[文章编号] 1007－7405(2021)02－0129－10 DOI:10.19715/j.jmuzr.2021.02.07

一种面向移动端的浅层CNN表情识别

张东晓，陈彦翔

（集美大学理学院，福建 厦门 361021）

[摘要] 移动端的表情识别有巨大需求，但是受算力限制，主流深度神经网络无法直接移植。为此，设计了一个浅层网络，在节约计算量的同时保证了识别率。网络中使用三组堆叠而成的卷积层，有助于增大感受野，便于更好地提取特征，这是提升识别率的关键；使用全局平均池化层，避免引入额外的全连接层，大幅降低参数量，在训练样本不足的情况下，降低模型过拟合风险。在FER－2013数据集进行训练，准确率超过现有大多数算法；在CK＋数据集上进行微调，测试集上的准确率可达到0.96。将所得模型转换为Core ML模型，结合Xcode平台在iOS端搭建了实时表情识别App，在iPhone 8 Plus上能够稳定、流畅运行，识别效果达到预期。

[关键词] 面部表情识别；卷积神经网络；全局平均池化；Google Colab；Core ML

[中图分类号] TP 391.41

Mobile-Oriented Facial Expression Recognition Based on Shallow CNN

ZHANG Dongxiao, CHEN Yanxiang

(School of Science, Jimei University, Xiamen 361021, China)

Abstract: There is a huge demand for facial expression recognition on mobile terminals. However, due to the limitation of computational power, most popular deep neural networks cannot be directly transplanted. Therefore, a shallow network is designed in this paper, which can not only save the amount of calculation, but also ensure the recognition rate. The network uses three groups of stacked convolution layer, which helps to increase the receptive field and facilitate better feature extraction. This is the key to improve the recognition rate. The network also uses the global average pooling layer instead of additional full connection layer, which greatly reduces the number of parameters, and reduces the risk of over-fitting in the case of insufficient training samples. The accuracy of model is higher than that of most existing algorithms on FER-2013 data set. The accuracy rate can reach 0.96 by fine tuning on CK + data set. The model is transformed into the core ML model, and a real-time expression recognition app is built on the iOS side with Xcode platform. It can run stably and smoothly on the iPhone 8 plus, and the recognition effect reaches the expectation.

Keywords: facial expression recognition; conrolutional neural networks(CNN); global average pooling; Google Colab; Core ML

0 引言

面部表情识别正受到越来越多的关注，涉及计算机视觉、机器学习和认知科学等领域[1]，应用于医

[收稿日期] 2020－10－02

[基金项目] 福建省自然科学基金项目（2020J01710）；国家自然科学基金资助项目（41971424）；福建省高校产学研重大项目（2017H6015）；集美大学国家基金培育计划项目（ZP2020063）

[作者简介] 张东晓（1980—），男，博士，副教授，主要研究方向为视频与图像处理、机器学习。E-mail:200661000115@jmu.edu.cn

疗[2]、工业互联网[3]、人机交互[4]、娱乐[5]、虚拟现实[6]、餐饮[7]等方面。一个完整的情绪识别系统通常包括三个部分：人脸区域检测、面部情绪特征提取和使用分类器进行预测。人脸检测技术已经比较成熟，所以相关研究主要集中在后两个方面。传统的研究思路是构造人工特征，如基于降维算法来构建特征[8-10]，基于眼睛、嘴巴、鼻子等部位构建几何特征[11-13]，基于颜色变化构建纹理特征[14-16]；然后采用传统机器学习算法进行分类，如支持向量机（support vector machine，SVM）[17-18]、K－近邻（K－nearest neighbor，KNN）[18]、随机森林[19]等。这些方法可以在特定数据库下取得良好的结果。但人工特征往往具有局限性，导致模型泛化能力不强，在现实任务中很容易受到光照、遮挡等情形的影响。

近几年深度学习在视觉领域取得优异成绩，在表情识别方面也涌现出很多研究成果，如卷积神经网络（convolutional neural networks，CNN）[20-22]、对抗神经网络[23-24]、成熟网络结构的迁移学习[25-26]，以及其他深度网络[27-28]。也有学者尝试将经典方法与深度学习相结合。如夏添等[29]将眉毛、眼睛、鼻子和嘴巴等部位的特征点作为深度网络的输入，张发勇等[30]将深度网络获取的特征用到随机森林模型中，所得结果对视角都具有较强的鲁棒性。由于可以自动学习特征，这些方法均取得良好的效果。

目前公开报道的应用主要集中在 PC 端，但是近几年移动互联网的快速发展也催生了移动端的巨大需求。如：在线教育中，可以通过自动检测学生的表情变化，及时掌握学习情况；购物平台的智能客服可以通过用户表情及时调整对话策略。所有这些应用场景均有两个前提条件——计算资源消耗小和实时性。但是深度学习框架下的主流网络结构往往比较深，需要消耗大量计算资源。受算力限制，这些模型均无法直接移植到移动端。而经典提取特征的方法泛化能力有限，又不能满足移动端的应用需求。因此，本文设计了一个浅层神经网络，期望在确保模型识别率的前提下，能大幅降低运算量。

