从骨架数据中识别情感：一项综合调查

HAIFENG LU, Shenzhen MSU-BIT University, China
JIUYI CHEN, Shenzhen MSU-BIT University, China
ZHEN ZHANG, Shenzhen MSU-BIT University, China
RUIDA LIU, Shenzhen MSU-BIT University, China
RUNHAO ZENG*, Shenzhen MSU-BIT University, China
XIPING HU*, Shenzhen MSU-BIT University, China

通过身体动作的情感识别已成为一种引人注目且隐私保护的替代方法，替代依赖面部表情或生理信号的传统方法。近年来，3D 骨架采集技术和姿态估计算法的进步极大地增强了基于全身运动的情感识别的可行性。本综述提供了骨架基础的情感识别技术的全面系统的回顾。首先，我们介绍了心理学的情感模型，并研究了身体动作与情感表达之间的关系。接下来，我们总结了公开可用的数据集，突出了数据采集方法和情感标注策略的差异。然后，我们将现有的方法分为基于姿势和基于步态的方法，从数据驱动和技术的角度分析它们。特别是，我们提出了一个统一的分类法，涵盖了四个主要的技术范式：传统方法、Feat2Net、FeatFusionNet 和 End2EndNet。每个类别中的代表性工作都进行了回顾和比较，并在常用数据集上进行了基准测试。最后，我们探讨了情感识别在心理健康评估中的扩展应用，如检测抑郁症和自闭症，并讨论了这一快速发展的领域中未解决的挑战和未来的研究方向。

CCS Concepts: • **Do Not Use This Code → Generate the Correct Terms for Your Paper**; *Generate the Correct Terms for Your Paper*; Generate the Correct Terms for Your Paper; Generate the Correct Terms for Your Paper.

Additional Key Words and Phrases: Emotion Recognition, Skeleton, Posture, Gait

ACM Reference Format:
Haifeng Lu, Jiuyi Chen, Zhen Zhang, Ruida Liu, Runhao Zeng, and Xiping Hu. 2025. 从骨架数据中识别情感：一项综合调查. *J. ACM* 37, 4, Article 111 (August 2025), 28 pages. https://doi.org/XXXXXXX.XXXXXXX

1 引言

情感在日常交流和行为决策中起着至关重要的作用，塑造个人认知、社会互动和群体动态。在人工智能（AI）领域，使机器能够识别人的情感已成为一个中心研究目标，因为这可以显著增强智能系统的互动性、适应性和决策能力。此外，情感识别在广泛的应用领域中具有重要价值，包括人机交互（HCI）、心理健康监测、教育技术、安全驾驶、异常行为检测、智能辅导系统、市场分析和安全监控 [1–3] 。因此，开发高效而准确的情感识别方法对于推动 AI 研究和提高智能系统在各个领域的功能性和实际应用性都至关重要。

*Runhao Zeng and Xiping Hu are corresponding authors.

Authors' Contact Information: Haifeng Lu, luhf18@lzu.edu.cn, Shenzhen MSU-BIT University, Shenzhen, Guangdong, China; Jiuyi Chen, ftchenjiuyi@mail.scut.edu.cn, Shenzhen MSU-BIT University, Shenzhen, Guangdong, China; Zhen Zhang, zhangzhen19@lzu.edu.cn, Shenzhen MSU-BIT University, Shenzhen, Guangdong, China; Ruida Liu, soduku645@gmail.com, Shenzhen MSU-BIT University, Shenzhen, Guangdong, China; Runhao Zeng, zengrh@smbu.edu.cn, Shenzhen MSU-BIT University, Shenzhen, Guangdong, China; Xiping Hu, huxp@bit.edu.cn, Shenzhen MSU-BIT University, Shenzhen, Guangdong, China.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.
© 2025 Copyright held by the owner/author(s). Publication rights licensed to ACM.
ACM 1557-735X/2025/8-ART111
https://doi.org/XXXXXXX.XXXXXXX

www.xueshuxiangzi.com

Table 1. 相关调查的比较

	Year	Posture	Gait
Kleinsmith et al. [9]	2013	☒	×
Noroozi et al. [10]	2018	☒	×
Deligianni et al. [13]	2019	×	☒
Xu et al. [21]	2023	×	☒
Mahfoudi et al. [22]	2023	☒	×
Our	-	☒	☒

目前，情感识别最广泛使用的方法依赖于面部表情分析 [4] 、音频信号处理 [5] 、文本分析 [6] 和生理信号监测 [7, 8] 。然而，这些方法通常依赖于基于接触的传感器或可穿戴设备，这些设备可能昂贵、使用不便，并且可能具有侵入性，从而引发隐私问题并限制其实际部署。总之，实现准确的情感识别，同时保护用户隐私并确保舒适、非侵入性的体验，仍然是该领域的一项重大挑战。大量研究表明，全身运动在传达情感状态方面起着重要作用，身体姿势是情感交流的重要非语言通道 [9–13] 。与面部表情或声音相比，基于身体的线索——特别是来自躯干和四肢的线索——使得长距离情感感知成为可能 [14] 。深度感知技术 [15] 的快速发展和人体姿态估计 [16, 17] 的进步，使得提取准确的 2D/3D 骨架数据变得越来越可行，这比传统的视觉或音频模式更能适应环境变化，并更适合隐私敏感的场景。这些技术进步激发了对基于骨架的情感识别日益增长的兴趣，这是情感计算中的一个新兴主题。

基于 3D 骨骼数据的情感识别大致可以分为两大类。第一类集中于基于姿态的情感识别，通过分析与特定动作（如敲击、挥手或跳跃）相关的运动特征来推断情感状态。第二类则以基于步态的情感识别为中心，探索自然行走过程中的动态特征，以建模潜在的情感状态。数据显示，利用姿态和步态进行情感识别的出版物数量稳步增加。这一增长突显了基于骨骼的方法在情感理解和人机交互中的日益兴趣和实际潜力 [18–20] 。

尽管该领域的研究兴趣不断增加，但对基于骨架的情感识别进行全面和统一的综述仍然缺乏。现有的评论往往采取零散的视角，常常忽略一个关键见解：姿势和步态都源于相同的骨架数据，因此共用共同的表示方法、特征提取方法和建模策略。它们的主要区别在于时间动态性而非数据模态。如表 1 所总结，大多数现有的综述集中于单一模态——例如身体表达 [9, 10, 22] 或基于步态的情感分析 [13, 21] 。

鉴于方法上的重叠和最近对基于骨架的情感识别的浓厚兴趣，建立一个整合基于姿态和步态方法的统一的综述框架是及时和必要的。在这项工作中，我们提出了一项综合性的研究，统一了姿态和步态分析到一个框架中，对现有的方法、数据集和应用提供结构化的比较，本文的结构如图 1 所示。第 2 节介绍了基本的情感模型，并探讨了身体动作与情感表达之间的关系。在第 3 节中，我们回顾了现有的公共数据集，强调数据收集方法的差异，比较了每个数据集的特征。第 4 节对方法进行了详细分析，将其分为基于姿态和基于步态的方法，并探讨了每种方法的技术特点。第 5 节探索了基于骨架的情感识别的扩展应用。最后，第 6 节总结了研究现状并概述了未来研究的有前景方向。

2 情感与身体运动的模型

本节全面概述了在 Sec. 2.1 中讨论的主要情感模型，包括离散模型、维度模型和组成模型。情感状态与基于姿势的表达之间的关系在 Sec. 2.2.1 中进行了回顾，而基于步态的情感表达则在 Sec. 2.2.2 中进行了探讨。

