

## 基于ME-ResNet人脸微表情识别方法江

江盛, 朱建鸿

引用本文

江盛, 朱建鸿. 基于ME-ResNet人脸微表情识别方法江[J]. 计算机科学, 2024, 51(11A): 231000053-7.

JIANG Sheng, ZHU Jianhong. Face Micro-expression Recognition Method Based on ME-ResNet[J]. Computer Science, 2024, 51(11A): 231000053-7.

---

### 相似文章推荐 (请使用火狐或 IE 浏览器查看文章)

Similar articles recommended (Please use Firefox or IE to view the article)

#### [基于CNN结合BiGRU的恶意流量分类算法研究](#)

Study on Malicious Traffic Classification Algorithm Based on CNN Combined with BiGRU  
计算机科学, 2024, 51(11A): 231100106-9. <https://doi.org/10.11896/jsjcx.231100106>

#### [FCTNet:基于双域深度学习的公交车到站时间预测方法](#)

FCTNet:Bus Arrival Time Prediction Method Based on Dual Domain Deep Learning  
计算机科学, 2024, 51(11A): 231000180-7. <https://doi.org/10.11896/jsjcx.231000180>

#### [Partition-Time Masking:一种唇语识别数据增强方法](#)

Partition-Time Masking:A Data Augmentation Method for Lip Reading  
计算机科学, 2024, 51(11A): 240300139-6. <https://doi.org/10.11896/jsjcx.240300139>

#### [基于图卷积网络的糖尿病视网膜病变分级模型](#)

Grading Model for Diabetic Retinopathy Based on Graph Convolutional Network  
计算机科学, 2024, 51(11A): 231000042-5. <https://doi.org/10.11896/jsjcx.231000042>

#### [基于改进残差网络的混凝土砂石骨料种类识别研究](#)

Study on Identification of Concrete Sand and Gravel Aggregate Types Based on Improved Residual Network  
计算机科学, 2024, 51(11A): 231000082-6. <https://doi.org/10.11896/jsjcx.231000082>

# 基于 ME-ResNet 人脸微表情识别方法

江 盛 朱建鸿

江南大学轻工过程先进控制教育部重点实验室 江苏 无锡 214122

(466597129@qq.com)

**摘 要** 人脸微表情具有持续时间短、动作幅度小的特点。数据集样本量较少等因素也给微表情识别带来了巨大挑战。针对上述问题,提出一种基于 ME-ResNet 残差网络的微表情识别方法。首先,在预处理阶段,等间隔提取微表情视频片段起始帧至顶点帧之间的关键帧序列,利用改进 Farneback 光流法提取微表情关键帧序列的面部光流运动特征;接着,构建基于 3D 卷积的 ResNet50 网络,并将空间通道注意力 CBAM 机制加入网络 Bottleneck 模块,以增强网络对面部关键运动特征的聚焦学习能力,并构建 ME-ResNet 网络模型,将所提取的面部光流运动特征送入网络进行训练;最后,使用数据增强增加网络训练样本量,将 ME-ResNet 网络模型用于微表情识别任务,并在 CASMEII,SMIC 和 SAMM 数据集上进行实验验证,所提算法识别率达到了 84.42%,72.56%,70.41%,与其他算法相比具有较高的识别能力。

**关键词:** 微表情识别;Farneback 光流法;卷积神经网络;运动特征;数据增强

**中图分类号** TP391

## Face Micro-expression Recognition Method Based on ME-ResNet

JIANG Sheng and ZHU Jianhong

Key Laboratory of Advanced Process Control for Light Industry, Ministry of Education, Jiangnan University, Wuxi, Jiangsu 214122, China

**Abstract** Face micro-expressions have the characteristics of short duration and small amplitude of movement. Factors such as the small sample size of dataset also bring great challenges to micro-expression recognition. To solve these problems, this paper proposes a micro-expression recognition method based on ME-ResNet residual network. First, in the pre-processing stage, extract the key frame sequence between the start frame and the vertex frame in the micro-expression video clip at equal intervals and then, use the improved Farneback optical flow method to extract the motion features of the micro-expression key frame sequence. Second, construct a ResNet50 network based on 3D convolution and add the spatial channel attention CBAM mechanism to the network Bottleneck module, so as to enhance the ability to focus on key facial motor features. Next, construct the ME-ResNet network model and sent the extracted facial optical flow motion features to the network for training. Finally, use the data enhancement to increase the sample size of network training and apply the ME-ResNet network model to micro-expression recognition tasks. Also, experimental results on CASME II, SMIC and SAMM datasets show that the recognition rate of the proposed algorithm reaches 84.42%, 72.56% and 70.41% respectively. It has higher recognition ability compared with other algorithms.

