中图法分类号:TP391　　**文献标识码**: A　　**文章编号**: 1006-8961(2024)01-0111-12

论文引用格式:Zhao M H, Dong S S, Hu J, Du S L, Shi C, Li P and Shi Z H. 2024. Attention-guided three-stream convolutional neural network for microexpression recognition. Journal of Image and Graphics, 29(01):0111-0122(赵明华, 董爽爽, 胡静, 都双丽, 石程, 李鹏, 石争浩. 2024. 注意力引导的三流卷积神经网络用于微表情识别. 中国图象图形学报, 29(01):0111-0122)[DOI:10.11834/jig.230053]

注意力引导的三流卷积神经网络用于微表情识别

赵明华[1,2],董爽爽[1],胡静[1*],都双丽[1],石程[1],李鹏[1],石争浩[1]

1. 西安理工大学计算机科学与工程学院,西安　710048; 2. 陕西省网络计算与安全技术重点实验室,西安　710048

摘　要: **目的**　微表情识别在心理咨询、置信测谎和意图分析等多个领域都有着重要的应用价值。然而,由于微表情自身具有动作幅度小、持续时间短的特点,到目前为止,微表情的识别性能仍然有很大的提升空间。为了进一步推动微表情识别的发展,提出了一种注意力引导的三流卷积神经网络(attention-guided three-stream convolutional neural network, ATSCNN)用于微表情识别。**方法**　首先,对所有微表情序列的起始帧和峰值帧进行预处理;然后,利用TV-L1(total variation-L1)能量泛函提取微表情两帧之间的光流;接下来,在特征提取阶段,为了克服有限样本量带来的过拟合问题,通过3个相同的浅层卷积神经网络分别提取输入3个光流值的特征,再引入卷积块注意力模块以聚焦重要信息并抑制不相关信息,提高微表情的识别性能;最后,将提取到的特征送入全连接层分类。此外,整个模型架构采用SELU(scaled exponential linear unit)激活函数以加快收敛速度。**结果**　本文在微表情组合数据集上进行LOSO(leave-one-subject-out)交叉验证,未加权平均召回率(unweighted average recall, UAR)以及未加权F1-Score(unweighted F1-score, UF1)分别达到了0.735 1和0.720 5。与对比方法中性能最优的Dual-Inception模型相比,UAR和UF1分别提高了0.060 7和0.068 3。实验结果证实了本文方法的可行性。**结论**　本文方法所提出的微表情识别网络,在有效缓解过拟合的同时,也能在小规模的微表情数据集上达到先进的识别效果。

关键词: 微表情识别;光流;三流卷积神经网络;卷积块注意力模块(CBAM);SELU激活函数

Attention-guided three-stream convolutional neural network for microexpression recognition

Zhao Minghua[1,2], Dong Shuangshuang[1], Hu Jing[1*], Du Shuangli[1], Shi Cheng[1], Li Peng[1], Shi Zhenghao[1]

1. School of Computer Science and Engineering, Xi'an University of Technology, Xi'an　710048, China;

2. Shaanxi Key Laboratory of Network Computing and Security Technology, Xi'an　710048, China

Abstract: **Objective** In recent years, microexpression recognition has remarkable application value in various fields such as psychological counseling, lie detection, and intention analysis. However, unlike macro-expressions generated in conscious states, microexpressions often occur in high-risk scenarios and are generated in an unconscious state. They are characterized by small action amplitudes, short duration, and usually affect local facial areas. These features also determine the difficulty of microexpression recognition. Traditional methods used in early research mainly include methods based on

收稿日期:2023-02-16;**修回日期**:2023-04-18;**预印本日期**:2023-04-25

*通信作者:胡静　jinghu@xaut.edu.cn

基金项目:国家重点研发计划资助(2017YFB1402103-3);国家自然科学基金项目(61901363,61901362);陕西省自然科学基金项目(2020JQ-648,2019JM-381,2019JQ-729);陕西省教育厅重点实验室基金项目(20JS086,20JS087)

Supported by: National Key R&D Program of China (2017YFB1402103-3); National Natural Science Foundation of China (61901363, 61901362); Natural Science Foundation of Shaanxi Province, China (2020JQ-648, 2019JM-381, 2019JQ-729); Key Laboratory Foundation of Shaanxi Education Department (20JS086, 20JS087)

local binary patterns and methods based on optical flow. The former can effectively extract the texture features of microexpressions, whereas the latter calculates the pixel changes in the temporal domain and the relationship between adjacent frames, providing rich, key input information for the network. Although the traditional methods based on texture features and optical flow features have made good progress in early microexpression recognition, they often require considerable cost and have room for improvement in recognition accuracy and robustness. Later, with the development of machine learning, microexpression recognition based on deep learning gradually became the mainstream of research in this field. This method uses neural networks to extract features from input image sequences after a series of preprocessing operations (facial cropping and alignment and grayscale processing) and classifies them to obtain the final recognition result. The introduction of deep learning has substantially improved the recognition performance of microexpressions. However, given the characteristics of microexpressions themselves, the recognition accuracy of microexpressions can still be improved considerably, while the limited scale of existing microexpression datasets also restricts the recognition effect of such emotional behaviors. To solve these problems, this paper proposes an attention-guided three-stream convolutional neural network (ATSCNN) for microexpression recognition. **Method** First, considering that the motion changes between adjacent frames of microexpressions are very subtle, to reduce redundant information and computation in the image sequence while preserving the important features of microexpressions, this paper only performs preprocessing operations such as facial alignment and cropping on the two key frames of microexpressions (onset frame and apex frame) to obtain a single-channel grayscale image sequence with a resolution of 128 × 128 pixels and to reduce the influence of nonfacial areas on microexpression recognition. Then, because optical flow can capture representative motion features between two frames of microexpressions, it can obtain a higher signal-to-noise ratio than the original data and provide rich, critical input features for the network. Therefore, this paper uses the total variation-L1 (TV-L1) energy functional to extract optical flow features between two frames of microexpressions (the horizontal component of optical flow, the vertical component of optical flow, and the optical strain). Next, in the microexpression feature extraction stage, to overcome the overfitting problem caused by limited sample size, three identical four-layer convolutional neural networks are used to extract the features of the input optical flow horizontal component, optical flow vertical component, and optical strain, (the input channel numbers of the four convolutional layers are 1, 3, 5, and 8, and the output channel numbers are 3, 5, 8, and 16), thus improving the network performance. Afterward, because the image sequences in the microexpression dataset used in this paper inevitably contain some redundant information other than the face, a convolutional block attention module(CBAM) with channel attention and spatial attention serially connected is added after each shallow convolutional neural network in each stream to focus on the important information of the input and suppress irrelevant information, while paying attention to both the channel dimension and the spatial dimension, thereby enhancing the network's ability to obtain effective features and improving the recognition performance of microexpressions. Finally, the extracted features are fed into a fully connected layer to achieve microexpression emotion classification (including negative, positive, and surprise). In addition, the entire model architecture uses the scaled exponential linear unit (SELU) activation function to overcome the potential problems of neuron death and gradient disappearance in the commonly used rectified linear unit (ReLU) activation function to speed up the convergence speed of the neural network. **Result** This paper conducted experiments on the microexpression combination dataset using the leave-one-subject-out (LOSO) cross-validation strategy. In this strategy, each subject serves as the test set, and all remaining samples are used for training. This validation method can fully utilize the samples and has a certain generalization ability. This method is the most commonly used validation in current microexpression recognition research. The results of this paper's experiments on the unweighted average recall (UAR) and unweighted F1-score (UF1) reached 0.735 1 and 0.720 5, respectively. Compared with the Dual-Inception model, which performed best in the comparative methods, UAR and UF1 increased by 0.060 7 and 0.068 3, respectively. To verify further the effectiveness of the ATSCNN neural network architecture proposed in this paper, several ablation experiments were also conducted on the combined dataset, and the results confirmed the feasibility of this paper's method. **Conclusion** The microexpression recognition network proposed in this paper can effectively alleviate overfitting, focus on important information of microexpressions, and achieve state-of-the-art (SOTA) recognition performance on small-scale microexpression datasets using LOSO cross-validation. Compared with other mainstream models, the proposed method achieved state-of-the-art recognition performance. In addition, the results