www.xueshuxiangzi.com

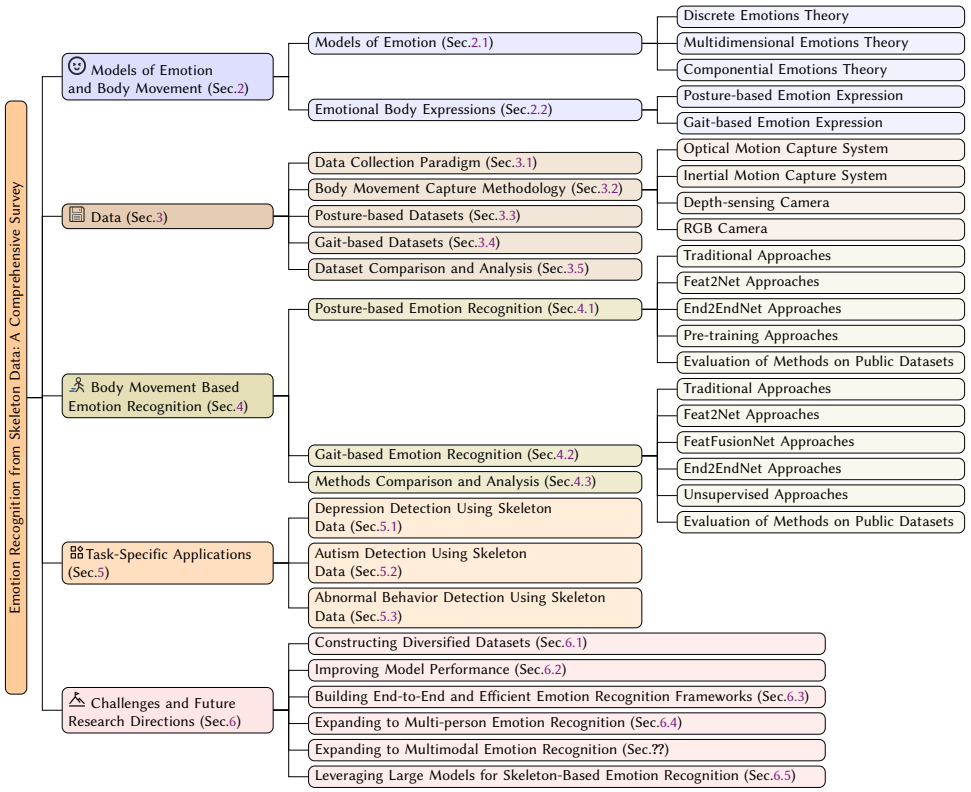


Fig. 1. 我们调查的整体结构

2.1 情感模型

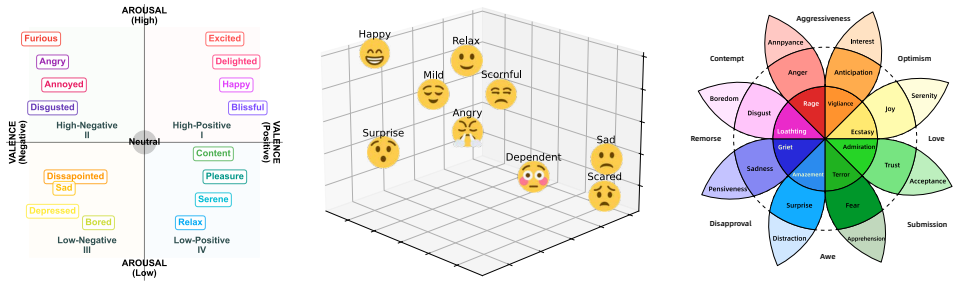
从心理学的角度来看，情绪是由刺激驱动的反应，具有明显的生理变化特征 [23]，通常被分类为反应性、激素性或自动过程 [24]。情感的建模一直是学术辩论的主题，其中出现了三种主要的方法：离散模型、维度模型和成分模型 [25]。

2.1.1 **离散情绪理论**. 将情感分类为独立且易于识别的类别一直是情感研究的基础方法。一种广泛接受的观点，受到 Paul Ekman [26] 工作的深刻影响，认为存在一组普遍的主要情感——通常包括快乐、悲伤、恐惧、愤怒、厌恶和惊讶——这些情感在生物学上是内在的，并在文化间普遍被认可。这种离散状态的观点由于其概念上的简单性及其跨文化适用性的主张，在情感计算领域获得了显著的关注。

然而，离散情感理论面临着越来越多的争议。虽然离散情感理论强调跨文化的情感一致性，但许多研究表明，情感的感知和表达受到诸如年龄、性别以及文化或语言环境 [9, 27, 28] 等因素的显著影响。这些因素对于实验设计以及在不同人群中对研究结果的准确解读至关重要，并且在文献中得到了广泛讨论。

2.1.2 **多维情感理论**. 另一种广泛采用的情感建模方法是维度模型，它将情感表示为一个连续的、多维的空间中的点 [29, 30]。在最常用的维度中，有效价（愉悦或不愉悦的程度）和唤醒（激活或强度的水平）。这一框架承认情感体验的复杂性，并通过捕捉情感状态间的细微变化，实现更为细致的分析。二维 VA 模型（效价–唤醒），最著名的是由 Russell 的环形模型表示的 [31]（见图 2 .(a)），将情感映射到效价-唤醒平面上，并被广泛应用于心理学、市场营销和教育等学科。同样，Mehrabian 提出的三维 PAD 模型（愉悦、唤醒、支配）[32]

www.xueshuxiangzi.com



(a) The 2D VA emotion model (b) The 3D PAD emotion model (c) The Plutchik’s model

Fig. 2. **各种情感模型**

(见图 2 .(b))，通过引入支配作为第三轴来扩展这一框架。这些连续模型为表示情感提供了丰富的描述空间，并已被证明与生理信号直观相关，特别是那些与效价和唤醒相关的信号 [9, 28] 。值得注意的是，分析身体运动与情绪关系的研究表明，唤醒维度在与运动相关的情感表达中占最大的方差 [33–35] 。

然而，尽管理论上非常丰富，维度模型对自动情感识别系统提出了挑战，因为将细微和连续的情感状态映射到离散的、可观察的身体表达仍然是一项不平凡的任务。

2.1.3 **成分情绪理论** . 成分模型在描述能力方面位于类别方法和维度方法之间。这些模型以层级结构组织情绪，认为复杂情绪是由更基本情绪组合而成的。一个众所周知的例子是 Plutchik 的模型（见图 2 .（c））[36] ，它将复杂情绪定义为二元组——基本情绪的对，它们的可能性随着复杂性的增加而减少。

尽管在情感计算中不太普遍，但诸如 “愉快地惊讶” 或 “愤怒地惊讶” 之类的复合情感由于其更大的表达灵活性而引起了越来越多的关注。成分模型提供了可解释性与丰富性之间的有用平衡，使其成为开发更具辨别力的情感计算系统的一个有前途的框架。

2.2 情感身体表达

2.2.1 **基于姿态的情感表达** . 人类情感与身体动作紧密相连，因为情绪状态通常通过身体行为的细微而稳定的变化表现出来 [37] 。科学研究已经表明，关节速度、身体摇摆和手势动态等参数会被情感状态调节。这些可观察到的动作变化作为重要的非言语信号，可以通过分析来推断个体的情感状态。

在关于姿势和情感表达的研究中，Dahl 等人 [38] 提出，音乐家通过不同的运动模式传达核心情感：快乐和愤怒通过大而快速的动作（流畅与抽动）表达，悲伤则以小而缓慢、平滑的动作为特征，恐惧则表现为最小的、断续的动作，由于可能的压抑，其识别度较低。在 [39] 中，作者观察到情感特定的运动模式在不同的情感状态中有所不同：愤怒涉及大而有力的动作和快速打击；悲伤通过较慢的、受限的动作且运动范围减小来表达；快乐则特点是流畅、频繁的动作和头部向上倾斜；焦虑则由紧张、受限的动作和急促的节奏所特征。Dael 等人 [40] 进一步确定了情感特定的运动模式通过不同的身体表现来体现：愤怒与前倾的姿势、伸出的双臂和攻击性的动作相关；娱乐包括间歇性的行为，如触碰物体和一个面朝交谈者的直立、侧偏头部；快乐通常涉及向上和偏离中心的头部倾斜以及不对称的手臂运动。

为更好地展示身体运动模式与离散情绪状态之间的强关联性，本文总结了表 2 中对应四种基本情绪的特征运动特征。

www.xueshuxiangzi.com

Table 2. 不同情绪状态下的典型姿态特征

Emotion	Posture Characteristics	3D Display
Happiness	Open and dynamic posture with moving arms and legs, feet oriented toward objects or people of interest. Often accompanied by spontaneous, non-goal-directed actions such as jumping, clapping, stamping, or rhythmic body movements. The body is typically held upright and may shake during intense laughter.	
Anger	Expansive posture with hands on hips or waist, clenched fists, palm-down or pointing gestures. Frequently includes trembling or shaking of hands or body, squared shoulders, erect head, and expanded chest. The stance is firm, with rigid or suspended arms. Aggressive actions such as pacing, striking objects, or making frantic, purposeless gestures may occur.	
Sadness	Contracted posture with bowed shoulders, forward-leaning trunk, and a lowered or motionless head. Hand movements are slow or minimal, often involving self-touching of the face, head, or neck. Overall body language is passive and withdrawn, indicating emotional withdrawal.	
Fear	Postures characterised by crossed limbs and visible muscular tension, such as clenched hands and inward-drawn elbows. Movements may be restrained, sudden, or bouncy. The head is typically lowered between the shoulders, or the body may adopt a crouched, motionless stance.	

Table 3. 不同情绪状态下的典型步态特征

Emotion	Gait Characteristics	3D Display
Happiness	Increased gait speed and cadence, greater step length, enhanced arm swing, noticeable vertical head movement, elevated thigh angles, and overall smoother motion patterns. The torso often appears more expansive or relaxed, frequently accompanied by a bouncy walking rhythm.	
Anger	Faster gait with higher cadence, longer stride length, pronounced thigh elevation, and vigorous arm swing. Inter-segment coordination is smoother but footfalls are heavier, conveying forceful intent.	
Sadness	Reduced walking speed and cadence, shorter strides, diminished arm swing, and limited upper-body involvement. Joint excursions—pelvic rotation, hip and shoulder flexion—are minimal. Posture tends to be slouched, with head flexion and chest contraction.	
Fear	Elevated cadence with shorter step length and reduced overall speed. Posture is tense and contracted, limbs show greater flexion, and motions can be sharp or jittery. Walk often appears cautious, with increased thigh lift and guarded foot placement.	

