改进残差网络与峰值帧的微表情识别

任 宇, 陈新泉, 王岱嵘, 陈新怡
安徽工程大学 计算机与信息学院, 安徽 芜湖 241000

摘 要:目的 微表情(Micro Expression, ME)是人们流露内心情感时展现出的细微面部表情。针对微表情识别的样本较少且不同类别数量分布不均导致难以识别和识别准确率较低的问题,提出能够提高微表情识别准确率的模型框架。**方法** 提取微表情视频序列中含有更多关键表情信息的峰值帧;使用加入 SE 模块的改进残差网络 SE-ResNeXt-50 对微表情的峰值帧进行特征提取,其中 SE 模块可以更好地学习特征中的关键信息,ResNeXt 通过分组卷积的方式用稀疏结构取代密集结构从而使结构更加简化,提升了识别效率。与此同时,使用 Focal Loss 损失函数可以更好地解决因微表情数据的不平衡带来的模型性能问题。**结果** 在微表情数据集 CASME Ⅱ 上进行了仿真实验,可以发现改进的残差网络与峰值帧提高了微表情识别的准确率与 *F*1 值。**结论** 改进的残差网络与峰值帧可以降低数据集较少所带来的影响,使模型有着良好的拟合效果,同时改善了在不同类别上表现差异较大的问题,提升了微表情的识别准确率,对于微表情识别有着更好的识别性能。

关键词:微表情识别; 残差网络; 峰值帧; 深度学习
中图分类号:O643 **文献标识码:**A doi:10.16055/j.issn.1672-058X.2024.0001.003

Micro-expression Recognition Based on Improved Residual Network and Apex Frame

REN Yu, CHEN Xinquan, WANG Dairong, CHEN Xinyi
School of Computer and Information, Anhui Polytechnic University, Anhui Wuhu 241000, China

Abstract: Objective Micro-expression (ME) is the subtle facial expression that reveals one's inner emotions. The number of samples for micro-expression recognition is small and the number of different categories is uneven, leading to difficulty in recognition and low recognition accuracy. In view of this, a model framework that can improve the accuracy of micro-expression recognition was proposed. **Methods** Peak frames containing more key expression information were extracted from the micro-expression video sequences. An improved residual network, SE-ResNeXt-50, incorporating the SE module was used to extract features from the apex frames of micro-expressions. The SE module learned the key information in the features better. ResNeXt simplified the structure by replacing the dense structure with a sparse one by means of group convolution, thus improving the recognition efficiency. At the same time, the Focal Loss function was used to better solve the model performance problems caused by the imbalance of micro-expression data. **Results** Simulation experiments were conducted on the micro-expression dataset CASME II, and it was found that the improved residual network and apex frames improved the accuracy and F1 value of micro-expression recognition. **Conclusion** The improved residual network and apex frames can reduce the impact caused by fewer data sets, so that the model has a good fitting effect. At the same time, it can mitigate the impact caused by the performance differences in different categories, improve

收稿日期:2022-12-13 **修回日期:**2023-02-21 **文章编号:**1672-058X(2024)01-0021-09
基金项目:安徽省自然科学基金项目(2108085MF213);安徽省高校自然科学研究项目(KJ2021A0516);国家自然科学基金面上项目(61976005);国家级大学生创新创业项目(202110363102,202210363094).
作者简介:任宇(1995—), 男, 安徽合肥人, 硕士研究生, 从事神经网络、机器学习研究.
通讯作者:陈新泉(1974—), 男, 湖南安仁人, 教授, 博士, 从事数据挖掘、机器学习研究. Email: chenxqscut@126.com
引用格式:任宇, 陈新泉, 王岱嵘, 等. 改进残差网络与峰值帧的微表情识别[J]. 重庆工商大学学报(自然科学版),2024,41(1):21—29.
REN Yu, CHEN Xinquan, WANG Dairong, et al. Micro-expression recognition based on improved residual network and apex frame[J]. Journal of Chongqing Technology and Business University (Natural Science Edition), 2024, 41(1): 21—29.

the accuracy of micro-expression recognition, and have better recognition performance for micro-expression recognition.
Keywords: micro-expression recognition; residual network; apex frame; deep learning

1 引 言

微表情是人们内心情感通过不受控制的面部活动露出的面部表情。这种自发产生的面部表情因其不可控性而难以克制,可以将内心的情感更直观地表达出来。微表情通常发生在1/25~1/3 s内,且在幅度很小的特定面部区域运动[1]。因此在生产生活中微表情识别被广泛应用于评估情绪状态、谎言检测、商务谈判等领域[2]。

最初的微表情识别是由有经验或受过专业训练的心理学家和专业人士完成,但是由于时间表现上的压缩以及理解上的困难[3],识别效率与准确度较低。随着机器学习与深度学习等方法的发展,现有的识别方法可分为基于手工描述特征的微表情识别算法与基于深度学习的微表情识别算法[4]。目前,微表情识别步骤大致可分为预处理、特征提取、分类。

