Lightweight Facial Expression Recognition Method Based on Enhanced Attention*

ZHUANG Jianjun[1,2*], CHEN Yongqiang[1], JIN Xin[1], ZHOU Xiaoyan[1]
(1.School of Electronic and Information Engineering, Nanjing University of Information Science and Technology, Nanjing Jiangsu 210044, China; 2.Institute for AI in Medicine, School of Artificial Intelligence, Nanjing University of Information Science and Technology, Nanjing Jiangsu 210044, China)

Abstract: To address the issues of model complexity, large parameter size, and difficulty in capturing key feature information in the field of facial expression recognition using deep learning models, a lightweight facial expression recognition network model is proposed. This model utilizes depthwise separable convolutions instead of traditional convolutions and replaces fully connected layers with a fusion of adaptive max and average pooling, significantly reducing parameter count and computational load. An SFS (scale feature selection) algorithm is introduced to enhance the expression of critical scale features within the CBAM attention mechanism. To alleviate the problem of imbalanced sample categories in facial expression datasets, a joint loss function is designed to enhance the model's focus on challenging samples. The proposed approach is validated on three datasets: FER2013, FER2013Plus, and CK+. The recognition rates achieved are 68.52%, 87.58%, and 95.51%, respectively, with a parameter count of only 0.33 M. The model's size is much smaller than that of most mainstream methods while maintaining high accuracy, providing valuable insights for the lightweight implementation of deep learning-based facial expression recognition models.

Key words: expression recognition; attention mechanism; depthwise separable convolution; joint loss
EEACC: 6135 **doi**: 10.3969/j.issn.1005-9490.2026.04.034

基于改进注意力的轻量化面部表情识别方法*

庄建军[1,2*], 陈永强[1], 金 鑫[1], 周晓彦[1]
(1.南京信息工程大学电子与信息工程学院, 江苏 南京 210044;
2.南京信息工程大学人工智能学院, 南京信息工程大学-中大医院智慧医疗研究院, 江苏 南京 210044)

摘 要: 为解决目前在表情识别领域中深度学习模型存在的模型复杂、参数量大、难以获得关键特征信息等问题, 提出一种轻量化面部表情识别网络模型。该模型使用深度可分离卷积代替传统卷积, 并使用自适应最大、平均池化融合的方式替换全连接层, 从而大幅减少参数量和计算量; 提出 SFS(Scale Feature Selection)算法对 CBAM 注意力进行改进, 增强关键尺度特征的表达; 为缓解人脸表情数据集样本类别不平衡问题, 设计了联合损失函数以提高模型对难分样本的关注。分别在 FER2013、FER2013Plus 和 CK+三个数据集上进行验证, 识别率分别达到 68.52%、87.58%和 95.51%, 参数量仅为 0.33 M, 在保证精度的前提下, 模型体积远小于多数主流方法, 为深度学习表情识别模型的轻量化实现提供了有价值的参考。

关键词: 表情识别; 注意力机制; 深度可分离卷积; 联合损失
中图分类号: TP183 **文献标识码**: A **文章编号**: 1005-9490(2026)04-0938-08

面部表情是人表达情感和意图最自然、最普遍的方式之一[1]。Mehrabian[2]通过研究发现, 在人类面对面的交谈中, 表情传达的信息占 55%的比例, 而只有 7%取决于说话者的内容。在人工智能领域, 特别是计算机视觉中, 人脸表情识别已经引起了图像处理研究人员的广泛关注, 并在人机交互、安全驾驶[3]和在线教育中找到了广泛的应用前景。

人脸表情识别的基础是从图像数据中准确提取特征, 传统的特征提取主要靠人工设计特征提取器, 比如局部二值模式(Local Binary Patterns, LBP)[4]、Gabor 小波变换[5]和梯度方向直方图(Histogram of Oriented Gradient, HOG)[6]等, 但它们只能提取人脸的浅层语义特征, 这限制了模型在复杂情况下的准确性。深度学习技术的出现提供了一种有效的方法来提取人脸的深层语义特征, 而卷积神经网络(Convolutional Neural Network, CNN)是深度学习领域中最经典的网络模型之一, 当前表情识别领域的主流方法多采用基于卷积神经网络的深度学习方案。如 He 等[7]提出一种基于三通道卷积神经网络(Three-Channel Convolutional Neural Network, HFT-CNN)的多层特征识别算法, 不同通道提取不同面部位置特征, 最后融合三个网络通道提取的特征, 得到表情分类结果。

