ISSN 1004-9037,CODEN SCYCE4
Journal of Data Acquisition and Processing Vol. 39,No. 6,Nov. 2024,pp. 1493－1504
DOI:10. 16337/j. 1004-9037. 2024. 06. 017

http://sjcj. nuaa. edu. cn
E-mail:sjcj@nuaa. edu. cn
Tel/Fax:＋86-025-84892742

基于金字塔分割注意力和联合损失的表情识别模型

谷　瑞[1,2]，顾家乐[2]，宋翠玲[1]

(1. 南京大学数字经济与管理学院,南京 210003;2. 苏州工业园区服务外包职业学院,苏州 215123)

摘　要：如何提取多尺度特征和建模远程通道间的语义依赖仍是表情识别网络面临的挑战。本文提出一种基于金字塔分割注意力的残差网络(Residual network based on pyramid split attention，PSA-ResNet)模型,该模型将ResNet50残差模块中的3×3卷积替换成金字塔分割注意力,以有效提取多尺度特征,增强跨通道语义信息的相关性。同时,为缩小同类表情之间的差异,扩大不同类表情之间的距离,在训练过程中引入了Softmax loss和Center loss联合损失函数优化模型参数。本文所提出的方法在Fer2013和CK＋两个公开的数据集上进行仿真实验,分别取得了74.26％和98.35％的准确率,进一步证实了该方法相比前沿算法具有更好的表情识别效果。

关键词：表情识别;金字塔分割注意力;多尺度特征;残差网络

中图分类号：TP183　　**文献标志码**:A

An Expression Recognition Model Based on Pyramid Split Attention and Joint Loss

GU Rui[1,2]，GU Jiale[2]，SONG Cuiling[1]

(1. School of Digital Economy and Management，Nanjing University，Nanjing 210003,China; 2. Suzhou Industrial Park Institute of Services Outsourcing，Suzhou 215123,China)

Abstract：How to extract multi-scale features and model semantic dependencies between remote channels remains a challenge for expression recognition networks. This paper proposes a residual network based on pyramid split attention (PSA-ResNet)，which replaces the 3 × 3 convolution in the ResNet50 residual module with PSA to effectively extract multi-scale features and enhance the correlation of cross channel information. In order to reduce the differences between similar expressions and expand the distance between different types of expressions，a joint loss function optimization parameter of Softmax loss and Center loss is introduced during the training process. The proposed model is simulated on two publicly available datasets，Fer2013 and CK＋，and achieves accuracies of 74.26％ and 98.35％，respectively，further confirming that this method has better recognition results compared to cutting-edge algorithms.

Key words: expression recognition; pyramid split attention(PSA); multi-scale feature; residual network

引　　言

人脸表情蕴含着十分丰富的情感信息,面部表情的变化反映人际交往中心理情绪的波动情况,据统计,日常交流中55％的信息通过面部表情传达[1]。人脸表情识别是视觉检测技术的进一步发展,如

基金项目:2023年江苏省高职院校教师专业带头人高端研修项目(2023TDFX010)。
收稿日期:2024-04-07;**修订日期**:2024-05-22

何让计算机正确识别人脸中蕴含的表情信息是计算机视觉领域一项具有重要意义和挑战性的工作。近年来,人脸表情识别在智慧课堂、侦查审讯、安全驾驶、医疗诊断等领域受到了广泛关注,逐渐成为学术界和工业界的研究热点[2]。

在传统的表情识别方法中,首先利用手工设计的特殊算子提取特征,然后将特征向量送入诸如支持向量机、K临近算法之类的分类器中输出识别结果。常见的特征提取算法有提取纹理特征的局部二值模式算法[3]、提取边缘特征的方向梯度直方图算法[4]、提取几何特征的主动形状模型算法[5]等。虽然传统的特征提取算法取得了一定的效果,但是手工设计的特征受限于设计者的经验和知识,且只能提取浅层特征,缺乏高层语义信息,导致分类准确率较低。

随着深度学习的发展,卷积神经网络在计算机视觉领域得到了广泛应用。不同于传统的特征提取算法,卷积神经网络通过多层卷积和非线性变换,自动提取图像的深层语义信息。在表情识别领域,程学军等[6]提出一种改进的VGG16人脸表情识别方法,实现表情在多场景下的精准区分。赵晓等[7]以ResNet18为主干网络,引入倒残差结构优化表情识别模型。关小蕊等[8]在ResNet50中嵌入深度可分离卷积,增强多视角下特征的有效提取。Mollahosseini等[9]以Inception为基础,增加网络的宽度与深度以提高模型的泛化性。Arul等[10]提出基于DenseNet152稠密深度神经网络模型,实现自然场景下人脸表情识别。Liu等[11]通过改进的LetNet和ResNet分别提取表情特征,然后将两个特征向量拼接起来用于分类。基于神经网络的方法提高了表情识别的准确率,然而可用于辨别的表情主要集中在眼睛、鼻子和嘴巴等部位,因此增加这些关键特征的权重有助于改善表情识别效果。

为表征有效的特征信息,一些研究者将注意力引入到表情识别中,以增强对关键特征的提取能力。Li等[12]基于通道注意力使用全局平均池化和全连接来重新校准通道权重,增强对重要表情通道的表征能力。Yao等[13]分别在表情特征的通道维度和空间维度上进行压缩和重新加权。Minaee等[14]认为主要的表情特征集中在某些显著部位,引入Spatial transformer注意力聚焦表情丰富的区域。Liu等[15]改进高效通道注意力将特征图的深度与空间信息结合。实践表明,虽然注意力机制可增强表情关键特征的表达,但通道或空间注意只能有效捕获局部信息,无法建立远程的通道依赖关系,而人脸表情复杂多样,单一尺寸的卷积核无法充分提取表情特征,导致大量有效特征丢失。为了充分提取表情特征,本文将金字塔分割注意力(Pyramid split attention,PSA)融入表情识别网络,通过提取多尺度特征图的通道方向注意权重建立远程依赖关系,提高表情识别准确率。其主要贡献如下:

(1) 引入金字塔分割注意力模块[16],使用多尺度分组卷积构建特征金字塔,生成不同分辨率和深度的特征图,以有效提取多尺度特征,增强网络的特征学习能力。

(2) 用PSA替换ResNet50网络残差模块中的3×3卷积,形成基于金字塔分割注意力的残差网络(Residual network based on PSA, PSA-ResNet)模型,将不同尺度的特征信息整合到每个通道的特征图上,使用通道注意力增强跨通道语义信息的相关性,产生更好的像素级注意力,提升模型识别的准确度。

(3) 为缩小同类表情之间的距离,降低模型误判的概率,使用Softmax loss和Center loss联合损失函数训练网络,并在Fer2013与CK+两个数据集上进行大量实验,证明了本文方法的有效性。

1 表情识别模型的构建

为提高多尺度关键特征的跨通道表征能力,捕捉不同表情间的细微变化,本文提出一种基于金字塔分割注意力的残差网络模型,对各种人脸表情进行识别,该模型的体系结构如图1所示。

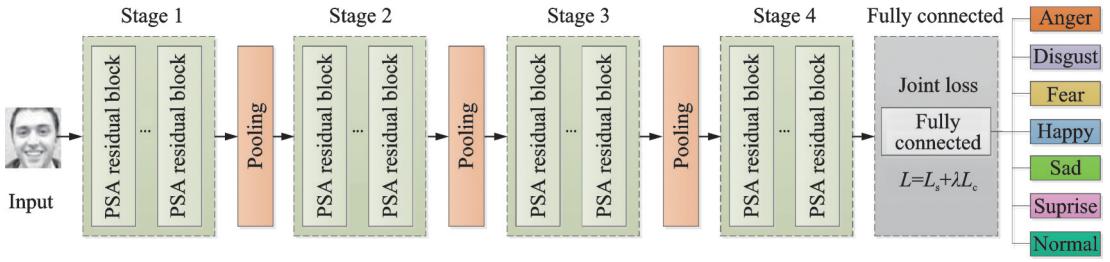

图1　PSA-ResNet网络结构

Fig.1　PSA-ResNet architecture

首先基于ImageNet数据集预训练ResNet50模型,得到网络初始权重参数;然后将其迁移到PSA-ResNet模型上,学习丰富的多尺度特征,将不同尺度的信息整合到通道级特征图上,重新校准多尺度特征跨通道注意力权重,增加人脸表情变化显著性区域的权重;最后在训练过程中采用Softmax loss和Center loss联合损失函数进行参数优化,降低表情识别误判的概率,提高模型识别的准确性。

1.1　PSA残差块

为了提取高层语义信息,卷积神经网络的层数越来越多,从8层的AlexNet[17],到19层的VGG-Net[18]、22层的GoogLeNet[19]。但单纯增加网络层数并不能增强模型的特征学习能力,当模型到达一定的深度后,会出现网络退化问题,从而导致精度的下降。ResNet[20]通过残差学习和恒等映射,有效地解决了深层神经网络中出现的梯度爆炸和梯度消失现象,提高了神经网络的训练效率。在ResNet系列中广泛应用的是ResNet34、ResNet50和ResNet101。与ResNet34相比,ResNet50使用3层残差块代替2层残差块,在保持模型精度的同时又能大幅度降低参数量,而ResNet101层数过多,过度关注语义信息而忽略细节特征,因此本文选取ResNet50作为表情识别特征提取的基线模型。

ResNet50的layer2、layer3、layer4、layer5分别包含3、4、6、3个残差块,ResNet残差块结构如图2(a)所示。将残差块中3×3卷积对应的位置替换为PSA,得到新的基于金字塔分割注意力的残差块,其结构如图2(b)所示。相应的,将PSA残差块按ResNet50网络风格进行堆叠,得到一种新的基于PSA-ResNet的网络模型,采用该模型对表情进行识别。

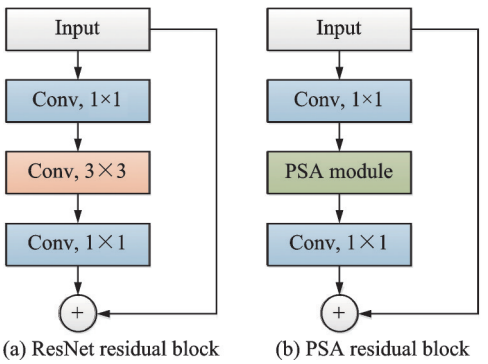

图2　ResNet和PSA残差块

Fig.2　ResNet and PSA residual blocks

1.2　金字塔分割注意力机制

注意力机制可以让神经网络对不同部分的数据,赋予不同的权重,从而选择对当前任务最关键的信息。为建立高效的注意力机制,本文引入PSA模块,该模块通过多尺度金字塔卷积结构和通道注意力机制增强多尺度特征和跨通道语义信息的相关性,捕获不同层次和不同粒度的表情信息,提升表情预测的精度。

