

融合注意力机制改进 ResNet 的人脸表情识别

张栋昱, 赵磊

(武汉大学 国家网络安全学院, 湖北 武汉 430072)

摘要: 鉴于现有脸部表情识别方法在表情识别过程中存在的诸多痛点, 比如对有效特征提取不够、泛化能力不强、识别准确性不高等, 提出了一种改进残差网络的人脸表情识别方法。首先, 引入卷积注意力机制, 对网络中间的特征图进行重构, 强调重要特征, 抑制一般特征; 其次, 使用激活函数 PReLU 替换 ResNet 中原有的 ReLU, 在提高模型拟合复杂数据能力的同时, 避免出现负值区域的梯度永远为 0, 进而导致模型训练时无法执行反向传播的问题; 然后, 在网络输出层的 avgpool 与 fc 之间加入 Dropout 抑制过拟合, 以进一步增加网络模型的鲁棒性与泛化性; 最后, 在公开数据集 CK+ 上的仿真实验结果表明, 该方法的准确识别率达到 96.12%。与现有多种经典算法, 以及 baseline 算法即 ResNet101 相比, 改进的网络模型具有更好的识别效果, 证明了该方法的有效性与优越性。

关键词: 人脸表情识别; 深度学习; 残差网络; 卷积注意力机制; Dropout

中图分类号: TP391

文献标识码: A

文章编号: 1673-629X(2023)05-0130-08

doi: 10.3969/j.issn.1673-629X.2023.05.020

Improved Facial Expression Recognition in ResNet by Integrating Attention Mechanism

ZHANG Dong-yu, ZHAO Lei

(School of National Cyber Security, Wuhan University, Wuhan 430072, China)

Abstract: In view of many pain points in the process of expression recognition of the existing facial expression recognition methods, such as the extraction of effective features is not enough, generalization ability is not strong, recognition accuracy is not high, a facial expression recognition method with improved residual network is proposed. Firstly, the convolutional attention mechanism is introduced to reconstruct the feature map in the middle of the network, emphasizing important features and suppressing general features. Secondly, the activation function PReLU is used to replace the original ReLU in ResNet, which can not only improve the ability of the model to fit complex data, but also avoid the problem that the gradient in the negative area is always zero, which leads to the failure of backpropagation in model training. Then, Dropout is added between avgpool and fc in the output layer of the network to suppress overfitting, so as to further increase the robustness and generalization of the network model. Finally, the simulation experiments on the open dataset CK+ show that the accurate recognition rate of the proposed method reaches 96.12%. The improved network model has better recognition effect than the existing classical algorithms and baseline algorithm, namely ResNet101, which proves the effectiveness and excellence of the proposed method.

Key words: facial expression recognition; deep learning; residual network; convolutional attention mechanism; Dropout

0 引言

人脸表情是人们用来表达情绪的一种基本手段,更是非语言交流中的一种有效方式,人们可以通过表情准确而微妙地表达自己的思想感情,也可以通过表情辨认对方的态度和内心世界^[1]。心理学家 Mehrabian^[2]研究表明,人类的感情表露 55% 来自于面部表情,可见人脸表情在感情表露中的重要地位。近

些年来,人脸表情识别在众多领域得到了广泛的关注,并已然成为学术界和工业界的重要研究热点之一^[3]。传统的人脸表情识别算法,在对各种特征进行抽取时,通常做法是事先通过人工设计特征提取器^[4],较为常见的方法有梯度方向直方图^[5] (histogram of oriented gradient, HOG)、局部二值模式^[6] (local binary patterns, LBP) 和主成分分析法^[7] (principal component analysis,

收稿日期: 2022-07-12

修回日期: 2022-11-16

基金项目: 国家自然科学基金联合基金项目 (U1936122); 武汉市应用基础前沿项目 (2018010401011295)

作者简介: 张栋昱 (2003-), 男, 通讯作者, 研究方向为网络空间安全与深度学习; 赵磊 (1985-), 男, 博士, 教授, 博导, 研究方向为软件及系统安全、可信计算与深度学习。

PCA)等。但非常遗憾的是,在提取人脸表情特征时,现有传统方法的共性是很容易漏掉对分类识别有较大影响的关键特征,结果使得它们对人脸表情识别准确率长期以来不太理想^[4]。