**Keywords** Micro-expression recognition, Farneback optical flow method, Convolutional neural networks, Motion features, Data augmentation

## 1 引言

微表情(Micro Expression, ME)是一种自发、无意识、短暂的面部表情,通常在人们试图掩饰自身情绪波动时产生。相比普通表情,微表情的持续时间更短,很难察觉,通常只持续 1/25~1/5 s,但它们仍然能够揭示人们真实情感状态,包括愤怒、快乐、悲伤、厌恶、惊讶等。例如,在面对刑讯审问<sup>[1]</sup>、法庭审判等场合中,检察官、警察可以使用微表情分析技术来揭示嫌疑人或证人的真实情感状态。在心理学和医学研究领域<sup>[2]</sup>中,微表情也可以作为一种诊断工具,帮助医生和心理学家分析患者情感和心理健康。近年来,随着计算机视觉技术

的不断突破,微表情识别成为充满挑战与价值的研究领域。

目前微表情识别算法可以分成两大类:基于传统方法手工提取特征方式和基于深度学习方式。早期微表情识别是通过传统手工提取特征方法,主要有基于光流特征方法和基于局部二值模式方法。Pfister 团队<sup>[3]</sup>实现了 3 个正交平面局部二元模式(Local Binary Patterns from Three Orthogonal Planes, LBP-TOP),通过描述动态纹理来识别面部微表情,有效利用压缩时空形式信息,首次成功实现面部自发微表情识别。Huang 等<sup>[4]</sup>提出基于时空完备局部量化模式的微表情分析方法,将完备局部量化模式扩展到三维空间进行微表情识别,在提取特征时加入了信号、大小和方向等因素,考虑了

基金项目:国家自然科学基金(61973139)

This work was supported by the National Natural Science Foundation of China(61973139).

通信作者:朱建鸿(466597129@qq.com)

更多信息,提高了微表情识别性能,但也引入了高维度。Shreve 等<sup>[5]</sup>利用运动引起面部时空应变,通过计算下巴、嘴巴、脸颊、前额等不同面部区域光流场的运动强度,度量区域中的张力幅值来识别微表情,但是这种方法需要通过手工设定阈值,有效性难以得到保证。Liu 等<sup>[6]</sup>提出主方向平均光流特征方法,基于光流特征进行微表情识别,利用光流场分析视频片段中人脸的不同子区域,计算各方向光流均值后对不同微表情进行分类。上述基于传统手工提取特征的方法,虽然在一定程度上能够较为准确地对微表情进行识别,但也存在着明显不足:产生计算量大,效率低,复杂度高,数据容易冗余,而且人工提取需要较强的专业性知识,往往会忽略一些微表情产生时的面部细节,造成识别准确率低下。

随着计算机视觉技术不断发展和广泛应用,越来越多研究人员将其应用在微表情识别领域,并且取得了更高的识别精度。Wang 等<sup>[7]</sup>提出一种名为迁移长期卷积神经网络的深度学习方法,该方法使用迁移的深层卷积神经网络提取面部特征,然后使用长短时记忆(Long Short-Term Memory Network)模型进行训练学习以识别微表情。Peng 等<sup>[8]</sup>提出基于顶点帧空间信息和起始帧时间信息的顶点时间网络,通过学习时间信息辅助微表情识别,在跨数据集验证中极具鲁棒性。Gan 等<sup>[9]</sup>介绍了一种新的特征描述符,该特征描述符将光流与 CNN 相结合,从起始帧至顶点帧获得光流信息后,将光流特征送入 Off-ApexNet 模型进行进一步的特征增强和表情分类。随后又设计了一种浅层三流三维卷积神经网络<sup>[10]</sup>(Shallow Triple Stream Three-dimensional CNN),该网络更小,却能提取更具差异性的高级特征。

然而,上述算法大多只利用起始帧和顶点帧两帧之间的差异进行运动特征<sup>[11]</sup>提取,而忽略了两帧之外其他帧序列的运动细节,从而造成信息提取不全,影响了整体网络识别微表情的准确率。除此之外,由于微表情数据样本集较少,建立过深的网络进行训练会造成过拟合,降低模型性能。

针对面部关键运动信息提取不全、样本量少的问题,本文提出了一种基于 ME-ResNet 残差网络的微表情识别方法。

1)首先,本文方法使用改进 Farneback 光流法,等间隔抽取微表情视频片段中的起始帧至顶点帧之间的帧序列,利用人脸面部感兴趣区域 ROI,裁剪出人脸重要区域,提取微表情产生时的光流运动特征;

2)然后,基于 3D 卷积构建出用于微表情识别的 ME-ResNet 网络模型,增强网络提取时间和空间维度信息能力;并在 Bottleneck 模块插入空间通道注意力机制,使网络能够更加聚焦学习面部运动光流特征信息;

3)最后,利用数据增强<sup>[12]</sup>方法来提高微表情样本量,使用迁移学习进行预训练,让网络训练出更为有效的分类器,提高微表情识别准确率。

## 2 本文算法

本文算法具体包括 3 个步骤:微表情视频片段预处理、面部运动特征提取和微表情分类。首先预处理部分包括人脸裁剪对齐,起始帧、顶点帧定位;然后从微表情序列中提取面部光流运动特征和空间位置特征送入网络进行训练;最后将经过网络处理后的信息输入分类器,完成微表情分类输出。其流程图如图 1 所示。

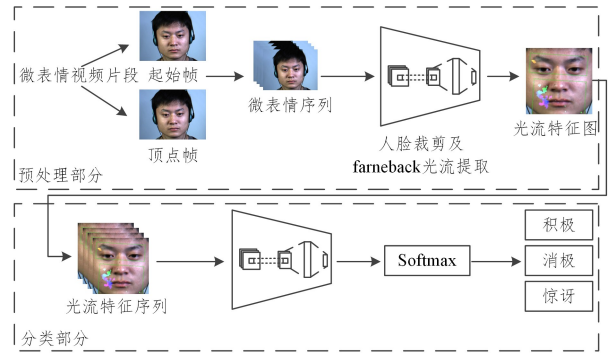


图 1 算法流程图

Fig. 1 Algorithm flow chart

### 2.1 光流运动特征提取

由于微表情是一种自发、无意识的面部表情,通常其持续时间短,动作幅度小,视频片段的相邻帧之间反映出的光流运动特征并不明显。若逐帧输入所有帧序列的光流特征图,不仅会造成数据冗余,引入过多无关背景信息,还会增加模型计算量,降低识别效率。此外,人脸微表情产生是面部多处区域共同作用的结果,仅提取起始帧和顶点帧两帧之间的光流运动差异会忽略面部其他位置运动细节。针对此问题,本文提出改进 Farneback 光流法<sup>[13]</sup>,通过面部感兴趣区域 ROI 进行光流运动特征提取。具体过程如下:

1)将每个微表情视频片段的起始帧至顶点帧之间等间距提取  $N$  帧图像序列。

2)将提取的帧序列进行人脸裁剪,去除无关区域的影响。

3)利用改进的 Farneback 光流法提取起始帧至顶点帧之间关键帧序列中的面部光流运动特征。

#### 2.1.1 改进 Farneback 光流法

Farneback 光流法是一种基于梯度的方法,假设目标图像梯度恒定且局部光流恒定。首先构建高斯图像金字塔,对输入的两帧连续图像进行金字塔处理。图像金字塔是一种多尺度表达,通过逐级降采样得到不同分辨率图像,从而能够处理不同大小和速度的运动目标。高斯金字塔可以表示为:

$$G_i = G_i' * K \quad (1)$$

其中,  $G_i$  表示金字塔的某一层,  $G_i'$  表示上一层,  $K$  表示用于平滑操作的低通滤波器。下采样操作方法如下:

$$G_{i+1,j+1} = \frac{G'_{2i,j} + G'_{2i+1,j} + G'_{2i,j+1} + G'_{2i+1,j+1}}{4} \quad (2)$$

其中,  $G_{i+1,j+1}$  表示金字塔下一层对应位置的像素值,  $G'_{2i,j}$  表示上一层中对应位置的像素值。下采样操作将前一层中每个相邻的 4 个像素平均值作为下一层对应位置的像素值。

在每个金字塔级别上,对图像进行空间和时间方向的梯度计算。然后通过在整个图像区域内进行积分和平均化来求解位移向量  $(u, v)$ :

$$u = \frac{\sum_{x,y} (I_x^2 \cdot I_t + I_x \cdot I_y)}{\sum_{x,y} (I_x^2 + I_y^2)} \quad (3)$$

$$v = \frac{\sum_{x,y} (I_y^2 \cdot I_t + I_x \cdot I_y)}{\sum_{x,y} (I_x^2 + I_y^2)} \quad (4)$$

其中,  $(x, y)$  表示图像中的像素位置,  $I_x$  和  $I_y$  分别表示图像在  $x$  方向和  $y$  方向上的梯度,  $I_t$  表示两帧图像之间的帧间差异。分子部分是对整个图像区域内梯度信息进行加权求和,用于估计位移向量,分母部分是对梯度的模长进行加权求和,用于

归一化位移向量。

从最低分辨率的金字塔级别开始,逐步向上优化光流场。在每个级别上,通过插值和缩放方式将上一级的光流场估计作为下一级初始估计,并继续进行光流近似和优化。通过逐级优化,得到每个像素点在  $x$  和  $y$  方向上的位移向量  $(u, v)$ ,从而得到一个密集的光流场  $F=(u, v)$ 。

实际情况中,由于图像中纹理变化和光照变化敏感会使此方法求得的结果噪声太多,光流图如图 2 所示,噪声点密集,分布广。直接输入网络训练会影响网络对关键光流运动的特征提取。

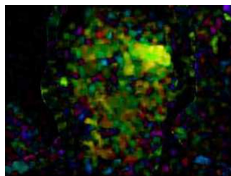


图 2 原 Farneback 光流特征图

Fig. 2 Optical flow feature diagram of original Farneback

针对上述噪声点密集问题,提出改进的 Farneback 算法。首先,利用人脸感兴趣区域 ROI,裁剪出需要计算光流运动特征的区域,去除无关背景;然后,根据微表情在空间和时间上的连续性进行时空连续性判别,将面部运动与无关噪声进行区分,提高网络对面部运动特征的提取准确性。

### 2.1.2 时空连续性判别

提取微表情视频起始帧至顶点帧之间关键帧序列进行灰度化和 Farneback 光流计算,即计算前后两帧图像之间像素点的运动矢量,得到目标图像  $I_m$ 。此时目标图像  $I_m$  包含面部运动目标和多余噪声点,需将目标图像进行时空连续性判别,以此去除噪声。

时空连续性判别是通过一个动态阈值  $U$ ,对目标图像中的像素运动轨迹进行过滤,去除运动轨迹不连贯的噪声,计算式如下:

$$Q(x, y) = \begin{cases} 1, & |I_m(x, y) - I_{m(t-1)}(x, y)| > U \\ 0, & \text{其他} \end{cases} \quad (5)$$

其中,  $Q(x, y)$  为目标图像  $I_m$  上指定位置像素点,  $t$  表示当前帧,  $(t-1)$  表示前一帧,若前后两帧图像灰度差的绝对值大于阈值  $U$ ,则判定该位置为有效运动,并保留其运动轨迹,否则视为噪声点并清除。若阈值  $U$  选择过高,会清除有效运动,导致面部运动特征提取不完整;若阈值  $U$  选择过低,会引入大量干扰噪声,干扰网络训练。

