

基于改进的残差网络面部表情识别^{*}

兰江海^{1,2} 林国军^{1,2} 游松^{1,2} 周顺勇^{1,2} 黄丹^{1,2}

(1. 四川轻化工大学人工智能四川省重点实验室 自贡 643000;

2. 四川轻化工大学自动化与信息工程学院 自贡 643000)

摘要: 为了解决人脸表情识别中存在的对关键信息获取能力不足、识别率偏低、模型容易出现过拟合等问题,以 ResNet18 作为基本网络进行改进得到一个新的表情识别模型(IERNet)。IERNet 通过引入 ECA 注意力机制构建出两种不同的注意力残差单元,并组成注意力残差模块,从而增强对深层的表情关键特征的提取能力;又引入 Iception 模块来提取图像的多尺度浅层信息,通过同时引入这两个模块的方式增强了网络的鲁棒性、提升了模型的识别率;最后使用全局平均池化结合 Dropout 技术取代全连接层,可以有效防止模型的过拟合问题还能简化模型。通过实验数据可知,在公开表情数据集 CK+ 和 FER2013 上取得了不错的成绩,准确率分别达到了 97.778% 和 73.558%。

关键词: 表情识别;残差网络;注意力机制;深度学习

中图分类号: TN2 **文献标识码:** A **国家标准学科分类代码:** 520.604

Facial expression recognition based on improved residual network

Lan Jianghai^{1,2} Ling Guojun^{1,2} You Song^{1,2} Zhou Shunrong^{1,2} Huang Dan^{1,2}

(1. Sichuan Key Laboratory of Artificial Intelligence, Sichuan University of Science & Engineering, Zigong 643000, China;

2. School of Automation and Information Engineering, Sichuan University of Science & Engineering, Zigong 643000, China)

Abstract: In order to solve the problems existing in facial expression recognition, such as insufficient ability to obtain key information, low recognition rate and easy to overfit the model, a new facial expression recognition model (IERNet) was developed by using ResNet18 as the basic network. By introducing ECA attention mechanism, IERNet constructs two different kinds of attention residual units and forms an attention residual module, so as to enhance the ability to extract the key features of deep expressions. Iception module is introduced to extract the multi-scale shallow information of the image. By introducing the two modules at the same time, the robustness of the network is enhanced and the recognition rate of the model is improved. Finally, we use the global average pool combined with Dropout technology to replace the full-connection layer, which can effectively prevent the overfitting problem of the model and simplify the model. According to the experimental data, this paper has achieved good results in the public expression data sets CK+ and FER2013, with the accuracy rate reaching 97.778% and 73.558% respectively.

Keywords: expression recognition; residual network; attention mechanism; deep learning

0 引言

人脸的面部表情是人际交往中除语言外一种非常重要的表达方式。快乐、愤怒、惊讶、恐惧、厌恶、悲伤以及蔑视,这 7 种情绪是人们研究表情识别的基础。面部表情识别作为计算机视觉领域的一个研究方向,与人脸检测^[1]和

识别密切相关,并已逐渐应用于日常生活中,如驾驶员疲劳驾驶检测、谎言测试、以及一些娱乐特效等。

而在人脸表情识别的研究中最重要的就是特征提取^[2],目前面部表情特征提取主要有两种方式,基于传统方法的人工特征提取和基于深度学习的研究。结合线性判别分析(LDA)和主成分分析(PCA)对表情进行分类^[3]。

收稿日期:2023-01-13

^{*} 基金项目:四川省科技计划项目(2020YFSY0027,2022YFSY0056)资助

Feng 等^[4]将局部二值模式 (LBP) 特征与支持向量机 (SVM) 分类相结合实现对面部表情的识别,并在实际应用中分辨率较低的情况下对模型进行了相应的改进。Zhi 等^[5]提出了一种新的用于人脸表情识别的图保持稀疏非负矩阵分解算法,该算法对部分遮挡具有较强的鲁棒性。不过人工提取特征的方法对于深层特征提取能力不强,容易使特征流失,难以应用于各种复杂场景。

深度学习在表情识别领域取得良好效果,常用的面部表情识别卷积神经网络 (CNN) 有 AlexNet (Alex Krizhevsky net)^[6]、VGGNet (visual geometry group net)^[7]、GoogleNet (Google inception net)^[8] 等。GoogleNet 模型中的 Inception 模块能获取图像丰富的语义特征,不过并不是网络层数越多效果就越好,可能会出现梯度趋于 0 或者梯度疯狂增大问题。He 等^[9]提出的残差网络 (Residual Network, ResNet) 模型,通过添加 Shortcut 连接实现恒等映射从而处理梯度爆炸的问题。