项目来源: 国家自然科学基金(62272234); 江苏高校"青蓝工程"项目
收稿日期: 2024-04-09 **修改日期**: 2024-05-29

然而随着 CNN 的不断发展,网络模型的结构复杂度呈不断提升的趋势,为了实现网络模型轻量化以满足现实场景的要求,一些轻量化模型被提出。MobileNet[8]采用深度可分离卷积结构,并引入宽度乘数实现模型大小的可调节,能够满足在移动设备上的运行。EfficientNet[9]通过复合缩放方法和全局深度可分离卷积等优化方法进一步提高了轻量化模型的性能,但相对于 MobileNet,EfficientNet 引入的复合缩放和全局深度可分离卷积的设计,使得整个模型结构相对较为复杂。GhostNet[10]通过在每个卷积核中选择一小部分权重,即采用低秩逼近的“Ghost”卷积,从而降低了模型的参数数量,但是“Ghost Module”等辅助分支的设计,同样使得模型整体结构更为复杂。注意力机制使得模型能够更加关注人脸表情识别任务中具有情感信息的重要区域,为人脸表情识别领域带来了显著的性能提升。Han 等[11]在 MobileNet 网络中引入高效通道注意力对网络特征通道进行重新校准。Guo 等[12]提出了一种基于双码 LBP 层空间注意力网络(Double-Code LBP-Layer Spatial-Attention Network,DLSANet)的方法,在主干网络中嵌入了一个空间注意力模块将网络注意力集中在各种表情差异显着的区域。Wu 等[13]在 VGG19 网络中加入改进的通道注意力机制,将原注意力中的 ReLU 激活函数替换为 Mish 激活函数,加速收敛,提升网络对面部细节关注度,但是 VGG 网络本身参数过多,计算成本高难以满足现实场景需求。以上方法都只关注到单一维度的特征注意力,而未能充分利用多维度信息。He 等[14]在所提模型中引入了 CBMA 注意力机制,构建了 DscCBAM 模块,使模型同时照顾到通道和空间的重要特征,发掘出面部重要区域。

基于以上内容,本文提出了一种轻量级深度可分离卷积残差网络,使用深度可分离卷积代替传统卷积构建了一个结构更加简洁高效的网络模型。对引入的注意力机制添加不同尺寸的卷积核并提出尺度特征选择(Scale Feature Selection,SFS)算法进行强特征选择,保留关键面部特征。放弃使用全连接层输出模型结果,使用自适应最大、平均池化融合的方式提取整个特征图的全局信息用于模型的最终预测。并且在训练过程中采用联合损失函数缓解数据集样本类别不均衡带来的问题,增强了模型的鲁棒性。实验结果表明所提方法具备有效性。

1　网络模型

1.1　网络模型设计

为了在保证模型性能的前提下减小模型体积,本文提出了一个轻量化的深度可分离残差网络模型。首先设计了深度可分离残差块(Depthwise Sparable Residual Block,DSRB)作为基础,DSRB 使用了残差连接技术,引入前向连接,有效预防了梯度消失问题,模块结构如图 1(a)所示,该模块由深度可分离卷积层、BatchNorm(归一化)、激活函数和最大值池化层构成,最终输出特征为输入特征尺寸的一半。在 DSRB 的基础上引入 ICBAM 注意力机制,构建了深度可分离混合注意力残差块(Depthwise Separable Residual Block with Hybrid Attention,DSRB-HA),结构如图 1(b)所示。为增加网络深度以提升对高层语义特征的提取,设计了 DSRB_NoPooling 残差块,该模块是在 DSRB 基础上去掉了最大池化层,使得输入尺寸和输出尺寸相同,结构如图 1(c)所示。

