



## 双流增强融合网络微表情识别

陈斌, 朱晋宁

引用本文:

陈斌,朱晋宁. 双流增强融合网络微表情识别[J]. 智能系统学报, 2023, 18(2): 360–371.

CHEN Bin,ZHU Jinning. Micro-expression recognition based on a dual-stream enhanced fusion network[J]. *CAAI Transactions on Intelligent Systems*, 2023, 18(2): 360–371.

在线阅读 View online: <https://dx.doi.org/10.11992/tis.202109036>

## 您可能感兴趣的其他文章

### 面向表情识别的重影非对称残差注意力网络模型

A ghost asymmetric residual attention network model for facial expression recognition

智能系统学报. 2023, 18(2): 333–340 <https://dx.doi.org/10.11992/tis.202201003>

### 深度多尺度融合注意力残差人脸表情识别网络

Deep multiscale fusion attention residual network for facial expression recognition

智能系统学报. 2022, 17(2): 393–401 <https://dx.doi.org/10.11992/tis.202107028>

### 基于循环互相关系数的CGAN温度值图像扩增

Image amplification for temperature value image based on cyclic cross-correlation coefficient CGAN

智能系统学报. 2022, 17(1): 32–40 <https://dx.doi.org/10.11992/tis.202106036>

### 基于双注意力模型和迁移学习的Apex帧微表情识别

Apex frame microexpression recognition based on dual attention model and transfer learning

智能系统学报. 2021, 16(6): 1015–1020 <https://dx.doi.org/10.11992/tis.202010031>

### 基于改进的Faster RCNN面部表情检测算法

Facial expression recognition based on improved Faster RCNN

智能系统学报. 2021, 16(2): 210–217 <https://dx.doi.org/10.11992/tis.201910020>

### 鲁棒的正则化编码随机遮挡表情识别

Recognition of facial expression in case of random shielding based on ro-bust regularized coding

智能系统学报. 2018, 13(2): 261–268 <https://dx.doi.org/10.11992/tis.201609002>

DOI: 10.11992/tis.202109036

网络出版地址: <https://kns.cnki.net/kcms/detail/23.1538.tp.20221019.1132.004.html>

# 双流增强融合网络微表情识别

陈斌, 朱晋宁

(南京师范大学 信息化建设管理处, 江苏 南京 210046)

**摘要:**为解决微表情识别领域数据集样本数量少, 样本类型分布不均导致识别率鲁棒性差的问题, 提出了一种基于双流增强网络的微表情识别模型。该模型基于单帧 RGB 图像流及光流图像流的双流卷积神经网络, 以权威数据集为基础, 数据增强为基准, 构建微表情识别模型。通过在 SoftMax 逻辑回归层融合单帧空域信息和光流时域信息, 对两个独立流的网络性能进行提升, 并通过引入基于带循环约束的生成对抗网络的图像生成方式对数据集进行扩充。通过将输入微表情视频帧序列进行分解, 将其分割为双流网络的灰度单帧序列与光流单帧序列, 对两类序列图进行数据增强, 再进行微表情识别模型构建的方法, 有效提高了微表情识别率。基于双流增强网络的微表情识别模型可以较好提升微表情识别准确度, 鲁棒性较好, 泛化状态较稳定。

**关键词:** 微表情; 双流网络; 生成对抗网络; 数据增强; 特征融合; 模式识别; 卷积神经网络; 循环约束

**中图分类号:** TP39; TH691.9   **文献标志码:** A   **文章编号:** 1673-4785(2023)02-0360-12

中文引用格式: 陈斌, 朱晋宁. 双流增强融合网络微表情识别 [J]. 智能系统学报, 2023, 18(2): 360-371.

英文引用格式: CHEN Bin, ZHU Jinning. Micro-expression recognition based on a dual-stream enhanced fusion network[J]. CAAI transactions on intelligent systems, 2023, 18(2): 360-371.

## Micro-expression recognition based on a dual-stream enhanced fusion network

CHEN Bin, ZHU Jinning

(Informatization Office, Nanjing Normal University, Nanjing 210046, China)

**Abstract:** We propose a micro-expression recognition model based on a dual-stream enhanced network in order to address the issues of insufficient samples from the dataset of micro-expression and uneven distribution of sample types leading to a low rate of robustness. Targeting a dual-stream convolutional neural network of single-frame RGB image flow and optical image flow, a micro-expression recognition model is built based on a fundamental authoritative dataset and data enhancement. Single-frame airspace information and optical time flow domain information are incorporated in the SoftMax logistic regression layer to improve the network performance for two independent streams. The dataset is augmented by introducing a method for image generation based on a generative adversarial network with loop constraints. After segmenting the input micro-expression video frame sequence into greyscale single frame sequences and optical flow single frame sequences of a dual-stream sequence diagram, augmenting the data of the two sequences, and constructing the micro-expression recognition model, the input micro-expression video frame sequence is subdivided. The rate of micro-expression recognition has been significantly enhanced by this method. The micro-expression recognition model based on dual-stream enhanced networks can effectively improve the recognition accuracy of micro-expressions with improved robustness and generalization state.

**Keywords:** micro-expression; dual-stream network; generative adversarial network; data enhancement; fusion of features; pattern identification; convolutional neural network; cycle constraint

收稿日期: 2021-09-19. 网络出版日期: 2022-10-19.

基金项目: 江苏省现代教育技术研究 2021 年度智慧校园专项课题 (2021-R-96609).

通信作者: 陈斌. E-mail: 60167@njnu.edu.cn.

©《智能系统学报》编辑部版权所有

面部微表情 (facial micro-expression, FME) 是引发隐藏某种真实情绪的短暂和不自主的快速面部表情<sup>[1]</sup>。微表情与宏表情不同, 它不受大脑控制, 脱离意识支配, 是最真实的内心感受和心理

状态的折射。标准的微表情持续时间在 1/5 s 到 1/25 s 之间,通常只发生在脸部的特定部位。以往对微表情的分析需要通过相当有经验的专业人士,长时间细致的观察才能做出判断。近年来,借助人工智能技术的发展和计算机视觉与模式识别技术在表情及微表情分析领域的不断突破,智能实时快速微表情识别成为既有挑战又有价值的研究领域。

纵观微表情的研究史,最早始于 1966 年,当时由 Haggard 和 Isaacs 发现了一种产生于一瞬间且不易被感知到的面部表情<sup>[2]</sup>。三年后,由 Ekman 与 Friesen 等人开始了对这类瞬间表情的系统性研究,并为之取名“微表情”<sup>[3]</sup>。在数代人持续的努力下,时至今日,针对微表情领域的研究已经获得了较为成熟的理论框架,也有了一定的方法论基础。

目前微表情识别算法总体上分为两大类:基于传统特征的微表情识别算法及基于深度学习的微表情识别算法。基于传统特征的识别方法主要有局部二值模式及其改良方法<sup>[4]</sup>,基于光流场特征及其改良方法<sup>[5]</sup>,基于张量变化分析及其改良方法<sup>[6]</sup>。Pfister 团队通过使用具有 3 个正交平面的局部二值模式(local binary pattern from three orthogonal planes, LBP-TOP)来处理动态特征,同时以支持向量机和多核学习随机森林方法进行分类识别,首次成功实现了自发面部微表情识别<sup>[7]</sup>。Wang 团队提出的 6 个交点的局部二值模式(LBP with six intersection points, LBP-SIP)的时空描述子,对 LBP-TOP 中特征冗余情况做出了改进<sup>[8]</sup>。Liong 团队使用光学应变加权特征提取方法和光学应变幅值特征提取方法,利用光学应变对运动信息进行表示,以此对面部细微肌肉运动进行计算,并通过时空信息进行微表情识别<sup>[9]</sup>。另外代表性的方法还有 Liu 团队提出的主方向平均光流(main directional mean optical-flow, MDMO)特征微表情识别<sup>[10]</sup>,Happy 团队提出的模糊光流方向直方图(fuzzy histogram of optical flow orientations, FHOFO)特征微表情识别<sup>[11]</sup>。