1 人脸表情数据集及识别模型

1.1 人脸表情数据集

伴随着表情识别研究的兴起，各国学者开始构建表情数据集。最早开始这项工作的是学者 Lyons 等[31]，他们于 1998 年构建了一个日本女性表情数据集（The Japanese Female Facial Expression Database，JAFFE），包括高兴、悲伤、中性、厌恶、生气、害怕及惊讶 7 种不同的面部表情，共计 213 张图像。

目前应用较为广泛的数据集是构建于 2010 年的 The Extended Cohn-Kanade Dataset（CK＋）[32]。该数据集包含来自 123 个人的总计 593 张正面照，它还包括对应于面部表情的每张照片的标签及动作单元编码标签。与 JAFFE 相比，该数据集的样本数量更多，且多了一个“蔑视”表情。

另一个知名数据集是 FER－2013[33]。该数据集是由 Carrier 等于 2013 年建立的公开数据集，数据主要来源于网络检索，总计 35 887 张灰度人脸图像，先后被发布在 Kaggle 和 ICML2013 等平台。其数据是已经裁剪好的人脸灰度图像，大小为 48 px × 48 px。表情种类与 JAFFE 一致，含有 7 种常见的面部表情：0－生气（angry）；1－厌恶（disgust）；2－恐惧（fear）；3－开心（happy）；4－伤心（sad）；5－惊讶（surprise）；6－中性（neutral）。每种表情的样本数量及对应的标签如表 1 所示。

表 1 FER－2013 各表情对应的标签及样本数量

Tab. 1 The label and amount of each expression in FER－2013

标签 Label	数量 Amount	表情 Expression
0	4 593	生气 Angry
1	547	厌恶 Disgust
2	5 121	恐惧 Fear
3	8 989	开心 Happy
4	6 077	伤心 Sad
5	4 002	惊讶 Surprise
6	6 198	中性 Neutral

国内也有学者开展这方面的工作，如吴丹等[34]建立了包括 70 个人的 1 000 段脸部表情视频数据库，为表情识别的深入研究做出贡献。对比这些数据集，FER－2013 有较大规模的数据量，基本可以满足本文所提模型对样本量的需求，所以本文最终选择使用 FER－2013 数据集进行模型的训练。

1.2 表情识别模型

经典机器学习方法在表情识别方面存在泛化能力弱的问题，不能满足移动端需求。而深度学习方

法在移动端又有计算资源的限制，所以本文将搭建一个浅层神经网络。

AlexNet[35]在 ILSVRC2012 的表现引起轰动，随后各种网络结构纷纷被提出，表现也越来越好。基本上沿着两个方向在前行：在纵向不断加深网络，在横向不断拓宽网络。在众多网络结构中，纵向加深的代表是 VGGNet[36]，该网络摒弃了 AlexNet 中 5 ×5 和 7 ×7 卷积核，改用堆积的 3 ×3 卷积核，在 ILSVRC2014 比赛中大放异彩，后续诸多研究表明 VGGNet 在特征提取方面相较于其他深度网络有明显优势，原因可能在于堆积的 3 ×3 卷积核具有较好的感受野。

更深层的网络具有更强的特征表达能力，其分类效果更好。但是深层网络结构的参数也随之大幅增加，这带来两个问题：需要更多的训练样本和大量的计算资源。考虑到公开的表情数据规模较小，且移动端算力有限，本文不直接使用深度网络，而是搭建一个如图 1 所示的浅层 CNN 模型。