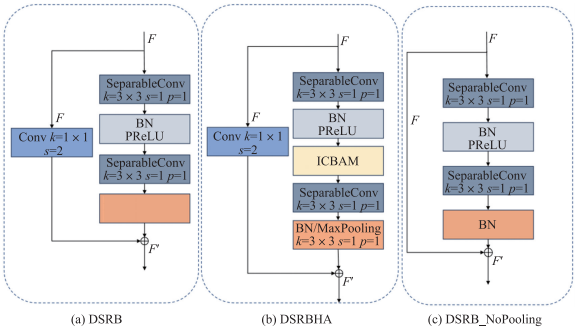

图 1　搭建网络模型的三种残差块

整体网络结构如图 2 所示,首先是两层 3×3 卷积层,用于初步的特征提取和局部信息处理,为后续

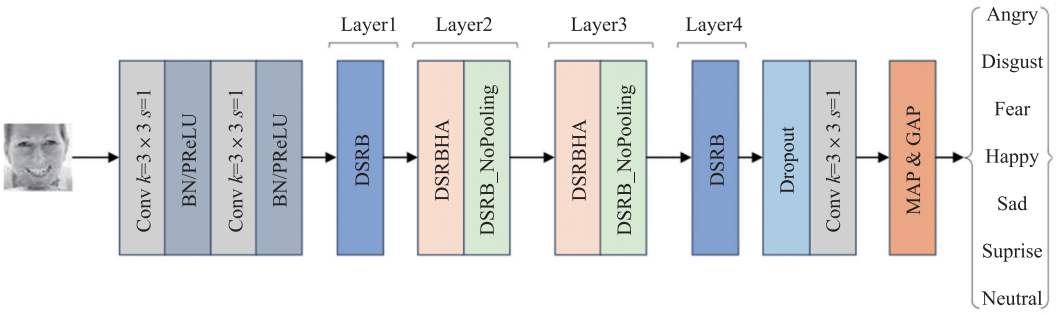

图 2　本文所提出的网络模型结构

的更深层次分析奠定基础。然后将提取到的特征送入到模型的 4 个层,其中第一层和第四层是 DSRB。第二和第三层是由一个 DSRBHA 和一个 DSRB_NoPooling 构成,在这些层之后,采用一个 3×3 卷积层使特征图输出尺寸为 7×3×3,在此卷积层之前加入了 Dropuot 层使训练过程中随机关闭网络中的一部分神经元以减少过拟合,提高模型的泛化能力。最后采取全局平均池化和全局最大值池化的融合代替全连接层,计算公式为:

$$output=Sigmoid(\alpha)\times x_{max}+(1-Sigmoid(\alpha))\times x_{mean} \quad (1)$$

式中:α 是一个可学习参数,随着网络训练进行更新寻找最优值,x_{max} 和 x_{mean} 分别是最大池化和平均池化,output 为模型输出。

1.2 深度可分离卷积

深度可分离卷积是一种可以减少卷积计算量和参数量的卷积方法。基本思想是将标准卷积拆分成对空间信息的提取和对通道信息的融合两个子步骤。第一步称为深度卷积(Depthwise Convolution),是对输入的每个通道分别进行卷积,如图 3 所示。在此步骤中,对于一个具有 D 个通道的输入,将有 D 个卷积核,每个核的尺寸是 $1\times k\times k$,这些卷积核只在单个通道上进行空间特征提取,这样可以提取输入各通道的空间信息,在这个步骤中不改变通道数。

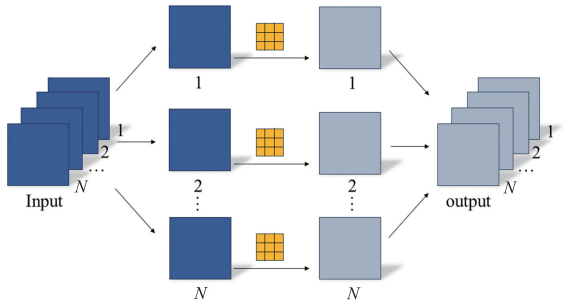

图 3 深度卷积

第二步是进行 1×1 卷积,即逐点卷积(Pointwise Convolution),如图 4 所示。这一步的作用是整合步骤一提取到的空间信息,并重新组合通道,逐点卷积可以自由设置输出通道数。相比于标准卷积,深度可分离卷积的总参数量减少可以通过式(2)近似计算:

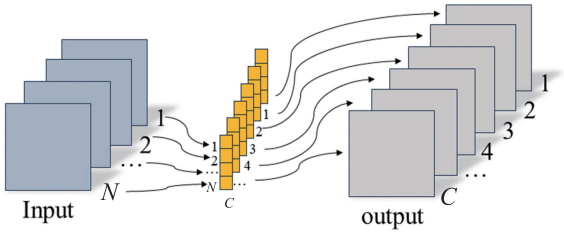

图 4 逐点卷积

$$\Delta=\frac{K\times K\times N\times M}{K\times K\times N+N\times M} \quad (2)$$

式中:K 是卷积核的尺寸,N 是输入通道数,M 是输出通道数,Δ 为标准卷积与深度可分离卷积参数的比值。特别是当 $K\times K$ 的值较大时(例如 3×3 或更大),由于深度可分离卷积将传统卷积分解为深度卷积和逐点卷积,其参数量可以减少 8 到 9 倍,这是因为在大卷积核尺寸下,标准卷积的参数数量随着卷积核尺寸的增大而显著增加,而深度可分离卷积的参数增加幅度较小。