### 2.1.3 自适应阈值

由于不同微表情片段中的运动强度不同,固定阈值不能充分提取面部光流运动特征,因此提出了一种自适应确定阈值方法。

首先将提取的图像帧序列转换为灰度图像,计算每两个相邻帧之间对应像素灰度差的绝对值  $P_i$ 。对于每个像素位置  $(x, y)$ ,假设  $I_{m2}$  和  $I_{m1}$  分别表示相邻帧的灰度值,则灰度差  $P_i$  可以表示为:

$$P_i(x, y) = |I_{m2}(x, y) - I_{m1}(x, y)| \quad (6)$$

再将两个相邻帧之间的所有像素产生灰度差  $P_i$  进行累加,则总灰度差  $S_i$  可以表示为:

$$S_i = \sum_x \sum_y P_i(x, y) \quad (7)$$

最后,求出 11 帧图像产生总灰度差累积值的均值  $p$ :

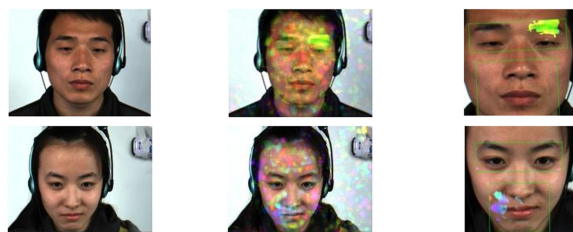
$$p = \frac{1}{M * K} * \sum_{i=1}^{10} S_i \quad (8)$$

其中,  $M$  的值为  $(N-1)$ ,表示  $N$  张帧序列共产生  $(N-1)$  个灰度差。  $K$  为 1 帧图像包含的像素值。把均值  $p$  乘以一个比例因子  $a$  后可作为动态阈值  $U$ :

$$U = a * p \quad (9)$$

理论上,微表情强度越小,连续帧序列之间灰度差越小,得到的阈值  $U$  也越小,符合实际需求。

通过改进的 Farneback 光流法可以充分提取面部关键运动信息。图 3 是 CASMEII 数据集中 sub19/EP02\_01 惊讶样本和 sub17/EP01\_06 积极样本对应的可视化光流应变图,惊讶样本面部运动集中在右边眉毛区域,而积极样本面部运动集中在左边嘴角区域,颜色越亮代表运动幅度越大。由此可以看出,改进后 Farneback 光流法很好地降低了面部其他无关噪声点的影响,准确地提取了微表情产生时面部关键区域光流运动特征。



(a) 原始图片 (b) 原 Farneback 法提取的光流图 (c) 改进 Farneback 提取的光流图

图 3 不同方法光流应变可视化图对比

Fig. 3 Comparison of optical flow strain visualizations by different methods

## 2.2 ME-ResNet 网络

随着网络深度增加,网络提取的特征更加丰富,但模型准确率会出现大幅度降低的现象。这是由于之前网络模型随着网络层不断加深会造成梯度爆炸和梯度消失问题,为解决此问题便诞生了 ResNet 网络。ResNet 网络是由残差块(Residual Building Block)构建的,而 ResNet50 网络是由 7 个部分组成,其中第 1 部分不包含残差块,主要用于对输入图像进行卷积、正则化、激活和池化等计算。而第 2-5 部分,一共有 16 个残差模块组成,用于对输入图像进行特征提取。第 6 部分和第 7 部分分别为全局平均池化层和全连接层,最后经过 SoftMax 输出分类结果。

本文提出的 ME-ResNet 网络模型以 ResNet50 网络<sup>[14]</sup>为主干结构,为了加强网络对微表情片段中时间维度信息的获取,本文构建了基于 3D 卷积的 ME-ResNet 网络<sup>[15-16]</sup>,加强网络对时间维度信息的获取。并在网络的 Bottleneck 模块中穿插 CBAM<sup>[17]</sup>通道空间注意力机制<sup>[18]</sup>,使网络可以聚焦学习面部运动信息,其结构如图 4 所示。

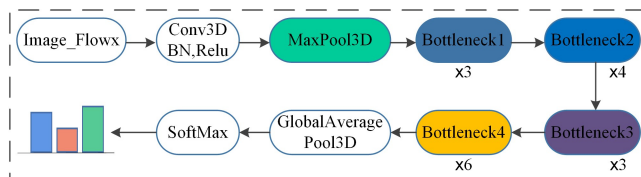


图 4 ME-Resnet 网络结构图

Fig. 4 Diagram of ME-Resnet network structure

### 2.2.1 3D 卷积层

2D 卷积通常作用于单张图像,可以充分提取单张图像的细节特征,但不能提取时间轴上的局部链接特征,即忽略了对图像时间维度的特征提取。而微表情视频是由多个相互联系的帧序列组合而成,时间轴上特征联系紧密,因此采用 3D 卷积可以充分提取微表情片段时空特征。3D 卷积原理是通过堆叠多个连续帧序列图像组成一个立方体,然后在立方体中运用 3D 卷积核( $K_w, K_h, K_d$ )进行计算。 $K_w, K_h$  和  $K_d$  分别表示卷积核的宽度、高度和深度。

3D 卷积层中每一个特征图都会与上一层中多个邻近连续帧相连,可以捕捉图像时间维度信息。而微表情片段是包含多帧图像的序列,时间特征和空间特征联系密切。因此,本文模型采用 3D 卷积,这样能更好提取关键帧序列中的时序信息,提高识别准确率。