of several ablation experiments made the proposed method more convincing. In conclusion, the proposed method remarkably improved the effectiveness of microexpression recognition.

Key words: microexpression recognition; optical flow; three-stream convolution neural network; convolutional block attention module(CBAM); SELU activation function

0 引 言

心理学家 Albert Mehrabian 的研究表明：面部表情在传递信息的过程中占据大约 55% 的比例(Mehrabian，1965)，远远高于语言沟通和肢体动作。因此，通过对面部表情进行分析可以很大程度上了解人们的情绪与心理。

面部表情按照持续时间和动作强度可以分为宏表情和微表情(Lu 等，2022)。近年来，学者们对宏表情的研究已经非常成熟，在很多领域都有实际应用。宏表情有两个明显的特点：一方面，宏表情通常是自发的且在有意识的状态下产生，适用于各种场景；另一方面，其动作幅度大，可作用于全脸。与宏表情不同，微表情有 3 个显著特点：1)往往在高风险场景中于无意识状态下触发且难以掩饰；2)动作强度低，仅作用于局部人脸区域(徐峰和张军平，2017)；3)持续时间很短，大约在 0.04～0.2 s 之间(Porter 和 Ten Brinke，2008；Zhang 等，2014)，人眼难以察觉。所以，相比宏表情，微表情的研究面临着更大的挑战。

鉴于微表情自身的特点，人工识别微表情存在很大的困难，即使是经过专业培训的心理学研究人员对微表情的识别准确率也仅仅约 47%(Frank 等，2009)。随着计算机视觉和机器学习技术的快速发展，越来越多的研究人员将机器学习算法应用到微表情识别问题中，解决了人工识别存在的很多困难，识别准确率也有了明显提高。但是，目前微表情识别在计算机视觉领域仍然是一个比较前沿的工作，针对这方面的研究主要存在两大挑战：1)由于微表情的数据采集与鉴定很不容易，现有的微表情数据集稀缺(Yan 等，2013，2014；Qu 等，2018；Li 等，2023；Li 等，2013；Davison 等，2018)，这使得深度学习在微表情识别中的应用存在困难；2)微表情持续时间短、动作强度低的特点导致特征难以提取，因此需要进行合适的数据预处理与特征提取操作。

本文主要针对上述微表情识别存在的第 2 个问题展开研究，采用 CASME(Chinese Academy of Sciences microexpression)、CASME II 和 SAMM(spontaneous actions and micromovements)的组合数据集(Yan 等，2013，2014；Davison 等，2018)，以起始帧和峰值帧的 3 个光流信息作为输入，设计并提出了一种注意力引导的浅层三流卷积神经网络(attention-guided three-stream convolutional neural network，ATSCNN)来识别微表情。本文的主要贡献概括如下：1)设计并提出了一种注意力引导的三流卷积神经网络(ATSCNN)用于微表情识别，可以有效防止过拟合、提升网络性能。2)使用 ReLU(rectified linear unit)激活函数不仅计算简单，而且能够获得强大的学习和拟合能力。但是该函数可能会造成神经元坏死和梯度消失的情况。因此，本文尝试使用 SELU(scaled exponential linear unit)激活函数来解决这一问题。3) 在 CASME、CASME II 和 SAMM 组合数据集中，通过实验验证了本文方法在未加权平均召回率(unweighted average recall，UAR)和未加权 F1-Score(unweighted F1-score，UF1)这两个评估指标下的有效性。与现有的主流模型相比，本文方法获得了先进的识别性能。

1 相关工作

自 Haggard 和 Isaacs(1966)在心理治疗研究中首次注意到微表情的存在之后，学术界针对微表情的研究逐步深入。然而直到 2009 年，自动微表情识别才开始出现在计算机视觉领域(Polikovsky 等，2009)，相比宏表情识别的研究要晚很多。

早期的研究以基于局部二值模式(local binary patterns，LBP)和光流的方法作为微表情识别的主要技术手段。Pfister 等人(2011)首次将来自 3 个正交平面的局部二值模式(LBP from three orthogonal planes，LBP-TOP)应用于微表情识别任务，能够有效提取时空域的纹理特征。此后，LBP-TOP 作为一种经典算法，为后续研究提供基础和验证基准。基于

光流的方法计算像素在时域中的变化以及相邻帧之间的相关性，从而找到前一帧和当前帧的对应关系。为了应对微表情持续时间短、动作强度低的挑战，Chaudhry等人(2009)提出了定向光流直方图(histograms of oriented optical flow, HOOF)，大多数基于光流的方法都是HOOF的变体。此外，Liong等人(2018)提出了加权定向光流(bi-weighted oriented optical flow，Bi-WOOF)，选择图像序列中的起始帧和峰值帧并提取这两帧之间的光流，对HOOF特征进行全局或局部加权。此后很多工作都利用起始帧和峰值帧来对微表情进行分析，以减少整个图像序列中可能存在的冗余现象。