1.3 ICBAM 注意力

1.3.1 通道注意力替换

CBAM 中的通道注意力(Squeeze-and-Excitation Networks,SENet)[15]采用了简单的多层全连接网络结构来学习各通道的注意力权重。这种结构对通道间复杂依赖关系的拟合能力较弱,难以建模通道间的相关性,对复杂非线性关系的拟合程度不高。这导致了 CBAM 学习到的通道注意力权重可能没有充分反映特征通道的重要性,降低了注意力机制的有效性。

高效通道注意力(Efficient Channel Attention,ECANet)是对 SENet 的一种改进,其结构如图 5 所示,ECANet 在全局平均池化后直接应用一维卷积来完成跨通道间的信息交互,卷积核大小 k 能够根据通道数自适应变化,k 与 C 之间存在如式的映射关系。

$$k=\psi(C)\left|\frac{\log_2(C)}{\gamma}+\frac{b}{\gamma}\right|_{odd} \quad (3)$$

式中:γ 和 b 分别为 2 和 1,C 表示通道数,odd 表示最接近的奇数。通过一维卷积消除对全连接层的依赖,并能够根据输入数据的特点自适应调整重点通道,实现跨通道的交互,避免因降维导致的信息损失。

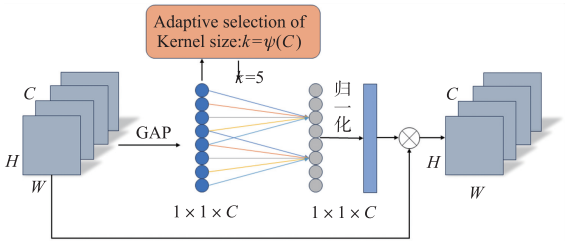

图 5 高效通道注意力

1.3.2 SFS(Scale Feature Selection)算法

CBAM 中的空间注意力只能在单一尺度上捕捉图像的局部关系,因此只能获得单一尺度的特征空间,为了提高模型对整体图像结构的理解能力,获得更丰富的特征表示,在原始空间注意力上引入了不同尺度的卷积核(3×3、5×5、7×7),以扩展感受野的范围,改进后的空间注意力结构如图 6 所示。

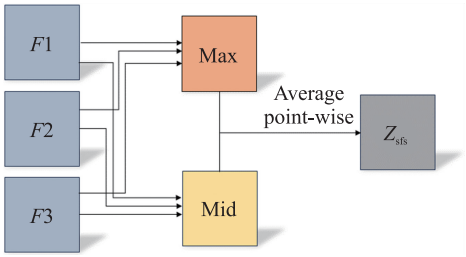

图 6　多尺度空间注意力

然而提取到的多尺度特征可能包含背景区域以及其他冗余信息从而导致关键面部细节被模糊。为此本文提出 SFS 算法对提取到的多尺度特征进行特征选择,具体来说就是在融合前消除不同尺度的弱特征,保留强特征,这样能够减少计算参数,并且尽可能地保留不同尺度特征中的有效特征。SFS 算法结构如图 6 所示,SFS 会将输入特征 $F1,F2,F3(1\times h\times w)$ 的相同位置进行比较,保留最高特征和中间特征,而最低特征将会被丢弃,然后将保留的特征进行逐点平均得到最后的特征图。SFS 算法的数学表达式如式:

$$Z_{sfs}=\frac{1}{2}(\max(x_{hw}^{F1},x_{hw}^{F2},x_{hw}^{F3})+\mathrm{mid}(x_{hw}^{F1},x_{hw}^{F2},x_{hw}^{F3})) \quad (4)$$

式中:Z_{sfs} 表示输出结果,x_{hw}^{F1} 表示在特征 $F1$ 中位于 (h,w) 位置的元素。

将改进后的 CBAM 命名为 ICBAM(improved CBAM),相比于原始的 CBAM,改进后的注意力有效地实现了跨通道的交互,能够对不同区域大小的特征进行筛选并强调其中更加重要的部分。

1.4　联合损失函数

交叉熵损失(Cross Entropy Loss)是一种标准的损失函数,用于衡量模型预测的概率分布和目标真实分布之间的差异。它的优点在于简单、直观且在多种任务上效果良好,特别是在类别分布相对平衡的数据集中。其公式如式所示。

$$L_C=-\sum_{i=1}^{C}x_i\ln P_i \quad (5)$$

式中:C 是类别数,x_i 是样本 x 的第 i 个标签,P_i 是模型预测样本 x 属于第 i 个类别的概率。

然而人脸表情之间的界限可能模糊不清,并且一些表情难以获取,如厌恶。这就导致数据集分布不平衡,包括不同类别样本数量上的不平衡,以及简单示例与困难示例之间的不平衡。焦点损失(Focal Loss)通过为每个样本添加两个基于预测正确性的调节因子来工作,这种方法使得模型在训练过程中更加关注难以预测正确的样本,在一定程度上缓解了数据集不平衡问题。焦点损失函数计算公式如式(6):