受益于近几年来计算机深度学习研究工作在图像识别等相关领域取得的巨大突破,逐步使得人脸表情识别由传统方法转为基于计算机深度学习的方法。作为深度学习领域最为经典的网络模型之一的卷积神经网络^[8](convolutional neural networks, CNN),其主要优点是能够共享卷积核,对多维数据计算速度快,且具备较好的特征抽取与分类识别效果,使得该类型的网络模型在图像识别与语音处理等领域得到了较为广泛的应用^[9]。进而,在传统的卷积神经网络基础之上,人们将一些引入某些特殊改进措施的模型,比如来源于 Alex Krizhevsky 的 AlexNet^[10]、来源于牛津大学计算机视觉组(visual geometry group)的 VGGNet^[11],还有来源于谷歌的 GoogleNet^[12]等网络模型,尝试用于人脸面部表情识别任务之中。但是,随着网络深度的不断加大,学习能力的不断加强,网络模型反而出现了所谓的“退化”,即造成了梯度爆炸与消失,优化效果也越来越差等不可控问题。在此背景之下,2016 年华人学者何恺明博士等人提出了深度残差网络模型^[13](deep residual networks, ResNet),它与当时其他卷积神经网络模型最大的不同在于,引入了残差的思想,解决了网络变深之后梯度消失的尴尬难题。

为更加准确且高效地区分人脸面部表情各个不同的类型,该文提出了一种基于改进的残差网络 ResNet 的网络模型,即首先引入了卷积注意力机制,对网络中间的特征图进行重构。同时,使用激活函数 PReLU 取代 ResNet 中原有的激活函数 ReLU,避免了对于小于 0 的值梯度永远都是 0,造成网络中较多的神经元都“死”掉的问题。然后,在输出层的 avgpool 与 fc 之间加入 Dropout,进一步抑制过拟合,并增加模型鲁棒性。实验结果表明,通过采用上述改进措施,很大程度上提升了人脸表情识别的准确率,验证了该方法的有效性与优异性。

1 ResNet 概述

一般认为,加大网络模型的深度,相对来讲比加大宽度更能提升其数据拟合能力,即使得应用于图像分类任务时的准确率更高,但在实践过程中,仅单纯地通过不断增加卷积神经网络 CNN 的层数,网络模型却出现性能不增反减即“退化”的尴尬情形。该文采用残差网络 ResNet 作为人脸面部表情的特征提取网络,其增加网络层数的具体操作方法,是将恒等映射层叠加在浅层网络上,以构建残差学习单元。残差单元

Bottleneck 结构如图 1 所示,通过学习残差,将部分原始输入的特征信息,经过恒等映射层直接输出到下一层,在一定程度上减少了卷积层在进行信息传递时的特征丢失问题,可以在输入特征的基础上学习到其他新的特征,以便拥有更好的性能^[14]。实践表明,残差网络 ResNet 的主要特点是非常容易优化,并且能够通过增加相当的深度来提高准确率,得益于其内部的残差块使用了跳跃连接,很大程度上缓解了在深度神经网络中仅简单地通过增加网络层数而带来的梯度消失问题。

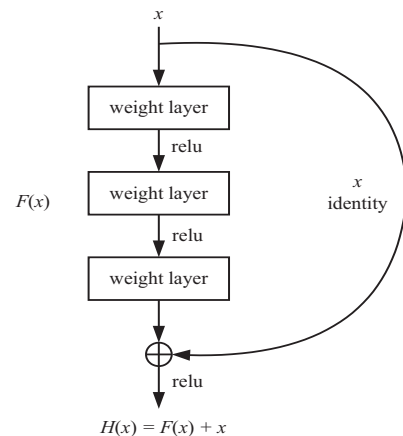


图 1 Bottleneck 残差单元

设输入为 x 时所求特征为 $H(x)$, $F(x) = H(x) - x$ 为网络学习残差,则残差单元可以表示为:

$$y_l = x_l + F(x_l, \{W_l\}) \quad (1)$$

$$x_{l+1} = f(y_l) \quad (2)$$

其中, x_l 与 y_l 分别表示第 l 层的输入和输出, W_l 为权重矩阵, $f(\cdot)$ 表示 Relu 激活函数。对于 L 层残差单元,从浅层到深层的学习特征公式为:

$$x_L = x_l + \sum_{i=l}^{L-1} F(x_i, \{W_i\}) \quad (3)$$

由于残差网络 ResNet^[15],不仅错误率较低,参数数量和计算量也较小,进而加速了模型训练的速度,使得训练模型的效果非常突出。ResNet 系列网络模型有两种类型的残差块,分别为 Buildingblock 与 Bottleneck,相对前者,后者既减少了参数量,保持了原有的精度,又优化了计算。文中的人脸表情识别网络模型选用 ResNet101,其网络结构如表 1 所示。残差块 Bottleneck 相当于积木,不算初始的 7×7 卷积层和 3×3 max pooling 层,整个网络模型由 4 个 layer 组成,而每个 layer 由若干残差块 Bottleneck 叠加而成。其中, conv2_x、conv3_x、conv4_x 与 conv5_x 分别对应 layer1、layer2、layer3 与 layer4 层。在网络输出层,通过 Average Pool 得到最终的特征,而不是通过全连接层,最后仅通过一层 Linear 层即完成了从特征到具体类别的映射过程。