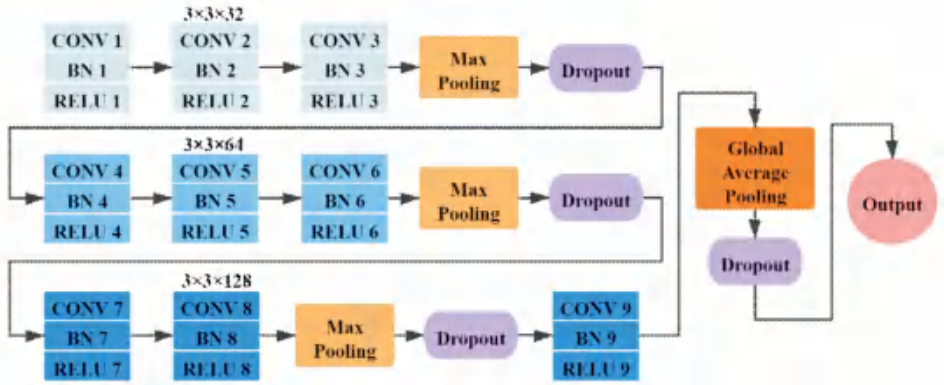

图 1　用于表情识别的网络结构

Fig.1　Network structure for facial expression recognition

在如图 1 所示的网络结构中，本文使用三组堆叠的卷积层来提取面部表情特征。堆叠的卷积层有助于增大感受野，便于更好地提取特征，这是提升识别率的关键。在三组卷积层中，卷积核大小均为 3 ×3，第一组由 3 个卷积核数为 32 的卷积层组成，第二组由 3 个卷积核数为 64 的卷积层组成，第三组由 2 个卷积核数为 128 的卷积层和一个卷积核数为 256 的卷积层组成。这里每组的卷积核数量递增是受 VGGNet 启发，以期获得更丰富的特征。第一组和第二组之后均有尺寸为 2 ×2 的最大池化层，第三组的池化层在中间。每个卷积层之后都使用 ReLU 激活函数 $f(x) = \max(x,0)$，即输入小于 0 时输出 0，否则原样输出。为了加速训练，以及避免学习率过高导致不良影响，在每组卷积层和激活层之间插入了 BN（batch normalization）层[37]。

公开报道的用于表情识别的卷积神经网络中，大多数在卷积层和输出层之间添加 1 个或者更多的全连接层，如文献［20 -22］均采用这种策略。其背后的考量是将末端的每个特征层（最后的卷积层的输出）的每个像素均视作一个神经元。与此截然不同的另外一种方式是 GAP（global average pooling），其思想源自文献［38］，它将末端的每个特征层的平均值作为一个神经元。相对而言，GAP 更符合卷积的工作机理。事实上，经过多次卷积和 Pooling，最后提炼出来的均是高级特征，每个特征层中的像素应该高度相关。将像素单独作为神经元，会破坏它们之间的相关性。GAP 策略正好可以避免这一问题，而且相比单像素策略，还能大幅节约参数。因此，本文选择 GAP 策略，即在第三组卷积后连接一个 GAP。

GAP 后直接连接输出层，其节点数为类别数。为了避免出现过拟合，在每一个最大池化层之后，以及 GAP 层和输出层之间添加一个 Dropout 层[39]，Dropout 比率在 0.3 ~0.6 之间调整。输出层的激活函数选择 Softmax，即 $S(z_i) = e^{z_i} / \sum_{j=1}^{C} e^{z_j}$，其中 z_i是输出层第 i 个神经元尚未激活的值，C 是类别个数。

类别标签表示为 one -hot 形式 $y = [y_1, y_2, \cdots, y_C]$，其中第 i 个标签对应于 $y_i = 1$ 且 $y_j = 0$ $(j \neq i)$。损失函数选择交叉熵损失函数，$L = -\sum_{i=1}^{C} y_i \log_2(S(z_i))$。其中，$y_i$是类别标签，$S(z_i)$ 是输出层第 i 个神

经元的 Softmax 激活值。实际训练时每次随机选取一个批次的样本用于参数更新，此时损失为该批次样本的平均损失。

Adam 优化算法[40]可以高效地求解非凸优化问题，且内存占用少，被广泛应用到深度学习的模型训练中。本文采用该算法反向传播更新参数。

2 模型训练及评价

2.1 数据预处理

FER－2013 数据集的每个样本包括“emotion”“pixels”和“usage”三个部分。其中：“emotion”是表情标签；“pixels”是图像数据；“usage”记录了对应样本的用途——用于训练的“Training”，用于训练过程中进行验证的“PrivateTest”和用于训练完成后测试模型的“PublicTest”。按照 Usage 记录将样本分为三个独立的部分，如表 2 所示。

表 2 数据集的分配

Tab. 2 Distribution of data sets

标签 Label	数量 Length	用途 Usage
Training	28 709	Train set
PrivateTest	3 589	Test set
PublicTest	3 589	Validation set

从表 1 来看，FER－2013 数据集各表情的样本分布不均衡，特别是“厌恶”的样本数只占其他类别数量的 10% 左右，这会严重影响该表情的识别效果。为此，本文先做样本均衡化处理。考虑到样本的特殊性，这里主要使用左右镜像、随机缩放、随机调整亮度、随机旋转等方式增加“厌恶”样本。

数据均衡化后，训练样本约 3 万个。为了增强模型的泛化能力，在训练过程中对样本进行随机增强。即，每读入一个批次的样本，先随机增强然后再输入到网络中进行参数更新。

2.2 训练平台及参数设置

Google Colab 是 Google 公司的一个研究项目，旨在为开发者提供一个云端的深度神经网络训练平台。其向开发者提供型号为 Tesla T4 的 GPU 设备，约 12 GB 的临时 RAM 和约 358 GB 的临时存储空间。这样的配置可以满足前述模型训练需求，所以本文使用该平台进行训练。

众所周知，参数设置对模型训练至关重要。经过尝试，本文的参数设置为：最大池化层后接的 Dropout 比率为 0.3，平均池化层后接的 Dropout 比率为 0.4；每个批次的样本数为 128，每轮训练次数取 224，即训练样本总数 28 709 除以每批样本数 128 再取整；在验证集上每批样本数也为 128；初始学习率设置为 0.01，并对学习率进行动态下调，以保证模型收敛。训练过程中模型的准确率和损失的可视化结果如图 2 所示。