毫无疑问,基于传统特征的微表情识别研究在相当长的一段时间内为该领域发展做出了积极和关键的贡献。但其仍然存在很多问题,例如由于结构性问题导致的计算量大、复杂性高、兼容性差、交叉学科领域知识门槛高、专业性要求高等,令微表情识别技术的发展逐现迟缓。基于深度学习的微表情识别研究正是在这样的背景下应运而生的。

在计算机视觉分析领域,基于深度学习的方

法提取特征的能力优于传统方法。深度学习是利用神经网络自主对特征进行学习的,其可以解析高层语意信息,故对微表情特征有更好的表征效果。Gan 团队针对微表情序列动幅最大的帧进行了计算,进而分别计算该帧与初始帧之间水平及垂直方向的光流坐标,再将其输入卷积神经网络进行特征提取,该方法实际效果优于单一光流法<sup>[12]</sup>。刘汝涵团队利用基于眼部干扰消除的视频放大方法,结合深度神经网络进行微表情识别,该方法实际效果优于单一神经网络方法<sup>[13]</sup>。微表情识别利用的深度学习方法通常为卷积神经网络或循环神经网络及其相结合的方式,其缺陷在于无法并行对时空域关系进行编排处理,并且由于微表情数据库的极度稀缺导致神经网络构建时规模及深度受限,训练过程中过拟合现象也非常普遍,直接导致微表情识别准确性的下滑。

针对以上问题,本文提出了一种双流增强融合网络微表情识别模型(dual-stream convolution enhanced networks, DSCEN),通过双流网络分别编排处理时空域信息以解决独立并行处理问题,并通过增强方式解决微表情数据集样本数量少,样本类型分布不均导致识别率鲁棒性差的问题。既利用双流网络的时空双域交互信息的叠加作用,同时利用图像增强后的分类效果提升作用,以提高微表情的识别准确度。

## 1 基础原理

### 1.1 微表情识别方法及原理

人类面部由 42 块肌肉组成,它们各自受控于不同的大脑区域,有些肌肉可以受意识控制,我们称之为“自由肌”,有些却不易受意识控制,我们称之为“非自由肌”<sup>[14]</sup>。其中在“非自由肌”的运动痕迹中蕴藏着深刻的微表情信息。微表情识别方法通常分为图像采集、图像预处理、微表情检测、特征提取、微表情分类几个步骤,框架流程如图 1 所示<sup>[15]</sup>。首先要对采集图像进行预处理,预处理阶段对图像进行降噪和局部放大,主要包括面部检测、面部校正、运动放大和时域归一化。微表情检测主要基于外观特征强度值,通过计算局部二值模式(local binary patterns, LBP)距离特征差或 3D 梯度直方图算子作为检测和提取的特征对象。通过提取类似光流或光强等细微表情非刚性的运动变化,或者使用面部动态图(facial dynamic map, FDM)对面部组件进行运动建模,结合多尺度滑动窗口的微表情检测结果与数据集样本进行比对,进而完成微表情的分类<sup>[16]</sup>。



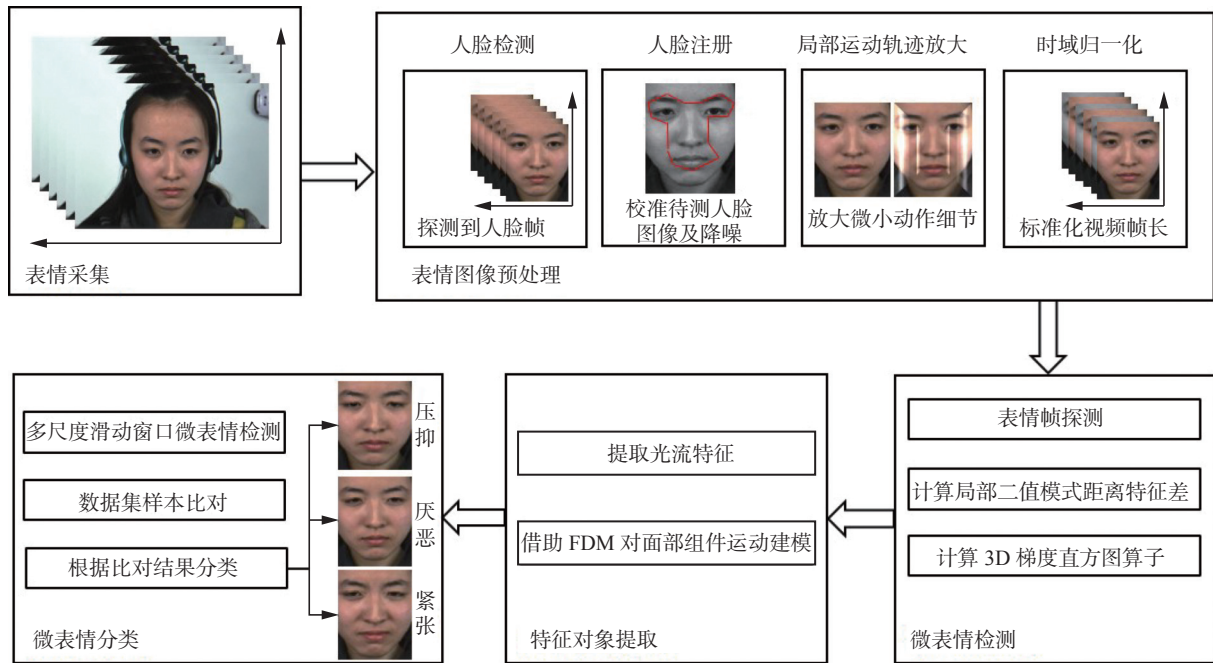


图 1 微表情识别方法常用流程

Fig. 1 Common processes of micro-expression recognition methods

### 1.2 典型微表情实验数据集及其比较

与宏表情相比,微表情数据集样本较为稀缺。以下简述几个具有代表性的微表情数据集。

#### 1.2.1 CASME 数据集

CASME 是一个自发式微表情数据集,由中国科学院心理研究所傅小兰团队建立<sup>[17]</sup>。通过 60 f/s 的高速摄像机对微表情进行图像采集,从中 1500 个微表情序列中遴选了 195 个,通过对序列中动作单元标定标签,实现微表情 8 种分类。

#### 1.2.2 CASMEII 数据集

CASMEII 也是由傅小兰团队建立的一个更为全面的自发式微表情数据集<sup>[18]</sup>,通过 200 f/s 的高速摄像机对微表情进行图像采集,从 3000 个微表情序列中精选出 255 个,由 26 名平均年龄为 22 岁的亚洲参与者组成,分类为 9 种微表情。

#### 1.2.3 SMIC 数据集

SMIC 数据集由芬兰奥卢大学的赵国英团队建立<sup>[19]</sup>。它包含了 SMIC-HS、SMIC-VIS 及 SMIC-NIR 3 个子数据集,通过 100 f/s 的高速摄像机对微表情进行图像采集,共含有 164 个微表情动作序列,划分为 3 个表情分类。