PSA模块结构如图3所示,对于高度为H、宽度为W、通道数为C的特征图,分别经过以下4个步骤:(1)使用分割融合模块(Split and concat,SPC)从通道方向将输入的特征向量分割为若干组,利用金字塔结构的不同尺度的卷积核提取不同通道级上的含有多尺度的特征信息;(2)将SPC模块的输出送

入SE weight模块，计算不同通道的权重值，得到每个通道特征图的注意力向量；(3)将注意力向量通过Softmax函数进行归一化，重新标定每个通道的注意力权重，得到新的跨通道注意力权重；(4)将上述得到的多尺度空间信息和跨通道注意力权重按元素进行点乘，从而提取丰富的多尺度特征并建模多尺度特征不同通道之间的相关性。

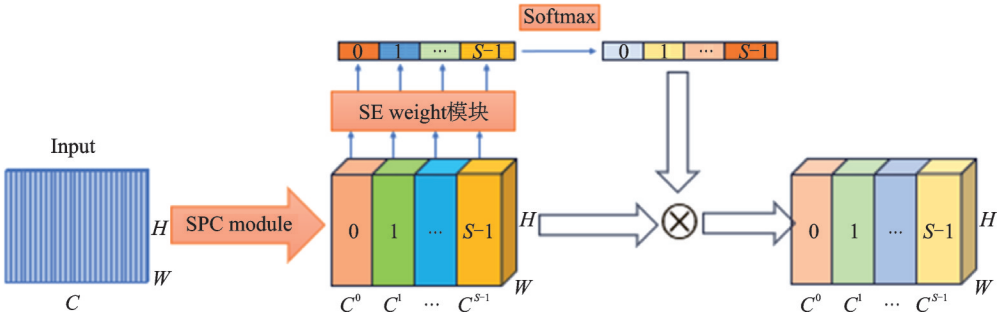

图3 PSA模块结构

Fig.3 PSA module structure

1.2.1 分割融合SPC模块

SPC模块计算过程如图4所示，使用多尺度分组卷积构建特征金字塔，提取每个通道特征图上不同尺度的空间信息，生成不同分辨率和深度的特征图，形成更加丰富的特征表示，增强网络的特征学习能力。

首先将输入的特征图从通道上拆分为S(本模型取$S=4$)个部分，表示为$[X_0, X_1, \cdots, X_{S-1}]$，每个部分的通道数量满足$C'=\frac{C}{S}$，分别使用不同大小的卷积核，每部分卷积核大小满足

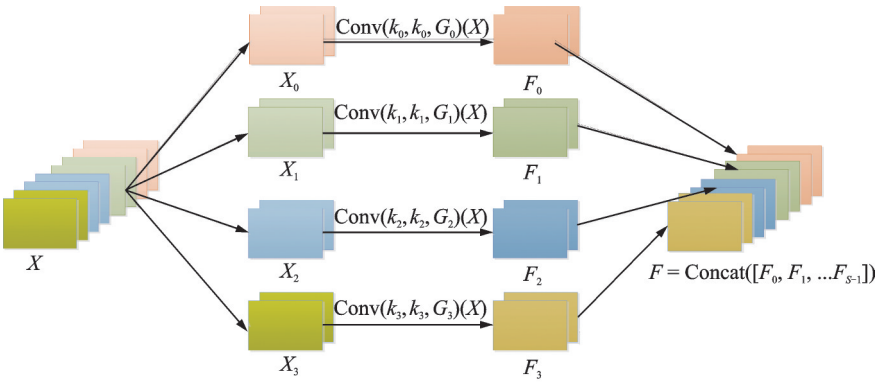

图4 分割融合模块计算过程

Fig.4 Calculation process of SPC

$$k_i=2\times(i+1)+1 \quad i=0,1,\cdots,S-1 \tag{1}$$

卷积核尺寸的增加会引起计算量的迅速增大，本文在SPC模块中对分割的每部分的特征应用分组卷积，有效地避免了此类问题，分组数量与卷积核的大小满足如下关系

$$G_i=2^{\frac{k_i-1}{2}} \tag{2}$$

因此，多尺度特征图的生成函数可以表示为

$$F_i=\mathrm{Conv}(k_i\times k_i, G_i)(X) \quad i=0,1,\cdots,S-1 \tag{3}$$

最后,在通道方向上拼接多尺度特征,得到一个叠加的多通道特征,拼接函数为

$$F=\mathrm{Concat}([F_0,F_1,\cdots,F_{S-1}]) \tag{4}$$

1.2.2　通道注意力SE模块

通道注意力机制能重点关注重要的特征通道,并抑制无意义的特征通道,达到提升网络性能的目的。SE模块结构如图5所示,主要由挤压和激励两部分组成。

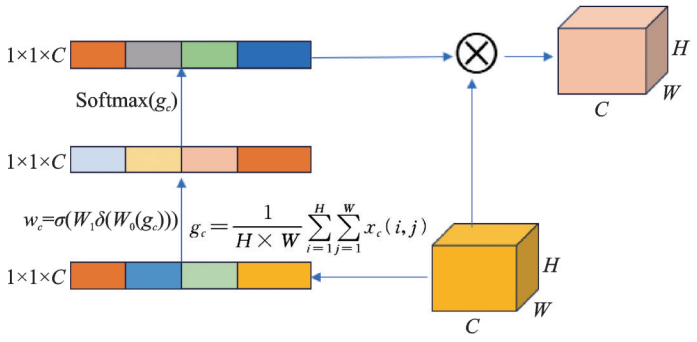

图5　SE模块结构

Fig.5　SE module structure

首先使用全局平均池化对输入特征图的通道维度进行压缩,得到一个全局特征图,其维度为$1\times1\times C$。若第c个通道的输入为x,则全局平均池化的计算公式为