较早提出的是基于手工描述特征的微表情识别方法。局部二进制模式(Local Binary Pattern, LBP)是基于方向梯度直方图[5]算法早期提出的手工描述特征的微情识别算法。Li等[6]提出了三正交平面局部二进制模式(LBP from Three Orthogonal Planes, LBP-TOP)进行微表情识别。Huang[7]提出了时空全局部量化模式(Spatial Temporal Completed Local Quantized Pattern, STCLQP),有效地避免了LBP-TOP特征仅考虑局部外观和运动特征的局限性。Wang等提出了六交点局部二值模式(LBP with Six Intersection Points, LBP-SIP)方法[8]和平均正交平面二进制模式(LBP Mean Orthogonal Planes, LBP-MOP)方法[9]。Xu等[10]提出面部动态图(Facial Dynamics Map, FDM)方法,Liu等[11]提出计算面部肌肉运动的主方向平均光流,这两种都是基于运动信息的光流特征来进行识别。FDM通过将人脸进行定位,之后将每个微表情进行人脸对齐和剪裁,再把抽取出来的光流场进行分割,最后抽取分割出的立方体主方向。上述基于手工描述特征的微表情识别算法可以大致分为外观特征和几何特征:LBP及其衍生的方法是较为典型的外观特征方法,而利用光流特征或是动态特征则是几何特征的方法。基于这两种特征的算法,可以将微表情内在特征更加直观地表现出来,同时对一些微小的特征进行放大,从而更加容易地进行微表情识别。但是由于其方法的局限性,而导致需要大量的计算和繁琐的参数调整,同时泛化能力与鲁棒性都比较低,难以处理复杂情况下的微表情识别,降低识别准确率[12]。

随着神经网络的发展,深度学习的方法在图像分类、人脸识别、语音识别[13-14]等方向取得了巨大的突破,基于深度学习的微表情识别方法也开始逐渐被提出。Patel等[15]使用深度学习方法,利用卷积神经网络用深层特征方式对微表情特征进行提取,减少了特征信息的冗余度,再使用传统的分类方法进行表情的分类。Peng等[16]提出了双时域尺度卷积神经网络(Dual Temporal Scale Convolutional Neural Network, DTSCNN),每条流都由独立的浅网络组成,同时注入光流序列获取更高层次的特征避免过拟合问题。Khor等[17]提出了一个增强的长期递归卷积网络(Enriched Long-term Recurrent Convolutional Network, ELRCN),该网络结合卷积神经网络和长短期记忆网络分别提取空间信息和时空信息。Cai等[18]将DenseNet网络用于微表情的识别,DenseNet以其独特的连接方式可以缓解常见的梯度消失的问题,从而提高了准确率。对比手工描述特征的方法,深度学习的方法可以更好地专注于微表情特征的提取,从不同的角度提取视频序列的信息,最后将这些信息用于识别任务,提高了模型识别的泛化能力与鲁棒性。

以上的研究方法都是将微表情视频片段从起始帧到终止帧所有帧送入模型进行训练与识别。微表情视频序列一般包含微表情开始的起始帧(Onset frame)和结束的终止帧(Offset frame),在起始帧与终止帧中间变化最为明显的一帧称为峰值帧(Apex frame)[19]。但是除了峰值帧之外,其余帧包含了较多无用信息,而使用峰值帧进行微表情的识别可以降低无用信息的干扰。因此本文首先对峰值帧进行提取,其次将改进的残差网络ResNeXt-50[20]模型用于微表情识别,可以使结构更加简化,减少参数数量,提升了识别效率,同时使用SE模块可以更好地学习特征中的关键信息,抑制无用信息,形成用于微表情识别的SE-ResNeXt-50模型[21]。最后在CASME Ⅱ微表情数据集上完成模型训练与实验,实验结果表明,本文提出的方法与其他的微表情识别方法相比,识别准确率与F1值得到提升,可以取得较好的识别效果。

2 模型结构框架

本节将介绍残差网络的残差模块与改进的残差模块、SE模块相关内容及其最后使用的整体模型结构框架。

2.1　残差模块

为了缓解在深度神经网络中增加深度带来的梯度消失问题,同时用更深更宽的网络框架提取到更加丰富的特征信息和语义信息,He 等[22]在模型中加入残差模块从而提出了 ResNet 模型。所提出的残差模块可以定义为如式(1)所示:

$$x_{i+1}=F(x,W_i)+x_i \tag{1}$$

式(1)中,$F(x,W_i)$ 为网络中的残差映射;x_{i+1} 为残差模块的输出;x_i 为残差模块的输入。

残差模块引入了一个恒等映射,其内部的残差块使用了跳跃连接, ReLU 函数作为激活函数。原本的恒等映射 $H(X)=X$ 直接去拟合是一件比较困难的事,但是如果将原本 $H(X)=X$ 转换成为如图 1 所示的 $H(X)=F(X)+X$,当 $F(X)=0$ 时就形成一个残差的恒等结构,用这种残差形式去拟合会更加容易。残差模块的加入可以避免梯度消失问题,进一步提高模型的拟合能力,减轻网络层数增加带来的影响。

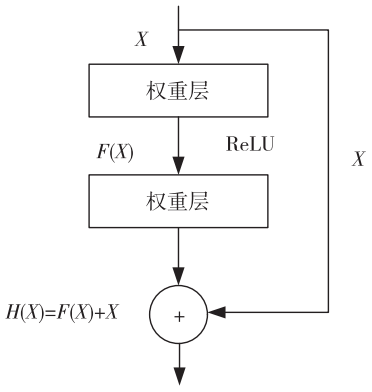
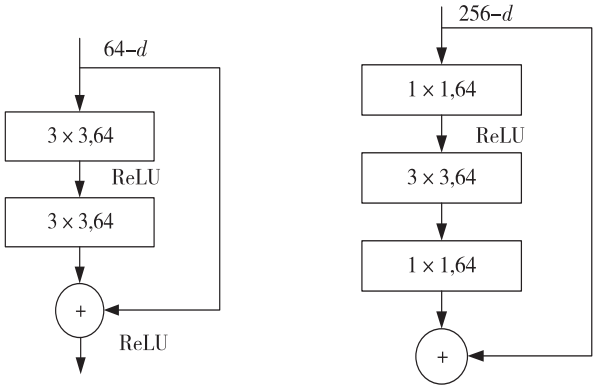