除此之外,单一的卷积神经网络^[10]不能针对性的提取部分关键信息,而注意力机制能够选择性的关注部分信息。冯婧等^[11]通过协调注意力机制可以同时获取人脸的位置和通道信息。Sun 等^[12]利用感兴趣区域 (region of interest, ROI) 使表情预测的可靠性有所提高。Li 等^[13]提出了一种端到端的注意力机制人脸表情识别网络从而获得更多的有用信息。Hu 等^[14]提出 SE-Net (squeeze and excitation networks),通过增加空间编码质量来实现对局部信息的捕获。而 Woo 等^[15]基于通道和空间分别加入注意力图从而得到 CBAM (convolutional block attention module)。但是这些网络模型都有一定的不足,有的识别率偏低,有的网络模型太复杂,运行速度慢。Wang 等^[16]提出了一种高效通道注意力网络模型 (efficient channel attention networks, EAC-Net),EAC-Net 可以凸出局部有用信息,抑制不相关信息,是一种有效且轻量的模型,而且此模型使用方便。

为了解决传统的神经网络对图像关键特征能力提取不足、识别率偏低、网络模型复杂容易过拟合等问题,本文基于 ResNet18 构建了一种新的表情识别网络模型 (IERNet)。通过对残差块进行改进,并引入 EAC-Net 注意力机制,增强了网络对面表情关键特征的提取能力和鲁棒性,同时有效解决了训练时取梯度消失的问题。在对残差块改进后的基础上又引入 Inception 模块扩大网络的感受野,使模型既能够充分提取关键特征同时也能获得更强的特征学习能力,从而进一步提高模型的识别率。使用全局平均池化和 Dropout 技术相结合,减少了计算量、增强了网络的拟合能力,提升了模型的性能。

1 IERNet 网络模型

1.1 Inception 模块

Inception 模块如图 1 所示,有多个卷积核大小不同的卷积,使用这些卷积可以获取到图像更丰富的不同尺度特

征,最后进行融合得到来自不同感受野的丰富特征图。其网络结构如图 1 所示,用两个 3×3 的卷积代替最初的 5×5 的卷积不仅能扩大感受野,同时还能减少参数量;在每条支路都加上一个 1×1 的卷积来增加网络的特征提取能力。

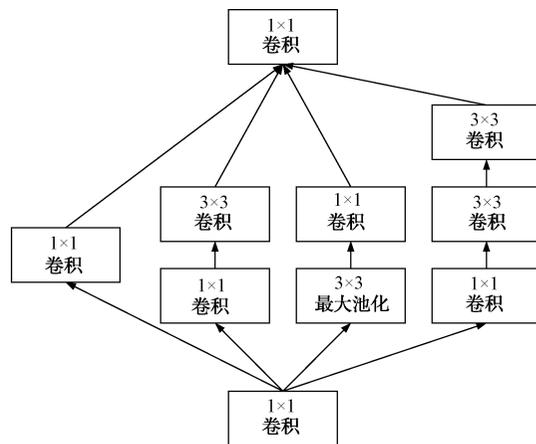


图 1 Inception 模块结构

1.2 ECA-Net

ECA-Net^[16]是在 SE-Net^[14]的基础上优化得来的一种通道注意力模型。SE-Net 中的降维存在缺陷,对获取通道之间的联系效率低。ECA-Net 注意力机制模块选择使用的卷积去学习通道信息,通过这样的方式来解决通道的降维问题。ECA-Net 只涉及少数参数就能达到很好的效果,ECA-Net 是一种超轻量级模型。为了进行有效的跨通道交流,让每个通道和附近的通道之间交换信息、生成权重来解决 SE-Net 中的降维问题。ECA-Net 可以更关注人脸面部表情相关部位,降低其他权重,优化网络,从而提高模型对重点部位特征的利用率,取得更高的识别准确率。ECA-Net 结构如图 2 所示。

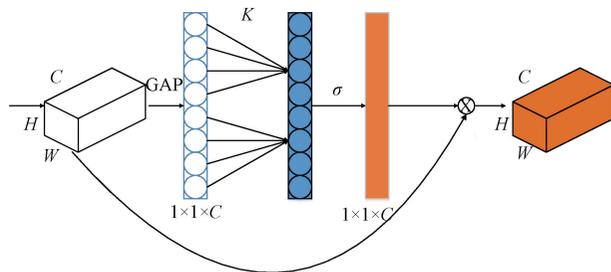


图 2 EAC-Net 结构

ECA-Net 的实现可以作两步处理,一方面首先将输入特征通过全局平均池化操作得到 $1 \times 1 \times C$ 的全局描述特征,然后通过卷积核尺寸为 K 的 1×1 卷积来进行部分通道特征学习,再通过 sigmoid 激活函数作用计算所有通道的权重 ω ;最后将通过这条路径得到的通道权重与另一步则不做处理的输入特征进行对应元素相乘从而实现对

