

基于多层感知机改进型 Xception 人脸表情识别

韩保金^{1†}, 任福继²

(1. 合肥工业大学 计算机与信息学院, 安徽 合肥 230601; 2. 德岛大学 先端技术科学教育部, 德岛 7708502, 日本)

摘要:针对使用深度学习提取人脸表情图像特征时易出现冗余特征,提出了一种基于多层感知机(MLP)的改进型 Xception 人脸表情识别网络.该模型将 Xception 网络提取的特征输入至多层感知机中进行加权处理,提取出主要特征,滤除冗余特征,从而使得识别准确率得到提升.首先将图像缩放为 48*48,然后对数据集进行增强处理,再将这些经过处理的图片送入本文所提网络模型中.消融实验对比表明:本文模型在 CK+数据集、JAFFE 数据集和 MMI 数据集上的正确识别率分别为 98.991%、99.02% 和 80.339%,Xception 模型在 CK+数据集、JAFFE 数据集和 MMI 数据集上的正确识别率分别为 97.4829%、90.476% 和 74.0678%,Xception+2lay 模型在 CK+数据集、JAFFE 数据集和 MMI 数据集上的正确识别率分别为 98.04%、84.06% 和 75.593%.通过以上消融实验对比,本文方法的识别正确率明显优于 Xception 模型与 Xception+2lay 模型.与其他模型相比较也验证了本文模型的有效性.

关键词:人脸表情识别;卷积神经网络(CNN);多层感知机;Xception;深度可分离卷积

中图分类号:TP 391.41

文献标志码:A

Improved Xception Facial Expression Recognition Based on MLP

HAN Baojin^{1†}, REN Fuji²

(1. School of Computer and Information, Hefei University of Technology, Hefei 230601, China;

2. Graduate School of Advanced Technology and Science, University of Tokushima, Tokushima 7708502, Japan)

Abstract: Aiming at the problem of redundant features when using deep learning to extract facial expression image features, an improved Xception facial expression recognition network based on multi-layer perceptron (MLP) is proposed. In this model, the features extracted from the Xception network are input into the multi-layer perceptron for weighting, the main features are extracted, and the redundant features are filtered out so that the recognition accuracy is improved. First, the image is scaled to 48*48, then the data set is enhanced, and these processed images are fed into the network model proposed in this paper. A comparison of ablation experiments show that: The correct recognition rates of this model on the CK + dataset, JAFFE dataset, and MMI dataset are 98.991%, 99.02% and 80.339% respectively. The correct recognition rates of Xception model on the CK + dataset, JAFFE dataset and MMI dataset are 97.4829%, 90.476%, and 74.0678%, respectively. The correct recognition rates of the Xception + 2lay

* 收稿日期:2021-03-31

基金项目:国家自然科学基金项目(61672202,61673156), National Natural Science Foundation of China(61672202,61673156);国家自然科学基金-深圳联合基金重点项目(U1613217), State Key Program of NSFC-Shenzhen Joint Foundation (U1613217)

作者简介:韩保金(1988—),男,合肥工业大学博士研究生

† 通信联系人:E-mail:hbj1104863230@126.com

model on the CK + dataset, JAFFE dataset and MMI dataset are 98.04% and 74.0678%, 84.06%, and 75.593%, respectively. By comparing the above ablation experiments, the recognition accuracy of this method is significantly better than the Xception model and the Xception + 2lay model. Compared with other models, the effectiveness of this model is also verified.

Key words: facial expression recognition; convolutional neural network (CNN); multilayer perceptron (MLP); Xception; depth separable convolution

人脸表情识别是人类情绪状态识别的有效技术之一,表情识别技术广泛应用于智能问答、在线学习、用户线上体验等智能化人机交互系统中,具有巨大的市场潜力与应用背景,成为了人工智能领域的研究热点^[1].

人脸表情识别根据研究方法的不同,分为传统方法与深度学习方法,传统方法包括 Gabor 小波、局部二值模式 (LBP, Local binary patterns)^[2]、方向梯度直方图 (HOG, Histogram of Gradient) 等. Gabor 小波,可获得最佳的局部化和类人式视觉接收场模型.局部二值模式具有旋转不变性和灰度不变性等显著的优点.方向梯度直方图对图像几何和光学形变都能保持很好的不变性,允许有一些细微的动作变化并不影响检测效果.李文辉等^[3]提出了一种多通道 Gabor 人脸识别方法:依据各通道特征可分离性判断确定特征提取区域,计算通道权值,采用模糊加权规则融合多通道的识别结果.基于特征的中性直方图(包括灰度直方图)特征和典型的表情特征, Mlakar 等^[4]提出一种有效的特征选择系统应用于人脸表情识别系统. Kwong 等^[5]提出了关键人脸检测、显著性映射、局部二值模式和方向梯度直方图的 12 种可能组合,以及 6 种机器学习分类算法,共生成 72 个模型. 伴随而来的又有许多辅助人脸表情识别的方法,如汤红忠等^[6]提出的人脸验证方法,判断是否为同一个身份. 这样为以后处理人脸表情识别时可加上身份验证,从而提高识别正确率. 综上所述,传统方法各有优点且取得了较好的实验结果,但是由于传统方法中提取的特征均在人工选定区域内,这样就使得提取出来的特征只能在特定空间中,在没有额外训练的情况下,很难形成其它特征用于提高人脸表情识别率. 同时,由于人脸姿态、图片光照、摄影角度与不同肤色的人种等各种外界因素的改变对

于识别正确率也造成了一些干扰,为了提高识别正确率需要在实验中加入更多的数据量,以提取充分的信息.