2.2.2 **基于步态的情感表达**. 步态指的是人类在行走过程中表现出的周期性运动模式，是一种重要的生物特征。研究表明，情绪状态可以显著影响步态的运动学参数，包括行走速度、加速度以及与特定情绪相关的其他运动特征 [13] 。

在步态和情感表达的研究中，Michalak 等人 [12] 发现，当人们感到快乐时，往往表现出更加轻快的步伐，并伴有更多的手臂摆动。相比之下，悲伤会导致步伐变得缓慢沉重，手臂运动减少，上半身姿势更加放松。怒火通常以快速有力的脚步来表现，而恐惧则与快速但步幅较短的步态有关。金等人 [41] 报告称，在快乐情绪下，足底压力峰值向前脚掌移动，而在悲伤时则集中在后脚跟的外部。Cross 等人 [42] 对 16 名个体在五种情感状态（快乐、满意、恐惧、愤怒和中立）下的运动捕捉数据进行了运动学分析。他们的结果显示，步态速度在快乐和愤怒状态下最高，而在悲伤时最低。此外，悲伤的参与者常常弯曲脖子和收缩胸部，而快乐的个体则倾向于伸展躯干或降低肩膀。Montepare 等人 [11] 研究了十名女性本科生在四种情感条件下的步态模式。他们观察到，在悲伤时手臂摆动显著减少，而快乐情绪导致较快的步态节奏和更具动感的手臂运动。愤怒和骄傲等情绪则与步幅增加相关。此外，观察者能够根据步态特征准确推断参与者的情感状态。

为了更直观地说明步态模式与情绪状态之间的强相关性，本文总结了与四种基本情绪相关的典型步态特征，如表 3 所示。

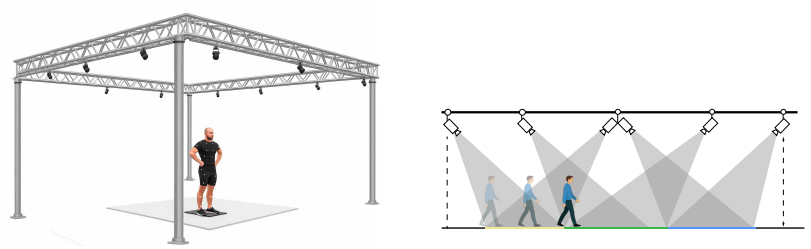
3 数据

本节在 Sec. 3.1 和 Sec. 3.2 中全面概述了获取身体运动数据的常用方法。在 Sec. 3.3 和 Sec. 3.4 中，分别回顾了基于一般身体运动和步态模式的情感识别数据集。最后，Sec. 3.5 对这些数据集进行了详细比较，重点介绍了它们的相似性和差异性，以指导和支持未来的研究工作。

3.1 数据收集范例

由于基于骨架的情感识别可以分为两种类型，因此相应地有两种数据收集范式，如图 3 所示。图 3 .(a) 描绘了姿态数据收集范式，其中演员通常在一个预先定义的区域内进行各种情感表达——通常小于一米的大小。相比之下，图 3 .(b) 展示了步态数据收集范式，参与者沿指定路径来回走动，覆盖一个较大的活动区域。

www.xueshuxiangzi.com



(a) Scene Settings for posture data collection

(b) Scene Settings for gait data collection

Fig. 3. **不同数据收集场景的比较**

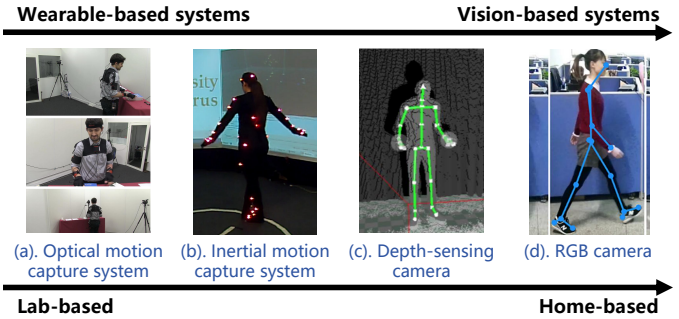


Fig. 4. **捕捉三维人体骨骼数据的常用方法**

3.2 身体运动捕捉方法学

捕捉三维人体骨骼数据的方法随着时间的推移发生了显著变化，从专业的实验室基础动作捕捉系统转变为更易获得的家用解决方案，如图 4 所示。最初，基于可穿戴设备的系统如光学动作捕捉系统（图 4 .(a)）被广泛用于受控的实验室环境。随后，惯性动作捕捉系统（图 4 .(b)）出现，利用惯性测量单元（IMUs）等可穿戴传感器记录身体运动。

随着传感器技术和计算机视觉的进步，基于视觉的系统因其便利性和无侵入性而越来越受到欢迎。例如，深度传感摄像头（图 4 .(c)）通过捕捉身体的三维结构信息，实现实时骨骼跟踪，而 RGB 摄像头（图 4 .(d)）通过深度学习算法支持基于图像的骨骼估计。

在接下来的章节中，我们将详细概述每种方法的特征和能力。

3.2.1 **光学动作捕捉系统** . 2006 年，Ma 等人 [43] 使用光学动作捕捉系统记录了三种类型的日常活动，并为其标注了相应的情绪标签。DMCD 数据集 [44] 则使用类似系统来捕捉舞者的动作。该系统依赖反光标记——通常为球形，并涂有回光材料——附着在受试者身体的关键解剖标志上。多个红外摄像机放置在捕捉空间周围，发出光线并检测从标记反射的信号。通过从不同的摄像机角度对标记位置进行三角测量，系统可以准确地重建关节的 3D 轨迹。该方法提供了高空间精度，并广泛应用于控制的实验室环境中进行详细的动作分析。

3.2.2 **惯性动作捕捉系统** . Emilya 数据集 [45] 、KDAE [46] 和 MEBED 数据集 [47] 利用惯性动作捕捉系统进行数据采集。这些系统采用可穿戴 IMU 传感器，通常附着在主要的身体部位，如四肢、躯干和头部。每个 IMU 包含加速度计、陀螺仪，有时还包含磁力计，从而能够测量线性加速度、角速度和方位。通过集成这些信号，系统可以在无需外部摄像机的情

www.xueshuxiangzi.com

Table 4. 基于骨架的情感识别中的姿势和步态数据集概述

Type	Dataset	Acquisition Device	Subjects	Samples	Joints	Frame Rate	Emotions
Posture	Emilya [45]	Xsens MVN	11	8206	28	125 Hz	8
	BML [43]	Optical MoCap	30	4080	30	16 Hz	4
	MEBED [47]	Xsens MVN	8	1447	23	120 Hz	11
	KDAE [46]	Noitom PN	22	1402	72	120 Hz	7
	EGBM [56]	Kinect V2	16	560	25	30 Hz	7
	UCLIC [57]	Optical MoCap	13	183	32	–	4
	DMCD [44]	Impulse X2	6	108	38	120 Hz	12
Gait	E-Gait I [58, 59]	Optical MoCap	–	1835	21	60 Hz	4
	E-Gait II [58, 59]	–	–	342	16	–	4
	BME I [60]	Vicon V8	8	200	12	120 Hz	5
	BME II [60]	Vicon V8	5	75	12	120 Hz	5
	EMOGAIT [61]	RGB Camera	60	1440	16	–	4

况下重建人体骨骼的三维运动。基于 IMU 的系统便携，适合在不受限制或真实世界环境中捕捉运动。

此外，崔等人 [48] 利用智能手机内置的三轴加速度计来监测参与者的日常活动并分类他们的情感状态。在这项研究中，使用了两部智能手机和一台平板电脑：智能手机佩戴在参与者的手腕和脚踝上，以每秒 5 次的采样率捕捉加速度计和重力传感器的原始数据。

3.2.3 **深度感知相机**. 2011 年 Kinect 深度相机的发布 [15] 大大简化了骨架序列获取的过程。从那时起，许多研究人员采用了 Kinect 系列进行数据收集 [49, 50] 。Microsoft Kinect 是一种深度感应相机，它集成了 RGB 相机、红外（IR）发射器和 IR 深度传感器，以捕捉颜色和深度信息。通过投射结构化红外光图案并分析其变形，Kinect 构建了场景的深度图。利用这些深度数据，内置的身体跟踪算法可以识别人形并实时估计关节的 3D 位置。

3.2.4 ***RGB* 相机**. 在 2017 年，人类姿态估计取得了突破，特别是实时多人姿态估计方法的发展，如 OpenPose [51] 、HRNet [52] 和 Vnect [53] ，进一步推动了基于身体动作的情感识别。这些技术使得能够直接从标准的 RGB 视频帧中提取二维或三维的骨骼关节坐标，消除了对可穿戴传感器或深度相机的需求。这些底层算法通常使用卷积神经网络（CNNs）来检测身体部位并估计关节位置，然后使用部位亲和场在多人场景中将关节与特定个体关联起来。到 2019 年，研究人员已经开始利用从 RGB 视频中提取的骨骼序列进行情感识别任务 [18, 54, 55] 。

3.3 基于姿势的数据集

表 4 总结了常用于基于骨架的情感识别研究的数据集。它们的独特特征将在下一节中详细讨论。

Emilya 数据集。EMotional body expression In daILY Actions 数据集（Emilya）[45, 62] 共包含 8,206 个样本。十一位演员（六位女性和五位男性）在七个日常动作的背景下表演了八种不同的情绪。这些动作经过精心选择，涉及全身运动，特别是强调上半身和手臂的动作，以确保情绪表达的多样性。这七种日常活动包括坐下、走路、走路时携带物品、用双手在桌子上移动书籍、敲门、抬起物品和投掷物品。

www.xueshuxiangzi.com

在行走任务中，演员被指示在房间的长边来回走动，分为两种不同的变体：正常行走和携带物品行走。每个样本根据八种情感类别之一进行标记：中性、喜悦、愤怒、恐慌、恐惧、焦虑、悲伤和羞愧。