通道权重的矫正。ECA-Net 没有对所有的通道进行信息交流,而是选择相近的通道进行相互作用,可以减少模型复杂度提高效率。其中每个通道的权重 ω 如下:

$$\omega = \sigma(C_1 D_K(y)) \quad (1)$$

式中: σ 为 *sigmoid* 激活函数; $C_1 D$ 代表一维卷积; y 表示通道; K 为卷积核大小,而 K 的大小和通道自适应函数有关,如式(2)所示。

$$C = 2^{(\gamma \cdot K - b)} \quad (2)$$

故可得:

$$K = \left\lfloor \frac{\log_2 C}{\gamma} + \frac{b}{\gamma} \right\rfloor_{\text{odd}} \quad (3)$$

式中: $\lfloor X \rfloor_{\text{odd}}$ 表示 K 的值只能为奇数,设 $\gamma=2, b=1$ 。

1.3 ResNet 残差块

对于深度学习,随着网络规模的扩大、网络层数的增加,网络模型训练出来的效果反而会下降。而 ResNet^[9] 可以有效解决这种问题。该网络主要通过一条恒等路径来实现跳跃连接从而处理的梯度更新的问题。残差基础结构如图 3 所示。

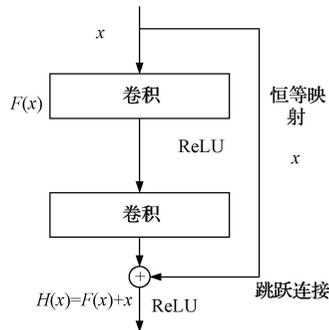


图 3 残差基础结构

图 3 中卷积操作公式如下:

$$C(x, y) = \sum_{k,l} F(k, l) G(x - k, y - l) \quad (4)$$

式中: $C(x, y) = F_{k,l} \otimes G_{u,n}$ 为卷积操作, $F_{k,l}$ 是卷积尺寸为 $k \times 1$ 的卷积核, $G_{u,n}$ 是 u, n 的输入矩阵, C 是卷积计算后得出的结果, $\forall_x \in [0, u], \forall_y \in [0, n]$. *ReLU* 是激活函数,其优点就是收敛快。*ReLU* 函数公式如下:

$$F(x) = \text{Max}(0, x) \quad (5)$$

式中: $F(x)$ 是 *ReLU* 激活结果。 $x < 0$ 时, $F(x) = 0$; $x \in [0, x]$ 时, $F(x) = x$ 。

1.4 改进的注意力残差模块

本文所提出的注意力残差模块,由两个不同的残差结构单元组成如图 4(a)和(b)所示。该残差结构单元分别是在两个不同的基础残差结构上得到的。其中注意力残差单元(图 4(a))是直接传统残差结构上加入了 ECA-Net 注意力模块,输入数据分为两条路径。一条路径依次经过一个 3×3 的卷积层, BN 层, *ReLU* 层, 3×3 的卷积层,再通过 ECA-Net 注意力模块,通过注意力机制获取局部关键信息,提取到更丰富的特征从而增加识别准确率;另一

条路径为恒等映射支路。而注意力残差学习单元(图 4(b))有所不同,在原恒等支路中加了一个 1×1 的卷积和一个 BN 层。这个 1×1 的卷积可以改变输入的维度,加深网络结构,以及特征提取能力。而注意力残差模块由这两个不同的注意力残差单元组成,这样既能继承原残差网络的优点又能提高特征提取的能力,从而实现性能上的提上。注意力残差模块由注意力残差结构单元(图 4(a)和(b))直接相连,输入图片依次经过这两个不同的注意力残差结构单元然后输出。

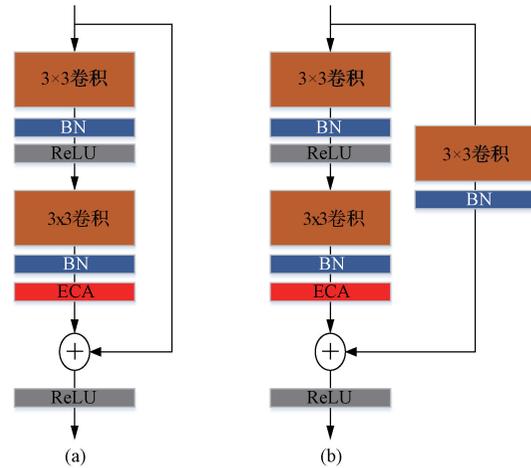


图 4 注意力残差学习单元结构

1.5 全局平均池化+Dropout 正则化

传统神经网络具有多个全连接层,多个全连接层会存在参数量太大、容易过拟合的问题,而且还会降低训练速度。本文在 ResNet18 的基础上,去掉了原有的全连接层,采用全局平均池化,这样可以有效减少网络参数和计算量;同时采用 Dropout 正则化方法来来解决模型的过拟合的问题。通过全局平均池化与 Dropout 技术来更换全连接层,可以有简化网络,有效减少模型出现过拟合现象。Dropout 技术简单来说就是随机去掉网络中的部分神经元。

