

基于Google MediaPipe的高精度情感计算框架：微表情识别与时序情感预测的深入研究

报告

1. 绪论：情感计算的范式转移与前沿挑战

情感计算 (Affective Computing) 作为人机交互 (HCI) 领域最具活力的分支之一，正经历着一场从实验室受控环境向“野外” (In-the-wild) 实时应用的历史性范式转移。传统的面部表情识别 (Facial Expression Recognition, FER) 主要依赖于高精度的深度传感器、多摄像头阵列或计算成本高昂的服务器端卷积神经网络 (CNN)。然而，随着边缘计算能力的提升以及轻量级机器视觉架构的突破，特别是Google MediaPipe框架的出现，使得在消费级设备上进行高保真、实时的面部几何分析成为可能¹。

本报告旨在响应对利用Google MediaPipe数据进行更深层次人类情绪分析的需求，特别是聚焦于两个高难度的子领域：微表情 (**Micro-Expression, ME**) 的识别以及未来情绪状态的预测 (**Emotion Prediction/Forecasting**)。微表情作为一种非自愿的、瞬间的 (持续时间通常少于500毫秒) 面部动作，往往揭示了人类试图压抑或隐藏的真实情感，因此在心理健康监测、安全审查及测谎等领域具有不可替代的价值³。而情绪预测则代表了情感计算的下一个前沿——从被动地“识别当下”进化为主动地“预判未来”，利用时序数据预测个体在未来数秒甚至数分钟内的情绪走向，为构建具有同理心和预见性的交互系统奠定基础。

本报告将系统性地剖析MediaPipe Face Mesh的底层架构及其生成的52个Blendshapes (混合形状) 语义数据，探讨其与面部动作编码系统 (FACS) 的映射关系，并深入论证结合长短期记忆网络 (LSTM) 与Transformer架构实现高精度微表情捕捉与时序预测的技术路径。

2. 数据基石：MediaPipe Face Mesh 的几何拓扑与特征空间

要实现超越传统宏表情 (Macro-expression) 的精细分析，必须深入理解数据的底层结构。MediaPipe Face Mesh 提供了一个密集且高维度的面部几何描述，这构成了后续所有高级分析的基石。

2.1 478个3D关键点的拓扑结构与意义

传统的面部特征点检测器 (如Dlib的68点模型) 仅能捕捉面部轮廓、眼睑边缘和嘴唇边界等稀疏特征，这种低分辨率的拓扑结构往往会遗漏微表情中极其细微的肌肉颤动 (例如眼轮匝肌的瞬间收缩)。相比之下，MediaPipe Face Mesh 能够实时估计**478个3D面部地标**¹。

这一高密度的网格结构是通过多阶段的机器学习管道实现的：

1. **BlazeFace** 检测器：首先，一个超轻量级的面部检测器在图像中定位面部区域(ROI)。该检测器专为移动GPU推理优化，能够以极高的帧率(100-1000 FPS)运行，确保了整个系统的高实时性¹。
2. **3D 网格回归**：在定位的面部区域内，一个深度神经网络直接回归出478个顶点的3D坐标(x, y, z)。值得注意的是，这些坐标是在一个度量的3D空间中定义的，其中Z轴代表深度，这使得模型能够区分面部旋转与实际的特征形变⁶。

2.1.1 虹膜追踪与注视方向的重要性

在468个基础地标之上，MediaPipe进一步集成了虹膜(Iris)追踪功能，增加了每只眼睛5个关键点(1个中心点，4个边缘点)，使总点数达到478个¹。注视方向(Gaze Estimation)是情感分析中不可或缺的维度。例如，羞愧、社交焦虑或欺骗往往伴随着回避性的目光接触，而愤怒则通常伴随着直视。通过结合虹膜的几何位置与眼睑的开合度(由相关Blendshapes描述)，系统可以区分“惊讶时的瞪眼”与“恐惧时的眼神游离”，这是仅凭稀疏特征点无法实现的深度语义分析。

2.2 坐标系标准化与Procrustes分析

在微表情分析中，头部姿态的微小变动往往会引入巨大的噪声，掩盖真实的肌肉运动。MediaPipe输出的坐标系分为归一化坐标(Normalized Coordinates)和世界坐标(World Coordinates)。对于微表情分析，世界坐标至关重要，因为它反映了面部的内禀几何结构(Intrinsic Geometry)，以米为单位，且剔除了距离因素的影响⁶。

