**长时微表情识别软件系统研究文档**

1221001027 裴一非

**一、项目问题提出及简介**

长时微表情指人在数分钟甚至更久、因情绪变化而反复出现并持续演变的面部细微动作，其持续时间远长于传统0.25–5秒的短暂微表情；在高压或需掩饰真实情绪的社交场景中，它以眼部颤动、嘴角轻动或面部局部抽搐等可捕捉的微弱线索，揭示焦虑、紧张、愉悦等深层情感。我们小组据此在经典算法与公开文献指导下，搭建实验环境、统一数据采集与特征提取流程，通过代码研读、参数调试和结果比对复现并验证原模型，进而优化特征提取与网络结构，在确保可重复性的同时探索提升识别精度与效率的新路径。

本研究复现并应用 SKD-TSTSAN（自知识蒸馏三流时序偏移注意力网络）。算法层面，以 CASME II 数据集为基准，融合运动放大、通道注意力与时间偏移模块，构建静态空间流（S-stream）、动态时间流（T-stream）与局部空间流（L-stream）三流架构，最终在长时微表情识别任务上取得 UF1 0.9288、UAR 0.9089 的精度。

交互设计层面，系统提供离线视频分析与实时摄像头检测双模式，并以 30 帧为周期滚动更新的动态情绪条可视化长时序情绪演变；同时指出数据多样性不足、预处理鲁棒性弱及模型复杂度高等缺陷，提出未来通过多数据集融合、轻量化网络与细粒度情绪分类进一步优化的方向。

**二、前期工作**

在前期我对《基于面部先验的一阶运动模型用于微表情生成Facial Prior Based First Order Motion Model (FOMM)》，进行了详细的阅读和了解，寻找实验模型与算法思路。它是一套“先定位-再估计-后生成”的微表情视频合成方案。

核心思路：先用 Region-Focusing 模块基于 68 点面部标志与高斯核生成 Facial Prior Map，把注意力锁定在眼周、嘴角等微表情活跃区；随后 Motion Prediction 模块将先验图与图像级联输入，用关键点+局部仿射变换估计源到目标的运动，并引入“虚拟参考帧”把外观对齐（T→R）与表情迁移（R→S）解耦，同时预测反向光流和遮挡图解决姿态差异与自遮挡；最后 Expression Generation 模块以编码器-解码器 GAN 做特征扭曲和遮挡修复，联合感知损失（VGG-19）与 MAE 训练，输出 256×256 灰度微表情视频。模型在 CASME II、SAMM、SMIC 上统一做灰度化、人脸裁剪、关键点适配，并对比单/混合数据集效果；主观由 3 位 FACS 专家评自然度，客观与 FOMM、MRAA 等基线比较，均取得更优的生成质量。关键超参：高斯核方差 σ、光流/遮挡权重、感知损失权重。

后续针对于全小组各自学习的不同论文，我们简述总结了以下的观点。

（一）实验研究现状  
 长时间表情识别与分析的研究仍处于“黑箱”阶段：学界对情绪在数分钟乃至更久时间内的动态演变、情绪间相互转化以及表情与真实情感的复杂映射关系所知甚少，亟需系统性的理论与技术突破。现实需求却日益迫切：人机交互（智能家居、客服、智能驾驶）、心理医疗、教育、安防等领域都希望借助机器对长时表情的捕捉来理解并响应人类情绪，例如通过监测驾驶者的愤怒或疲劳情绪实现主动安全干预，或利用学生情绪曲线指导个性化教学。在特征提取层面，传统手工特征（LBP-TOP、HWP-TOP、MDMO 等）已能刻画微表情的时空纹理与光流，但维度高、计算重、泛化差。深度学习方法（ELRCN、Dual-Inception 等）通过 CNN+LSTM 或多尺度卷积网络显著提升了时空表征力，已成为主流，但其对大规模标注数据的依赖使得在微表情这类小样本场景下依旧受限。

（二）研究内容  
 针对“小样本+长时序”双重挑战，本项目拟从算法与软件两条主线展开：算法层面，将融合手工特征的先验知识与深度网络的表征优势，构建对长时视频中微表情细微变化更敏感、鲁棒性更高的特征提取与识别框架；同时通过数据增强、迁移学习等手段缓解微表情数据集稀缺带来的泛化难题。软件层面，开发集“图像情绪识别、视频离线分析、实时摄像头监控”于一体的交互系统：前端提供直观、易用的可视化界面，支持长时间情绪曲线的动态展示；后端采用多线程并行、批量处理及高效压缩存储技术，实现对大规模视频流的快速处理与低占用存储，从而满足真实场景对实时性、易用性与可扩展性的综合需求。