$$g_c=\frac{1}{H\times W}\sum_{i=1}^{H}\sum_{j=1}^{W}x_c(i,j) \tag{5}$$

然后利用全连接层和激励函数计算不同通道之间相互关系,获得注意力权重。第c个通道的注意力权重可以表示为

$$w_c=\sigma(W_1\delta(W_0(g_c))) \tag{6}$$

式中:σ和δ分别代表ReLU和sigmoid激活函数的操作,两个全连接层权重分别用W_0和W_1表示。全连接层通过将输入数据与权重矩阵进行线性组合,并通过激活函数引入非线性变换,更加有效地建模远程通道间的语义依赖。

最后使用Softmax函数将注意力权重向量归一化,与多尺度特征提取模块的输出进行加权融合,从而更好地表达图像的特征。

1.3　联合损失函数

Softmax loss损失函数虽然保证了不同类别清晰可分,但是没有考虑类别内部的差异性。在表情识别任务中,不同的人做同一个表情的差异可能比一个人做不同的表情差别大得多,若仅仅使用Softmax loss可能导致表情的误判,进而影响模型对表情的正确识别。为了获得更好的表情分类效果,在损失函数设计上需要考虑如何缩小同类表情的距离、扩大不同类别表情间的距离。

因此本文引入Softmax loss和Center loss联合损失函数。Center loss缩小类内距离的能力较强,使得同类表情的数据表现得更加紧凑,有利于提高分类效果,本文将其引入到网络模型中,计算过程为

$$L_c=\frac{1}{2m}\sum_{i=1}^{m}\left\|x_i-c_{yi}\right\|_2^2 \tag{7}$$

式中:m表示数据集中样本个数,x_i表示输入的第i个表情特征,c_{yi}表示该类表情所有样本的中心。每个

样本特征距该类样本中心距离的平方和越小，类内差距就越小。PSA-ResNet模型联合损失函数可表示为

$$L = L_s + \lambda L_c \tag{8}$$

式中：L_s表示Softmax loss损失函数，λ为权重参数，用于调节Center loss在联合损失函数中的比重。

2 实验与分析

为有效地对本文提出的模型进行评估，分别在Fer2013和CK+两个数据集上进行训练和测试，并与当前主流的方法进行对比实验。

2.1 实验数据集

(1) Fer2013数据集：该数据集是由35 886张分辨率为48像素×48像素的灰度表情图组成，其中训练集28 708张，验证集和测试集各3 589张。每张图片均被打上标签类别，共包含7类表情，分别对应数字0～6，具体对应标签的中英文如下：0生气(Anger)、1厌恶(Disgust)、2恐惧(Fear)、3开心(Happy)、4伤心(Sad)、5惊讶(Surprise)、6中性(Normal)。在训练集中出现次数最多的图片是高兴，高达7 215幅，而厌恶表情图片仅仅有436幅，这种不均匀的数据分布以及一些标签噪声和非人脸区域的图片会影响模型的训练效率和准确性，本文对其进行如下处理：(a)对厌恶表情图片进行仿射变换、随机翻转、旋转等数据增强操作，扩充该类图片数量，以平衡数据分布；(b)对部分错误标签图像进行删除；针对(a)和(b)处理后的数据，重新执行(a)中的操作进行数据增强，数据增强前后训练集分布如图6所示。

(2) CK+数据集：CK+数据集是CK数据集的扩展，该数据集共包含了123个对象的327个标记好的面部表情序列。本文从每个序列中随机提取4帧，组成1 308幅带标签的表情图片作为训练集。为了提高模型的泛化能力和鲁棒性，同样对数据集实施了一系列的增强操作，包括随机旋转、翻转、亮度调整和颜色调整。将增强后的图片按各类表情2∶1的比例划分训练集和验证集。数据增强之后的训练集分布如图7所示。

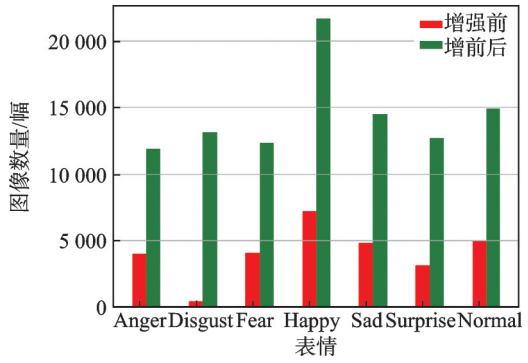

图6 Fer2013数据增强前后的分布

Fig.6 Fer2013 distributions before and after augmentation

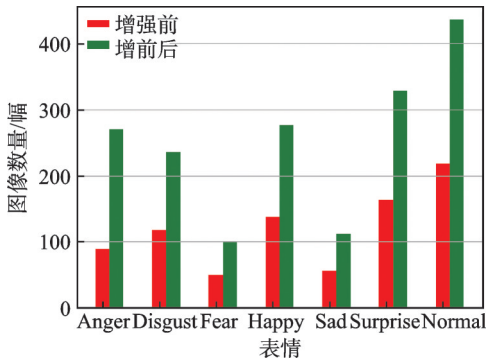

图7 CK+数据增强前后的训练集分布

Fig.7 CK+ distributions before and after data augmentation

2.2 实验环境及参数设置

本实验使用的软件平台如下：编程语言使用Python3.8版本，采用PyTorch2.0搭建深度学习网络框架，操作系统是64 bit的Microsoft Windows 10；硬件环境的配置是：GPU是i7-9700K，内存是128 GB，