所有数据均由 Xsens MVN 动作捕捉系统采集，该系统以 120 Hz 的帧率记录 3D 骨骼序列。每个骨骼序列由 28 个关节组成。

BML 数据集。BML 数据集 [43] 包含 4,080 个身体运动序列。三十位演员（15 名女性和 15 名男性）进行了四种不同的动作——行走、敲击、举起和投掷——每种动作在四种不同的情绪状态下以行走片段交替进行：愤怒、快乐、中性和悲伤。所有运动数据均采用 Falcon Analog 光学动作捕捉系统捕捉，该系统能够提供 16 个关节的 3D 位置数据。

MEBED 数据集。MPI 情感肢体表达数据库 (MEBED) 包含 1,477 个样本。数据收集由八名演员（四名男性和四名女性）参与，他们年龄均在 25 岁左右。参与者被要求在四种场景类型中表现出情感表达，每种场景类型包含十种不同的情感类别。四种场景类型分别是：孤立的非语言情感场景、交流的非语言情感场景、不含直接语言的短句和含直接语言的短句。十种情感包括：娱乐、快乐、自豪、宽慰、惊讶、愤怒、厌恶、恐惧、悲伤和羞耻，另外还有一个中性情感类别。

所有运动数据均使用 Xsens MVN 动作捕捉系统以 120 Hz 的采样率记录，捕捉由 28 个关节组成的 3D 骨骼序列。在录制期间，所有参与者都坐在凳子上。因此，分析时通常会排除十个下半身关节。

MEBED 包含双标签注释：来自表演者的自我报告情绪和来自外部观察者查看每个动作序列的第三方注释。值得注意的是，该数据集存在类别不平衡，各情绪类别的样本数量差异显著。

运动学情感表达演员数据集（KDAE）[46] 总共包含 1402 个样本。22 名半专业演员（11 名女性和 11 名男性）被要求在 70 个日常事件场景中表演七种不同的情感——快乐、悲伤、中性、愤怒、厌恶、恐惧和惊讶（每种情感对应 10 个场景）。演员在一个 1 米 × 1 米的方形舞台上进行情感表演，距离墙壁 0.5 米。

所有动作数据均使用可携带的无线动作捕捉系统记录，该系统能够以 125 Hz 的帧率追踪 72 个身体标记。值得注意的是，这 72 个标记中近一半对应于手部。然而，由于本研究专注于全身的情绪表达，因此手部标记被排除在分析之外。仅保留了 24 个关键身体标记。

情感手势和身体动作语料库（EGBM）[56] 包含 560 个样本。该数据集包含 16 位专业波兰演员的表演（8 位女性和 8 位男性），他们在没有任何预定义指令或提示的情况下表达了七种不同的情感——快乐、悲伤、中性、愤怒、厌恶、恐惧和惊讶。

所有运动数据均使用 Kinect V2 摄像机以 30 Hz 的帧率捕获。每个情感类别包含 80 个样本，每个姿势序列提供 25 个关节的 3D 位置。

UCLIC 数据集。UCLIC 数据集 [57] 包含 108 个样本。它的特点是由 13 名演员进行表演，其中包括 11 名日本演员、1 名斯里兰卡演员和 1 名美国演员。演员们自由地基于他们自己的理解表达了四种基本情感——愤怒、恐惧、快乐和悲伤，没有任何强加的限制。

所有运动数据均使用 MoCap 系统收集，每个姿态序列提供 32 个关节的 3D 位置。

DMCD 数据集。舞蹈动作捕捉数据库 (DMCD) [44] 包含 108 个样本。该数据集展示了六位背景各异的女性舞者的表演，包括体操、芭蕾、戏剧舞蹈及其他风格。每位参与者都表演了一个舞蹈序列，每个序列都与特定的情感状态相关联。

总共有 12 种情感被表达：兴奋、高兴、满意、满足、放松、疲倦、无聊、悲伤、痛苦、烦恼、生气和害怕。所有的动作序列都是使用 PhaseSpace Impulse X2 动作捕捉系统以 120 Hz 的帧率捕获的。每个姿势序列包含 38 个关节的 3D 位置。

3.4 基于步态的数据集

E-Gait 数据集。情感步态（E-Gait）数据集 [58] 包含 2,177 个真实世界的步态序列，每个序列都标注为四种情感类别之一：快乐、悲伤、中性或生气。该数据集被分为两个子集：

www.xueshuxiangzi.com

- 子集 A：包含 342 个步态序列，来源于包括 BML [43] 、ICT [63] 、CMU-MOCAP [64] 和 Human3.6M [65] 的数据集。
- 子集 B：包含 1,835 个步态序列，这些序列从爱丁堡运动捕捉数据库（ELMD）中提取 [59] 。

所有运动数据均使用动作捕捉（MoCap）系统以 60Hz 的帧率捕获，记录 21 个关节的 3D 位置。为了确保两个子集的一致性，作者将骨骼数据标准化为包括 16 个关节。E-Gait 数据集中每个步态序列都由 10 个人独立标注。每个序列的最终情感标签是通过标注者间的多数投票确定的。

BME 数据集。身体运动-情感数据集 (BME)，数据集 [60] 数据集涉及两个实验。

- 情感步态：8 位专业演员（4 男 4 女）在行走时表演了五种情绪——中性、快乐、愤怒、悲伤和恐惧。每种情绪至少录制了 5 次试验，共 200 个样本。演员们被指导去体现情绪（例如：对于恐惧，要表现成“在一个黑暗危险的房间里行走”），以标准化表达。
- 控制步态：5 名初次参与的男性受试者以三种速度行走——慢速、正常速度和快速——每种速度重复 5 次，总共获得 75 个样本。该子集隔离了与速度相关的运动变化以便进行比较。

所有运动数据都是使用 Vicon V8 光电运动捕捉系统捕捉的，该系统拥有 24 台摄像机（实验 2 中为 16 台），帧率为 120 Hz。该系统追踪 12 个身体部分上的标记的 3D 位置，包括肩膀、肘部、臀部和膝盖等关节。

EMOGAIT 数据集。EMOGAIT 数据集 [61] 包含从 60 名受试者（33 名女性和 27 名男性）收集的 1440 个真实世界的步态序列。每个序列都标注了四种情绪标签之一：快乐、悲伤、中性或愤怒。所有数据都是使用标准的 RGB 摄像机捕获的，并通过一个姿态估计算法从视频中提取了 2D 人体骨架。每个骨架序列包含 16 个关节。

3.5 数据集比较与分析

表 5 提供了几个广泛使用的数据集中情感诱发方法和表现指导的比较概述。大多数数据集——如 Emliya、BML、MEBED、KDAE 和 EGBM——利用预先选择的充满情感的情景来引导参与者表现特定的情感表达。虽然像 Emilya 和 BML 这样的数据集提供明确的动作指导以确保一致性，但其他如 KDAE 和 EGBM 则采用更自由的表现风格，让参与者更自然地表达情感。表演者的专业程度在各个数据集中也有所不同，从大学生和业余演员到训练有素的专业人员，这可能对记录的数据的表现力和可靠性产生影响。

EMOGAIT 数据集因其使用情感唤起电影片段作为引导方法而显著，旨在捕捉情感诱导后的步态变化。相反，UCLIC 数据集的独特之处在于其既没有定义的情感诱导协议，也没有特定的表现指令，而是依靠参与者自发的情感姿势。这些差异反映了每个数据集不同的设计理念和预期应用，最终影响了所捕获身体运动数据的自然性、一致性和情感丰富性。

4 基于身体运动的情感识别

本节全面概述了基于骨架的情感识别研究进展。从数据特征和技术发展的角度来看，现有方法可以大致分为两大类：基于姿态的方法（在第 4.1 节）和基于步态的方法（在第 4.2 节）。在每一组中，根据其技术设计，方法进一步分为四个有代表性的类别：

- 传统方法（参见图 5 .(a)），涉及从骨架数据中提取手工制作的情感特征（例如，关节角、身体对称性、速度），然后通过机器学习分类器进行情感识别；
- Feat2Net 方法（见图 5 .(b)），该方法从骨架序列中提取手工设计的特征，并仅利用神经网络进行分类；
- FeatFusionNet 方法（见图 5 .(c)），该方法提取手工特征并将其集成到深度学习模型的训练过程中，以增强区分性能；
- End2EndNet 方法（见图 5 .(d)），直接从原始骨架数据中学习身体运动表示并进行情感识别，而不依赖于手动设计的特征。

www.xueshuxiangzi.com

Table 5. 不同数据集的情感诱导方法和性能说明的比较

Dataset	Performer	Emotion Induction Method	Performance Instructions
Emilya [45]	College students	Pre-selected emotional scenarios	Perform specific movements while expressing the given emotions
BML [43]	College students	Pre-selected emotional scenarios	Perform specific movements while expressing the given emotions
MEBED [47]	Amateur actors	Pre-selected emotional scenarios	Sit on a chair and read aloud with matching gestures
KDAE [46]	College students	Pre-selected emotional scenarios	Performers act freely based on the given scenarios
EGBM [56]	Professional actors	Pre-selected emotional scenarios	Performers act freely based on the given scenarios
UCLIC [57]	Unknown	None	Perform emotional postures freely with no specific constraints
EMOGAIT [61]	College students	Watching various emotional film clips	Walk back and forth after watching the clips

