

多分支精简双线性池化的人脸表情识别

王彬¹,徐杨^{1,2*},石进¹,张显国¹

(1. 贵州大学大数据与信息工程学院, 贵州 贵阳 550025;

2. 贵阳铝镁设计研究院有限公司, 贵州 贵阳 550009)

摘要:针对人脸表情识别研究中特征提取不充分、难以辨别人脸表情细微的类间差异等问题,提出了一种多分支精简双线性池化的人脸表情识别方法。该方法以 ResNet-18 为基础,在避免大幅度增加计算复杂度的前提下提升 ResNet-18 的特征提取能力,提出了一个新的多样化分支块(diverse branch block)对 ResNet-18 进行改进;为使改进后的 ResNet-18 更方便地聚焦人脸图像中产生表情区域的特征,提出了残差空间注意力;为了减少人脸表情细微的类间差异带来的不利影响,增强人脸表情类间的区别性,设计了多分支精简双线性池化结构。最后用所提的方法分别在公开的人脸表情数据集 CK+、RAF-DB 进行实验,识别率分别达到了 98.46%、82.99%。实验结果表明,该方法的识别率优于 DLP-CNN、MA、DeepExp3D 等诸多的表情识别方法,具有一定的竞争性。

关键词:人脸表情识别;多样化分支块;残差空间注意力;多分支精简双线性池化;ResNet-18

中图分类号:TP391.4

文献标识码:A

文章编号:1673-629X(2023)03-0027-07

doi:10.3969/j.issn.1673-629X.2023.03.005

Multi-branch Compact Bilinear Pooling for Facial Expression Recognition

WANG Bin¹, XU Yang^{1,2*}, SHI Jin¹, ZHANG Xian-guo¹

(1. School of Big Data and Information Engineering, Guizhou University, Guiyang 550025, China;

2. Guiyang Aluminum Magnesium Design and Research Institute Co., Ltd., Guiyang 550009, China)

Abstract: Aiming at the problems of insufficient feature extraction and difficulty in distinguishing subtle inter-class differences in facial expression recognition in the study of facial expression recognition, a multi-branch compact bilinear pooling facial expression recognition method is proposed. Based on ResNet-18, this method improves the feature extraction capability of ResNet-18 without greatly increasing the computational complexity, and proposes a new diverse branch block to improve ResNet-18. To make the improved ResNet-18 more convenient to focus on the features of expression regions in face images, the residual spatial attention is proposed. In order to reduce the adverse effects of subtle inter-class differences in facial expressions and enhance the distinction between facial expression classes, a multi-branch compact bilinear pooling structure is designed. Finally, the proposed method was used to conduct experiments on the public facial expression datasets CK+ and RAF-DB, and the recognition rates reached 98.46% and 82.99%, respectively. The experimental results show that the recognition rate of the proposed method is better than that of DLP-CNN, MA, DeepExp3D and many other expression recognition methods, with certain competitiveness.

Key words: facial expression recognition; diverse branch block; residual spatial attention; multi-branch compact bilinear pool; ResNet-18

0 引言

人脸表情作为一种非语言的交流方式,由于它承载信息量胜过语言,因此在人类日常沟通交流中有至关重要的作用。近年来,人脸表情识别在汽车安全驾驶^[1]、人类情绪认知^[2]等方面应用广泛,使其成为计算机视觉研究领域的热点之一。

人脸表情识别包括三个步骤:人脸检测、人脸特征

提取、表情分类。人脸检测作为人脸表情识别研究的基础,主要是从图像和视频中捕捉人脸。最具有代表性的检测方法为 Zhang 等人^[3]提出的多任务级联卷积神经网络(Multi-Task Cascade Convolutional Network, MTCNN)人脸检测模型。人脸特征提取作为人脸表情识别最重要的一部分,常见的特征提取方法有:局部二值模式(Local Binary Pattern, LBP)^[4]、E-Gabor 过

收稿日期:2022-05-18

修回日期:2022-09-21

基金项目:贵州省科技计划项目(黔科合支撑[2021]一般176)

作者简介:王彬(1997-),男,硕士研究生,研究方向为计算机视觉、机器学习;通讯作者:徐杨(1980-),男,博士,副教授,研究方向为数据采集、机器学习。

滤波器^[5]等。对于人脸表情识别的最后一步,常采用支持向量机(Support Vector Machines, SVM)^[6]、随机森林(NFS)^[7]等方法。

