

【电子与信息科学 / Electronics and Information Science】

基于递增注意力的微表情识别方法

战子为^{1, 2, 3}, 孙兆才^{1, 2, 3}, 李翔^{1, 2, 3}, 吴镇东^{1, 2, 3}

1) 山东中医药大学青岛中医药科学院, 山东青岛266112; 2) 山东中医药大学医学人工智能研究中心, 山东青岛266112;
3) 山东中医药大学青岛市中医人工智能技术重点实验室, 山东青岛266112

摘要: 微表情是个体无意识发生的表情变化, 能够反映人们潜在的情绪和内心状态. 微表情发生时动作强度低且涉及面部范围小, 导致在微表情识别过程中存在着提取特征不充分、定位有效特征不准确的问题, 影响识别精度. 构建一种递增注意力多尺度卷积网络, 该网络融合了多尺度卷积模块和递增注意力模块. 利用多尺度卷积模块学习不同感受野下的细粒度特征, 提取丰富的细节特征, 同时设计一种递增注意力模块, 通过多个注意力图间的特征共享与增强, 准确定位面部运动区域, 稳健提取微表情图像中的运动特征. 所提网络在数据集SMIC、CASMEII及SAMM上进行实验, 准确率分别达到0.826、0.880和0.787, F_1 值分别达到0.817、0.864和0.761. 研究结果可为谎言检测、心理健康早期筛查等提供参考.

关键词: 人工智能; 微表情识别; 深度学习; 注意力机制; 卷积神经网络; 多尺度卷积; 谎言检测; 心理健康早期筛查

中图分类号: TP391; R318

文献标志码: A

DOI: 10.3724/SP.J.1249.2024.06756

Micro-expression recognition method based on progressive attention

ZHAN Ziwei^{1, 2, 3}, SUN Zhaocai^{1, 2, 3}, LI Xiang^{1, 2, 3}, and WU Zhendong^{1, 2, 3}

1) Qingdao Academy of Traditional Chinese Medicine, Shandong University of Traditional Chinese Medicine, Qingdao 266112, Shandong Province, P. R. China

2) Center for Medical Artificial Intelligence, Shandong University of Traditional Chinese Medicine, Qingdao 266112, Shandong Province, P. R. China

3) Qingdao Key Laboratory of Traditional Chinese Medicine Artificial Intelligence Technology, Shandong University of Traditional Chinese Medicine, Qingdao 266112, Shandong Province, P. R. China

Abstract: Micro-expression is unconscious expression changes that reflect people's underlying emotions and inner states. When micro-expressions occur, their low intensity and the small facial range result in insufficient feature extraction and inaccurate localization of effective features during the recognition process, which affects recognition accuracy. To address this issue, a progressive attention multi-scale convolutional network was constructed. The network integrates a multi-scale convolutional module and a progressive attention module. First, the multi-scale convolutional module is used to learn fine-grained features from different receptive fields, extracting rich details. Then, the progressive attention module is designed to accurately locate facial motion areas and robustly extract motion features from micro-expression images through information sharing and enhancement across multiple attention

Received: 2024-01-12; **Accepted:** 2024-04-21; **Online (CNKI):** 2024-08-27

Foundation: National Natural Science Foundation of China (62372280, 61872225); Natural Science Foundation of Shandong Province (ZR2020KF013, ZR2020QF043, ZR2023QF094); Science and Technology Benefiting the People Demonstration Project of Qingdao (23-2-8-smjk-2-nsh)

Corresponding author: Associate professor SUN Zhaocai (zhcsun@hotmail.com)

Citation: ZHAN Ziwei, SUN Zhaocai, LI Xiang, et al. Micro-expression recognition method based on progressive attention [J]. Journal of Shenzhen University Science and Engineering, 2024, 41(6): 756-764. (in Chinese)



maps. The proposed network was tested on SMIC, CASMEII and SAMM datasets, achieving accuracy rates of 0.826, 0.880 and 0.787, and F_1 scores of 0.817, 0.864 and 0.761, respectively. The proposed method can serve as an auxiliary tool for lie detection and early screening of mental health conditions.

Key words: artificial intelligence; micro-expression recognition; deep learning; attention mechanism; convolutional neural network; multi-scale convolution; lie detection; early screening for mental health

面部表情是人类最直接的表达方式之一, 传递的信息量占人们传递信息总量的55%^[1]. 面部表情分为宏表情与微表情(如图1). 微表情是对刺激做出反应的微妙且短暂的面部动作, 当人们试图隐藏真实情绪, 微表情就会自发地、不受控地显现^[2]. 因此, 微表情可作为提供相关个体真实情感和内心状态的重要线索^[3]. 高精度的微表情识别方法能为谎言检测及心理健康早期筛查等提供技术支持, 在相关领域有重要应用价值^[4-5].



图1 微表情和宏表情示例

Fig. 1 (Color online) Examples of various expressions of (a) micro-expressions and (b) macro-expressions.

