DOI:10.19651/j.cnki.emt.2211334

基于 MobileNetV3 多尺度特征融合的人脸表情识别*

薛志超 伊力哈木·亚尔买买提 闫天星
(新疆大学电气工程学院 乌鲁木齐 830017)

摘 要:针对人脸表情识别中普通卷积神经网络特征提取能力不足且识别效率低下的情况,本文提出了一种基于 MobileNetV3 多尺度特征融合的人脸表情识别。首先利用 MobileNetV3 进行特征提取以获得高层次情感信息;其次在骨干网络中借鉴 DenseNet 结构,增强特征复用并提升网络重要面部特征表达能力;然后利用特征金字塔模块充分获取人脸图像的深层和浅层多尺度融合特征,从而提高了 MobileNetV3 的特征提取能力和实时性;最后利用全连接层构建分类器对表情进行分类,从而完成了人脸表情识别。通过实验验证,结果表明,在 CK+和 FERPlus 数据集上识别准确率可以达到 88.3%和 98.8%,与现有方法相比分别提高了 2.3%和 1.5%,表明了所提方法识别效果好,泛化能力强。

关键词:人脸表情识别;情感分析;MobileNetV3 模型;特征金字塔;DenseNet 结构
中图分类号:TP391 **文献标识码**:A **国家标准学科分类代码**:520.60

Facial expression recognition based on MobileNetV3 multi-scale feature fusion

Xue Zhichao Yilihamu Ya'ermaimaiti Yan Tianxing
(School of Electrical Engineering, Xinjiang University, Urumqi 830017, China)

Abstract: In view of the lack of feature extraction ability and low recognition efficiency of common convolutional neural network in facial expression recognition, this paper proposes a facial expression recognition based on multi-scale feature fusion of MobileNetV3. Firstly, MobileNetV3 was used for feature extraction to obtain high-level emotion information. Secondly, the DenseNet structure is used in the backbone network to enhance feature reuse and improve the expression ability of important facial features. Then the feature pyramid module is used to fully obtain the deep and shallow multi-scale fusion features of face images, so as to improve the feature extraction ability and real-time performance of MobileNetV3. Finally, the full connection layer is used to construct a classifier to classify the facial expression, so as to complete the facial expression recognition. Through experimental verification, the results show that the recognition accuracy on CK+ and FERPlus datasets can reach 88.3% and 98.8%, which are improved by 2.3% and 1.5% respectively compared with the existing methods, indicating that the proposed method has good recognition effect and strong generalization.

Keywords: facial expression recognition; sentiment analysis; MobileNetV3 model; feature pyramid networks; DenseNet structure

0 引 言

人脸表情是人们沟通交流中主要的一种信息传达手段。日常生活中,面部表情虽然作为一种非语言的交流方式,但是其相较语言和动作却能够更好地表达人们的情绪状态,反映人们的情感。近些年来,由于计算机技术的快速发展和人机交互的渐渐兴起,人脸表情识别系统正在被医疗辅助、交互游戏和智能交通等领域广泛应用。

人脸表情识别主要包括以下 4 大组成部分:图像采集、图像预处理、特征提取和表情识别分析。特征提取就是将点阵转化成更高级别图像表述如形状、运动、颜色、纹理、空间结构等,在尽可能保证稳定性和识别率的前提下,对庞大的图像数据进行降维处理。特征提取作为其中最为关键的一步,对表情识别效果具有决定性影响。传统的人脸表