网络中改进 Bottleneck3D 模块的结构图如图 5 所示。

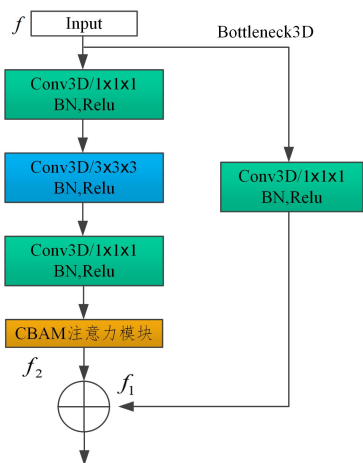


图 5 Bottleneck3D 模块结构图

Fig. 5 Diagram of of Bottleneck3D module structure

首先输入特征图  $f$  依次通过一个  $1 \times 1 \times 1$  卷积、 $3 \times 3 \times 3$  卷积和  $1 \times 1 \times 1$  卷积。每次经过卷积后使用 ReLU 函数激活和批归一化,再将输出特征图传入 CBAM 注意力模块,得到特征图  $f_2$ 。此外,输入特征图  $f$  通过右边  $1 \times 1 \times 1$  卷积后得到  $f_1$ ,将特征图  $f_2$  与  $f_1$  按通道叠加后再输入下一节网络。

### 2.2.2 注意力机制

为使网络更好地聚焦于面部关键区域,提取其光流运动特征,本文使用 CBAM 注意力机制。CBAM 是一种结合空间和通道的注意力机制,如图 6 所示。其主要由两部分组成:通道注意力模块和空间注意力<sup>[19]</sup>模块。目标图像通过此注意力机制得到的权重乘以输入特征图进行自适应特征细化,使网络同时学习输入特征图中的通道依赖性和空间依赖性,加强网络对关键特征区域中重点信息的关注。此外,CBAM 是一个轻量级通用模块,可以在保证效率的同时提高识别准确率。

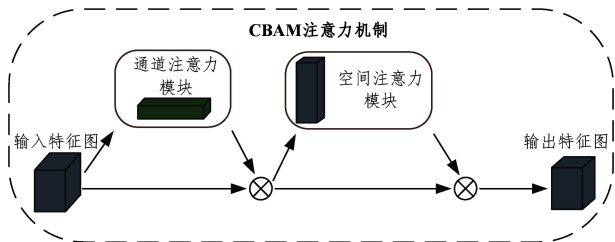


图 6 CBAM 注意力模块

Fig. 6 CBAM attention module

其中,通道注意力部分的作用是使网络学习输入特征图通道间的依赖性,强调重要的通道信息,如图 7 所示。特征图  $f$  有  $C$  个通道分别经过全局最大池化和全局平均池化,形成两个权重向量  $z_c$  :

$$z_c = \frac{1}{H \times W \times D} \sum_{h=1}^H \sum_{w=1}^W \sum_{d=1}^D X_{hwdc} \quad (10)$$

其中, $H, W$  和  $D$  分别表示特征图的高度、宽度和深度。然后将两个权重向量分别经过同一共享 MLP 网络,映射成每个通道权重:

$$s_c = MLP(z_c) \quad (11)$$

再将映射后的权重相加,最终使用 sigmoid 函数输出,从而得到通道权重  $\alpha_c$  :

$$\alpha_c = \frac{\exp(s_c)}{\sum_{i=1}^C \exp(s_i)} \quad (12)$$

最后与原特征图按通道相乘,这样就实现了对输入特征图通道信息的注意力加权。

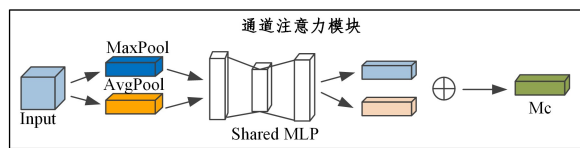


图 7 通道注意力模块

Fig. 7 Channel attention module

空间注意力部分可进一步增强网络对图像帧序列空间特征信息的提取,如图 8 所示,使网络可学习输入特征图的空间依赖性。首先,特征图  $f$  按通道轴经过最大池化和平均池化,形成两个权重向量  $M$  和  $A$ 。然后,将得到的两张特征图按照通道进行堆叠,通过一个卷积层  $K$  来对连接后的特征图进行映射,得到空间注意力得分:

$$s = K(Concat(M, A)) \quad (13)$$

得到空间注意力得分后,将其进行归一化,得到注意力权重  $\beta$  :

$$\beta = \frac{\exp(s)}{\sum_{i=1}^H \sum_{j=1}^W \sum_{k=1}^D \exp(s_{ijk})} \quad (14)$$

最后将得到的空间注意力权重  $\beta$  与原特征图相乘,即特征图上每一个点都被赋予了权重,实现了对输入特征图空间信息的注意力加权。

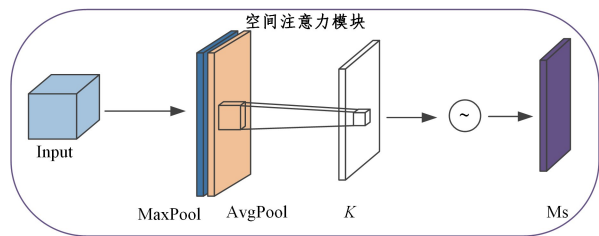


图 8 空间注意力模块

Fig. 8 Spatial attention module

## 3 实验

### 3.1 数据集

为验证本文算法的有效性,构建人脸微表情识别系统,完成微表情分类工作。实验共使用了 3 种公共的微表情自发数据集,分别是 SAMM 数据集、SMIC 数据集和 CASMEII 数据集。3 种公共数据集的具体信息如表 1 所列。按照情绪种类