随着计算机技术的发展,计算机软硬件在性能上得到了显著提升,为深度学习的提出与应用创造了环境,其中以卷积神经网络为主的一系列网络变体在理论与实际应用中得到了广泛的研究与应用. 其优势在于不用针对特定的图像数据集或分类方式提取具体的人工特征,而是用类人式的视觉处理机制对图像进行抽象化处理,自动进行特征提取并筛选,这就能实现批量式的图像处理,从而完成了对图像自动化处理的操作,免去了大量的人工劳动且相较于之前的正确率得到了提升. 卷积神经网络是一种有监督的学习模型,具有局部连接、权值共享、下采样的特点,能够有效地挖掘出数据局部特征,对图像的平移缩放、旋转都有较好的稳定性. 它能以原始数据作为输入,通过卷积、池化与非线性激活函数等一系列操作,用于提取数据集中的特征. 常用于图像分类的 CNN (Convolutional Neural Networks)^[7-8] 结构模型种类繁多,如 AlexNet^[9]、VGG、ResNet^[10]、BDBN. AlexNet 网络模型创新性地采用 ReLU 激活函数,加快了模型的收敛速度. VGG-Net 模型使用较小 3*3 卷积核代替大卷积核,同时增加了模型深度. ResNet 模型解决了深度网络的退化问题. TANG 等^[11]提出一种基于表情识别的课堂智能教学评价方法,该方法具有实时性、客观性和细粒度的特点,该方法充分考虑了学生的情绪状态,将情绪状态模型与传统的教学评价方法相结合,利用经典的卷积神经网络 AlexNet 完成了人脸表情识别的预训练,并在相应的数据集上取得了良好的效果.

FEI 等^[12]提出通过一种新的解决方案来处理面部图像并解释情绪的时间演变过程,从 AlexNet 的完

全连通的第 6 层提取深层特征,并利用标准的线性判别分类器来获得最终的分类结果. SARKAR 等提出^[13]一种基于 VGGNet 的卷积神经网络和一种新的处理技术,所提出的方法可显著提高数据集的性能,与不同数据集比较也证明了该方法的优越性. TRIPATHI 等^[14]提出了一种基于语音特征并在聚焦损失下训练的残差卷积神经网络(ResNet)来识别语音情感.

上述研究方法针对情感分类问题从多个方向进行了改进,如网络深度、激活函数、损失函数等,但没有使用能够提取特征信息较丰富的网络结构且对于冗余特征也未做处理.

本文针对人脸表情识别中提取的特征信息丰富度较低与冗余特征未被处理等问题,进行了两个方面的改进:1)选择了能够提取较丰富特征信息的卷积神经网络作为基础;2)增加了多层感知机,通过标定不同特征的权重来提取主要特征,抑制冗余特征.

1 Xception 算法介绍与改进

1.1 Xception 算法介绍

本文主要的研究框架是基于 Xception 模型构建,如图 1 所示,其框架由 3 个主要部分组成,分别为

输入层,中层和输出层.输入层主要作用是用来不断下采样,减少空间维度.中层的主要作用是为了不断学习关联关系,优化特征.输出层的主要作用为最终汇总,整理特征,交由全连接层(FC, fully connected layer)进行表达.

Xception 算法是在 inceptionv3 基础上进行改进的,把 inceptionv3 中的 3*3 模块全部换成了 depthwise separable convolution(深度可分离卷积).通常标准卷积操作将特征图的空间相关性与通道间相关性一并处理,而深度可分离卷积则将空间与通道信息处理过程完全分开. Depthwise 卷积的主要作用

为将每个输入特征通道单独卷积,若输入特征图数量为 n,卷积核大小为 m*m,这样每个输入特征图都将对应一个独立的 m*m 卷积核进行卷积,输出 n 个特征图.而 Pointwise 卷积使用 1*1 的标准卷积来关联特征通道之间的相关性输出特征.其结构如图 2 所示.

深度可分离卷积可以在保留较高准确率的情况下减少大量的模型参数和计算量.虽然深度可分离卷积减少了参数量,但是 Xception 模型的总参数量与 InceptionV3 相差不大,主要原因为 Xception 模型旨在提高分类效果,在网络其他位置增加了参数量.