表 1 ResNet101 网络结构

Layer Number	Layer Name	Output Size	Configuration	Residual Unit Type
	Conv1	112×112	7×7 conv, 64, stride 2 3×3 max pool, stride 2	
layer1	Conv2_x	56×56	$\begin{bmatrix} 1 \times 1, 64 \\ 3 \times 3, 64 \\ 1 \times 1, 256 \end{bmatrix} \times 3$	Bottleneck
layer2	Conv3_x	28×28	$\begin{bmatrix} 1 \times 1, 128 \\ 3 \times 3, 128 \\ 1 \times 1, 512 \end{bmatrix} \times 4$	Bottleneck
layer3	Conv4_x	14×14	$\begin{bmatrix} 1 \times 1, 256 \\ 3 \times 3, 256 \\ 1 \times 1, 1024 \end{bmatrix} \times 23$	Bottleneck
layer4	Conv5_x	7×7	$\begin{bmatrix} 1 \times 1, 512 \\ 3 \times 3, 512 \\ 1 \times 1, 2048 \end{bmatrix} \times 3$	Bottleneck
		1×1	average pool 1 000-d fc	

2 相关改进

2.1 融合注意力机制

神经网络中的卷积注意力机制 (convolutional attention mechanism, CAM) 已经广泛应用于多种场景, 尤其是在基于深度学习的计算机视觉领域, 比如图像分类 (image classification)、目标检测 (object detection)、目标跟踪 (target tracking)、语义分割 (semantic segmentation)、实例分割 (instance segmentation) 与全景分割 (panoptic segmentation) 等实际场合, 其有效性与实用性已经得到了广泛与充分的验证。卷积注意力主要由空间注意力和通道注意力两个方面组成, 在卷积神经网络 CNN 中, 注意力机制作用于特征图 (feature map) 上, 用于获取特征图中可用的注意力信息^[16]。如图 2 所示, 卷积注意力模块^[17] (convolutional block attention module, CBAM) 由通道注意力模块 (channel attention module, CAM) 和空间注意力模块 (spatial attention module, SAM) 组成。这两个子模块分别关注了通道信息和空间信息, 对网络中间的一系列的特征图进行再次构建, 它突出一些重要特征, 抑制其他一般特征, 可以达到增强图像识别效果的最终目的。

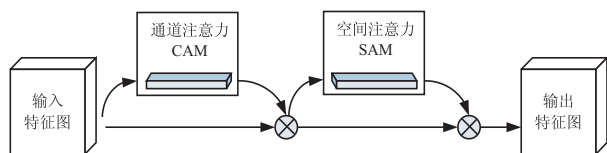


图 2 CBAM 结构

注意力重构具体过程如公式 (4) 所示, 其中 F 表示网络模型中某一网络层的特征图, $M_c(F)$ 表示使用 CAM 对 F 进行一维通道注意力重构, F' 则表示经过通道注意力重构的特征图, $M_s(F')$ 表示使用 SAM 对 F' 进行二维空间注意力重构, F'' 则表示同时结合了通道与空间两个方面注意力的输出特征图, 而 \otimes 表示逐元素乘法。针对卷积神经网络 CNN 中某一网络层的三维特征图 $F \in R^{C \times H \times W}$, 先后依次从 F 推理出一维通道注意力特征图 M_c 和二维空间注意力特征图 M_s , 并分别进行逐元素相乘, 最终得出与 F 同等维度的输出特征图 F'' 。

$$\begin{cases} F' = M_c(F) \otimes F \\ F'' = M_s(F') \otimes F' \end{cases} \quad (4)$$

卷积注意力机制最为重要的功能, 是重构网络模型中间层的特征图, 即抽取特征图中的重要信息, 并抑制其他的一般信息。对于残差网络 ResNet, 提取特征的最关键位置应该是在各个 Bottleneck。综合考虑, 该文将 CBAM 融合在各个 layer 之间, 这么做的主要原因是残差网络模型 ResNet 在各个 Bottleneck 中已经完成了特征提取, CBAM 在此处再进行注意力重构, 可以起到承上启下的作用, 融入了 CBAM 模块的 ResNet 101 结构如图 3 所示。