为了进一步剥离头部刚性运动(Rigid Motion)对表情分析的干扰，必须引入**Procrustes**分析。这是一种统计形状分析方法，通过平移、旋转和缩放变换，将每一帧的检测到的面部网格对齐到一个标准的“典型面部模型”(Canonical Face Model)上⁶。只有在完成这一对齐后，顶点位置的偏移量(Displacement)才能被解释为由面部肌肉收缩引起的非刚性形变，即真实的表情信号。

3. 语义映射：从 Blendshapes 到 FACS 动作单元 (AUs)

MediaPipe的强大之处不仅在于其几何输出，更在于其内建的语义理解能力。其Face Landmarker任务能够输出**52个Blendshape**分数，这些分数是基于Apple ARKit标准定义的浮点系数(范围0.0至1.0)，代表了特定面部肌肉运动的强度⁵。然而，心理学界和情感计算领域的通用标准是保罗·艾克曼(Paul Ekman)提出的面部动作编码系统(**FACS**)。为了进行科学严谨的情绪分析，我们需要构建从MediaPipe Blendshapes到FACS动作单元(Action Units, AUs)的精确映射。

3.1 上半脸的精细映射与肌肉机制

上半脸的表情对于识别负面情绪(如悲伤、恐惧、愤怒)至关重要，且受主观控制程度较低，因此是微表情出现的高频区域。

- **内眉提升 (AU1 - Inner Brow Raiser)**
 - **MediaPipe** 对应项: browInnerUp

- 解剖学基础: 额肌(Frontalis)内侧部分的收缩。
- 情感语义: 当AU1独立出现或与AU4结合时, 是悲伤、痛苦或内疚的典型标志。在微表情中, 瞬间的眉头内侧上扬往往泄露了极度的心理压力。
- 数据洞察: ARKit标准直接提供了browInnerUp, 这与FACS的AU1是一对的强映射关系⁹。
- **外眉提升 (AU2 - Outer Brow Raiser)**
 - **MediaPipe** 对应项: browOuterUpLeft, browOuterUpRight
 - 解剖学基础: 额肌(Frontalis)外侧部分的收缩。
 - 情感语义: 通常与AU1共同出现表示惊讶。如果在对话中出现单侧的、快速的AU2(挑眉), 则可能表示质疑或怀疑。
- **皱眉/眉毛下压 (AU4 - Brow Lowerer)**
 - **MediaPipe** 对应项: browDownLeft, browDownRight
 - 解剖学基础: 皱眉肌(Corrugator supercilii)、降眉肌(Depressor supercilii)和降眉间肌(Procerus)的复合运动。
 - 情感语义: AU4是愤怒、专注、困惑或深思的核心特征。它是微表情检测中最关键的信号之一, 因为即使是极力控制表情的人, 也往往难以完全抑制皱眉肌的微弱收缩⁹。

3.2 眼部区域的动态特征

眼部肌肉的运动速度极快, 是微表情捕捉的重点。

- **上眼睑提升 (AU5 - Upper Lid Raiser)**
 - **MediaPipe** 对应项: eyeWideLeft, eyeWideRight
 - 解剖学基础: 提上睑肌(Levator palpebrae superioris)的收缩。
 - 情感语义: 区分“惊讶”(AU1+2+5)与“恐惧”(AU1+2+4+5)的关键。在恐惧微表情中, 眼睑的瞬间张大往往比眉毛的运动更早出现。
- **眼轮匝肌收缩 (AU6 & AU7)**
 - **MediaPipe** 对应项: cheekSquint (对应AU6), eyeSquint (对应AU7)
 - 解剖学基础: AU6涉及眼轮匝肌的眼眶部(Orbital part), 导致脸颊提升和鱼尾纹, 是“杜兴式微笑”(Duchenne Smile, 即真笑)的标志; AU7涉及眼轮匝肌的眼睑部(Palpebral part), 导致眼睑收紧。
 - 技术细节: MediaPipe通过cheekSquint参数能够很好地捕捉到AU6的特征, 这是判断微笑真实性的核心依据⁹。