**三、算法模块**

（一）数据采集

本小组阅读了近几年有关于微表情识别的文章，了解到当前微表情识别领域较为主流的四个公开数据集：SMIC、CASME II、SAMM 和 MMEW。

SMIC 提供了三种模态（高速摄像、可见光、红外），情绪标签分为积极、消极和惊讶三类，适用于基础的微表情检测任务，但其标签较粗、样本数量有限，

缺乏详细的动作单元标注，难以满足我们对图像和视频时序信息的建模需求；

SAMM 具有更高的图像分辨率（2040×1088）和较丰富的实验者种族，支持跨人种的分析研究。其面部细节保留充分，适合高质量图像下的情绪建模，但样本数量偏少，情绪类别虽多但分布不均；

MMEW收集了微表情与宏表情视频共1200条，并提供了情绪标签与FACS动作单元的双重标注。然而，该数据集较新，尚缺乏统一的对比方法和评价标准，并且与我们的研究有出入。综合考虑后，我们决定在现阶段使用CASME II数据集作为研究的数据集输入来源。

CASME II采集过程中采用高帧率高速相机（200fps），在受控环境下诱导自发微表情，具有较高的数据质量。该数据集包含来自26位被试者的255个微表情片段，覆盖7种情绪类别，样本分布相对均衡，且每个样本均提供了起始帧、峰值帧、结束帧、情绪类别以及面部动作单元（AU）等详尽标注信息。

对我们研究目标而言，CASMEII提供了比较全面的数据背景，既适用于静态帧的面部特征提取，也适用于短视频中情绪随时间演化的建模。同时，CASME II已被广泛应用于微表情识别研究领域中，结果评价稳定且成熟，为我们进一步开展研究有很好的帮助。



图1 CASMEII数据集中样例sub03\_EP18\_06的部分图像

（二）算法架构

我们对广泛的表情识别论文进行了研究，最终选择论文《Three-Stream Temporal-Shift Attention Network Based on Self-Knowledge Distillation for Micro-Expression Recognition》中提出的算法SKD-TSTSAN作为长时表~情识别系统中的核心识别算法。

该算法系统性地、创新性地解决了微表情识别的两大核心挑战。1、特征层面： 通过运动放大（Mag） 增强输入信号强度，通过通道注意力（ECA） 聚焦关键区域，通过零成本时序建模（TSM） 有效捕捉运动动态。2、学习层面： 通过自知识蒸馏（SKD） 结合多任务损失（FL, KL, L2），最大化利用有限数据，提升模型各层的判别能力，有效防止过拟合，实现强大的小样本学习能力。该算法融合了多种技术增强特征表达与模型泛化能力，具有极强的微表情识别能力。

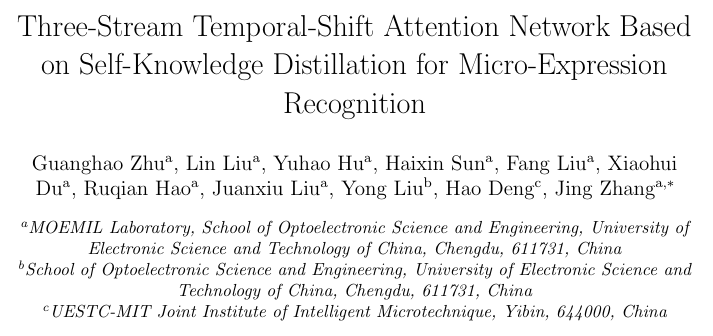


图2 提出算法SKD-TSTSAN的论文标题与作者信息

（三）算法实现

我们通过对SKD-TSTSAN算法进行复现，使用经过预处理与划分后的CASMEii数据集，我们对模型进行100轮训练后，对测试集进行测试，最终得到的UF1 (Unweighted F1 Score)为0.9288，UAR (Unweighted Average Recall)为0.9089，接近论文结果，认为有较好的表情识别能力，可用于后续系统设计。

