

# 基于轻量级注意力残差网络的 面部表情识别方法

邵高飞, 邵党国, 马磊, 易三莉

(昆明理工大学 信息工程与自动化学院, 云南省计算机技术应用重点实验室, 昆明 650504)

**摘要:** 针对卷积神经网络参数量大、训练时间长的问题, 提出一种基于轻量级注意力残差网络的面部表情识别方法. 首先, 以残差网络为骨架重新搭建模型, 通过减少层数并改进残差模块提高模型性能; 其次, 引入深度可分离卷积减少模型的参数量和计算工作量; 最后, 采用 Mish 函数替代 ReLU 函数的挤压激励模块自适应地调整通道权重. 该模型在两个公共数据集 CK+ 和 JAFFE 上采用经典的十折交叉验证方式进行验证, 分别获得了 98.16% 和 96.67% 的准确率. 实验结果表明, 该方法在模型识别精度和复杂度之间进行了较好权衡.

**关键词:** 面部表情识别; 轻量级; 残差网络; 深度可分离卷积; 注意力机制

**中图分类号:** TP391.41 **文献标志码:** A **文章编号:** 1671-5489(2025)02-0437-08

## Facial Expression Recognition Method Based on Lightweight Attention Residual Network

GAO Gaofei, SHAO Dangguo, MA Lei, YI Sanli

(Yunnan Key Laboratory of Computer Technologies Application, Faculty of Information Engineering and Automation, Kunming University of Science and Technology, Kunming 650504, China)

**Abstract:** Aiming at the problems of a large number of parameters and the long training time of convolutional neural networks, we proposed a facial expression recognition method based on a lightweight attention residual network. Firstly, we rebuilt the model by using the residual network as a skeleton, and improved the model performance by reducing the number of layers and improving the residual module. Secondly, the depthwise separable convolution was introduced to reduce the number of model parameters and computational effort. Finally, the squeeze and excitation module of ReLU function was replaced by Mish function to adaptively adjust the channel weight. The model was validated by using the classical ten-fold cross-validation mode on two public datasets CK+ and JAFFE, and obtained accuracies of 98.16% and 96.67%, respectively. The experimental results show that the proposed method provides a better trade-off between model identification accuracy and complexity.

**Keywords:** facial expression recognition; lightweight; residual network; depthwise separable convolution; attention mechanism

收稿日期: 2023-11-03.

**第一作者简介:** 邵高飞(1997—), 男, 汉族, 硕士研究生, 从事机器视觉的研究, E-mail: 1520871443@qq.com. **通信作者简介:** 邵党国(1979—), 男, 汉族, 博士, 副教授, 从事人工智能的研究, E-mail: orangebear152@gmail.com.

**基金项目:** 国家自然科学基金(批准号: 62266025)和云南省计算机技术应用重点实验室开放基金(批准号: CB22144S078A).

人类的面部表情包含了丰富的情感信息和心理过程,是人类除语言外最有效的情感交流方式.面部表情识别(facial expression recognition, FER)在智能安全<sup>[1]</sup>、医疗监测<sup>[2]</sup>、驾驶疲劳监测<sup>[3]</sup>、机器人<sup>[4]</sup>等多领域应用广泛.近年来,随着深度学习(deep learning, DL)的快速发展,FER方法研究也得到快速发展,逐渐形成了完善的FER系统<sup>[5]</sup>.

FER的本质是情绪特征的提取并将特征划分为离散的类别<sup>[6]</sup>.卷积神经网络(convolutional neural network, CNN)作为深度学习最重要的模型之一,可自动提取不同级别的特征.利用CNN中对人脸图像空间的局部感受野、参数共享、稀疏连接和降采样,从输入数据中提取局部性和其他特征,可完成自主学习,从而隐式地获得更抽象的图像全局特征表情.随着网络层数的加深,CNN可以提取更细微的特征,在提取鲁棒性和具有代表性的特征方面起重要作用.但随着网络的加深,模型所需的参数量也会增加,从而延长了训练所需的时间并加重了计算机负担.计算和存储资源的需求使CNN部署在资源有限的设备上具有挑战性.