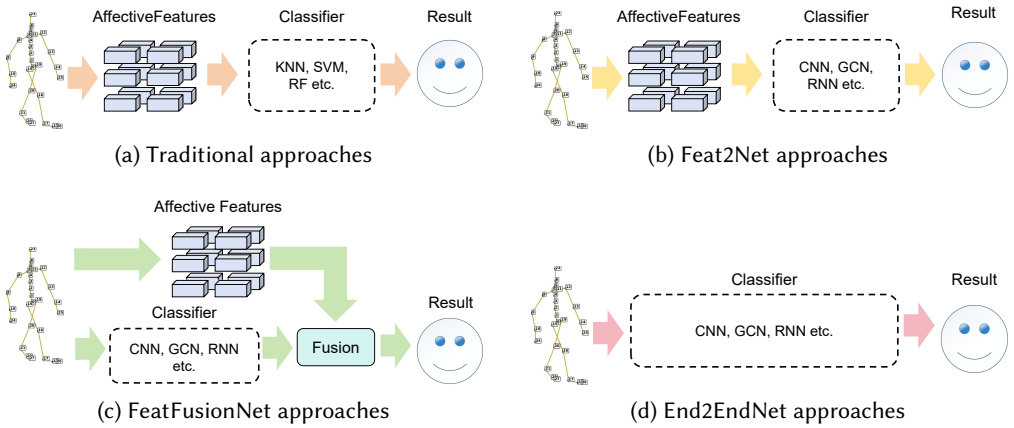


Fig. 5. 不同技术方法的比较

以下各小节的结构基于此分类法，详细分析了演变过程和主要代表性工作。

4.1 基于姿势的情感识别

4.1.1 **传统方法**. 基于机器学习的方法长期以来一直作为通过身体动作进行情感识别领域的基础组成部分。表 6 总结了近年来使用人体运动数据进行自动情感识别的一些代表性研究。

www.xueshuxiangzi.com

Table 6. 使用机器学习方法从姿势中识别情感

Study	Dataset	Feature	Classifier	Protocol	Accuracy
Kapur et al. [66]	5 subjects, 500 samples	Mean/SD of position, velocity, acceleration	LR / NB / DT / MLP / SVM	10-fold, LOSO	91.8 % ; 84.6 % ± 12.1 %
Bernhardt et al. [67]	BML	Max hand distance/speed/acceleration/jerk	SVM	LOSO	81.1 % (Sensitivity)
Fourati et al. [45]	Emilya	Power, Fluidity, Speed, Quantity/Regularity, Body openness, Leaning, Straightness	MLR	3-fold	36–48 %
Fourati et al. [68]	Emilya	110 expressive body cues (multi-level notation)	RF	OOB	67.9–84.8 %
Fourati et al. [69]	Emilya	114 expressive body cues (multi-level notation)	RF	OOB	78.29–91.53 %
Fourati et al. [70]	Emilya	114 cues; 80 position; 80 kinetic-energy feats	RF / SVM	3-fold	73.93–74.47 % (F1)
Crenn et al. [71]	BML / UCLIC / SIGGRAPH	68 low-level geometric, motion, Fourier feats	SVM	10-fold	BML 57 % / UCLIC 78 % / SIGGRAPH 93 %
Crenn et al. [72]	BML / MEBED / UCLIC / SIGGRAPH	Spectral amplitude differences (neutral vs expressive)	SVM / RF / KNN	10-fold	57 % / 67 % / 83 % / 98 %
Crenn et al. [73]	Emilya / MEBED / UCLIC / SIGGRAPH	Posture, Temporal, Residue features	SVM / RF / KNN	10-fold	82.2 % / 78.6 % / 74 % / 98.8 %
Saha et al. [74]	10 subjects	Hand–spine dist., max acc., various joint angles	DT / KNN / SVM / NN	N/A	76.63–90.83 %
Piana et al. [75]	Qualisys (12 ppl, 310 seg.) / Kinect (579 seg.)	Holistic (kinetic, contraction, symmetry ⋯) and local features	SVM	Hold-out, LOSO	Qualisys 62.3 ± 25.3 % , Kinect 68.5 ± 18.5 % ; Qualisys 54.2 ± 29.1 % , Kinect 61.6 ± 23.1 %

LR: Logistic Regression; NB: Naive Bayes; DT: Decision Tree; MLP: Multilayer Perceptron; SVM: Support Vector Machine; MLR: Multinomial Logistic Regression; RF: Random Forest; KNN: K-Nearest Neighbor; NN: Neural Network; LOSO: Leave-One-Subject-Out; OOB: Out-of-Bag.

为了更清楚的理解和比较基于姿势的情感识别方法，本节是根据用于捕捉姿势信息的数据获取技术来组织的。具体来说，我们根据姿势数据是通过动作捕捉系统、深度传感器还是RGB 摄像机收集的，对现有的工作进行了分类。

动作捕捉系统。若干研究专注于从骨骼或运动数据中提取低级运动学特征，以识别情感状态。Kapur 等人 [66] 通过提取位置、速度和加速度的均值和标准差作为输入特征，奠定了基础，用于训练机器学习模型来分类人类情感。Bernhardt 等人 [67] 通过将复杂运动分割成基本单元，分析关键动态特征，并应用归一化技术来减少个体运动偏差，推进了这方面的研究。基于这些努力，Samadani 等人 [76] 提出了一种混合生成-判别方法用于情感运动识别，该方法考虑了运动学变化、人与人之间的差异以及随机噪声。他们的方法采用隐马尔可夫模型（HMMs）来捕捉情感运动的时间动态，并生成 Fisher Score 表示以编码运动学和动态特征。

来自 Fourati 等人的工作对通过身体动作进行情感识别做出了显著贡献，他们开发了一个多层次框架，用于系统地描述和分析情感身体行为。在他们 2014 年的研究中，作者介绍了多层次身体标注系统（MLBNS）以及相应的情感身体行为数据集。该系统采用基于动作质量的编码方案，包含三个描述层次，使身体运动特征得以结构化表述。除了编码框架外，他们还记录并验证了一个新数据集以支持进一步分析。为了简化感知评分任务，后续研究从 MLBNS 中选择了一组关键变量。在此基础上，Fourati 等人扩展了该框架，从三个层次结构

www.xueshuxiangzi.com

中提取了 110 个与情感相关的特征：解剖描述、方向描述和姿势/运动动态。这些特征包括 38 个身体部位描述符、30 个肢体间关系度量以及 42 个局部肢体运动描述符。虽然 [68] 和 [69] 都追求类似的目标，但后者更详细地分析了各个特征在区分情感类别中的相对重要性。在进一步扩展 MLBNS 框架时，作者后来引入了四个附加特征组——时间模式、基于斜率的描述符、峰值动态、全局运动统计和时间规律性度量——这显著提高了分类准确性。

另一个具有影响力的研究方向由 Crenn 等人领导，他们通过对骨骼运动特征的详细分析系统地研究了情感识别。在他们早期的工作中，Crenn 等人 [71] 提出了一个综合的特征提取框架，该框架整合了几何关系、运动关联和基于频率的周期性。提取的特征包括关节间距离、特定关节形成的三角形面积、关节角度以及各种关节的速度和加速度。实验结果表明这些特征与情感表达紧密相关。在此基础上，Crenn 等人 [72] 引入了中性姿势的概念，建议通过分析观察到的骨骼数据与这种中性参考之间的残余差异来推断情感状态。这一概念在 2020 年的研究中得到进一步完善 [73]，该研究采用了成本函数优化策略来更有效地合成中性运动，从而提高了情感识别性能。在后续研究中，由 Crenn 提出的运动-情感特征框架被广泛采用并进一步扩展，突显其在该领域的持久影响。

基于完整身体动作数据的情感识别在游戏场景中已经成为一个重要且影响力日益增长的研究方向。Kleinsmith 等人 [77] 研究了视频游戏环境中非基本情绪状态的识别。他们收集了参与任天堂 Wii 体育游戏的玩家的姿势数据，通过人类观察者评估建立了真实标签，并根据捕获的动作开发了自动识别模型。Savva 等人 [78] 扩展了这一想法，探索了玩家在游戏环境中的全面身体动作是否能反映他们的审美和情感体验。使用类似的任天堂体育游戏，他们的研究表明，自动系统的情感识别准确率与人类评估者相当。进一步推进这一研究方向，Garber 等人 [79] 在游戏过程中收集了动作捕捉数据，并提取了身体对称性、头部位移和姿势开放度等特征，以推断玩家的情感状态。

深度传感器。Saha 等人 [74] 使用 Kinect 传感器进行了基于手势的情感识别研究。他们从上半身和手部选择了十一个位置信息以提取与关节距离、加速度和角度相关的九个特征，用于识别五种基本情感：愤怒、恐惧、快乐、悲伤和放松。结果表明，集成决策树实现了最高平均分类准确度 90.83%。Piana 等人 [75] 处理动作捕捉数据以提取特征，如关节能量、运动方向、姿势和身体对称性。他们通过字典学习构建自适应表示，并采用线性支持向量机进行情感分类。Ahmed 等人 [80] 计算了一套全面的身体运动特征，分为十组。在第一阶段，他们应用方差分析（ANOVA）和多因素方差分析（MANOVA）去除无关特征并将剩余特征分布到各组。在第二阶段，使用基于二元染色体的遗传算法选择最佳特征子集以最大化情感识别性能。最后，应用得分和排名级别融合技术进一步提高分类准确性。