1.6 IERNet 网络结构

基于 Resnet18 网络,本文提出了一种改进的 IERNet 模型,该网络模型由 Inception 模块、改进的注意力残差模块、最大池化层、全局平均池化层、Dropout 正则化以及 Softmax 分类层构成。首先对图片进行预处理,然后输入网络模型处理,提取表情特征,最后进行分类。设输入图片尺寸大小为 48×48 ,先经过两个 Inception 模块提取图片的多尺度浅层信息,然后经过 3 个注意力残差模块进行深层关键特征的提取;并在 Inception 模块后面以及每个注意力残差模块后加入一个最大池化层来过渡冗余信息,减少卷积层参数误差带来的预测影响;在经过最后一个残差注意力模块后得到 $3 \times 3 \times 512$ 输出特征图。然后通过全局平均池化加 Dropout 处理,最后直接通过 Softmax 层

进行分类。IERNet 网络结构如图 5 所示,网络模型参数如表 1 所示。

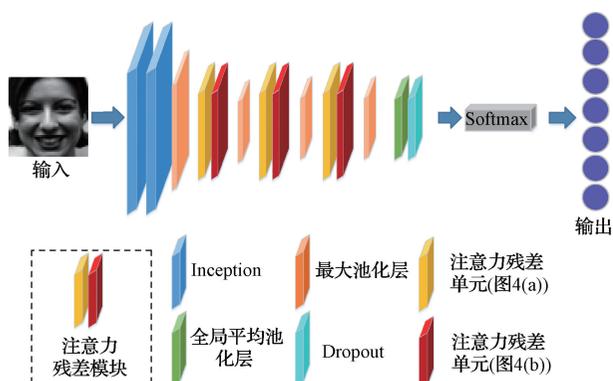


图 5 IERNet 网络结构

表 1 模型参数

输入类型	卷积尺寸	输入尺寸	输出尺寸	步长
Inception	1×1,3×3	48×48×3	48×48×32	
Inception	1×1,3×3	48×48×32	48×48×64	
最大池化层	2×2	48×48×64	24×24×64	2
注意力残差模块	1×1,3×3	24×24×64	24×24×128	
最大池化层	2×2	24×24×128	12×12×128	2
注意力残差模块	1×1,3×3	12×12×128	12×12×256	
最大池化层	2×2	12×12×256	6×6×256	2
注意力残差模块	1×1,3×3	6×6×256	6×6×512	
最大池化层	2×2	6×6×512	3×3×512	2

其中卷积尺寸就是每个模块中的卷积核大小;输入输出尺寸为图像尺寸×通道数,通道数的设置则是参考原 ResNet18 中的数据多次试验后得出的,由于数据集中图片的尺寸为 48×48,故初始输入尺寸为 48×48×3,然后每通过一个步长为 2 的最大池化层图片尺寸则变为之前的 1/2。

2 实验及结果

2.1 实验环境

实验环境为 Windows10,64 位操作系统, GPU 为 NVIDIA GeForce RTX 2060, CPU 为 11th Gen Intel(R) Core(TM) i7-11700 @2.50 GHz,使用 python 编写代码,通过 Pytorch 搭建深度学习环境。

2.2 数据集

本次实验选择了在人脸表情识别中使用较多的两个公开数据集,CK+数据集^[17]和 FER2013^[18]数据集。CK+数据集为标准环境下采集的,该数据集总共采集了 981 张不同的面部情绪照片,训练集和测试集分别有 882 张、99 张照片。该数据集人脸表情类别有高兴、吃惊、恐惧、难过、生气、讨厌以及藐视这 7 种表情,每张照片的尺

寸大小都是统一的 48×48。FER2013 人脸表情数据集是比较具有挑战性的表情数据集,主要来源于各种网络图片,里面存在一些卡通、文本框等非人脸的错误标签图片,图片中还有许多水印等噪声。该数据集的总共采集了 35 886 张不同的照片,且 FER2013 人脸表情数据集由训练集、公开验证集以及私有测试集 3 个部分组成。这 3 个部分对应的人脸表情照片数量分别为 28 708 张、3 589 张、3 589 张;该数据集人脸表情类别有高兴、吃惊、恐惧、难过、生气、讨厌以及中性这 7 种表情;同样的该数据集中的照片大小也全都是统一的 48×48 的灰度图像。两个数据集的表情图像如图 6 所示。