可以将数据集分成积极、消极和惊讶三大类,其中积极样本中包含积极、快乐 2 种标签,消极样本中包含消极、压抑、恐惧、悲伤 4 种标签,而惊讶样本中只包含惊讶标签。

表 1 3 种微表情数据集

数据集	SAMM	SMIC	CASMEII
发布机构	曼彻斯特大学	奥卢大学	中国科学院
发布年份	2018	2012	2014
微表情种类	7	3	7
积极样本	26	51	32
消极样本	92	70	86
惊讶样本	15	43	28
总样本个数	133	164	146

SMIC 数据集是全球首个公开发布的自发性微表情数据集<sup>[20]</sup>,由芬兰 Oulu 大学的赵国英团队在 2012 年收集,该数据集共有 16 位被试者,包含积极、消极、惊讶 3 种微表情,共有 164 个微表情序列。CASMEII 数据集是由中科院心理研究所的傅小兰团队于 2014 年采集<sup>[21]</sup>,该数据集包含 26 位被试者,且被试者均来自亚洲,共有 255 个微表情序列。SAMM 数据集是英国曼彻斯特大学 Moi Hoon Yap 团队于 2018 年收集<sup>[22]</sup>,该数据集中受试者男女比例为 1:1,其中的 32 位被试者包含 16 位男性和 16 位女性,并且还包含 13 个种族,年龄范围在 19~57 岁,共有 133 个微表情序列。

## 3.2 实验结果及分析

### 3.2.1 实验环境及设置

本文使用 Keras 框架搭建神经网络,采用数据增强方法,将数据集进行镜像、翻转操作,可以扩大样本量,以解决训练的样本量稀疏问题,以免造成网络过拟合。数据增强后的样本数量及各种类样本数量如表 2 和表 3 所列。

表 2 数据增强后的样本量

数据集	扩充前样本量	扩充后样本量
CASMEII	146	584
SMIC	164	656
SAMM	133	532

表 3 各种类样本数量

数据集	CASMEII	SMIC	SAMM
积极样本量	32	51	26
消极样本量	86	70	92
惊讶样本量	28	43	15

预训练采用 Fer2013 宏表情数据集,可一定程度上提高网络训练效率。然后对微表情数据集进行留一交叉验证(Leave-one-subject-out Cross-validation, LOSOCV),测试对微表情的识别效果。本实验的参数设置如下:网络的初始学习率  $\alpha=0.0001$ ,  $batch\_size=32$ ,使用 Adam 优化器。本网络选用 Focal Loss 损失函数,Focal Loss 是一种用于解决分类问题中类别不平衡的损失函数。它可以通过对易分类样本的权重进行调整来减少对非常规样本的过度拟合,该函数的定义如下:

$$FL(p_i) = -\alpha_i (1 - p_i)^\gamma \log(p_i) \quad (15)$$

其中, $\alpha$  为权重,帮助处理类别的不均衡; $(1 - p_i)^\gamma$  是调节因子; $\gamma > 0$  是可调的聚焦参数。为防止网络过拟合,在训练时加入 Early Stopping 机制,以便得到最佳训练模型。

### 3.2.2 评价指标

本实验用到的评价指标是准确率(Accuracy)、未加权 F1 得分(Unweighted F1-score,UF1)和未加权平均召回率(Unweighted Average Recall,UAR)。

UF1 指标是衡量多分类模型性能的一种指标,适用于处理样本类别分布不均衡的情况,因此在微表情分类问题上是一个很好的评估标准。

$$F1_i = \frac{2TP_i}{2TP_i + FP_i + FN_i} \quad (16)$$

$$UF1 = \frac{\sum_{i=1}^C F1_i}{C} \quad (17)$$

其中, $C$  为微表情总类别, $i$  表示某一类别, $TP$  为被模型预测为正类的正样本; $TN$  为被模型预测为负类的负样本; $FP$  为被模型预测为正类的负样本; $FN$  为被模型预测为负类的正样本。

未加权平均召回 UAR 是针对每个微表情类别计算召回率的平均值,其评估标准如下:

$$UAR = \frac{1}{C} \sum_{i=1}^C \frac{TP_i}{TP_i + FP_i + TN_i + FN_i} \quad (18)$$

其中, $i$  表示某一类别, $TN$  表示真阴性。

### 3.2.3 设定等间距提取图像序列最优帧数

由于微表情产生时面部动作幅度十分微小,原始视频图像帧序列之间的运动差异不明显。如果直接将原始视频图像帧序列全部进行光流特征提取,面部的有效运动和噪声点不易区分,造成网络对面部光流运动特征提取不全,进而无法准确识别微表情类型。针对上述问题,本文提出了等间距提取图像帧的方法,并通过实验来寻找合适的采样帧数  $N$ 。图 9 给出了实验所得识别率和采样帧数之间的关系。根据实验结果,当采样帧数为 11 帧时,网络识别准确率最高,相比仅采样 3 帧时在 3 种数据集上的识别率分别提高了 2.78%,2.54%,2.63%。采样帧数越高,网络提取的细节信息就越多。但采样帧数在 11 帧之后,识别精度在逐渐降低。这是由于本文网络采用 3D 卷积层,输入帧序列太少会造成时空维度信息缺失;输入帧序列太多会造成运动特征信息冗余,相邻帧之间光流运动差异不明显,进而影响网络对光流运动特征的提取,降低识别准确率。