近年来，深度学习已经成为人们研究的热点，很多学者开始尝试将深度学习应用于微表情识别。在第1届微表情挑战赛(Yap等，2018)中，Khor等人(2018)使用富集长期递归卷积网络(enriched long-term recurrent convolutional network，ELRCN)来预测微表情。通过光流信息(光流水平分量、光流垂直分量和光学应变)学习空间关系，并利用VGG16(Visual Geometry Group)和预训练的VGGFace模型来丰富输入通道和深层特征叠加。该论文在深度微表情识别工作中具有代表性，但是其相比传统方法仍然不具备优势。此后的第2届微表情挑战赛(See等，2019)中，深度学习方法在微表情识别任务中已经初步显示出相比传统方法的优越性：Off-ApexNet(optical flow features from apex frame network)(Gan等，2019)网络及改进版本浅层三流三维卷积神经网络(shallow triple stream three-dimensional CNN，STSTNet)(Liong等，2019)从每个视频的起始帧和峰值帧提取光流特征，然后将得到的光流特征送入浅层卷积神经网络中分类，由于网络层数较浅，减少了因微表情数据稀缺造成的过拟合；EMR(expression magnification and reduction)(Liu等，2019)采用迁移学习的方法，提取微表情起始帧和峰值帧之间的光流，利用两种域自适应策略达到SOTA(state-of-the-art)实验结果；另外，Dual-Inception网络(Zhou等，2019)同样是提取起始帧和峰值帧的光流特征，之后再送入两个Inception网络拼接并分类，在CASME II数据集上的表现最好；而CapsuleNet(capsule network)网络(van Quang等，2019)仅将峰值帧的像素值输入到预训练的ResNet18(residual network)(He等，2016)中得到特征图，再将所有特征图输入到胶囊网络中提取特征并分类，最终结果在整体上优于LBP-TOP基准网络和几种比较强大的卷积神经网络模型。多尺度卷积神经网络(multi-scale convolutional neural network，MACNN)模型(Lai等，2020)利用残差块和空洞卷积在低响应时间内对微表情进行分类；唐宏等人(2022)基于光流信息并采用伪三维残差网络学习微表情的时空特征，识别性能优于一些较先进的方法。上述几种网络模型都是以关键帧的光流值作为输入，并且取得了不错的效果。基于此，本文考虑将起始帧和峰值帧这两个关键帧的光流信息作为输入，观察在利用浅层卷积网络进行特征提取的优势。

上述几种神经网络架构均比较简单，没有用到复杂的模块(如注意力机制)。注意力机制最初用于计算机视觉，之后逐渐广泛应用于自动问答、情感分类等机器学习任务中。注意力模块可以给输入的不同部分分配不同的权重。卷积块注意力模块(convolutional block attention module，CBAM)从通道维度和空间维度进行双重特征权重标定，其根据多方向的特征来改善网络的识别效果(Woo等，2018)。后来，Chen等人(2020)提出用于微表情识别的带有CBAM的时空卷积神经网络，该网络以自适应的方式分配特征权重。Micro-attention(Wang等，2020)将微注意力和残差网络相结合，设法通过增加微注意力模块来提高ResNet在微表情识别中的性能。Off-TANet(optical flow feature-triplet attention net)(Zhang等，2021)是一个基于光流的包含空间注意力、通道注意力和自注意力的轻量级神经网络，利用三重注意力机制改善微表情识别效果。mini-AORCNN(mini-attention-based optical flow residual convolutional neural network)(张嘉淏 等，2022)将残差模块与Bottleneck Transformer相结合，可以有效避免过拟合并减少模型的参数复杂度。以上几种方法中都用到了注意力机制，在识别性能方面均表现不错，说明注意力机制在微表情识别中的应用是可行的。所以，本文尝试将注意力机制引入微表情识别中，以期获得良好的结果。

2 本文方法

本文的微表情识别算法框架如图1所示。主要包括两个步骤：光流特征提取和神经网络学习。首

先，为了便于提取微表情特征并减少非面部区域的影响，对原始图像进行简单的预处理，并利用（total variation-L1）能量泛函（Pérez 等，2013）提取光流信息；然后，将光流水平分量、光流垂直分量和光学应变这 3 个光流值分别送入 ATSCNN 中提取微表情特征并得到分类结果。

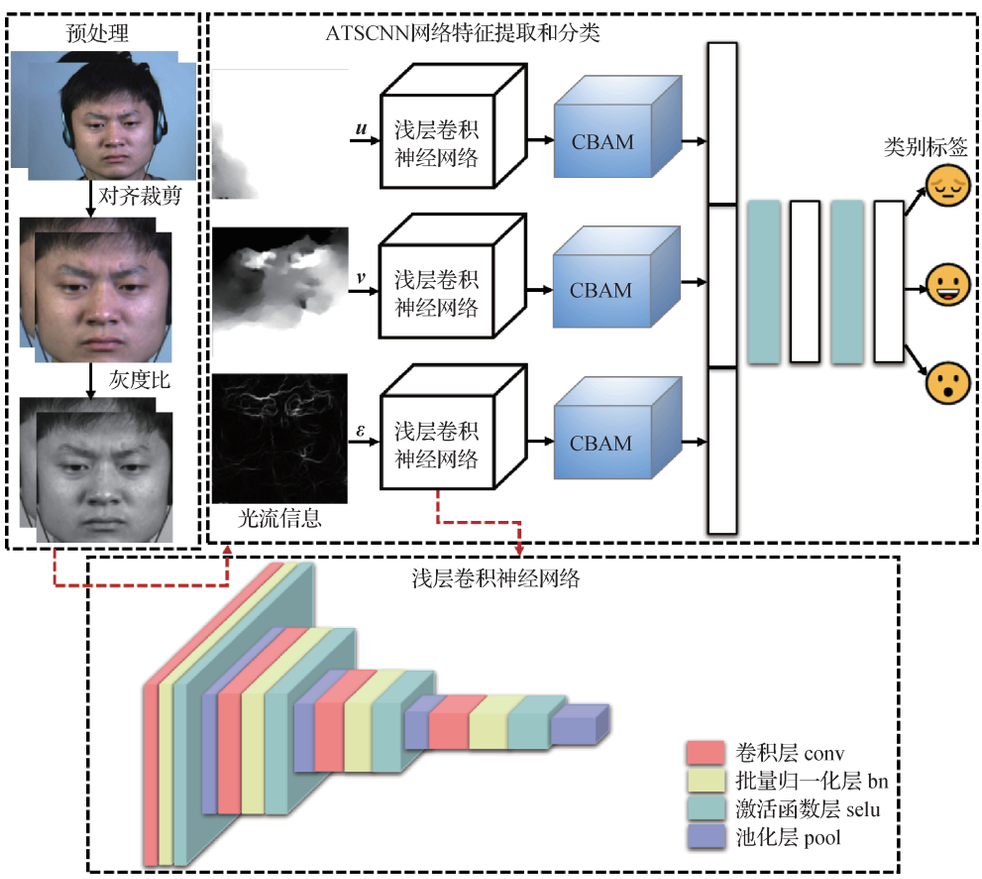

图 1　算法框架流程图

Fig. 1　Algorithm framework flow chart

2.1　光流特征提取

光流法通常基于两个前提假设：1）前后帧的光照亮度保持恒定；2）同一像素在相邻帧之间的运动变化很小。对于微表情而言，一方面，其从起始帧到偏移帧之间的时间间隔很短，所以帧间图像序列的亮度基本保持不变；另一方面，微表情相邻像素之间存在相似的运动，帧之间的相对位移很小。基于此，微表情的特点决定了其满足上述两个基本假设。

虽然微表情的持续时间短、变化速度快，但光流仍然可以捕捉到微表情相邻帧之间具有代表性的运动特征。与原始数据相比，其能够获得更高的信噪比，为网络提供丰富且关键的输入信息。即，光流对面部区域微弱运动的估计很大程度上决定了微表情识别的准确性。本文拟使用噪声鲁棒性较好的 TV-L1 能量泛函进行光流估计。设 $\boldsymbol{u}$ 和 $\boldsymbol{v}$ 分别代表光流场的水平分量和垂直分量。其可以表示为

$$\boldsymbol{u}=\frac{\mathrm{d}x}{\mathrm{d}t},\ \boldsymbol{v}=\frac{\mathrm{d}y}{\mathrm{d}t} \tag{1}$$