#### 1.2.4 CAS(ME)<sup>2</sup> 数据集

CAS(ME)<sup>2</sup> 数据集由中国科学院建立,它提供宏表情和微表情两个部分(Part A 与 Part B)<sup>[20]</sup>。A、B 两部分共包含了 387 个裁剪的自发宏表情样本或微表情样本<sup>[21]</sup>。情感标签基于动作单位、自我报告情感以及情感类型的组合<sup>[22]</sup>。局部二值模式

LBP 用于宏表情和微表情的检测和识别<sup>[23]</sup>。CAS(ME)<sup>2</sup> 数据库提供了长视频以及裁剪的表情样本,这将有助于科研人员用于开发有效的算法来识别宏表情和微表情<sup>[24]</sup>。它涵盖了 4 种微表情类别。

#### 1.2.5 典型实验数据集的比较

表 1 为几种典型代表型微表情数据集关键属性参数比对<sup>[25-26]</sup>。

表 1 典型实验数据集关键属性参数比对

Table 1 Compare of key attribute parameters in typical experimental datasets

| 属性参数                    | CASME  | CASMEII  | SMIC           | CAS(ME) <sup>2</sup> |
|-------------------------|--|--|----------------|----------------------|
| 参与者数                    | 35   | 35   | 16/8/8         | 22                   |
| 样本数                     | 195  | 247  | 164/71+        | 57                   |
| 表情分类                    | 蔑视<br>厌恶<br>恐惧<br>高兴<br>压抑<br>伤心<br>惊讶<br>紧张 | 高兴<br>惊讶<br>厌恶<br>恐惧<br>伤心<br>生气<br>压抑<br>紧张<br>消极 | 积极<br>消极<br>惊讶 | 积极<br>消极<br>惊讶<br>其他 |
| 分辨率/ppi                 | 720×1280                                     | 640×480  | 640×480        | 640×480              |
| 帧率/(f·s <sup>-1</sup> ) | 60   | 200  | 100/25+        | 30                   |

### 1.3 微表情识别共性问题及解决思路

在微表情识别领域中,各类方法均不能有效处理人脸局部变亮的问题;特征提取过程中利用

各类过程与方式,均存在数据集样本数量少,样本类型分布不均从而导致识别率鲁棒性差的问题。为解决该问题,本文提出基于单帧 RGB 图像流及光流图像流的双流卷积神经网络,以典型权威数据集为基础,数据增强处置条件为基准,构建微表情识别模型并辅以实验验证。致力于借助有效图片增强为载体,充分结合 CNN 共享卷积核,处理高维数据高性能的特点,为微表情研究领域的共性问题建立解决途径。

## 2 理论原型及构架

### 2.1 双流卷积神经网络图像特征处理

微表情的开始、变化、直至结束可以视作一个视频序列<sup>[27]</sup>。视频序列可以被分解为空域和时域两个维度<sup>[28]</sup>。空域维度由单帧构成,即 RGB 图像帧,作为空间信息的载体,包含环境及物体等空间信息<sup>[29]</sup>。时域维度以光流图像信息构成,承载了待观测对象的运动情况,将光流信息作为时序信息载体输入卷积神经网络,用以理解动作的动态特征<sup>[30]</sup>。空域和时域两个独立的信息流在 SoftMax 逻辑回归层通过以多元 L2 范式标准化的 SoftMax 逻辑回归分数作为特征的多类别线性 SVM 方式进行分数融合,组成双流卷积神经网络架构<sup>[31]</sup>。

双流卷积神经网络目的在于模仿与理解视频信息的视觉分析过程,通过卷积神经网络对获得的数据样本进行特征提取及分类<sup>[32]</sup>。通过将单帧彩色图像与单帧光流图像以及叠加后的光流图像作为网络的输入,再分别对图像进行分类,并对不同模型分类结果进行融合,可以显著提高两个独立流网络的性能。为提升双流神经网络效能,原始双流网络的时空域信息结构采用中等规模的卷积神经网络结构。一个密集的光流序列可以看作是连续帧  $f$  与  $f+1$  之间的一组位移向量场  $d_f$ 。光流场在水平和垂直的分量,可以视为图像通道。 $d_f(x,y)$  表示在第  $f$  帧中坐标点  $(x,y)$  处的位移矢量。光流可以进行叠加,将  $K$  个连续帧的光流通道相堆叠,形成  $2K$  个输入通道,表示了一系列帧之间的运动。任意帧的卷积网络输入卷可表示为

$$I_{\sigma}(x,y,2f-1)=d_{\sigma+f-1}^m(x,y)$$

$$I_{\sigma}(x,y,2f)=d_{\sigma+f-1}^n(x,y)$$

$$I_{\sigma} \in \mathbb{R}^{u \times v \times 2K}, x=[1:u], y=[1:v], t=[1:K]$$

输入卷堆叠了  $K$  帧序列的水平向量场  $m$  和垂直向量场  $n$ ,  $\sigma$  为位移偏量。

双流神经网络神经元权重初始化是神经网络

训练任务中一个非常关键的因素。神经元初始权重的选择将直接影响到网络的收敛速度及训练的效果。常用的初始化方法包括全局复位法、高斯随机法、标准方差法等。在实践中,因数据集的体量有限,造成迭代次数较少,很容易引起网络过拟合或不收敛的情况,实际效果并不理想。所以通常并不直接训练卷积神经网络,而是使用迁移学习法对预训练网络在不同数据集上的运行结果进行微调,从而改善训练效果。

光流特征图的提取方式是利用以 TV-L1<sup>[33]</sup> 的近似方法,其优势在于对噪声的鲁棒性及对持续的运动流处置的稳定性。图像的运动变化轨迹及流动方向与强度采用光流编码作为表征,光流编码使用矢量参数作为标记。光流编码的水平量值与垂直量值表示如下:  $l = \frac{dh}{dc}$ ,  $p = \frac{dv}{dc}$ 。其中,  $dh$  表示待观测对象沿着水平维度的运动轨迹,  $dv$  表示待观测对象沿着垂直维度的运动轨迹,  $dc$  代表时间周期。光流强度可以通过光流的水平量值与垂直量值计算得到:  $S = \sqrt{l^2 + p^2}$ 。将  $l$ 、 $p$  及  $S$  在通道层面进行融合,得到光流融合通道特征图。

双流网络在表情识别中的作用主要在图像的时空域信息的提取上,卷积网络在表情识别中的作用主要在表情特征提取及分类上。体现在各自信息提取的输入、计算、分类和输出的差异上。

### 2.2 数据增强融合

在人脸表情识别领域,有效高质量训练样本数量的缺乏为普遍的共识<sup>[34]</sup>。在微表情识别领域,该问题显得更为突出<sup>[35]</sup>,为数不多的典型数据集存在着类别分布不平衡,分类质量差,数据量少,模型训练结果容易出现过拟合现象,识别效果鲁棒性差等情况。通过数据增强的方式针对数据集有效性提升方面在表情识别领域可以获得较好的补充效果,这部分内容在本研究团队前期工作中已经得到有效的验证<sup>[36]</sup>。在本文微表情识别的实现中,也引入了这一机制,通过图像生成的方式,将微表情数据集进行大规模扩容增补。图像的生成方式同样使用了具有代表性的生成对抗网络 (generative adversarial network, GAN) 改进方法。该生成方法的基本思路为二人零和博弈,即二人的利益之和为零,一方的所得必然为另一方的所失<sup>[37]</sup>。它属于无监督概率学习方法,可以学习数据的实际分布情况,继而产生同其有关的新数据集<sup>[38]</sup>。