3.3 下半脸与嘴部动作的复杂性

嘴部是面部运动最复杂的区域, 包含了大量与言语相关的动作, 因此在微表情分析中需要复杂的过滤机制来区分微表情与说话动作。

- **嘴角拉伸 (AU12 - Lip Corner Puller)**
 - **MediaPipe** 对应项: mouthSmileLeft, mouthSmileRight
 - 解剖学基础: 颧大肌(Zygomaticus major)。
 - 情感语义: 快乐的主要标志。但在微表情中, 单侧的、不对称的AU12可能暗示轻蔑或嘲笑。
- **嘴角下压 (AU15 - Lip Corner Depressor)**

- **MediaPipe** 对应项: mouthFrownLeft, mouthFrownRight
- 解剖学基础: 降口角肌 (Depressor anguli oris)。
- 情感语义: 悲伤的典型特征。AU15通常伴随着下巴肌肉的颤动 (AU17), 是极其难以伪造的悲伤信号。
- **鼻唇沟加深 (AU9 & AU10)**
 - **MediaPipe** 对应项: noseSneer (对应AU9), mouthUpperUp (对应AU10)
 - 解剖学基础: 提上唇鼻翼肌 (Levator labii superioris alaeque nasi) 与提上唇肌。
 - 情感语义: 厌恶 (Disgust) 的核心信号。厌恶是一种原始的排斥反应, 其微表情往往表现为鼻翼两侧极短促的皱起⁹。

3.4 映射表总结

下表总结了MediaPipe Blendshapes与FACS AUs的关键对应关系, 为后续的数据处理提供了标准化的字典:

FACS 动作单元 (AU)	描述	MediaPipe/ARKit Blendshape	关联情绪/语义
AU1	内眉提升	browInnerUp	悲伤, 痛苦
AU2	外眉提升	browOuterUpLeft, browOuterUpRight	惊讶, 质疑
AU4	皱眉	browDownLeft, browDownRight	愤怒, 专注, 恐惧
AU5	上眼睑提升	eyeWideLeft, eyeWideRight	惊恐, 惊讶
AU6	脸颊提升	cheekSquintLeft, cheekSquintRight	真实快乐 (Duchenne marker)
AU7	眼睑收紧	eyeSquintLeft, eyeSquintRight	愤怒, 怀疑
AU9	皱鼻	noseSneerLeft, noseSneerRight	厌恶
AU12	嘴角拉伸	mouthSmileLeft, mouthSmileRight	快乐

AU15	嘴角下压	mouthFrownLeft, mouthFrownRight	悲伤
AU24	嘴唇紧闭	mouthPressLeft, mouthPressRight	愤怒控制, 决心
AU45	眨眼	eyeBlinkLeft, eyeBlinkRight	生理反应, 压力指标

4. 微表情识别(MER)的深度分析方法

微表情具有三个显著特征, 使其识别难度远超普通表情: 持续时间短(<0.5秒)、强度低(肌肉运动幅度极小)和局部性(往往只出现在面部的某一部分)。针对这些挑战, 基于MediaPipe的分析系统需要采用特定的算法策略。

4.1 克服时间分辨率的限制: 时间插值

标准的网络摄像头通常以30 FPS运行, 这意味着每帧间隔约33毫秒。一个持续200毫秒的微表情可能仅占据6-7帧, 这使得捕捉其“起始-顶点-结束”(Onset-Apex-Offset)的完整演变过程变得极其困难。

- 时间插值模型(TIM): 为了解决这一问题, 研究者通常采用时间插值技术(Temporal Interpolation Model), 在原始帧之间生成中间帧, 人为地提高采样率。
- MediaPipe的高帧率优势: MediaPipe的一个核心优势在于其极高的推理速度。在移动GPU上, 它可以达到100 FPS以上的处理速度¹。如果配合高帧率摄像头(如60 FPS或120 FPS的USB摄像头), MediaPipe可以直接提供足够的时间分辨率, 从而避免插值带来的伪影, 这对于捕捉稍纵即逝的微表情至关重要。

4.2 特征提取: 从光流法到几何动力学

传统的微表情识别常使用光流法(Optical Flow)或LBP-TOP(局部二值模式-正交平面)来捕捉纹理变化。然而, 光流法对光照变化极其敏感。MediaPipe的出现使得基于**几何动力学(Geometric Dynamics)**的方法成为主流。