目前关于FER的研究已取得了许多成果.Yao等<sup>[7]</sup>提出了一种基于空间和通道-高度性能模块实现(high performance module implementation, HPMI)的混合注意力机制,以实现面部表情的自动识别.HPMI的加入可以增强关键特征的权重,使模型集中在训练过程中对表情分类有用的特征上,有效缓解了网络的溢出现象,抑制无用信息,同时也解决了输入维度与输出维度不一致的问题.Liu等<sup>[8]</sup>通过端到端的方式结合了4种典型的特征提取方法,进一步鉴别特征,提升了模型的特征提取能力.Podder等<sup>[9]</sup>采用最小参数CNN和迁移学习相结合的方式,设计了一个实时的应用程序,可同时检测多个人类的普遍情绪水平.Sadeghi等<sup>[10]</sup>利用一个直方图计算层,在卷积层的输出处提供特征映射的统计描述,在卡方距离方程中引入了一个可学习的矩阵,以此在直方图空间中训练所提出的CNN,提高了模型识别的准确度.Zhang等<sup>[11]</sup>采用生成对抗网络生成数据的方式生成了面部表情图像,扩充了数据集,简化了过拟合的问题.

近年来,深度卷积神经网络(deep convolutional neural network, DCNN)在FER方面取得了许多成果.Zhang等<sup>[12]</sup>提出了视频FER的混合DCNN模型,其两个独立的DCNN分别学习时间和空间特征的分割频带,然后在视频均衡器中学习到的深度信念网络的分割特征进行平均并合并,生成一个全局的视频特征表示.Singh等<sup>[13]</sup>将CNN应用于静态人脸图像,并对预处理前后的结果进行比较,验证了预处理可进一步提高精度.Verma等<sup>[14]</sup>通过修改CNN网络架构提高FER精度.Ma等<sup>[15]</sup>通过定义运动单元划分规则对专家先验知识进行编码,提出了多标签监督信息自适应学习区域特征的运动单元区域卷积神经网络,为检测提供了准确的监督信息和细粒度的指导.Li等<sup>[16]</sup>提出了一种新的多特征联合学习集成框架,将全局特征与几种不同的局部关键特征相结合进行联合迭代优化,以解决多标签表情识别不准确的问题.

为减少DCNN的计算和存储资源需求,目前已提出了几种使体系结构更轻、性能损失较小的方法:修剪连接<sup>[17]</sup>删除了神经元之间不重要的连接;量化<sup>[18]</sup>本质上是一种共享权值的策略,而量化的权值张量是一个具有多个共享权值的高系数矩阵;张量分解<sup>[19]</sup>利用权值的冗余性,通过张量分解卷积网络参数矩阵,并利用其低秩特征进行逼近,从而减少参数和计算;通过使用较大模型指导较小模型,知识蒸馏<sup>[20]</sup>增强了较小模型的功能,从而提高了较小模型的性能.

对设备的高需求刺激了网络结构的发展,目前已出现了一系列紧凑的高效模型:MobileNet<sup>[21]</sup>是一种采用深度可分离卷积设计的轻量级深度神经网络;Zhang等<sup>[22]</sup>利用逐点群卷积和通道洗牌机制改进信道间的信息流交换,以较低的计算成本获得了较高的准确率提升;Ma等<sup>[23]</sup>进一步考虑了在目标硬件上的实际速度;GhostNet<sup>[24]</sup>通过简单的操作生成更多的特性映射,以较小的代价从细微特征中提取了所需的信息.

虽然上述各种方法提高了FER任务的准确率,但大多数是通过建立复杂网络提高模型的性能,延长了训练的周期.因此,本文提出一种轻量级结构的FER模型,以减少参数的数量和计算资源需求.首先,以残差网络ResNet18<sup>[25]</sup>为骨架,调整其架构,重新搭建一个轻量级别的残差网络,并对残差模块进行改进;其次,在模型中引入深度可分离卷积(depthwise separable convolution, DSC)模块,通过