$$L_f=-\alpha_i\times(1-P_i)^{\gamma}\times\ln(P_i) \quad (6)$$

式中:P_i 是样本属于类别 i 的概率,i 取值为 1,2,3…C,C 是类别数,α_i 是类别 i 的权重系数,γ 是可调参数。式中$(1-P_i)^{\gamma}$ 是焦点因子,它的作用是调整样本的权重,特别是对难以分类的样本。具体来说:当 P_i 接近 1 时,表示模型对正样本的预测很准确,$(1-P_i)^{\gamma}$ 的值趋近于 0,对应的权重较小,损失减小。当 P_i 接近 0 时,表示模型对正样本的预测不准确,$(1-P_i)^{\gamma}$ 的值增大,对应的权重较大,损失增大。这种机制使得焦点损失函数更加关注那些对模型来说难以分类的样本,而不是让所有样本的损失都对模型的更新产生相同的影响。

参数 γ 决定了这个焦点因子的调整程度,进而控制模型对难以分类样本的关注度。文献[16]中提到,γ 取值范围应该在 0~5 之间,本文通过对 γ 取值为 1、2、3、4、5 进行对比实验,旨在找出使模型性能最好的 γ 值。通过实验发现,当 γ 取值为 2 时,模型性能(准确率)最佳。因此,在后续实验中,选择将 γ 设定为 2。参数 α 用来缓解各类别样本数不均衡的问题,其取值与该类别的样本数相关,在实验中本文将 α 取值为类别样本数逆频率。

联合损失函数是交叉熵损失和焦点损失的结合,对整体结果起到联合优化效果,既能保证了整体表情识别性能,又减轻样本不均衡导致的长尾效应。具体为式:

$$L_{all}=L_c+\lambda L_f \quad (7)$$

式中:L_{all} 为联合损失函数,λ 为焦点损失的占比参数。在联合损失函数作用下,通过调节参数 γ 和 α 能够降低难分类样本和样本数量不平衡的影响,调节参数 λ 能保证整体表情识别性能。

2　实验结果与分析

2.1　实验环境

本文所使用的实验环境为配置为:Win11 操作系统,i5-13400F CPU,NVIDIA GeForce RTX 3070 显卡,1.7 版本的 PyTorch 和 3.8 版本的 Python。所有试验均在同一环境下进行。

2.2　数据集

FER2013[17]是一份用于面部表情识别的数据集,该数据集包含 35 887 张 48×48 像素的面部图像。图像标签包括 7 种基本面部表情:生气、厌恶、恐惧、高兴、难过、惊讶、中性。但是不同类别的数据分布非常不均匀,例如高兴类别有 8 989 张图片,占总数的近 25%,而厌恶只有 547 张图片,占比仅 1.5%。这种数据分布不均衡可能会导致在模型训练中对某些类别样本的学习不足,影响模型的鲁棒

性和泛化能力。这就要求模型具备处理类别不均衡数据的能力,有较好的鲁棒性和泛化能力。图 7 是 FER2013 数据集样本分布。

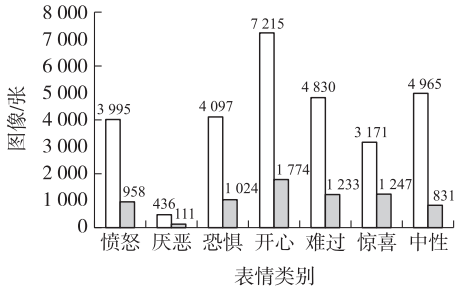

图 7 FER2013 数据集样布数量分布

FER2013Plus 数据集是通过对原始 FER2013 数据集重新标注而获得的,对 FER2013 中的所有图像进行了二次标注,并将表情分为 10 类,通过多标注者的标注结果进行汇总和多数表决的方式,修正了部分 FER2013 原数据集的错误标签,得到了标签质量更高的 FER2013Plus 数据集。本实验选取了其中 7 类基本表情进行测试。

CK+[18] 是表情识别研究中最常用的数据集之一,它包含了 123 名受试者,327 个含表情标签的序列,对于每一个图片序列,只有最后一帧被提供了表情标签,共有 327 个图像被标记。