式中，$\mathrm{d}x$ 和 $\mathrm{d}y$ 分别表示沿 x 和 y 维度的像素变化，$\mathrm{d}t$ 表示时间变化。

光学应变能够近似面部变形强度，通过计算光流的导数可以得到光学应变，并且可以定义为

$$\boldsymbol{\varepsilon}=\frac{1}{2}\left[\nabla\boldsymbol{U}+\nabla(\boldsymbol{U})^{\mathrm{T}}\right] \tag{2}$$

式中，$\boldsymbol{U}=[\boldsymbol{u},\boldsymbol{v}]^{\mathrm{T}}$ 是位移向量，因此 $\boldsymbol{\varepsilon}$ 也可以表示为

$$\boldsymbol{\varepsilon}=\begin{bmatrix}\varepsilon_{xx}=\dfrac{\partial\boldsymbol{u}}{\partial x} & \varepsilon_{xy}=\dfrac{1}{2}\left(\dfrac{\partial\boldsymbol{u}}{\partial y}+\dfrac{\partial\boldsymbol{v}}{\partial x}\right)\\ \varepsilon_{yx}=\dfrac{1}{2}\left(\dfrac{\partial\boldsymbol{v}}{\partial x}+\dfrac{\partial\boldsymbol{u}}{\partial y}\right) & \varepsilon_{yy}=\dfrac{\partial\boldsymbol{v}}{\partial y}\end{bmatrix} \tag{3}$$

式中，对角线应变分量中，$(\varepsilon_{xx},\varepsilon_{yy})$ 是法向应变分量、$(\varepsilon_{xy},\varepsilon_{yx})$ 是剪切应变分量。可以通过取法向应变分量和剪切应变分量的平方和来计算每个像素的光学

应变大小，从而有

$$|\boldsymbol{\varepsilon}| = \sqrt{\varepsilon_{xx}^2 + \varepsilon_{yy}^2 + \varepsilon_{xy}^2 + \varepsilon_{yx}^2} \tag{4}$$

最后，将$\boldsymbol{u}$，$\boldsymbol{v}$，$\boldsymbol{\varepsilon}$这3个包含丰富微表情的光流信息作为3个输入流分别送入融合了CBAM注意力机制的浅层三流卷积神经网络中。

2.2 浅层三流卷积神经网络

为了避免过拟合，深度卷积神经网络通常需要大量样本进行训练，然而，目前可用于微表情识别的样本量十分有限。为了在有限的数据集中训练卷积神经网络，本文采用由4个子网络组成的浅层三流卷积神经网络进行特征提取，每个子网络包括卷积层（conv）、批量归一化层（bn）、激活函数层（selu）和池化层（pool）。利用浅层三流卷积神经网络提取输入3个光流信息的特征，以缓解过拟合，改善网络性能。表1显示了每个子网络的具体配置。

表1 每个子网络的具体配置

Table 1 The specific configuration of each sub-network

层	输入形状	输出形状	内核大小
conv1	1×128×128	3×128×128	5
bn1	3×128×128	3×128×128	–
selu1	3×128×128	3×128×128	–
pool1	3×128×128	3×64×64	2
conv2	3×64×64	5×64×64	5
bn2	5×64×64	5×64×64	–
selu2	5×64×64	5×64×64	–
pool2	5×64×64	5×32×32	2
conv3	5×32×32	8×32×32	5
bn3	8×32×32	8×32×32	–
selu3	8×32×32	8×32×32	–
pool3	8×32×32	8×16×16	2
conv4	8×16×16	16×16×16	5
bn4	16×16×16	16×16×16	–
selu4	16×16×16	16×16×16	–
pool4	16×16×16	16×8×8	2

注：“–”表示不涉及内核大小。

2.3 CBAM

传统的卷积网络模型可能无法对微小的特征进行有效的捕捉和学习，从而限制了模型的识别性能。此外，微表情通常在时空上分布不均匀，这也需要网络能够有效关注到重要的时间和空间区域。基于上述分析，本文在卷积神经网络之后引入了CBAM注意力机制。

CBAM（Woo等，2018）是一种轻量级卷积块注意力模块，可以集成到任何卷积神经网络中，且能与卷积神经网络一起进行端到端训练。如图2所示，CBAM将通道注意力和空间注意力两个子模块串联，同时关注通道维度和空间维度的信息，以提高网络获取有效特征的能力。下面将详细说明本文使用的CBAM中通道注意力和空间注意力这两个子模块。

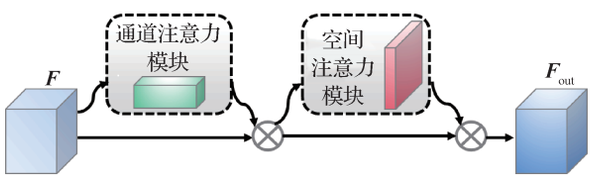

图2 卷积块注意力模块

Fig. 2 Convolutional block attention module

通道注意力模块主要关注什么样的特征是有意义的，让神经网络自动判断哪个通道重要，从而帮助网络提取包含有用信息的通道。如图3所示，为了完成特征提取，减少数据丢失，首先，通道注意力机制同时使用平均池化操作和最大池化操作这两种方式将输入特征图$\boldsymbol{F}$在空间维度进行压缩，保留通道信息，得到两个空间维度为1的特征图；然后，将这两个空间维度为1的特征图分别送入共享的多层感知器（shared multilayer perceptron，Shared MLP）中进行处理，得到两个特征图并将其相加；最后，通过sigmoid激活函数，得到最终输出的特征图$\boldsymbol{M}_C$。通道注意力具体可以表示为

$$\boldsymbol{M}_C(\boldsymbol{F}) = \delta(MLP(AvgPool(\boldsymbol{F})) + MaxPool(\boldsymbol{F}))) \tag{5}$$

式中，δ表示sigmoid激活函数。

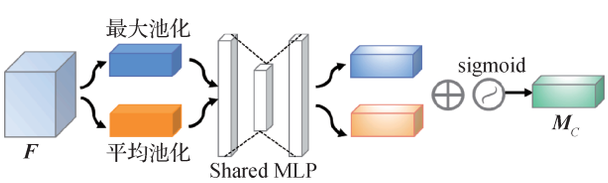

图3 通道注意力模块

Fig. 3 Channel attention module

空间注意力模块主要关注空间中哪部分的特征是有意义的。如图4所示，首先对原始输入的特征图$\boldsymbol{F}'$分别进行一个通道维度的最大池化和平均池

化，然后将得到的结果按照通道维度拼接成两个通道，再经过一个 3 × 3 的卷积层降为一个通道，最后通过 sigmoid 激活函数得到最终的特征图 $\boldsymbol{M}_S$。空间注意力具体可以表示为

$$\boldsymbol{M}_S(\boldsymbol{F}') = \delta(f^{3\times3}([AvgPool(\boldsymbol{F}'), MaxPool(\boldsymbol{F}')])) \quad (6)$$