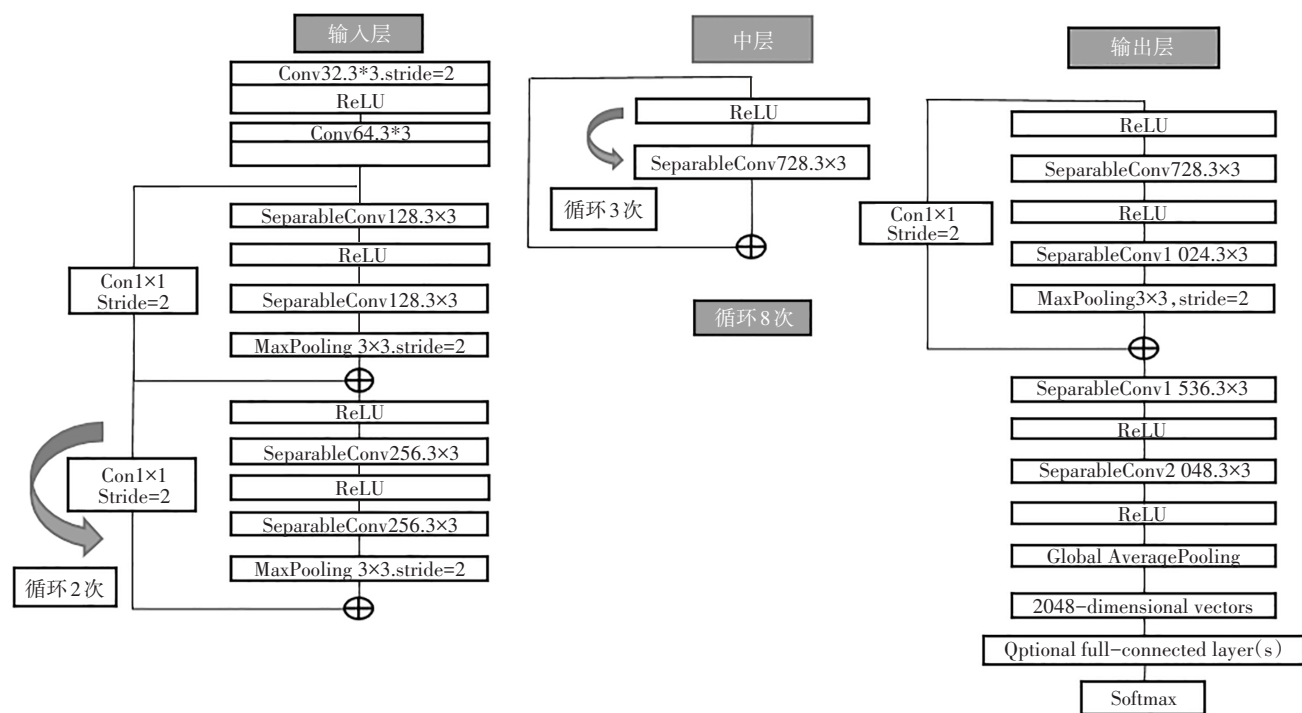


图 1 Xception 模型的主要网络模块示意图^[15]

Fig.1 Diagram of main network module of Xception model^[15]

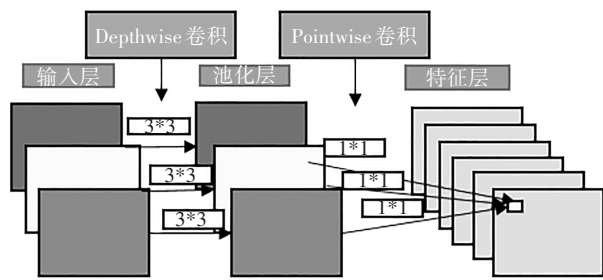


图2 深度可分离卷积结构图^[16]
Fig.2 Architecture of depth separable convolution^[16]

1.2 多层感知机

多层感知机(multilayer perceptron, MLP)由感知机发展而来,其主要特征是有多个神经层.其基本结构包括输入层、隐含层与输出层,其隐含层的数量可多可少,输入层到隐含层可看作一个全连接层,隐含层到输出层可看作一个分类器.

图3所示的多层感知机模型中,输入和输出个数分别为4和3,中间的隐藏层中包含了5个隐藏单元(hidden unit).由于输入层不涉及计算,所以图3中的多层感知机的层数为2.由图3可见,隐藏层中的神经元和输入层中各个输入完全连接,输出层中的神经元和隐藏层中的各个神经元也完全连接.因此,多层感知机中的隐藏层和输出层都是全连接层.