传统的人脸表情识别方法大多数为手工提取特征,而手工特征提取器容易丢失人脸关键位置的特征信息,近年来逐渐被卷积神经网络(Convolutional Neural Networks, CNN)所替代。Kuo 等人^[8]将一种精简的 CNN 模型用于人脸表情识别,然而随着卷积神经网络层数不断增加会产生梯度爆炸和过拟合现象。为了克服该问题,He 等人^[9]提出了残差网络(Residual Network, ResNet)。在此基础上,石敏等人^[10]用 3×3 和 5×5 的卷积层替换 ResNet-18 的 7×7 卷积层,并在残差块中加入注意力机制以改进残差网络,提升了人脸表情的识别率。高涛等人^[11]将改进后残差网络用于人脸表情识别,并在公开数据集 CK+ 上识别率达到 93.96%,但该方法未研究人脸表情细微的类间差异对模型识别率、鲁棒性的影响。付小龙等人^[12]以 ResNet-18 为基础研究不同的损失函数,以减少类内差异,增大类间差异,取得了不错的识别效果。通过对上述文献的研究发现,提取充分的人脸特征和辨别人脸表情样本之间细微的差异至关重要。

为此,该文在 ResNet-18 基础上进行改进,提出了多分支精简双线性池化网络(Multi-branch Compact Bilinear Pooling network, MCBP)。主要改进如下:

(1) 提出了残差空间注意力,该注意力对输入图

像不同维度的像素赋予不同的权重,使网络更好地凸显人脸关键区域的特征。

(2) 在 ResNet-18 的基础上,设计了 DBB-Res18 网络架构,使网络能获得更丰富的人脸表情。

(3) 设计了一个多分支精简双池化结构。该结构通过将精简双线性池化引入到 DBB-Res18 的多个高层卷积分支中,增强不同尺度的表情图像的内聚性。最后融合不同分支高层卷积的特征信息,提升网络对人脸表情类间差异的分辨能力。

1 多分支精简双线性池化网络

该文在 DBB-Res18 基础上提出了 MCBP 网络,架构如图 1 所示。首先,选取 DBB-Res18 模型中的高层卷积 res4_1、res4_2、res5_1、res5_2 构成不同尺度的多分支特征支路。其次,将残差空间注意力引入每个分支,为避免残差空间注意力加入后模型收敛速度变慢,对残差空间注意力机制输出进行批归一化处理。经过上述操作后形成携带人脸关键区域不同尺度的特征: F1、F2、F3、F4。最后,将精简双线性池化引入到多分支之路中,同时将不同分支的特征融合后送入到 softmax 分类器中进行分类, softmax 函数如式(1)所示。

$$\text{Softmax}(z_i) = \frac{e^{z_i}}{\sum_{c=1}^c e^{z_c}} \quad (1)$$

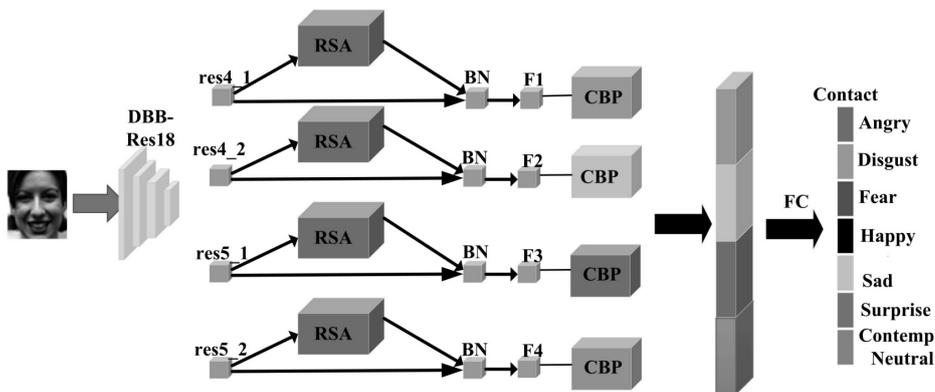


图 1 MCBP 网络

1.1 DBB-Res18 网络

常见的人脸表情之间存在着极大的相似性,而这一现象大部分是由人脸表情类别之间的细微的差异造成,要解决这一问题,提取丰富的表情特征成为了至关重要的工作。为了增强单个卷积的表征能力,Ding 等^[13]提出了多样化分支块(Diverse Branch Block, DBB)。DBB 通过组合不同规模和复杂度的多样化分支来获得丰富的特征空间,以代替单个卷积层。如图 2 所示,六种不同的 DBB 转换组合等效为单个 $k * k$ 的卷积层。