显卡的型号是NVIDIA GTX 2080Ti。

本实验超参数设置如下:训练批次设为200轮,32张图片为一个batch,初始学习率为0.001,采用联合损失函数,并在实验过程中使用Adam优化器优化训练过程。当验证损失函数在第30个批次内没有下降时,则按照10倍的速率降低学习率。

2.3　实验结果分析

2.3.1　数据集实验结果

图8是模型在Fer2013数据集上的实验结果,可以看到:随着训练批次的增加,训练集和验证集的准确率都在稳步增加;在训练初期,准确率增长迅猛;当训练到第30轮至第50轮时,由于训练集和验证集的数据分布不同,出现了波动;但100轮以后,模型的识别准确率变得非常平稳。需要说明的是由于在训练集删除了部分错误样本和非人脸区域,模型表现了较高的准确率。

图9是模型在CK+数据集上的准确率,可以看到:模型在训练初期,准确率也同样保持了比较高的增长速度;在第25轮至第75轮之间,虽然出现了波动,但准确率依然在不断提升,由于训练集和验证集数据分布一致,两个数据集的增长趋势保持一致;在第100轮后,模型识别准确率变得非常平稳。在引入了金字塔分割注意力和联合损失函数后,在Fer2013和CK+上的准确率稳步提升,且未出现过拟合和欠拟合现象,证明了本文所提方法的有效性。

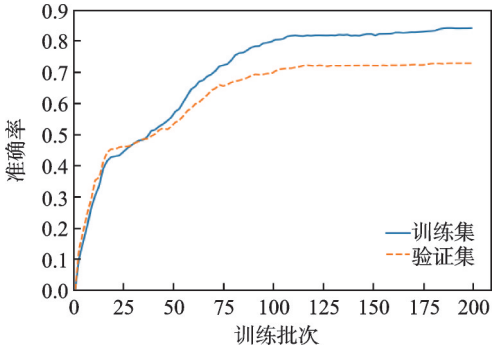

图8　Fer2013数据集上的准确率

Fig.8　Accuracy for Fer2013 dataset

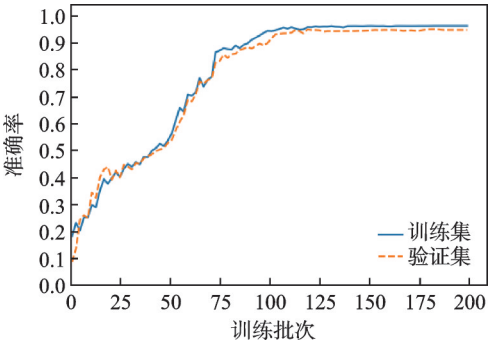

图9　CK+数据集上的准确率

Fig.9　Accuracy for CK+ dataset

为进一步分析各种表情识别的准确率,本文将模型在不同类别表情上得到的识别结果使用混淆矩阵进行可视化分析。由图10可知,相对于其他表情,本文所提出的方法对于高兴和惊讶具有较高的识别率,分别达到了84%和82%。与此同时,对愤怒、悲伤和恐惧的识别的准确率表现欠佳。主要原因如下:高兴和惊讶两种表情具有鲜明的面部特征;悲伤、愤怒和恐惧表情中会带有皱眉和张嘴巴的动作,模型难以区分,从而导致这3类表情的识别度较低。

图11为CK+验证集7种表情识别率的混淆矩阵,相比于Fer2013,每种表情识别准确率都大幅度提升。产生这种现象的原因是CK+数据集是参与者在实验室条件下摆拍指定表情获得,数据采集严谨可靠,大大降低了人为因素和环境的干扰,从而使得模型在CK+数据集的检测结果更加准确。

2.3.2　消融实验

为了评估各个模块的有效性,本文对金字塔分割注意力、联合损失函数进行验证。其中对不添加任何模块的ResNet50网络标记为Base,将各个模块依次加入Base进行对比实验,实验结果如表1所示。

从表1可以看出,在Base上加入联合损失函数后,CK+和Fer2013两个数据集上的准确率分别提升

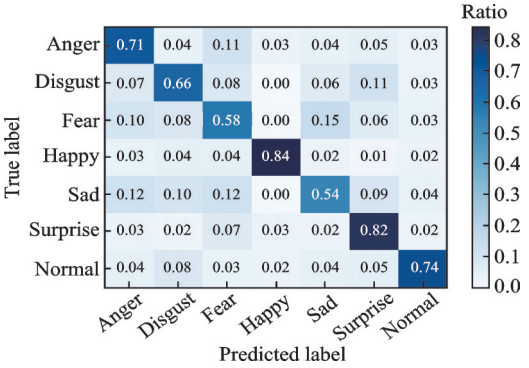

图 10 Fer2013验证集混淆矩阵

Fig.10 Confusion matrix for Fer2013 validation dataset

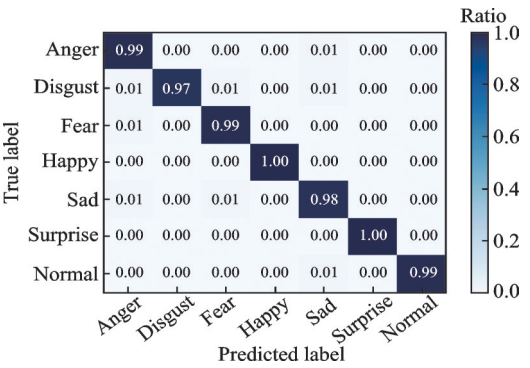

图 11 CK+验证集混淆矩阵

Fig.11 Confusion matrix for CK+ validation dataset

了0.63%和1.45%。分析发现,在CK+数据集中仅有少量同类表情存在差异较大的情况,因此加入Center loss后的提升幅度较小。在Base上加入PSA模块后,准确率提升了2.76%和2.53%,因为该模块有效提取了多尺度特征,增强跨通道信息的相关性,产生更好的像素级注意力,提升了模型对表情的判别能力。结果表明将联合损失函数和PSA模块都融合到Base中能得到最好的识别结果,对比Base,本文方法在CK+数据集上提升了3.55%,在Fer2013数据集上提升了3.34%。