图 1　残差模块结构

Fig. 1　Structure of the residual module

ResNet 模型对残差模块的结构进行了优化,降低因为深层次的网络模型而带来的计算成本相对较大的大问题。将原结构中 2 个 3×3 的卷积层替换为图 2(b)所示的只保留一个 3×3 卷积层,但是在这个 3×3 卷积层前后各添加一个 1×1 的卷积层分别进行降维操作与升维还原的操作的新结构,同时在三个卷积层中间也将使用线性整流单元作为激活函数。这种优化以后的残差结构相比较未优化的残差结构,既减少了参数量降低了计算成本也保持了模型精度。

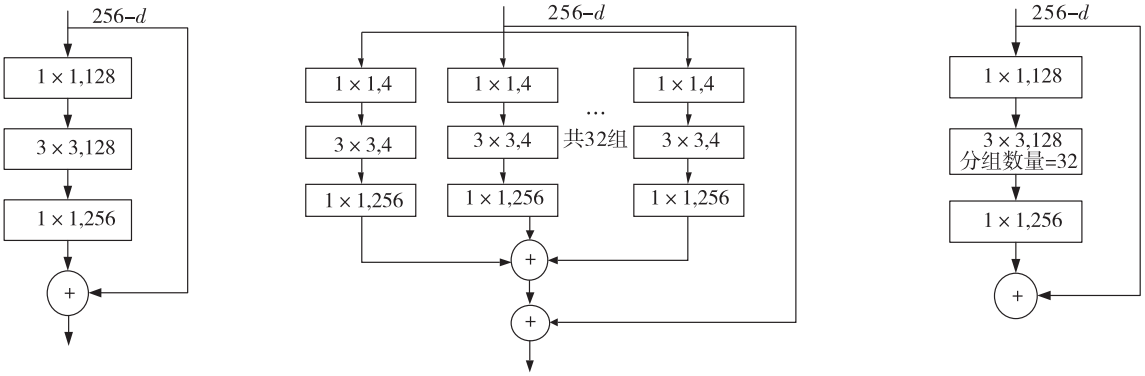


(a) 残差模块结构　　**(b) 优化后残差模块结构**

图 2　残差模块的优化

Fig. 2　Optimization of the residual module

ResNeXt 模型借鉴了 GoogLeNet 模型[23]中提出的 Inception 多尺度处理模块的思想对残差模块进行了改进,将原本的卷积结构进行分组,用稀疏结构取代密集结构,这样可以融合不同尺度的信息,避免因提高模型性能而加大模型的规模所带来的过拟合与大量参数计算浪费计算资源的问题。与 Inception 模块不同,ResNeXt 分组卷积层使用了相同结构,将原本图 3(a)中的卷积层进行分组,分解成 32 组,形成图 3(b)中的网络结构,之后再融合各个组的结果。这种使用相同分组结构的卷积层使网络设计更加的简化,从而可以避免参数的迅速膨胀。同时 ResNeXt 网络引入了表示残差结构中卷积层的分组数量的新超参数“Cardinality”。图 3(c)所示的是分组卷积的简易表达形式,如图 3(c)中使用的是分组数量为 32 个输入输出维度为 4 维的 3×3 卷积层。



(a) 普通模块　　**(b) 分组卷积**　　**(c) 简易形式**

图 3　ResNeXt 网络残差模块

Fig. 3　ResNeXt network residual module

2.2 SE 模块

SE(Squeeze and Excitation)模块是一种从特征的通道维度上面考虑的注意力机制。SE 模块由两部分组成,压缩(Squeeze)部分和激励(Excitation)部分,每个卷积操作实际上是在输入的空间维度和通道维度上面进行的乘加操作。SE 模块的大体结构如图 4 所示。

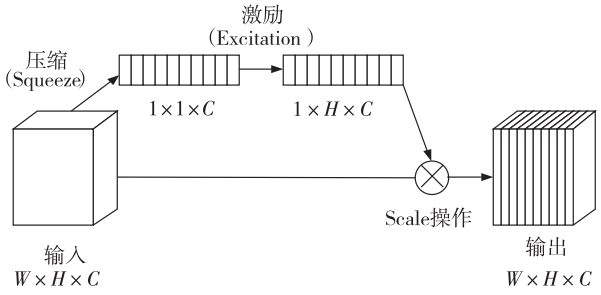


图 4 SE 模块结构

Fig. 4 SE module structure

Squeeze 部分通过一个全局平均滤波实现全局信息的获取,如式(2)所示:

$$z_c = F_{sq}(u_c) = \frac{1}{H \times W}\sum_{i=1}^{H}\sum_{i=1}^{W} u_c(i,j) \tag{2}$$

其中,u_c 为由输入 X 经过分离卷积 $\boldsymbol{U}=[u_1,u_2,\cdots,u_C]$ 得到的特征中第 c 个特征图,C 为通道数;H 与 W 是 $\boldsymbol{U}$ 的空间维度,$F_{sq}(\cdot)$ 为压缩操作。