逐通道卷积和逐点卷积, 最大限度地减小模型参数, 保留模型的性能; 最后引入挤压-激励模块 (squeeze-and-excitation, SE)<sup>[26]</sup>, 并使用 Mish 函数<sup>[27]</sup>替代 ReLU 函数, 提升了通道权重调整的稳定性, 从通道角度自适应地调整特征权重。

### 1 本文方法

本文设计一种基于轻量级注意力残差网络的面部表情识别模型。该模型主要由三部分组成: 首先, 利用残差网络的优越性和架构的可调整性, 以残差网络为骨架重新搭建模型, 并对核心的残差模块进行改进; 其次, 引入 DSC 模块减少模型的参数量; 最后, 嵌入通道注意力模块 SE 自适应地调整通道权重信息, 并采用 Mish 函数替代 ReLU 函数, 提升通道权重调整的平稳性。

#### 1.1 优化的网络架构

残差网络克服了梯度消失问题, 减少了过拟合的程度, 性能较好。本文模型以残差网络为骨架进行搭建。残差模块作为残差网络的核心, 以跳层连接的形式实现, 即将单元的输入直接与单元输出加在一起进行激活, 其原理如图 1 所示。将期望的底层映射表示为  $H(x)$ , 使堆叠的非线性层拟合另一个映射  $F(x) = H(x) - x$ , 于是原始映射被重铸为  $F(x) + x$ ,  $F(x) + x$  可通过具有捷径连接的前馈神经网络实现。

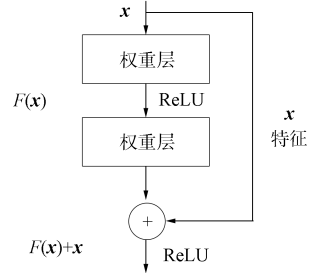


图 1 残差模块结构示意图

Fig. 1 Schematic diagram of residual module structure

网络的整体流程如图 2 所示。由图 2 可见, 该模型有 13 层深度, 包括 7 个卷积层 ( $C_1, C_2, C_3^*$ , 残差层 1, 残差层 2)、3 个池化层 ( $P_1, P_2, P_3$ )、1 个深度可分离卷积层 (SC)、1 个全连接层和 1 个分类器。输入层是一个面部表情图像, 大小为  $48 \times 48$  像素。卷积层  $C_1$  和  $C_2$  均使用 64 个通道进行卷积;  $C_3^*$  和 SC 则使用 128 个通道进行维度增强; 残差层 1 和残差层 2 使用 256 个卷积内核, 增加通道数量并接收更多特性, 卷积层连接使用一个批处理归一化层进行处理。最后, 将全连接层的输出分为 7 个情绪类别。其中, 残差层 1 和残差层 2 是在多次实验中改进的残差结构, 都具有两层卷积层, 跨层连接通过卷积核大小为  $3 \times 3$  的平均池化。改进的残差模块架构如图 3 所示。