2.3.3 不同注意力模块对比分析

为了验证金字塔分割注意力机制的优越性,把去除PSA模块后的模型标记为ResNet,依次嵌入嵌入通道注意力(Squeeze-and-excitation, SE)[21]、卷积块注意力(Convolutional block attention module, CBAM)[22]、高效通道注意力(Efficient channel attention module, ECA)[23]和SA[24]注意力模块,使用CK+数据集进行对比实验,实验结果如表2所示。从表2可以看出,本文模型PSA-ResNet因在ResNet中加入了PSA模块,所以能够以较少的参数量、较高的计算效率和准确率实现对表情的识别,与ResNet相比,模型运行一次需要的浮点运算量减少0.49 G、参数量减少3.0 MB,准确率提高3.35%。从参数量上看,SE和CBAM注意力计算都使用了全连接层,使参数量分别高达28.09 MB和28.10 MB,因此模型具有较高的计算复杂度;ECA和SA分别使用一维卷积和置换矩阵实现跨通道信息交互,相比之下参数量有所减少,模型计算复杂度也随之降低;PSA使用分组卷积大幅降低了模型的参数量,因此计算复杂度最低。从准确率上看,SE使用自适应挤压降维和激励机制提升重要通道的权重,然而降维不利于学习通道之间的依赖关系,影响了模型准确率的提升;CBAM利用通道和空间注意力捕获不同维度的特征信息,但过度聚焦局部区域,导致模型泛化能力有限;ECA采用一维卷积提取通道之间的依赖关系,实现跨通道交互,但

表1 消融实验

Table 1 Ablation experiments

方法	准确率/%	
	CK+	Fer2013
Base	94.80	70.92
Base+联合损失函数	95.43	72.37
Base+PSA	97.56	73.45
Base+ PSA +联合损失函数	98.35	74.26

表2 不同注意力机制准确率对比

Table 2 Accuracy comparison of different attention mechanisms

注意力模型	参数量/MB	FLOPs/10^9	准确率/%
ResNet	25.56	4.11	94.80
+SE	28.09	4.13	96.80
+CBAM	28.10	4.14	97.03
+ECA	25.56	4.13	97.82
+SA	25.56	4.13	98.07
+PSA	22.56	3.62	98.35

对一些背景复杂的表情图像,无法捕捉图像的中重要信息;SA采用置换矩阵对特征图进行重组,使模型理解不同通道特征之间的关系,但在信息交换和重组时容易导致信息丢失;PSA通过金字塔注意力机制,使用多尺度卷积有效提取不同尺度的特征信息,增强模型的感受野,帮助网络建模远程通道间的语义依赖,从而提升模型对特征的区分和判别能力。PSA与SE、CBAM、ECA和SA相比,准确率分别提升1.55%、1.32%、0.53%和0.28%,证明了本文所提金字塔分割注意力的有效性。

2.3.4　与其他先进模型的对比分析

为验证本文所选择ResNet50基线模型的特征提取能力,本文在CK+和Fer2013数据集上复现经典深度神经网络模型,实验结果如表3所示。

VGG19通过一系列小尺寸的3×3卷积核替换5×5的大尺寸卷积核,有利于细微表情特征的提取,结果达到了人类的识别水平,然而模型参数过多,需要消耗大量的训练时间。ResNet通过两个1×1卷积对通道进行升维和降维,既能保证模型的精度又减少了网络参数,表3显示ResNet34的参数量比VGG19减少了65.38 MB,准确率分别提升1.07%和1.14%。DenseNet121通过密集连接和特征复用大幅度降低了参数量,与类似层级的ResNet101相比,参数量减少了近50%,但特征复用需消耗大量内存,训练周期过长。实际上网络层数与处理速度和精度都有很大关系,浅层神经网络能够利用更多的细粒度表情信息,网络层数越深,越容易提取丰富的语义信息,但容易忽略眼睛,鼻子等小目标关键特征信息,导致识别精度的下降。与ResNet50比,ResNet101参数量增加了18.99 MB,而准确率下降了0.3%和1.07%,证明了ResNet50网络具有较强的特征提取能力。

表3　不同经典神经网络模型对比

Table 3　Comparison of different classical neural network models

模型	参数量/MB	准确率/%	
		CK+	Fer2013
VGG19[18]	77.23	92.18	68.33
ResNet34	21.85	93.25	69.47
ResNet101	44.55	94.50	69.85
DenseNet121[10]	27.60	95.01	70.93
ResNet50	25.56	94.80	70.92