Excitation 部分首先通过一个全连接层将特征压缩到$\frac{C}{r}$通道,然后使用 ReLU 层进行非线性操作,接着使用全连接层将特征还原至 C 通道,其中 r 为压缩比,最后使用一个 Sigmoid 函数激活,如式(3)所示:

$$s=F_{ex}(z,W)=\sigma(g(z,W))=\sigma(W_2\delta(W_2z)) \tag{3}$$

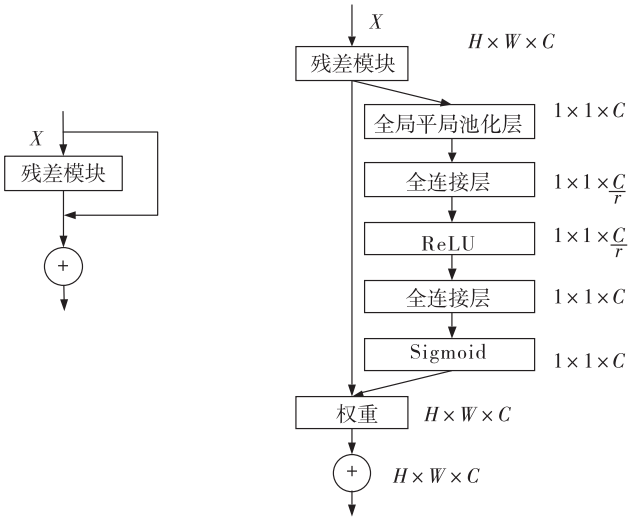
其中,σ 为 Sigmoid 激活函数;δ 为 ReLU 线性激活函数;$F_{ex}(\cdot,W)$ 为激励操作;W_1 与 W_2 是两个全连接层参数。

Squeeze 部分和 Excitation 部分进行完以后,之后进行 scale 操作。将 SE 模块计算出来的各通道权重值 S 分别和原特征图对应通道相乘,得出的结果进行输出,如公式(4)所示:

$$\tilde{x}_c = F_{scale}(u_c,s_c) = s_c u_c \tag{4}$$

其中,$F_{scale}(\cdot,\cdot)$ 为 scale 操作;$\tilde{\boldsymbol{X}}=[\tilde{x}_1,\tilde{x}_2,\cdots,\tilde{x}_C]$ 是每个通道进行 scale 操作后得到结果的集合。

SE block 可以嵌入到目前提出的所有经典的网络结构中,实现模型的改造。将 SE 模块加入 ResNeXt 模型中,如图 5 所示。对比原始的残差模块结构,加入了 SE 模块新的残差结构中为了计算每个通道权重值 s,在残差结构的后面添加了一条新路径。在新的路径中为了进行每个通道中的 Excitation 操作,在残差模块之后加入了全局平均池化层进行 Squeeze 操作,在第一个全连接层中加入降维压缩操作,然后使用 ReLU 激活函数,之后再加入一个全连接层进行还原后使用 Sigmoid 激活函数,最后将每个通道的信息进行权重融合计算。把 SE 模块加入到残差模块后可以更好的将一些没有用的信息进行抑制,突出特征中信息量大和需要注意到的特征,从而提高模型最后的识别精度。



(a) 原始结构 (b) SE 残差结构

图 5 SE 残差模块与原始残差模块对比

Fig. 5 Comparison of SE residual module and original residual module

2.3 整体模型结构框架

本文使用 SE-ResNeXt-50 模型,即在 ResNet 的改进网络模型 ResNeXt 上加入 SE 模块。整个网络由卷积层、池化层、残差模块和全连接层构成,模型结构如表 1 所示。

表 1 SE-ResNeXt-50 架构

Table 1 SE-ResNeXt-50 architecture

层	输出尺寸	网络结构
Conv1	112×112	Conv, 7×7, 64, Stride 2 Max pool, 3×3, Stride 2
Conv2_x	56×56	$\begin{pmatrix}1\times1,128\\3\times3,128\ C=32\\1\times1,256\\f_c,[16,256]\end{pmatrix}\times3$
Conv3_x	28×28	$\begin{pmatrix}1\times1,256\\3\times3,256\ C=32\\1\times1,512\\f_c,[32,512]\end{pmatrix}\times3$
Conv4_x	14×14	$\begin{pmatrix}1\times1,512\\3\times3,512\ C=32\\1\times1,1024\\f_c,[64,1024]\end{pmatrix}\times3$
Conv5_x	7×7	$\begin{pmatrix}1\times1,1024\\3\times3,1024\ C=32\\1\times1,2048\\f_c,[128,2048]\end{pmatrix}\times3$
	1×1	5 维全连接层,Softmax

表 1 展示的网络结构中,残差模块采用的是 32×4 分组卷积结构,括号外部参数表示残差模块重复堆叠的数量,括号内是残差模块的构建参数和结构。为了适用微表情识别的任务,将原本在原始网络架构中经过全局平均池化层输出,全连接层输出维度为 5,将 Softmax 函数用于最后分类。同时为了更好地平衡精确度与复杂度之间的关系,将其中 SE 模块的降维压缩比例 r 值设置为 16。

3　实验及结果分析

3.1　实验数据集

本文实验使用 CASME Ⅱ数据集[24]。CASME Ⅱ数据集是中国科学院心理研究所创建的自发式微表情数据库。CASME Ⅱ数据集使用 200 fps 的高速摄像机拍摄,分辨率为 280 像素×340 像素。该数据集收集了平均年龄在 22.03 岁左右亚洲人脸的微表情数据,这些表情都是在一个控制良好的实验室环境下激发出来的。CASME Ⅱ要求 26 个受试者观看基于厌恶、高兴、压抑、惊讶、悲伤、恐惧及其他这 7 种基本情感的短片进行情感诱发,最后得到共包括这些基本情感的 255 个微表情样本,其中包括其他类型 99 个样本、厌恶类型 63 个样本、高兴类型 32 个样本、惊讶类型 28 个样本、压抑类型 27 个样本、悲伤与恐惧类型分别 4 个和 2 个样本。如表 2 所示,在一些 3 分类实验中,会将数据集统一分为消极,积极,惊讶 3 种类型,其中厌恶、压抑、悲伤、恐惧统一归为消极样本,高兴为积极样本,惊讶保留为原有的惊讶样本。在进行 5 分类实验中,由于悲伤和恐惧两种类型的样本数量过少,不能更好地训练其中的特征,因此本实验仅使用其余 5 类共 249 个样本实验。