2.2 激活函数改进

激活函数对于神经网络去学习理解复杂的非线性的业务逻辑来说具有非常重要的作用, 它们将非线性特性引入到特定的网络模型中, 使得神经网络具有拟合非线性情形的能力, 即使得网络模型学习数据之

间更为复杂的关系成为可能。残差网络 ResNet 的隐藏层激活函数默认采用线性修正单元函数^[18] (rectified linear unit, ReLU), 其数学表达式如公式(5)

$$\text{ReLU}(x) = \begin{cases} 0, & x \leq 0 \\ x, & x > 0 \end{cases} \quad (5)$$

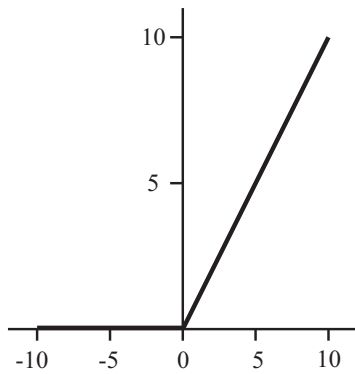


图4 ReLU 函数曲线

从图4 函数曲线可以看出, ReLU 函数非常简单且有效, 其优点是计算效率高与收敛速度快, 但最大缺点是存在神经元坏死问题, 即当输入处于区间 $[0, -\infty)$ 内, 函数的梯度变为 0, 模型无法执行反向传播, 进而导致相应权重参数 weight 与偏移 bias 永远得不到更新, 无法进行深度学习, 也即这部分神经元彻底“死”掉了。

为了解决以上问题, 该文采用带参数的线性修正单元函数^[19] (parametric rectified linear unit, PReLU) 替代残差网络 ResNet 中原有的 ReLU 激活函数, 如公式(6)所示。通常参数 a_i 相对较小, 一般位于 0 到 1 之间, 它是通过在训练中自学习的, 其对应的函数曲线如图5 所示。

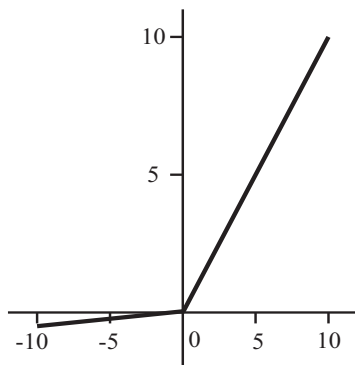


图5 PReLU 函数曲线

所示, 即当 x 大于 0 时函数值等于 x 本身, 而当 x 小于或等于 0 时其值均为 0。对应的函数曲线如图4 所示。

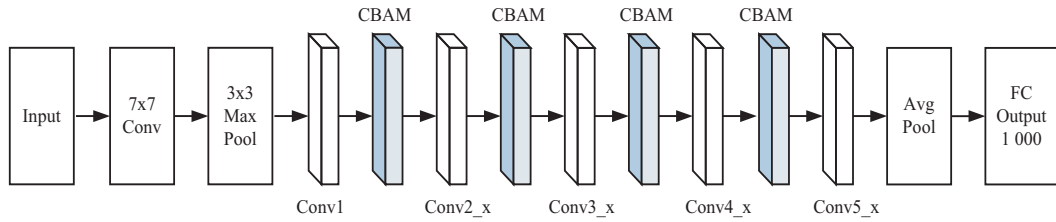


图3 ResNet 融合 CBAM

$$\text{PReLU}(x) = \begin{cases} x, & x > 0 \\ a_i x, & x \leq 0 \end{cases} \quad (6)$$

从图5 中可以看出, 在负值区间 $[0, -\infty)$ 内, PReLU 也存在一个较小的可学习的斜率, 这样既保留了在原有正值区间 $[0, \infty)$ 内 ReLU 函数的优点, 强化了模型的表达能力, 同时又能避免上述的“Dead ReLU”问题。

2.3 使用 Dropout

通常在训练网络模型时, 如果该网络模型越深, 也即卷积层与线性变换层越多, 每层的神经元数就越多, 而训练时提供的训练数据集与测试数据集的样本量又太少, 那么训练出来的模型就非常容易发生过拟合问题, 较为明显的矛盾在于, 模型在训练数据上尽管损失较小且推理准确率较高, 但是在全新的测试数据上推理的准确率却比较低, 损失也比较大。当把一个小样本的数据集提交给一个较为复杂的神经网络进行训练时, 非常容易发生过拟合, 基于此背景, Hinton^[20] 于 2012 年在论文《Improving neural networks by preventing co-adaptation of feature detectors》中首次提出了 Dropout, 其示意图如图6 所示。