为了进一步验证PSA-ResNet的先进性,将本文所提PSA-ResNet模型与近几年比较先进的面部表情识别模型进行对比,对比结果如表4所示("—"表示相关文献未提供此项数据)。由表4可知,本文所提PSA-ResNet模型在CK+和Fer2013两个数据集上均优于其他模型,但Fer2013数据集中存在标注错误、水印和遮挡等问题,导致其准确率远远低于CK+数据集。ExpressionNet设计多通道融合不同的特征信息,Parallel CNN引入Inception模块以增加图像的宽度和深度,PyConv利用双向金字塔结构同时提升网络对细节信息和全局信息的感知能力,虽然上述3种算法都考虑了图像的多尺度信息,但没有考虑不同特征的重要性,对面部表情无关信息的抑制不够明显。MIANet利用金字塔卷积神经网络提取多尺度特征,使用全局注意力模块得到重要特征信息。与本文模型相比,忽略对不同通道表情特征的建模,以此形成跨通道依赖关系。APRNET引入深度可分离卷积减少模型参数,嵌入SE模块学习不同通道的权重,并通过空间金字塔池化增强模型的鲁棒性,然而没有考虑到同类表情之

表4　不同表情识别方法对比

Table 4　Comparison of different expression recognition methods

模型	参数量/MB	准确率/%	
		CK+	Fer2013
ExpressionNet[25]	—	95.25	70.39
Parallel CNN[26]	—	95.50	70.56
PyConv[27]	24.72	95.31	70.21
MIANet[28]	—	96.37	71.53
APRNET[29]	31.07	97.29	72.00
VTFF[30]	51.80	—	74.08
SimFLE[31]	48.89	—	74.13
TransFER[32]	—	—	74.21
PSA-ResNet	22.56	98.35	74.26

间巨大的差异,与本文模型相比,在CK+和Fer2013数据集上的准确率分别降低了1.06%和2.26%。

Transformer中的自注意力机制可以将特征信息映射到多个空间,增强了模型的感知能力,因此将本文模型与基于Transformer的表情识别模型进行了对比。VTFF将表情图像分割成小块,然后转换成序列输入自注意力模块。SimFLE通过语义掩码和自注意力来重构掩码的面部表情图像,探索信道的丰富语义。TransFER使用多头自注意力机制在不同位置关注不同表情子空间的特征,提升表情识别的效果。在Fer2013数据集上,本文模型与VTFF、SimFLE、TransFER相比,准确度分别提升了0.18%,0.13%和0.05%。

除准确率之外,参数量也是衡量模型性能的重要因素,本文模型与PyConv、APRNET、VTFF、SimFLE相比,参数量分别减少2.16 MB、8.51 MB、28.52 MB和26.33 MB。综上所述,本文所提方法能在保持较少参数量的同时,实现较高的识别率,验证了该模型的先进性。

3 结束语

本文提出了基于金字塔分割注意力和联合损失的表情识别网络,利用SPC实现多尺度特征提取,通过SE增强跨通道之间的信息相关性,提高了表情边缘及远距离预测的精度。为扩大不同种类表情的距离,减少同类表情的距离,在训练中使用Softmax loss和Center loss联合损失函数优化网络模型,进一步提升识别效果。该模型结构简单,训练过程稳定,在训练过程中没有出现欠拟合或过拟合现象,从实验结果可以得出,与前沿算法相比,该模型取得了更好的准确度,然而模型对于某些类别的表情识别的准确度依然不够理想,是下一步需要优化的方向。

参考文献:

[1] PANTIC M, ROTHKRANTZ L J M. Expert system for automatic analysis of facial expressions[J]. Image and Vision Computing, 2000, 18(11): 881-905.

[2] LI S, DENG W. Deep facial expression recognition: A survey[J]. IEEE Transactions on Affective Computing, 2022, 13(3): 1195-1215.

[3] OJALA T, PIETIKAINEN M, MAENPAA T. Multiresolution gray-scale and rotation invariant texture classification with local binary patterns[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2002, 24(7): 971-987.

[4] DALAL N, TRIGGS B. Histograms of oriented gradients for human detection[C]//Proceedings of the 2015 IEEE Conference on Computer Vision and Pattern Recognition. Boston, USA: IEEE, 2015: 3431-3440.

[5] COOTES T F, TAYLOR C J, COOPER D H, et al. Active shape models—Their training and application[J]. Computer Vision and Image Understanding, 1995, 61(1): 38-59.

[6] 程学军, 邢萧飞. 利用改进型VGG标签学习的表情识别方法[J]. 计算机工程与设计, 2022, 43(4): 1134-1144.
CHENG Xuejun, XING Xiaofei. Expression recognition method using improved VGG tag learning[J]. Computer Engineering and Design, 2022, 43(4): 1134-1144.

[7] 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.
ZHAO Xiao, YANG Chen, WANG Ruonan, et al. Facial expression recognition based on attention mechanism ResNet light-weight network[J]. Chinese Journal of Liquid Crystals and Displays, 2023, 38(11): 1503-1510.

[8] 关小蕊, 高璐, 宋文博, 等. 深度残差卷积下多视角特征融合的人脸表情识别[J]. 哈尔滨理工大学学报, 2023, 28(2): 117-127.
GUAN Xiaorui, GAO Lu, SONG Wenbo, et al. Facial expression recognition with multi-perspective feature fusion under deep residual convolution[J]. Journal of Harbin University of Science and Technology, 2023, 28(2): 117-127.

[9] MOLLAHOSSEINI A, CHAN D, MAHOOR M H. Going deeper in facial expression recognition using deep neural networks [C]//Proceedings of the 2016 IEEE Winter Conference on Applications of Computer Vision (WACV). Lake Placid, USA: IEEE, 2016: 1-10.