表 2　CASME Ⅱ数据集分类与数量

Table 2　Classification and quantity of CASME Ⅱ dataset

3 分类	消　极				积　极	惊　讶	
原始	厌恶	压抑	悲伤	恐惧	高兴	惊讶	其他
5 分类	厌恶	压抑			高兴	惊讶	其他
样本数	63	27	4	2	32	28	99

3.2　数据集预处理

数据集的预处理包括峰值帧的提取和人脸的对齐剪裁。微表情序列是从表情的起始帧开始,到达具有最大面部信息的峰值帧,最后在终止帧结束。峰值帧的提取使用的是 Quang 等[25]提出的峰值帧查找算法。该算法使用开源工具包根据面部 68 个特征点,定位 10 个微表情肌肉移动发生频繁的区域。如图 6 所示。

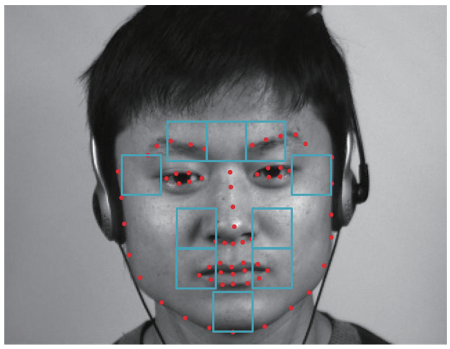


图 6　定位点与特征区域

Fig. 6　Anchor points and feature areas

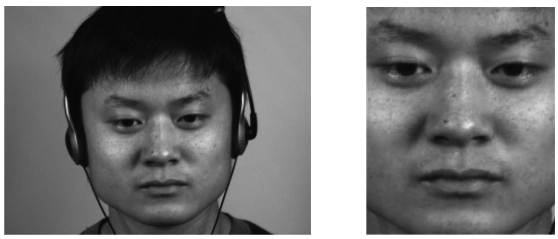
为了确定序列中的峰值帧,计算当前帧与在这 10 个区域中的起始帧和偏移帧之间的绝对像素差。

同时将两个差值的和除以所考虑的帧与其连续帧之间的差值进行归一化,从而减少环境噪声。之后可以得到微表情序列中每一帧的每像素平均值 M_i,M_i 如式(5)、式(6)所示:

$$f(frame_i, frame_j) = \frac{|frame_i - frame_j| + 1}{|frame_i - frame_{i-3}| + 1} \tag{5}$$

$$M_i = f(frame_i, frame_{oneset}) + f(frame_i, frame_{offset}) \tag{6}$$

最后遍历计算对应帧的变化强度 M_i,取值最大的一帧作为 Apex 帧。将峰值帧提取出之后,对峰值帧进行对齐裁剪,同样使用开源的工具包对人脸 68 个关键特征点检测并进行标注,最后裁剪掉没有意义的背景部分,去掉背景的干扰信息,通过人脸对齐裁剪后使之更符合微表情识别的需求,如图 7 所示。



(a) 原始峰值帧　　(b) 面部剪裁后峰值帧

图 7　剪裁前后峰值帧对比

Fig. 7　Comparison of apex frames before and after cropping

3.3　数据增强

数据增强可以运用在深度学习的各个领域。当进行实验时出现数据量较小的问题,可以使用数据增强来增加样本的数量,让模型泛化和拟合的能力得到提高,从而增加模型的性能,同时这些变化并不会影响神

经网络对特征的提取。本文使用数据增强方法包括灰度化、水平翻转、随机旋转。详细来说,灰度化就是将彩色图片变为灰度图片,随机旋转是将图片在15°内的小角度之间随机旋转,水平翻转是指将图片翻转为镜像图片。

3.4 实验设置

为了验证SE-ResNeXt-50模型在针对峰值帧微表情识别的有效性和适用性,本实验在Windows10系统环境下,模型的训练和测试均在深度学习框架PyTorch下完成神经网络的模型搭建。实验硬件为Intel Core i7-6700HQ,内存16 GB,显卡型号为NVDIA GeForce GTX 960。实验的软件环境为Python 3.7;NVDIA CUDA框架10.1;cuDNN 7.6库。

实验中各项设置如下:使用Adam方法;学习率设为0.000 1;训练时批处理数量为2,epoch为100,训练测试过程中损失函数为Focal Loss函数[26]。Focal Loss函数可以更好地解决图像领域数据不平衡造成的模型性能问题。

3.5 对比实验结果分析

本文采用留一交叉验证法作为微表情识别的评估方法。每次用1名受试者的微表情样本作为测试集,其余为训练集,最后将所有的测试结果合并作为最终的实验结果[27]。这样有利于在小样本的微表情中经行验证,可以最大程度保证训练和评估模型的客观性不受个体影响。

通过准确率(Accuracy)和$F1$值作为评估模型识别效果的指标,并与其他算法的结果进行对比。准确率和$F1$值可以公平地客观地评估模型在微表情识别中的表现。