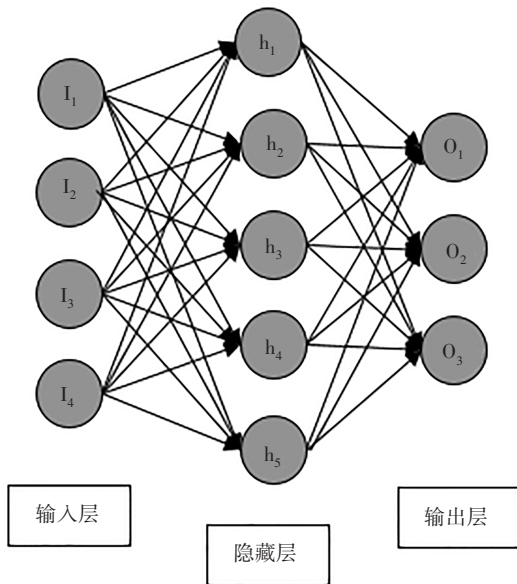


图3 多层感知机结构图
Fig.3 Structure diagram of multilayer perceptron

1.3 采用MLP改进Xception模型

由图1,输入层在不断下采样,减少空间维度,同时也在保持原始特征的学习,这样就可以提取更深

层的特征避免了原始特征的丢失.中层在不断学习关联关系,优化特征,同时也在学习输入层中所学习的特征,如此既能学习到中层特征也能学习到输入层中的特征,进而能够获得较丰富的特征信息.由图3可以直观看出多层感知机的结构,输入层与隐藏层直接相连,隐藏层与输出层直接相联,而且多层感知机中隐藏层的参数随着训练的进行随时变动,更加有利于输入层信息量与输出层信息量达到尽量相一致,同时又可以对特征进行加权处理,从而加强了重要特征的权重.

将由Xception网络模型提取出的特征向量送入多层感知机中,并利用softmax函数对所提取的特征向量进行权重学习,从而得到一组最优的权重分布.通过训练集对网络进行有监督的训练,不断地学习图像中的内容.在训练过程中,通过不断地调整MLP中的层数与层内的各参数,最终确定在层数为3时,整个模型效果最好,所以本文提出了[(Xception+3lay)如下图4(a)所示的网络结构],为对比网络效果,还作出了图4(b)(Xception+2lay)的结构作为一种对比网络结构.

2 实验与分析

2.1 实验数据集

为了评估本文提出的算法,本节将在三个公开的面部表情数据集上进行实验,这三个数据集分别是日本女性面部表情库(JAFFE)、扩展的Cohn-kanade库(CK+)表情库和MMI数据集.由于本文实验针对静态图像,因此截取视频序列((CK+)表情库与MMI表情库)的表情变化的三个峰值作为图像样本,及对JAFFE数据库的表情对其眼部周围添加噪声,所有图片缩放为48*48.如图5所示为处理后照片的形式,前三张为(CK+)表情库处理后的照片形式,中间三张为JAFFE表情库处理后的照片形式,后三张为MMI表情库处理后的照片形式.如表1所示,各个表情数据集的分布与对应情感类图像数量,其中CK+图像总数量为981张,JAFFE图像总数量为639张.MMI图像总数量为609张,形成的总图片数为2229张.