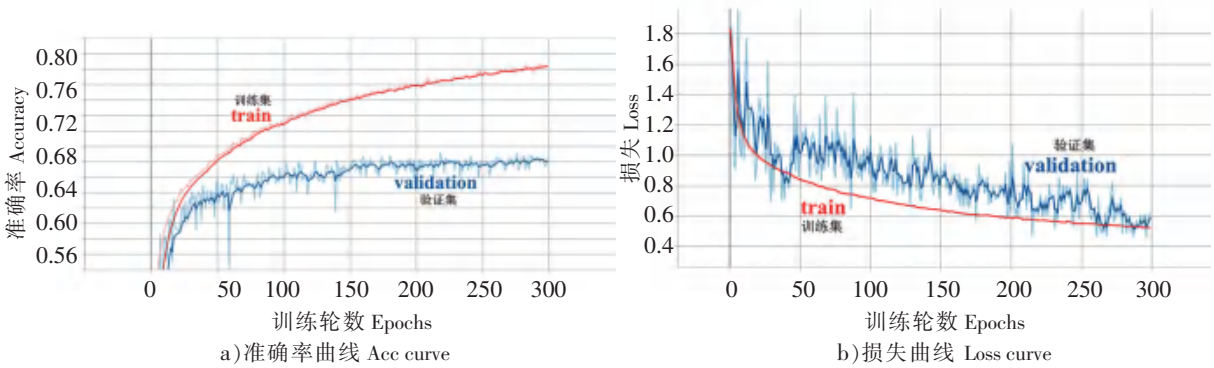

图 2 模型训练的准确率曲线和损失曲线

Fig.2 Acc curve & loss curve of model training

分析图 2 可知，训练集上的准确率随着训练轮数在增加，损失在减少，说明模型训练状况良好；验证集上的准确率和损失在经过 300 轮训练后均趋于平缓，说明此时模型的识别能力已经达到极限，故训练到 300 轮左右时得到最终的模型。

2.3　模型评价

首先考虑模型在 FER－2013 测试集上的表现，为此绘制了如图 3 所示的混淆矩阵图，横轴为模型预测结果，纵轴是真实标签。图 3 是标准化的混淆矩阵（normalized confusion matrix），对角线的数值称为召回率，表示成功预测出该表情的样本量占该表情所有样本的比例，该值越接近 1，对应的矩阵块颜色越深，模型对该表情的分类准确程度越高。从结果来看，该模型在“开心（happy）”和“惊讶（surprise）”这两类表情的分类准确程度最高，“厌恶（disgust）”和“中性（neutral）”次之，其余表情的准确程度较差。

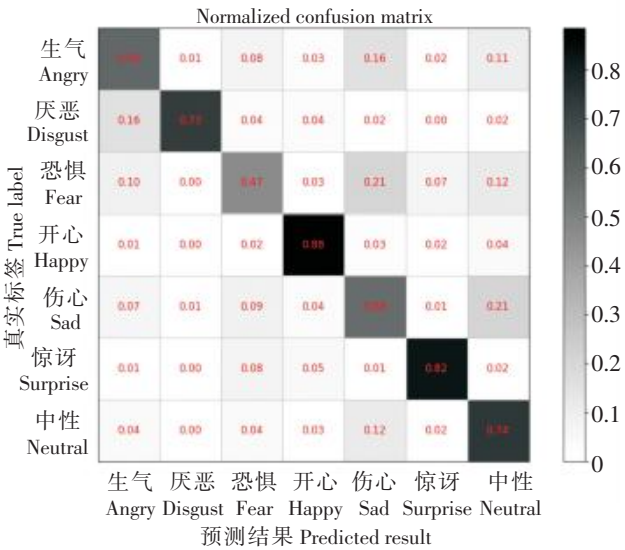

图 3　模型在 FER–2013 测试集上的混淆矩阵

Fig.3　Confusion matrix of the model on the FER–2013 test set

为了探究“恐惧（fear）”“伤心（sad）”和“生气（angry）”表情召回率偏低的原因，进一步分析图 3。从“恐惧（fear）”所在行可以看出，约 21% 的“恐惧（fear）”表情预测为“伤心（sad）”，约 12% 预测为“中性（neutral）”，约 10% 预测为“生气（angry）”，约 7% 预测为“惊讶（surprise）”，约 3% 预测为“开心（happy）”。从每种预测错误的表情中挑选 2 张，结果如图 4 所示。从图 4 来看，与这些样本的真实标签“恐惧（fear）”相比，预测的结果反倒更为合理。

图 4　标注为“Fear”的部分样本（预测结果显示在下方）

Fig.4　Some samples of “fear”(the texts below are the predicted results)

这种情况在标注为“伤心（sad）”和“生气（angry）”的样本中也有出现。如图 5 所示，原本标注为“伤心（sad）”和“生气（angry）”的表情从图像上来看并不符合实际情况，而模型预测的结果却更贴近真实表情。通过查看预测有误的全部样本，得出结论：FER－2013 测试集中关于“恐惧（fear）”“伤心（sad）”和“生气（angry）”的标注有较多错误。这也正是这三种表情召回率偏低的主要原因。