准确率如式(7)所示:

$$U_{\text{Accuracy}}=\frac{T}{N} \tag{7}$$

其中,N为样本的总数量;T为样本中预测正确的数量。

$F1$值使用的是未加权$F1$值。未加权$F1$值在多分类评估中是一个很好的评判标准,它不会受类别数量所影响,对于每个类别都平等对待。$F1$值如式(8)所示:

$$F1=\frac{1}{K}\sum_{k=1}^{K}\frac{2TP_k}{2TP_k+FP_k+FN_k} \tag{8}$$

其中,TP_k、FP_k、FN_k分别为类别k中真正、假正和假负的数量,对K个类别的比值求平均得到未加权$F1$值。

使用本文提出的网络结构SE-ResNeXt-50模型,在CASME Ⅱ数据集中得到的峰值帧上进行5分类实验。得到的准确率与未加权$F1$值如表3所示,对比其他几种微表情识别的算法在CASME II数据集上进行5分类实验的实验结果,可以发现本文提出的网络结构实验结果要优于其他现有方法。

表3 结果比较

Table 3 Comparison of results

方 法	U_{Accuracy}	$F1$
LBP-TOP+AdaBoost[28]	0.437 8	0.333 7
LBP-SIP[8]	0.465 6	0.448 0
ELRCN[17]	0.524 4	0.500 0
STCLQP[29]	0.583 9	0.583 6
FDM[10]	0.419 6	0.297 2
SDF[30]	0.473 0	-
本文方法	0.592 6	0.585 3

本文模型基于改进的残差网络并将注意力机制运用到改进的残差网络中,使用含有更多面部信息的峰值帧。将峰值帧的特征赋以不同权重,突出重要信息,抑制无用信息,同时由于ResNeXt-50本身的特性,由稀疏结构代替密集结构,避免了大量参数计算的问题,更好地融合不同尺度的信息。使用Focal Loss函数作为损失函数,将较难分类的样本较大的权重,使得模型在面对较难的样本时,可以更好地对较难样本的特征进行提取,从而在进行类似于微表情分类这种数据集较少且分布不平衡的训练任务时,降低因数据集的问题带来的影响。由表3可以看出,在5分类实验中,各个方法的准确率都是相对较低,说明微表情本身细微的表情变化和微表情数据集数据的不平衡给识别工作带来了一定的影响,并且未加权$F1$值也同样处于较低的水平,说明现有的微表情识别方法在不同类别上的表现差异相对较大。使用本文提出的方法,准确率达到了0.592 6,未加权$F1$值也达到了0.585 3,在所有展示方法中实验效果最好,同时也说明本文方法改善了在不同类别上的表现差异较大的问题。实验结果说明本文提出的微表情识别方法具有较好的效果,对于样本数据较少的问题也有着良好的拟合效果,数据不平衡问题也在一定程度上得到了解决。

3.6　消融实验结果分析

3.6.1　验证 SE-ResNeXt-50 的有效性

为了进一步验证 SE 模块与改进的残差网络 ResNeXt-50 在实验中的性能与有效性,在数据集上进行消融实验。这里进行 3 组实验,分别是不使用 SE 模块只使用初始的 ResNet-50 进行时实验;不使用 SE 模块只使用改进的残差网络 ResNeXt-50 进行实验;使用本文提出的同时使用 SE 模块和改进的残差网络 SE-ResNeXt-50 进行实验。

不同方法在 CASME Ⅱ数据集上的精确度与未加权 $F1$ 值实验结果对比如图 8 所示。由图 8 中的信息可以知道,使用 SE 模块和改进的残差网络 SE-ResNeXt-50 进行实验比不使用 SE 模块只使用初始的 ResNet-50 进行实验准确率提升 0.111 2;比不使用 SE 模块只使用改进的残差网络 ResNeXt-50 进行实验的准确率提升了 0.037 0。对于未加权 $F1$ 值,本文提出的方法较另外两种方法分别提升了 0.079 4 和 0.125 4。实验结果表明同时使用改进的残差网络和 SE 模块要比单独使用或者不使用的准确率与 $F1$ 的实验结果要有所提升。通过使用改进的残差网络可以较好地提高微表情识别的性能与精度。此外由于 SE 模块的加入,模型可以更好地学习峰值帧的中的关键信息,进一步地提高模型的识别准确度。改进的残差网络 SE-ResNeXt-50 通过将 SE 模块与 ResNeXt-50 相结合,用分组卷积的方式可以在不明显增加参数的情况下提升准确率,进而使网络结构也更加的简单,进一步利用注意力机制关注更有用更需要学习的信息,使得准确率得到提升,通过实验验证了其有效性。

3.6.2　验证峰值帧的有效性

为了进一步验证峰值帧在实验中的性能与有效性,在数据集上进行消融实验。进行两组实验,分别是使用峰值帧对网络进行训练和不使用峰值帧对网络进行训练,其中非峰值帧的选取方式为初始帧到峰值帧中间的一帧。对于不使用峰值帧的实验,也使用上一节实验所使用的实验方式,进行 3 组对比实验,即不使用 SE 模块只使用初始的 ResNet-50 进行时实验;不使用 SE 模块只使用改进的残差网络 ResNeXt-50 进行实验;使用本文提出的同时使用 SE 模块和改进的残差网络 SE-ResNeXt-50 进行实验。通过不同的组合实验,综合多种原因对比峰值帧的有效性,使实验的结果更加客观。