生成对抗网络模型由生成模型与判别模型共同组成,生成模型的作用是确定训练集数据分

布属性,判别模型则用来判别待检测数据是源数据还是衍生数据<sup>[39]</sup>。传统的生成对抗网络训练稳定性较差,生成的数据样本分类较为单一,继而导致生成的图像质量并不理想,缺乏可用性。一般的处理方式是对生成网络赋予生成约束条件,也即对生成网络进行有条件的补充。较为典型的是带循环约束的生成对抗网络,在本研究团队此前的基于残差整流增强卷积神经网络的表情识别实验中选择该网络获得了较好的实验效果,在本文微表情识别训练中同样选择了该方法。带循环约束的生成对抗网络由两个生成模型和两个判别模型组成,两组生成判别模型分别对应输入源和目标域的映射关系。该类生成对抗网络的损失主要由对抗损失和循环一致性损失构成<sup>[40]</sup>。理论上,对抗损失不会对目标域分布结果偏离源域分布形成影响,但实际上受学习效果的影响,通常偏离范围会超出预期,循环一致性损失却可以修

复该问题。本文将会在实验中采用该循环一致性生成对抗网络进行图像的增强,由原始微表情数据集图像和增强后的生成图像共同支撑实验的进行。

### 2.3 双流增强网络的微表情识别模型

为了解决人脸微表情发生时间短,变化不明显的问题,本文引入了基于双流增强卷积神经网络的微表情识别模型(DSCEN),既利用双流网络的时空双域交互信息的叠加作用,又利用图像增强后的分类效果提升作用,以提高微表情的识别准确度。

本文将输入微表情视频帧序列进行分解,通过分割为双流网络的灰度单帧序列与光流单帧序列,对两类序列图进行数据增强,再进行微表情识别模型的构建,其中包括了微表情模型的训练和测试,以及微表情识别模型分类与决策,整体模型框架如图2所示。

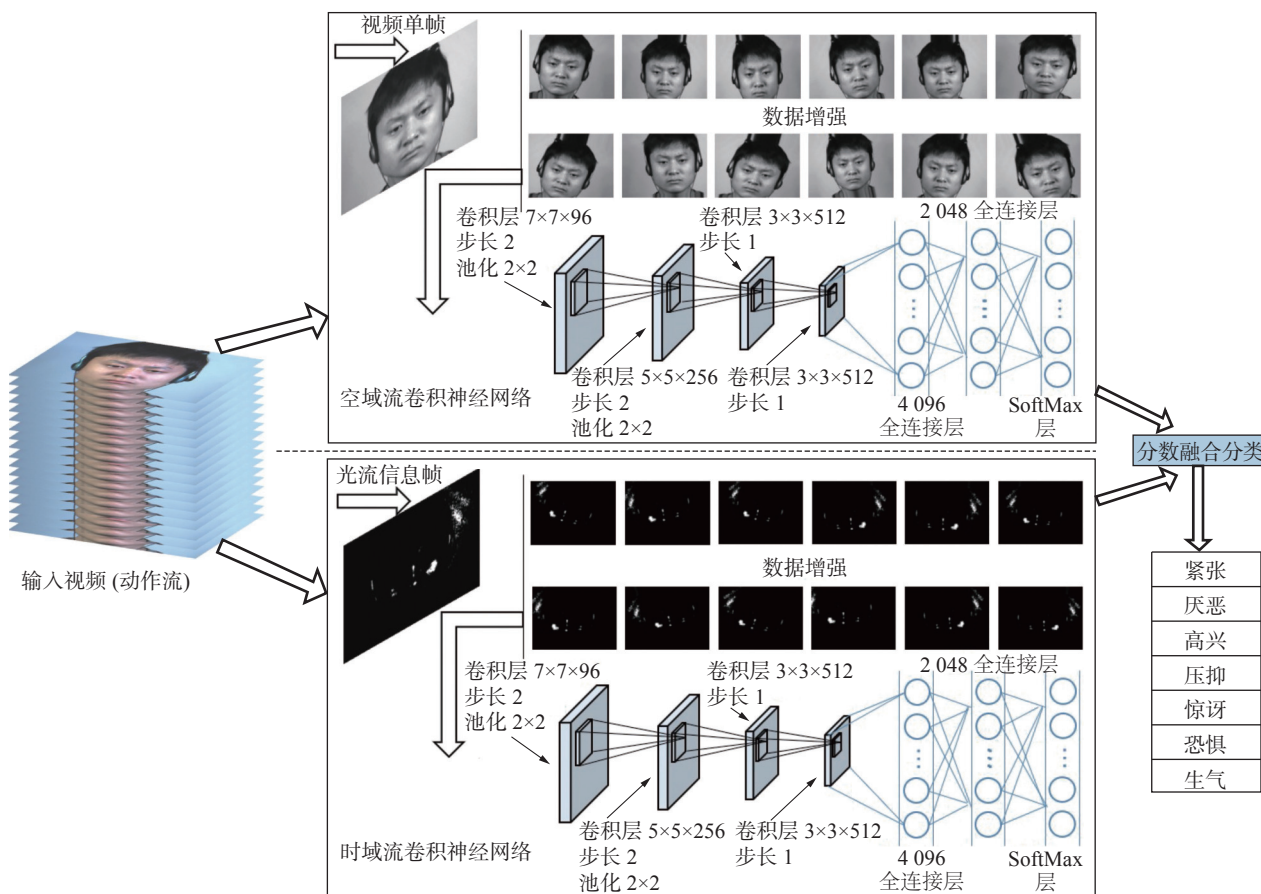


图2 双流增强网络的微表情识别模型

Fig. 2 Micro-expression recognition model of dual-stream enhanced network

双流增强网络的微表情识别模型主框架由空域流卷积神经网络与时域流卷积神经网络的双流网络结构构成,其中融合了数据增强处理,并将两个独立神经网络的处理结果分别独立输出到分

数融合分类处理,最终得到微表情分类结果。在模型训练过程中,当框架模型接收到外部输入微表情基准视频即动作流信号时,将分别转发到空域流卷积神经网络及时域流卷积神经网络。空域



流卷积神经网络首先对视频流进行单帧 RGB 图像流切片及递归处理,继而对 RGB 单帧信息进行数据增强处理,扩容微表情图像数据集,进一步进行多轮卷积处理以及全连接处理,并在 SoftMax 逻辑回归层进行多元 L2 范式标准化处理。类似的,在时域流卷积神经网络上,当收到外部输入微表情基准视频即动作流信号时,首先对视频流进行图像流切片并生成相应的光流信息帧,再对光流信息帧进行数据增强处理以扩充微表情光流数据集,同样的经过多轮卷积处理及全连接处理后,在 SoftMax 逻辑回归层进行多元 L2 范式标准化处理。对双流网络 SoftMax 逻辑回归层结果进行分数融合分类,形成双流增强网络的微表情识别模型。由于该模型继承了空域维度的空间信息用以识别对象的物理属性,同时容纳了时域维度的光流信息用以识别对象的运动特征,并在两个维度中融合了数据增强处理用以提升和稳定拟合

