



面向表情识别的重影非对称残差注意力网络模型

闫河, 李梦雪, 张宇宁, 刘建骐

引用本文:

闫河,李梦雪,张宇宁,刘建骐. 面向表情识别的重影非对称残差注意力网络模型[J]. 智能系统学报, 2023, 18(2): 333–340.

YAN He,LI Mengxue,ZHANG Yuning,LIU Jianqi. A ghost asymmetric residual attention network model for facial expression recognition[J]. *CAAI Transactions on Intelligent Systems*, 2023, 18(2): 333–340.

在线阅读 View online: <https://dx.doi.org/10.11992/tis.202201003>

您可能感兴趣的其他文章

基于混合卷积与三重注意力的高光谱图像分类网络

Hyperspectral image classification based on hybrid convolutional neural network with triplet attention

智能系统学报. 2023, 18(2): 260–269 <https://dx.doi.org/10.11992/tis.202204002>

深度多尺度融合注意力残差人脸表情识别网络

Deep multiscale fusion attention residual network for facial expression recognition

智能系统学报. 2022, 17(2): 393–401 <https://dx.doi.org/10.11992/tis.202107028>

基于双向LSTM卷积网络与注意力机制的自动睡眠分期模型

Automatic sleep staging model based on the bi-directional LSTM convolutional network and attention mechanism

智能系统学报. 2022, 17(3): 523–530 <https://dx.doi.org/10.11992/tis.202103013>

改进MobileNet的图像分类方法研究

Research on the improved image classification method of MobileNet

智能系统学报. 2021, 16(1): 11–20 <https://dx.doi.org/10.11992/tis.202012034>

基于改进的Faster RCNN面部表情检测算法

Facial expression recognition based on improved Faster RCNN

智能系统学报. 2021, 16(2): 210–217 <https://dx.doi.org/10.11992/tis.201910020>

基于双注意力模型和迁移学习的Apex帧微表情识别

Apex frame microexpression recognition based on dual attention model and transfer learning

智能系统学报. 2021, 16(6): 1015–1020 <https://dx.doi.org/10.11992/tis.202010031>

DOI: 10.11992/tis.202201003

网络出版地址: <https://kns.cnki.net/kcms/detail/23.1538.TP.20221019.1317.006.html>

面向表情识别的重影非对称残差注意力网络模型

闫河, 李梦雪, 张宇宁, 刘建骐

(重庆理工大学 两江人工智能学院, 重庆 401135)

摘要: 针对 ResNet50 中的 Bottleneck 经过 1×1 卷积降维后主干分支丢失部分特征信息而导致在表情识别中准确率不高的问题, 本文通过引入 Ghost 模块和深度可分离卷积分别替换 Bottleneck 中的 1×1 卷积和 3×3 卷积, 保留更多原始特征信息, 提升主干分支的特征提取能力; 利用 Mish 激活函数替换 Bottleneck 中的 ReLU 激活函数, 提高了表情识别的准确率; 在此基础上, 通过在改进的 Bottleneck 之间添加非对称残差注意力模块 (asymmetric residual attention block, ARABlock) 来提升模型对重要信息的表示能力, 从而提出一种面向表情识别的重影非对称残差注意力网络 (ghost asymmetric residual attention network, GARAN) 模型。对比实验结果表明, 本文方法在 FER2013 和 CK+表情数据集上具有较高的识别准确率。

关键词: 表情识别; 特征提取; ResNet50; Ghost 模块; Mish; 非对称残差注意力; 深度可分离卷积; 深度学习
中图分类号: TP391 **文献标志码:** A **文章编号:** 1673-4785(2023)02-0333-08

中文引用格式: 闫河, 李梦雪, 张宇宁, 等. 面向表情识别的重影非对称残差注意力网络模型 [J]. 智能系统学报, 2023, 18(2): 333-340.

英文引用格式: YAN He, LI Mengxue, ZHANG Yuning, et al. A ghost asymmetric residual attention network model for facial expression recognition[J]. CAAI transactions on intelligent systems, 2023, 18(2): 333-340.