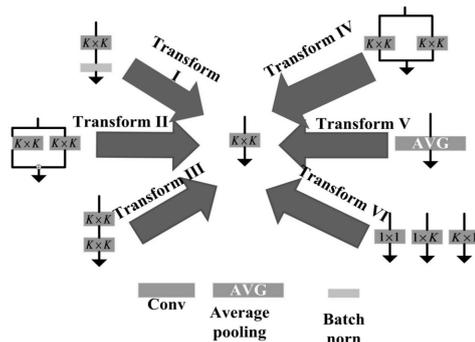


图 2 DBB 六种转换组合

依据六种转换组合,该文设计了一个 DBB 结构,如图 3 所示,以代替基础的单个卷积层。从图 3 可知, DBB 结构由 1×1 卷积-BN 组合、两个 1×1 卷积-平均池化-BN 组合以及一个 $k \times k$ 卷积-BN 组合进行并行多尺度融合组成。经图 2 的六种转换组合作用后,可以将 DBB 等效地转换为 $k \times k$ 的卷积层,如图 3 右侧所示。

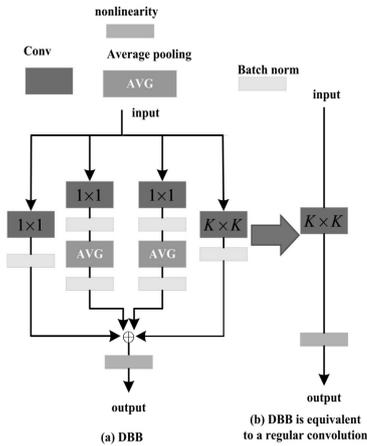


图 3 多样化分支块

为进一步丰富人脸表情图像的特征空间,增强 ResNet-18 的特征提取能力,使用图 2 中设计的 DBB 结构改进 ResNet-18 的所有瓶颈块中的 3×3 的卷积,构成 DBB-Res18 网络。DBB-Res18 网络的瓶颈块如图 4 所示。

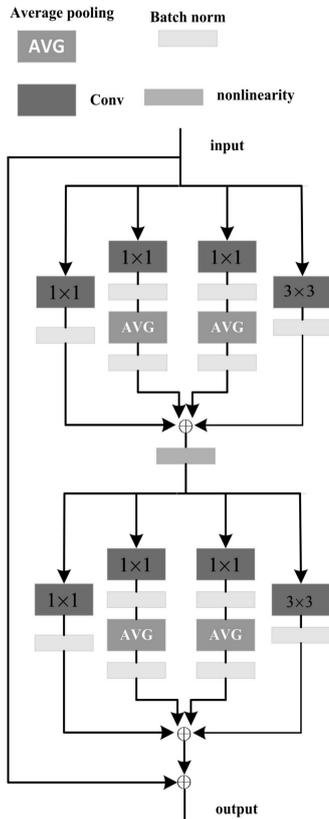


图 4 DBB-Res18 瓶颈块

1.2 残差空间注意力

人脸表情所包含的鼻子、嘴巴、眼睛、眉毛等部分的变化会产生不同类别的表情,而区分表情之间的细微差异,突出上述人脸关键部位的特征尤为重要。受文献 [14] 的启发,该文提出了残差空间注意力 (Residual Spatial Attention, RSA),如图 5 所示。

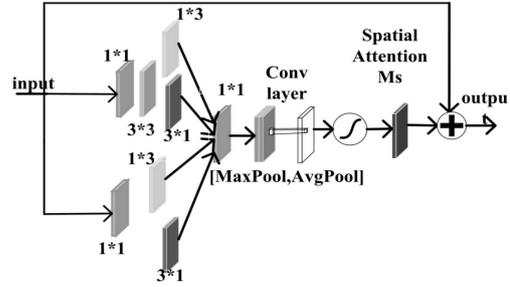


图 5 残差空间注意力

RSA 由两条多尺度卷积序列,空间注意力^[15]、恒等映射支路组成。以 DBB-Res18 的高层卷积 res4_1 产生的人脸图像为例,图像输入后将分为三条支路 F_A 、 F_B 、 F_C 。 F_B 和 F_C 支路使用 1×1 的卷积对输入的人脸图像进行降维; F_B 、 F_C 支路在降维后对输入图像进行并行非对称 1×3 和 3×1 卷积操作,获取不同尺度的特征人脸;在融合 F_B 和 F_C 支路的不同尺度特征后,将其送入 1×1 卷积升维获得特征 F 。将 F 引入最大池化层和平均池化层,生成两个特征描述符 $F_{avg}^S \in \mathbb{R}^{1 \times H \times W}$ 和 $F_{max}^S \in \mathbb{R}^{1 \times H \times W}$ 。将两者拼接送入 7×7 的卷积,经 sigmoid 函数生成空间特征 $M_s(F) \in \mathbb{R}^{H \times W}$,具体计算如式(2)所示:

$$M_s(F) = \delta(f^{7 \times 7} [F_{avg}^S; F_{max}^S]) \quad (2)$$

最后将输出的空间特征 $M_s(F)$ 与输入的特征映射相 F_A 叠加得到输出特征。从上述描述可知,输入的图像经 RSA 作用后,可以使 DBB-Res18 更容易获得人脸关键区域的特征。