目前微表情识别方法的研究可分为基于手工特征的方法^[6-7]和基于深度学习的方法. 基于手工提取特征的方法提取的特征仅包含表面信息, 局部纹理特征存在鲁棒性和稀疏采样的问题, 无法很好地捕捉微表情的细微变形. 近年来深度学习技术发展快速, 被广泛用于微表情识别, 有学者通过从视频帧序列中提取特征来实现微表情识别^[8-9], 能够捕捉微表情序列的时空变形, 但也存在着数据冗余、特征提取复杂度高的问题. KHOR等^[10]从关键帧出发, 构造了输入为异构光流特征的双流卷积神经网络(convolutional neural network, CNN); ZHOU等^[11]从关键帧学习微表情的共有和特异特征. 为充分考虑局部区域内细粒度特征提取和表达, 更好地识别微小且瞬时的微表情, WANG等^[12]采用跳跃连接级联图像和局部特征来保证细节特征的有效传输; LI等^[13]提出基于光流图像的多尺度联合特征网络. 此

外, 针对微表情运动幅度小带来的有效特征定位困难的问题, 有学者将注意力机制引入到深度学习网络中. CHEN等^[14]在3维卷积神经网络中添加了卷积注意力模块, 以自适应地分配特征的权重来提取关键特征. WANG等^[15]结合微注意力和残差网络, 定位不同运动单元(action unit, AU)的面部感兴趣区域. LI等^[16]提出了能有效地识别出面部AU的空间通道注意力网络. ZHOU等^[17]基于Inception构建了双流注意力网络. 尽管引入注意力机制提高了识别精度, 但在处理程度微弱的微表情时, 仍存在着关键特征定位不准确的问题.

本研究提出一种递增注意力模块, 通过引入前层注意力作为先验知识来动态更新注意力生成过程, 用以解决有效特征定位不准确的问题. 考虑到多尺度特征对不同尺度目标的捕获能力不同^[18-19], 又设计了多尺度卷积模块, 用以解决微表情识别中细节特征提取不充分的问题. 该模块能通过多分支获取不同感受野的特征, 保留微表情的细节特征. 将递增注意力模块与多尺度卷积模块融合构成递增注意力多尺度卷积网络, 能够准确捕捉微表情的细微形变, 实现更精准微表情识别.

1 基本原理

1.1 网络结构

本研究提出的基于递增注意力的微表情方法流程如图2, 包括数据预处理、网络训练和性能评估等关键环节. 在构建网络模型时, 因微表情数据集样本量有限, 为防止网络层数过多造成过拟合, 采用ResNet18网络作为主干网络, 并分别构建多尺度卷积模块和递增注意力模块. 该网络中含有1个多尺度卷积模块、4个残差结构和4个递增注意力模块. 多尺度卷积模块用于提取细粒的细节特征, 递增注意力模块用于定位重要特征区域, 提取丰富的语义特征. 最后, 将提取到的特征输入到全连接层, 通过softmax层输出微表情识别结果. 图3为具体构建的网络结构.

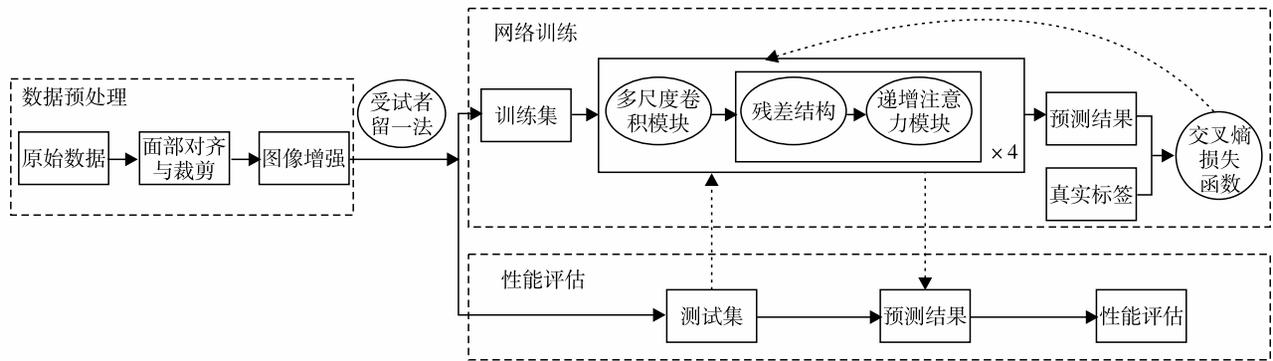


图2 基于递增注意力的微表情识别方法流程

Fig. 2 The flow chart of micro-expression recognition algorithm based on progressive attention.