A ghost asymmetric residual attention network model for facial expression recognition

YAN He, LI Mengxue, ZHANG Yuning, LIU Jianqi

(School of Artificial Intelligence, Chongqing University of Technology, Chongqing 401135, China)

Abstract: In this paper, a solution is proposed to address the low accuracy in facial expression recognition that results from the 1×1 convolution dimensionality reduction of the Bottleneck in ResNet50. To do so, the authors introduce the Ghost module and depth separable convolution to replace the 1×1 and 3×3 convolutions in the Bottleneck, respectively, in order to preserve more of the original feature information and improve the feature extraction ability of the trunk branch. The Mish activation function is also used to replace the ReLU activation function in the Bottleneck, further enhancing the accuracy of facial expression recognition. To further improve the ability of the model to express important information, the authors also introduce an asymmetric residual attention block (ARABlock) between the improved Bottlenecks. The proposed method, which is referred to as the ghost asymmetric residual attention network (GARAN) model, shows high recognition accuracy on the FER2013 and CK+ facial expression datasets based on comparative experimental results.

Keywords: expression recognition; feature extraction; ResNet50; Ghost module; Mish; asymmetric residual attention; depthwise separable convolution; deep learning

表情识别是通过人类的面部表情来判断情绪的一种手段。在最近的工作中, 微表情识别逐渐成为该领域的研究热点。微表情持续时间短、难以捕捉, 却表达了人试图压抑与隐藏的真实情感, 在自动测谎、心理咨询以及临床诊断等方面

有着很高的研究价值及应用前景^[1]。目前, 基于卷积神经网络 (convolutional neural networks, CNN) 的表情识别方法取得了不错的进展^[2-4]。文献 [5] 提出一种改进 LeNet-5 的表情识别方法, 采用跨连接的方式将低层次特征与高层次特征结合起来, 使网络能够提取到更多的细节特征信息。文献 [6] 将 VGG16 中的全连接模块替换为深度卷积和全局平均池化, 较好地避免了过拟合问题。深层网络更容易提取到具有丰富语义信息的深层

收稿日期: 2022-01-04. 网络出版日期: 2022-10-19.

基金项目: 国家重点研发计划“智能机器人”重点专项 (2018YFB-1308602); 国家自然科学基金面上项目 (61173184); 重庆市自然科学基金项目 (cstc2018jcyjAX0694).

通信作者: 闫河. E-mail: cqyanhe@163.com.

特征,但是,随着网络的加深,梯度消失和梯度爆炸的现象越来越明显。深度残差网络(residual network, ResNet)^[7]通过在残差模块间使用跳跃连接使得梯度正常回传,有效解决了这一问题。文献[8]将 ResNet34 与 U-Net 结构相结合,生成一个残差掩膜网络,通过掩膜块来细化输入特征映射,并为相应的特征映射生成权重,使网络能够聚焦关键的空间信息以做出正确的决策,提高了模型的表情识别性能。文献[9]提出的深度多尺度融合注意力残差网络(deep multi-scale fusion attention residual network, DMFA-ResNet)以 ResNet50 为主干网络,使用 3 个 3×3 卷积代替原 7×7 卷积,在保证网络具有相同感受野的同时进一步提升网络深度;通过引入三支路的注意力残差模块突出重点局部区域,提高识别准确率。文献[10]受金字塔卷积(pyramidal convolution, PyConv)^[11]的启发,结合通道和空间注意力以及 PyConv 构成一个新的模块 PyConv-CSneck,以此来替换 ResNet50 中的 Bottleneck,在提取多尺度特征的同时突出面部细节信息,强化了网络的特征提取能力。ResNet50 中的 Bottleneck 采用 1×1 卷积进行先降维后升维的操作增加了网络的非线性,同时降低了模型计算量。但是,经过降维操作后再使用 ReLU 函数会造成激活空间坍塌,进而使得输入特征信息丢失严重,在一定程度上限制了主干网络的特征提取能力^[12],而升维操作则会产生过多的相似特征图^[13],导致了资源的浪费。

针对以上问题,部分学者进行了相关研究。Ghost 模块^[13]利用线性变化生成更多的特征图,能够以很小的代价从原始特征中挖掘更多特征信息,从而提升网络的特征提取能力。文献[14]将 Ghost-Net 作为目标跟踪的主干特征提取网络,文献[15]在 YOLOv4 中引入 Ghost 模块来提升目标识别的精度,在各自的任务中都具有良好的表现。注意力机制的作用是使网络学到更多重要的细节信息,同时忽略不重要的信息。文献[16]提出的残差注意力学习模块使用恒等映射单元来减轻信号的衰减,并且随着网络的深入其注意力感知功能会自适应变化,更好地感知特征。结合 Ghost 模块与残差注意力模块的思想,本文以 ResNet50 为主干网络,提出一种面向表情识别的重影非对称残差注意力网络模型(ghost asymmetric residual attention network, GARAN)。本文的主要贡献如下:

1)采用 Ghost 模块替换 Bottleneck 中的 1×1 卷积,采用 Depthwise 卷积替换 Bottleneck 中的 3×3 卷积,保留更丰富的特征信息,提高表情识别准确率,并且降低了模型的参数和计算量。

2)引入 Mish 激活函数^[17]替换原 Bottleneck

中的 ReLU 激活函数,避免了网络结构加深时带来的梯度消失问题。

3)在 Ghost_Bottleneck 之间引入残差注意力模块来聚焦面部关键特征,有效地解决了细节信息丢失的问题;并将残差注意力模块中的 3×3 标准卷积分解为 1×3 和 3×1 连续两层不对称卷积,降低模型计算量的同时提升了网络的识别性能。

4)在 Fer2013 数据集和 CK+数据集上进行训练和测试,并与 ResNet50 以及当前的主流方法进行了对比实验。结果表明本文所提方法在表情识别的准确率上具有明显的提升。

1 GARAN 算法

1.1 Ghost 模块

深度卷积神经网络中大量的卷积运算通常会导致整个网络的计算成本过大以及特征冗余问题。文献[13]提出的 Ghost 模块将标准卷积分为两步,第一步使用普通的 1×1 卷积生成部分原始特征,第二步对生成的部分原始特征进行低成本的线性运算(本文采用 3×3 的通道卷积)来增强特征和增加信道,然后将两组特征拼接在一起。Ghost 模块与标准卷积相比能更好地提取特征,并且可以减少模型的参数和计算量,有效提高网络性能。Ghost 模块的结构如图 1 所示。

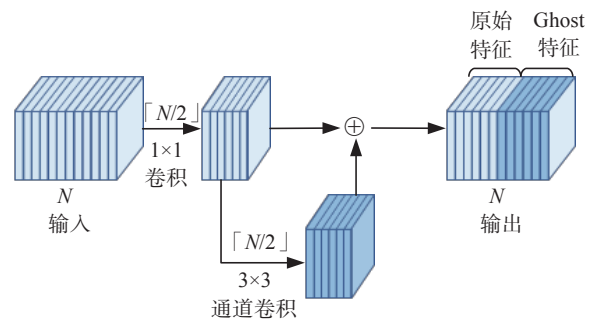


图 1 Ghost 模块

Fig. 1 Ghost module

1.2 改进的 Ghost_Bottleneck

ResNet50 中的 Bottleneck 结构如图 2(a) 所示,由 2 个 1×1 卷积和 1 个 3×3 卷积构成,但是通过步长为 2 的 1×1 卷积下采样后会造成部分特征信息丢失^[18]。本文使用的 Ghost_Bottleneck 由 2 个 Ghost 模块和 1 个步长为 2 的 Depthwise 卷积构成,其中第 1 个 Ghost 模块用来扩大通道数,Depthwise 卷积将特征图大小降维到输入特征图的 $1/2$,第 2 个 Ghost 模块将通道数恢复到与输入通道数一致,然后通过残差边与输入进行跳跃连接得到最终的输出,保证网络精度的同时保留更多有效特征信息。

注意模块 $A(x)$ 的输出为

$$A_{i,c}(x) = (1 + S_{i,c}(x)) \cdot F_{i,c}(x) \quad (3)$$

式中: i 代表神经元的空间位置; c 代表神经元的通道位置; $S_{i,c}(x)$ 的取值在 $[0, 1]$ 区间内, 当 $S(x)$ 接近 0 时, $A(x)$ 的输出接近原始特征 x 。

2 重影非对称残差注意力网络模型

本文提出的网络模型通过交叉堆叠多个 Ghost_

Bottleneck 和多个非对称残差注意力模块 ARABlock 构建而成, 完整的网络结构如图 4 所示。采用 ResNet50 作为主干网络结构, 使用改进后的 Ghost_Bottleneck 替换原 ResNet50 中的 Bottleneck, 在减少计算量的同时最大程度保留原始图像细节信息; 在多个 Ghost_Bottleneck 之间增加非对称残差注意力模块, 进一步加深网络, 增强局部重点区域的特征表达能力。