1.3 多分支精简双线性池化结构

双线性池化^[16] (Bilinear Pooling, BP)通过式(3)生成全局图像描述符:

$$B(X) = \sum_{s \in S} x_s x_s^T \quad (3)$$

其中, X 为局部描述符,表示如式(4):

$$X = (x_1, \dots, x_{|S|}, x_s \in \mathbb{R}^C) \quad (4)$$

S 为空间位置集合。由式(3)可知, $B(X)$ 为一个 $c \times c$ 的矩阵。在图像分类中通常使用逻辑回归去实现使用双线性描述符的图像分类任务,而逻辑回归被看作线性核机。假设 X 、 Y 为两组局部图像描述符,那么线性核机将对其做下述比较:

$$\langle \langle B(X), B(Y) \rangle \rangle = \langle \sum_{s \in S} x_s x_s^T, \sum_{u \in U} y_u y_u^T \rangle = \sum_{s \in S} \sum_{u \in U} \langle x_s x_s^T, y_u y_u^T \rangle =$$

$$\sum_{s \in S} \sum_{u \in U} \langle x_s, y_u \rangle^2 \quad (5)$$

从式(5)可知,双线性描述符的比较算子为二阶多项式核。设 $k(x, y)$ 表示二阶多项式核,若存在低维映射函数 $\varphi(x) \in R^d$, 其中 $d \ll c^2$, 满足 $\langle \varphi(x), \varphi(y) \rangle \approx k(x, y)$, 则可以用式(6)近似式(5)的内积:

$$\begin{aligned} \langle \langle B(X), B(Y) \rangle \rangle &= \sum_{s \in S} \sum_{u \in U} \langle x_s, y_u \rangle^2 = \\ &= \sum_{s \in S} \sum_{u \in U} \langle \varphi(x_s), \varphi(y_u) \rangle^2 = \\ &= \langle C(X), C(Y) \rangle \end{aligned} \quad (6)$$

其中:

$$C(X) = \sum_{s \in S} \varphi(x_s) \quad (7)$$

为精简双线性池化的特征。

以输入的人脸图像特征 F1 为例,经精简双线性池化后,生成式(7)所示的精简的人脸全局特征描述符。

如图 1 所示,该文选取 DBB-Res18 高层卷积构成了一个多分支支路,在经 RSA 后得到了人脸特征 F1、F2、F3、F4。双线性池化在“类内差异小,类间差异大”的细粒度图像分类任务中取得不错的成绩,而人脸图像中不同类别的表情之间差异小,相似度高。但双线性池化参数数量太大,不利于后续分析。为此,该文将多分支路的所携带的人脸特征送入精简双线性池化中构成多分支精简双线性池化网络结构。该结构对不同尺度人脸图像使用精简双线性池化,增强了不同尺度的表情图像的内聚性,使 MCBP 网络获取更丰富的人脸关键部位的特征。经实验表明,该多分支精简双线性池化网络结构明显改善了基础网络捕捉人脸细微部分特征能力,更好地辨别了表情之间的相似性、促进了面

部表情识别精度的提升。

2 实验结果和分析

2.1 实验环境与实验参数

实验操作系统为 Ubuntu16.04.4, GPU 为 NVIDIA GeForce GTX 1080Ti * 2, 实验环境 python 3.5, 深度学习框架 PyTorch1.0.0, 显存 22G。

使用随机裁剪、中心裁剪、批归一化处理等方法对 RAF-DB、CK+数据集的测试集和训练集表情图像进行了预处理。同时使用随机梯度下降算法对网络进行优化,此时网络的初始学习率为 0.01, 权重衰减为 1×10^{-4} , 动量为 0.9, 总共迭代 240 次。由于 CK+和 RAF-DB 数据集样本量的差异,故 CK+数据集上批量处理的大小为 128, 每 40 次学习率衰减 10 倍。在 RAF-DB 数据集上批量处理大小为 32, 每 40 次学习率衰减 0.1。

2.2 数据集

该文选用的数据集为 CK+、RAF-DB。CK+ 总共拥有 981 张 44x44 像素的人脸图片, 包含 7 种基本的人脸表情类别。RAF-DB 是在日常环境下组成的数据集, 该数据集由 7 种基本的表情类别和 11 种复合情感表情类别组成。该文使用的 RAF-DB 数据集是经 Face++ 定位后的人脸图片。单标签总共 15 339 张图片, 其中 12 271 张用作训练样本, 3 068 张用作测试样本。上述数据集样本分布如表 1 所示, 包含 Neutral (中性)、Happy (快乐)、Surprise (惊讶)、Sad (悲伤)、Fear (害怕)、Angry (生气)、Disgust (厌恶)、Contempt (鄙视) 等表情类别。