图 6 CK+与 FER-2013 数据集的表情示例

进行实验时,可以对图片做一些数据增强,在训练之前,对原始输入照片进行随机旋转、放缩、镜像等几何处理,增大数据集。在测试阶段,先对初始图片进行中心裁剪操作,然后在对图片进行镜像操作,这样就可以将数据增大为原始数据集的 10 倍。在将这些照片加载进入模型,取平均概率,最大的输出分类与其表情分类相对应,该操作的好处就是能够较好的减少错误分类的情况。

2.3 实验结果与分析

对于 CK+数据集,由于该数据集本身所采集的人脸表情照片不多,因此在进行测试时选择十折交叉验证的方法。在这里把 CK+数据集中的图片划分为 10 份,训练集占 9 份,照片数量为 882,而测试集占据 1 份,照片数量为 99。然后为了减少偶然性,还对该验证方法进行了多次的重复实验,同时通过生成的随机子样本进行反复的训练和测试,最后取平均值的平均概率。实验共设置训练 60 个循环, batch size 设为 32,学习率设置为 0.01,当模型跑完 20 个循环后,每间隔 2 轮就把学习率调整为之前的 0.9 倍。而对于 FER-2013 数据集,实验的参数设置有所不同,对网络进行 200 个循环训练,初始学习率设为 0.01, batch size 设为 32,当网络跑 80 个循环后,每间隔 5 轮就把学习率调整为之前的 0.9 倍,最后在测试集上展示模型性能。

IERNet 网络在 CK+数据集和 FER-2013 数据集上的准确率和损失曲线如图 7、8 所示。通过这两种曲线图可以知道,IERNet 模型具有较强的拟合能力和鲁棒性,模型在测试集上的准确率同迭代次数一直在增加,损失也不断减小,最后都收敛。