收稿日期:2022-09-08
* 基金项目:国家自然科学基金(61866037,61462082)项目资助

情识别算法都是先通过人工设计特征提取器进行特征提取,然后再利用分类器实现表情识别,如主成分分析法[1](principal component analysis,PCA),局部二值模式[2](local binary pattern,LBP)和 Gabor 小波变换[3]等。Zhu 等[4]通过等效主成分分析进行表情特征提取,以线性回归作分类器的方法,大幅度提高了表情特征提取的鲁棒性。胡敏等[5]提出一种基于小波的多尺度中心化二值模式(multi-scale centralized binary pattern,MCBP)的人脸表情识别技术,对经小波分解后的两幅低频图像的特征区域进行中心化二值模式(centralized binary pattern,CBP)变换,获取到多个局部 CBP 直方图序列特征,提高了 MCBP 的表情辨别力。Han 等[6]提出一种由 LBP 和有监督局部保持投影(supervised locality preserving projection,SLPP)两者相结合的模型,该模型使用 LBP 获取图像直方图,并通过 SLPP 减少维度,来提高模型识别能力。吴昊等[7]提出一种利用典型相关分析法(canonical correlation analysis,CCA)融合双编码局部二值模式(double coding local binary pattern,DCLBP)算子和绝对梯度直方图(histogram of absolute gradients,HOAG)算子分别提取到的人脸图像局部纹理和形状特征,充分发挥单个特征的有效性,提升了人脸表情识别效果。虽然传统识别方法效果很好,但是由于人工特征提取受不同姿态、遮挡等因素的影响较为严重,导致识别性能有限,难以有效提升分类效果。

目前,深度学习凭借其自主特征学习能力逐渐应用于人脸表情识别领域中,并且取得了良好效果。Chen 等[8]为了有效提高突出表情变化区域的关注能力,将主网络与辅助网络共享结构参数。申毫等[9]提出一种改进的残差倒置网络,并通过筛选浅层特征与深层特征相融合的方式,实现轻量化的多特征融合模型。王建霞等[10]引入 Inception 结构增强网络特征提取能力,融合了高维和低维的特征,获得了相当好的表情识别效果。Li 等[11]给出一种基于深度残差网络 ResNet-50 的特征提取方案,其依赖人脸关键点的检测,当脸部区域被遮挡过多时,易出现预测错误的情况,因此多尺度特征融合方式显得尤为重要。

上述研究均使用完整特征图作为输入,然而实际分类任务中,特征的作用程度不同,为了突出有效特征信息,部分研究中引入了注意力机制。Hu 等[12]设计了挤压和激励注意力机制网络(squeeze excitation network, SENet),它通过自动收集各通道的特征权重信息,以提升有用特征抑制无效特征。Li 等[13]提出一个基于注意力机制的人脸表情识别网络,该网络将 LBP 特征与注意力机制相结合,增强了模型特征提取性能。为了有效的获取人脸表情上的有效特征,提高识别精度的同时减少网络模型参数,本文提出一种新的基于 MobileNetV3 多尺度特征融合的人脸表情识别网络,工作概括如下:

1)本文借鉴并改进了特征金字塔模块(feature pyramid networks,FPN)模块使其应用于人脸表情识别。多尺度特征融合模块获取到人脸图像的多尺度特征,通过特征在不同尺度上进行卷积再聚合的方法,既考虑了图像的高层次特征,又考虑了图像的低层次特征,使得浅层特征具备深层特征的语义能力从而达到加强网络特征提取能力的目的。

2)本文在既没有加深网络层数又没有加宽网络结构的同时将具有特征复用功能的 DenseNet 结构替换到 MobileNetV3 网络中。从而获取更丰富的图像特征,以加强特征的有效性,在减轻梯度消失现象的同时,又减少计算量,从而有效提升表情识别能力。

3)提出基于 MobileNetv3 多尺度特征融合的人脸表情识别模型在公共人脸表情数据集 CK+、FERPlus 进行实验,实验结果验证了该方案的有效性、可行性和泛化性。

1 方法设计

1.1 网络结构

本文模型主要由 FPN 模块、DenseNet 结构、分类器 3 部分组成,如图 1 所示。首先,将尺寸为 224×224×3 的人脸表情图像作为模型输入,通过 MobileNetV3 骨干网络进行特征提取从而获得高层次的情感特征 $C1$;其次使用步长为 2 的卷积操作来将图像尺寸缩小一倍,再将其输入到 DenseNet 结构中。DenseNet 将网络中的所有层两两连接起来,并让网络中每一层都接受其前面所有层的特征信息来作为输入,然后再把每层的特征信息和作为输出,以增强对表情特征的复用能力;然后重复此操作,将得到的特征图分别命名为 $C1$、$C2$、$C3$;其中特征层 $C1$ 为包含丰富细微信息的高层特征;特征层 $C2$ 的感受野为中等;特征层 $C3$ 感受野较大,但是因多次下采样操作使得其只有较少的特征语义信息。此时网络已经构成了一个 3 层的特征金字塔结构。首先对 $C1$ 进行下采样与 $C2$ 进行拼接获得 $P1$,然后再对 $P1$ 进行下采样与 $C3$ 进行拼接并卷积得到低语义、高分辨率信息与高语义、低分辨率信息相结合的特征 $P2$,尺寸为 7×7×160,此时完成了对图像的特征提取。最后,再将特征 $P2$ 输入到全连接层构建的分类器中实现情感预测。