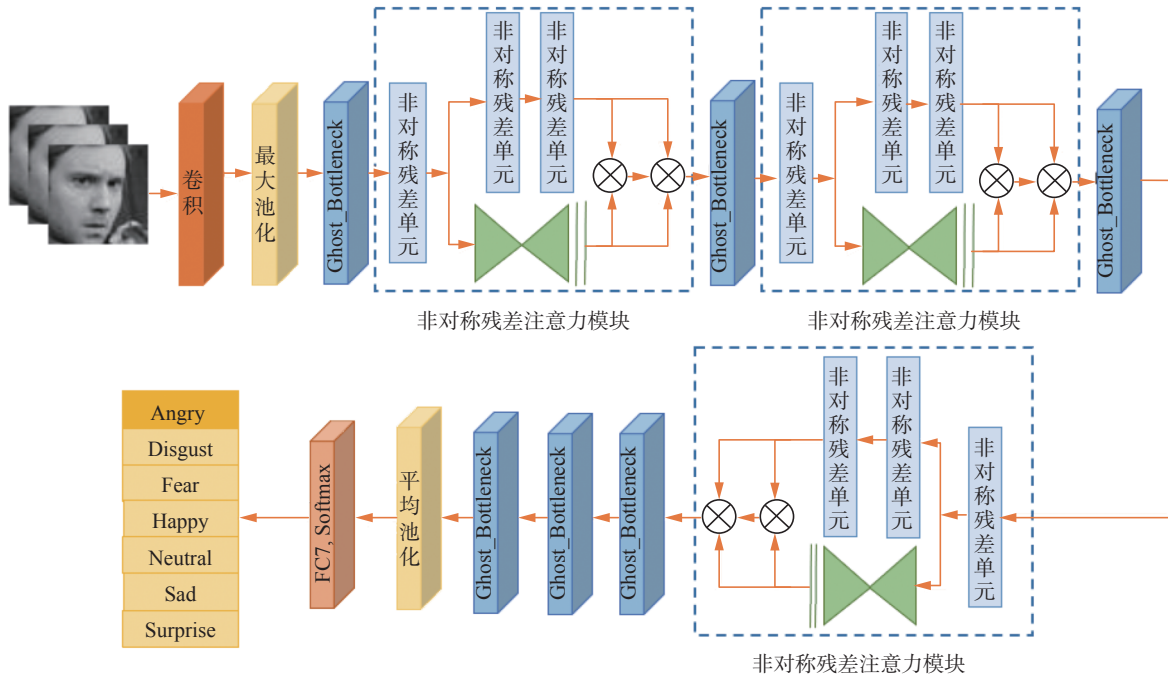


图 4 GARAN 网络结构

Fig. 4 GARAN network structure

本文所提网络模型的详细信息如表 1 所示。首先, 大小为 224×224 的输入图像经过一个 7×7 的卷积和 2×2 的最大池化下采样, 将空间缩小到 56×56 。然后, 经过 6 个 Ghost_Bottleneck 和 3 个非对称残差注意力模块堆叠, 将输出的特征图空间尺寸降为 7×7 。再经过平均池化层使空间大小变为 1×1 , 最后使用一个 7 路全连接层以及 Soft-Max 激活函数生成对应于 7 种面部表情的输出。

表 1 网络模型的详细信息

Table 1 Details of the network model

网络层	输出	卷积核
Conv1	$112 \times 112 \times 64$	7×7 , stride=2
Max Pooling	$56 \times 56 \times 64$	3×3 , stride=2
Ghost_Bottleneck1	$56 \times 56 \times 256$	$\begin{pmatrix} \text{Ghost module} \\ \text{DWConv} \\ \text{Ghost module} \end{pmatrix} \times 1$
ARABlock1	$56 \times 56 \times 256$	Attention $\times 1$

续表 1

网络层	输出	卷积核
Ghost_Bottleneck2	$28 \times 28 \times 512$	$\begin{pmatrix} \text{Ghost module} \\ \text{DWConv} \\ \text{Ghost module} \end{pmatrix} \times 1$
ARABlock2	$28 \times 28 \times 512$	Attention $\times 1$
Ghost_Bottleneck3	$14 \times 14 \times 1024$	$\begin{pmatrix} \text{Ghost module} \\ \text{DWConv} \\ \text{Ghost module} \end{pmatrix} \times 1$
ARABlock3	$14 \times 14 \times 1024$	Attention $\times 1$
Ghost_Bottleneck4	$7 \times 7 \times 2048$	$\begin{pmatrix} \text{Ghost module} \\ \text{DWConv} \\ \text{Ghost module} \end{pmatrix} \times 3$
Average Pooling	$1 \times 1 \times 2048$	7×7 , stride=1
FC, SoftMax	7	/