式中，δ 表示 sigmoid 激活函数，$f^{3\times3}$ 为 3 × 3 的卷积运算。

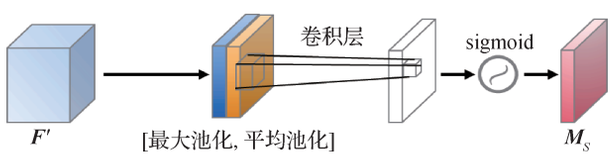

图 4　空间注意力模块

Fig. 4　Spatial attention module

本文将 CBAM 模块添加在浅层卷积神经网络之后，同时关注通道维度和空间维度，期望聚焦输入的重要信息且抑制不相关信息，从而提升微表情的识别效果。

2.4　SELU 激活函数

激活函数（Scardapane 等，2020）在神经网络中主要用来完成数据的非线性变换，通过加入非线性激活函数促使网络获得更强的学习和拟合能力。典型的 ReLU（rectified linear unit）激活函数（Yarotsky，2017）就是其中之一。其主要包括两个优点：一方面，该函数只需要一个阈值就能计算得到激活值，非常简单；另一方面，ReLU 会使一部分神经元的输出为 0，减少参数间的相互依赖关系，能缓解过拟合。但是，ReLU 仍有其局限性。当函数在负半轴上的值都为 0 时，负梯度经过 ReLU 单元不会被激活，导致该神经元坏死；且 ReLU 在正半轴的导数恒为 1，当梯度过小的时候，经过该函数容易导致梯度消失。ReLU 以及 ReLU 的导数可以分别表示为

$$ReLU(x) = \begin{cases} x & x > 0 \\ 0 & x \leqslant 0 \end{cases} \quad (7)$$

$$ReLU'(x) = \begin{cases} 1 & x > 0 \\ 0 & x \leqslant 0 \end{cases} \quad (8)$$

SELU 激活函数（Klambauer 等，2017）是基于 ReLU 的一种变体，其正半轴导数 $\lambda > 1$，因此可以在方差过小的时候通过乘以 λ 增大方差，能有效避免梯度爆炸和梯度消失现象；同时，在负半轴不再简单地用 0 来表示，而是采用两个超参数 λ 和 α 进行调整，从而解决 ReLU 中可能存在的神经元坏死问题，增强了网络的表现力和泛化性能。SELU 以及 SELU 的导数可以分别表示为

$$SELU(x) = \lambda \begin{cases} x & x > 0 \\ \alpha(e^x - 1) & x \leqslant 0 \end{cases}, \quad \lambda > 1 \quad (9)$$

$$SELU'(x) = \lambda \begin{cases} 1 & x > 0 \\ \alpha e^x & x \leqslant 0 \end{cases}, \quad \lambda > 1 \quad (10)$$

式中，超参数 λ 和 α 的值是给定的，即：$\lambda \approx 1.050\,7$，$\alpha \approx 1.673\,26$。

因此，本文尝试使用 SELU 激活函数以弥补常用的 ReLU 激活函数中存在的问题，提高神经网络的收敛速度。

3　实验结果与分析

3.1　数据集与实验设置

3.1.1　数据集

CASME（Yan 等，2013）包含 19 名受试者自发产生的 195 个微表情样本，其中包括从起始到偏移的所有帧。CASME 由高速摄像机以 60 帧/s 的速度记录，样本分辨率为 1 280 × 720 像素（section A）或 640 × 480 像素（section B）。CASME 数据集中的样本分为 8 个微表情情感类别（“happiness”、“surprise”、“disgust”、“repression”、“sadness”、“contempt”、“fear”和“tense”）。情绪的标注部分基于 AUs（action units），同时考虑了受试者的自我报告和视频片段的内容。除了起始帧和偏移帧，也标记了贡献情绪最多的峰值帧。

CASME II（Yan 等，2014）是 CASME 数据集的改进版本。其微表情样本增加到 255 个，包含 26 名有效的受试者，由高速摄像机以 200 帧/s 记录，样本分辨率为 640 × 480 像素。因此，与 CASME 相比，CASME II 具有更高的时间和空间分辨率。样本标记了起始帧、峰值帧和偏移帧。情绪类别包括：“happiness”、“surprise”、“disgust”、“repression”、“fear”、“others”和“sadness”，共 7 类。

SAMM（Davison 等，2018）从 32 名受试者中收集了 159 个微表情样本。与之前缺乏种族多样性的数据集不同，受试者来自 13 个不同的种族。样本是在受控照明条件下以 200 帧/s 的速度采集的，分辨率为 2 040 × 1 088 像素。实验过程中捕捉到的微表情类别包括：“happiness”、“surprise”、“anger”、“contempt”、“disgust”、“fear”、“sadness”和“other”。样本

基于AUs标记了起始帧、峰值帧和偏移帧。

本文使用上述3个数据集中包含丰富微表情的样本得到新的组合数据集进行实验。舍弃了CASME(Yan等,2013)中情绪类别为“contempt”、“fear”和“tense”的所有样本、CASME II(Yan等,2014)中情绪类别为“fear”、“others”和“sadness”的所有样本及SAMM(Davison等,2018)中的“other”样本。为了缓解所使用的数据集之间存在的类别不平衡问题,本文将原始情绪标签映射到“negative”、“positive”和“surprise”3个类别中。表2列出了本文使用的微表情组合数据集的分类信息。

为了最大程度上减少图像序列中非面部区域对微表情识别的影响,本文利用dlib库提供的68个人脸关键点检测的模型来实现面部对齐操作,并将样本分辨率统一调整为128×128像素。由于颜色容易受到光照影响,造成RGB变化较大;且三通道转换为单通道,无需进行加和操作,计算量大大减少。因此,在提取光流特征之前,将RGB图像转换为灰度图以减少计算量。另外,考虑到微表情相邻帧之间的变化非常微小,为了减少冗余,本文选取每个样本的起始帧和微表情变化最大的峰值帧进行处理。

表2　微表情组合数据集的情绪分类信息

Table 2　Emotion classification information of micro expression combination dataset

微表情类别	CASME	CASME II	SAMM	合计
negative	82	88	92	262
positive	7	32	26	65
surprise	15	25	15	55
合计	104	145	133	382

3.1.2　数据集性能评估指标

由于表1中选取的组合数据集情绪类别依然具有不平衡性,使用传统的性能评估指标(如:精确度)将会导致过拟合。因此,本文使用UAR和UF1这两个性能度量指标以减少所提方法潜在的偏差。

UF1也称为宏观平均F1分数,通过对每类F1分数进行平均来确定。其中,F1分数可以解释为精度和召回率的加权平均值,当F1分数等于1时达到最佳值,等于0时达到最差值。精度(precision,P)和召回率(recall,R)对F1分数的相对贡献是一样的。F1分数可表示为

$$F1 = \frac{2 \times P \times R}{P + R} \tag{11}$$

$$P = \sum_{i=1}^{C} \frac{\sum_{j=1}^{S} TP_i^j}{\sum_{j=1}^{S} TP_i^j + \sum_{j=1}^{S} FP_i^j} \tag{12}$$

$$R = \sum_{i=1}^{C} \frac{\sum_{j=1}^{S} TP_i^j}{\sum_{j=1}^{S} TP_i^j + \sum_{j=1}^{S} FN_i^j} \tag{13}$$