效果及弱化归一化差异,故而较好地提升了微表情识别的准确度,使得模型识别效果拥有较稳定的鲁棒性,且增强了模型泛化处理状态的稳定性。

### 3 实验实现及结果分析

#### 3.1 实验环境及设置

本实验采用 64 位 Ubuntu 操作系统, 64 GB DDR SDRAM 内存, NVIDIA GeForce GTX 1070Ti GPU 加速。基础平台采用 Anaconda3 2020.02(64-bit) 集成包, 开发平台采用在搭建深度学习模型方面具有独特优势的 PyCharm(版本号: 2021.181.3263.12), 神经网络架构搭建采用的是 TensorFlow 2.3。通过内嵌 Python Programming API, 实现图像元素多节点均摊负载的并行计算模式, 使得计算效率大幅度提升。实验在光源充足无干扰环境下, 通过近距离高分辨率正脸表情帧流捕捉条件下进行。图 3 给出了现场实验环境。

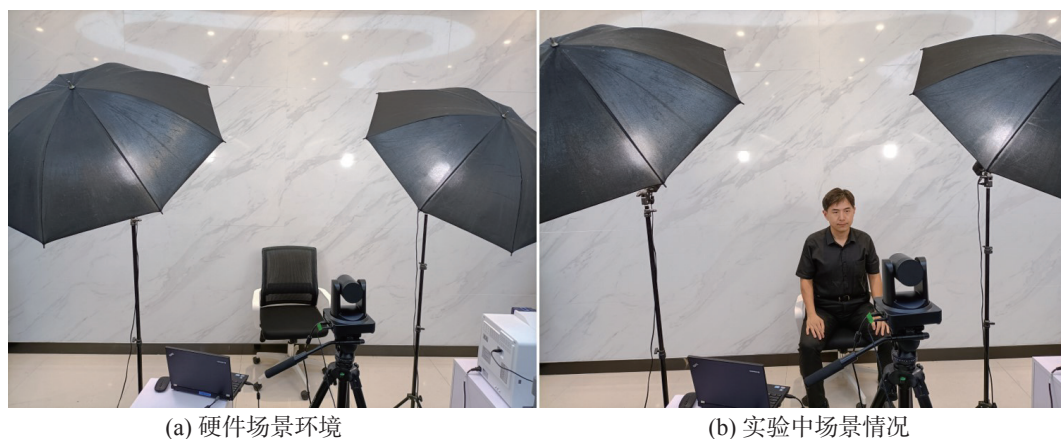


图 3 实验环境展示

Fig. 3 Exhibition of the experiment environment

#### 3.2 实验数据集

在通过实验验证基于双流增强网络的微表情识别方法的效能的过程中,首先需要选择并构建实验所需数据集用于训练和测试。本文选取了领域内具备权威代表性的微表情数据集 CASMEII 作为基础数据集,通过数据增强的方式,对数据集进行了扩充。之所以选择 CASMEII 数据集,是因为其五大特性: 1) 该数据集的自发性和动态性比同类其他数据库强。2) 其具备较高的时间分辨率(200 f/s)和较高的面部分辨率(280 像素×340 像素)。3) 微表情标签是基于面部表情编码系统 FACS,在领域内具有绝对的权威性。4) 该数据集的录制过程中光照柔和饱满,面部高光区较少,噪声占比较低,有助于多元情况下衡量实验结果的稳定性。5) 数据集分布广泛,覆盖面

宽。虽然 CASMEII 数据集包含了 9 种微表情分类,但重点分布集中在紧张(Tense)、厌恶(Disgust)、高兴(Happiness)、惊讶(Surprise)、压抑(Repression)这 5 类,另外恐惧(Fear)和生气(Anger)也有较多原始样本,而悲伤(Sadness)和消极(Negative)样本数量非常有限。本文实验数据集仅聚焦于前 7 类微表情的训练和识别。图 4 是 CASMEII 微表情数据集样例,图 5 是基于该数据集的通过带循环约束的生成对抗网络模型所产生的增补数据样例,它们共同作为本文实验部分所使用的训练数据集和测试数据集。本实验模型的训练使用了 10 356 个样本,其中用于训练集的为 6 784 张,测试集的为 3 572 张,为了避免过拟合,训练中采用 64 epoch 为一组的单网格模式,通过 GPU 加速,训练用时 17.6 h。

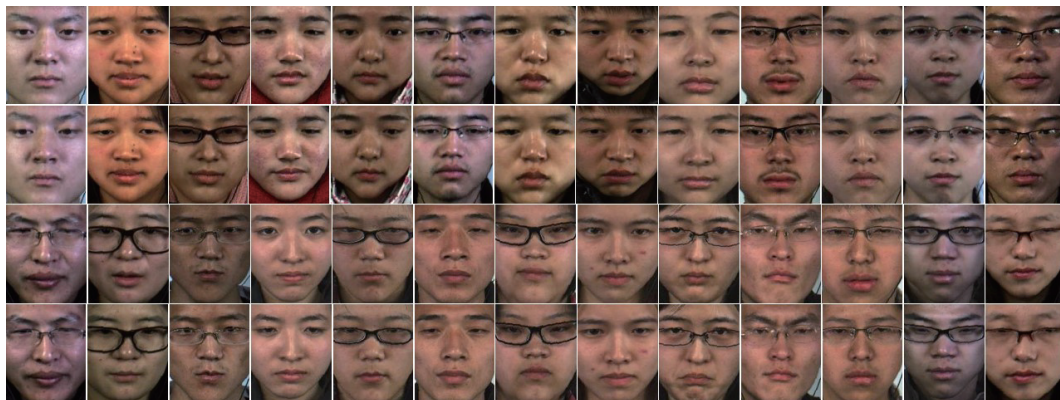


图 4 CASMEII 数据集微表情图像示例

Fig. 4 Example of the micro-expression images of the CASMEII dataset



图 5 CASMEII 数据集微表情生成图示例

Fig. 5 Example of the micro-expression generate images of the CASMEII dataset

### 3.3 实验结果及分析

本实验基于非开放条件下固定采集模式, 待测对象是基于近距离非干扰环境下的受控检测。测试环境的特点在于待测对象位置恒定, 光照均匀充足, 受光线变化的影响较小, 近距离的高分辨率快速采集设备对瞬间发生的微表情捕捉效果较好。

我们进行了两组实验, 第 1 组采用双流网络框架做模型训练和识别分类, 第 2 组在实验中结合了数据增强处置方法。根据实验数据结果看, 第 2 组识别率效果明显优于第 1 组。实验通过增强处理后的数据集进行训练, 并经过多轮算法优化及回归堆叠, 识别率在单一表情象限最高达到 91.81%, 未增强前最高仅为 69.05%, 综合分布平



均最高识别率最高达到 57.01%, 未增强前最高仅为 30.48%。两组实验结果比对见表 2。图 6 为基于本文算法模型的实验系统例图截取。根据实验结果分析, 7 类微表情分类中, 识别准确率最高的是高兴 (Happiness), 对其最高识别率达到 91.81%, 对其平均识别率为 86.35%。识别率最低的是压抑 (Repression), 对其最高识别率仅为 34.55%, 平均识别率为 9.43%。导致这种结果主要是因为训练样本集中压抑类别的样本数量最小, 在微表情特征中, 压抑表情特征较不明显, 其运动方向是从眉毛内沿肌肉连接到额骨上端, 也就是眉毛内侧会被牵拉向上微动。从视觉上看, 多数情况一般人在这种表情下会形成一个类八字眉, 并且前额中部的皮肤会产生水平的褶皱, 这些褶皱通常不会贯穿全部的前额而是局限在前额中部, 褶皱一般呈弧线, 皱褶中部比尾端提升略高一点, 并非是维持水平。由于这些特征与生气 (Anger)、恐惧 (Fear) 和紧张 (Tense) 微表情的交集较多, 界定不清晰, 易引起交叉分类。另外由于其特征在数据增强后不易保留, 继而增强后的