图 5　标注为“sad”（上）和“angry”（下）的部分样本（预测结果显示在下方）

Fig.5　Some samples of “sad”(top) and “angry”(bottom)(the texts below are the predicted results)

当衡量模型在每个类别上的性能时，可以考察每个类别的精确率 P、召回率 R、F_1 值，其计算公式为：

$$P = P_T/(P_T + P_F), R = P_T/(P_T + N_F), F_1 = 2PR/(P + R)。$$

其中：针对某种表情，P_T 表示正确预测为该表情的样本（true positive）数；P_F 表示预测为该表情但不是该表情的样本（false positive）数；N_F 表示是该表情但预测为其他表情的样本（false negative）数；N_T 表示不是该表情且预测结果也不是该表情的样本（true negative）数。

当衡量模型在所有类别上的综合性能时，可以考察准确率 A，$A = (P_T + N_T)/N$。其中，N 为测试样本总数。

模型在 FER－2013 测试集上的各项指标如表 3 所示。

表 3 模型在 FER－2013 测试集上的指标

Tab. 3 Evaluation of model on the FER－2013 test set

表情 Expression	精确率 Precision	召回率 Recall	F_1 F_1 value	样本数 Number of samples
生气 Angry	0.67	0.59	0.63	491
厌恶 Disgust	0.70	0.73	0.71	55
恐惧 Fear	0.60	0.47	0.53	528
开心 Happy	0.90	0.88	0.89	879
伤心 Sad	0.54	0.58	0.56	594
惊讶 Surprise	0.81	0.82	0.82	416
中性 Neutral	0.62	0.74	0.67	626

各表情的精确率和召回率，基本相差不大，说明模型对各种表情没有明显的偏好。至于“恐惧（fear）”和“中性（neutral）”两种表情的数据相差超过 10%，主要是如上所述标签不准所致。整体准确率为 0.7，该值偏低的原因除了上述标签错误以外，还在于 FER－2013 数据集存在异常样本。这里分别从测试集和训练集中针对每个表情选取一个样本，结果如图 6 所示。其中，测试集的“厌恶（disgust）”表情没有发现异常样本。

图 6 每种表情中测试集（上）和训练集（下）中的异常样本

Fig.6 Abnormal samples in the test set (top) and training set (bottom) of each expression

异常的情况有“文字图像”“纯色图像”“图标图像”“少部分人脸”“人脸太小”“面具或者雕塑图像”等，这些异常样本对训练和测试均构成干扰。正如 Goodfellow 等[33]指出的那样：FER－2013 数据集存在标注错误、样本异常等问题，人类的识别率约为 65%±5%。本文所提模型已经达到人类识别率的上限，同时超越 FER－2013 面部表情识别挑战赛的亚军成绩：在 FER－2013 测试集上，并行卷积网络[20]、MobileNets＋Softmax Loss[22]、挑战赛的亚军[33]、挑战赛的冠军[33]，以及本文方法的准确率分别为 0.66、0.69、0.69、0.71、0.70。

为了衡量模型的泛化能力，考虑模型在 CK＋数据集上的表现。CK＋数据集与 FER－2013 数据集重合的表情有 6 种，所以这里只考虑这 6 种表情。为了公平起见，在 CK＋中随机抽取 20% 的样本，运行 3 次取各指标的平均值，准确率为 0.76，其余各指标如表 4 所示。

结果表明，模型在“开心（happy）”和“惊讶（surprise）”上的泛化能力较强，而在其他表情上的泛化能力较弱。原因除了 FER－2013 样本异常以及标注问题以外，还有面部表情本身也具有歧

义性，不同人的不同情绪导致的面部特征可能会大相径庭，例如：“生气（anger）”和“厌恶（disgust）”均可能包含“皱眉”和“抿嘴”等，多数情况下人眼都难以准确辨别。而精确率较高的“开心（happy）”和“惊讶（surprise）”则均含有比较突出的面部特征，如“咧嘴笑”、“嘴部呈 O 形”等。所以该模型的分类效果基本符合预期。

表 4　模型在 CK + 测试集上的指标

Tab. 4　Evaluation of model on CK + test set

表情 Expression	精确率 Precision	召回率 Recall	F_1 F_1 value	样本数 Number of samples
生气 Angry	0.50	0.78	0.61	23
厌恶 Disgust	1.00	0.52	0.68	31
恐惧 Fear	0.38	0.26	0.31	19
开心 Happy	0.92	1.00	0.96	36
伤心 Sad	0.48	0.61	0.54	18
惊讶 Surprise	0.93	0.93	0.93	59

为了排除数据集的影响，更为客观地评估模型的效果，冻结模型的卷积层，只在最后一层进行微调（fine tuning）。随机抽取 CK + 中 80% 的样本作为训练集，在剩余样本上测试，准确率为 0.96，其余指标如表 5 所示。与表 4 对比，各项指标均大幅提升。