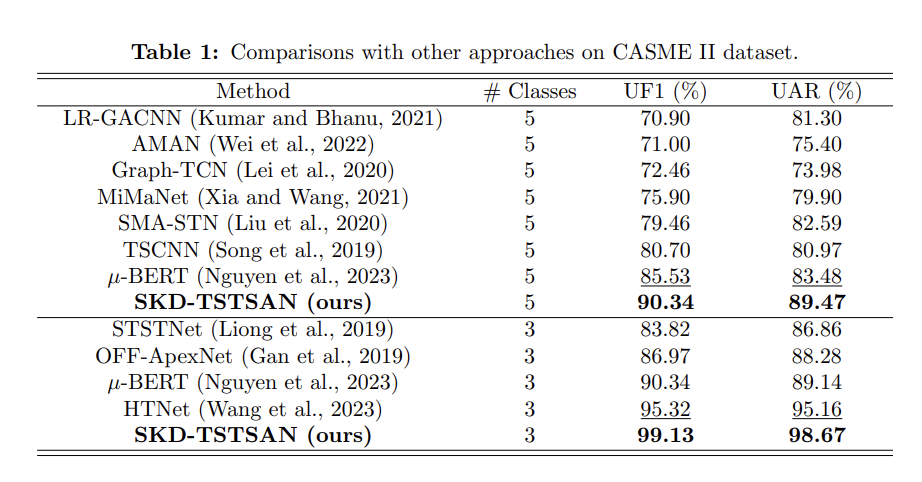


图3 算法SKD-TSTSAN论文中给出的评价指标结果

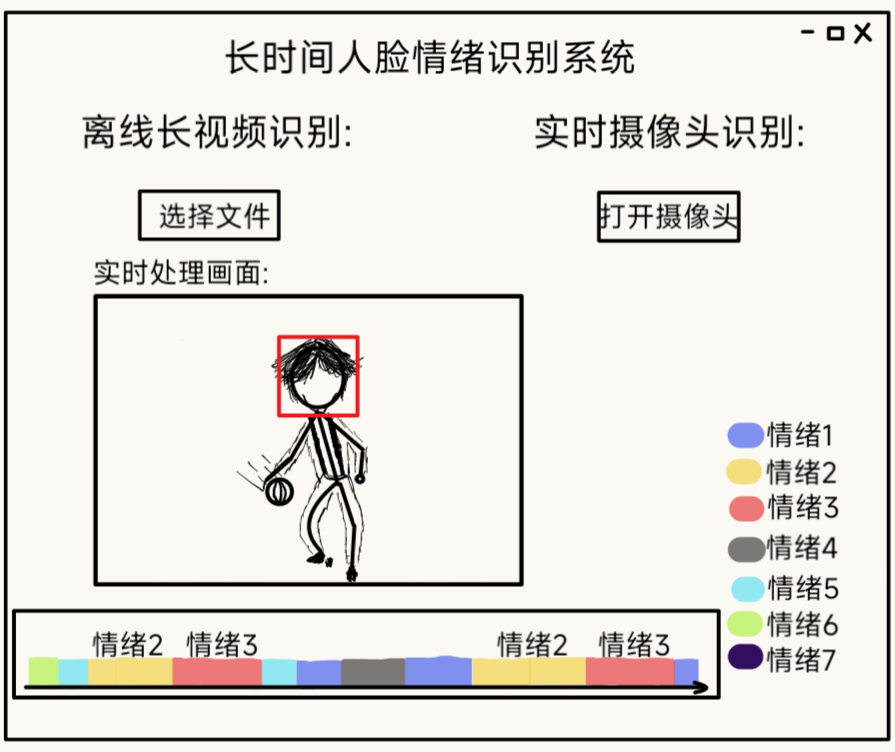
**四、系统设计**

分为数据预处理、神经网络情绪预测、长时预测结果可视化三部分。

1. 数据预处理  
 系统以25 fps视频为输入，按20帧切分微表情片段，自动抽取onset（第1帧）、apex（第11帧）、offset（第20帧）三关键帧。先用RetinaFace检测人脸并统一裁切成128×128；随后把onset与apex再缩成48×48灰度图，供静态空间流S-stream。用TV-L1光流法计算onset→apex、apex→offset两组光流，生成48×48×2双流图，供动态时间流T-stream。最后将apex与onset各自划分为4×4网格的16个局部块，每块再缩至48×48，按通道拼接成48×48×32张量，供局部空间流L-stream，完成三路输入准备。  
2. 神经网络情绪预测  
 三流共享前端：每流先接入轻量级运动放大模块（仅保留形状分支）增强肌肉微动；L-stream额外插入ECA通道注意力抑制冗余信息；随后使用双向Temporal Shift Module（TSM）在2D CNN复杂度下融合时空上下文。三路特征级联后经全连接得到256维表示，经Softmax输出7类情绪概率。训练阶段引入自知识蒸馏（SKD）：深层网络作为教师，浅层网络为学生，联合交叉熵、KL散度与L2损失，在只有255段CASME II样本的小数据上防止过拟合。100 epoch后测试集UF1=0.9288，UAR=0.9089，与原文相当。  