在 FER-2013 数据集上,训练时的效果与进行测试出

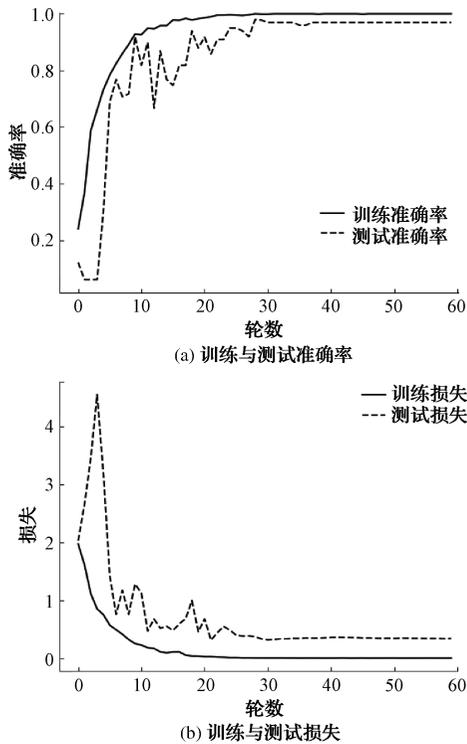


图7 CK+数据集准确率和损失曲线

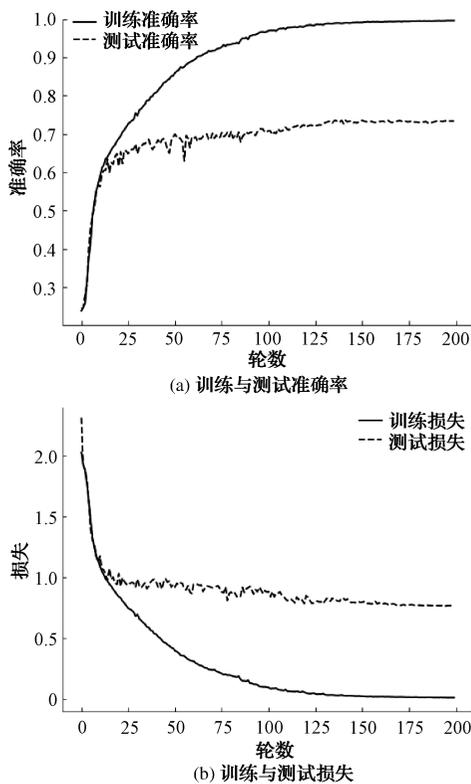


图8 FER-2013数据集准确率和损失曲线

来的结果差距较大,网络在训练时学习到了训练集的表情特征,当模型达到优化后,由于测试集上由许多的非人脸

的错误标签样本,从而使模型在测试集上与训练集上识别出的结果存在较大差异。

IERNet网络在CK+数据集和FER-2013数据集上的混淆矩阵分别如图9、10所示,总体准确率分别为97.778%和73.558%。

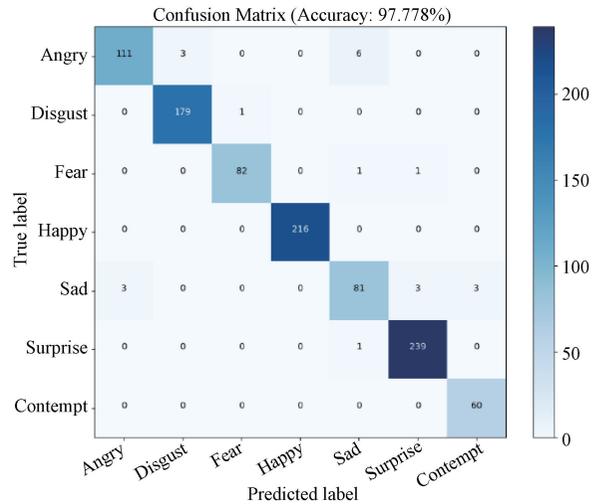


图9 CK+数据集混淆矩阵

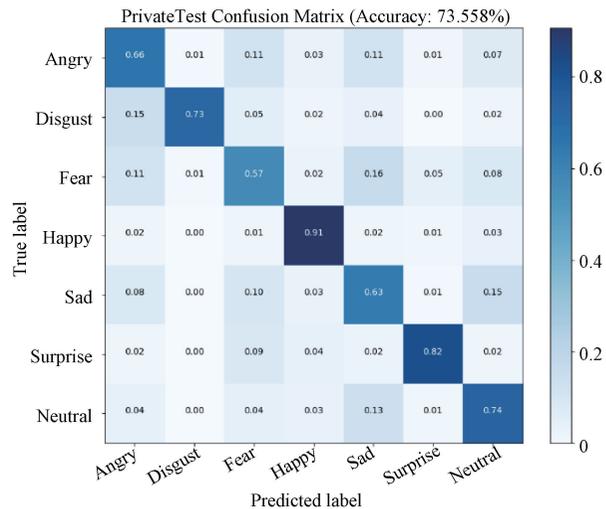


图10 FER-2013数据集混淆矩阵

因为CK+数据集中的图片都是在标准实验室环境下得到的,图片清晰且不存在标记错误的情况,因此整体准确率较高。通过混淆矩阵可以发现惊喜、高兴、厌恶这3类表情的识别率是比较高的,这3类表情都是属于面部动作作较大容易区分的,而且在CK+数据集中的样本数量也比较多,所以识别率较高。然而生气、害怕、悲伤、藐视这几类消极表情的识别准确率就相对较低,其中藐视最低。一方面是因为这几类表情都是属于消极类情绪,这些表情的面部动作本来就有许多相似的特征,容易产生混淆;另一方面的原因就是CK+这个数据集本身对这几类消极表

情的采集就不够多,训练样本较少,类别分布不够均衡,网络对特征的训练不够,所以对这几类消极表情的预测效果不太理想。

而 FER-2013 数据集来源广泛各种图片都有,存在一些非人脸表情的错误标签,有许多被遮挡的、非正面的人脸。除此之外,图片中还存在着许多的水印,所以该数据集的整体识别率不高。通过 FER-2013 数据集的混淆矩阵可以知道,高兴类表情的测试效果最好,达到了 90% 以上,由于高兴类表情本身的面部动作幅度较大容易区分,而且该类数据样本数量多,所以测试效果比较理想;而对于生气、害怕和难过 3 类消极表情的测试效果就不太理想,同样存在与 CK+数据集相似的问题。一方面是这 3 类消极情绪本身具有较强的相似性,面部关键特征区分度不够,互相容易出现错分。惊讶与害怕容易混淆,这两种表情都有睁大眼睛、张大嘴巴等特征;而害怕类表情与悲伤类表情也不容易区分,这两类表情的面部都会出现皱眉这样相似的情况,且这些消极表情中还有侧脸、水印等噪声,使得区分难度更大。另一方面就是数据集本身这几类消极表情图片的采集较少,与其他表情相比在数据集中所占比例太低,导致在该数据集中的类别分布出现不均衡的问题,网络模型得不到足够的数据进行训练导致识别率下降。