4.2.1 顶点应变与形变分析

利用478个关键点, 我们可以构建面部的三角网格。微表情可以通过计算特定区域(如眉间、嘴角)三角网格的**应变(Strain)**来检测。

- 计算逻辑: 对于每一个三角形单元, 计算其在时间 t 相对于中性表情基准帧的面积变化率或边长变化率。
- 优势: 几何应变特征对于光照变化具有天然的鲁棒性。即使在光线闪烁的情况下, 只要面部

关键点定位准确，计算出的物理形变就是稳定的。

4.2.2 主方向最大差异(MDMD)算法的几何适配

MDMD(Main Directional Maximal Difference)是微表情识别中的经典算法，原基于光流特征。在MediaPipe框架下，可以将其适配为基于地标向量的分析：

1. 计算关键特征点(如眉毛中心点)在当前帧 t 与基准帧 $t - k$ 之间的位移向量。
2. 将位移向量分解为主要运动方向(如眉毛的垂直运动)。
3. 如果该方向上的位移幅度超过阈值 θ ，且在极短时间窗口内迅速回落，则判定为微表情候选片段¹²。

4.3 顶点定位(Apex Spotting)算法

微表情分析的核心不仅是分类，更是定位(Spotting)——即准确找出微表情发生的起始点和顶点。

基于MediaPipe Blendshapes的顶点定位算法流程如下：

1. 基准线估计(**Baseline Estimation**): 维护一个滑动窗口(例如过去1秒)，计算该窗口内Blendshape系数的平均值作为当前的“中性面孔”基准。这能适应用户面部的自然变化。
2. 差异计算: 计算当前帧特征向量 F_t 与基准向量 F_{base} 的差值范数 $D_t = ||F_t - F_{base}||$ 。
3. 峰值检测: 在时间序列 D_t 中寻找局部极大值。如果该极大值满足微表情的时空约束(持续时间 < 500ms, 且具有快速的上升斜率)，则标记为微表情顶点。

4.4 性能评估与数据集验证

在标准的微表情数据集(如CASME II和SAMM)上的测试表明，结合MediaPipe的几何特征与深度学习模型能够取得卓越的性能。

- **CASME II**: 这是一个包含自发微表情的数据集。研究表明，采用混合架构(MediaPipe提取特征 + 规则过滤 + 深度学习分类)的系统在CASME II上实现了高达**93.30%**的分类准确率³。
- **与OpenFace的对比**: 虽然OpenFace长期以来是学术界的AUs检测基准，但在实时性方面，MediaPipe展现了压倒性优势(2-5ms推理延迟 vs OpenFace的30-50ms)，这使其更适合实际部署。尽管OpenFace在某些静态图像的AUs检测精度上略高(约93% vs MediaPipe的改进版)，但MediaPipe通过密集网格对细微形变的捕捉能力在动态微表情分析中弥补了这一差距¹³。

5. 信号处理与数据增强：平滑与去噪

MediaPipe的高灵敏度是一把双刃剑。虽然它能捕捉到微小的肌肉运动，但也容易受到传感器噪

声(Jitter)的影响。在微表情分析中, 1-2毫米的抖动可能被误判为表情信号。因此, 先进的信号处理是必不可少的。

5.1 抖动消除: 一欧元滤波器 (One Euro Filter)

简单的移动平均滤波器会引入延迟, 导致系统错过微表情的快速起始阶段。**一欧元滤波器(One Euro Filter)**是HCI领域的标准解决方案, 它是一种自适应的一阶低通滤波器¹⁶。

其核心数学原理基于动态调整截止频率 f_c :

$$f_c = f_{min} + \beta \cdot |\dot{x}|$$

其中:

- f_{min} 是最小截止频率, 用于在信号静止时强力平滑噪声。
- β 是速度系数, 当信号变化速度 $|\dot{x}|$ 增加时(即微表情发生时), 截止频率线性增加, 从而减少延迟。
- 应用策略: 在MediaPipe输出的52个Blendshape系数上分别应用一欧元滤波器。对于微表情分析, 需要精细调节 β 参数, 以确保在保持“静如处子”(滤除静止噪声)的同时, 能够“动如脱兔”(零延迟捕捉瞬间表情)。