图 8 数据集中不同表情类别的部分样本

2.3 实验结果及分析

在 FER2013 数据集训练过程中,采用动量优化的随机梯度下降算法更新网络参数,动量因子设置为 0.9,有助于加速训练过程并减少振荡。初始学习率为 0.01,总共迭代 500 次,在前 25 轮中实施了 Warmup 操作,即前 25 轮逐渐增大学习率,从较小值增长到设定的最大学习率。批量大小设置为 64,采用余弦退火学习率衰减策略(Cosine Annealing LR)学习率策略。使用 FER2013 中的训练集 28 709 张图片进行训练,使用测试集 3 859 张图片验证模型的性能。在进行训练之前,对训练样本进行数据增强,进行随机的缩放、旋转、偏移与反转等操作。FER2013 数据集上的实验结果如图 9 所示。

FER2013Plus 数据集与 FER2013 数据集相似,在实验中采用相同的训练策略,实验结果如图 10。在 CK+数据集训练时,由于训练数据较小,在保持其他参数不变的情况下,将批量大小设置为 32,总共迭代 150 次。使用 CK+中训练集的 720 张图片进行训练,测试集的 91 张图片进行测试验证模型的性能。训练前同样对数据样本增强,进行随机缩放、旋转、偏移与反转等操。图 11 是本文网络模型在 CK+数据集上的实验结果。

最终本文模型在仅有 0.33 MB 的参数量下,在 FER2013、FER2013Plus 和 CK+三个数据集上分别取得了 68.52%、87.58%和 95.51%的识别率,模型在保持轻量化的同时仍保持较高识别性能。FER2013 数据集中,恐惧和悲伤表情的识别率较低。分析其数据集存在错误标注、不同年龄、肤色、遮挡等干扰,增大了网络学习有效特征的难度。FER2013Plus 数据集更正了 FER2013 数据集中的错误标注信息,因此模型在 FER2013Plus 数据集上的识别准确率明显提高。在 CK+数据集中,生气表情的识别准确率相对较低,推测是因为生气和厌恶表情外观比较相似,模型难以分辨。其他表情都有较高的识别率,尤其是害怕、伤心和惊喜这三类表情,达到了 100%的识别率。

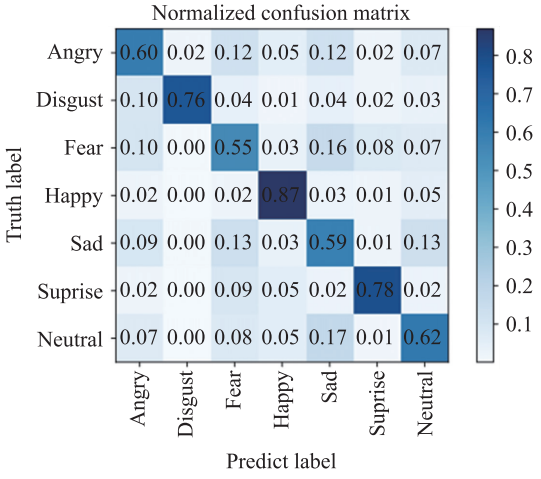

图 9 FER2013 数据集的实验结果

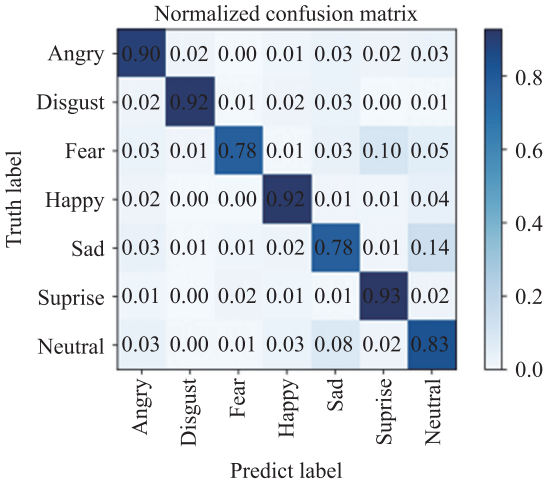

图 10　FER2013Plus 数据集的实验结果

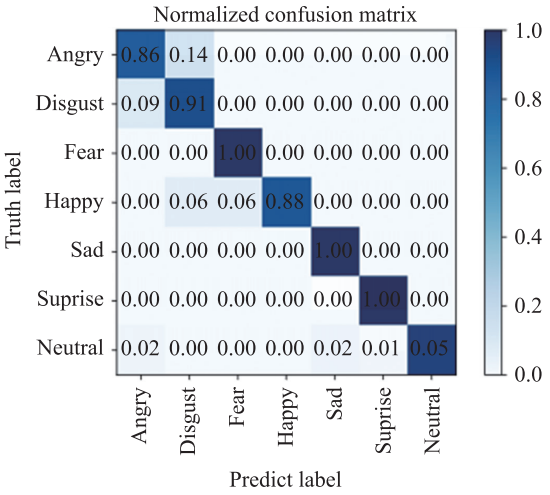

图 11　CK+数据集的实验结果

2.4　消融实验

为了验证本文所提出的改进模型的有效性,本文分别在三个数据集上进行了消融实验,其中 Basic 表示所提模型不加入注意力模块,并且训练时使用交叉熵损失函数。实验结果如表 1 所示。