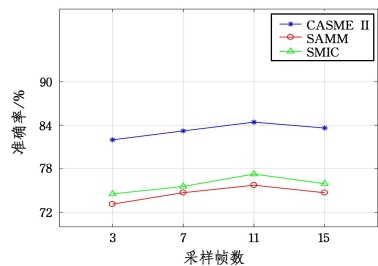


图 9 不同采样帧数在 3 种数据集上的识别精度

Fig. 9 Recognition accuracy of different sampling frames on three datasets

### 3.2.4 设置比例因子 $\alpha$

动态阈值  $U$  的比例因子  $\alpha$  通常是一个用于调整局部阈值的参数,以适应不同情况。这个因子可以根据具体应用和图像特点来调整。不同比例因子下在 3 种数据集上的识别精度如图 10 所示,当比例因子  $\alpha=0.8$  时,网络的识别准确率最高,此时改进的 Farneback 光流法可充分提取面部关键区域

的光流运动特征。当比例因子  $a$  选择过低时,面部某些关键运动会被当作噪声点清除,影响网络学习。当比例因子  $a$  选择过高时,提取的光流图噪声点密集,会造成特征冗余,影响网络识别。

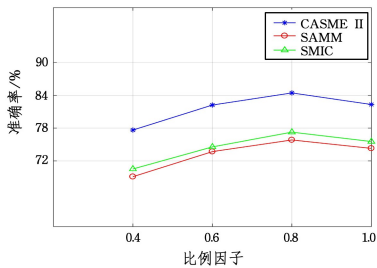


图 10 不同比例因子  $a$  在 3 种数据集上的识别精度

Fig. 10 Recognition accuracy of different scale factors  $a$  on three datasets

### 3.2.5 消融实验

消融实验方法的设计如表 4 所列,为验证 ME-ResNet 网络识别微表情有效性,消融实验在 CASMEII 微表情数据集上进行,不同方法的识别结果如图 11 所示。

表 4 4 种实验方法的设计

Table 4 Design of four experimental methods

方法	网络模型	是否提取 farneback 光流	是否引用 3D 卷积	是否添加 CBAM 模块
1	ResNet50	×	×	×
2	ResNet50	✓	×	×
3	ResNet50	✓	✓	×
4	ME-ResNet	✓	✓	✓

由图 11 可知,ME-ResNet 网络在 CASMEII 数据集上 UF1 达到了 83.02%,UAR 达到了 84.19%,准确率达到 84.42%。通过改进的 Farneback 光流法提取面部关键运动的光流特征,极大提高了网络的识别性能;此外,通过引用 3D 卷积和添加 CBAM 注意力模块,加强了网络对图像序列在时间和空间维度特征信息的关注,也在一定程度上提高了模型的识别准确率。

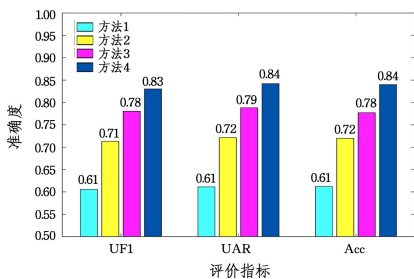


图 11 不同方法在 CASMEII 数据集上的精度对比

Fig. 11 Precision comparison of different methods on CASME II dataset

### 3.3 与其他算法的对比

将本文算法与其他算法进行精度对比,选取 UF1 和 UAR 两个指标,结果如表 5 所列。结果表明,本文算法取得了良好性能,ME-ResNet 网络模型在 CASME II 数据库上的准确率最高,这是因为 CASME II 数据库样本精确的起始帧和峰值帧标注及高帧率可以提供更准确的光流计算,能更好地表征面部光流运动变化。本文算法在 SMIC 数据库和 SMM 数据集上的准确率略低,这是因为数据集受试者年龄

差范围较大,面部皱纹和肌肉松弛会影响网络识别。针对 SMIC 数据集,ME-ResNet 网络的 UF1 和 UAR 指标分别提高了 1.85% 和 2.12%;在 SMM 数据集中 UF1 和 UAR 指标分别提高了 2.43% 和 2.13%。总体上本文算法有着不错的竞争力。

表 5 与其他算法的精度对比

Table 5 Accuracy comparison with other algorithms

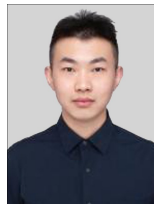
方法	CASMEII		SMIC		SMM	
	UF1	UAR	UF1	UAR	UF1	UAR
OFF-ApexNet <sup>[9]</sup>	<b>0.8764</b>	0.8681	0.6817	0.6695	0.5409	0.5392
STSTNet <sup>[10]</sup>	0.8382	<b>0.8686</b>	0.6801	0.7013	0.6588	0.6810
RCN-W <sup>[23]</sup>	0.8131	0.8522	0.6600	0.6584	0.6164	0.7164
AT-Net <sup>[8]</sup>	0.7980	0.7750	0.5530	0.5430	0.4960	0.4820
H-SVM <sup>[24]</sup>	0.7166	0.7525	0.5465	0.6398	0.4180	0.4141
本文方法	0.8302	0.8419	<b>0.7002</b>	<b>0.7225</b>	<b>0.6831</b>	<b>0.7023</b>