5.2 卡尔曼滤波 (Kalman Filter) 的预测性平滑

对于更高级的应用, 卡尔曼滤波器不仅能平滑数据, 还能估计系统的状态(位置与速度)。通过建立面部肌肉运动的运动学模型, 卡尔曼滤波器可以基于上一时刻的状态预测下一时刻的Blendshape值, 并利用当前的MediaPipe观测值进行修正。这种“预测-修正”机制不仅降低了噪声, 还为后续的情绪预测模块提供了可靠的速度(Velocity)和加速度(Acceleration)特征¹⁶。

6. 前沿探索: 时序情感预测 (Temporal Emotion Forecasting)

用户的核心需求不仅在于识别“现在是什么情绪”, 更在于预测“未来会是什么情绪”。这标志着从分类任务(Classification)向时序预测任务(Time-Series Forecasting)的跨越。

6.1 定义预测任务

情感预测的目标是利用过去 N 帧的观测序列 $X_{t-N:t}$, 预测未来 K 帧的情绪状态 $E_{t+1:t+K}$ 或面部特征 $X_{t+1:t+K}$ 。这在人机交互中具有极高的价值: 如果机器人能预判用户即将发怒, 它

可以在用户爆发前改变交互策略。

6.2 基于循环神经网络 (RNN/LSTM) 的预测模型

长短期记忆网络 (LSTM) 是处理此类问题的经典架构。

- 输入层: 输入通常是经过平滑处理的Blendshape序列或由MediaPipe提取的关键地标坐标向量。
- 隐藏层: LSTM单元维护一个“状态向量”(Cell State), 该向量编码了长期的情感上下文。例如, 一个持续的、低强度的皱眉 (AU4) 可能在LSTM的内部状态中积累, 预示着即将到来的愤怒爆发。
- 输出层: 模型输出未来时间步的情绪概率分布。
- 实证研究: 在手语识别 (类似的时序动作任务) 中, MediaPipe + LSTM架构已证明能达到90%以上的准确率¹⁹。对于情绪预测, 这种架构能够捕捉表情演变的“动量”。

6.3 基于Transformer的表情片段建模

随着深度学习的发展, Transformer架构正逐渐取代RNN成为时序建模的首选。

- 自注意力机制 (**Self-Attention**): 与LSTM按顺序处理不同, Transformer能够同时关注输入序列中的所有帧, 并计算帧与帧之间的关联权重。这意味着模型可以发现长距离的依赖关系——例如, 5秒前的一个微小眼神 (Frame $t - 150$) 可能与当前的嘴角动作 (Frame t) 存在强因果关系, 共同预示着某种复杂情绪的产生。
- **EST (Expression Snippet Transformer)**: 最新的研究提出了EST架构, 将视频分解为“表情片段”(Snippets), 利用Transformer分别对片段内 (Intra-snippet) 的细微运动和片段间 (Inter-snippet) 的演变进行建模。这种方法在捕捉微小的面部运动变化方面优于传统的CNN和RNN²¹。
- **Swin Transformer**: 亦有研究利用Swin Transformer作为骨干网络, 处理MediaPipe生成的图像块或特征图, 在面部属性分析和表情识别中实现了SOTA (State-of-the-art) 性能, 特别是在处理自闭症患者微妙的面部表情模式时展现了优越性²²。

6.4 生成式预测: 预见未来的面孔

最前沿的研究正在尝试生成式预测 (**Generative Forecasting**), 即直接预测未来的面部几何网格。

- 机制: 采用Encoder-Decoder架构 (通常基于Transformer或GAN), 输入过去的面部网格序列, 解码器输出未来数帧的网格坐标。
- 应用: 系统可以在内部“想象”用户1秒后的表情。如果生成的网格显示嘴角下垂、眉毛紧锁, 系统即可判定为“悲伤预警”。这种方法比单纯的标签预测更具解释性, 因为我们可以可视化系统所预测的未来表情²⁴。

7. 系统架构设计与实施建议

基于上述分析，本报告提出一套完整的高精度情感分析与预测系统架构建议：

7.1 系统流水线 (Pipeline)