RGB 相机。Glowinski 等人 [81] 提出了一种通过分析上半身手势，特别是头部和手部动作进行自动情感识别的方法。研究人员使用了来自 GEMEP 数据库 [82] 的 40 个情感片段——由专业演员表演，代表愤怒、喜悦、宽慰和悲伤——通过双摄像机装置拍摄视频，并提取了动态特征，如运动能量（速度大小的总和）和空间范围（包围三角形的周长）。在后续研究中，Glowinski 等人 [83] 从头部和手部的轨迹中提取了其他运动学特征，包括能量、空间范围、平滑度和对称性。他们应用主成分分析（PCA）来将特征维数减少到一个四维表示，这有效地将情感按效价（正 vs. 负）和唤醒度（高 vs. 低）的轴线分组。

在 [84] 中，提出了一种多分支深度学习架构，该架构由基于堆叠长短期记忆（LSTM）单元的局部分支和基于多层感知机（MLP）的全局分支组成。局部分支通过堆叠的 LSTM 层处理时间局部特征，以实现更高层次的抽象并捕捉细粒度的动作细节。同时，全局分支利用 MLP 处理时间全局特征并提取显著模式。两个分支的输出连接在一起，并经过一个全连接层，然后通过 softmax 函数进行情感分类。

在最近的一项研究中，O ğ uz 等人 [85] 调查了广泛的时域、频域和统计特征，应用特征选择技术以识别每帧的四个显著特征。这些选择的特征被汇总成一个特征矩阵，用于后续的情感识别。

王等人 [14] 提出了一种基于伪能量模型的多尺度特征选择算法，并引入了一种多尺度时空网络来解码情绪状态与全身运动之间的复杂关系。该方法有效地捕捉了长期姿势变化和

www.xueshuxiangzi.com

短期动态，从而提高了情绪识别性能。在这一研究方向的进一步延伸中，王等人 [86] 通过构建身体表达能量模型和设计多输入对称正定矩阵网络，探讨了关节能量特征。该框架有助于提取可解释的时空特征，有利于提高情绪分类的稳健性和可解释性。

4.1.2 **端到端网络方法**. 近年来，随着神经网络在广泛的人工智能领域取得显著成功，深度学习模型已成为从 3D 骨骼运动序列进行情感识别的主导范式。在本节中，我们根据网络架构的差异总结现有方法。

基于 RNN 的模型。Sapiński 等人 [87] 使用 Kinect V2 收集专业演员的身体运动数据，并利用 LSTM 网络基于骨骼序列对情感进行分类。Zhang 等人 [19] 提出了一种基于注意力机制的堆叠 LSTM（AS-LSTM）模型，用于在虚拟现实（VR）环境中对全身运动的情感识别。通过将注意力机制融入传统的 LSTM 框架中，该模型对运动帧中的关节点序列赋予不同的权重，从而能够集中关注关键关节，同时抑制冗余信息。此增强不仅提高了网络的学习能力，还提升了整体识别准确性。

基于 CNN 的模型。Karumuri 等人 [88] 率先提出将骨骼信息编码为图像表示。他们使用粗位置、细位置、逻辑位置和逻辑速度四种编码方案将三维关节坐标转换为 8 位 RGB 图像。然后设计了两种卷积神经网络 (CNN) 架构进行训练——单输入架构 (SIA-CNN) 和多输入架构 (MIA-CNN)。Cui 等人 [89] 借鉴肢体语言文献中的研究结果，将特定姿态与情感相关联。他们使用卷积姿态机器 (CPM) 算法提取人体关键点坐标，通过聚类生成简化线条图，并使用结合 Softmax 层的 CNN 进行情感分类。Beyan 等人 [20] 提出了一种双分支 CNN 架构，可以并行处理粗粒度 (例如 4 秒) 和细粒度 (例如 1 秒) 的时间特征，利用逻辑位置图像表示和数据增强技术来提高分类性能。

基于 GCN 的模型。Shen 等人 [90] 采用时间段网络（TSN）来提取 RGB 特征，并使用 ST-GCN 提取骨架特征。这些特征通过残差编码器统一，并通过一个残差全连接网络进行分类。Ghaleb 等人 [18] 开发了一个基于空间时间图卷积网络（ST-GCN）的情感识别框架，增强了空间注意机制，以突出关节空间模式与情感状态之间的关系。在 ST-GCN 架构的基础上，Shi 等人 [91] 引入了一种自注意机制，该机制动态调整骨架连通性结构，展示了自适应拓扑在捕捉情感线索方面的有效性。Shirian 等人 [92] 提出了可学习图 Inception 网络 (L-GrIN)，其中包含非线性谱图卷积、图 Inception 层、可学习的邻接关系和可学习的池化函数。该模型通过优化结合分类和图结构损失的复合损失函数，联合学习图结构并进行情感分类。

随着大型模型的快速发展，研究人员越来越认识到预训练模型在跨不同领域提供可转移的先验知识的价值，从而显著提升了任务表现。因此，人们对将预训练策略纳入基于骨架的情感识别的兴趣日益增加。

Paiva 等人在 [93] 采用了两阶段的方法：首先使用掩码自编码器（MAE）对大规模无标签骨骼数据集（MPOSE2021 和 Panoptic Studio）进行自监督预训练，以学习时空依赖关系，然后使用 MLP 分类器在 BoLD 数据集上进行微调。在 [94] 中，作者提出基于大语言模型的情感动作解释器（EAI-LLM）。该模型首先使用图卷积网络（GCN）提取骨骼特征，然后通过多粒度骨骼标注器和统一骨骼标注模块等组件将这些特征映射到语义空间中。通过利用大语言模型的推理能力，EAI-LLM 不仅执行情感识别，还生成可解释的文本解释。

4.1.3 **在公共数据集上的方法评估**. 为了方便不同方法的清晰比较，我们在表格 7 到 9 中总结了代表性方法在三个广泛使用的公共数据集——EGBM、KDAE 和 Emilya——上的性能。

在 EGBM 数据集上，早期基于 RNN 的方法取得了中等精度（分别为 69 % 和 74 %）。相比之下，近期一些结合手工特征与全连接网络（FC）的新方法，例如 Wang 等人所提出的 [14, 86]，报告显著更高的性能，超过 95 %。在 KDAE 数据集上也观察到了类似的模式。虽然基于 GCN 的模型 [18] 只取得了较为平淡的结果（65 %），但结合手工特征与浅层分类器或神经网络的方法表现出了强劲的性能。值得注意的是，O ğ uz 等人 [85] 在留出验证方案下实现了令人印象深刻的 99.99 % 精度。在 Emilya 数据集上，传统与深度学习模型表现均优异；基于 CNN 的方法 [20] 和驱动于手工特征的全连接网络 [14] 都超过了 94 % 的精度。

www.xueshuxiangzi.com

Table 7. 在 EGBM 数据集上的方法比较

Study	Backbone	Protocol	Accuracy
Sapiński et al.[87]	RNN	10 - fold	69.00 %
AS-LSTM [19]	RNN	10 - fold	74.00 %
Wang et al. [86]	Manual features+FC	10 - fold	97.43 %
Wang et al. [14]	Manual features+FC	10 - fold	95.55 %
EAI-LLM [94]	LLMs	Hold-out	66.97 %

Table 8. 方法在 KDAE 数据集上的比较

Study	Backbone	Protocol	Accuracy
Ghaleb et al. [18]	GCN	10 - fold	65.00 %
Wang et al. [86]	Manual features+FC	10 - fold	96.67 %
Wang et al. [14]	Manual features+FC	10 - fold	95.60 %
Oğuz et al. [85]	Manual features+NN	Hold-out	99.99 %
EAI-LLM[94]	LLMs	Hold-out	71.17 %

Table 9. 方法在 Emilya 数据集上的比较

Study	Backbone	Protocol	Accuracy
Beyan et al. [20]	CNN	5 - fold	96.59 %
Wang et al. [14]	Manual features+FC	10 - fold	94.42 %
EAI-LLM [94]	LLMs	Hold-out	85.44 %