证明本文所构建的网络是有效果的,与一些现有的 FER 特征提取方法进行比较,分别在 CK+和 FER-2013 数据集测试。测试结果如表 2、3 所示。

表 2 不同模型在 CK+数据集上的识别率

方法	识别率/%
DCMA-CNNs ^[19]	93.46
Parallel CNN ^[20]	94.03
文献[21]	94.82
IACNN ^[22]	95.38
VGG-19 ^[23]	96.47
ResNet18	94.04
IERNet	97.77

表 3 不同模型在 FER-2013 数据集上的识别率

方法	识别率/%
SHCNN ^[24]	69.10
文献[25]	70.30
文献[26]	70.91
文献[27]	73.00
IcRL ^[28]	72.36
ResNet18	72.97
IERNet	73.55

其中,Parallel CNN 提出使用两个并行卷积池化单元的卷积神经网络,分别提取不同的图像特征并进行融合,

但效果一般识别率不高;文献[21]的模型将 Haar 与自商图像滤波器结合,这样虽然提高了特征提取能力同时也会造成网络太复杂导致模型过拟合等问题;VGG-19 模型的效果虽然可以,但是存在多个全连接层,网络参数太多,运行时间长。而本文在 ResNet18 模型的基础上提出的 IERNet 网络模型,通过 EAC-Net 模块可以充分提取关键特征,且该网络还属于轻量型注意力机制模型,又采用了全局平均池化加 Dropout 技术来简化网络,减少模型的过拟合情况。在 CK+数据集上取得了 97.77% 的识别率,比原网络高 3.74%,通过这一系列对比实验证明了 IERNet 网络模型的有效性。

通过表 3 可以发现,本文的 IERNet 网络模型在 FER2013 数据集上的识别率比 SHCNN 模型、文献[25]的模型、文献[26]的模型提高了 2.64%~4.45%。与文献[27]的模型相比,IERNet 模型不需要人工提取面部表情图像的 LBP 特征,网络模型参数更少,复杂度更低。与 IcRL 模型相比,IERNet 模型可以通过 Inception 模块来提取图像更丰富的多尺度语义特征,同时通过 EAC-Net 注意力模块提高了对部分关键特征的提取能力。本文提出的 IERNet 网络模型在保证对关键特征的提取的同时又能减少参数量,且具有较强的鲁棒性和拟合能力。综上所述,本文提出的 IERNet 网络模型在 CK+数据集和 FER-2013 数据集上都体现出了有效性。

2.4 模块有效性验证实验

为了证明本文网络结构中使用模块的作用,依次进行消融实验:1)在未引入注意力残差模块和 Inception 模块的网络中实验,这种方法记为 Base;2)用 Base+ARM 代表在模型中只加入注意力残差模块;3)用 Base+ Inception 表示在模型中只加入 Inception 模块;4)用 IERNet 表示在模型中同时加入注意力残差模块和 Inception 模块。消融实验结果如表 4 所示。

表 4 消融实验结果 (%)

方法	CK+	FER-2013
Base	94.04	72.97
Base+ARM	96.73	73.24
Base+ Inception	95.82	73.18
IERNet	97.77	73.55

通过消融实验验证了 IERNet 模型中注意力残差模块和 Inception 模块的有效性,对两个模块分别进行了对比试验。从表 4 可以看出,引入注意力残差模块和 Inception 模块后,模型的识别率分别都有所提高,注意力残差模块通过增加对关键信息的获取来提高识别率,Inception 模块通过提取图像的多尺度特征来提高识别效果,这里注意力残差模块所带来的效果比 Inception 模块略高一点,而本文的 IERNet 模型同时引入这两个模块,从而进一步提高了识别率,都证明了模块的有效性。

3 结论

本文提出了一种自定义的神经网络模型,首先通过两个 Inception 模块提取图像浅层的丰富语义信息;再通过 3 个改进的注意力残差模块来充分提取深层的面部表情的关键信息,从而解决了对关键特征提取能力不足、识别率较低的问题;最后通过采用全局平均池化与 Dropout 技术相结合,降低了网络复杂度、有效缓解了过拟合问题,增强了模型性能。本文设计的 IERNet 模型在两个公开数据集上验证了有效性,与部分已有的先进方法对比也具有更高的识别率。本文方法对部分消极类表情的识别效果还不够理想,后续研究工作将集中在如何减少对消极表情分类的误判。