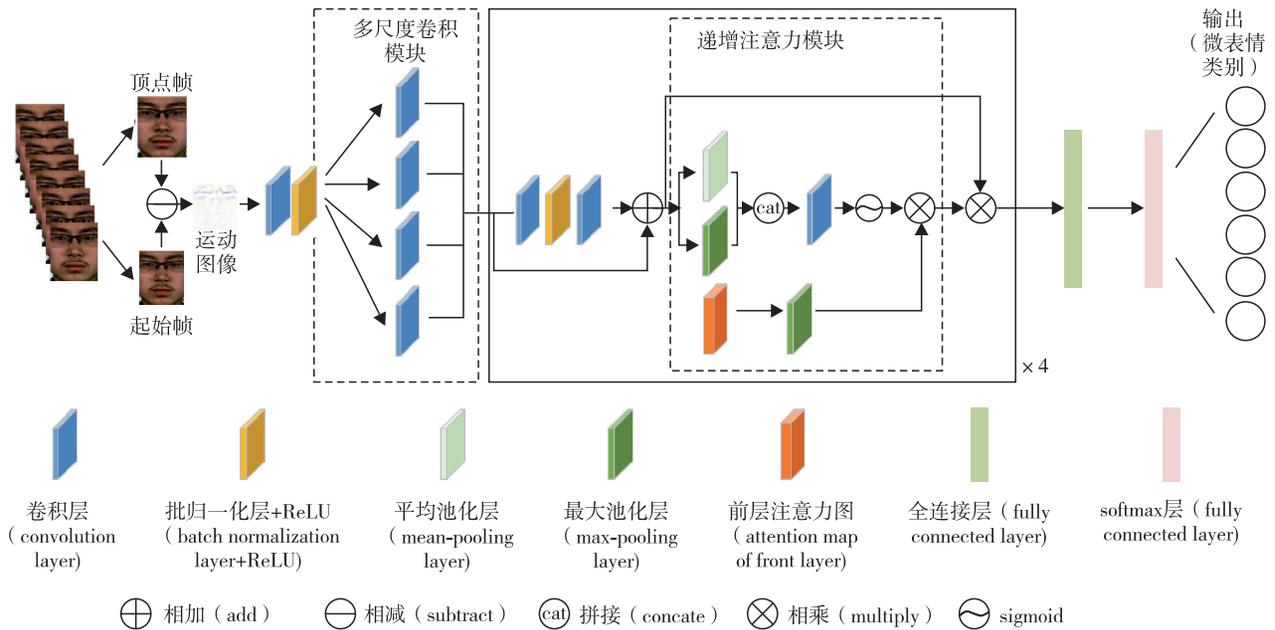


图3 基于递增注意力的微表情识别方法的网络结构

Fig. 3 (Color online) Structure of micro-expression recognition algorithm based on progressive attention.

1.2 图像预处理

1.2.1 面面对齐与裁剪

本研究使用常用数据集 SMIC^[20]、CASMEII^[21] 及 SMM^[22] 进行实验. 由于 SMM 数据集中的原始图像帧中含有干扰识别任务的图像背景, 为减轻图像背景中噪声和干扰物的影响, 先对原始图像进行面面对齐与裁剪. 如图 4, 首先通过 Dlib 检测器对面部区域 68 个关键点进行检测, 获取关键点的坐标信息; 然后利用坐标信息计算出两眼中心与水平线之间的角度, 再采用仿射变换实现面面对齐; 最后将检测出的面部区域裁剪并保存. 经过面面对齐与裁剪, 能够使后续特征提取更精确地聚焦于面部区域.



图4 面面对齐与裁剪效果

Fig. 4 (Color online) Visualization of face alignment and cropping effects. (a) Original images, (b) key point detection images, and (c) face alignment and cropping images.

1.2.2 图像增强

为避免网络因样本数量少导致训练过拟合, 需先对输入图像进行增强(水平翻转、随机旋转及随机裁剪等操作), 再将图像大小统一裁剪为 224 × 224 像素(图 5). 然后, 对所有数据作归一化处理,

确保输入的图像数据在训练过程中具有相似分布, 利于网络训练和收敛. 接着, 通过数据索引在图像序列中定位到起始帧与顶点帧. 由于数据集 SMIC 未提供顶点帧的标签信息, 借鉴文献[23]将起始帧与结束帧的中间帧作为顶点帧, 再将两帧间逐元素相减生成面部运动图像, 用于后续进行特征提取. 经过图像反转处理的运动图像的可视化效果示例可扫描论文末页右下角二维码查看补充材料图 S1.

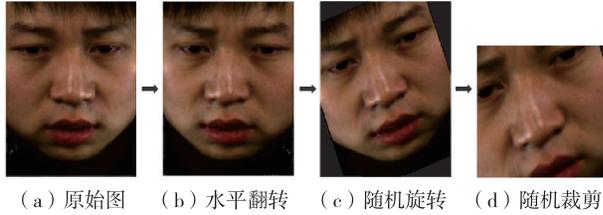


图5 图像增强效果

Fig. 5 (Color online) Visualization of image enhancement effects. (a) Original images, (b) horizontal flip images, (c) random rotation images, and (d) random images.

1.3 多尺度卷积模块

在网络提取特征阶段, 所得运动图像需先经过卷积核为 3×3 大小的卷积层、批归一化层和 ReLU 层, 从而增加通道数量. 运动图像 x 的特征为

$$X = f_{3 \times 3}(x) \quad (1)$$

其中, $f_{3 \times 3}$ 表示 3×3 卷积操作.

将 X 输入多尺度卷积模块, 以期探索微表情图像的纹理等细节特征, 有效捕获面部特征. 卷积层使用卷积核对输入数据进行滑动窗口的卷积操作, 得到有用特征. 对于输入尺寸为 $W \times H \times D$ (宽度 \times 高度 \times 深度) 的特征图, 若以步长 s 、填充大小 p 经过 k 个 $w \times w$ 大小的卷积核, 得到的输出特征图尺寸为 $W_{out} \times H_{out} \times D_{out}$ (宽度 \times 高度 \times 深度). 其中,

$$W_{out} = \frac{W - w + 2p}{s} + 1 \quad (2)$$

$$H_{out} = \frac{H - w + 2p}{s} + 1 \quad (3)$$

$$D_{out} = k \quad (4)$$

微表情的细微变化常常隐藏在面部的不同区域, 多尺度卷积模块通过多分支使用不同大小的卷积核, 自适应地学习不同感受野的特征, 并通过融合多级特征, 更全面地提取细粒度的细节特征, 从而增强了网络的多尺度表达能力. 多尺度卷积模块结构如图6. 具体来说, 对于通道维度为 C 的输入, 经过多尺度卷积模块时, 被并行传输给4个分支通道, 在每个分支中通道维度均被压缩为 $C/4$, 且各