1.2 MobileNetV3 网络

Mobilenet 系列网络[14]核心思想是深度可分离卷积操作。使用深度可分离卷积相比于传统卷积计算可以显著的减少模型运算参数量。假设卷积核的尺寸为 $D_K \times D_K$,输入特征图的尺度为 $D_F \times D_F$,M 为输入通道数,N 为输出通道数。那么标准卷积所需要的参数量为:

$$P_1 = D_K^2 \times M \times N \tag{1}$$

深度可分离卷积若也进行一样的步骤,则需要的参数量为:

$$P_2 = D_K^2 \times 1 \times M + 1 \times 1 \times M \times N \tag{2}$$

深度可分离卷积与标准卷积运算的参数量比值为:

$$\frac{P_2}{P_1} = \frac{D_K^2 + N}{D_K^2 \times N} = \frac{1}{N} + \frac{1}{D_K^2} \tag{3}$$

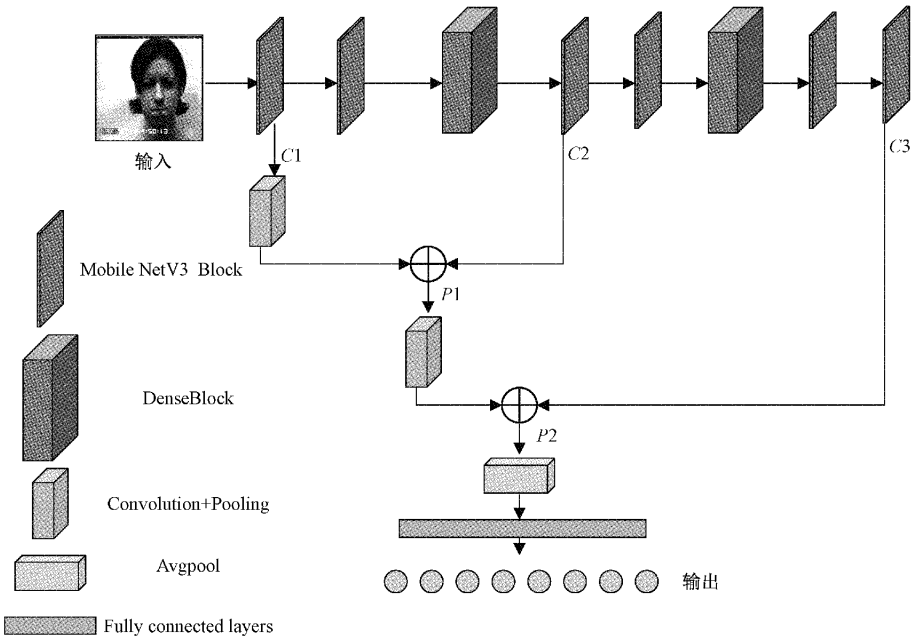

图 1　网络结构

根据上式得到，如果卷积核的尺寸为 3×3，与标准卷积相比，深度可分离卷积的参数量低了 8～9 倍。所以，使用深度可分离卷积可以在显著降低模型参数量的同时增加检测效率。

在 MobileNetV3 网络中引入注意力机制则是其另一改善之处，该机制由两部分构成：挤压模块和激励模块。挤压模块的作用是将输入的多通道矩阵压缩成一维向量，再将获得的一维向量与可训练的权重进行乘法运算来实现激励，权重值则可以通过误差反向传播来调整，从而筛选出所需要的特征，其中挤压模块的计算公式为：

$$z_c = F_{sq}(u_c) = \frac{1}{HW}\sum_{i=1}^{H}\sum_{j=1}^{W} u_c(i,j) \tag{4}$$

