**长时微表情识别软件系统研究文档**

鲁达 李林洋 沈宏霓 裴一非 邵行健

（首都师范大学信息工程学院）

**摘要**

本研究基于深度学习方法，针对长时微表情识别中的算法优化与系统设计展开探索。通过复现SKD-TSTSAN（自知识蒸馏三流时序偏移注意力网络）算法，结合运动放大、通道注意力与时间偏移模块，构建了包含静态空间流（S-stream）、动态时间流（T-stream）和局部空间流（L-stream）的三流网络架构。系统采用CASME II数据集进行训练与验证，实现了UF1 0.9288和UAR 0.9089的高精度识别性能。在软件设计层面，开发了支持离线视频分析与实时摄像头检测的双模式系统，通过动态情绪条可视化技术直观展示长时序情绪演变（以30帧为周期滚动更新）。研究同时指出当前系统在数据多样性、预处理鲁棒性及模型复杂度方面的不足，并提出引入多数据集融合、轻量化网络优化及细粒度情绪分类的未来改进方向。

**关键词：**长时微表情识别；情绪动态可视化；实时情绪监测；SKD-TSTSAN算法

目录

[一、问题提出与研究目标 - 3 -](#_Toc201853540)

[（一）什么是长时微表情？ - 3 -](#_Toc201853541)

[（二）研究目标 - 3 -](#_Toc201853542)

[二、国内外研究现状以及问题 - 3 -](#_Toc201853543)

[（一）研究现状 - 3 -](#_Toc201853544)

[（二）现有研究存在的问题 - 3 -](#_Toc201853545)

[三、算法模块 - 4 -](#_Toc201853546)

[（一）数据采集 - 4 -](#_Toc201853547)

[（二）算法架构 - 4 -](#_Toc201853548)

[（三）算法实现 - 5 -](#_Toc201853549)

[四、系统模块 - 5 -](#_Toc201853550)

[（一）系统设计： - 5 -](#_Toc201853551)

[1、数据预处理模块： - 5 -](#_Toc201853552)

[（2）人脸检测与裁切： - 5 -](#_Toc201853553)

[（3）光流计算： - 6 -](#_Toc201853554)

[（4）局部空间流输入准备： - 6 -](#_Toc201853555)

[（5）网络输入整合： - 6 -](#_Toc201853556)

[2、神经网络情绪预测模块： - 7 -](#_Toc201853557)

[3、长时间预测结果可视化模块： - 7 -](#_Toc201853558)

[（二） 系统功能展示 - 8 -](#_Toc201853559)

[1．离线视频情绪识别 - 8 -](#_Toc201853560)

[2.摄像头实时表情识别 - 8 -](#_Toc201853561)

[五、研究总结 - 9 -](#_Toc201853562)

[（一）存在不足 - 9 -](#_Toc201853563)

[（二）未来展望 - 9 -](#_Toc201853564)

# 一、问题提出与研究目标

## （一）什么是长时微表情？

长时微表情是指个体在数分钟甚至更长时间跨度内，因内在情绪被压抑、掩饰，而不自觉流露在面部的细微表情变化 。与传统微表情仅持续 0.25 - 5 秒的短暂瞬间不同，长时微表情在相对较长的时间周期内反复出现、持续演变。

从产生机制来看，当人们处于复杂的社交场景、高压环境或刻意隐藏真实情绪时，即便通过语言、行为进行掩饰，内心的情感波动仍会以细微面部动作的形式泄露。例如眼部肌肉的轻微颤动、嘴角不自主的下拉或上扬、面部肌肉的局部抽搐等。这些表情虽然幅度微小、不易察觉，但通过专业设备和分析方法，能够捕捉到其在长时间内的变化规律，从而反映出个体深层次、持续的情绪状态，如焦虑、紧张、愉悦等。

## （二）研究目标

本次我们小组的研究旨在基于经典算法与模型，结合现有文献和技术方案，参照原始文献构建实验环境，规范数据采集、特征提取及模型训练流程，验证原研究成果的可靠性与可重复性。通过代码研读、参数调试及结果比对，剖析原算法模型的设计原理，与此同时，尝试改进特征提取方法、优化模型结构，在确保复现准确性的前提下，力求寻找提升识别精度识别与效率的创新思路。