3. 长时预测结果与可视化  
 系统提供离线视频与实时摄像头两种模式。推理时以30帧（≈1.2 s）为滑动窗口进行一次前向推断，输出情绪类别及置信度。结果以彩色情绪条呈现：每个单位条对应30帧，七种颜色区分情绪类别；若连续同类情绪则叠加文字标签。离线模式下完整展示整段视频的情绪曲线；实时模式维护环形缓冲区，仅保留最近600条（≈12 min）结果，新数据到达时自动淘汰最早记录，实现动态滚动更新，用户可直观追踪被试者长时间情绪演变。

**五、软件前端设计**

我设计和参与了整体的软件前段的设计。在软件设计方面，我们计划开发一个用户友好的软件界面，支持多种图像和视频输入格式，包括实时摄像头捕获，以适应不同的应用场景。软件将具备高效的多线程数据处理能力，能够快速读取和解码视频数据，并对每帧图像进行实时分析和处理。为应对长时间视频产生的大量数据，软件将设计一个高效的数据存储模块，采用合适的数据存储结构和压缩技术，减少存储空间，提高数据读写效率。此外，软件将提供多种数据可视化选项，如将检测结果以表格形式展示，包括表情类型、置信度和时间戳等信息，方便用户查看和分析。用户可以根据需要选择显示检测画面和原始画面的方式，并可以动态调整检测算法的阈值，以获得最佳的检测效果。软件还将支持将检测结果导出为CSV文件，便于用户进行进一步的数据分析和记录。同时，用户可以将带有表情检测标记的图像或视频保存为指定格式的文件，方便保存和分享。（下面是设计初稿）