通道分别使用尺寸 k_1 、 k_2 、 k_3 和 k_4 的卷积核, 经卷积扩充后得到相同尺寸的特征图 F_1 、 F_2 、 F_3 和 F_4 , 再拼接所有通道的特征得到通道维度为 C 的模块输出. 多尺度输出为

$$F_{cat} = \text{cat}(F_1, F_2, F_3, F_4) \quad (5)$$

$$F_i = f_{k_i \times k_i}(X), \quad i = 1, 2, 3, 4 \quad (6)$$

其中, cat 函数为沿着深度(通道数)维度进行拼接; $f_{k_i \times k_i}$ 为 $k_i \times k_i$ 的卷积操作函数.

对比模块中不同卷积核尺寸的实验结果后发现, 证明卷积核大小为3、5、7和9时效果最好. 因此, 本研究实验设 $k_i = 2i + 1$.

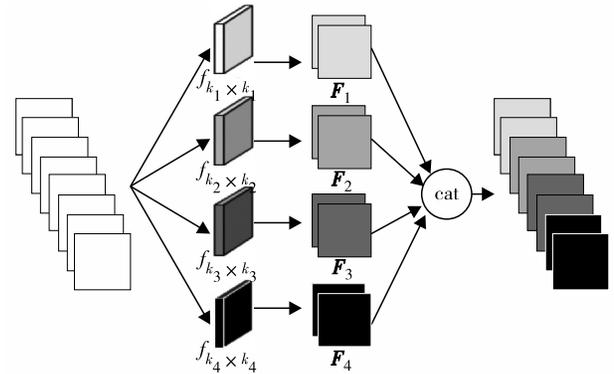


图6 多尺度卷积模块结构

Fig. 6 Concrete structure of multi-scale convolutional module.

1.4 递增注意力模块

在微表情识别任务中, 运动特征的获取对于准确识别微表情有重要意义. 为缓解现有的微表情识别方法难以准确聚焦细微面部肌肉运动的问题, 本研究设计了递增注意力模块, 在生成注意力图时引入前层注意力图作为先验知识, 再利用注意力图之间的特征共享生成新注意力图, 以实现有效特征动态定位. 所提网络采用4个递增注意力模块, 为防止网络退化, 在每个递增注意力模块前连接1个残差结构. 残差结构直接将输入与拟合的残差相加. 对于第 i 个残差结构输入 D_i , 可得残差结构的输出 $F_{conv,i}$, 即第 i 个递增注意力模块的输入为

$$F_{conv,i} = f_{1 \times 1}(f_{3 \times 3}(D_i)) + D_i, \quad i = 1, 2, 3, 4 \quad (8)$$

其中, $f_{1 \times 1}$ 和 $f_{3 \times 3}$ 分别为 1×1 和 3×3 卷积操作.

为得到更全面的空间注意力图, 受文献[24]启发, 递增注意力模块在生成注意力图时将最大池化和平均池化相结合, 分别获得每个空间位置的最大值和平均值, 以期捕获图像中显著特征与整体特征, 再通过拼接操作实现特征融合. 通过 1×1 卷积层和 sigmoid 函数, 得到能够指导网络关注重要

面部区域的注意力图，再将此新生成的注意力图与前层注意力图逐点相乘，得到进一步加权的当前层注意力图。这种前层注意力图的引入能够更新注意力分布，在空间上加权前一层的注意力图，从而确保所提网络的关注点能更精确地集中于图像中的关键部位。由于第 1 个递增注意力模块在生成注意力图时未引入前层注意力图，因此产生的注意力图为

$$A_1 = \sigma \left(f_{1 \times 1} \left(\left[P_{\max}(\mathbf{F}_{\text{conv},1}); P_{\text{avg}}(\mathbf{F}_{\text{conv},1}) \right] \right) \right) \quad (9)$$

其中， $\mathbf{F}_{\text{conv},1}$ 为第 1 个递增注意力模块的输入； σ 为 sigmoid 函数； P_{\max} 和 P_{avg} 分别为最大池化和平均池化。其他模块在生成注意力图时均引入了前层注意力图，因此得到的注意力图为

$$A_i = \sigma \left(f_{1 \times 1} \left(\left[P_{\max}(\mathbf{F}_{\text{conv},i}); P_{\text{avg}}(\mathbf{F}_{\text{conv},i}) \right] \right) \right) \odot P_{\max}(A_{i-1}), \quad i = 2, 3, 4 \quad (10)$$

其中， \odot 为哈达玛积运算符。将注意力图与原始特征图逐元素相乘，就可得到具有增强空间区域重要性的特征图。这种多模块间的串联使得所提网络能在注意力图的更新引导下，重点关注具有细微运动的区域，从中稳健地提取并学习微表情图像中的运动特征，实现更加准确和鲁棒的特征表达。

2 实验结果分析

2.1 实验配置

实验设置学习率为 1×10^{-4} ，训练过程采用受试者留一法交叉验证来分析所提网络的性能。操作系统为 Windows10，在型号 Tesla V100-SXM2-32 Gbyte 的 GPU 上运行。