样本数量和质量与其他分类相比不够丰富有效, 数据增强效果差, 较难得到更优的训练效果。同时我们观察到恐惧 (Fear) 与生气 (Anger) 微表情在增强前后的结果甚至有反转情况, 这是与生成图像的特征继承有效性相关的, 特征继承性高则效果更佳。

表 2 实验组 1 与实验组 2 微表情识别率比较

Table 2 Comparison of micro-expression recognition rates between experimental group 1 and experimental group 2

| 微表情分类 | 第1组   | 第2组   |
|-------|-------|-------|
| 紧张    | 22.61 | 51.58 |
| 厌恶    | 39.68 | 73.75 |
| 高兴    | 69.05 | 91.81 |
| 压抑    | 12.17 | 34.55 |
| 惊讶    | 32.84 | 59.61 |
| 恐惧    | 19.26 | 43.16 |
| 生气    | 17.75 | 44.63 |

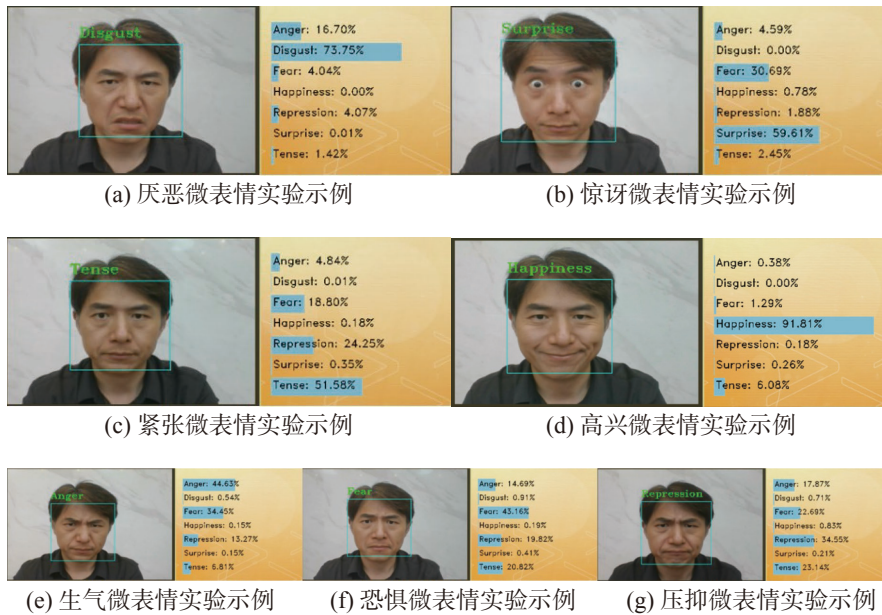


图 6 微表情识别分析实验系统结果示例

Fig. 6 Example results of the experimental system for micro-expression recognition and analysis

### 3.4 跨数据集实验结果比较

为了验证双流增强网络方法训练后的识别效果, 分析算法在多元条件下鲁棒性的表现, 以及在不同运行环境中的泛化状态, 本实验的整个过程也选择在 CASME 和 CAS(ME)<sup>2</sup> 2 个数据集上分别进行了验证。因为 CASME 数据集较 CASMEII 数据集样本量少, CAS(ME)<sup>2</sup> 数据集在数量

上差距更为悬殊, 且从样本分类角度而言, 两者分类结构均覆盖于 CASMEII 分类范围中, 所以选中了 CASME 的 195 个全量样本和 CAS(ME)<sup>2</sup> 的 57 个全量样本投入实验。实验程序与 CASMEII 类似, 同样的双流结构卷积神经网络在对 2 个数据集分布进行增强后实施训练和识别实验。每个数据集上同样分别进行了增强前后的两组实验,

从最终实验数据看,同一数据集增强前后识别率有显著提升,不同数据集之间在同一表情归类上的识别率有一定的差别,增强后的 3 个数据集上的实验结果比对见表 3(“—”代表数据集不支持的微表情类型)。由此可见,双流卷积神经网络在增强前后对数据集训练效果有较显著的影响,但在不同数据集之间影响效果有较明显的差异,造成这种差异的原因与数据集规模和质量均有着直接的关系。由于不同数据集对微表情定义及名称的差别,在本实验比对中,对部分名称不同但特征交集较大的微表情划分为横向比对类,具体表现为: CAS(ME)<sup>2</sup> 数据集与 CASMEII 数据集之间, Positive 类别映射 Happiness 类别, Negative 类别映射 Repression 类别。

表 3 跨数据集微表情识别率实验结果比较

Table 3 Comparison of experimental results on micro-expression recognition rates in different datasets %

| 微表情分类 | CASMEII | CASME | CAS(ME) <sup>2</sup> |
|-------|---------|-------|----------------------|
| 紧张    | 51.58   | 34.29 | —                    |
| 厌恶    | 73.75   | 67.02 | —                    |
| 高兴    | 91.81   | 80.27 | 68.50                |
| 压抑    | 34.55   | 20.05 | 37.69                |
| 惊讶    | 59.61   | 41.91 | 19.76                |
| 恐惧    | 43.16   | 38.00 | —                    |
| 生气    | 44.63   | —     | —                    |

### 3.5 与其他代表性研究成果的比较

将本文的研究成果与其他代表性研究成果进行比较。主要以如下 8 个典型方法为例: LBP-TOP<sup>[41]</sup>、LBP-SIP<sup>[42]</sup>、ELRCN (enriched long-term recurrent convolutional network)<sup>[43]</sup>、FDM (facial dynamics map)<sup>[44]</sup>、MRW (monogenic risez wavelet)<sup>[45]</sup>、STCLQP (spatio temporal completed local quantization patterns)<sup>[46]</sup>、MDMO (main directional mean optical-flow)<sup>[47]</sup>、Bi-WOOF (bi-weighted oriented optical flow)<sup>[48]</sup>。有关研究成果之间的指标比较如表 4 所示,所有实验均在相同数据集 CASMEII 条件下进行比较。从对比结果可以得出,本实验所提出的模型算法取得了较为理想的微表情分类识别效果。主要原因有以下几个方面: 1) 从网络结构角度,本文的双流网络模型将时空域分别形成独立网络,各自承载不同的特征信息单独处理,更接近于实际理解视频信息的视觉分析过程。2) 从网络空间角度,本文通过数据增强方式将双流网络实验数据集进行了拓展,延伸了网络训练的宽度,增强了网络

分类的准确性和鲁棒性。3) 从增强融合的效果角度,时空双域交互叠加融合图像增强,无论在 RGB 图像流信息帧,还是光流信息帧均起到了较好的作用。

表 4 有关研究成果之间的指标比较

Table 4 Comparison of the indicators between the relevant research results %

| 代表性研究成果典型方法 | 最高识别准确率 | 平均识别准确率 |
|-------------|---------|---------|
| LBP-TOP     | 58.98   | 47.00   |
| LBP-SIP     | 65.34   | 46.56   |
| ELRCN       | 50.00   | 39.37   |
| FDM         | 45.93   | 40.53   |
| MRW         | 46.15   | 43.07   |
| STCLQP      | 58.39   | 55.88   |
| MDMO        | 56.26   | 44.16   |
| Bi-WOOF     | 61.25   | 57.89   |
| DSCEN(本文)   | 91.81   | 57.01   |