表 1 数据集样本分布

数据集	Neutral	Happy	Surprise	Sad	Fear	Angry	Disgust	Contempt	总样本
CK+	-	207	249	84	75	135	177	54	981
RAF-DB	3 204	5 957	1 619	2 460	355	867	877	-	15 339

2.3 多分支结构性选择分析

DBB-Res18 在分类时仅使用最后一层卷积进行分类, 这显然不足以对整个输入图像的各部分语义信息进行表征, 同时最后一层卷积不可避免地会丢失骨干网络中间部分的卷积层所含有可以区别人脸表情

类间差异的特征信息。为了明确 DBB-Res18 的不同卷积层关注人脸表情图像的位置, 该文采用 Grad-CAM^[17] 方法将 DBB-Res18 的部分卷积层进行可视化, 如图 6 所示。

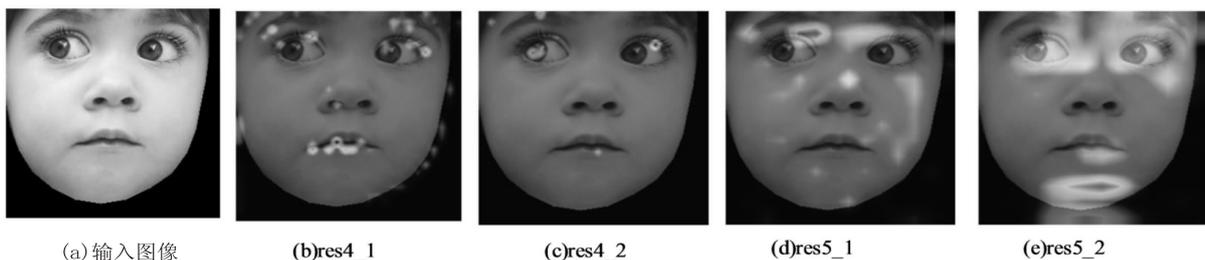


图 6 DBB-Res18 部分卷积层可视化

从图6可知, DBB-Res18 的高层卷积对人脸的眼睛、嘴巴、眉毛等部分反应较为强烈。而这些关键区域特征变化正是辨别表情之间细微区别、解决不同类别表情之间差异的关键因素。为此, 该文选取 DBB-Res18 的几组高层卷积组合构成多分支结构并在 CK+、RAF-DB 数据集上开展实验, 多分支结构组合方式如

表2 多分支结构组合方式

res4_1	res4_2	res5_1	res5_2	RSA	CBP	Accuracy in CK+/%	Accuracy in RAF-DB/%
√	×	√	√	√	√	97.43	81.29
×	√	√	√	√	√	97.95	81.94
×	×	√	√	√	√	94.35	78.87
√	√	√	√	√	√	98.46	82.99

2.4 MCBP 网络与现有方法的对比结果与分析

为了体现 MCBP 网络的优势和竞争力, 该文将 MCBP 网络与近年现有的人脸表情识别研究工作在 CK+、RAF-DB 数据集上进行识别率的对比实验, 实验结果如表3、表4所示。

表3 在 CK+数据集上与现有方法识别率的比较

方法	年份	Accuracy/%	标准方差/%
Lan ^[18]	2020	94.46	0.279
Shao ^[19]	2020	95.29	0.438
Gan ^[20]	2022	95.71	0.471
Lan ^[21]	2022	98.27	0.351
MCBP	2022	98.46	0.249

从表3实验结果可知, 所提的 MCBP 网络在 CK+数据集上, 相比于 Lan^[18]、Shao^[19]、Gan^[20]等人提出的方法, 分别提升了4个百分点、3.17个百分点、2.75个百分点。由于 CK+数据集样本数量不多, 为了更进一步展示 MCBP 网络在 CK+的稳定性, 将 MCBP 与现有方法分别在 CK+上开展了5次实验, 然后再求取标准方差。表3的结果充分地展示了文中方法在 CK+数据集上有效性、竞争性。

表4 在 RAF-DB数据集上与现有方法识别率的比较

方法	年份	Accuracy/%
DLP-CNN ^[22]	2019	74.20
MA ^[23]	2020	75.59
Gan ^[20]	2022	76.59
DeepExp3D ^[24]	2020	82.06
MCBP	2022	82.99