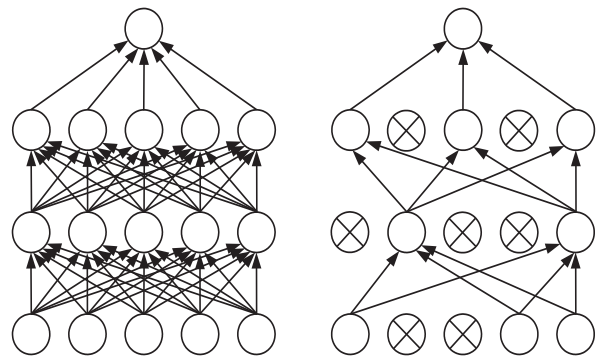


图6 Dropout 示意图

Dropout 的具体做法是在训练深度神经网络模型过程中, 按照一定的概率, 随机地将一部分神经元从网络中暂时进行逻辑隔离。对于随机梯度下降来说, 由于采用的是随机与暂时的隔离策略, 因此每一个小批次都在训练不同的网络, 但是在模型训练完成之后, 每个神经元都不会缺席, 都会参与到实际的推理工作中。

实践表明,Dropout 可以减少神经元相互之间的依赖性,较为有效地缓解网络模型过拟合现象的发生,在一定程度上可以达到降低结构风险、提高模型泛化能力的效果。

原本残差网络 ResNet 中并没有使用 Dropout,但受 Google 提出的 GoogleNet^[12] 网络模型的启发,它虽然不是通过全连接层 (fully connected layer),而是通过

Average Pooling 得到最终的特征,网络中依然使用了 Dropout,即在全连接层前面加入 Dropout 层,较为容易提升网络模型的泛化能力。基于上述考虑,为进一步防止模型过拟合,提高网络的鲁棒性,该文尝试在 ResNet 输出层的 avgpool 与 fc 之间使用 Dropout 策略,融入完 Dropout 层之后三者之间的关系如图 7 所示。

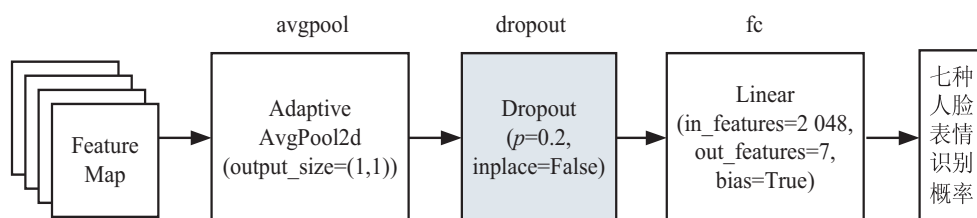


图 7 ResNet 融合 Dropout

3 实验结果与分析

3.1 整体流程

采用改进的 ResNet101 网络模型进行训练与推理过程的整体流程如图 8 所示。

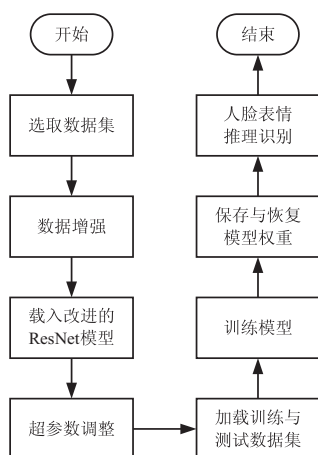


图 8 训练与推理流程

从图中可知,依次进行数据集选取,数据增强,载入改进的模型,超参数调整,然后加载数据集进行训练与验证,在此过程中保存最优权重,最后利用最优的权重文件进行人脸表情推理识别。

3.2 实验环境

关于实验环境的硬件方面,CPU 为 Intel 酷睿 i7-12700H,内存为 32 GB,硬盘为 512 GB SSD,GPU 为 GeForce RTX 3060(8 GB GDDR6 显存)。软件方面,操作系统为 Windows11 旗舰版,开发环境为 PyCharm 2022,基于 Python 3.8 与当前学术界最热门的深度学习框架之一的 Pytorch 1.11 设计、开发、训练与测试。另外,为了提升网络模型训练与测试的速度,模型与数据均加载到 GPU,并通过英伟达的统一计算设备架构 (compute unified device architecture,CUDA) 技术完成运算加速。