## 4 结束语

针对基于深度学习的微表情识别算法并行时空域关系的处理障碍,以及因微表情数据库的缺乏导致网络构建时的功能障碍,进而导致微表情识别准确度下降的问题,本文提出了一种基于双流增强网络的微表情识别模型(DSCEN)。实验采用了 CASMEII 数据集作为基础验证数据集,并以 CASME 和 CAS(ME)<sup>2</sup> 数据集辅以有效性验证。根据多轮分组综合实验情况显示,本模型识别率在单一表情象限最高达到 91.81%,较未增强前提高了 22.76%,综合分布平均最高识别率最高达到 57.01%,较未增强前提高了 26.53%,由此看出本模型可以较好提升微表情识别准确度,鲁棒性较好,泛化状态较稳定。

未来的工作重点将集中在提高微表情特征提取能力方面,主要体现在网络结构优化和数据增强效能提升上。希望从网络参数着手,找到一种与本双流网络结构更加适配的微表情特征描述算子,以期获得更准确、稳定及高效的微表情识别效果。

## 参考文献:

- [1] 朱伟杰,陈莹. 双流网络信息交互机制下的微表情识别[J]. 计算机辅助设计与图形学学报, 2021, 33(4): 545-552.  
ZHU Weijie, CHEN Ying. Micro-expression recognition based on dual-stream networks information interaction[J].



- Journal of computer-aided design & computer graphics, 2021, 33(4): 545–552.
- [2] HAGGARD E A, ISAACS K S. Micromomentary facial expressions as indicators of ego mechanisms in psychotherapy[M]//Methods of Research in Psychotherapy. Boston: Springer US, 1966: 154–165.
  - [3] EKMAN P, FRIESEN W V. Nonverbal leakage and clues to deception[J]. *Psychiatry*, 1969, 32(1): 88–106.
  - [4] KARIS M S, RAZIF N R A, ALI N M, et al. Local binary pattern (LBP) with application to variant object detection: a survey and method[C]//12th IEEE International Colloquium on Signal Processing & its Applications. Melaka: IEEE, 2016: 221–226.
  - [5] XU Feng, ZHANG Junping, WANG J Z. Microexpression identification and categorization using a facial dynamics map[J]. *IEEE transactions on affective computing*, 2017, 8(2): 254–267.
  - [6] WANG Sujing, CHEN Huiling, YAN Wenjing, et al. Face recognition and micro-expression recognition based on discriminant tensor subspace analysis plus extreme learning machine[J]. *Neural processing letters*, 2014, 39(1): 25–43.
  - [7] PFISTER T, LI Xiaobai, ZHAO Guoying, et al. Recognising spontaneous facial micro-expressions[C]//2011 International Conference on Computer Vision. Barcelona: IEEE, 2011: 1449–1456.
  - [8] WANG Yandan, SEE J, PHAN R C W, et al. LBP with six intersection points: reducing redundant information in LBP-TOP for micro-expression recognition[M]//Computer Vision-ACCV 2014. Cham: Springer International Publishing, 2015: 525–537.
  - [9] LIONG S T, SEE J, PHAN R C W, et al. Subtle expression recognition using optical strain weighted features[M]//Computer Vision-ACCV 2014 Workshops. Cham: Springer International Publishing, 2015: 644–657.
  - [10] LIU Yongjin, ZHANG Jinkai, YAN Wenjing, 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.
  - [11] HAPPY S L, ROUTRAY A. Fuzzy histogram of optical flow orientations for micro-expression recognition[J]. *IEEE transactions on affective computing*, 2019, 10(3): 394–406.
  - [12] 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.
  - [13] 刘汝涵, 徐丹. 视频放大和深度学习在微表情识别任务上的应用[J]. 计算机辅助设计与图形学学报, 2019, 31(9): 1535–1541.
  - LIU Ruhan, XU Dan. Video amplification and deep learning in micro-expression recognition[J]. *Journal of computer-aided design & computer graphics*, 2019, 31(9): 1535–1541.
  - [14] 徐峰, 张军平. 人脸微表情识别综述[J]. *自动化学报*, 2017, 43(3): 333–348.
  - XU Feng, ZHANG Junping. Facial microexpression recognition: a survey[J]. *Acta automatica sinica*, 2017, 43(3): 333–348.
  - [15] 苏育挺, 王蒙蒙, 刘婧, 等. 基于多运动特征融合的微表情识别算法[J]. *激光与光电子学进展*, 2020, 57(14): 353–358.
  - SU Yuting, WANG Mengmeng, LIU Jing, et al. Micro-expression recognition algorithm based on multiple motive feature fusion[J]. *Laser & optoelectronics progress*, 2020, 57(14): 353–358.
  - [16] 诗雨桐, 袁德成. 基于深度学习的面部微表情识别[J]. *沈阳化工大学学报*, 2021, 35(4): 380–384.
  - SHI yutong, YUANG decheng. Facial microexpression recognition based on deep learning[J]. *Journal of Shenyang university of chemical technology*, 2021, 35(4): 380–384.
  - [17] 朱伟杰, 陈莹. 双流时间域信息交互的微表情识别卷积网络[J]. *计算机科学与探索*, 2022, 16(4): 950–958.
  - ZHU Weijie, CHEN Ying. Micro-expression recognition convolutional network for dual-stream temporal-domain information interaction[J]. *Journal of frontiers of computer science and technology*, 2022, 16(4): 950–958.
  - [18] 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.
  - [19] LI Xiaobai, PFISTER T, HUANG Xiaohua, 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. Shanghai: IEEE, 2013: 1–6.
  - [20] QU Fangbing, WANG Sujing, YAN Wenjing, et al. CAS(ME)2: A database of spontaneous macro-expressions and micro-expressions[C]//International Conference on Human-Computer Interaction. Cham: Springer, 2016: 48–59.
  - [21] 潘通, 李文国. 基于特征拼接的行人重识别[J]. *激光与光电子学进展*, 2019, 56(16): 193–199.
  - PAN Tong, LI Wenguo. Person re-identification based on feature stitching[J]. *Laser & optoelectronics progress*, 2019, 56(16): 193–199.