从表4实验结果可知, MCBP 网络在 RAF-DB 数据集上相较于 DLP-CNN^[22]、MA^[23]、DeepExp3D^[24]等方法分别提升了8.79百分点、7.4百分点、0.93百分点, 相较于 Gan^[21]等人提出方法准确率提升了6.4百分点。上述结果可见, 所提的方法在 RAF-DB 相较于现有的方法有更高的识别率, 具有一定竞争性和优势。

表2所示, 其中“Accuracy”为准确率, 表示正确分类的人脸表情样本与总的人脸表情样本之比。表2结果表明 res4_1、res4_2、res5_1、res5_2 多分支结构组合在人脸表情数据集上展示了优越的性能, 故选其为文中的多分支结构。

人脸表情类别之间细微差异的存在导致了表情之间的相似性, 而这种相似性导致骨干网络对单个类别样本识别精度不高, 进而影响网络对整个数据集的识别率。为了探究 MCBP 网络对上述问题的作用, 分别展示了 ResNet-18 和 MCBP 在 CK+、RAF-DB 数据集的混淆矩阵, 如图7和图8所示。

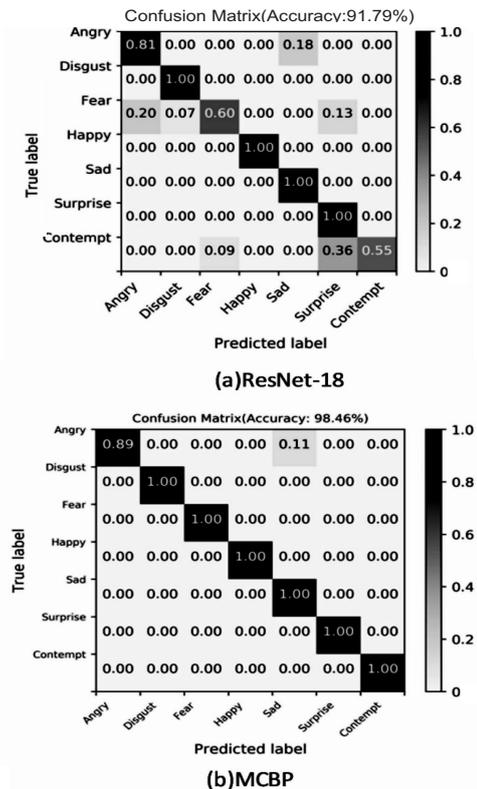


图7 CK+数据集上的混淆矩阵

从图7(a)可知, ResNet-18 对 CK+数据集中的生气 (Angry) 以及鄙视 (Contempt) 两类表情识别率不太理想。其中有18%生气表情的数据样本被误识别为悲伤表情, 36%鄙视类表情的数据样本被错误分类为惊讶 (Surprise) 表情。这极大地反应了残差网络 ResNet-18 对人脸表情细微差别判断的难度。从图7

(b)可以看出,提出的 MCBP 网络有效地减少了这种类间细微差异对表情识别工作的影响,提升了人脸表情识别率。

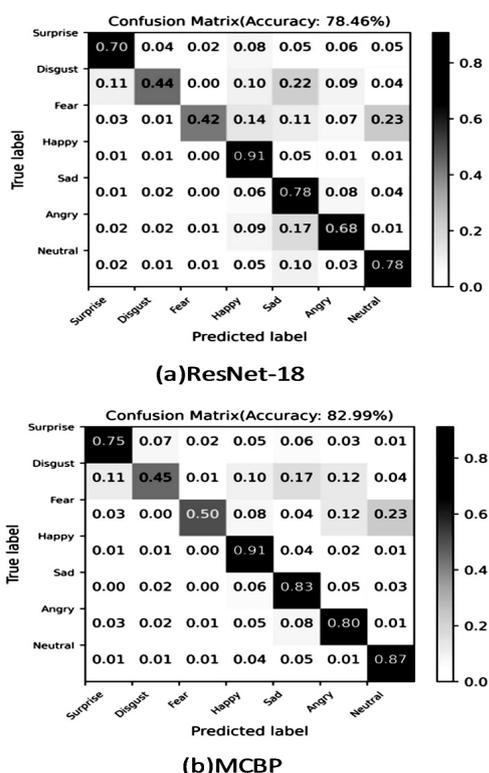


图 8 RAF-DB 数据集上的混淆矩阵

从图 8(a)可知,在 RAF-DB 数据集上 ResNet-18 对害怕(Fear)、厌恶(Disgust)、惊讶(Surprise)、生气(Angry)等数据样本分类精度不高。其中 22% 的厌恶表情类别被误分类为伤心(Sad),8% 的惊讶表情被误分类为厌恶和快乐(Happy)、23% 的害怕表情被错误分类为中性(Neutral)、17% 的生气表情被错误分类为伤心。该数据是由日常生活环境下构成人脸表情组合,更能反映现实生活中不相同的人脸表情类别之间的细微差异。由图 8(b)可知,MCBP 网络有效地缓减了人脸表情类别之间的细微差异对识别率的不利影响。