式中：H，W 为空间维度；z_c 为网络输出的特征向量；u_c 则为输入矩阵的第 c 个通道特征；F_{sq} 是挤压模块的挤压函数；i 为二维特征图的行坐标；j 为二维特征图的列坐标。

激励模块是由两个全连接层组成。当输入层和输出层的神经元个数相等时，采用 ReLU 激活函数，而激励运算则采用 Sigmoid 函数来进行计算，其表达式为：

$$s = F_{ex}(z,W) = \sigma(W_2\delta(W_1 z)) \tag{5}$$

式中：σ 为 Sigmoid 激活函数；δ 为 ReLU 函数；F_{ex} 为激励函数；g 是全连接层；z 为网络输出特征值；W_1、W_2 为全连接层权重值。

最后，将输出向量和原始特征图的矩阵通过 Scale 运算相乘，获得施加权重的特征图，从而增加有效特以及抑制无效特征，使获得的特征更具有表达性，其表达式为：

$$\tilde{x}_c = F_{scale}(u_c, s_c) = s_c \times u_c \tag{6}$$

式中：$\tilde{x}_c$ 是输出特征图的第 c 个通道特征；F_{scale} 则为注意力权值。

MobileNetV3 为了在确保准确率的基础上再次减少计算成本，所以使用由 Swish 函数改善而成的 H-Swish 来代替部分 ReLU6 函数。其具体变化过程如下：

$$\text{swish}x = x \cdot \sigma(x) \tag{7}$$

$$h-\text{swish}[x] = x\,\frac{\text{ReLU6}(x+3)}{6} \tag{8}$$

MobileNetV3 网络可以进行移动分类、检测以及分割，该网络能够在确保模型轻量化的基础上，进一步提升模型精确度和运行效率，所以本文采用 MobileNetV3 网络作为人脸表情分类模型的基础网络，以获得大量人脸表情特征。

1.3　多尺度特征融合模块

在人脸表情识别中，存在着各种大小不一、背景复杂的人脸表情图像，因此需要识别网络具有较高的提取能力来充分获取多尺度的人脸表情从而提高识别准确率。本文多尺度特征融合模块构建原理如图 2 所示。

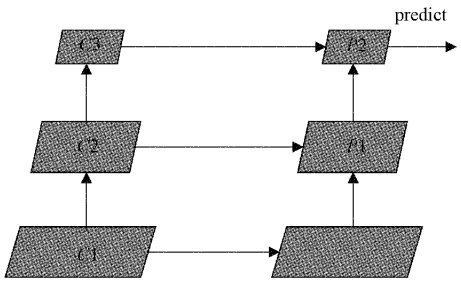

图 2　多尺度特征融合模块

本文选用 MobileNetV3 网络作为 FPN[15] 的基本网络，在自下而上的网络中，选择有较强语义特征的{C1，C2，C3}3 层特征图。网络中生成的特征金字塔分辨率最

高的映射特征图 $C1$ 是自下向上的特征金字塔结构中所构成的最底层特征图;$C1$ 采用步长为 2 的卷积层代替了最大池化操作实现下采样之后,将 $C1$ 与 $C2$ 逐元素相加,进而实现深层特征与稍浅层特征相互融合得到 $P1$。再将此过程迭代,最后完成了特征金字塔层级间最高层特征与底层特征的融合得到 $P2$,在融合后,特征金字塔的特征层为 $\{P1,P2\}$。图 2 左侧为图像特征提取步骤,以 224×224 尺寸大小的图像作为输入,经过两个卷积核大小为 3×3、步长为 2 的卷积层实现下采样后就得到了 56×56 的 4 倍下采样特征图。继续下采样得到 7×7 的下采样特征图。图 2 右侧为图像多尺度特征融合过程,就是对每层的特征图都使用大小为 1×1,步长为 2 的卷积核来升维,然后再经过池化操作使其与上层的特征图保持相同的尺寸和维度,随后,将两张特征图逐元素添加,进行特征融合。在融合后的特征图上使用 3×3 卷积,消除融合的混叠效应从而获得低语义、高分辨率信息与高语义、低分辨率信息相结合的 $P2$。