# 二、国内外研究现状以及问题

## （一）研究现状

目前，我们对人类表情与情绪关系的理解还存在诸多未解之谜，尤其是在长时间维度下情绪的动态演变、不同情绪之间的相互转化以及表情与内心真实情感之间的复杂关联等方面。要完善人类情绪表达与认知的理论体系，对长时间的表情识别与分析进行深入研究具有十足的重要性。不仅如此，在许多场景中，掌握人类的情绪变化都具有至关重要的地位：在未来的人机交互中，长时间表情识别与分析可以使计算机从多角度接收人类情绪表达信号，推动智能家居、智能客服、智能驾驶等行业发展，对人类微表情的深刻理解可以帮助计算机自动感知和回应情绪，尤其在智能驾驶中，驾驶者的疲劳、愤怒、悲伤等情绪造成的许多危险驾驶行为都可通过车辆的主动介入避免；在心理医疗领域，机器对长时间表情的分析可以帮助心理医生捕捉患者的细微情绪变化，为病情评估提供更丰富的依据；在教育领域，对学生的情绪分析有助于老师明白学生实际的心理变化过程，依此为学生提供更有助于心理成长的教育方法，避免心理压力过大甚至产生心理问题；在公安领域，情绪分析能够帮助公安机关捕捉嫌疑人的微妙情绪变化，通过焦虑、紧张等情绪的积累缩小嫌疑范围……在当今这个信息爆炸、人际交往日益复杂而人工智能技术蓬勃发展的时代，基于深度学习研究长时间表情识别与分析具有深远且多维度的必要性。

在表情识别的特征提取方面，国内外已经有许多优秀的学者做了大量的工作。基于手工特征的方法，如局部二值模式（LBP）和光流特征，通过提取图像的纹理和运动信息，能够有效捕捉表情的动态特征。例如，G. Zhao 和 M. Pietikäinen 提出的 LBP-TOP（Local Binary Patterns on Three Orthogonal Planes） 通过在三个正交平面上提取LBP特征，能够有效建模微表情的时空特征，但其特征维度较高，计算复杂度较大。X. Ben 等人提出的 HWP-TOP（Hot Wheel Patterns on Three Orthogonal Planes） 和 DCP-TOP（Dual-Cross Patterns on Three Orthogonal Planes）在 LBP-TOP 的基础上引入了方向信息和双交叉模式，进一步提高了特征的判别能力。Y.-J. Liu 等人提出的 MDMO（Mean Directional Mean Optical Flow） 和 Sparse MDMO（Sparse Mean Directional Mean Optical Flow）通过计算主方向上的光流向量的均值和稀疏表示，有效地捕捉了微表情中的运动信息，并对光照变化和头部运动具有更好的鲁棒性。然而，现有的手工特征方法往往难以适应复杂的实际场景，泛化能力有限。为了克服这一局限，基于深度学习的方法逐渐成为表情识别的主流。H.-Q. Khor等人提出的 ELRCN（Enriched Long-term Recurrent Convolutional Network）结合CNN和LSTM，提取视频中的时空特征，能够有效处理视频中的时间信息，适用于动态表情识别。X. Jia等人提出的Dual-Inception Network\*\* 通过Inception模块提取多尺度特征，提高了模型的表征能力，进一步提升了识别的准确性。

## （二）现有研究存在的问题

传统方法虽能有效捕捉微表情动态特征，但在处理小样本图像空间时，难以捕捉微表情的细微变化，特征提取多样性不足。现有的基于深度学习的算法在特征提取和模型优化上虽有显著进展，但对数据依赖性强，导致模型训练和泛化能力受限，且模型可解释性较差。同时，现有研究多集中在算法改进和性能提升上，对用户交互和结果呈现的用户友好度关注不够，缺乏对多维度数据的整合和关联展示，难以满足实际应用场景中用户对直观、易用软件界面的需求。

因此，本项目所研究的问题主要分为算法优化和软件设计两方面。首先在算法方面，本项目将致力于更有效地捕捉长时间视频序列中的微表情变化，提高识别准确率。将结合多种特征提取技术，充分利用微表情的时间和空间信息，设计更鲁棒的特征提取和识别算法。同时，针对微表情数据集样本稀缺问题，研究数据增强、迁移学习等方法，提升模型泛化能力。软件设计方面，本项目希望开发一款结合多种功能——包括图像情绪识别分析、视频情绪识别分析、实时情绪监控分析等——并且拥有对用户更加友好的软件界面的软件。除去优秀多样的功能，软件还将具备高效的数据处理与存储模块，以应对长时间视频产生的大量数据。采用多线程读取、批量处理等技术快速处理视频数据，并运用合适的数据存储结构和压缩技术，减少存储空间，提高读写效率。

# 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（Micro-Expression）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 数据集最早发布于2013年，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 数据集于2016年发布，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 数据集是近年来发展较快，且较新的数据集之一，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由中科院心理研究所发布，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。