3 实验与分析

本文对所提方法进行了系列对比实验, 实验环境如下: 操作系统是 64 位 Windows 10, 编程语

言使用 Python3.8, 深度学习框架为 Pytorch1.7.0, CPU 为 18 核心 36 线程 3.09 GHz、96 GB, 显卡为 NVIDIA Quadro RTX 8000、48 GB。在训练阶段, 设置 batch_size 为 48, 学习率为 10^{-4} , 动量为 0.9, 权重衰减为 10^{-3} , 优化器采用 Adam。为了防止过拟合以及增强模型的泛化能力, 在全连接层使用 Dropout 技术, 设置 dropout=0.5; 当模型在验证集上的表现开始下降时, 及时停止训练, 使模型达到一个较好的拟合效果。

3.1 表情数据集

本文采用 Fer2013 和 CK+ 表情数据集进行实验。Fer2013 数据集由 35 886 张人脸表情图片组成, 其中, 训练图片 (Training) 28 708 张, 公共测试图片 (PublicTest) 和私有测试图片 (PrivateTest) 各 3 589 张, 每张图片是由大小固定为 48×48 的灰度图像组成, 共有 7 种表情, 具体表情对应的标签和中英文如下: 0 anger 生气, 1 disgust 厌恶, 2 fear 恐惧, 3 happy 开心, 4 sad 伤心, 5 surprise 惊讶, 6 neutral 中性。每种表情的示例如图 5。



图 5 Fer2013 数据集中 7 种表情示例

Fig. 5 Seven emoticons from the Fer2013 dataset

CK+数据集包含了 123 位受试者的 593 个视频序列, 其中 327 个视频序列带有表情标签。本实验选取除轻蔑表情外的其他 7 种基本表情 (愤怒、厌恶、恐惧、高兴、悲伤、惊讶和中性), 每张图像的原始尺寸为 640×490, 将其处理为 48×48 大小。图 6 给出了经过裁剪处理后的其余 7 种表情示例。

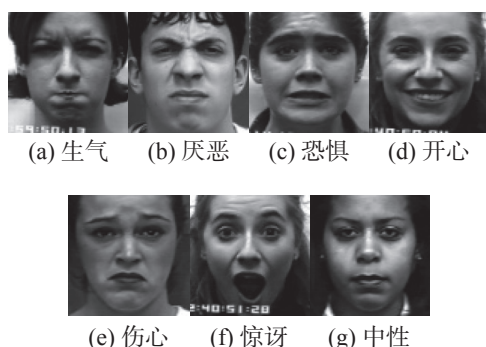


图 6 CK+数据集中 7 种表情示例

Fig. 6 Seven emoticons from the CK+ dataset

由于表情数据集的样本较少, 本文在数据集预处理阶段, 分别对图像进行了缩放、旋转以及水平翻转操作, 增加训练样本的多样性。

3.2 评价指标

为了评估本文提出的方法在面部表情识别问题中的有效性, 采用识别准确率来衡量模型的性能, 将其定义为

$$\text{Accuracy} = \sum_{i=1}^7 \frac{N_i}{n_T} \quad (4)$$

式中: Accuracy 表示识别准确率; N_i 表示第 i 类识别正确的表情数量; n_T 表示为目标域的表情数量。

3.3 消融实验

为了验证 Ghost_Bottleneck 与非对称残差注意力模块在特征提取中的有效性, 本文设计了如下实验。其中 ResNet50 是没有进行任何改进的基础网络, ARABlock 表示非对称残差注意力模块, Ghost_Bottleneck 表示将 Bottleneck 中的 1×1 和 3×3 卷积分别替换为 Ghost 模块和 DeepWise 卷积, Mish 表示将 Ghost_Bottleneck 中的 ReLU 激活函数替换为 Mish 激活函数, 将各个模块进行对比实验, 实验结果如表 2 所示。

表 2 不同模块识别准确率对比

方法	Fer2013	CK+
ResNet50	69.71	95.23
ResNet50 + ARABlock	71.94	97.41
ResNet50 + Ghost_Bottleneck	71.35	96.93
ResNet50 + Ghost_Bottleneck + Mish	71.58	97.35
The Proposed	72.14	98.67