总的来说，尽管端到端模型通常表现更为优越，但基于人工特征的方法仍然具有很强的竞争力，特别是在数据有限的情况下。

4.2 基于步态的情感识别

4.2.1 **传统方法**. 基于机器学习的方法一直是基于步态进行情感识别领域的基石。表格 10 总结了近年来一些关注基于人类步态的自动情感识别的代表性研究。为了便于更清晰地理解和比较基于步态的情感识别方法，本节根据捕捉步态信息的数据采集方法进行组织。具体来说，我们根据步态数据是通过运动捕捉系统、深度传感器（如 Kinect）、可穿戴设备还是基于视频的姿态估计来获得的，对现有工作进行分类。运动捕捉系统。在基于骨架的步态情感识别研究的早期阶段，运动捕捉系统，特别是 VICON 系统，被广泛用于获取高保真度的步态数据。Omlor 等人 [95] 是最早利用 VICON 系统从参与者处收集步态数据的研究者之一。他们引入了一种非线性源分离技术，从复杂全身运动的关节角轨迹中提取时空基元，特别关注于与情感步态相关的模式。Venture 等人 [96, 97] 最初进行了心理实验，研究了影响人类对情感步态感知的因素。随后，他们使用特征向量和主成分分析（PCA）来展示数值情感识别的可行性，并提出了一种基于相似性指数的分类算法。他们的实验同时采用了 6 自由度（DOF）和 12 自由度模型。

www.xueshuxiangzi.com

Table 10. 使用深度学习方法从步态中识别情感

Study	Dataset	Feature	Classifier	Protocol	Accuracy
Venture et al. [97]	4 actors / 100 samples	Lower torso movement (6DOF), waist rotations (3DOF), head inclinations (3DOF)	Similarity Index	LOSO	69 %
Karg et al. [98]	13 actors / 1300 samples	Statistical params: velocity, stride length; PCA, KPCA, LDA, GDA; eigenposture via Fourier	NB / NN / SVM	LOSO	60 % -69 %
Daoudi et al. [101]	BME	Cov. matrix of posture/velocity vectors	KNN	LOSO	71.12 %
Li et al. [49]	59 students	42 main frequencies + phases (Fourier)	NB / RF / SVM / SMO	10-fold	51.69 % -80.51 %
Li et al. [102]	59 students	44 time features, 2880 freq features via DFT of 6 joints	LDA / NB / DT / SVM	N/A	46 % -88 %
Ferdous et al. [103]	7 subjects	17 body/effort/shape/space features, quantized and normalized	SVM / KNN / LDA / DT	LOSO	62.86 % -74.29 %
Zhang et al. [104]	123 students	Temporal (skew, kurtosis), freq (PSD), time-freq (FFT) features	DT / SVM / RF	10-fold	56.60 % -81.20 %
Chiu et al. [107]	11 students	L2, angular, speed features	SVM / MLP / DT / NB / RF / LR	12-fold	53.3 % -62.1 %

SMO: Sequential Minimal Optimization; LDA: Linear Discriminant Analysis

Karg 等人 [98–100] 提取了与步态相关的特征，例如行走速度、步幅长度，以及关键关节角度的最小、平均和最大值（例如脖子、肩膀和胸部）。他们应用了包括 PCA 在内的降维技术，以便在使用机器学习模型进行情感分类之前管理特征的复杂性。在相关研究中，Daoudi 等人 [101] 将关节位置和速度数据转换为协方差矩阵，然后将其映射到对称正定 (SPD) 矩阵的非线性黎曼流形上。情感分类通过计算流形上的测地距离和几何平均来进行。深度传感器。2010 年，微软发布的深度传感摄像机引入了一种获取和分析步态数据的新技术方法。中国科学院的研究人员设计了一项实验，参与者观看引发情感的视频片段，然后进行步行，使用 Kinect v2 传感器记录这些步行过程 [49, 102] 。通过傅立叶变换提取了 168 个频域特征，以区分快乐、悲伤和中性情绪状态 [49] 。在此基础上，Li 等人 [102] 进一步结合了步幅长度和步态周期持续时间等时域特征，显著提高了分类性能。

在相关研究 [103] 中，从人体步态数据中提取了几何和运动学特征。这些特征包括与身体有关的特征（例如，头部倾斜角度、关节屈曲角度），与努力相关的特征（例如，动能、平均关节速度），与形状相关的度量（例如，密度指数），以及空间描述符（例如，身体收缩指数、对称性）。一共提取了 17 个特征，并通过矢量量化和标准化进行处理。然后使用基于二进制染色体的遗传算法为四个专家模型选择最优特征子集，从而提高每个模型的情感识别性能。可穿戴设备。随着智能手机和可穿戴设备的广泛采用，这些技术的内置传感器和摄像头为步态数据采集提供了更便捷和方便的方法。张等人 [104] 和崔等人 [48] 使用了定制设计的智能腕带来捕捉来自各种关节（包括右手腕和脚踝）的 3D 加速度数据。他们计算了如偏度、峰度和标准差之类的时域特征，并进一步使用功率谱密度和快速傅里叶变换（FFT）提取时域和频域特征，以分类三种情感状态。

同样地，Quiroz 等人 [105, 106] 让参与者在一个 250 米的 S 型走廊上行走时佩戴智能手表。他们的研究集中在从通过智能手表收集的传感器数据推断个体的情感状态，分析运动传感器捕捉到的步态模式与情感表达之间的关系。他们从加速度计数据中提取了常见的统计特征，包括平均值、标准差、最大值和最小值，并应用了随机森林和逻辑回归等机器学习算法进行情感分类。

RGB 相机。除了使用智能手机传感器之外，Chiu 等人 [107] 还利用智能手机摄像头来记录参与者的步态视频，并应用 OpenPose [51] 提取骨骼数据。通过分析从骨架中得出的欧几里得距离特征、角度特征和基于速度的特征，他们成功地基于步态实现了情感分类。

www.xueshuxiangzi.com

4.2.2 Feat2Net **方法** . 自 2018 年以来，使用骨骼数据进行基于步态的情感识别的研究进入了一个新阶段，其特点是将人工特征与深度学习技术相结合。

Randhavane 等人 [108] 使用 LSTM 网络提取步态特征，然后通过支持向量机 (SVM) 和随机森林 (RF) 等传统机器学习算法进行情感分类。Bhatia 等人 [109] 提取了手工特征，包括关节角度和关节间距离，并使用 LSTM 网络进行情感分类。Zhang 等人 [110] 提出了一种基于层次注意力机制的神经网络（MAHANN）用于步态情感识别。其框架通过运动情感模块提取包括相对位置、21 个关节的速度以及行走速度在内的运动特征，并通过动作情感模块提取包括五个选定关节角度的动作特征。然后这些特征使用三层全连接网络融合，以实现最终的情感分类。

4.2.3 FeatFusionNet **方法** . Bhattacharya 等人在 [58] 中通过提取 29 个与情感相关的特征（例如，步幅长度、关节角度）并将其与空间-时间图卷积网络（ST-GCN）的最终层进行连接，进一步将手工特征与深度学习模型结合，以提高分类性能。在另一项研究中，他们提出了一种基于自编码器的半监督框架，该框架结合分层注意力池化和潜在嵌入学习来进行情感识别 [111] 。

孙等人 [112] 提取了关节角度和加速度等特征，将视觉信息与原始骨架数据融合，并采用基于双向 LSTM 的分类器来识别四种情感类别。胡等人 [113] 将骨架关节和情感特征编码为图像表示，应用了一个双流 CNN 来提取特征，并使用基于 Transformer 的互补模块（TCM）通过利用两个流之间的互补信息来捕获长程依赖性。

张等人 [114, 115] 利用 GCN 建模骨架动态，同时采用 CNN 处理手工特征；两个流的输出在决策层融合以提高分类准确性。翟等人 [116] 提出了一个结合姿态流和运动流的双流框架。姿态流应用基于手工特征的回归约束，将先验情感知识嵌入到深度模型中，而运动流构建了一个高阶速度-加速度关系图以捕捉情感强度。最终分类是通过融合两种流的输出实现的。

4.2.4 **端到端网络方法** . 随着深度学习技术的进步，基于步态的情感识别逐渐从传统的特征工程转向端到端的学习方法。最近的研究不再依赖手工提取的时域和频域特征，而是更关注于使用诸如 Transformers 和图卷积网络（GCNs）等架构直接建模步态与情感状态之间的映射。从 2020 年开始，基于步态骨架数据的情感识别的端到端方法已成为主要范式。

基于 CNN 的模型。Narayanan 等人通过将 3D 骨架序列嵌入到 244×244 的 RGB 图像中，将多视图骨架数据编码为伪图像。具体来说，每个时间步的骨架关节的 Z 、Y 和 X 坐标分别被映射到 R 、G 和 B 通道。然后结合时间变压器的卷积神经网络（CNN）被用来提取跨模态情感特征。

基于 GCN 的模型。Zhuang 等 [117] 提出了一种扩展的联合连接方案，该方案结合了根节点全连接策略和收缩去噪模块，使模型性能提高了 12.6 %。Lu 等 [118] 深入研究了关节点拓扑对情感识别的影响，并引入了一种优化的关节点连接设计以提高分类准确性。Sheng 等 [61] 开发了一种增强注意力的时空图卷积网络（AE-STGCN），具有编码器-解码器架构，可以同时建模空间依赖性和时间动态。他们的框架支持关节身份识别和情感分类的多任务学习。Yin 等 [119] 提出了用于步态情感识别的多尺度自适应图卷积网络（MSA-GCN）。他们的模型集成了自适应选择时空图卷积（ASST-GCN），根据情感上下文动态选择卷积核，并应用跨尺度映射交互融合多尺度信息。Chen 等 [120] 提出了时空自适应图卷积网络（STA-GCN），解决了传统模型在捕捉隐式关节点关系和刚性多尺度时间特征聚合方面的局限性。这是通过专门的空间和时间特征学习模块实现的。