表 5　模型微调后在 CK + 测试集上的指标

Tab. 5　Evaluation of model on CK + test set after fine-tuning

表情 Expression	精确率 Precision	召回率 Recall	$F1$ $F1$ value	样本数 Number of samples
生气 Angry	0.83	0.87	0.85	23
厌恶 Disgust	0.91	0.97	0.94	31
恐惧 Fear	1.00	0.95	0.97	19
开心 Happy	1.00	1.00	1.00	36
伤心 Sad	1.00	0.83	0.91	18
惊讶 Surprise	0.98	1.00	0.99	59

在 186 个测试样本中，8 个样本预测错误。这些错误样本如图 7 所示。有 3 个标注为“生气（angry）”的样本预测为“厌恶（disgust）”，有 1 个标注为“厌恶（disgust）”的样本预测为“生气（angry）”。从图 7 上来看，预测为“厌恶（disgust）”的后两个和标注为“厌恶（disgust）”的样本基本没有区别，所以很难下定“预测错误”的结论；比较标注为“伤心（sad）”和“生气（angry）”的样本，特别是第 1 个与标注为“生气（angry）”的样本区分度也不是很大，难怪模型将其预测为“生气（angry）”；将“恐惧（fear）”表情预测为“惊讶（surprise）”的情形，似乎人眼也很难确定这个表情到底是哪一个。这也进一步验证了前文所述表情存在歧义性的事实。

图 7　CK+数据集预测错误的样本(箭头前为真实标签,后为预测结果)

Fig.7　The samplespredicted incorrectly on the CK+ data set
(before the arrow is the real label,and behind the arrow is the predicted result)

3　移植到移动端

3.1　移动端开发环境及开发工具

本文主要针对 iOS 系统阐述开发流程。移动端移植的基础是 Core ML，这是一个由 Apple 公司发

布的机器学习框架，其提供了一系列可以将机器学习模型集成到 iOS App 中的工具。使用 Xcode 进行移动端程序的开发，编程语言选用 Swift，使用 Coremltools 将训练所得的 TensorFlow 模型转换为 Core ML 格式的表情识别模型，用于后续的移动端开发。

3.2 移动端开发流程

步骤 1：搭建平台。在 Xcode 开发平台，建立一个 iOS 工程项目，并创建一个“Single View App”（单视图应用），开发语言选择 Swift。进入开发界面后，对主程序界面进行简单的排版，并添加启动页图片和 App 图标；在“Info. plist”文件中，添加“Privacy-Camera Usage Description”条目，为程序授予摄像头使用权限。

步骤 2：人脸检测。Xcode 中集成的 CIDetector API 是 Core Image 框架中所提供的一个识别器，可以进行包括对人脸、物体、文本等对象的识别。为了简化开发流程，本文直接使用 CIDetector 进行人脸检测。使用时需要实例化一个 CIDetector 对象，并将摄像头获取的图像数据传入到 CIDetector 中，若在图像中检测到人脸，则会返回人脸区域的位置信息，再根据位置信息，对人脸图像裁剪、预处理，为下一步做准备。

步骤 3：表情识别。1）加载 CIDetector 人脸检测器和表情识别模型；2）在 sessionPrepare 函数中，创建摄像头会话，用于捕捉摄像头画面，捕捉到的画面将被输出到 captureOutput 函数中；3）重写 viewDidLayoutSubviews 函数和 viewDidLoad 函数，触发摄像头会话的捕获动作；4）在 captureOutput 函数中，将捕捉到的摄像头画面依次传入人脸检测器和表情识别模型，将检测到的人脸区域和表情识别结果输出到视频帧中；5）定义一个文本标签组件，用于实时显示预测结果。

3.3 移动端程序的运行效果

实时表情识别 App 顺利运行后，将自动开启前置摄像头，从摄像头中实时检测人脸区域，并将表情的预测结果显示在屏幕下方。在 iPhone 8 Plus 上能够流畅运行，效果截屏如图 8 所示。图 8 中，第一幅图是应用启动界面，其余图是程序运行结果。

本研究依次做了“中性（neutral）”“惊讶（surprise）”“开心（happy）”“生气（angry）”“伤心（sad）”“恐惧（fear）”和“厌恶（disgust）”7 种表情，前面 5 种的识别结果与实际表情相符，后两种与实际表情不符。本意做的“恐惧（fear）”表情被识别为“伤心（sad）”，本意做的“厌恶（disgust）”表情被识别为“生气（angry）”。原因一方面是如前所述的模型对这两种表情的识别率偏低，另一方面这两个表情也确实具有歧义性，如果用人眼来判断，也可能将它们判定为“伤心（sad）”和“生气（angry）”。