3.3 训练策略

优化器 (optimizer)^[21] 就是在深度学习反向传播过程中,指引损失函数的各个参数往正确的方向更新合适的大小,使得更新后的各个参数让损失函数值不断逼近全局最小。该文在训练中选用算法 Adam 优化网络模型,它可以看作是 Adaptive 与 Momentum 的结合体,是一种可以替代随机梯度下降算法的优化算法。它能基于训练数据不断迭代并更新神经网络的模型参数,并且其通常被认为对超参数选择的健壮性相当好。模型训练的超参数配置如表 2 所示。

表 2 模型训练的超参数配置

训练超参数		参数值
Epoch	训练周期	2 000
Batch_size	批尺寸	64
Optimizer	优化器	Adam
Initial learning rate	初始学习率	1×10^{-2}
Last learning rate	终止学习率	1×10^{-4}
Step_size	衰减步长	10
Gamma	衰减因子	1×10^{-2}
p	Dropout 舍弃概率	0.2

3.4 数据集

该文选用 CK+人脸表情数据集,它是数据集 Cohn-Kanade^[22] 的扩展版本,该数据集是在实验室条件下获取的,相对较为严谨可靠,是人脸表情识别中比较常用且标准的一个公开数据集。它包含 123 个对象的 327 个被标记的表情图片,共分为愤怒、蔑视、厌恶、害怕、高兴、悲伤和惊讶共七种表情,这七种表情的示例图片依次如图 9 所示。



图 9 CK+七种表情示例

另外,该数据集7种表情的标签与英文、中文的对应关系如表3所示。

表3 CK+数据集标签对应关系

标签	英文	中文
0	anger	愤怒
1	contempt	蔑视
2	disgust	厌恶
3	fear	害怕
4	happiness	高兴
5	sadness	悲伤
6	surprise	惊讶

3.5 数据增强

实际上,由于CK+数据集样本的数量并不大,在网络模型的训练学习过程中,就较为容易发生拟合现象,相应的模型泛化能力与鲁棒性也会较差。为了应对此问题,该文对样本数据进行数据增强处理。首先,将数据集所有原始图片样本统一 Resize 成 224×224 大小,然后进行随机饱和度、对比度以及明暗度等调整,同时也进行随机的缩放、旋转、偏移与反转等操作。通过此过程,使数据集的数据量增加了几十倍以上。另外,为了使实验结果避免偶然性,首先将数据集的图片随机排序,接下来对整个数据集依照留出法分割为没有交集的两个集合,并按照2:8的比例随机划分为测试集与训练集。

3.6 对比实验

为了验证该方法的有效性与优越性,下面与现有多种经典算法在CK+数据集上的识别准确率进行比较。

文献[23]将模型所提取的低、高两种层次特征相结合构造分类器,并基于跨连接网络进行人脸表情识别。

文献[24]首先采用加博尔滤波器提取特征,然后用主成分分析法降低特征图维数,最后用卷积神经网络进行人脸表情识别。

文献[25]首先在常规卷积池化操作基础上引入四层 Inception 结构,然后用跨数据库的方式来评估网络模型。

文献[26]提出了一个具有两个并行卷积池化单元的卷积神经网络分别提取不同的人脸特征,并将这些特征进行融合。

文献[27]提出了融合尺度不变特征变换(scale-invariant feature transform, SIFT)和卷积神经网络所提取的特征,并用支持向量机(support vector machine, SVM)进行分类。

文献[28]首先对密集的人脸标志点进行精确定

位,然后将以标志点为中心的斑点中提取所有描述子串联起来,形成高维特征,最后基于深度稀疏自编码器对高维特征进行有效提取。

将文中网络模型与上述算法在CK+数据集上的表现进行对比,详细结果如表4所示。相比上述其他方法,文中模型对人脸表情识别更有针对性,它强化了重要特征,抑制了一般特征,取得了更好的实验效果。

表4 多种经典算法在CK+数据集上的识别率比较

算法	识别率/%
Cross-connect LeNet-5 ^[23]	83.74
AlexNet	87.03
DNN ^[24]	91.70
Mollahosseini ^[25]	93.20
Parallel CNN ^[26]	94.03
Hybrid CNN-SIFT features ^[27]	94.82
Deep sparse autoencoders ^[28]	95.79
文中方法	96.12

3.7 消融实验

按照 Batch_size 大小,按批依次读入数据,在经过2000个epoch迭代之后,模型在训练集与测试集上的loss均接近于0,在训练集上的正确率接近于99.81%,而在测试集上的正确率也接近于96.32%,各方法在CK+人脸数据集上的识别结果如表5所示。