那么,UF1可以表示为

$$UF1 = \frac{1}{|C|} \sum_{i=1}^{C} \frac{2 \times P_i \times R_i}{P_i + R_i} \tag{14}$$

$$P_i = \frac{\sum_{j=1}^{S} TP_i^j}{\sum_{j=1}^{S} TP_i^j + \sum_{j=1}^{S} FP_i^j} \tag{15}$$

$$R_i = \frac{\sum_{j=1}^{S} TP_i^j}{\sum_{j=1}^{S} TP_i^j + \sum_{j=1}^{S} FN_i^j} \tag{16}$$

UAR可以表示为每个类别的平均精度除以类的数量,而不考虑每个类的样本。UAR能减少由于类别不平衡而引起的偏差,也称为平衡精度。UAR具体可以表示为

$$UAR = \frac{1}{|C|} \sum_{i=1}^{C} \frac{\sum_{j=1}^{S} TP_i^j}{\sum_{j=1}^{S} TP_i^j + \sum_{j=1}^{S} FN_i^j} \tag{17}$$

式(12)—式(17)中,C是所划分的情绪类别数($C=3$),S是受试者个数($S=67$)。TP(true positive)表示真阳性,即实际是正样本、分类器预测也是正样本。FN(false negative)表示假阴性,即实际是正样本,分类器预测是负样本。FP(false positive)表示假阳性,即实际是负样本、分类器预测是正样本。

3.1.3　实验设置

本文网络模型基于Python3.7.15版本、利用PyTorch1.13.0框架训练,显存为10 GB。本文方法使用的初始学习率为0.000 1,batch_size设置为64,epoch为200。在训练模型时,使用交叉熵损失函数和自适应矩估计优化器(Adam)优化。

本文采用留一被试(leave-one-subject-out,LOSO)交叉验证策略,其中,每个受试者都被作为测试集,所有剩余的样本用于训练。本文使用的组合

数据集一共包括 67 个受试者(15 个受试者来自 CASME,24 个受试者来自 CASME II,28 个受试者来自 SAMM),因此,一轮实验需要重复评估 67 次。这种验证方法能充分利用样本并且具有一定的泛化能力,在目前微表情识别的研究工作中普遍应用。

3.2 算法比较

将本文方法与其他主流方法进行对比,在组合数据集上得到的结果如表 3 所示,表 3 中包括性能评价指标 UAR 和 UF1。此外,还提供了所有模型的总参数量(total params)、总浮点运算次数(total flops)和总内存读写量(total MemR+W)。表 3 中涉及的所有实验结果均是重新实现的。

根据表 3 中的结果可以看出,本文提出的 ATSCNN 模型在组合数据集上获得的 UAR(0.735 1)和 UF1(0.720 5)性能最优。具体来讲,与应用浅层卷积网络的 STSTNet(shallow triple stream three-dimensional CNN)(Liong 等,2019)相比,UAR 提高了 0.074 8,UF1 提高了 0.077 1,证明本文在设计的卷积网络基础上引入注意力机制效果显著。与使用两个 Inception 网络进行特征提取与分类的 Dual-Inception(Zhou 等,2019)相比,UAR 提高了 0.060 7,UF1 提高了 0.068 3。一方面说明光学应变信息对于提取面部特征的重要性;另一方面也说明了注意力机制对重要信息的关注度。与采用残差块和空洞卷积的 MACNN(multi-scale convolutional neural network)(Lai 等,2020)模型相比,UAR 提高了 0.143 1,UF1 提高了 0.189 6,识别性能大幅提升,进一步证实了本文方法的有效性。与结合微注意力和残差网络的 Micro-Attention(Wang 等,2020)模型相比,UAR 提高了 0.111 4,UF1 提高了 0.129 2,说明本文方法将注意力与浅层卷积网络相结合在特征提取方面的效果更好。与基于三重注意力机制的微表情识别网络 Off-TANet(optical flow feature-triplet attention network)(Zhang 等,2021)相比,UAR 提高了 0.084 7,UF1 提高了 0.096 6,这可能得益于本文提出的浅层卷积网络在有限样本量上较好的特征提取能力。与融合了残差模块和 Bottleneck Transformer 的 mini-AORCNN(mini-attention-based optical flow residual convolutional neural network)(张嘉淏 等,2022)网络模型相比,UAR 提高了 0.075 0,UF1 提高了 0.085 0,虽然该方法的参数量是几个对比方法中最少的,但是在识别性能上仍然低于本文方法。综上所述,本文提出的 ATSCNN 模型在对比方法中取得了最好的微表情识别效果。

表 3 本文方法与主流算法的对比

Table 3 Comparison between this method and mainstream algorithms

网络模型	UAR	UF1	总参数量	总浮点运算次数/MFlops	总内存读写量/MB
STSTNet(Liong 等,2019)	0.660 3	0.643 4	4.98×10^7	57.45	192.98
Dual-Inception(Zhou 等,2019)	0.674 4	0.652 2	1.34×10^8	264.15	546.62
MACNN(Lai 等,2020)	0.592 0	0.530 9	8.63×10^7	1 040	366.72
Micro-Attention(Wang 等,2020)	0.623 7	0.591 3	1.88×10^7	327.35	82.89
Off-TANet(Zhang 等,2021)	0.650 4	0.623 9	67 083	39.29	7.33
mini-AORCNN(张嘉淏 等,2022)	0.660 1	0.635 5	**46 251**	**16.39**	**4.84**
ATSCNN(本文)	**0.735 1**	**0.720 5**	4.21×10^6	19.01	20.5

注:加粗字体表示各列最优结果。

3.3 消融实验

如表 4 所示,为了进一步证实本文 ATSCNN 架构的有效性,使用 LOSO 交叉验证方法设计了 6 组对比实验。

实验 1 中,不添加注意力机制,仅使用浅层三流卷积神经网络进行实验,对比发现本文引入 CBAM 注意力机制明显提高了微表情识别结果:UAR 增加了 0.035 8,UF1 增加了 0.045 7。

实验 2 中,仅利用通道注意力模块进行实验,结果表明引入通道注意力模块对识别效果有一定的提升:UAR 增加了 0.008 5,UF1 增加了 0.012 8。

实验 3 中,只使用空间注意力模块进行实验,以证明同时关注通道和空间两个维度确实能够提升识别性能:UAR 增加了 0.012 7,UF1 增加了 0.023 5。

表4　消融实验结果

Table 4　Results of ablation experiments

实验	方法	UAR	UF1
实验1	卷积 + SELU	0.699 3	0.674 8
实验2	卷积 + 通道注意力 + SELU	0.726 6	0.707 7
实验3	卷积 + 空间注意力 + SELU	0.722 4	0.697 0
实验4	卷积 + CBAM + ReLU	0.690 0	0.668 5
实验5	卷积 + CBAM + SELU(无光学应变信息)	0.722 8	0.696 6
实验6	卷积 + CBAM + SELU(无光流信息)	0.592 1	0.530 9
本文	卷积 + CBAM + SELU	**0.735 1**	**0.720 5**