2.2 实验数据集

表 1 给出了实验所用微表情数据集 SMIC、CASMEII 及 SAMP 的具体分布。SMIC 数据集是最早发布的自发微表情数据集，包含了来自 16 名参与者的 167 个微表情样本，样本由 100 帧/s 的高速相机采集记录，标签分为积极、惊讶及消极 3 类。CASMEII 数据集由中国科学院大学发布，所含样本是采用 200 帧/s 的高速相机对 26 名参与者进行采集，人脸区域的图像分辨率为 280×340 像素。数据集 SAMP 中的样本来自 29 名受试者，使用高速相机以 200 帧/s 的速率采集样本，所得面部区域图像的分辨率为 400×400 像素。

2.3 评价指标

采用受试者留一法选取训练集和测试集。依次

表 1 微表情数据集 SMIC、CASMEII 及 SAMP 的类别分布
Table 1 The class distribution of SMIC, CASMEII, and SAMP micro-expression datasets

SMIC		CASMEII		SAMP	
类别	样本数	类别	样本数	类别	样本数
积极	51	快乐	32	快乐	26
惊讶	45	厌恶	63	生气	57
消极	71	惊讶	28	蔑视	12
		压抑	27	惊讶	15
		其他	99	其他	26

选择数据集中 1 名受试者样本作为测试集，并将剩余受试者样本作为训练集，直到测试集遍历所有受试者样本。考虑到数据集类别不平衡的问题，实验采用准确率 (A) 与 F_1 值作为评价指标。准确率指预测的结果中被正确预测的样本数量与总样本数量的占比。准确率越高代表网络的预测能力越好，计算公式为

$$A = \frac{1}{N} \sum_{i=1}^N [y_i = \hat{y}_i] \quad (11)$$

其中， N 为样本总量； y_i 为第 i 个样本的预测结果； \hat{y}_i 为第 i 个样本的标签结果； $[y_i = \hat{y}_i]$ 为指示函数，当 $y_i = \hat{y}_i$ 时， $[y_i = \hat{y}_i]$ 返回 1，否则， $[y_i = \hat{y}_i]$ 返回 0。

F_1 值是一个综合考虑精确率和召回率的指标，准确率指预测为正样本中被预测正确的样本量占比，召回率指标为标签为正样本中被正确预测的样本量占比。实验为多分类任务，可得

$$F_1 = \frac{1}{M} \sum_{a=1}^M F_{1a} \quad (12)$$

$$F_{1a} = 2 \times \frac{\sum_{i=1}^N [y_i = a, \hat{y}_i = a]}{\sum_{i=1}^N [y_i = a] + \sum_{i=1}^N [\hat{y}_i = a]} \quad (13)$$

其中， M 为标签类别量； a 为标签类别； F_{1a} 为类别 a 的 F_1 值； $[y_i = a, \hat{y}_i = a]$ 、 $[y_i = a]$ 和 $[\hat{y}_i = a]$ 为指示函数；当 y_i 为 a 且 \hat{y}_i 为 a 时， $[y_i = a, \hat{y}_i = a]$ 返回 1，否则返回 0；当 y_i 为 a 时， $[y_i = a]$ 返回结果 1，否则返回 0；当 \hat{y}_i 为 a 时， $[\hat{y}_i = a]$ 返回 1，否则返回 0。

2.4 对比实验

将本研究所提网络与其他微表情识别任务中较为先进的方法进行对比，包括基于手工特征的 3 个

正交平面的局部二元模式(local binary patterns from three orthogonal planes, LBP-TOP)^[25]、双加权定向光流(bi-weighted oriented optical flow, Bi-WOOF)^[26], 以及基于深度学习的双流浅层网络(dual-stream shallow network, DSSN)^[10]、DConv+ETran+AUGCN+AUFsuion^[27]、LBP+attention^[28]、面向人脸微表情识别的多模态注意力动态融合网络(multimodal attention dynamic fusion network, MADFN)^[29]和微表情识别双分支注意力网络(dual-branch attention network, Dual-ATME)^[17]。LBP-TOP从3个正交平面上提取纹理特征实现识别任务; Bi-WOOF通过光流应变对微表情的运动特征进行编码; DSSN输入异构光流到多分支CNN以进一步提

取并融合特征用于分类; DConv+ETran+AUGCN+AUFsuion使用CNN从地标标签中学习面部节点和边缘特征, 并嵌入图卷积网络(graph convolutional networks, GCN)学习AU矩阵; LBP+attention提取图像的RGB特征与LBP纹理特征, 并添加注意力机制来进行特征融合; MADFN通过注意力机制形成掩码去除微表情图像中情绪表达能力较低的局部区域, 融合AU来提高图像特征的潜在表达能力; Dual-ATME采用双支结构, 利用先验知识从感兴趣区域中手动提取特征, 并基于注意机制自动从数据中提取深层特征。通过对比实验, 可更全面地评估所提网络的性能, 对比实验结果汇总如表2。

表2 对比实验结果

Table 2 Results of the comparative experiments with advanced methods of deep learning