表5 文中模型与baseline的对比

方法	识别率/%
(Baseline) ResNet101	94.41
ResNet101+Dropout	94.56
ResNet101+PReLU	94.85
ResNet101+CBAM	95.73
(Proposed) ResNet101+CBAM+PReLU+Dropout	96.12

把 ResNet101 作为 baseline,依次单独分别融入 Dropout, PReLU 与 CBAM 时,相对原始 baseline,识别率都分别有所提升,尤其是融入了卷积注意力 CBAM 模块后,提高的百分比更为明显。最后,将这三个措施同时作用于改进 ResNet101 时,识别率直接提升到了96.12%,改善效果较为明显。可见,所设计的基于改进 ResNet 的方法对人脸表情识别任务较为有效。

接下来,利用混淆矩阵评估文中模型对人脸面部表情各个类型在识别过程中的混淆程度。通常一般认为,如果混淆矩阵的对角线上的数值越高,说明模型的识别混淆程度越低;相反,如果混淆矩阵反对角线上的数值越高,则说明模型的识别混淆程度越高。对文中模型的表情识别情况,表6是经过周密试验后进行整理的混淆矩阵的详细结果数据。

表 6 CK+数据集混淆矩阵 %

	anger	contempt	disgust	fear	happiness	sadness	surprise
anger	97.51	0	3.06	0	0	2.42	0
contempt	1.72	95.78	0	0	0	2.31	0
disgust	1.82	0	95.71	0.31	0	1.47	0
fear	3.32	0	0.08	93.89	0	2.72	0
happiness	0	0	0	0	99.82	0	0.30
sadness	4.54	2.21	0	0	0	92.89	0
surprise	0	0	0	1.42	0	0	99.53

从表 6 统计结果可以得出,文中网络模型对恐惧 fear 和悲伤 sadness 两种表情的识别效果相对较差,因为恐惧和悲伤均具备拉开嘴唇和紧张前额的特征,这两者之间具有较大的相似性,相对来讲更容易发生错误分类,而对于愤怒 anger、惊讶 surprise 与高兴 happiness 这三种面部表情的识别准确率最高,分别达到了 97.51%、99.53% 与 99.82%,这是因为这三种表情的面部特征均较为夸张,相对来讲面部特征更加容易被提取。但是,从整体上来讲,文中网络模型对上述七种人脸面部表情类型的识别分类具有较低的混淆效果。

4 结束语

针对人脸表情识别问题,提出了一种改进残差网络 ResNet101 的人脸表情识别方法。首先,为了精确提取面部表情特征,融合了卷积注意力机制 CBAM,以重构网络中间的特征图,目的是突出重要特征,抑制其他一般特征,进而提升模型对有用特征的抽取能力。其次,使用带参数的线性修正单元激活函数 PReLU 替代原有的 ReLU 函数,在提高网络模型表达能力的同时避免出现负值区域的函数的梯度变为 0,进而使模型无法执行反向传播的问题。最后,在网络输出层的 average pool 层与 fc 层之间加入 Dropout,以增加模型的鲁棒性与泛化性。实验结果表明,与现有多种经典算法,以及 baseline 算法即 ResNet101 相比,改进的网络模型具备更好的识别效果,证明了该方法的有效性与优越性。后续,将进一步研究复杂环境下的人脸表情识别,以不断增强网络模型的特征表达与学习能力。

参考文献:

- [1] 刘晓旻,谭华春,章毓晋. 人脸表情识别研究的新进展[J]. 中国图象图形学报,2006,11(10):1359-1368.
- [2] MEHRABIAN A. Communication without words[J]. Psychology Today,1968,2(4):53-56.
- [3] 李振东,言有三. 人脸表情识别研究[EB/OL]. [2021]. https://blog.51cto.com/u_14122493/3338225.
- [4] 张鹏,孔韦韦,滕金保. 基于多尺度特征注意力机制的人脸表情识别[J]. 计算机工程与应用,2022,58(1):182-189.
- [5] DALAL N,TRIGGS B. Histograms of oriented gradients for human detection[C]//Proceedings of 2005 IEEE computer society conference on computer vision and pattern recognition. San Diego:IEEE,2005:886-893.
- [6] 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.
- [7] WOLD S,ESBENSEN K,GELADI P. Principal component analysis[J]. Chemometrics and Intelligent Laboratory Systems,1987,2(1/3):37-52.
- [8] WANG C Y,LIAO H Y M,WU Y H,et al. CSPNet:a new backbone that can enhance learning capability of CNN[C]//Proceedings of the 2020 IEEE/CVF conference on computer vision and pattern recognition workshops (CVPRW). Seattle:IEEE,2020:1571-1580.
- [9] 周婕,马明栋. 基于改进的 ResNet 网络的人脸表情识别[J]. 计算机技术与发展,2022,32(1):25-29.
- [10] 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.
- [11] ZENG N Y,ZHANG H,SONG B Y,et al. Facial expression recognition via learning deep sparse autoencoders[J]. Neurocomputing,2018,273:643-649.
- [12] SZEGEDY C,WEI L,JIA Y,et al. Going deeper with convolutions[C]//IEEE conference on computer vision and pattern recognition (CVPR). Boston:IEEE,2015:1-9.
- [13] HE K,ZHANG X,REN S,et al. Deep residual learning for image recognition[C]//IEEE conference on computer vision and pattern recognition (CVPR). Las Vegas:IEEE,2016:770-778.
- [14] 陈娅娅,刘全香,王凯丽,等. 基于 ResNet 和迁移学习的古印章文本识别[J]. 计算机工程与应用,2022,58(10):125-131.
- [15] HE K,ZHANG X,REN S,et al. Identity mappings in deep residual networks[C]//14th European conference on computer vision. Amsterdam:Springer,2016.

- [16] 张宸嘉,朱磊,俞璐. 卷积神经网络中的注意力机制综述[J]. 计算机工程与应用,2021,57(20):64-72.
- [17] WOO S, PARK J, LEE J Y, et al. CBAM: convolutional block attention module [C]//Proceedings of the European conference on computer vision (ECCV). Munich: Springer International Publishing,2018:3-19.
- [18] 王红霞,周家奇,辜承昊,等. 用于图像分类的卷积神经网络中激活函数的设计[J]. 浙江大学学报:工学版,2019,53(7):1363-1373.
- [19] 田娟,李英祥,李彤岩. 激活函数在卷积神经网络中的对比研究[J]. 计算机系统应用,2018,28(7):43-49.
- [20] HINTON G E, SRIVASTAVA N, KRIZHEVSKY A, et al. Improving neural networks by preventing co-adaptation of feature detectors [J]. Computer Science,2012,3(4):212-223.
- [21] 杰奏. 优化器 (Optimizer) [EB/OL]. [2021]. <https://zhuanlan.zhihu.com/p/261695487>.
- [22] 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]//Proceedings of 2010 IEEE computer society conference on computer vision and pattern recognition - workshops. San Francisco: IEEE, 2010:94-101.
- [23] 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.
- [24] LI J N, LAM E Y. Facial expression recognition using deep neural networks [C]//Proceedings of 2015 IEEE international conference on imaging systems and techniques (IST). Macau: IEEE,2015:1-6.
- [25] MOLLAHOSSEINI A, CHAN D, MAHOOR M H. Going deeper in facial expression recognition using deep neural networks [C]//2016 IEEE winter conference on applications of computer vision (WACV). Lake Placid: IEEE,2016:1-10.
- [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,2019,24(2):227-236.
- [27] 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.
- [28] ZENG N, ZHANG H, SONG B, et al. Facial expression recognition via learning deep sparse autoencoders [J]. Neurocomputing,2018,273:643-649.
- +++++
- (上接第80页)
- (ABC) algorithm [J]. Journal of Global Optimization,2007,39(3):459-471.
- [14] JONES D R, PERTTUNEN C D, STUCKMAN B E. Lipschitzian optimization without the Lipschitz constant [J]. Journal of Optimization Theory and Applications,1993,79(1):157-181.
- [15] LIU Q, CHENG W. A modified DIRECT algorithm with bi-level partition [J]. Journal of Global Optimization,2014,60(3):483-499.
- [16] LIU Q. Linear scaling and the DIRECT algorithm [J]. Journal of Global Optimization,2013,56(3):1233-1245.
- [17] BANDYOPADHYAY R, KUNDU R, OLIVA D, et al. Segmentation of brain MRI using an altruistic Harris Hawks' optimization algorithm [J]. Knowledge-Based Systems,2021,232:107468.
- [18] ARBELAEZ P, FOWLKES C, MARTIN D. The Berkeley segmentation dataset and benchmark [EB/OL]. (2007-06) [2022-07]. <https://www2.eecs.berkeley.edu/Research/Projects/CS/vision/bsds/>.
- [19] ZHANG L, ZHANG L, MOU X, et al. FSIM: a feature similarity index for image quality assessment [J]. IEEE Transactions on Image Processing,2011,20(8):2378-2386.