注:加粗字体表示各列最优结果。

实验4中,利用ReLU激活函数进行实验,发现相比ReLU,本文使用SELU激活函数能够提高神经网络的收敛速度,识别性能显著提升:UAR增加了0.045 1,UF1增加了0.052 0。原因是SELU激活函数解决了ReLU激活函数中可能会存在的神经元坏死以及梯度消失问题。

实验5中,不计算光流的导数,即不添加光学应变,只利用光流水平分量和光流垂直分量两种光流信息,发现本文增加光学应变这一输入流可以很好地近似面部变形强度,微表情识别效果有一定幅度提升:UAR增加了0.012 3,UF1增加了0.023 9。

实验6中,只利用单通道灰度图作为输入,不计算光流信息,从而证明利用光流信息能够提供更加丰富的输入,获得与原始数据相比更高的信噪比,从而显著增强后续网络的特征提取能力:UAR增加了0.143 0,UF1增加了0.189 6。

4　结　论

微表情自身运动幅度微弱、持续时间短暂的特性决定了对其进行特征提取的难度。本文提出了一种基于光流的用于微表情识别研究的神经网络架构ATSCNN,由融合了CBAM注意力机制的浅层三流卷积神经网络组成,在有限的样本量下实现了对微表情的高效特征提取。其中,浅层卷积神经网络主要用于缓解小数据集上存在的过拟合问题,同时,考虑到浅层卷积神经网络在捕捉微小特征方面的局限性,引入CBAM注意力机制以聚焦重要特征并抑制无关特征。此外,在整个模型中采用SELU激活函数替代常用的ReLU激活函数,明显提高了神经网络的表达能力和收敛速度。本文采用LOSO交叉验证策略,在使用的组合数据集(CASME、CASME II和SAMM)中测试了ATSCNN模型,并与现有基于深度学习的6个主流模型进行对比,UAR和UF1都达到了最优结果,证实了所提方法在识别性能方面具有一定的优势。同时,执行6组消融实验以进一步确保本文网络模型在微表情识别中的有效性。

虽然本文方法取得了不错的识别效果,但是距离投入工程应用仍需进一步研究与探索。因此,未来的工作将着力解决微表情数据集中样本稀缺以及类别不平衡的问题,使现有方法可以呈现出更好的识别效果。

参考文献(References)

Chaudhry R, Ravichandran A, Hager G and Vidal R. 2009. Histograms of oriented optical flow and Binet-Cauchy kernels on nonlinear dynamical systems for the recognition of human actions//Proceedings of 2009 IEEE Conference on Computer Vision and Pattern Recognition. Miami, USA: IEEE: 1932-1939 [DOI: 10.1109/CVPR.2009.5206821]

Chen B Y, Zhang Z H, Liu N, Tan Y, Liu X Y and Chen T. 2020. Spatiotemporal convolutional neural network with convolutional block attention module for micro-expression recognition. Information, 11(8): #380 [DOI: 10.3390/info11080380]

Davison A K, Lansley C, Costen N, Tan K and Yap M H. 2018. SAMM: a spontaneous micro-facial movement dataset. IEEE Transactions on Affective Computing, 9(1): 116-129 [DOI: 10.1109/TAFFC.2016.2573832]

Frank M, Herbasz M, Sinuk K, Keller A and Nolan C. 2009. I see how you feel: training laypeople and professionals to recognize fleeting

emotions//Proceedings of 2009 Annual Meeting of the International Communication Association. Sheraton, USA: [s.n.]: 1-35

Gan Y S, Liong S T, Yau W C, Huang Y C and Tan L K. 2019. Off-ApexNet on micro-expression recognition system. Signal Processing: Image Communication, 74: 129-139 [DOI: 10.1016/j.image.2019.02.005]

Haggard E A and Isaacs K S. 1966. Micromomentary facial expressions as indicators of ego mechanisms in psychotherapy//Gottschalk L A, Auerbach A H, eds. Methods of Research in Psychotherapy. New York, USA: Springer: 154-165 [DOI: 10.1007/978-1-4684-6045-2_14]

He K M, Zhang X Y, Ren S Q and Sun J. 2016. Deep residual learning for image recognition//Proceedings of 2016 IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas, USA: IEEE: 770-778 [DOI: 10.1109/CVPR.2016.90]

Khor H Q, See J, Phan R C W and Lin W Y. 2018. Enriched long-term recurrent convolutional network for facial micro-expression recognition//Proceedings of the 13th IEEE International Conference on Automatic Face and Gesture Recognition. Xi'an, China: IEEE: 667-674 [DOI: 10.1109/FG.2018.00105]

Klambauer G, Unterthiner T, Mayr A and Hochreiter S. 2017. Self-normalizing neural networks//Proceedings of the 31st International Conference on Neural Information Processing Systems. Long Beach, USA: Curran Associates Inc.: 972-981

Lai Z, Chen R, Jia J and Qian Y. 2020. Real-time micro-expression recognition based on resnet and atrous convolutions. Journal of Ambient Intelligence and Humanized Computing, 1-12 [DOI: 10.1007/s12652-020-01779-5]

Li J T, Dong Z Z, Lu S Y, Wang S J, Yan W J, Ma Y H, Liu Y, Huang C B and Fu X L. 2023. CAS(ME)^3: a third generation facial spontaneous micro-expression database with depth information and high ecological validity. IEEE Transactions on Pattern Analysis and Machine Intelligence, 45(3): 2782-2800 [DOI: 10.1109/TPAMI.2022.3174895]

Li X B, Pfister T, Huang X H, Zhao G Y and Pietikäinen M. 2013. A spontaneous micro-expression database: inducement, collection and baseline//Proceedings of the 10th IEEE International Conference and Workshops on Automatic Face and Gesture Recognition. Shanghai, China: IEEE: 1-6 [DOI: 10.1109/FG.2013.6553717]

Liong S T, Gan Y S, See J, Khor H Q and Huang Y C. 2019. Shallow triple stream three-dimensional CNN (STSTNet) for Micro-expression Recognition//Proceedings of the 14th IEEE International Conference on Automatic Face and Gesture Recognition (FG 2019). Lille, France: IEEE: 1-5 [DOI: 10.1109/FG.2019.8756567]

Liong S T, See J, Wong K S and Phan R C W. 2018. Less is more: micro-expression recognition from video using apex frame. Signal Processing Image Communication, 62: 82-92 [DOI: 10.1016/j.image.2017.11.006]

Liu Y C, Du H M, Zheng L and Gedeon T. 2019. A neural micro-expression recognizer//Proceedings of the 14th IEEE International Conference on Automatic Face and Gesture Recognition (FG 2019). Lille, France: IEEE: 1-4 [DOI: 10.1109/FG.2019.8756583]

Lu S Y, Li J T, Wang Y, Dong Z Z, Wang S J and Fu X L. 2022. A more objective quantification of micro-expression intensity through facial electromyography//Proceedings of the 2nd Workshop on Facial Micro-Expression: Advanced Techniques for Multi-Modal Facial Expression Analysis. Lisboa, Portugal: ACM: 11-17 [DOI: 10.1145/3552465.3555038]