方法	模型	SMIC数据集		CASMEII数据集		SAMM数据集	
		A	F_1	A	F_1	A	F_1
基于手工特征的方法	LBP-TOP ^[25]	0.544	0.567	0.651	0.674	0.462	0.488
	Bi-WOOF ^[26]	0.583	0.573	0.803	0.781	0.514	0.521
基于深度学习的方法	DSSN ^[10]	0.634	0.646	0.708	0.729	0.574	0.464
	DConv+ETran+AUGCN+AUFsuion ^[27]	0.719	0.722	0.743	0.705	0.743	0.704
	LBP+attention ^[28]	0.812	—	0.745	—	—	—
	MADFN ^[29]	0.817	0.810	0.821	0.801	0.772	0.749
	Dual-ATME ^[17]	0.646	0.645	0.817	0.765	0.714	0.560
	本研究网络	0.826	0.817	0.880	0.864	0.787	0.761

注: 灰底数据为对应指标下的最优值。

由表2可知, 基于手工特征的LBP-TOP^[25]与Bi-WOOF^[26]整体表现较差, 原因是基于手工特征的方法将提取的纹理、形状或光流特征拼接为特征向量, 再输送到分类器中完成分类, 提取特征时缺乏对于深层空间特征的考虑, 未能有效地捕获微表情中的细微变化和动态特征。相比之下, 基于深度学习的方法能够自动学习并提取更丰富、更高层次的特征表示, 在识别任务中取得更好的性能。本研究网络在数据集SMIC、CASMEII和SAMM上准确率分别为0.826、0.880和0.787, F_1 值分别为0.817、0.864和0.761, 在所有方法中均为最优。综上所述, 所提网络通过融合多尺度卷积模块和递增注意力模块, 取得了较好的微表情识别结果。

2.5 消融实验

为进一步验证所提网络中多尺度卷积模块和递增注意力模块的有效性, 在3个数据集上进行消融

实验, 结果如表3。其中, 实验A为基准实验, 用于验证主干网络的效果, 实验B、C和D分别添加递增注意力模块、多尺度卷积模块和同时添加两个模块。由表3可见, 在主干网络上添加递增注意力模块后, 数据集SMIC、CASMEII和SAMM上的准确率分别提升了0.066、0.040和0.059, 表明该模块有助于网络更好地关注微表情中的关键特征, 从而提高识别性能; 在主干网络上添加多尺度卷积模块后, 数据集SMIC、CASMEII和SAMM上的准确率分别提升了0.125、0.188和0.125, 表明该模块能够更全面地捕捉微表情的空间特征; 而在主干网络上同时添加两个模块时, 数据集SMIC、CASMEII和SAMM上的准确率分别提升了0.173、0.210和0.154。结果证明了所提网络中两个模块对于微表情识别任务均有提高效果, 且模块间的相互作用能够进一步增强网络性能。

表3 消融实验结果

Table 3 Results of ablation experiments

实验	SMIC数据集		CASMEII数据集		SAMM数据集	
	A	F_1	A	F_1	A	F_1
A	0.653	0.552	0.671	0.598	0.610	0.607
B	0.719	0.615	0.711	0.658	0.669	0.652
C	0.778	0.759	0.859	0.827	0.735	0.721
D	0.826	0.817	0.880	0.864	0.787	0.761

注：灰底数据为对应指标下的最优值。

为探究多尺度卷积模块中不同卷积核大小对网络性能的影响，并验证多个相同卷积是否能带来类似的性能提升，在不同数据集上各设置 5 组实验，结果如表 4。由表 4 可见，当卷积核大小为 3、5、7 和 9 时，所提网络的分类识别能力最优秀，在数据集 SMIC、CASMEII 及 SAMM 上均达到最优效果。

2.6 混淆矩阵结果

为评估本研究网络在不同情绪类别上的分类性能，使用混淆矩阵直观地展示网络在 3 个数据集上的分类性能，结果如图 7。从图 7 可见，所提网络在 CASMEII 数据集上的分类结果最好，各类别的分类准确性整体较高，尤其是“惊讶”的准确性达到 0.93，原因可能是数据集中的参与者均为亚洲人且年龄差异较小，此因素有助于网络更好地捕捉情绪

表4 不同卷积核大小的实验结果

Table 4 Results of different convolution kernel sizes

实验	卷积核大小	SMIC数据集		CASMEII数据集		SAMM数据集	
		A	F_1	A	F_1	A	F_1
E	1,3,5,7	0.796	0.782	0.835	0.822	0.750	0.742
F	3,5,7,9	0.826	0.817	0.880	0.864	0.787	0.761
G	5,7,9,11	0.808	0.796	0.855	0.846	0.742	0.730
H	3,3,3,3	0.802	0.780	0.859	0.839	0.746	0.733
I	9,9,9,9	0.784	0.775	0.831	0.797	0.735	0.726

注：灰底数据为对应指标下的最优值。

特征，提高识别的准确性。在数据集 SMIC 中，各类别分类准确率相对稳定，但“惊讶”被错误识别为“消极”的可能性较高，原因可能是这两种情绪在表情和生理反应上存在一定的相似性，且标签主要基于参与者的自我报告，增加了错误识别的可能性。相比之下，数据集 SAMM 中各类别的分类准确率较低，特别是“惊讶”的分类准确率为 0.67，这可能是由于该数据集中涵盖了 13 个不同种族的参与者，多样化的个体差异给网络的训练带来了挑战，且“惊讶”数据量较少，造成网络无法充分提取类别的特征，在分类时对该类别的识别能力较弱。