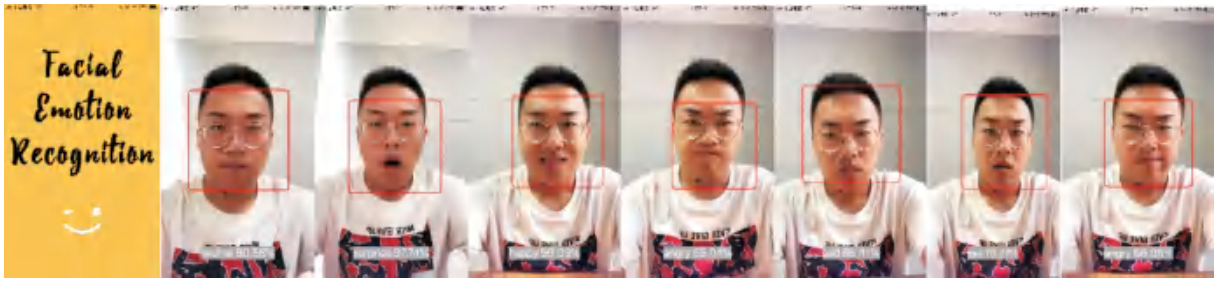

图 8 iOS 端 App 运行示例

Fig.8 Examples of App running on iOS

4 结束语

为了满足移动端的应用需求，本文搭建了一个浅层卷积神经网络，网络结构为三组堆叠的卷积层外加一个全局平均池化层。基于 FER－2013 表情数据集，在 Google Colab 平台使用 TensorFlow 进行训

练，在 FER－2013 测试集和 CK＋数据集上均取得不错的识别效果。使用 Core ML 将训练好的模型移植到 iOS 移动端，在 iPhone 8 Plus 上能够稳定、流畅地运行。FER－2013 数据集上的实验结果表明该数据集存在异常样本和标注错误的情况，这在一定程度上影响了模型的性能，影响到移动端的识别效果。在 CK＋数据集上微调的测试结果表明，规范的数据集可以进一步大幅提升模型的性能，这也说明本文所提模型具有较好的特征提取能力。未来只要有更加规范的数据集，数量足够的各类样本，部署在移动端的应用就会有更好的识别效果，也能更好地应对复杂应用场景。

［参考文献］

［1］叶继华，祝锦泰，江爱文，等. 人脸表情识别综述［J］. 数据采集与处理，2020，35(1)：21-34.

［2］KIM K H，PARK K，KIM H，et al. Facial expression monitoring system for predicting patient's sudden movement during radiotherapy using deep learning［J］. Journal of Applied Clinical Medical Physics，2020，21(8)：191-199.

［3］ZHENG H X，NIU Y H，CHEN J Y，et al. Facial expression recognition of industrial internet of things by parallel neural networks combining texture features［J］. IEEE Transactions on Industrial Informatics，2021，17(4)：2784-2793. DOI：10.1109/TLL.2020.3007629.

［4］TUYEN N T V，ELIBOL A，CHONG N Y. Learning bodily expression of emotion for social robots through human interaction［J］. IEEE Transactions on Cognitive and Developmental Systems，2020(99)：1. DOI：10.1109/TCDS.2020.3005907.

［5］SEKHAVAT Y A，ROOHI S，MOHAMMADI H S，et al. Play with one's feelings：a study on emotion awareness for player experience［J］. IEEE Transactions on Games，2020(99)：1. DOI：10.1109/TG.2020.3003324.

［6］CAO Q，YU H，NDUKA C. Perception of head motion effect on emotional facial expression in virtual reality［C］//2020 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops（VRW）. Atlanta：IEEE，2020：751-752.

［7］SIVABALASELVAMANI D，SOORYA B. Convolution neural network based specialized restaurant rating using facial expression detection［C］//2020 International Conference on Inventive Computation Technologies（ICICT）. Coimbatore：IEEE，2020：739-744. DOI：10.1109/ICICT48043.2020.9112518.

［8］张瑞，蒋晨之，苏剑波. 基于稀疏特征挑选和概率线性判别分析的表情识别研究［J］. 电子学报，2018，46(7)：1710-1718.

［9］MENG N D，CAO N G，HE N Z，et al. Facial expression recognition based on LLENet［C］//IEEE International Conference on Bioinformatics & Biomedicine. Shenzhen：IEEE，2017：1915-1917.

［10］SIDDIQI M H，ALI R，KHAN A M，et al. Human facial expression recognition using stepwise linear discriminant analysis and hidden conditional random fields［J］. IEEE Transactions on Image Processing，2015，24(4)：1386-1398.

［11］ZHAO K，CHU W S，TORRE F D L，et al. Joint patch and multi-label learning for facial action unit detection［C］//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Boston：IEEE，2015：2207-2216.

［12］吴珂，周梦莹，李高阳，等. 基于角度几何特征的人脸表情识别［J］. 计算机应用与软件，2020，37(7)：120-124.

［13］BARMAN A，DUTTA P. Facial expression recognition using distance and shape signature features［J］. Pattern Recognition Letters，2017：S0167865517302246.