实验中,在训练阶段,采用随机切割44*44的图像,并将图像进行随机镜像,然后送入训练.在测试

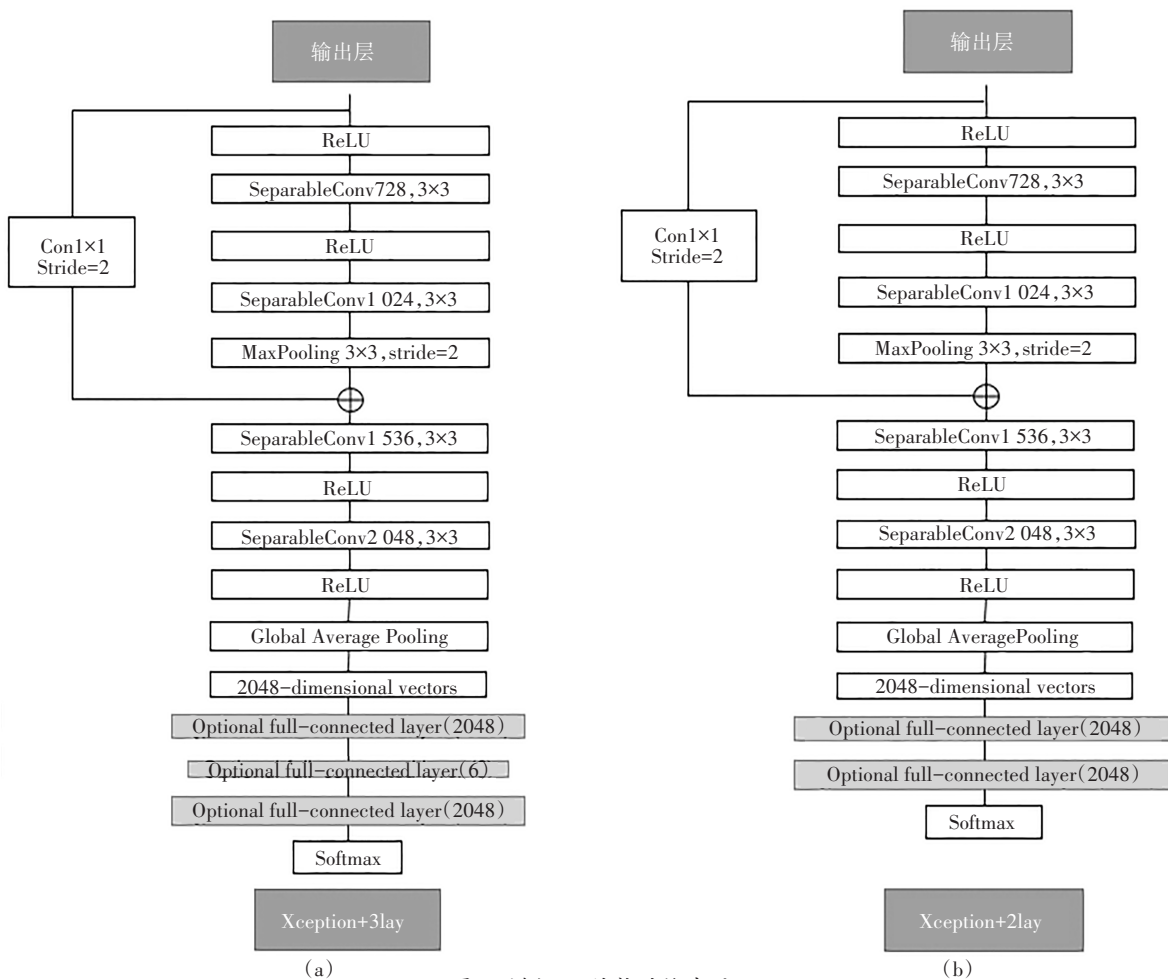


图 4 增加 lay 结构的输出层
Fig.4 Add the output layer of lay structure



图 5 处理后照片的形式
Fig.5 Form of processed photos

阶段,避免训练集中的图像进入测试集中.将图片在左上角、左下角、右上角、右下角、中心进行切割并做镜像操作,这样的操作使得所要训练的数据集的数量扩大了10倍,再将这10张图片送入模型中.然后将得到的概率值取平均,最大的值即为对应表情类别,这种方法有效地降低了误识别率.

表 1 JAFFE、CK+和MMI样本分布

Tab. 1 Sample distribution of JAFFE、CK+ and MMI

| 表情类别 | 高兴 | 生气 | 悲伤 | 厌恶 | 惊讶 | 害怕 | 中性 | 蔑视 |
|-------|----|----|----|----|----|----|----|----|
| JAFFE | 31 | 30 | 30 | 31 | 31 | 30 | 30 | 0 |
| CK+ | 69 | 45 | 28 | 59 | 83 | 25 | 0 | 18 |
| MMI | 42 | 32 | 32 | 28 | 41 | 28 | 0 | 0 |

2.2 实验环境与设置相关参数

实验需用相关环境如下:操作系统为 18.04.1-Ubuntu 版本,显卡为 12 G 显存 Nvidia GeForce GTX1080Ti 一块,CPU 为 Intel(R)Xeon(R) CPU E5-2620 v3@ 2.40 GHz,Python 版本为 Python3.6.10,深度学习框架安装 pytorch1.6.0, TensorFlow 版本为 1.14.0.实验使用 GPU 加快模型计算速度,减少训练时间,选择小批次带动量参数的随机梯度下降法(stochastic gradient descent,SGD)作为模型参数优化器(Optimizer).学习率更新采用固定周期缩减策略,将初始学习率设置为 0.01;其余相关设置见表 2,本实验所需相关参数是在实验中不断调试后才最终确

定.为了扩大数据集并增强模型泛化能力,本文将采用十折交叉法,将JAFFE、CK+和MMI数据集上的样本总量大致平均分为10份,每次试验从中选取其中9份即样本总量的90%作为训练样本,剩余的被用作测试.

表2 相关参数设置

Tab.2 Related parameter settings

| 设置项 | 设置 |
|-----------------------|----------|
| 随机梯度下降法 | 动量参数为0.9 |
| 损失函数 | 交叉熵损失 |
| 训练周期 | 60次 |
| 每批次训练的图像数量(Batchsize) | 32张 |

2.3 消融实验

为了验证本文所提出方法的有效性,将本文方法与Xception和Xception+2lay作对比分析.1)在三个数据集上使每个模型由随机参数开始训练,直至收敛,并且保证三个模型在相同条件下完成训练.2)在模型训练过程中,训练集每迭代完一个周期,对测试数据集进行一次测试,达到所需训练周期后,迭代终止.3)为了保证实验数据的可靠性,实验重复进行了十次,计算平均值作为识别结果,表3为对比数据的结果.

