learn[M]//MICHAEL A N. Neural networks and deep learning. Pittsburgh:Academic Press,2015.
- [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]//2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops(CVPRW). San Francisco,USA:IEEE,2010:94–101.
- [24] EKMAN P. Strong evidence for universals in facial expressions;a reply to Russell's mistaken critique[J]. Psychological bulletin,1994,115(2):268–287.
- [25] LYONS M,AKAMATSU S,KAMACHI M,et al. Coding facial expressions with gabor wavelets[C]//Proceedings Third IEEE International Conference on Automatic Face and Gesture Recognition. Nara,Japan:IEEE,1998:200–205.
- [26] ALY S,ABBOTT A L,TORKI M. A multi-modal feature fusion framework for kinect-based facial expression recognition using Dual Kernel Discriminant Analysis(DKDA)[C]//IEEE Winter Conference on Applications of Computer Vision. NewYork,USA:IEEE,2016:1–10.
- [27] RIVERA A R,CASTILLO J R,CHAE O O. Local directional number pattern for face analysis:face and expression recognition[J]. IEEE transactions on image processing,2013,22(5):1740–1752.

[责任编辑:严海琳]