1.4　DenseNet 网络结构

DenseNet 网络是由 Huang 等[16]提出的一种具有密集连接的卷积神经网络,致力于从特征重用角度提升网络性能,其核心思想是通过建立一个跨层连接,来连通网络中所有的前后层。为了最大限度地进行网络各层之间信息交换,DenseNet 把网络中的所有层都两两连接了起来,并让网络中每一层都接受其前面所有层的特征信息来作为输入,DenseNet 网络的密集连接机制如图 3 所示。

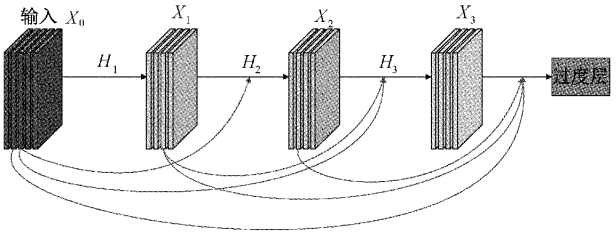

图 3　DenseNet 密集连接机制图

由图 3 可以看出,DenseNet 抛弃了传统神经网络在特征传递时先将特征叠加然后再传给某一层的思路,而是通过将前面所有层的特征输入加以拼接的方法,再将得到的输出特征图传输给后面网络中的每个层里。通过这种方法可以在保证前馈特性的同时,网络中每层都能传递出最大信息且网络训练速度也加快了。这种密集连接机制在减少网络学习冗余信息的同时也有效的提高了特征传播的复用能力。若一幅图像 X_0 在网络中传递,那么如式(9)所示为 DenseNet 网络在第 L 层的输出情况:

$$X_L = H_L([X_0,X_1,\cdots,X_{L-1}]) \tag{9}$$

式中:H_L 是一个非线性转换函数。第 L 层的输入是通过将之前每一层的特征图 $[X_0,X_1,\cdots,X_{L-1}]$ 进行合并获得。

本文设计的网络模型受到该网络结构的启发,在 MobileNetV3 网络基础上,通过结合人脸表情识别的特点及所使用的数据集,对网络模型进行修改以及对参数进行优化,使其能够更有效地实现特征的重用,从而使网络在提高网络性能的同时还能保证网络计算量和参数更少。

2　实　　验

2.1　实验准备

本文的实验操作系统是 Windows 10,实验环境是采用了 Pytorch 框架搭建,硬件平台 CPU 为 i7-11800H,GPU 为 6 G 的 NVIDA GeForce GTX 3060Ti。采用了 Adam 优化器,学习率设为 0.001,实验 batch size 设置为 32,训练 epoch 为 150。

2.2　数据集

实验中分别使用 CK+表情数据集和 FERPlus 表情数据集进行实验。

CK+数据集[17]为含有 123 个人共 593 个表情的视频序列样本。本实验只取了视频中人脸表情较为明显状态的 3 幅图像,图像裁剪到 48×48 pixels 尺寸。其共涉及 7 种表情,依次是愤怒(anger)、厌恶(disgust)、害怕(fear)、开心(happy)、悲伤(sad)、惊讶(surprise)及中性(neutral)表情。如图 4 所示是 CK+数据集中的 7 种表情样例图像。

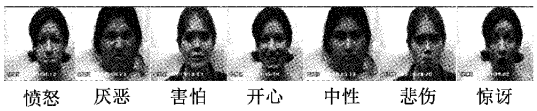

图 4　CK+表情库部分图像

FERPlus 数据集[18]为 2013 年 Kaggle 比赛中所提供 FER2013 的延伸,共由 31 313 幅图像所构成,从中挑出 24 941 张图像作为训练集,作为测试集与验证集的图像则各有 3 186 张,图像均为 48×48 像素的灰度图。重新标注的 FERPlus 数据集共被分成了 10 类:中性、开心、惊讶、悲伤、愤怒、厌恶、恐惧、蔑视、未知以及非人脸。其中未知类别和非人脸类别对人脸表情识别没有显著的影响,所以本实验选取 7 种常见的人脸表情类别作为数据集,但由于其包含了不同年龄、光照、遮挡、侧脸、姿势等,这使得在该数据集上进行人脸表情鉴别具有相当大的挑战性。如图 5 所示是 FERPlus 数据集中的 7 种表情样例图像。