- [22] PAN Hang, XIE Lun, LI Juan, et al. Micro-expression recognition by two-stream difference network[J]. *IET computer vision*, 2021, 15(6): 440–448.
- [23] 徐玮, 郑豪, 杨种学. 基于双注意力模型和迁移学习的Apex 帧微表情识别 [J]. *智能系统学报*, 2021, 16(6): 1015–1020.
- XU Wei, ZHENG Hao, YANG Zhongxue. Apex frame micro-expression recognition based on dual attention model and transfer learning[J]. *CAAI transactions on intelligent systems*, 2021, 16(6): 1015–1020.
- [24] LI Jing, WANG Yandan, SEE J, et al. Micro-expression recognition based on 3D flow convolutional neural network[J]. *Pattern analysis and applications*, 2019, 22(4): 1331–1339.
- [25] 张爱梅, 徐杨. 注意力分层双线性池化残差网络的表情识别 [J]. *计算机工程与应用*, 2020, 56(23): 161–166.
- ZHANG Aimei, XU Yang. Attention hierarchical bilinear pooling residual network for expression recognition[J]. *Computer engineering and applications*, 2020, 56(23): 161–166.
- [26] TEJA REDDY S P, TEJA KARRI S, DUBEY S R, et al. Spontaneous facial micro-expression recognition using 3D spatiotemporal convolutional neural networks[C]//2019 International Joint Conference on Neural Networks (IJCNN). Budapest: IEEE, 2019: 1–8.
- [27] ZHOU Ling, MAO Qirong, XUE Luoyang. Dual-inception network for cross-database micro-expression recognition[C]//14th IEEE International Conference on Automatic Face & Gesture Recognition. Lille: IEEE, 2019: 1–5.
- [28] 张艺涵, 张朝晖, 霍丽娜, 等. 结合双流特征融合及对抗学习的图像显著性检测 [J]. *计算机辅助设计与图形学学报*, 2021, 33(3): 376–384.
- ZHANG Yihan, ZHANG Zhaohui, HUO Lina, et al. Image saliency detection via two-stream feature fusion and adversarial learning[J]. *Journal of computer-aided design & computer graphics*, 2021, 33(3): 376–384.
- [29] 张红颖, 安征. 基于改进双流时空网络的人体行为识别 [J]. *光学精密工程*, 2021, 29(2): 420–429.
- ZHANG Hongying, AN Zheng. Human action recognition based on improved two-stream spatiotemporal network[J]. *Optics and precision engineering*, 2021, 29(2): 420–429.
- [30] YANG Dun, ZHOU Qing, WEN Ju. Interactive two-stream graph neural network for skeleton-based action recognition[J]. *Journal of electronic imaging*, 2021, 30: 033025.
- [31] ZHU Dehui, DU Bo, ZHANG Liangpei. Two-stream convolutional networks for hyperspectral target detection[J]. *IEEE transactions on geoscience and remote sensing*, 2021, 59(8): 6907–6921.
- [32] KIM N, CHO S, AHN C H, et al. Facial micro-expression recognition in video using squeezed landmark feature maps[C]//2021 International Conference on Information and Communication Technology Convergence. Jeju Island: ICT convergence, 2021: 1107–1110.
- [33] ZACH C, POCK T, BISCHOF H. A duality based approach for realtime TV-L 1 optical flow[M]//Lecture Notes in Computer Science. Berlin: Springer Berlin Heidelberg, 2007: 214–223.
- [34] 李勇, 林小竹, 蒋梦莹. 基于跨连接 LeNet-5 网络的面部表情识别 [J]. *自动化学报*, 2018, 44(1): 176–182.
- LI Yong, LIN Xiaozhu, JIANG Mengying. Facial expression recognition with cross-connect LeNet-5 network[J]. *Acta automatica sinica*, 2018, 44(1): 176–182.
- [35] GOYAL S J, UPADHYAY A K, JADON R S, et al. Retraction note to: real-life facial expression recognition systems: a review[M]//Smart Computing and Informatics. Singapore: Springer Singapore, 2019: C1–C1.
- [36] 陈斌, 朱晋宁, 东一舟. 基于残差整流增强卷积神经网络的表情识别 [J]. *液晶与显示*, 2020, 35(12): 1299–1308.
- CHEN Bin, ZHU Jinning, DONG Yizhou. Expression recognition based on residual rectifier enhanced convolution neural network[J]. *Chinese journal of liquid crystals and displays*, 2020, 35(12): 1299–1308.
- [37] 王秀友, 范建中, 刘华明, 等. 基于概率协作表示的多表情序列融合识别 [J]. *激光与光电子学进展*, 2019, 56(13): 87–93.
- WANG Xiuyou, FAN Jianzhong, LIU Huaming, et al. Multi-expression sequence fusion recognition based on probabilistic cooperative representation[J]. *Laser & optoelectronics progress*, 2019, 56(13): 87–93.
- [38] 张红颖, 王汇三, 胡文博. 基于双模型的相关滤波跟踪 [J]. *光学精密工程*, 2019, 27(11): 2450–2458.
- ZHANG Hongying, WANG Huisan, HU Wenbo. Correlation filter tracking algorithm based on double model[J]. *Optics and precision engineering*, 2019, 27(11): 2450–2458.
- [39] HUU P N, NGOC T L, MINH Q T. Proposing gesture recognition algorithm using two-stream convolutional network and LSTM[C]//IEEE Eighth International Conference on Communications and Electronics. Phu Quoc Island: IEEE, 2020: 427–432.
- [40] 宋立飞, 翁理国, 汪凌峰, 等. 多尺度输入 3D 卷积融合双流模型的行为识别方法 [J]. *计算机辅助设计与图形学学报*, 2018, 30(11): 2074–2083.



- SONG Lifei, WENG Ligu, WANG Lingfeng, et al. Multi-scale 3D convolution fusion two-stream networks for action recognition[J]. *Journal of computer-aided design & computer graphics*, 2018, 30(11): 2074–2083.
- [41] 卢官明, 杨成, 杨文娟, 等. 基于 LBP-TOP 特征的微表情识别 [J]. *南京邮电大学学报(自然科学版)*, 2017, 37(6): 1–7.
- LU Guanming, YANG Cheng, YANG Wenjuan, et al. Micro-expression recognition based on LBP-TOP features[J]. *Journal of Nanjing university of posts and telecommunications (natural science edition)*, 2017, 37(6): 1–7.
- [42] 姜万, 周晓彦, 徐华南, 等. 基于 LBP 与双时空神经网络的微表情识别 [J]. *信息与控制*, 2020, 49(6): 673–679.
- JIANG Wan, ZHOU Xiaoyan, XU Huanan, et al. Micro-expression recognition based on LBP and two-stream spatial-temporal neural network[J]. *Information and control*, 2020, 49(6): 673–679.
- [43] KHOR H Q, SEE J, PHAN R C W, et al. Enriched long-term recurrent convolutional network for facial micro-expression recognition[C]//13th IEEE International Conference on Automatic Face & Gesture Recognition. Xi'an: IEEE, 2018: 667–674.
- [44] 刘德志, 梁正友, 孙宇. 结合空间注意力机制与光流特征的微表情识别方法 [J]. *计算机辅助设计与图形学学报*, 2021, 33(10): 1541–1552.
- LIU Dezhi, LIANG Zhengyou, SUN Yu. Micro-expression recognition method based on spatial attention mechanism and optical flow features[J]. *Journal of computer-aided design & computer graphics*, 2021, 33(10): 1541–1552.
- [45] OH Y H, LE NGO A C, SEE J, et al. Monogenic Riesz wavelet representation for micro-expression recognition[C]//IEEE International Conference on Digital Signal Processing. Singapore: IEEE, 2015: 1237–1241.
- [46] HUANG Xiaohua, ZHAO Guoying, HONG Xiaopeng, et al. Spontaneous facial micro-expression analysis using spatiotemporal completed local quantized patterns[J]. *Neurocomputing*, 2016, 175: 564–578.
- [47] 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 Xuesen, JIA Jingping. 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.
- [48] LIONG S T, SEE J, WONG K, et al. Less is more: micro-expression recognition from video using apex frame[J]. *Signal processing: image communication*, 2018, 62: 82–92.

#### 作者简介:



陈斌, 高级工程师, 博士, 主要研究方向为模式识别、机器学习、大数据分析。主持江苏省高等学校信息化重点项目 2 项和厅局级项目 3 项。发表学术论文 20 余篇。



朱晋宁, 高级工程师, 主要研究方向为大数据分析、智能系统。