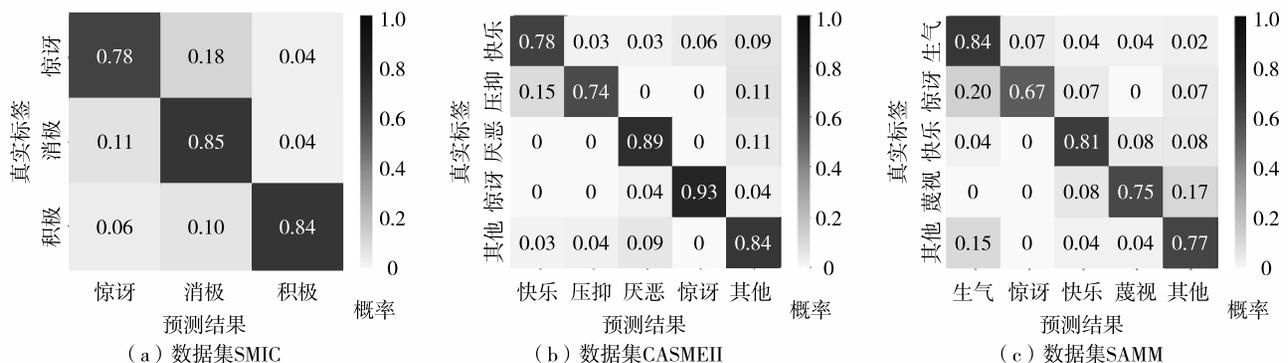


图7 各数据集的分类混淆矩阵

Fig. 7 Confusion matrix for classification on datasets (a) SMIC, (b) CASMEII, and (c) SAMM.

结 语

提出一种递增注意力多尺度卷积网络，实验结果证明该网络具有较强的鲁棒性，在微表情识别任务中展现出优越的识别性能。该网络融合了多尺度卷积模块和递增注意力模块，通过多分支特征融合

提取不同尺度下的细粒度特征，实现了不同尺度下的特征补偿，充分捕获面部细节特征，获得丰富的细节表达；采用递增式多层注意力机制动态关注对微表情运动区域，精准定位面部微小运动区域，捕捉面部微小变化并学习微表情运动的语义特征。对比实验结果证明了本研究网络性能优越，有较强的多尺度表达能力和定位重要特征的能力，有效缓解

了其他网络因特征提取不充分、特征定位不准确带来的识别问题, 有助于提升对微表情情感判定的准确性。

微表情数据集数量少且类别不平衡仍是目前制约微表情识别研究的重要因素。微表情与宏表情在数据上存在一定的相似性, 而现有宏表情识别的研究已经相对成熟, 因此, 未来可考虑通过迁移学习, 将宏表情识别的经验和网络知识迁移到微表情识别任务中, 以应对数据稀缺和类别不平衡带来的挑战, 提升网络的泛化能力和稳定性。

基金项目: 国家自然科学基金资助项目(62372280, 61872225); 山东省自然科学基金资助项目(ZR2020KF013, ZR2020QF043, ZR2023QF094); 青岛市科技惠民示范专项资助项目(23-2-8-smjk-2-nsh)

作者简介: 战子为(2021111401@stutem.edu.cn), 山东中医药大学硕士研究生。研究方向: 医学人工智能。

引文: 战子为, 孙兆才, 李翔, 等. 基于递增注意力的微表情识别方法[J]. 深圳大学学报理工版, 2024, 41(6): 756-764.

参考文献 / References:

- [1] 于明, 钟元想, 王岩. 人脸微表情分析方法综述[J]. 计算机工程, 2023, 49(2): 1-14.
YU Ming, ZHONG Yuanxiang, WANG Yan. A survey of facial micro-expression analysis methods [J]. Computer Engineering, 2023, 49(2): 1-14. (in Chinese)
- [2] QUANG N V, CHUN J, TOKUYAMA T. CapsuleNet for micro-expression recognition [C]// The 14th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2019). Piscataway, USA: IEEE, 2019: 1-7.
- [3] 宋明芮, 唐海, 徐洪胜. 基于改进YOLOv5s的微表情识别研究[J]. 重庆科技学院学报自然科学版, 2023, 25(6): 65-71.
SONG Mingrui, TANG Hai, XU Hongsheng. Research on micro-expression recognition based on improved YOLOv5s [J]. Journal of Chongqing University of Science and Technology Natural Sciences Edition, 2023, 25(6): 65-71. (in Chinese)
- [4] 黄树成, 罗德广. 基于光流和自编码器的微表情检测方法[J]. 计算机应用与软件, 2023, 40(9): 171-176.
HUANG Shucheng, LUO Deguang. Micro-expression detection method based on optical flow and autoencoder [J]. Computer Applications and Software, 2023, 40(9): 171-176. (in Chinese)
- [5] 杜芳芳, 王福忠, 高继梅. 空洞卷积网络下微表情实时识别方法仿真[J]. 计算机仿真, 2023, 40(7): 172-175, 461.
DU Fangfang, WANG Fuzhong, GAO Jimei. Simulation of real-time recognition method of micro-expression under empty convolution network [J]. Computer Simulation, 2023, 40(7): 172-175, 461. (in Chinese)
- [6] HUANG Xiaohua, ZHAO Guoying. Spontaneous facial micro-expression analysis using spatiotemporal local radon-based binary pattern [C]// International Conference on the Frontiers and Advances in Data Science. Piscataway, USA: IEEE, 2017: 159-164.
- [7] LIU Yongjin, LI Bingjun, LAI Yukun. Sparse MDMO: learning a discriminative feature for micro-expression recognition [J]. IEEE Transactions on Affective Computing, 2021, 12(1): 254-261.
- [8] ZHOU Guanqun, YUAN Shusen, XING Hongbo, et al. Micro-expression action unit recognition based on dynamic image and spatial pyramid [J]. The Journal of Supercomputing, 2023, 79(17): 19879-19902.
- [9] ZHENG Yufeng, BLASCH E. Facial micro-expression recognition enhanced by score fusion and a hybrid model from convolutional LSTM and vision transformer [J]. Sensors, 2023, 23(12): 5650.
- [10] KHOR H Q, SEE J, LIONG S T, et al. Dual-stream shallow networks for facial micro-expression recognition [C]// IEEE International Conference on Image Processing. Piscataway, USA: IEEE, 2019: 36-40.
- [11] ZHOU Ling, MAO Qirong, HUANG Xiaohua, et al. Feature refinement: an expression-specific feature learning and fusion method for micro-expression recognition [J]. Pattern Recognition, 2022, 122: 108275.
- [12] WANG Mengyue. Micro-expression recognition based on multi-scale attention fusion [C]// IEEE International Conference on Data Science and Computer Application. Piscataway, USA: IEEE, 2021: 853-861.
- [13] LI Xinyu, WEI Guangshun, WANG Jie, et al. Multi-scale joint feature network for micro-expression recognition [J]. Computational Visual Media, 2021, 7(3): 407-417.
- [14] CHEN Boyu, ZHANG Zhihao, LIU Nian, et al. Spatiotemporal convolutional neural network with convolutional block attention module for micro-expression recognition [J]. Information, 2020, 11(8): 380.
- [15] WANG Chongyang, PENG Min, BI Tao, et al. Micro-attention for micro-expression recognition [J]. Neurocomputing, 2020, 410: 354-362.
- [16] LI Yante, HUANG Xiaohua, ZHAO Guoying. Micro-