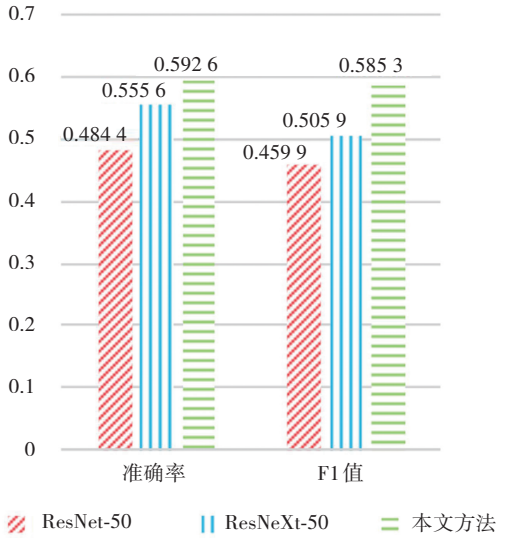


图 8　不同网络结构对比

Fig. 8　Comparison of different network structures

如表 4 所示,展示了是否使用峰值帧的实验结果对比。从表中可以看出使用峰值帧的实验结果要好于未使用峰值帧的实验结果。使用本文提出的方法,采用峰值帧的准确率与 $F1$ 值比不采用峰值帧分别提升了 0.071 与 0.084。使用 ResNet-50 与 ResNeXt-50 的方法,采用峰值帧的准确率与 $F1$ 值比不采用峰值帧也有所提升。实验结果证实了峰值帧可以提升微表情识别的准确率与 $F1$ 值,对于微表情识别具有一定的有效性。使用峰值帧的实验结果提升主要是因为峰值帧包含了更多微表情识别所需要的关键表情信息,相对于其余帧这些信息更加的紧凑。同时峰值帧的无用信息也更少,可以减少无用信息对于识别的干扰,进一步帮助模型更好地学习与挖掘到识别时所需要的更深层次的信息。

表 4　使用峰值帧结果比较

Table 4　Comparison of results using apex frames

方　法	准确率			$F1$ 值		
	RseNet-50	RseNeXt-50	本文方法	RseNet-50	RseNeXt-50	本文方法
使用峰值帧	0.481 4	0.555 6	0.592 6	0.459 9	0.505 9	0.585 3
不使用峰值帧	0.445 7	0.518 5	0.575 6	0.373 9	0.453 4	0.501 3

4 结论与展望

4.1 结论

本文提出了一种基于改进的残差网络与峰值帧的微表情识别算法。在微表情数据集 CASME Ⅱ上进行了仿真实验,根据相同评价指标对比各类微表情识别方法,可以发现改进的残差网络与峰值帧的微表情识别的准确率达到0.592 6,未加权 $F1$ 值达到了0.585 3。进一步进行消融实验,在不同的方法中也可以发现同时使用改进的残差网络与峰值帧相较于不使用的方法提高了准确率与 $F1$ 值。改进的残差网络与峰值帧可以降低数据集较少所带来的影响,数据不平衡问题在一定程度上得到了解决,模型有着良好的拟合效果,同时改善了在不同类别上的表现差异较大的问题,提升了微表情的识别准确率,对于微表情有着更好的识别性能。

4.2 展望

本文针对微表情识别提出了一种深度学习的方法,首先提取微表情 CASME Ⅱ数据集的峰值帧;对峰值帧预处理后,使用改进的残差网络 ResNeXt-50,同时加入 SE 模块形成 SE-ResNeXt-50 模型,用于微表情识别。ResNeXt 网络通过分组卷积的方式用稀疏结构取代密集结构从而使结构更加简化,融合不同尺度的信息,减少参数数量,提升了识别准确率;SE 模块可以更好地增强有用的特征,在更好地利用这些特征时,抑制一些无用的特征,进一步提升模型的性能;使用 Focal Loss 函数可以更好地解决图像领域因数据不平衡而引起的模型性能问题。由于现有的微表情数据集的样本数量太少,网络无法学习到更多的信息,所以识别的效果仍有较大的提升空间。后续计划收集更多的年龄、种族跨度更大的微表情数据,优化提升模型的性能与鲁棒性,建立端到端的实用型微表情识别系统,提高其实用价值。

参考文献(References):

[1] SHEN X B, QI W, FU X L. Effects of the duration of expressions on the recognition of micro expressions[J]. Journal of Zhejiang University SCIENCE B, 2012, 13(3): 221—230.

[2] 周伟航, 肖正清, 钱育蓉, 等. 微表情自动分析方法研究综述[J]. 计算机应用研究, 2022, 39(7): 1921—1932.
ZHOU Wei-hang, XIAO Zheng-qing, QIAN Yu-rong, et al. Review on automatic analysis methods of micro-expression[J]. Application Research of Computers, 2022, 39(7): 1921—1932.

[3] 贲晛烨, 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 Xian-ye, YANG Ming-qiang, ZHANG Peng, et al. Survey on automatic micro expression recognition methods[J]. Journal of Computer-Aided Design & Computer Graphics, 2014, 26(9): 1385—1395.

[4] ZHOU L, SHAO X, MAO Q. A survey of micro-expression recognition[J]. Image and Vision Computing, 2020, 105(1): 104043.

[5] POLIKOVSKY S, KAMEDA Y, OHTA Y. Facial micro-expressions recognition using high speed camera and 3D-gradient descriptor[C]//3rd International Conference on Imaging for Crime Detection and Prevention, ICDP. UK: IET, 2010: 1—6.

[6] Li X, PFISTER T, HUANG X, et al. A spontaneous micro-expression database: Inducement, collection and baseline[C]// 2013 10th IEEE International Conference and Workshops on Automatic Face and Gesture Recognition. Washington USA: IEEE, 2013: 1—6.

[7] HUANG X, ZHAO G, HONG X, et al. Spontaneous facial micro-expression analysis using spatiotemporal completed local quantized patterns[J]. Neurocomputing, 2016, 175(29): 564—578.