[10] ARUL VINAYAKAM RAJASIMMAN M, MANOHARAN R K, SUBRAMANI N, et al. Robust facial expression

recognition using an evolutionary algorithm with a deep learning model[J]. Applied Sciences, 2023, 13(1): 468.

[11] LIU K C, HSU C C, WANG W Y, et al. Facial expression recognition using merged convolution neural network[C]// Proceedings of the 2019 IEEE 8th Global Conference on Consumer Electronics (GCCE). Osaka, Japan: IEEE, 2019: 296-298.

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

[13] YAO L, HE S, SU K, et al. Facial expression recognition based on spatial and channel attention mechanisms[J]. Wireless Personal Communications, 2022, 13(56): 1483-1500.

[14] MINAEE S, MINAEI M, ABDOLRASHIDI A. Deep-emotion: Facial expression recognition using attentional convolutional network[J], Sensors, 2021, 21(9): 3046.

[15] LIU Y, DAI W, FANG F, et al. Dynamic multi-channel metric network for joint pose-aware and identity-invariant facial expression recognition[J]. Information Sciences, 2021, 578: 195-213.

[16] ZHANG H, ZU K, LU J, et al. EPSANet: An efficient pyramid squeeze attention block on convolutional neural network[C]// Proceedings of the 2022 Asian Conference on Computer Vision. Macau, China: [s.n.], 2022:1161-1177.

[17] KRIZHEVSKY A, SUTSKEVER I, HINTON G E. ImageNet classification with deep convolutional neural networks[J]. Advances in Neural Information Processing Systems, 2012, 25(3): 142-156.

[18] SIMONYAN K, ZISSERMAN A. Very deep convolutional networks for large-scale image recognition[EB/OL]. (2014-09-04). http://arXiv.1409.1556v6.

[19] CHOLLET F. Xception: Deep learning with depthwise separable convolutions[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Honolulu, USA: IEEE, 2017: 1800-1807.

[20] HE K M, ZHANG X Y, et al. Deep residual learning for image recognition[C]//Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition. New York, USA: IEEE, 2016: 770-778.

[21] HU J, LI S, SUN G, et al. Squeeze-and-excitation networks[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2020, 42(8): 2011-2023.

[22] WOO S, PARK J, LEE J Y, et al. CBAM: Convolutional block attention module[C]//Proceedings of the European Conference on Computer Vision. Berlin, Germany: Springer, 2018: 3-19.

[23] WANG Q, WU B, ZHU P, et al. ECA-Net:Efficient channel attention for deep convolutional neural networks[C]//Proceedings of the 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle, USA: IEEE, 2020: 11531-11539.

[24] ZHANG Q L, YANG Y B. SA-Net: Shuffle attention for deep convolutional neural networks[C]//Proceedings of the 2021 IEEE International Conference on Acoustics. Toronto, Canada: IEEE, 2021: 2235-2239.

[25] ZHOU Y, FENG Y, ZENG S Y, et al. Facial expression recognition based on convolutional neutral network[C]//Proceedings of 2019 IEEE 10th International Conference on Software Engineering and Service Science (ICSESS). Beijing, China: IEEE, 2019: 410-413.

[26] 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, 2020, 24(2): 227-236.
XU Linlin, ZHANG Shumei, ZHAO Junli. Expression recognition algorithm for parallel convolutional neural networks[J]. Journal of Image and Graphics, 2020, 24(2): 227-236.

[27] 李军, 李明. 改进多尺度卷积神经网络的人脸表情识别研究[J]. 重庆邮电大学学报(自然科学版), 2022, 34(2): 201-207.
LI Jun, LI Ming. Research on facial expression recognition based on improved multi-scale convolutional neural networks[J]. Journal of Chongqing University of Posts and Telecommunications (Natural Science Edition), 2022, 34(2): 201-207.

[28] 陈加敏, 徐杨. 注意力金字塔卷积残差网络的表情识别[J]. 计算机工程与应用, 2022, 58(22): 123-131.
CHEN Jiamin, XU Yang. Expression recognition based on convolution residual network of attention pyramid[J]. Computer Engineering and Applications, 2022, 58(22): 123-131.

[29] 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, 2023, 59(1): 199-206.
LUO Sishi, LI Maojun, CHEN Man. Multi-scale integrated attention mechanism for facial expression recognition network[J]. Computer Engineering and Applications, 2023, 59(1): 199-206.

[30] MA F, SUN B, LI S. Facial expression recognition with visual transformers and attentional selective fusion[J]. IEEE

Transactions on Affective Computing, 2023, 14(2): 1236-1248.

[31] MOON J H, PARK S. Simple facial landmark encoding for self-supervised facial expression recognition in the wild[C]// Proceedings of the 2023 IEEE Conference on Computer Vision and Pattern Recognition. New York, USA: IEEE, 2023: 1120-1135.

[32] XUE F, WANG Q, GUO G. Transfer: Learning relation-aware facial expression representations with transformers[C]// Proceedings of the 2023 IEEE Conference on Computer Vision and Pattern Recognition. New York, USA: IEEE, 2023: 3601-3610.

作者简介:

谷瑞(1982-),男,硕士,副教授,研究方向:计算机视觉、数据挖掘等,E-mail:gur@siso.edu.cn。

顾家乐(1981-),**通信作者**,男,硕士,副教授,研究方向:机器学习,E-mail:gujl@siso.edu.cn。

宋翠玲(1980-),**通信作者**,女,博士,教授,研究方向:金融数据挖掘,E-mail:songcl@siso.edu.cn。

(编辑:张蓓,王婕)