从表 2 中的数据可以看出, 改进的 Ghost_Bottleneck、非对称残差注意力模块、Mish 激活函数都能够有效地提升表情识别的准确率, 并且同时引入这几个模块对模型进行改进后的识别准确率提升最为显著, 在 Fer2013 和 CK+数据集上的表情识别准确率分别为 72.14% 和 98.67%, 与基础网络 ResNet50 相比提升了 2.43% 和 3.44%, 从而验证了本文所提方法的有效性。

图 7 和图 8 分别给出了本文所提方法在 Fer2013 和 CK+数据集上的混淆矩阵。由图 7 的混淆矩阵可知, 在数据集 Fer2013 中, 本文方法在高兴 (90%)、惊讶 (81%)、厌恶 (75%) 和中性 (72%) 表情的识别准确率较高, 这是因为这几种表情的面部动作变化明显, 能够产生更多差异性的特征点, 因此更容易识别。而生气 (66%)、害怕 (55%) 和悲伤

(59%)表情的识别准确率较低,且它们之间的误判率相对较高,其主要原因是这几种表情具有较高的相似性,易混淆,并且面部关键点特征变化不明显。

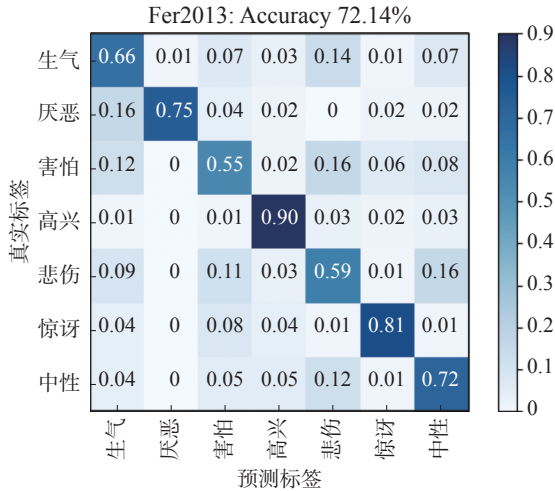


图 7 Fer2013 数据集的混淆矩阵

Fig. 7 Confusion matrix of Fer2013 dataset

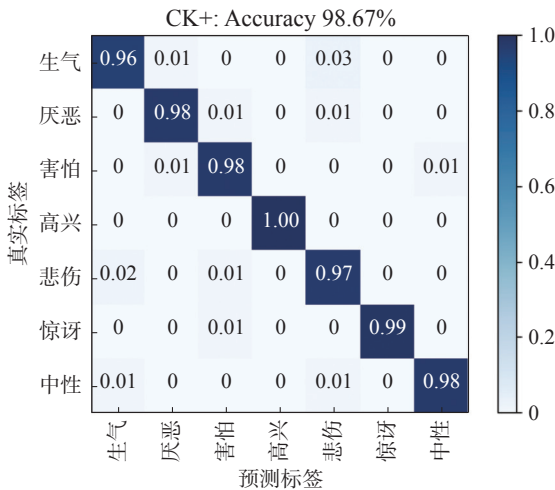


图 8 CK+数据集的混淆矩阵

Fig. 8 Confusion matrix of CK+ dataset

由图 8 的混淆矩阵可知,本文方法在 CK+数据集上能取得较好的识别结果,其中高兴和惊讶的识别准确率最高,分别为 99% 和 98%;而生气和悲伤的识别准确率较低,分别为 96% 和 97%,且容易将生气表情错误地判断为悲伤表情,这是因为两者都属于消极情绪,类间差异小,具体表现为表情发生时都伴随着眉头紧皱、嘴角朝下的动作,因此分类相对困难。

3.4 其他主流方法对比实验

为了进一步验证所提方法的有效性,本文在 Fer2013 和 CK+数据集上与当前比较新的几种表情识别方法分别进行了对比实验。对比结果如表 3 和表 4 所示。

表 3 不同方法在 Fer2013 数据集上的识别结果

Table 3 Recognition results of different methods on the Fer2013 dataset

方法	准确率/%
I-VGG16 ^[6]	71.84
ResMasking ^[8]	71.97
APRNET50 ^[10]	71.05
Deep-Emotion ^[21]	70.02
MBCC-CNN ^[22]	71.52
ResNet50(baseline)	69.71
The Proposed	72.14