[8] WANG Y, SEE J, PHAN W, et al. LBP with six intersection points: Reducing redundant information in LBP-TOP for micro-expression recognition[C]//Asian conference on computer vision, ACCV2014. Switzerland: Springer, 2015: 525—537.

[9] WANG Y D, SEE J, PHAN R C W, et al. Efficient spatiotemporal local binary patterns for spontaneous facial micro-expression recognition[J]. PloS One, 2015, 10(5): 1—9.

[10] XU F, ZHANG J, WANG J Z. Micro expression identification and categorization using a facial dynamics map[J]. IEEE Transactions on Affective Computing, 2017, 8(2): 254—267.

[11] LIU Y J, ZHANG J K, YAN W J, et al. A main directional mean optical flow feature for spontaneous micro-expression recognition[J]. IEEE Transactions on Affective Computing, 2016, 7(4): 299—310.

[12] 李星燃, 张立言, 姚树婧. 结合特征融合和注意力机制的微表情识别方法[J]. 计算机科学, 2022, 49(2): 4—11.
LI Xing-ran, ZHANG Li-yan, YAO Shu-jing. Micro expression recognition method combining feature fusion and attention

mechanism[J]. Computer Science, 2022, 49(2): 4—11.

[13] DING C X, TAO D. Trunk-branch ensemble convolutional neural networks for video-based face recognition[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2018, 40(4): 1002—1014.

[14] HOURRI S, KHARROUBI J. A deep learning approach for speaker recognition[J]. International Journal of Speech Technology, 2020, 23(2): 123—131.

[15] PATEL D, HONG X, ZHAO G. Selective deep features for micro-expression recognition[C]//2016 23rd International Conference on Pattern Recognition, ICPR. Washington, USA: IEEE, 2017: 2258—2263.

[16] PENG M, WANG C Y, CHEN T, et al. Dual temporal scale convolutional neural network for micro-expression recognition [J]. Frontiers in Psychology, 2017, 8(1): 1745—1757.

[17] KHOR H Q, SEE J, PHAN R, et al. Enriched long-term recurrent convolutional network for facial micro-expression recognition[C]// 2018 13th IEEE International Conference on Automatic Face & Gesture Recognition, FG2018. Washington, USA: IEEE, 2018: 667—674.

[18] CAI LQ, LI H, DONG W, et al. Micro-expression recognition using 3D densenet fused squeeze-and-excitation networks[J]. Applied Soft Computing, 2022, 119(1): 1—12.

[19] 张学森, 贾静平. 基于三维卷积神经网络和峰值帧光流的微表情识别算法[J]. 模式识别与人工智能, 2021, 34(5): 423—433.
ZHANG Xue-sen, JIA Jing-ping. Micro-expression recognition algorithm based on 3D convolutional neural network and optical flow fields from neighboring frames of apex frame[J]. Pattern Recognition and Artificial Intelligence, 2021, 34(5): 423—433.

[20] XIE S, GIRSHICK R, DOLLÁR P, et al. Aggregated residual transformations for deep neural networks[C]//2017 IEEE Conference on Computer Vision and Pattern Recognition, CVPR. Washington, USA: IEEE, 2017: 1492—1500.

[21] HU J, SHEN L, ALBANIE S. Squeeze-and-excitation networks[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2020, 42(8): 2011—2023.

[22] HE K, ZHANG X Y, REN S P, et al. Deep residual learning for image recognition[C]//2016 IEEE Conference on Computer Vision and Pattern Recognition, CVPR. Washington, USA: IEEE, 2016: 770—778.

[23] SZEGEDY C, LIU W, JIA Y, et al. Going deeper with convolutions[C]//2015 IEEE Conference on Computer Vision and Pattern Recognition, CVPR. Washington, USA: IEEE, 2015: 1—9.

[24] YAN W J, LI X B, WANG S J, et al. CASME Ⅱ: An improved spontaneous micro-expression database and the baseline evaluation[J]. PLoS ONE, 2014, 9(1): 1—8.

[25] QUANG N V, CHUN J, TOKUYAMA T. CapsuleNet for micro-expression recognition[C]//2019 14th IEEE International Conference on Automatic Face & Gesture Recognition. Washington, USA: IEEE, 2019: 1—7.

[26] LIN T Y, GOYAL P, GIRSHICK R, et al. Focal loss for dense object detection[J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2017(99): 2999—3007.

[27] 牛瑞华, 杨俊, 邢斓馨, 等. 基于卷积注意力模块和双通道网络的微表情识别算法[J]. 计算机应用, 2021, 41(9): 2552—2559.
NIU Rui-hua, YANG Jun, XING Lan-xin, et al. Micro-expression recognition algorithm based on convolutional block attention module and dual path networks[J]. Journal of Computer Applications, 2021, 41(9): 2552—2559.

[28] LE NGO A C, PHAN R C W, SEE J. Spontaneous subtle expression recognition: Imbalanced databases and solutions[C]// 12th Asian Conference on Computer Vision, ACCV. Switzerland: Springer, 2014: 33—48.

[29] HUANG X, ZHAO G, HONG X, et al. Spontaneous facial micro-expression analysis using spatiotemporal completed local quantized patterns[J]. Neurocomputing, 2016, 175(29): 564—578.

[30] PATEL D, HONG X P, ZHAO G Y. Selective deep features for micro-expression recognition[C]//2016 23rd International Conference on Pattern Recognition, ICPR. Washington, USA: IEEE, 2016: 2258—2263.

责任编辑:陈　芳