参考文献

- [1] 侯学良,单腾飞,薛靖国. 深度学习的目标检测典型算法及其应用现状分析[J]. 国外电子测量技术, 2022, 41(6): 165-174.
- [2] 鲍光海,林善银,徐林森. 基于改进型卷积网络的汽车高度调节器缺陷检测方法[J]. 仪器仪表学报, 2020, 41(2): 157-165.
- [3] LI S, DENG W H. Deep facial expression recognition: A survey[J]. IEEE Transactions on Affective Computing, 2020, 13(3): 1195-1215.
- [4] FENG X Y, PIETIKAINEN M, HADID A. Facial expression recognition based on local binary patterns[J]. Pattern Recognition and Image Analysis, 2007, 17(4): 592-598.
- [5] ZHI R C, FLIERL M, RUAN Q Q, et al. Graph-preserving Sparse nonnegative matrix factorization with application to facial expression recognition[J]. IEEE Transactions on Systems, Man, and Cybernetics Part B, 2011, 41(1): 38-52.
- [6] KRIZHEVSKY A, SUTSKEVER I, HINTON G E, et al. ImageNet classification with deep convolutional neural networks[J]. Communications of the ACM, 2017, 60(6): 84-90.
- [7] ZENG N Y, ZHANG H, SONG B Y, et al. Facial expression recognition via learning deep sparse autoencoders [J]. Neurocomputing, 2018, 273: 643-649.
- [8] SZEGEDY C, WEI L, JIA Y, et al. Going deeper with convolutions [C]. IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2015: 1-9.
- [9] HE K M, ZHANG X Y, REN S Q, et al. Deep residual learning for image recognition[C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016.
- [10] 严佩敏,唐婉琪. 基于可分离卷积神经网络的文本分类[J]. 电子测量技术, 2020, 43(13): 7-12.
- [11] 冯婧,顾梅花,刘晓龙,等. 课堂场景下改进 MobileNetv3 人脸识别算法[J]. 国外电子测量技术, 2022, 41(10): 47-55.
- [12] SUN X, XIA P P, ZHANG L M, et al. A ROI-guided deep architecture for robust facial expressions recognition[J]. Information Sciences, 2020, 522: 35-48.
- [13] LI J, JIN K, ZHOU D L, et al. Attention mechanism-based CNN for facial expression recognition[J]. Neurocomputing, 2020, 411: 340-350.
- [14] HU J, SHEN L, ALBANIE S, et al. Squeeze-and-excitation networks[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2019, 42(8): 2011-2023.
- [15] WOO S, PARK J, LEE J Y, et al. CBAM: Convolutional block attention module[C]. Proceedings of the European Conference on Computer Vision. Berlin: Springer, 2018: 3-19.
- [16] WANG Q L, WU B G, ZHU P F. EAC-Net: Efficient channel attention for deep convolutional neural networks[C]. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020: 11531-11539.
- [17] LUCEY P, COHN J F, KANADE T, et al. The extended CohnKanade dataset (CK+): A complete dataset for action unit and emotion-specified expression[C]. Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2010: 94-101.
- [18] GOODFELLOW I J, ERHAN D, CARRIER P L, et al. Challenges in representation learning: A report on three machine learning contests[J]. Neural Network, 2015, 64: 59-63.
- [19] XIE S Y, HU H F. Facial expression recognition using hierarchical features with deep comprehensive multipatches aggregation convolutional neural networks[J]. IEEE Transactions on Multimedia, 2019, 21(1): 211-220.
- [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): 73-82.
- [21] SUN X, LV M. Facial expression recognition based on a hybrid model combining deep and shallow features[J]. Cognitive Computation, 2019, 11(4): 587-597.
- [22] MENG Z, LIU P, CAI J, et al. Identity-aware convolutional neural network for facial expression recognition[C]. IEEE International Conference on Automatic

- Face and Gesture Recognition, 2017: 558-565.
- [23] SIMONYAN K, ZISSERMAN A. Very deep convolutional networks for large-scale image recognition [J]. Computer Science, 2014, DOI: 10.48550/arXiv.1409.1556.
- [24] MIAO S, XU H Y, HAN Z Q, et al. Recognizing facial expressions using a shallow convolutional neural network[J]. IEEE Access, 2019(7): 78000-78.
- [25] 梁华刚, 雷毅雄. 增强可分离卷积通道特征的表情识别研究 [J]. 计算机工程与应用, 2022, 58(2): 184-192.
- [26] ZHOU J C, JIA X, SHEN L L, et al. Improved Softmax loss for deep learning-based face and expression recognition[J]. Cognitive Computation and Systems, 2019, 1(4): 97-102.
- [27] QU J S, ZHANG R J, ZHANG Z W. Facial expression recognition based on deep residual network[J]. Journal of Computers, 2020, 31(2): 12-19.
- [28] CHEN Y, HU H. Facial expression recognition by inter-class relational learning [J]. IEEE Access, 2019(7): 94106-94117.

作者简介

兰江海, 硕士研究生, 主要研究方向为图像处理、计算机视觉等。

E-mail: 1617114793@qq.com

林国军, 讲师, 主要研究方向为人脸识别、图像处理和计算机视觉等。

E-mail: 3836988463@qq.com