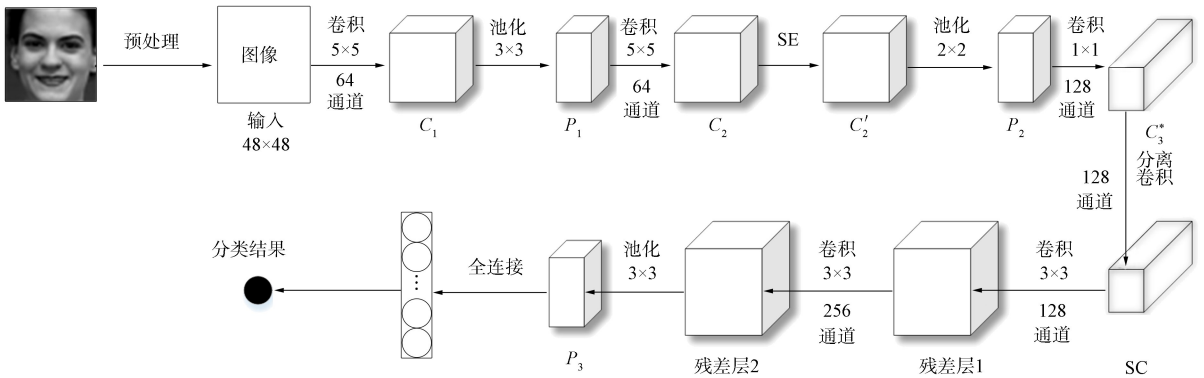


图 2 网络的整体架构

Fig. 2 Overall architecture of network

#### 1.2 深度可分离卷积

正则卷积同时考虑映射跨通道相关性和空间相关性。DSC 在单独映射空间相关性后, 再单独映射跨通道相关性。DSC 首先执行逐通道卷积, 当使用逐通道卷积时, 卷积内核通过仅卷积一个通道映射空间相关性, 从而产生具有完全相同数量通道的特征图。然后执行逐点卷积过程, 逐点卷积和正则卷积具有相同数量的通道或内核大小, 只是将卷积核尺寸变为  $1 \times 1 \times M$ , 其中  $M$  为通道数。逐点卷积在

深度方向上对逐通道卷积产生的特征图进行加权组合,生成新的特征图,同时计算的复杂度也因逐点卷积而改变.通过应用  $1 \times 1$  大小的卷积核对不同通道中的每个像素进行线性组合,DSC 不仅实现了跨通道信息交换,还保持了网络非线性映射能力的准确性,同时增强了网络结构的降维.DSC 的应用极大减少了模型的参数量,随着通道数的增加,参数减少的比例会正向提升.

1.3 注意力机制

为提高模型的特征提取性能,实现精确的人脸表情识别,本文利用信道注意机制对 CNN 提取的深度特征图进行重新校准.通过 SE 操作,网络可以从全局信息的角度分配不同的权重比,有选择性地放大有价值的特征信道,抑制无用的特征信道,以减少冗余信息.改进后的 SE 模块结构如图 4 所示.

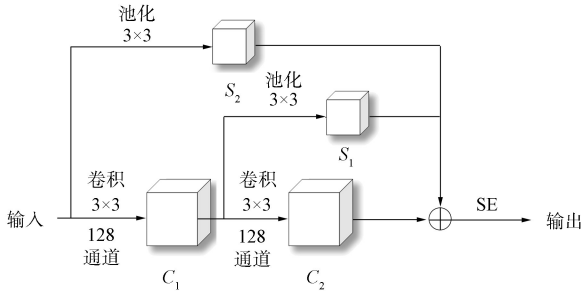


图 3 改进的残差模块架构

Fig. 3 Architecture of improved residual module

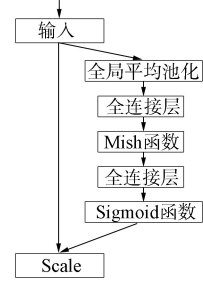


图 4 改进的注意力模块结构

Fig. 4 Structure of improved attention module

对特征图  $X$  经过任意给定的变换  $F_u: X \rightarrow U, X \in \mathbb{R}^{H' \times W' \times C'}, U \in \mathbb{R}^{H \times W \times C}$  得到的特征图  $U$ ,先将其每个通道的空间信息压缩到一个单值,即从  $H \times W \times C$  大小的  $U$  压缩到  $1 \times 1 \times C$  大小的向量,该操作作用公式表示为

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

其中  $H$  和  $W$  分别为特征图的高和宽,  $u_c(i, j)$  为特征图的坐标位置.对得到的  $1 \times 1 \times C$  大小的向量,应用一组具有门控机制的全连接(fully connected, FC)层对其进行权重调整,得到一个  $1 \times 1 \times C'$  的通道注意力向量.最后将该向量加权到特征图  $U$ ,形成一个加权后的新特征图  $U'$ ,即  $H' \times W' \times C'$ .该操作作用公式表示为

$$\tilde{X}_c = F_{scale}(u_c, s_c) = s_c \cdot u_c, \tag{2}$$

其中  $s_c$  为具有门控机制的 Sigmoid 函数对其调整后的通道特征标量,  $F_{scale}(u_c, s_c)$  为通道级乘法.对 SE, 本文将其中的 ReLU 函数更改为 Mish 函数, ReLU 函数和 Mish 函数分别为