- 数据采集层：
 - 输入：60 FPS+ 视频流。
 - 核心引擎：MediaPipe Face Landmarker (Stream Mode)。
- 预处理层：
 - 几何对齐：Procrustes分析，剔除头部姿态。
 - 信号平滑：对52个Blendshapes应用一欧元滤波器 (β 参数需针对微表情调优)。
- 特征工程层：
 - FACS映射：将Blendshapes转换为AU强度向量。
 - 动力学特征：计算特征的一阶导数(速度)和二阶导数(加速度)，高加速度是微表情的强特征。
- 分析与预测核心：
 - 分支A (实时识别)：使用规则引擎(检测强AU组合)+ 轻量级CNN/GNN(基于图的微表情分类)。
 - 分支B (时序预测)：采用滑动窗口(Window Size=30-60帧)输入Transformer模型(如EST变体)，输出未来10帧的情绪概率分布。
- 应用层：
 - 根据预测结果调整系统反馈(例如：教育软件检测到“挫折感”预测上升，自动降低题目难度)。

7.2 伦理与隐私考量

构建此类系统必须严格遵守伦理规范。

- 隐私权：预测用户未表达的情绪涉及深层隐私。系统设计应遵循“边缘处理”原则，即所有面部数据仅在本地设备处理，不上传云端，仅输出匿名的情绪标签。
- 偏见风险：现有的微表情数据集(如CASME II)样本量较小且人口统计学分布不均。模型可能对特定种族或性别产生偏见。在部署前，必须在多样化的人群数据上进行广泛的验证及偏差测试²⁶。

8. 结论

Google MediaPipe 以其高密度(478点)、高语义(52 Blendshapes)和高实时性的特征，为情感计算领域带来了革命性的工具。通过结合几何动力学分析、一欧元滤波去噪以及Transformer时序预测模型，我们不仅能够捕捉到人类肉眼难以察觉的微表情，更能实现对情绪状态的超前预测。

这一技术路径将情感计算从单纯的“识别”提升到了“理解与预判”的新高度，为下一代具有高度情商(EQ)的人工智能系统奠定了坚实的基础。未来的研究应致力于解决野外环境下的光照与遮挡挑战，并进一步探索多模态(结合语音、心率等)的融合预测模型。

参考文献摘要与数据来源

本报告综合分析了大量最新文献与技术文档，主要数据来源包括：

- **MediaPipe**架构与特性：¹
- 微表情识别算法与数据集：³
- **FACS**与**Blendshapes**映射：⁸
- 时序预测与**Transformer**模型：²⁰
- 信号处理技术：¹⁶
- 伦理与偏差研究：²⁶

引用的著作

1. Deep Learning-Based Real-Time Sequential Facial Expression Analysis Using Geometric Features - arXiv, 访问时间为 一月 29, 2026, <https://arxiv.org/html/2512.05669v1>
2. A Comprehensive Survey and Evaluation of MediaPipe Face Mesh for Human Emotion Recognition | Request PDF - ResearchGate, 访问时间为 一月 29, 2026, https://www.researchgate.net/publication/387923091_A_Comprehensive_Survey_and_Evaluation_of_MediaPipe_Face_Mesh_for_Human_Emotion_Recognition
3. Seeing the Unseen: Real-Time Micro-Expression Recognition with Action Units and GPT-Based Reasoning - MDPI, 访问时间为 一月 29, 2026, <https://www.mdpi.com/2076-3417/15/12/6417>
4. Microexpression - Wikipedia, 访问时间为 一月 29, 2026, <https://en.wikipedia.org/wiki/Microexpression>
5. Face landmark detection guide | Google AI Edge, 访问时间为 一月 29, 2026, https://ai.google.dev/edge/mediapipe/solutions/vision/face_landmarker
6. layout: forward target:
https://developers.google.com/mediapipe/solutions/vision/face_landmarker/ title: Face Mesh parent: MediaPipe Legacy Solutions nav_order: 2 — MediaPipe v0.7.5 documentation - Read the Docs, 访问时间为 一月 29, 2026, https://mediapipe.readthedocs.io/en/latest/solutions/face_mesh.html
7. mediapipe/docs/solutions/face_mesh.md at master - GitHub, 访问时间为 一月 29, 2026, https://github.com/google-ai-edge/mediapipe/blob/master/docs/solutions/face_mesh.md
8. MediaPipe Blendshapes recording and filtering | by Samer Attrah - Medium, 访问时间为 一月 29, 2026, <https://medium.com/@samiratra95/mediapipe-blendshapes-recording-and-filtering-29bd6243924e>
9. ARKit 52 Facial Blendshapes: The Ultimate Guide with Anatomy ..., 访问时间为 一月 29, 2026, <https://pooyadeperson.com/the-ultimate-guide-to-creating-arkits-52-facial-blendshapes/>