表 1　消融实验

方法	FER2013/%	FER2013 Plus/%	CK+/%
Basic	65.21	83.66	92.32
Basic+ICBAM 注意力	67.54	85.52	94.15
Basic+联合损失	66.62	84.64	93.28
Basic+All	68.52	87.58	95.51

从消融实验结果可以看出,本文提出的两种改进措施:ICBAM 注意力、联合损失函数在不同数据集上均使模型性能有一定的提升,其中在 FER2013 数据集上,ICBAM 注意力使得模型预测准确率提高 2.33 个百分点,使用联合损失训练网络模型预测准确率提升了 1.41 个百分点,将所有改进加入 Basic 后模型预测准确率提升 3.31 个百分点,证明了所提改进的有效性。

为了验证本文所设计 ICBAM 注意力的有效性,分别使用 SE、CBAM、ECA 替换模型中的 ICBAM 注意力并在 CK+数据集上进行对比。表 2 的实验结果表明,相比于单一注意力,多注意力融合的方式能使模型有更大的提升。而改进后的 ICBAM 拥有多尺度特征提取能力并能够跨通道交互,使得模型特征提取更充分从而获得了更高的识别准确率。

表 2　不同注意力对比实验结果

注意力	准确率/%
SE	92.71
ECA	92.59
CBAM	94.46
ICBAM	95.51

2.5　对比试验

本文将所提模型与其他模型进行对比试验。从表 3 可以看出,与主流的轻量级模型如 MobileNetV3 以及 Xception 相比,本文所提出的模型在 FER2013 数据集上准确率提高的情况下,以极低的参数量取得绝对的优势,在 CK+数据集上准确率相比于 MobileNetV3 提高了 8.16 个百分点,与 Zhao[19] 所提出的基于注意力机制 ResNet 轻量网络持平。相比于同样轻量级网络 mini_Xception,参数量仅仅增加了 0.09 MB,但是在 FER2013、FER2013Plus 和 CK+数据集上分别提升了 2.96、6.72 和 6.29 个百分点。对比最近提出的一些方法,本文所提模型的准确率没有明显提高,但是参数量大大减小,如本文模型以 9%的参数量达到了 Yan[20] 提出改进 MobileNetV2 模型 98%的性能,表明了所提模型在边缘计算应用场景下具有一定的竞争性。

表 3　本文模型与主流模型以及其他方法对比实验结果

模　型	参数量/MB	准确率/%		
		FER 2013	CK+	FER2013Plus
ResNet18	42.73	71.06	92.72	87.92
MobileNetV3[21]	21.54	68.24	87.35	83.68
MobileNetV2	19.82	68.42	88.56	83.62
Xception[22]	5.28	67.32	86.63	82.52
miniXception[23]	0.24	65.56	89.22	80.86
Yan[20]	3.56	68.96	95.96	/
Zhao[19]	7.72	68.58	95.81	/
Sun[24]	0.55	68.32	94.88	86.42
Ours	0.33	68.52	95.51	87.58

3 结论

本文提出的基于改进注意力可分离残差网络模型,将传统的卷积替换成深度可分离卷积大大减少了模型的参数量,引入的 ICBAM 注意力模块使模型更加专注面部重要的区域,同时提出联合损失函数缓解数据集样布分布不均匀的问题。通过在 FER2013、FER2013Plus 和 CK+数据集上的验证以及与主流轻量化模型方法的对比,证明了本文方法的有效性,展示了在表情识别领域以极低的参数量实现良好效果的可能性,为其他领域深度学习模型的轻量化提供了有价值的参考。由于模型对于数据集中遭到光照、遮挡表情样本识别率较低,未来的研究可以探索更有效的注意力融合方式和网络结构,提高模型的抗干扰能力,进一步提高表情识别的准确率。

参考文献:

[1] Ekman P. Universality of Emotional Expression? A Personal History of the Dispute[M]. Darwin,C. The Expression of the Emotions in Man and Animals,3rd Editon Oxford University Press:New York,NY,USA,1998:363-393.

[2] Mehrabian A. Silent Messages:Implicit Communication of Emotions and Attitudes[M]. 2nd ed. Belmont, Calif: Wadsworth Pub. Co.,1981:165-189.