$$h(x) = \begin{cases} x, & x > 0, \\ 0, & x \leq 0, \end{cases} \tag{3}$$

$$f(x) = x \cdot \tanh(\ln(1 + e^x)), \tag{4}$$

其中  $x$  为特征向量.由于 ReLU 函数在  $x < 0$  时梯度为 0,在  $x = 0$  点处不可微,导致负的梯度通过 ReLU 被置零,该神经元可能再也不会被任何数据激活,相应参数也就不会被更新. Mish 函数在零点处是平滑且可微的,弥补了 ReLU 函数的不足.因此,本文使用 Mish 函数取代 ReLU 函数的通道注意力机制,从通道的维度抑制不重要的通道特征,提高了通道权重分布的稳定性.

2 实验及结果分析

2.1 数据集

数据集 CK+<sup>[28]</sup>是 Cohn-Kande 的扩展版本,包含了 123 名受试者在实验室环境中采集的 593 个序列,共有 981 张带有标记的图像,包括 7 种基本表情:愤怒、中立、厌恶、恐惧、快乐、悲伤和惊讶.数据集 JAFFE<sup>[29]</sup>包含 10 名日本女性拍摄的 213 张  $256 \times 256$  大小的灰度图像,包括 7 种基本表情:愤怒、厌恶、恐惧、快乐、自然、悲伤和惊讶.图 5 为两个数据集的不同表情示例.



图 5 数据集 CK+ 和 JAFFE 的 7 种不同表情示例

Fig. 5 Examples of seven different facial expressions on CK+ and JAFFE datasets

### 2.2 数据预处理及训练

本文研究的目的是提升网络的基础性能, 并最大程度保留数据的原始性. 对数据集 CK+ 仅做了随机水平翻转处理, 然后转化为张量. 将数据集 JAFFE 的图像缩放为  $48 \times 48$  大小并随机水平翻转处理, 然后转化为张量输入网络进行训练.

由于数据集 CK+ 和 JAFFE 的数据量较少, 因此采用十折交叉验证的方式进行训练和测试, 固定学习率为 0.000 1 的 Adam 优化器进行 300 轮训练, 两个数据集的批处理大小均为 16. 采用更稳定的 Ten-crop 方式进行数据集测试, 以提升测试结果的可靠性. 模型基于 Pytorch 框架建立, 所有实验结果均在具有 24 GB 显存的 NVIDIA GeForce RTX 3090 GPU 上实现.

### 2.3 消融实验

本文模型在数据集 CK+ 和 JAFFE 上分别取得了 98.16% 和 96.67% 的准确率. 下面在两个公共数据集上对所提出的轻量级注意残差网络中的子模块的不同组合进行消融实验, 以验证各网络模块的有效性. 表 1 和表 2 分别列出了子模块的不同组合在数据集 CK+ 和 JAFFE 上的实验结果, 其中: 以残差网络为骨架搭建的 13 层深度的轻量级模型架构记为 ResNet13; 改进的残差模块 (improved residual module, IRM) 记为 IRM; 传统的 SE 模块记为 SE(ReLU); 使用 Mish 函数替代 ReLU 函数的 SE 模块记为 SE(Mish).

表 1 子模块的不同组合在数据集 CK+ 上的实验结果

Table 1 Experimental results of different combinations of submodules on CK+ dataset

实验组别	ResNet13	IRM	SE(ReLU)	SE(Mish)	准确率/%
1	✓	✓		✓	98.16
2	✓	✓	✓		97.35
3	✓	✓			96.63
4	✓				95.92

表 2 子模块的不同组合在数据集 JAFFE 上的实验结果

Table 2 Experimental results of different combinations of submodules on JAFFE dataset

实验组别	ResNet13	IRM	SE(ReLU)	SE(Mish)	准确率/%
1	✓	✓		✓	96.67
2	✓	✓	✓		95.72
3	✓	✓			94.77
4	✓				93.30