对我们现阶段以图片作为输入、结合时序的微表情特征变化所反映的情绪变化的研究目标而言，CASMEII数据集提供了较为优越的数据基础，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，结果评价手段成熟，在微表情特征提取的结果对比和重复实验的结果具有稳定性，为我们进一步开展研究有很大利好。



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。

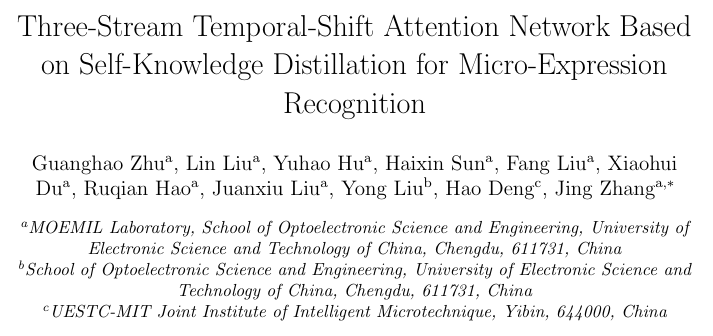


图2 提出算法SKD-TSTSAN的论文标题与作者信息

## （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。

表格

AI 生成的内容可能不正确。

图3 算法SKD-TSTSAN论文中给出的评价指标结果

# 四、系统模块

## （一）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、数据预处理模块：

#### （1）关键帧提取：

认为在帧率25的常规视频中，微表情完成约在20帧（0.8s），对每段20帧的表情序列， 选取第1帧作为起始帧 (onset), 选取第11帧作为峰值帧 (apex)，选取第20帧作为结束帧 (offset) 。

### （2）人脸检测与裁切：

使用 RetinaFace 算法及预训练模型检测每张关键帧中的人脸区域，并将检测到的人脸区域裁切后缩放至 128×128像素（onset、apex、offset），其中起始帧（onset）与峰值帧（apex）进一步缩放至48×48像素并转化为灰度图，作为静态空间流（S-stream）的输入。

图形用户界面, 网站

AI 生成的内容可能不正确。

图4 图像裁切结果

### （3）光流计算：

使用 TV-L1方法计算“起始帧 (onset) 到峰值帧 (apex) 的光流”和“峰值帧 (apex) 到结束帧 (offset) 的光流”两对光流场，将每对光流的两个分量（u和v）组合成 2通道图像，并缩放至 48×48像素，最终得到两个48×48×2的光流图像，作为动态时间流（T-stream）的输入。

日程表

AI 生成的内容可能不正确。

图5 光流图计算示例

### （4）局部空间流输入准备：

将128×128像素的 峰值帧 (apex) 和 起始帧 (onset) 分别划分为 4×4网格（共16个局部块），并将每个局部块独立缩放到 48×48像素，将峰值帧的16个局部块和起始帧的16个局部块按通道维度拼接，形成局部空间流（L-stream）的输入（48×48×32的图像）。

图形用户界面, 应用程序

AI 生成的内容可能不正确。

图6 局部空间流准备过程

### （5）网络输入整合：

整合以上数据，最终得到静态空间流输入：48×48×2的起始帧与峰值帧灰度图、动态时间流输入：32个48×48局部块拼接成的多通道图像（48×48×32），以及局部空间流输入：两对光流图拼接成的48×48×4图像。

### 2、神经网络情绪预测模块：

图示

AI 生成的内容可能不正确。

图7 表情推理网络

1. 运动放大模块：我们采用运动放大网络为三个流添加单独的运动放大模块，增强肌肉运动，同时移除了纹理表示提取部分，仅保留形状表示提取部分，降低了运动放大模块的复杂性。
2. 高效通道注意力模块（ECA）：我们在L-stream中引入了高效的通道注意力（ECA）模块，抑制了冗余信息对情绪分类的影响。
3. 时间偏移模块：我们采用双向TSM在保持2D CNN复杂度的同时提取时空信息，实现了过去和未来帧的信息与当前帧的混合。
4. 特征融合与分类模块：将三个流的特征链接在一起，形成最终表示，然后进行分类，最后转化为概率分布。
5. 自知识蒸馏模块：我们采用自蒸馏框架（SKD），在保持低复杂度的同时提高网络性能。

### 3、长时间预测结果可视化模块：

我们设计了一个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：用户可以选择1、上传长视频进行离线情绪识别，也可以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，如图8所示。



图8 软件界面示例

对于离线长视频，系统将完整展示其全部时长的情绪变化，为用户提供全面的情绪分析结果。而对于实时摄像头模式，系统将记录最近600个单位时间的情绪数据。当摄像头总记录时间超过600帧时，系统将采用动态更新机制：每新增一个单位时长的情绪数据时，会自动删除最早记录的第一个单位的情绪数据，并即时生成新的情绪图像，确保用户始终能查看到最新的情绪变化趋势。