[3] Li W,Cui Y,Ma Y,et al. A Spontaneous Driver Emotion Facial Expression(Defe) Dataset for Intelligent Vehicles:Emotions Triggered by Video-Audio Clips in Driving Scenarios[J]. IEEE Transactions on Affective Computing, 2023, 14(1):747-760.

[4] Jia Q,Gao X,Guo H,et al. Multi-Layer Sparse Representation for Weighted LBP-Patches Based Facial Expression Recognition[J]. Sensors,2015,15(3):6719-6739.

[5] Zhang Y, Wang S, Huo Y, et al. Feature Extraction of Brain MRI by Stationary Wavelet Transform and Its Applications[J]. Journal of Biological Systems, 2010, 18(spec01):115-132.

[6] Déniz O,Bueno G,Salido J,et al. Face Recognition Using Histograms of Oriented Gradients[J]. Pattern Recognition Letters,2011,32(12):1598-1603.

[7] He Y,Zhang Y,Chen S,et al. Facial Expression Recognition Using Hierarchical Features with Three-Channel Convolutional Neural Network[J]. IEEE Access, 2023, 11: 84785-84794.

[8] Howard A G,Zhu M,Chen B,et al. MobileNets:Efficient Convolutional Neural Networks for Mobile Vision Applications[J]. arXiv Preprint arXiv:170404861,2017.

[9] Tan M,Le Q. Efficientnet:Rethinking Model Scaling for Convolutional Neural Networks[C]//Proceedings of the International Conference on Machine Learning, Long Beach,CA,USA:PMLR,2019:6105-6114.

[10] Han K,Wang Y,Tian Q,et al. GhostNet:More Features from Cheap Operations[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition,Seattle,WA,USA:IEEE,2020:1577-1586.

[11] 韩兴,张红英,张媛媛. 基于高效通道注意力网络的人脸表情识别[J]. 传感器与微系统,2021,40(1):118-121.

[12] Guo X,Lu S,Wang S,et al. DLSANet:Facial Expression Recognition with Double-Code LBP-Layer Spatial-Attention Network[J]. IET Image Processing, 2023, 17(9):2659-2672.

[13] 吴怡啄,杨定礼,周辉,等. 基于改进 VGG 网络的面部表情识别方法[J]. 电子器件,2023,46(4):1062-1069.

[14] He Y. Facial Expression Recognition Using Multi-Branch Attention Convolutional Neural Network[J]. IEEE Access,2022,11:1244-1253.

[15] Hu J,Shen L,Sun G. Squeeze-and-Excitation Networks[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence,2020,42(8):2011-2023.

[16] Lin T Y,Goyal P,Girshick R,et al. Focal Loss for Dense Object Detection[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence,2020,42(2):318-327.

[17] Goodfellow I 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.

[18] 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,San Francisco,CA,USA: IEEE,2010:94-101.

[19] 赵晓,杨晨,王若男,等. 基于注意力机制 ResNet 轻量网络的面部表情识别[J]. 液晶与显示,2023,38(11): 1503-1510.

[20] 严春满,张翔,王青朋. 基于改进 MobileNetV2 的人脸表情识别[J]. 计算机工程与科学, 2023, 45(6): 1071-1078.

[21] Howard A,Sandler M,Chu G,et al. Searching for MobileNetv3[C]//2019 IEEE/CVF International Conference on Computer Vision, Seoul, Korea (South): IEEE, 2019: 1314-1324.

[22] Chollet F. Xception:Deep Learning with Depthwise Separable Convolutions[C]//2017 IEEE Conference on Computer Vision and Pattern Recognition,Honolulu,HI,USA:IEEE, 2017:1251-1258.

[23] Arriaga O,Valdenegro-Toro M,Plöger P. Real-Time Convolutional Neural Networks for Emotion and Gender Classification[J]. arXiv Preprint arXiv:171007557,2017.

[24] Sun Q, Chen Y, Yang D, et al. EC-RFERNet: An Edge Computing-Oriented Real-Time Facial Expression Recognition Network[J]. Signal, Image and Video Processing, 2023:1-17.

庄建军(1973—),男,教授,博士,主要研究方向为机器视觉、智能信号处理等,jjzhuang@nuist.edu.cn;

陈永强(1999—),男,研究生,主要研究方向为图像处理、表情识别等,2629132045@qq.com;

金　鑫(1998—),男,研究生,主要研究方向为计算机视觉、目标检测等,1278491940@qq.com;

周晓彦(1980—),女,副教授,博士,主要研究方向模式识别、机器学习,001385@nuist.edu.cn。