比较实验组 1 和实验组 4 的结果可见, 所添加的子模块与仅搭建的 ResNet13 对系统性能有提升, 对数据集 CK+ 和 JAFFE 识别准确率分别提升了 2.24, 3.37 个百分点. 比较实验组 3 和实验组 4 可见, 改进的残差模块 IRM 对 ResNet13 在两个数据集的识别准确率上分别提升了 0.71, 1.47 个百分点, 以较小的工作量提高了模型的识别准确率. 比较实验组 1 和实验组 2 可见, 使用 Mish 函数替代 ReLU 函数对两个数据集在其他模块的基础上识别准确率分别提升了 0.81, 0.95 个百分点. Mish 函数在零点处的光滑性使通道注意力模块获得了更好的泛化能力和优化能力, 在很大程度上提高了通道权

重分布的稳定性和准确性. 实验结果表明, 不同子模块均对系统性能有提升, 验证了本文模块的有效性.

图 6 为本文方法在数据集 CK+ 和 JAFFE 上的混淆矩阵. 有图 6 可见, 愤怒、厌恶、悲伤和恐惧几种情绪承担了主要混淆, 这主要是因为通用表情的相似性. 这几种表情都属于负面表情, 具有相似的特征. 例如: 这 4 种表情都包含皱眉的特征, 愤怒、厌恶和悲伤都会出现皱眉头和撇嘴的特征, 愤怒和恐惧都有睁大眼睛和张大嘴巴的动作. 这些表情的相似性导致面部特征难以提取, 因此, 这几种表情的识别更具挑战性.

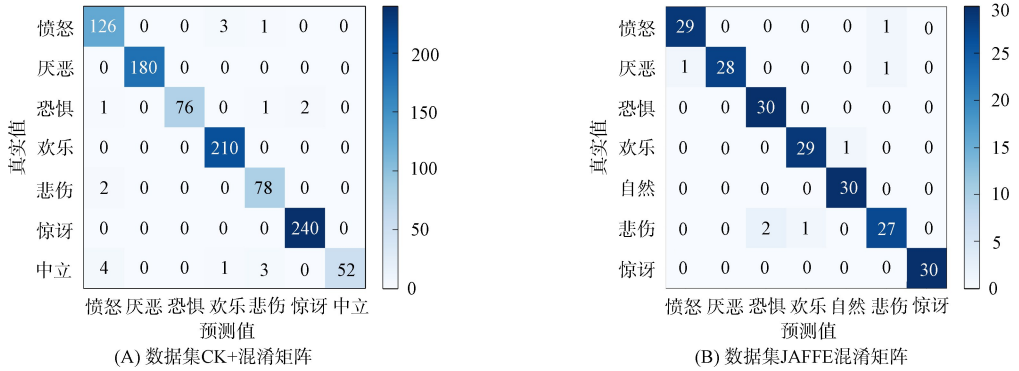


图 6 本文方法在数据集 CK+ 和 JAFFE 上的混淆矩阵

Fig. 6 Confusion matrix of proposed method on CK+ and JAFFE datasets

2.4 与主流算法的比较实验

将本文算法与 5 种主流算法 AlexNet<sup>[30]</sup>, VGG16, VGG19<sup>[31]</sup>, ResNet18, ResNet50<sup>[12]</sup> 进行对比实验, 在两个数据集上比较参数的大小和识别精度, 结果列于表 3.

表 3 本文算法与 5 种主流算法的对比结果

Table 3 Comparison results between proposed algorithm and 5 mainstream algorithms

模型	准确率/%		参数数量/M
	CK+	JAFFE	
AlexNet <sup>[30]</sup>	87.59	89.83	60.92
VGG16 <sup>[31]</sup>	95.46	91.04	14.75
VGG19 <sup>[31]</sup>	92.18	90.37	20.06
ResNet18 <sup>[12]</sup>	89.39	92.55	11.69
ResNet50 <sup>[12]</sup>	92.46	93.08	25.56
本文	98.16	96.67	1.52

由表 3 可见, 本文轻量级模型在数据集 CK+ 上的面部表情识别精度最高, 比 5 种主流算法中最高的 VGG16 模型高 2.70 个百分点, 而参数量仅为 VGG16 模型的约 1/10. 本文模型在数据集 JAFFE 上的识别准确率与 5 种主流算法中最高的 ResNet50 模型相比提高 3.59 个百分点, 而参数量仅为 ResNet50 模型的 1/17. 与 5 种主流算法相比, 本文轻量级模型在面部表情识别方面具有最高的精度和最好的算法性能. 在模型参数量上, 本文轻量级模型参数量为 1.52 M, 远低于其他 5 种主流算法, 并能保持较高的识别精度.