图 5　FERPlus 表情库部分图像

2.3　实验结果分析

为了验证本文所改进 MobileNetV3 网络对人脸表情特征提取更具优势,选用了 JAIN[19]、Parallel CNN[20]、Em-AlexNet[21]、MIANet[22]和 SCAN[23]几种网络来作对比,在 CK+数据集上作测试,网络输入图像大小为 224×224。实验结果如表 1 所示。

表 1　CK＋数据集上不同 backbone 网络的准确率

算法	识别率/%
JAIN	93.20
Parallel CNN	94.03
Em-AlexNet	94.25
MIANet	95.76
SCAN	97.31
本文	98.80

由表 1 可以看出，本文所提出的模型在 CK＋数据集的准确率为 98.8%，与 JAIN、Parallel CNN、Em-AlexNet、MIANet、SCAN 这 5 种方法相比分别提升 5.8%，4.77%，4.55%，3.04%，1.49%。这表明本文所改进 MobileNetV3 网络模型人脸表情识别能力更强，可达到较高的识别精度。

本文模型与其他的经典模型 PLD[24]、ResNet＋VGG[25]、SHCNN[26]、LDR[27]、SCN[28] 在 FERPlus 数据集上进行实验对比。实验结果如表 2 所示。

表 2　与 FERPlus 数据集上一些最新方法进行比较

算法	识别率/%
PLD	85.10
ResNet＋VGG	87.40
SHCNN	86.25
LDR	87.60
SCN	88.01
本文	88.30

从表 2 对比结果中可以看出，本文所提出的模型具有更好的表现。在 FERPlus 上准确率为 88.30%，分别高于 PLD 的 85.10%和 ResNet＋VGG 的 87.40%，此外同样高于 SHCNN 的 86.25%、LDR 的 87.60%和 SCN 的 88.01%。可见，本文所提算法的先进性和泛化性比当前先进的方法更好，且对人脸表情的识别具有更高的准确率。

混淆矩阵能够详细描述每种表情识别精度和被误区分为其他表情的比例，其中对角线项表示每个表情的识别精度。如图 6 和 7 所示，分别为本文算法对两个数据集的混淆矩阵。

由图 6 和 7 可知，两个数据集中高兴表情的识别率最高，均达到 90%以上，主要原因为高兴表情样本基数大，表情幅度大，相较其他表情更易辨别。从 CK＋混淆矩阵可以看出在表达开心、中性和惊讶方面的表情识别率可达到 100%，伤心这类表情被错分为愤怒的数量较高，通过分析测试集图像可以发现，伤心和愤怒两类表情都伴随额头紧皱或者皱眉等特点。恐惧这类表情体现在嘴角张大和眼睛微微皱起，这使得预测结果易被错分为愤怒和悲伤两类表情，导致检测时识别率有所下降。相较 CK＋混淆矩阵，