表3 消融实验对比

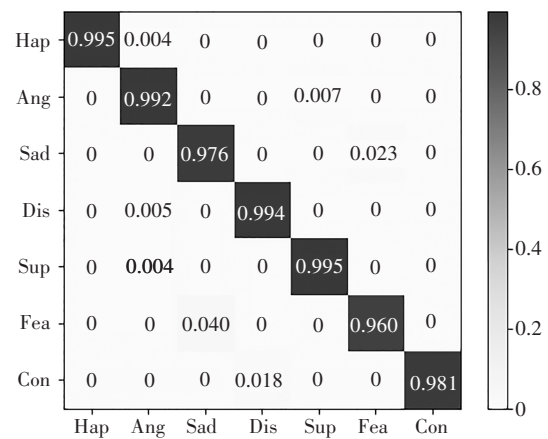
Tab.3 Comparison of ablation experiments

| 分支 | CK+ | JAFFE | MMI |
|---------------|---------|--------|---------|
| Xception | 97.4829 | 90.476 | 74.0678 |
| Xception+2lay | 98.04 | 84.06 | 75.593 |
| 本文 | 98.991 | 99.02 | 80.339 |

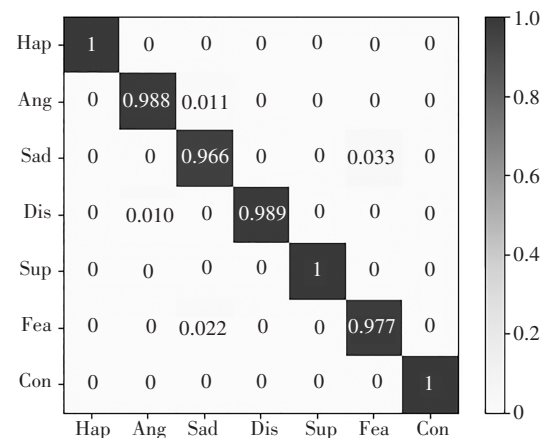
由表3可以得出,三个数据集在Xception、Xception+2lay与本文所提算法表现都良好,但前两个网络没有本文所提网络的效果好.从单个分支来看,在CK+数据集上Xception、Xception+2lay与本文提出的算法相差不大且分类正确率是递增的状态,可推测多层感知机确实能提高特征信息利用率.三个网络在JAFFE数据集上的表现差异较大.前两个网络结构表现不佳,一方面是由于JAFFE数据集本身数据量较少,提取特征信息中含有较多冗余信息,另一方面也可得出前两个网络无法有效地除去冗余特征,导致识别正确率表现不佳.而本文所提出的网络结

构中含多层感知机,能够有效减少冗余特征,所以即使在有较多冗余特征的情况下也能表现出较好的结果.在MMI数据集上,总体识别率并不高,一方面是由于个体表情之间差异较大,另一方面可能是由于遮挡物(如化妆、配戴眼镜)导致的识别率总体一般.但是本文所提方法与前两种方法相比,具有明显优势,这就从侧面说明多层感知机能提取重要特征,提高正确识别率.

由图6可知,在CK+数据集总体正确识别率为98.9%,在JAFFE数据集总体正确识别率为98.9%,在MMI数据集总体正确识别率为80.1%.在表情分类中,高兴与惊讶的识别效果最好,而害怕与伤心表情的识别效果较低,出现这种结果,主要是由于高兴与惊讶表情的变化较为明显,而害怕与悲伤的变化不太明显,图像中的表情反应的特征上可以看出,高兴与惊讶表情图像所提取的特征差异较大,易于区分,从而为以后识别正确率高奠定了基础.而害怕与悲伤图像所提取的特征差异较小,在分类过程中易出现混淆,这样就导致其分类正确率相对较低一些.