2.5 不同姿态的人脸表情的鲁棒性

在自然环境下,不同视角获取表情图像有不同的姿态,而这也往往会给人脸表情识别效果带来影响。为此,该文从 RAF-DB 数据集选取了侧脸、低头、仰脸等人脸图像,构成带有侧脸、低头、仰脸的 RAF-DB 子集以测试 MCBP 网络对不同姿态人脸表情的鲁棒性。从表 5 实验结果可知,MCBP 网络能够有效解决不同姿态对人脸表情带来的影响,有较好的鲁棒性。

表 5 不同姿态下 MCBP 网络的识别率 %

方法	侧脸	低头	仰脸
MCBP	71.43	74.67	92.95

2.6 消融实验

MCBP 网络包括 DBB-Res18、残差空间注意力、多分支精简双线性结构。为了检验上述各模块对人脸表情识别率的影响,分别在 CK+、RAF-DB 数据集上开展消融实验,实验的具体详情如表 6 所示。

表 6 在 CK+、RAF-DB 消融实验对比结果

Method	RSA	CBP	Accuracy in	Accuracy in
			CK+/%	RAF-DB/%
ResNet-18	×	×	91.79	78.46
DBB-Res18	×	×	92.82	79.14
DBB-Res18	✓	×	96.41	79.72
DBB-Res18	×	✓	94.87	81.48
res4_1, res4_2, res5_1, res5_2	✓	×	97.44	79.49
res4_1, res4_2, res5_1, res5_2	×	✓	97.95	81.1
res4_1, res4_2, res5_1, res5_2	✓	✓	98.46	82.99

从表 6 可知,DBB-Res18 相较于残差网络 ResNet-18 在 CK+、RAF-DB 识别率分别提高了 1.03 百分点、0.68 百分点。这表明利用该文设计的 DBB 结构替代 ResNet-18 的单个卷积层,能改善 ResNet-18 的特征提取能力,使其获取更丰富的人脸表情特征。在 DBB-Res18 的基础上,添加残差空间注意力(RSA),实验结果表明相较于 DBB-Res18 在 CK+、RAF-DB 识别率分别提高了 3.59 百分点,0.58 百分点。这反映了残差空间注意力能很好地抑制非关键的人脸特征信息,突出人脸关键区域位置的特征。在 DBB-Res18 的基础上,实验添加了精简双线性池化(CBP),实验结果表明相较于 DBB-Res18 在 CK+、RAF-DB 识别率分别提高了 2.05 百分点、2.34 百分点。侧面反映了 CBP 能够使网络提取更加全面的人脸表情特征。最后,选取 DDB-Res18 的高层卷积构成多分支结构,分别验证融入残差空间注意力,精简双线性池化网络,以及同时融入二者构成所提的多分支精简双线性池化网络。表 6 的结果表明,MCBP 在 CK+、RAF-DB 数据集的识别率达到了 98.46%、82.99%,相较于 ResNet-18 在 CK+、RAF-DB 数据集上提高了 6.67 百分点、4.53 百分点,表明在残差网络 ResNet-18 的改进的有效性,也验证了 MCBP 网络有利于获取更全面的人脸表征,有效缓减了人脸表情细微的类间差异带来的问题。

3 结束语

为了解决人脸不同类别的表情细微差异和特征提

取不充分等问题,提出了多分支精简双线性池化网络用于人脸表情识别研究工作,并在公开的数据集 RAF-DB、CK+上验证了模型的有效性。以提升残差网络 ResNet-18 对人脸表情特征的提取能力角度出发,分别利用所设计新的 DBB 结构组成新的瓶颈块替换残差网络原有的瓶颈块,选取改进后的 DBB-Res18 搭建多分支网络以及将残差空间注意力和精简双线性池化融入到多分支的每条支路中。这样的层次化和结构化的设计,有利于获取更加丰富的人脸表情特征,提升了网络判别具有细微差异的不同类别的人脸表情图像能力,明显提升了人脸表情识别的准确率,具有一定的应用价值。

下一步将深入探讨 MCBP 对不同光照、遮挡以及数据集样本之间不均匀人脸表情数据样本的作用,并验证网络的有效性、泛化能力和鲁棒性。

参考文献:

- [1] 甘路涛. 基于面部表情的驾驶员状态分析方法研究[D]. 成都:电子科技大学,2019.
- [2] SHAKYA S, SHARMA S, BASNET A. Human behavior prediction using facial expression analysis[C]//2016 international conference on computing communication and automation (ICCCA). Greater Noida; IEEE, 2016; 399-404.
- [3] XIANG J A Z G. Joint face detection and facial expression recognition with MTCNN[C]//2017 4th international conference on information science and control engineering (ICISCE). Changsha; IEEE, 2017; 424-427.
- [4] LAKSHMI D, PONNUSAMY R. Facial emotion recognition using modified HOG and LBP features with deep stacked autoencoders[J]. Microprocessors and Microsystems, 2021, 82: 103834.
- [5] ALPHONSE A S, DHARMA D. Enhanced Gabor (E-Gabor), hypersphere-based normalization and Pearson general kernel-based discriminant analysis for dimension reduction and classification of facial emotions[J]. Expert Systems with Applications, 2017, 90: 127-145.
- [6] KOTSIA I, PITAS I. Facial expression recognition in image sequences using geometric deformation features and support vector machines[J]. IEEE Transactions on Image Processing, 2007, 16(1): 172-187.
- [7] KONTSCIEDER P. Deep neural decision forests[C]//2015 IEEE international conference on computer vision (ICCV). Santiago; IEEE, 2015; 1467-1475.
- [8] KUO C, LAI S, SARKIS M. A compact deep learning model for robust facial expression recognition[C]//2018 IEEE/CVF conference on computer vision and pattern recognition workshops (CVPRW). Salt Lake City; IEEE, 2018; 2202-22028.
- [9] HE K, ZHANG X, REN S, et al. Deep residual learning for image recognition[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. Las Vegas; IEEE, 2016; 770-778.
- [10] 石敏. 基于深度残差网络与人脸关键点的表情识别[J]. 软件导刊, 2022, 21(4): 226-232.
- [11] 高涛, 邵倩, 张亚南, 等. 基于深度残差网络的人脸表情识别研究[J]. 电子设计工程, 2020, 28(23): 101-104.
- [12] 付小龙, 李志丹, 程吉祥, 等. 残差网络和损失函数集成的人脸表情识别[J]. 控制工程, 2022, 29(3): 522-529.
- [13] DING X, ZHANG X, HAN J, et al. Diverse branch block: building a convolution as an inception-like unit[C]//2021 IEEE/CVF conference on computer vision and pattern recognition (CVPR). Nashville; IEEE, 2021; 10881-10890.
- [14] 高涛, 杨朝晨, 陈婷, 等. 深度多尺度融合注意力残差人脸表情识别网络[J]. 智能系统学报, 2022, 17(2): 393-401.
- [15] WOO S, PARK J, LEE J Y, et al. CBAM: convolutional block attention module[C]//Computer vision - ECCV 2018. [s. l.]; Springer, 2018.
- [16] LIN T, ROYCHOWDHURY A, MAJI S. Bilinear CNN models for fine-grained visual recognition[C]//2015 IEEE international conference on computer vision (ICCV). Santiago; IEEE, 2015; 1449-1457.
- [17] SELVARAJU R R, COGSWELL M, DAS A, et al. Grad-CAM: visual explanations from deep networks via gradient-based localization[J]. International Journal of Computer Vision, 2020, 128(2): 336-359.
- [18] 兰凌强, 李欣, 刘洪缘, 等. 基于联合正则化策略的人脸表情识别方法[J]. 北京航空航天大学学报, 2020, 46(9): 1797-1806.
- [19] SHAO J, QIAN Y. Three convolutional neural network models for facial expression recognition in the wild[J]. Neurocomputing, 2019, 355: 82-92.
- [20] GAN C, XIAO J, WANG Z, et al. Facial expression recognition using densely connected convolutional neural network and hierarchical spatial attention[J]. Image and Vision Computing, 2022, 117: 104342.
- [21] 蓝峥杰, 王烈, 聂雄. 一种基于词频-逆文档频率和混合损失的表情识别算法[J/OL]. 计算机工程, 2022; 1-10.
- [22] LI S, DENG W. Reliable crowdsourcing and deep locality-preserving learning for unconstrained facial expression recognition[J]. IEEE Trans. Image Process, 2019, 28(1): 356-370.
- [23] GAN Y, CHEN J, YANG Z, et al. Multiple attention network for facial expression recognition[J]. IEEE Access, 2020, 8: 7383-7393.
- [24] KOUJAN M R, ALHARBAWEE L, GIANNAKAKIS G, et al. Real-time facial expression recognition "in the wild" by disentangling 3D expression from identity[C]//2020 14th IEEE international conference on automatic face and gesture recognition (FG 2020). Buenos Aires; IEEE, 2020.