**结束语** 本文提出了一种基于改进的 Farneback 光流法的微表情识别算法。首先,在视频预处理方面,等间隔抽取从起始帧到顶点帧之间的 11 帧序列,利用改进的 Farneback 光流法,提取帧序列之间面部关键区域肌肉运动的光流特征,去除无关背景噪声的影响。然后,构建基于 3D 卷积的 ME-ResNet 网络模型,同时在网络中添加 CBAM 注意力模块,使网络更专注于提取面部微弱运动特征以用于分类,并加强网络对空间维度和时间维度的特征提取能力。最后,使用数据增强增加样本量,训练出用于微表情识别的网络模型。实验结果表明,本文算法对微表情的识别效果显著,和其他主流算法相比具有不错的竞争力。由于微表情样本种类分布严重不均,跨数据集识别率不高,因此解决这些问题提高微表情识别的准确性和鲁棒性也是未来的研究方向与思路。

### 参考文献

- [1] PENG Y W. Research on the application of micro-expression analysis technology in investigation and interrogation[J]. Chinese Journal of Criminal Law, 2015(2):95-103.
- [2] ZHU C L, CHEN X Y, ZHANG J X, et al. Comparison of Ecological Micro-Expression Recognition in Patients with Depression and Healthy Individuals[J]. Frontiers in Behavioral Neuroscience, 2017, 11(6):199-224.
- [3] PFISTER T, LI X, ZHAO G, et al. Recognising Spontaneous Facial Micro-Expressions[C]// 2011 International Conference on Computer Vision. IEEE, 2011:1449-1456.
- [4] HUANG X H, ZHAO G Y, HONG X P, et al. Spontaneous facial micro-expression analysis using spatiotemporal completed local quantized patterns[J]. Neurocomputing, 2016, 175: 564-578.
- [5] SHREVE M, GODAVARTHY S, MANOHAR V, et al. Towards Macro-and Micro-Expression Spotting in Video Using Strain Patterns[C]// 2009 Workshop on Applications of Computer Vision(WACV). IEEE, 2009:1-6.
- [6] 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.
- [7] WANG S J, LI B J, LIU Y J, et al. Micro-Expression Recogni-

- tion with Small Sample Size by Transferring Long-Term Convolutional Neural Network[J]. *Neurocomputing*, 2018, 312: 251-262.
- [8] PENG M, WANG C, BI T, et al. A Novel Apex-Time Network for Cross-Dataset Micro-Expression Recognition[C]// 2019 8th International Conference on Affective Computing and Intelligent Interaction (ACII). 2024.
- [9] GAN Y S, LIONG S T, YAU W C, et al. OFF-ApexNet on micro-expression recognition system[J]. *Signal Processing: Image Communication*, 2019, 74: 129-139.
- [10] GAN Y S, LIONG S T, YAU W C, et al. OFF-ApexNet on micro-expression recognition system[J]. *Signal Processing: Image Communication*, 2019, 74: 129-139.
- [11] LIANG Z Y, LIU D Z, SUN Y. Micro-expression recognition method combining transfer learning and separable three-dimensional convolution[J]. *Computer Engineering*, 2022, 48(1): 228-235.
- [12] SUN S K, FAN J, SUN Z Q, et al. A review of image data enhancement based on deep learning[J/OL]. *Computer Science*, 1-23 [2023-10-06]. <http://kns.cnki.net/kcms/detail/50.1075.TP.20230926.0857.012.html>.
- [13] FARNEBACK, BIGUN J, GUSTAVSSON T. Two-Frame Motion Estimation Based on Polynomial Expansion[J]. *Image Analysis Volume*, 2003: 363-370.
- [14] LI B, LIMA D. Facial expression recognition via ResNet-50[J]. *International Journal of Cognitive Computing in Engineering*, 2021. DOI:10.1016/j.ijce.2021.02.002.
- [15] LIU S H, REN Y S, LI L T, et al. Micro-expression recognition based on SqueezeNet and C3D[J]. *Multimedia Systems*, 2022, 28(6): 2227-2236.
- [16] WANG C, CHANG J. CSI Cross-domain Gesture Recognition Method Based on 3D Convolutional Neural Network[J]. *Computer Science*, 2021, 48(8): 322-327.
- [17] WANG C, CHANG J. CSI Cross-domain Gesture Recognition Method Based on 3D Convolutional Neural Network[J]. *Computer Science*, 2021, 48(8): 322-327.
- [18] WOO S, PARK J, LEE J Y, et al. CBAM: Convolutional Block Attention Module[C]// ECCV 2018. Cham: Springer, 2018.
- [19] LI X R, ZHANG L Y, YAO S J. Micro-expression Recognition Method Combining Feature Fusion and Attention Mechanism[J]. *Computer Science*, 2022, 49(2): 4-11.
- [20] LI X R, ZHANG L Y, YAO S J. Micro-expression Recognition Method Combining Feature Fusion and Attention Mechanism[J]. *Computer Science*, 2022, 49(2): 4-11.
- [21] ZHU X, CHENG D, ZHANG Z, et al. An Empirical Study of Spatial Attention Mechanisms in Deep Networks[C]// 2019 IEEE/CVF International Conference on Computer Vision (ICCV). IEEE, 2020.
- [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] 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): e86041.
- [24] 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.



**JIANG Sheng**, born in 1998, postgraduate. His main research interests include computer vision and micro-expression recognition.



**ZHU Jianhong**, born in 1964, Ph.D, professor, Ph.D supervisor. His main research interests include visual Internet of Things and deep learning.