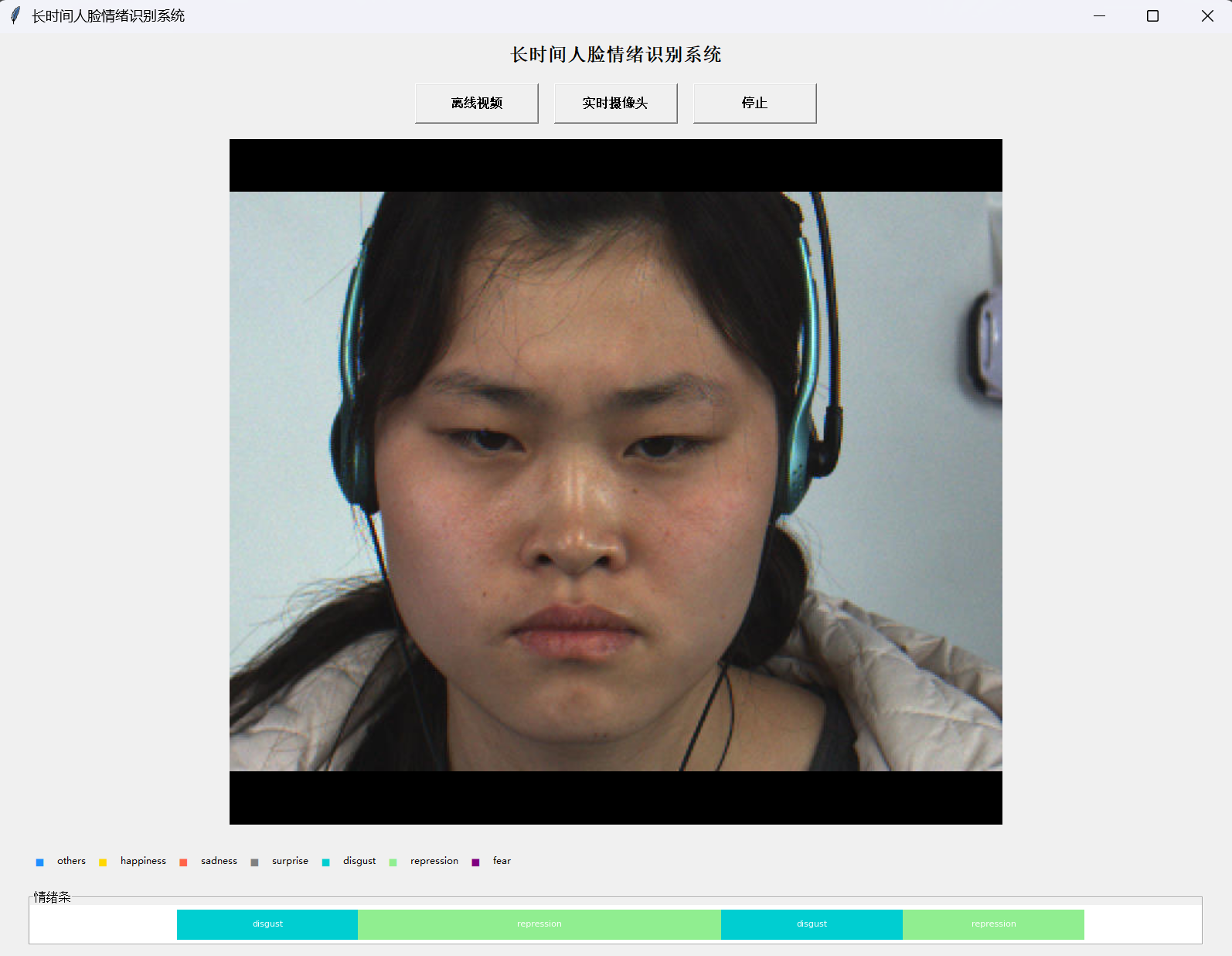


最终我们利用PYQT设计了长时间视频人脸情绪识别系统。系统支持两种情绪识别方式：

1)上传长视频进行离线情绪识别

2)直接通过实时摄像头进行情绪检测，满足不同场景下的使用需求。

系统的功能之一是生成直观的情绪条。在系统中，以“30帧”为一次情绪识别周期，情绪条中用一个长度随总数量变化地单元记录一个周期内识别出的情绪，每个单元使用七种不同颜色中的一种标注该单元内识别出的特定情绪，方便用户能够快速、直观地看出情绪的动态变化。在同一情绪较为稳定地出现时，将在颜色上面通过文字标注情绪。



图左:离线模式 图右:实时模式

**六、总结与展望**

本次研究我们基于CASMEII数据集，成功复现了SKD-TSTSAN算法的基本框架，构建了包含数据预处理、神经网络情绪预测和结果可视化的完整系统。系统在测试集上取得比较好的准确率，验证了算法在长时微表情识别中的有效性。通过运动放大模块增强肌肉特征、引入ECA注意力机制抑制冗余信息，以及采用双向TSM提取时空特征等创新设计，系统在复杂场景下仍能保持较高的识别精度。当然也存在以下不足：

1)存在不足数据局限性：复现实验数据集少，样本量较小且场景单一，可能导致模型泛化能力不足。

2)预处理粗糙：该系统默认每30帧为一个表情，并将中间帧认为是表情峰值帧，对表情的识别效果与预期不同，存在较大的预测错误概率。

3)情绪粒度问题：当前系统仅支持七种基本情绪分类，对于复杂情绪（如尴尬、轻蔑等）的识别能力有待提升。

4)模型复杂度：尽管采用了自知识蒸馏（SKD）优化，但多流网络结构仍存在参数冗余，硬件资源消耗较大，不利于移动端部署。

提出了上述问题之后，我们有以下的展望。对数据集扩充，引入其他公开微表情数据集，丰富样本多样性，提升模型泛化能力。对算法进行优化，尝试更高效的光流计算方法，优化运动放大模块与特征融合结构，减少计算耗时和资源消耗。对可视化改进，在情绪条展示中增加情绪强度指标，采用不同亮度或饱和度区分，增强可视化结果的信息丰富度 。