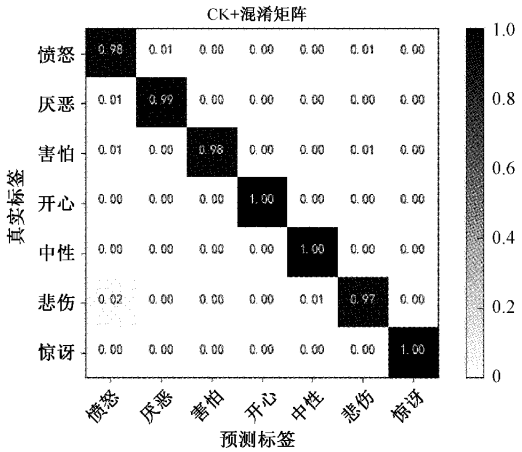

图 6　基于 CK＋数据集的人脸表情识别混淆矩阵

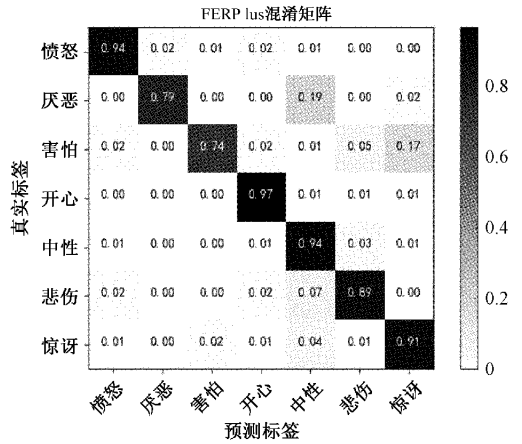

图 7　基于 FERPlus 数据集的人脸表情识别混淆矩阵

FERPlus 混淆矩阵中的表情识别率更低，这是由于 FERPlus 样本量大，且选择于互联网，包含错误样本较多，造成整体识别率下降。其中厌恶和恐惧识别率最低，厌恶易误识别为自然，恐惧识别错误率中最高为惊讶，这是由于这几类表情具有相似的外观特征。

2.4　消融实验

本文研究的主要改进重点在特征金字塔模块与 DenseNet 结构这两个部分，为了证明文中所提各模块的有效性与必要性，本文采用了消融实验来进行验证。采用了原始 MobileNetV3 网络与分别加入特征金字塔模块和 DenseNet 结构以及同时加入特征金字塔模块和 DenseNet 结构的改进模型进行了对比实验，实验结果如表 3 所示。

表 3　CK＋和 FERPlus 数据集消融实验结果

算法		识别率/%	
FPN	Densenet	CK＋	FERPlus
×	×	96.50	86.80
√	×	97.70	87.40
×	√	97.70	87.50
√	√	98.80	88.30

由表 3 可以看出,若只引入特征金字塔模块,识别准确率相较于 MobileNetV3 基础网络在 CK+和 FERPlus 数据集分别提升 1.2%和 0.6%。这表明 FPN 模块可以获得有效多尺度特征信息,使网络从中可以学习到不同人脸表情中的细微区别;若只引入 DenseNet 结构相较于基础网络准确率分别提升 1.2%和 0.7%,表明该结构确实提高了特征传播复用能力,从而提取到丰富的特征信息;FPN 模块与 DenseNet 结构的共同作用下相较于基础网络准确率分别提高 2.3%和 1.5%,且相比添加单一模块模型性能均有所提升,表明二者在共同作用下效果更优。

3 结 论

本文提出一种新的 MobileNetV3 网络融合 FPN 模块和 DenseNet 结构的人脸表情识别模型。其中,FPN 模块能够充分提取多尺度特征信息从而提高模型的特征提取能力,DenseNet 结构利用其密集连接机制以加强特征的传播复用,提升了网络识别效率。在已公开的人脸表情数据集 FER2013 以及 CK+上进行了对比实验来进行模型评估。实验结果显示,本文中所提方法在 CK+数据集上显著提升了表情识别的精度。但在 FER2013 数据集上识别率虽有提高,但提高幅度并不是很高,这是因为本文网络为轻量级网络,虽然相比于笨重模型,有更少的参数和计算量,对硬件更友好,但其抗干扰性能较差。因此在识别人脸表情时遭到光照、遮挡、侧脸、姿势等干扰会导致人脸表情识别率提升不明显。所以本文方法仅在训练样本处于干扰较少的情况下更具有优势,在未来的研究中应考虑如何将特征进行更深层次的融合,以及将本文模型应用到更复杂、真实的现实场景中,将表情识别从室内转向室外,使得理论研究能够与实际相结合。

参考文献

[1] 杨勇,蔡舒博. 一种基于两步降维和并行特征融合的表情识别方法[J]. 重庆邮电大学学报(自然科学版),2015,27(3):377-385.

[2] 周宇旋,吴秦,梁久祯,等. 判别性完全局部二值模式人脸表情识别[J]. 计算机工程与应用,2017,53(4):163-169,194.

[3] 徐峰,张军平. 人脸微表情识别综述[J]. 自动化学报,2017,43(3):333-348.

[4] ZHU Y, LI X, WU G. Face expression recognition based on equable principal component analysis and linear regression classification [C]. 2016 3rd International Conference on Systems and Informatics (ICSAI), 2016: 876-880.

[5] 胡敏,陈杏,王晓华,等. 基于小波 MCBP 和 WEF 的人脸表情识别[J]. 电子测量与仪器学报,2012,26(11):927-932.