10. ARKit to FACS: Blendshape Cheat Sheet - Face the FACS, 访问时间为 一月 29, 2026, <https://melindaozel.com/arkit-to-facs-cheat-sheet/>
11. Facial Action Coding System (FACS) Cheat Sheet, 访问时间为 一月 29, 2026, <https://melindaozel.com/facs-cheat-sheet/>
12. Macro- and Micro-Expressions Facial Datasets: A Survey - PMC - NIH, 访问时间为 一月 29, 2026, <https://pmc.ncbi.nlm.nih.gov/articles/PMC8879817/>
13. A Non-Invasive Approach for Facial Action Unit Extraction and Its Application in Pain Detection - NIH, 访问时间为 一月 29, 2026, <https://pmc.ncbi.nlm.nih.gov/articles/PMC11851526/>
14. Comparison of Facial Landmark Detection Methods for Micro-Expressions Analysis, 访问时间为 一月 29, 2026, https://dblib.rsreu.ru/data/publications/8358_text.pdf
15. serengil/deepface: A Lightweight Face Recognition and Facial Attribute Analysis (Age, Gender, Emotion and Race) Library for Python - GitHub, 访问时间为 一月 29, 2026, <https://github.com/serengil/deepface>
16. Setting-up Smoothing Filters for MediaPipe Pose Estimation Pipeline using Python: A Practical Guide | by Debasish Raut | Medium, 访问时间为 一月 29, 2026, <https://medium.com/@debasishraut.dev/setting-up-smoothing-filters-for-mediapipe-pose-estimation-pipeline-a-practical-guide-fcc03f462196>
17. How to reduce the jittering of face landmarks? · Issue #825 · google-ai-edge/mediapipe, 访问时间为 一月 29, 2026, <https://github.com/google/mediapipe/issues/825>
18. OneEuro Filter, 访问时间为 一月 29, 2026, <https://mohamedalirashad.github.io/FreeFaceMoCap/2021-12-25-filters-for-stability/>
19. (PDF) MediaPipe-LSTM-Enhanced Framework for Real-Time Dynamic Sign Language Recognition in Inclusive Communication Systems - ResearchGate, 访问时间为 一月 29, 2026, https://www.researchgate.net/publication/393918618_MediaPipe-LSTM-Enhanced_Framework_for_Real-Time_Dynamic_Sign_Language_Recognition_in_Inclusive_Communication_Systems
20. MediaPipe with LSTM Architecture for Real-Time Hand Gesture Recognition* - NIT Rourkela, 访问时间为 一月 29, 2026, https://dspace.nitrkl.ac.in/dspace/bitstream/2080/4092/1/2023_CVIP_SBiswas_MediaPipe.pdf
21. Expression Snippet Transformer for Robust Video-based Facial Expression Recognition - Yuanyuan Liu, 访问时间为 一月 29, 2026, <https://cvlab-liuyuanyuan.github.io/pdf/journalpaper/Expression%20Snippet%20Transformer%20for%20Robust%20Video-based%20Facial%20Expression%20Recognition.pdf>
22. Emotion Recognition in Autistic Children Through Facial Expressions Using Advanced Deep Learning Architectures - MDPI, 访问时间为 一月 29, 2026, <https://www.mdpi.com/2076-3417/15/17/9555>
23. Official Pytorch Implementation of the paper, "SwinFace: A Multi-task Transformer for Face Recognition, Facial Expression Recognition, Age Estimation

and Face Attribute Estimation" - GitHub, 访问时间为 一月 29, 2026,

<https://github.com/lxq1000/SwinFace>

24. FaceLift: Semi-supervised 3D Facial Landmark Localization - arXiv, 访问时间为 一月 29, 2026, <https://arxiv.org/html/2405.19646v1>

25. Hierarchical Video Prediction | video_prediction, 访问时间为 一月 29, 2026,

https://zhongxiayan.github.io/video_prediction/

26. An ingroup disadvantage in recognizing micro-expressions - PMC - NIH, 访问时间为 一月 29, 2026, <https://pmc.ncbi.nlm.nih.gov/articles/PMC9732534/>

27. [Literature Review] Emotion estimation from video footage with LSTM - Moonlight, 访问时间为 一月 29, 2026,

<https://www.themoonlight.io/en/review/emotion-estimation-from-video-footage-with-lstm>