综上所述, 针对卷积神经网络参数量大、训练时间长的问题, 本文提出了一种基于轻量级注意力残差网络的 FER 方法. 首先, 基于残差网络骨架并改进残差模块重新设计了轻量级网络; 其次, 引入了 DSC 代替正则卷积运算, 极大减少了模型参数量和计算工作量; 最后, 采用 SE 模块调整通道权重分布, 并使用 Mish 函数替代 ReLU 函数提升调整权重的稳定性, 在数据集 CK+ 和 JAFFE 上分别取得了 98.16% 和 96.67% 的准确率. 消融实验结果以及与其他方法进行对比的结果验证了本文方法的可靠性. 由于数据集提供的面部图像和野外情况下的面部表情具有差异性, 因此未来需要进一步采集更多数据提升模型的性能.

## 参 考 文 献

- [1] SAJJAD M, NASIR M, ULLAH F U M, et al. Raspberry Pi Assisted Facial Expression Recognition Framework for Smart Security in Law-Enforcement Services [J]. *Information Sciences*, 2019, 479: 416-431.
- [2] KALANTARIAN H, JEDOUI K, WASHINGTON P, et al. Labeling Images with Facial Emotion and the Potential for Pediatric Healthcare [J]. *Artificial Intelligence in Medicine*, 2019, 98: 77-86.
- [3] LIU Z M, PENG Y X, HU W J. Driver Fatigue Detection Based on Deeply-Learned Facial Expression Representation [J]. *Journal of Visual Communication and Image Representation*, 2020, 71: 102723-1-102723-7.
- [4] PARK U, KIM M, JANG Y, et al. Robot Facial Expression Framework for Enhancing Empathy in Human-Robot Interaction [C]//2021 30th IEEE International Conference on Robot & Human Interactive Communication (RO-MAN). Piscataway, NJ: IEEE, 2021: 832-838.
- [5] TZIRAKIS P, TRIGEORGIS G, NICOLAOU M A, et al. End-to-End Multimodal Emotion Recognition Using Deep Neural Networks [J]. *IEEE Journal of Selected Topics in Signal Processing*, 2017, 11(8): 1301-1309.
- [6] CHEN Q Q, JING X J, ZHANG F P, et al. Facial Expression Recognition Based on a Lightweight CNN Model [C]//2022 IEEE International Symposium on Broadband Multimedia Systems and Broadcasting (BMSB). Piscataway, NJ: IEEE, 2022: 1-5.
- [7] YAO L S, HE S X, SU K, et al. Facial Expression Recognition Based on Spatial and Channel Attention Mechanisms [J]. *Wireless Personal Communications*, 2022, 125(2): 1483-1500.
- [8] LIU J, WANG H X, FENG Y J. An End-to-End Deep Model with Discriminative Facial Features for Facial Expression Recognition [J]. *IEEE Access*, 2021, 9: 12158-12166.
- [9] PODDER T, BHATTACHARYA D, MAJUMDAR A. Time Efficient Real Time Facial Expression Recognition with CNN and Transfer Learning [J]. *Sādhanā*, 2022, 47(3): 177-1-177-20.
- [10] SADEGHI H, RAIE A A. HistNet: Histogram-Based Convolutional Neural Network with Chi-Squared Deep Metric Learning for Facial Expression Recognition [J]. *Information Sciences*, 2022, 608: 472-488.
- [11] ZHANG X, ZHANG F F, XU C S. Joint Expression Synthesis and Representation Learning for Facial Expression Recognition [J]. *IEEE Transactions on Circuits and Systems for Video Technology*, 2021, 32(3): 1681-1695.
- [12] ZHANG S Q, PAN X Z, CUI Y L, et al. Learning Affective Video Features for Facial Expression Recognition via Hybrid Deep Learning [J]. *IEEE Access*, 2019, 7: 32297-32304.
- [13] SINGH S, NASOZ F. Facial Expression Recognition with Convolutional Neural Networks [C]//2020 10th Annual Computing and Communication Workshop and Conference (CCWC). Piscataway, NJ: IEEE, 2020: 324-328.
- [14] VERMA A, SINGH P, ALEX J S R. Modified Convolutional Neural Network Architecture Analysis for Facial Emotion Recognition [C]//2019 International Conference on Systems, Signals and Image Processing (IWSSIP). Piscataway, NJ: IEEE, 2019: 169-173.
- [15] MA C, CHEN L, YONG J H. Au R-CNN: Encoding Expert Prior Knowledge into R-CNN for Action Unit Detection [J]. *Neurocomputing*, 2019, 355: 35-47.
- [16] LI W Z, LUO M Y, ZHANG P, et al. A Novel Multi-feature Joint Learning Ensemble Framework for Multi-label Facial Expression Recognition [J]. *IEEE Access*, 2021, 9: 119766-119777.
- [17] RATHI N, PANDA P, ROY K. STDP-Based Pruning of Connections and Weight Quantization in Spiking Neural Networks for Energy-Efficient Recognition [J]. *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, 2018, 38(4): 668-677.
- [18] THONGLEK K, TAKAHASHI K, ICHIKAWA K, et al. Retraining Quantized Neural Network Models with Unlabeled Data [C]//2020 International Joint Conference on Neural Networks (IJCNN). Piscataway, NJ: IEEE, 2020: 1-8.
- [19] WANG J L, ZHAO G S, WANG D H, et al. Tensor Completion Using Low-Rank Tensor Train Decomposition by Riemannian Optimization [C]//2019 Chinese Automation Congress (CAC). Piscataway, NJ: IEEE, 2019:

3380-3384.

- [20] YUAN M K, PENG Y X. CKD: Cross-Task Knowledge Distillation for Text-to-Image Synthesis [J]. IEEE Transactions on Multimedia, 2019, 22(8): 1955-1968.
- [21] HOWARD A G, ZHU M, CHEN B, et al. Mobilenets: Efficient Convolutional Neural Networks for Mobile Vision Applications [EB/OL]. (2017-04-17)[2023-02-10]. <https://arxiv.org/abs/1704.04861>.
- [22] ZHANG X Y, ZHOU X Y, LIN M X, et al. Shufflenet: An Extremely Efficient Convolutional Neural Network for Mobile Devices [C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE, 2018: 6848-6856.
- [23] MA N, ZHANG X, ZHENG H T, et al. Shufflenet V2: Practical Guidelines for Efficient CNN Architecture Design [C]//Proceedings of the European Conference on Computer Vision (ECCV). Piscataway, NJ: IEEE, 2018: 116-131.
- [24] 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. Piscataway, NJ: IEEE, 2020: 1580-1589.
- [25] 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. Piscataway, NJ: IEEE, 2016: 770-778.
- [26] HU J, SHEN L, SUN G. Squeeze-and-Excitation Networks [C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE, 2018: 7132-7141.
- [27] MISRA D. Mish: A Self Regularized Non-monotonic Activation Function [EB/OL]. (2019-08-23)[2023-03-15]. <https://arxiv.org/abs/1908.08681>.
- [28] 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. Piscataway, NJ: IEEE, 2010: 94-101.
- [29] LYONS M J, AKAMATSU S, KAMACHI M, et al. The Japanese Female Facial Expression (JAFPE) Database [C]//Proceedings of Third International Conference on Automatic Face and Gesture Recognition. Piscataway, NJ: IEEE, 1998: 14-16.
- [30] KRIZHEVSKY A, SUTSKEVER I, HINTON G E. Imagenet Classification with Deep Convolutional Neural Networks [J]. Communications of the ACM, 2017, 60(6): 84-90.
- [31] SIMONYAN K, ZISSERMAN A. Very Deep Convolutional Networks for Large-Scale Image Recognition [EB/OL]. (2014-09-04)[2023-04-10]. <https://arxiv.org/abs/1409.1556>.

(责任编辑:韩 啸)