表 4 不同方法在 CK+数据集上的识别结果

Table 4 Recognition results of different methods on the CK+ dataset

方法	准确率/%
C-LeNet5 ^[5]	83.74
ResMasking ^[8]	98.46
APRNET50 ^[10]	94.95
DTAGN ^[23]	97.25
DeRL ^[24]	97.30
ResNet50(baseline)	95.23
The Proposed	98.67

分析表 3 和表 4 可知,与现有的表情识别方法相比,本文提出的方法在 Fer2013 和 CK+两个标准数据集上均取得了不错的识别效果。在 Fer2013 数据集上,本文对比了 5 种主流的表情识别算法,其中 Deep-Emotion^[21]利用注意卷积网络来专注面部的重要部位,仅用一个很浅的网络就达到了较好的识别效果。MBCC-CNN^[22]将残差连接、网中网以及树状多分支结构结合在一起,提高了模型的特征提取能力。本文所提方法的识别准确率相比文献[6]提高了 0.3%,相比文献[8]提高了 0.17%,相比文献[10]提高了 1.09%,相比文献[21]提高了 2.12%,相比文献[22]提高了 0.62%。在 CK+数据集上,本文也对比了 5 种主流的表情识别方法,DTAGN^[23]采用两个子网络分别提取时间外观特征和时间几何特征,DeRL^[24]通过 cGAN 生成中性表情进而学习生成模型中间层的残留表情。本文方法的识别准确率相比文献[5、8、10、23-24]分别提高了 14.93%、0.21%、3.72%、1.42%以及 1.37%。可见,本文所提方法能够很好地学习表情特征,具有更高的识别准确率。

4 结束语

本文提出了一种面向表情识别的重影非对称残差注意网络(GARAN),该网络在ResNet50的基础上进行改进,将Bottleneck中的 1×1 卷积替换为Ghost模块, 3×3 卷积替换为深度可分离卷积,将ReLU激活函数替换为Mish激活函数,利用线性运算来增强特征表现,降低模型的参数量和计算量;通过引入非对称残差注意机制,增强模型对有用特征信息的响应,减少对无用特征信息的干扰。分别在Fer2013和CK+数据集上进行消融实验,验证了本文所提方法的有效性;通过与其他主流方法的对比进一步验证了本文方法的优越性。但是,由于表情的类间差异较小,生气、悲伤、害怕等表情仍具有较高的误判率,在之后的工作中将致力于解决高度相似表情间的误判问题,进一步提高表情的识别准确率。

参考文献:

- [1] 贾晔烨,杨明强,张鹏,等.微表情自动识别综述[J].计算机辅助设计与图形学学报,2014,26(9):1385-1395.
BEN Xianye, YANG Mingqiang, ZHANG Peng, et al. Survey on automatic micro expression recognition methods[J]. Journal of computer-aided design & computer graphics, 2014, 26(9): 1385-1395.
- [2] LI Shan, DENG Weihong. Deep facial expression recognition: a survey[J]. IEEE transactions on affective computing, 2022, 13(3): 1195-1215.
- [3] WANG Kai, PENG Xiaojang, YANG Jianfei, et al. Suppressing uncertainties for large-scale facial expression recognition[C]//2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle: IEEE, 2020: 6896-6905.
- [4] RUAN Delian, YAN Yan, LAI Shenqi, et al. Feature decomposition and reconstruction learning for effective facial expression recognition[C]//2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Nashville: IEEE, 2021: 7656-7665.
- [5] 李勇,林小竹,蒋梦莹.基于跨连接LeNet-5网络的面部表情识别[J].自动化学报,2018,44(1):176-182.
LI Yong, LIN Xiaozhu, JIANG Mengying. Facial expression recognition with cross-connect LeNet-5 network[J]. Acta automatica sinica, 2018, 44(1): 176-182.
- [6] 崔子越,皮家甜,陈勇,等.结合改进VGGNet和Focal Loss的人脸表情识别[J].计算机工程与应用,2021,57(19):171-178.
CUI Ziyue, PI Jiatian, CHEN Yong, et al. Facial expression recognition combined with improved VGGNet and focal loss[J]. Computer engineering and applications, 2021, 57(19): 171-178.
- [7] HE Kaiming, ZHANG Xiangyu, REN Shaoqing, et al. Deep residual learning for image recognition[C]//2016 IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas: IEEE, 2016: 770-778.
- [8] PHAM L, VU T H, TRAN T A. Facial expression recognition using residual masking network[C]//2020 25th International Conference on Pattern Recognition. Milan: IEEE, 2020: 4513-4519.
- [9] 高涛,杨朝晨,陈婷,等.深度多尺度融合注意力残差人脸表情识别网络[J].智能系统学报,2022,17(2):393-401.
GAO Tao, YANG Zhaochen, CHEN Ting, et al. Deep multiscale fusion attention residual network for facial expression recognition[J]. CAAI transactions on intelligent systems, 2022, 17(2): 393-401.
- [10] 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.
- [11] DUTA I C, LIU LI, ZHU FAN, et al. Pyramidal convolution: rethinking convolutional neural networks for visual recognition[C]//Proceedings of the 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle: IEEE, 2020.
- [12] SANDLER M, HOWARD A, ZHU Menglong, et al. MobileNetV2: inverted residuals and linear bottlenecks[C]//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City: IEEE, 2018: 4510-4520.
- [13] HAN Kai, WANG Yunhe, TIAN Qi, et al. GhostNet: more features from cheap operations[C]//2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle: IEEE, 2020: 1577-1586.
- [14] 刘超军,段喜萍,谢宝文.应用GhostNet卷积特征的ECO目标跟踪算法改进[J].激光技术,2022,46(2):239-247.
LIU Chaojun, DUAN Xiping, XIE Baowen. Improvement of eco target tracking algorithm based on GhostNet convolution feature[J]. Laser technology, 2022, 46(2): 239-247.
- [15] 符惠桐,王鹏,李晓艳,等.面向移动目标识别的轻量化网络模型[J].西安交通大学学报,2021,55(7):124-131.
FU Huitong, WANG Peng, LI Xiaoyan, et al. Light-

- weight network model for moving object recognition[J]. *Journal of Xi'an Jiaotong university*, 2021, 55(7): 124–131.
- [16] WANG Fei, JIANG Mengqing, QIAN Chen, et al. Residual attention network for image classification[C]//2017 IEEE Conference on Computer Vision and Pattern Recognition. Honolulu: IEEE, 2017: 6450–6458.
- [17] MISRA D. Mish: a self regularized non-monotonic neural activation function[EB/OL]. (2020–08–13)[2021–01–01]. <https://arxiv.org/abs/1908.08681v1>.
- [18] HE Tong, ZHANG Zhi, ZHANG Hang, et al. Bag of tricks for image classification with convolutional neural networks[C]//2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Long Beach: IEEE, 2019: 558–567.
- [19] 崔珂璠, 熊淑华, 陈洪刚, 等. 基于非对称残差注意网络的目标跟踪算法 [J]. 微电子学与计算机, 2021, 38(9): 8–16.
- CUI Kefan, XIONG Shuhua, CHEN Honggang, et al. An object tracking algorithm based on asymmetric residual attention network[J]. *Microelectronics & computer*, 2021, 38(9): 8–16.
- [20] SZEGEDY C, VANHOUCKE V, IOFFE S, et al. Rethinking the inception architecture for computer vision[C]//2016 IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas: IEEE, 2016: 2818–2826.
- [21] MINAE S, MINAEI M, ABDOLRASHIDI A. Deep-emotion: facial expression recognition using attentional convolutional network[J]. *Sensors*, 2021, 21(9): 3046.
- [22] SHI Cuiping, TAN Cong, WANG Ligu. A facial expression recognition method based on a multibranch cross-connection convolutional neural network[J]. *IEEE access*, 2021, 9: 39255–39274.
- [23] JUNG H, LEE S, YIM J, et al. Joint fine-tuning in deep neural networks for facial expression recognition[C]//2015 IEEE International Conference on Computer Vision. Santiago: IEEE, 2015: 2983–2991.
- [24] YANG Huiyuan, CIFTCI U, YIN Lijun. Facial expression recognition by de-expression residue learning[C]//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City: IEEE, 2018: 2168–2177.

作者简介:



闫河, 教授, 主要研究方向为小波分析、目标跟踪、计算机视觉与视觉测量。发表学术论文 90 余篇。



李梦雪, 硕士研究生, 主要研究方向为计算机视觉、情感感知。



张宇宁, 硕士研究生, 主要研究方向为计算机视觉、语音处理。