- expression action unit detection with spatial and channel attention [J]. *Neurocomputing*, 2021, 436: 221-231.
- [17] ZHOU Haoliang, HUANG Shucheng, LI Jingting, et al. Dual-ATME: dual-branch attention network for micro-expression recognition [J]. *Entropy*, 2023, 25(3): 460.
- [18] WANG Jie, PAN Xiao, LI Xinyu, et al. Single trunk multi-scale network for micro-expression recognition [J]. *Graphics and Visual Computing*, 2021, 4: 200026
- [19] YANG Jun, WU Zilu, WU Renbiao. Micro-expression spotting based on VoVNet, driven by multi-scale features [J]. *Electronics*, 2023, 12(21): 4459.
- [20] LI Xiaobai, PFISTER T, HUANG Xiaohua, et al. A spontaneous micro-expression database: inducement, collection and baseline [C]// *The 10th IEEE International Conference and Workshops on Automatic Face and Gesture Recognition (FG)*. Piscataway, USA: IEEE, 2013: 1-6.
- [21] YAN Wenjing, LI Xiaobai, WANG Sujing, et al. CASME II: an improved spontaneous micro-expression database and the baseline evaluation [J]. *PLoS One*, 2014, 9(1): e86041.
- [22] DAVISON A K, LANSLEY C, COSTEN N, et al. SAMM: a spontaneous micro-facial movement dataset [J]. *IEEE Transactions on Affective Computing*, 2018, 9(1): 116-129.
- [23] ZHOU Ling, MAO Qirong, XUE Luoyang. Dual-inception network for cross-database micro-expression recognition [C]// *The 14th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2019)*. Piscataway, USA: IEEE, 2019: 1-5.
- [24] WOO S, PARK J, LEE J Y, et al. CBAM: convolutional block attention module [C]// *Computer Vision-ECCV 2018*. Cham, Switzerland: Springer International Publishing, 2018: 3-19.
- [25] ZHAO Guoying, PIETIKAINEN M. Dynamic texture recognition using local binary patterns with an application to facial expressions [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2007, 29(6): 915-928.
- [26] LIONG S T, SEE J, WONG K S, et al. Less is more: micro-expression recognition from video using apex frame [J]. *Signal Processing: Image Communication*, 2018, 62: 82-92.
- [27] LEI Ling, CHEN Tong, LI Shigang, et al. Micro-expression recognition based on facial graph representation learning and facial action unit fusion [C]// *IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*. Piscataway, USA: IEEE, 2021: 1571-1580.
- [28] 李帅超, 李明泽, 孙嘉傲, 等. 融合LBP与并行注意力机制的微表情识别方法[J/OL]. *北京航空航天大学学报*. (2023-07-13) [2024-01-01]. <https://doi.org/10.13700/j.bh.1001-5965.2023.0215>.
LI Shuaichao, LI Mingze, SUN Jiaao, et al. Combining LBP and parallel attention mechanism micro expression recognition [J/OL]. *Journal of Beijing University of Aeronautics and Astronautics*. (2023-07-13) [2024-01-01]. <https://doi.org/10.13700/j.bh.1001-5965.2023.0215>. (in Chinese)
- [29] YANG Hongling, XIE Lun, PAN Hang, et al. Multimodal attention dynamic fusion network for facial micro-expression recognition [J]. *Entropy*, 2023, 25(9): 1246.

【中文责编：英子；英文责编：木柯】



补充材料