(a)CK+



(b)JAFFE

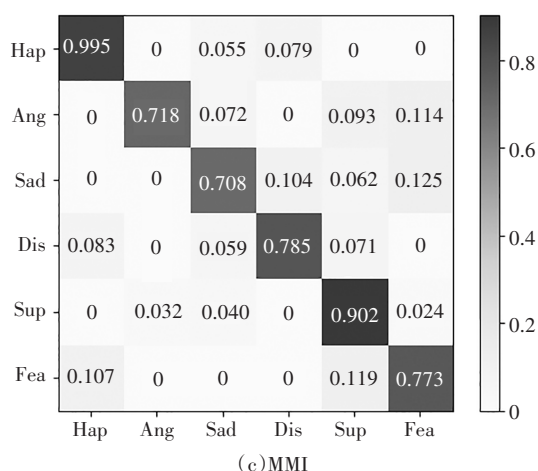


图 6 CK+, JAFFE, MMI 表情分类混合矩阵

Fig.6 Mixed matrix of CK+, JAFFE and MMI expression classification

CK+数据集与 JAFFE 的正确率相对于 MMI 数据集要高一些,这是因为 MMI 数据集中人脸遮挡物所致和个体表情之间差异较大的原因所致。

2.4 与其它方法对比

本次采用了不同方法与本文方法进行对比,表 4 给出了对比结果。从表 4 中可看出 Lenet-5 在 CK+、JAFFE 与 MMI 正确率分别为 85.31%、82.67% 与 68.644%;本文所提算法正确率最高,由此可得本文所提算法是一种性能优良的网络模型。文献[2]提出了一种基于特征局部纹理编码算子——中心对称局部梯度编码,然后通过训练一台极限学习机来估计预期的值。这样通过局部纹理与极限学习相结合,可以提高局部特征利用率,但是由于极限学习的训练速度慢,易陷入局部极小值对于学习率较敏感,所以在训练后的效果上可能提升不高。其中文献[7]提取图像的多对角 HOG 特征,并与 CNN 合并提出了一种优化算法的分类器融合方法,通过迭代取得了很好的结果。文献[8]提出了基于静态图像的双通道加权混合深度卷积神经网络。文献[9]提出了一种基于多方向梯度计算 HOG (moo-HOG)特征和深度学习特征的分类器迭代融合了面部表情识别方法 (AlexNet)。文献[7]、[9]都是通过传统方法与卷积神经网络结合起来,由于提取的特征过于单一,所以可以适当改进提取特征方法。文献[17]提出一种表达式分类利用概率图模型提高小尺度样本集学习算法的准确性的算法。文献[18]提出了一种基于奇异值分解的共聚类特征选择策略,利用该策略寻找表

情特征中具有较高识别能力的最显著区域。文献[19]提出了一种基于局部的分层双向递归神经网络 (PHRNN)来分析时序序列的面部表情信息。文献[20]提出了一种具有注意机制的卷积中立网络 (CNN),它能够感知人脸的遮挡区域,并将焦点集中在最具辨别性的非遮挡区域,针对不同的 ROI,介绍了 ACNN 的两种版本:基于补丁的 ACNN (pACNN)和基于全局局部的 ACNN (gACNN)。pACNN 只关注局部面片。gACNN 将补丁级的局部表示与图像级的全局表示集成在一起。文献[21]提出了一种期望最大化算法来可靠地估计情感标签,揭示了现实世界中的人脸往往表现出复合甚至混合的情感。文献[22]提出了一个多任务的深度框架,借助关键点特征识别人脸表情。

表 4 与其他方法对比表

Tab. 4 Comparison with other methods

| 算法 | CK+ | JAFFE | MMI |
|-------------|--------|---------|---------|
| Lenet-5 | 85.31 | 82.67 | 68.644 |
| AlexNet | 87.28 | 84.9 | 65.0847 |
| VGGNet | 94.4 | 86.7 | 63.235 |
| ResNet | 92.01 | 96.57 | 72.712 |
| GoogleNet | 97.89 | 98.6431 | 73.5594 |
| InceptionV2 | 95.482 | 83.4832 | 62.0338 |
| InceptionV3 | 94.024 | 88.285 | 66.1016 |
| [2] | 97.64 | 97.7 | |
| [7] | 97.18 | 97.38 | |
| [8] | 92.3 | 98.5 | |
| [9] | 97.64 | 97.7 | |
| [17] | 97.78 | 98.95 | |
| [18] | 95.31 | 96.05 | |
| [19] | | | 77.1 |
| [20] pACNN | | | 70.4 |
| [20] gACNN | | | 69 |
| [21] | | | 78.5 |
| [22] | | | 78.44 |
| 本文 | 98.991 | 99.02 | 80.339 |

3 结论

本文所提出的方法,在 CK+、JAFFE 和 MMI 数据集上的实验结果表明,相比于其他常用模型,本文方法的识别效果最好;在分支消融实验结果与其它方

法的结果对比中发现,Xception网络对于表情分类来说确实有一定的优势,增加了多层感知机后可以有效提高分类正确率.以上都是针对静态图片,而随着视屏广泛在日常生活中出现,如何把语音与图像相结合进行情感分析将是下一步要做的工作.