Mehrabian A. 1965. Communication without words. The Lancet, 286(7401): #30 [DOI: 10.1016/S0140-6736(65)90194-7]

Pérez J S, Meinhardt-Llopis E and Facciolo G. 2013. TV-L1 optical flow estimation. Image Processing on Line, 3: 137-150 [DOI: 10.5201/ipol.2013.26]

Pfister T, Li X B, Zhao G Y and Pietikäinen M. 2011. Recognising spontaneous facial micro-expressions//Proceedings of 2011 International Conference on Computer Vision. Barcelona, Spain: IEEE: 1449-1456 [DOI: 10.1109/ICCV.20116126401]

Polikovsky S, Kameda Y and Ohta Y. 2009. Facial micro-expressions recognition using high speed camera and 3D-gradient descriptor//Proceedings of the 3rd International Conference on Imaging for Crime Detection and Prevention (ICDP 2009). London, UK: IET: 1-6 [DOI: 10.1049/ic.2009.0244]

Porter S and Ten Brinke L. 2008. Reading between the lies: identifying concealed and falsified emotions in universal facial expressions. Psychological Science, 19(5): 508-514 [DOI: 10.1111/j.1467-9280.2008.02116.x]

Qu F B, Wang S J, Yan W J, Li H, Wu S H and Fu X L. 2018. CAS(ME)^2: a database for spontaneous macro-expression and micro-expression spotting and recognition. IEEE Transactions on Affective Computing, 9(4): 424-436 [DOI: 10.1109/TAFFC.2017.2654440]

Scardapane S, van Vaerenbergh S, Hussain A and Uncini A. 2020. Complex-valued neural networks with nonparametric activation functions. IEEE Transactions on Emerging Topics in Computational Intelligence, 4(2): 140-150 [DOI: 10.1109/TETCI.2018.2872600]

See J, Yap M H, Li J T, Hong X P and Wang S J. 2019. MEGC 2019-the second facial micro-expressions grand challenge//Proceedings of the 14th IEEE International Conference on Automatic Face and Gesture Recognition (FG 2019). Lille, France: IEEE: 1-5 [DOI: 10.1109/FG.2019.8756611]

Tang H, Zhu L J, Fan S and Liu H M. 2022. Micro-expression recognition based on optical flow method and pseudo three-dimensional residual network. Journal of Signal Processing, 38(5): 1075-1087 (唐宏, 朱龙娇, 范森, 刘红梅. 2022. 基于光流法与伪3维残差网络的微表情识别. 信号处理, 38(5): 1075-1087) [DOI: 10.

16798/j.issn.1003-0530.2022.05.020]

van Quang N, Chun J and Tokuyama T. 2019. Capsulenet for micro-expression recognition//Proceedings of the 14th IEEE International Conference on Automatic Face and Gesture Recognition (FG 2019). Lille, France: IEEE: 1-7 [DOI: 10.1109/FG. 2019. 8756544]

Wang C Y, Peng M, Bi T and Chen T. 2020. Micro-attention for micro-expression recognition. Neurocomputing, 410: 354-362 [DOI: 10. 1016/j.neucom.2020.06.005]

Woo S, Park J, Lee J Y and Kweon I S. 2018. CBAM: convolutional block attention module//Proceedings of the 15th European Conference on Computer Vision. Munich, Germany: Springer: 3-19 [DOI: 10.1007/978-3-030-01234-2_1]

Xu F and Zhang J P. 2017. Facial microexpression recognition: a survey. Acta Automatica Sinica, 43(3): 333-348 (徐峰, 张军平. 2017. 人脸微表情识别综述. 自动化学报, 43(3): 333-348) [DOI: 10.16383/j.aas.2017.c160398]

Yan W J, Li X B, Wang S J, Zhao G Y, Liu Y J, Chen Y H and Fu X L. 2014. CASME II: an improved spontaneous micro-expression database and the baseline evaluation. PLoS One, 9(1): #e86041 [DOI: 10.1371/journal.pone.0086041]

Yan W J, Wu Q, Liu Y J, Wang S J and Fu X L. 2013. CASME database: a dataset of spontaneous micro-expressions collected from neutralized faces//Proceedings of the 10th IEEE International Conference and Workshops on Automatic Face and Gesture Recognition. Shanghai, China: IEEE: 1-7 [DOI: 10.1109/FG. 2013. 6553799]

Yap M H, See J, Hong X P and Wang S J. 2018. Facial micro-expressions grand challenge 2018 summary//Proceedings of the 13th IEEE International Conference on Automatic Face and Gesture Recognition (FG 2018). Xi'an, China: IEEE: 675-678 [DOI: 10.1109/FG.2018.00106]

Yarotsky D. 2017. Error bounds for approximations with deep ReLU networks. Neural Networks, 94: 103-114 [DOI: 10.1016/j. neunet. 2017.07.002]

Zhang J H, Liu F and Qi J Y. 2022. Lightweight micro-expression recognition architecture based on Bottleneck Transformer. Computer Science, 49(S1): 370-377 (张嘉淏, 刘峰, 齐佳音. 2022. 一种基于Bottleneck Transformer的轻量级微表情识别架构. 计算机科学, 49(S1): 370-377) [DOI: 10.11896/jsjkx.210500023]

Zhang J H, Liu F and Zhou A M. 2021. Off-TANet: a lightweight neural micro-expression recognizer with optical flow features and integrated attention mechanism//Proceedings of the 18th Pacific Rim International Conference on Artificial Intelligence. Hanoi, Vietnam: Springer: 266-279 [DOI: 10.1007/978-3-030-89188-6_20]

Zhang M, Fu Q F, Chen Y H and Fu X L. 2014. Emotional context influences micro-expression recognition. PLoS One, 9(4): #e95018 [DOI: 10.1371/journal.pone.0095018]

Zhou L, Mao Q R and Xue L Y. 2019. Dual-inception network for cross-database micro-expression recognition//Proceedings of the 14th IEEE International Conference on Automatic Face and Gesture Recognition (FG 2019). Lille, France: IEEE: 1-5 [DOI: 10.1109/FG. 2019.8756579]

作者简介

赵明华,女,教授,主要研究方向为数字图像处理、模式识别、计算机图形学、计算机视觉。E-mail: mh_zhao@126.com

胡静,通信作者,女,副教授,主要研究方向为高光谱遥感影像处理。E-mail: jinghu@xaut.edu.cn

董爽爽,女,硕士研究生,主要研究方向为数字图像处理。E-mail: 2201220007@stu.xaut.edu.cn

都双丽,女,讲师,主要研究方向为雨雾图像与低照度图像处理。E-mail: dusl@xaut.edu.cn

石程,女,副教授,主要研究方向为遥感图像分类和模式识别。E-mail: c_shi@xaut.edu.cn

李鹏,男,副教授,主要研究方向为图像分类。E-mail: lip2006@xaut.edu.cn

石争浩,男,教授,主要研究方向为机器视觉及图像处理、机器学习及其应用。E-mail: ylshi @xaut.edu.cn