[6] HAN D, MING Y. Facial expression recognition with LBP and SLPP combined method[C]. 2014 12th International Conference on Signal Processing(ICSP), 2014: 1418-1422.

[7] 吴昊,胡敏,高永,等. 融合 DCLBP 和 HOAG 特征的人脸表情识别方法[J]. 电子测量与仪器学报,2020,34(2):73-79.

[8] CHEN W, HU H. Joint prominent expression feature regions in auxiliary task learning network for facial expression recognition[J]. Electronics Letters, 2019, 55(1): 22-24.

[9] 申毫,孟庆浩,刘胤伯. 基于轻量卷积网络多层特征融合的人脸表情识别[J]. 激光与光电子学进展,2021,58(6):148-155.

[10] 王建霞,陈慧萍,李佳泽,等. 基于多特征融合卷积神经网络的人脸表情识别[J]. 河北科技大学学报,2019,40(6):540-547.

[11] LI Y, ZENG J, SHAN S, et al. Occlusion aware facial expression recognition using CNN with attention mechanism [J]. IEEE Transactions on Image Processing, 2018, 28(5): 2439-2450.

[12] HU J, SHEN L, SUN G. Squeeze and excitation networks[C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018: 7132-7141.

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

[14] HOWARD A, SANDLER M, CHU G, et al. Searching for MobileNetV3[C]. Proceedings of the IEEE/CVF International Conference on Computer Vision, 2019: 1314-1324.

[15] LIN T Y, DOLLAR P, GIRSHICK R, et al. Feature pyramid networks for object detection [C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017: 2117-2125.

[16] HUANG G, LIU Z, VAN DER MAATEN L, et al. Densely connected convolutional networks [C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017: 4700-4708.

[17] 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, 2010: 94-101.

[18] BALTRUSAITIS T, ZADEH A, LIM Y C, et al. Openface 2.0: Facial behavior analysis toolkit[C]. 2018 13th IEEE International Conference on Automatic Face & Gesture Recognition(FG 2018),

2018：59-66.

[19] JAIN D K，SHAMSOLMOALI P，SEHDEV P. Extended deep neural network for facial emotion recognition[J]. Pattern Recognition Letters，2019，120：69-74.

[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] 杨旭，尚振宏. 基于改进 AlexNet 的人脸表情识别[J]. 激光与光电子学进展，2020，57(14)：243-250.

[22] 罗思诗，李茂军，陈满. 多尺度融合注意力机制的人脸表情识别网络[J]. 计算机工程与应用，2022，80(2)：255-261.

[23] GERA D，BALASUBRAMANIAN S. Landmark guidance independent spatio-channel attention and complementary context information based facial expression recognition [J]. Pattern Recognition Letters，2021，145：58-66.

[24] BARSOUM E，ZHANG C，FERRER C C，et al. Training deep networks for facial expression recognition with crowd-sourced label distribution[C]. Proceedings of the 18th ACM International Conference on Multimodal Interaction，2016：279-283.

[25] HUANG C. Combining convolutional neural networks for emotion recognition [C]. 2017 IEEE MIT Undergraduate Research Technology Conference (URTC)，2017：1-4.

[26] MIAO S，XU H，HAN Z，et al. Recognizing facial expressions using a shallow convolutional neural network[J]. IEEE Access，2019，7：78000-78011.

[27] FAN X，DENG Z，WANG K，et al. Learning discriminative representation for facial expression recognition from uncertainties [C]. 2020 IEEE International Conference on Image Processing(ICIP)，2020：903-907.

[28] WANG K，PENG X，YANG J，et al. Suppressing uncertainties for large-scale facial expression recognition [C]. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition，2020：6897-6906.

作者简介

薛志超，硕士研究生，主要研究方向为图像处理、人脸识别等。

E-mail：1418612862@qq.com

伊力哈木·亚尔买买提(通信作者)，副教授，硕士生导师，主要研究方向为图像信息处理、人脸识别、模式识别等。

E-mail：65891080@qq.com

闫天星，硕士研究生，主要研究方向为计算机视觉、图像处理、人脸三维重建等。

E-mail：1394698398@qq.com