参考文献

- [1] ALHARBI M, HUANG S H. A survey of incorporating affective computing for human-system co-adaptation [C]//Proceedings of the 2020 The 2nd World Symposium on Software Engineering. New York, NY, USA : ACM, 2020 : 72-79.
- [2] YANG J, WANG X, HAN S, *et al.* Improved real-time facial expression recognition based on a novel balanced and symmetric local gradient coding [J]. *Sensors(Basel)*, 2019, 19(8) : 1899.
- [3] 李文辉, 高璐, 林逸峰, 等. 特征选择模糊加权多通道 Gabor 人脸识别[J]. *湖南大学学报(自然科学版)*, 2013, 40(4):87-93.
LI W H, GAO L, LIN Y F, *et al.* Feature selection fuzzy weighted multi-gabor face recognition [J]. *Journal of Hunan University (Natural Sciences)*, 2013, 40(4) : 87-93. (In Chinese)
- [4] MLAKAR U, FISTER I, BREST J, *et al.* Multi-objective differential evolution for feature selection in facial expression recognition systems [J]. *Expert Systems With Applications*, 2017, 89 : 129-137.
- [5] KWONG J C T, GARCIA F C C, ABU P A R, *et al.* Emotion recognition via facial expression: utilization of numerous feature descriptors in different machine learning algorithms [C]//TENCON 2018 - 2018 IEEE Region 10 Conference. October 28-31, 2018. Jeju, Korea (South). IEEE, 2018 : 2045-2049.
- [6] 汤红忠, 张小刚, 陈华, 等. 结合加权子空间和相似度度量学习的人脸验证方法研究 [J]. *湖南大学学报(自然科学版)*, 2018, 45(2) : 152-160.
TANG H Z, ZHANG X G, CHEN H, *et al.* Face verification based on weighted subspace and similarity metric learning [J]. *Journal of Hunan University (Natural Sciences)*, 2018, 45(2) : 152-160. (In Chinese)
- [7] 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*, 2020, 36(2) : 391-404.
- [8] ZHANG H P, HUANG B, TIAN G H. Facial expression recognition based on deep convolution long short-term memory networks of double-channel weighted mixture [J]. *Pattern Recognition Letters*, 2020, 131 : 128-134.
- [9] KRIZHEVSKY A, SUTSKEVER I, HINTON G E. ImageNet classification with deep convolutional neural networks [J]. *Communications of the ACM*, 2017, 60(6) : 84-90.
- [10] YANG J X, ADU J H, CHEN H G, *et al.* A facial expression recognition method based on dlib, RI-LBP and ResNet [J]. *Journal of Physics: Conference Series*, 2020, 1634(1) : 012080.
- [11] TANG X Y, PENG W Y, LIU S R, *et al.* Classroom teaching evaluation based on facial expression recognition [C]//Proceedings of the 2020 9th International Conference on Educational and Information Technology. Oxford United Kingdom. New York, NY, USA : ACM, 2020 : 62-67.
- [12] FEI Z X, YANG E F, LI D D U, *et al.* Deep convolution network based emotion analysis towards mental health care [J]. *Neurocomputing*, 2020, 388 : 212-227.
- [13] SARKAR R, CHOUDHURY S, DUTTA S, *et al.* Recognition of emotion in music based on deep convolutional neural network [J]. *Multimedia Tools and Applications*, 2020, 79(1/2) : 765-783.
- [14] TRIPATHI S, KUMAR A, RAMESH A, *et al.* Focal loss based residual convolutional neural network for speech emotion recognition [EB/OL]. 2019 : arXiv: 1906.05682 [eess. AS]. <https://arxiv.org/abs/1906.05682>.
- [15] CHOLLET F. Xception: deep learning with depthwise separable convolutions [C]//2017 IEEE Conference on Computer Vision and Pattern Recognition. July 21-26, 2017, Honolulu, HI, USA. IEEE, 2017 : 1800-1807.
- [16] HOWARD A G, ZHU M L, CHEN B, *et al.* MobileNets: efficient convolutional neural networks for mobile vision applications [EB/OL]. 2017 : arXiv: 1704.04861 [cs. CV]. <https://arxiv.org/abs/1704.04861>.
- [17] SUN J G, LI T, YAN H, *et al.* Research on an expression classification method based on a probability graph model [J]. *Multimedia Tools and Applications*, 2020, 79(45/46) : 34029-34043.
- [18] KHAN S, CHEN L J, YAN H. Co-clustering to reveal salient facial features for expression recognition [J]. *IEEE Transactions on Affective Computing*, 2020, 11(2) : 348-360.
- [19] ZHANG K H, HUANG Y Z, DU Y, *et al.* Facial expression recognition based on deep evolutionary spatial-temporal networks [J]. *IEEE Transactions on Image Processing: a Publication of the IEEE Signal Processing Society*, 2017, 26(9) : 4193-4203.
- [20] LI Y, ZENG J, SHAN S, *et al.* Occlusion aware facial expression recognition using CNN with attention mechanism [J]. *IEEE Transactions on Image Processing: a Publication of the IEEE Signal Processing Society*, 2018 : 2018Dec14.
- [21] LI S, DENG W H. Reliable crowdsourcing and deep locality-preserving learning for unconstrained facial expression recognition [J]. *IEEE Transactions on Image Processing: a Publication of the IEEE Signal Processing Society*, 2019, 28(1) : 356-370.
- [22] 王善敏, 帅惠, 刘青山. 关键点深度特征驱动人脸表情识别 [J]. *中国图象图形学报*, 2020, 25(4) : 813-823.
WANG S M, SHUAI H, LIU Q S. Facial expression recognition based on deep facial landmark features [J]. *Journal of Image and Graphics*, 2020, 25(4) : 813-823. (In Chinese)