## 系统功能展示

### 1．离线视频情绪识别

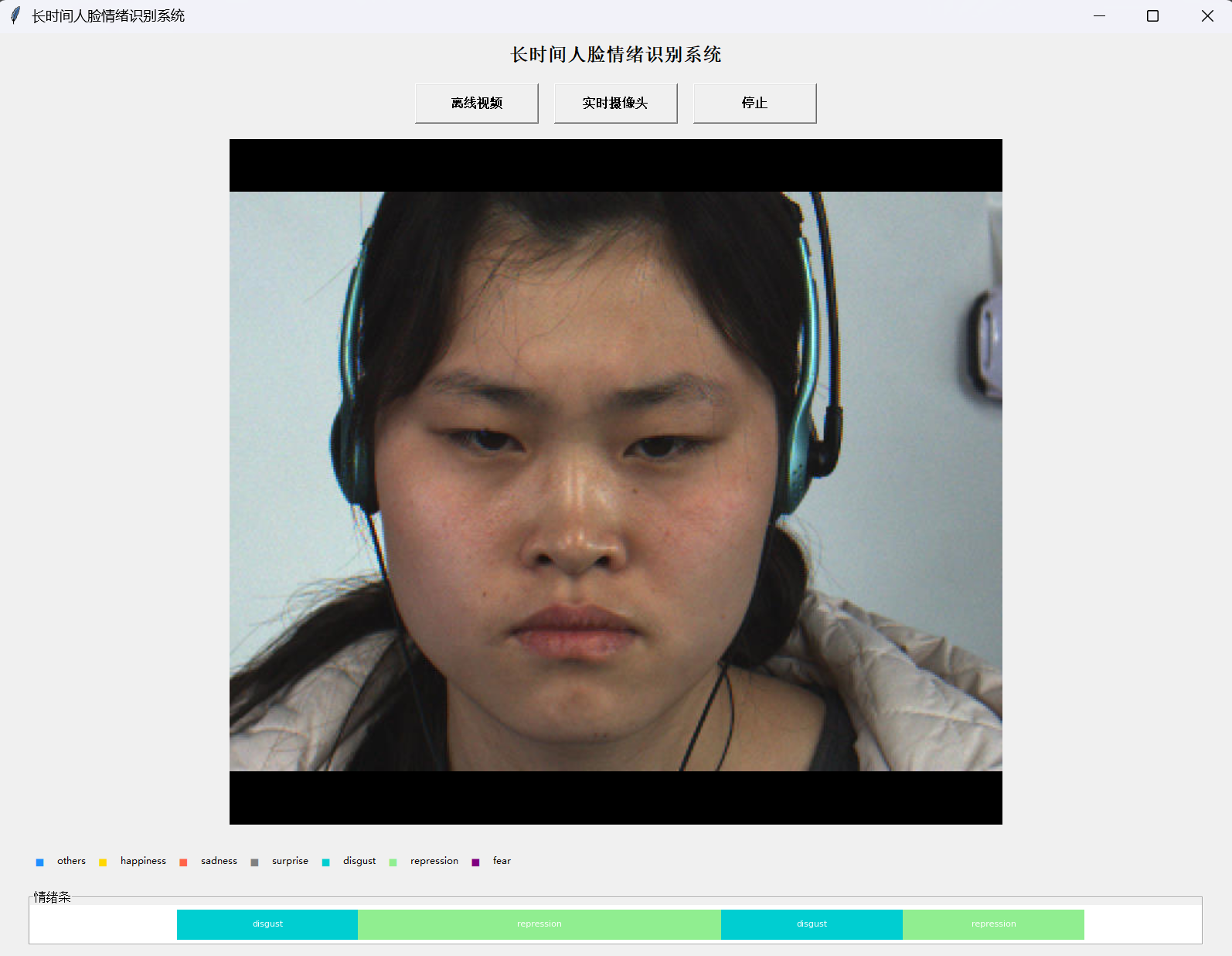


图9 离线视频表情分析结果

### 2.摄像头实时表情识别



图10 摄像头实时表情分析结果

# 五、研究总结

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了较为可观的识别准确率，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。

## （一）存在不足

数据局限性：复现实验仅基于CASMEii数据集，样本量较小且场景单一，可能导致模型泛化能力不足。真实应用场景中的光照变化、面部遮挡、多角度问题等未得到充分验证。

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。

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。且情绪条可视化方式过于简化，未能充分展示微表情强度的动态变化。

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。

## （二）未来展望

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。

算法优化：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。

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 。