［14］JABID T，KABIR M H，CHAE O. Robust facial expression recognition based on local directional pattern［J］. ETRI Journal，2010，32(5)：784-794.

［15］LI H，LI G. Research on facial expression recognition based on LBP and DeepLearning［C］//2019 International Conference on Robots & Intelligent System（ICRIS）. Haikou：IEEE，2019：94-97.

［16］RUBEL A S，CHOWDHURY A A，KABIR M H. Facial expression recognition using adaptive robust local complete pattern［C］//2019 IEEE International Conference on Image Processing（ICIP）. Taiwan：IEEE，2019：41-45.

［17］NIU Z，QIU X. Facial expression recognition based on weighted principal component analysis and support vector machines［C］//International Conference on Advanced Computer Theory & Engineering. Chengdu：IEEE，2010，3：174. DOI：10.1109/ICACTE.2010.5579670.

［18］DINO H I，ABDULRAZZAQ M B. Facial expression classification based on SVM，KNN and MLP classifiers［C］//2019 International Conference on Advanced Science and Engineering（ICOASE）. Zakho：IEEE，2019：70-75.

[19] 罗珍珍，陈靓影，刘乐元，等. 基于条件随机森林的非约束环境自然笑脸检测 [J]. 自动化学报，2018，44(4)：696-706.

[20] 徐琳琳，张树美，赵俊莉. 构建并行卷积神经网络的表情识别算法 [J]. 中国图象图形学报，2019，24(2)：227-236.

[21] LI J，JIN K，ZHOU D，et al. Attention mechanism-based CNN for facial expression recognition [J]. Neurocomputing，2020，411(21)：340-350.

[22] HU L，GE Q. Automatic facial expression recognition based on MobileNetV2 in Real-time [J]. Journal of Physics Conference Series，2020，1549：022136.

[23] 张飞飞，张天柱，毛启容，等. 基于生成对抗网络的多姿态人脸表情识别 [J/OL]. 计算机学报，2019：1-16 [2020-07-28]. http://kns.cnki.net/kcms/detail/11.1826.TP.20191205.1151.002.html.

[24] 姚乃明，郭清沛，乔逢春，等. 基于生成式对抗网络的鲁棒人脸表情识别 [J]. 自动化学报，2018，44(5)：865-877.

[25] KAYA H，GURPINAR F，SALAH A A. Video-based emotion recognition in the wild using deep transfer learning and score fusion [J]. Image and Vision Computing，2017，65：66-75.

[26] NG H W，NGUYEN D，VONIKAKIS V，et al. Deep learning for emotion recognition on small datasets using transfer learning [C] //ACM International Conference on Multimodal Interaction. New York：ACM，2015：443-449.

[27] LI H D，XU H. Deep reinforcement learning for robust emotional classification in facial expression recognition [J]. Knowledge-Based Systems，2020，204：106-172.

[28] LIU D Z，OUYANG X，XU S J，et al. SAANet：Siamese action-units attention network for improving dynamic facial expression recognition [J]. Neurocomputing，2020，413：145-157.

[29] 夏添，张毅锋，刘 袁. 基于特征点与多网络联合训练的表情识别 [J]. 计算机辅助设计与图形学学报，2019，31(4)：552-559.

[30] 张发勇，刘袁缘，李杏梅，等. 基于多视角深度网络增强森林的表情识别 [J]. 计算机辅助设计与图形学学报，2018，30(12)：2318-2326.

[31] LYONS M J，KAMACHI M，GYOBA J，et al. The Japanese female facial expression (JAFFE) database [DB]. 1997. DOI:10.6084/mq.figshare.5245003.

[32] 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. San Francisco：IEEE，2010：94-101.

[33] GOODFELLOWI J，ERHAN D，CARRIER P L，et al. Challenges in representation learning：a report on three machine learning contests [J]. Neural Networks，2015，64：59-63.

[34] 吴丹，林学闾. 人脸表情视频数据库的设计与实现 [J]. 计算机工程与应用，2004，40(5)：177-180.

[35] ALEX K，ILYA S. HG E. Imagenet classification with deep convolutional neural networks [C] //Proceedings of Neural Information Processing System. Lake Tahoe：IEEE，2012：1097-1105.

[36] SIMONYAN K，ZISSERMAN A. Very deep convolutional networks for large-scale image recognition [C]. //Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Boston：IEEE，2015：2207-2216.

[37] IOFFE S，SZEGEDY C. Batch normalization：accelerating deep network training by reducing internal covariate shift [C] //Proceedings of the 32nd International Conference on Machine Learning. Lille：Curran Associates，2015：448-456.

[38] LIN M，CHEN Q，YAN S. Network in network [C] //2nd International Conference on Learning Representations. Banff：ICLR，2014：1-10.

[39] SRIVASTAVA N，HINTON G，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.

[40] KINGMA D，BA J. Adam：a method for Stochastic Optimization [C] //3rd International Conference on Learning Representations. San Diego：ICLR，2015：1-13.

(责任编辑 朱雪莲 英文审校 黄振坤)