基于 Transformer 的模型。Zeng 等人 [121] 引入了 GaitCycFormer，一个基于 Transformer 的框架，该框架结合了周期位置编码，以及由循环内和循环间 Transformer 组成的双层架构。此设计使模型能够有效捕捉基于步态的情感识别中局部循环内和全球循环间的时间特征。

www.xueshuxiangzi.com

Table 11. 在 E-Gait 数据集上的方法比较

Study	Backbone	Protocol	Accuracy
TAEW [111]	RNN	Hold-out	84.00 %
STEP [58]	GCN	Hold-out	82.15 %
MSA-GCN [119]	GCN	Hold-out	93.51 %
T2A [114]	GCN	5-fold	82.91 %
TT-GCN [115]	GCN	5-fold	80.11 %
BPM-GCN [116]	GCN	5-fold	88.94 %
EIPC [118]	GCN	5-fold	82.25 %
G-GCSN [117]	GCN	10 - fold	81.50 %
STA-GCN [120]	GCN	N/A	85.80 %
MAHANN [110]	CNN	Hold-out	93.40 %
VFL [112]	CNN	Hold-out	89.29 %
Proxemo [123]	CNN	5-fold	80.01 %
TNTC [113]	Transfromer	5-fold	85.97 %
Gaitcycformer [121]	Graph-Transformer	Hold-out	86.30 %

4.2.5 **无监督方法**. 由于步态情感数据集的有限性，一些研究人员开始探索无监督的情感识别方法。主要策略是训练一个编码器，在没有标签监督的情况下提取与情感相关的步态特征，随后通过一系列下游任务评估学习到的表示的质量。

Lu 等人 [122] 提出了一种基于步态的情感识别自监督对比学习框架，旨在解决现有方法中步态多样性和语义一致性有限的挑战。该框架包括两个核心组件：(1) 模糊对比学习，通过修改步态速度和关节角度生成模糊样本，并将其整合到记忆库中以丰富语义多样性；(2) 跨坐标对比学习（C^3L），在笛卡尔坐标系和球面坐标系之间进行对比学习，以利用互补表示来提高语义不变性。

同样地，宋等人提出了一种自监督对比框架，该框架引入了选择性强增强，包括上半身抖动和随机时空遮掩等技术，以生成多样的正样本并促进鲁棒特征学习。他们进一步设计了一个互补特征融合网络（CFFN），以结合来自图域（通过 ST-GCN）的拓扑特征和来自图像域（通过自适应频率滤波器）的全局自适应特征，从而增强表示能力。为了确保一般样本与强增强样本之间的分布一致性，应用了分布性差异最小化损失。

为了提供一种清晰和系统的方法性能比较，我们总结了在两种广泛使用的基于步态的情感识别数据集：E-Gait 和 EMOGAIT 上评估的典型模型的结果。比较结果在表格 11 和表格 12 中展示。

在 E-Gait 数据集中，基于 GCN 的方法，如 MSA-GCN [119] 和 BPM-GCN [116] 展示了强劲的性能，其中 MSA-GCN 实现了最高的报告准确率 93.51 % 。基于 CNN 和 Transformer 的模型，包括 MAHANN [110] 和 GaitCycFormer [121]，也表现得具有竞争力，表明深度学习模型在捕捉局部和全局步态动态方面的有效性。

在 EMOGAIT 数据集上，最近的 GCN 变体——如 T2A [114] 和 TT-GCN [115] ——取得了超过 90 % 的准确率，进一步证实了基于图的时间建模在步态情感识别中的有效性。

www.xueshuxiangzi.com

Table 12. 在 EMOGAIT 数据集上的方法比较

Study	Backbone	Protocol	Accuracy
T2A [114]	GCN	5-fold	91.87 %
TT-GCN [115]	GCN	5-fold	90.25 %
AT-GCN [61]	GCN	Hold-out	86.80 %

Table 13. 常见姿势-情感特征列表

Feature Type		Detailed Description
Postural Features	Angle Features	Left shoulder–neck–right shoulder, left shoulder–neck–left elbow, Left hip–waist–left knee *, left shoulder–left elbow–left hand * left hip–left knee–left foot *, Head–neck–torso
	Distance Features	Left hand–torso *, left hand–left shoulder *, left hand–left hip * Left hand–neck *, left elbow–torso *, left foot–right foot
	Area Features	Left hand–neck–right hand, left shoulder–neck–right shoulder Left hand–hip–right hand, left elbow–neck–right elbow Left foot–waist–right foot, left knee–neck–right knee
Motion Features	Velocity Features	Head, hands, shoulders, knees, feet
	Acceleration Features	Head, hands, shoulders, knees, feet

* The corresponding feature is also computed on the right side (e.g. for left shoulder–left elbow–left hand, compute right shoulder–right elbow–right hand).

4.3 方法比较与分析

尽管数据来源和分类算法存在差异，大多数传统方法在从骨架数据中提取与情感相关的特征时采用了共同的策略。通过对众多具有代表性的研究进行详细调查，我们观察到经常使用的特征包括关节速度、关节之间的角度和相对距离—这些参数能够有效反映姿势和动作中的细微情感变化。我们在表 13 中总结了这些常用特征。

尽管深度学习方法被广泛认为是端到端的解决方案，但我们的分析表明，在数据稀缺或可解释性受到关注的情况下，许多方法仍依赖于人工特征。例如，一些 Feat2Net 和 FeatFusionNet 模型将域特定的特征（如运动能量或关节角度动态）作为神经网络的输入。此类混合设计反映了深度学习方法向完全自动化表示学习演变过程中的过渡阶段。

总体而言，基于图的方法在深度学习方法中扮演着越来越重要的角色，尤其是那些结合了时间注意机制或多尺度建模策略的方法。基于 CNN 和 Transformer 的模型也展现出显著的潜力，特别是在设计用于捕捉细粒度运动模式或长距离依赖时。这些趋势反映出一种更广泛的转变，即向能够从骨架步态序列中直接学习具有表现力和判别力的表示的端到端、数据驱动的框架转变。

同时，尽管 Lu 等人提出的方法 [94] 在分类准确率上略逊一筹，但它引入了一个独特的优势：能够在生成描述性文本解释的同时识别情感。这种基于 LLM 的方法为未来的研究开辟了一个有前景的方向，不仅使情感识别系统能够检测情感状态，还能提供增强透明度和用户信任的人类可解释的见解。

5 任务特定应用

5.1 使用骨架数据检测抑郁症

情绪通常被认为是一种短期的心理状态，而抑郁则是一种长期的、多方面的心理状况。尽管这两者有区别，但它们之间关系密切 [13] 。经历抑郁的人往往持续遭受消极情绪和情绪

www.xueshuxiangzi.com

低落的困扰。因此，研究人员试图将从基于步态的情绪识别中获得的见解扩展到基于步态的抑郁检测领域，旨在探索步态特征作为抑郁症状指标的潜力。

中国科学院心理研究所的一个研究团队 [124–126]，利用快速傅立叶变换（FFT）和希尔伯特-黄变换（HHT）等方法在频域中分析步态特征，建立了步态特征与抑郁水平之间的映射关系。Lu 等人 [127, 128] 研究了抑郁患者的步态模式，并提出了一种联合能量特征，有效地区分抑郁患者和健康对照组。方等人 [129] 来自中国科学院深圳先进技术研究院，提取了 12 种时域步态特征，如步行速度、步幅宽度、步幅长度和头部垂直运动，并分析这些特征与抑郁症状之间的相关性。

在早期研究的基础上，Wang 等人 [130] 整合了时域、频域和空间几何特征，开发了一种新颖的基于步态的抑郁评估算法。此外，Yang 等人 [131] 对骨架序列应用了各种数据增强策略，以研究不同形式的骨架数据对抑郁识别性能的影响。Shao 等人 [132] 通过在骨架序列旁加入视频数据，并引入骨架与步态轮廓特征之间的融合机制，进一步增强了模型的鲁棒性。

基于从视频数据集 [133] 中提取的骨架，Li 等人 [134] 提出了一种新颖的时空多粒度网络（STM-Net），用于基于骨架的抑郁风险识别，结合多粒度时间聚焦（MTF）模块和多粒度空间聚焦（MSF）模块，以捕捉步态模式中的动态时间信息和空间特征。

5.2 使用骨骼数据检测自闭症

自闭症谱系障碍（ASD）是一种长期的神经发育状况，其特征是社交沟通困难以及存在有限的、重复的行为。新兴研究表明，ASD 个体通常表现出放大的情绪反应和受损的情绪调节，这可能会反映在不同于神经典型个体的特定身体行为中 [135]。因此，研究人员开始通过行为提示——特别是步态模式——来研究识别 ASD，这是理解和检测自闭症相关特征的一种非侵入性方法。这一研究方向有望推动早期诊断和干预策略的发展。

研究 [136] 首次引入了一个框架，用 Kinect v2 深度相机捕捉和构建自闭症谱系障碍（ASD）儿童的 3D 步态和全身运动数据集，为行为分析和 ASD 相关研究提供了宝贵的资源。张等人。[137] 使用 OpenPose 算法从视频录制中提取初始骨架数据。然后使用一个骨架距离匹配算法进行多人跟踪，以关联同一场景中多名 ASD 儿童的骨架数据。最后，应用 LSTM 网络对去噪后的骨架序列进行分类，并自动提取时序特征。

Zahan 等人 [138] 利用带角嵌入的 GCN 从骨架数据中捕获时空特征，并结合骨架图像表示（Skepxels）和 Vision Transformer (ViT) 进行辅助训练。他们的研究表明，自闭症儿童表现出非典型的步态特征，例如更大的关节角度变异性和增加的步态不对称性。Yang 等人 [139] 使用 HRNet [52] 提取骨架关键点坐标，并应用 PoTion 算法生成关节运动轨迹图。他们的研究揭示，自闭症谱系障碍 (ASD) 儿童往往表现出比神经发育典型的同龄人更大的运动范围和更不规则的运动模式。

5.3 使用骨骼数据检测异常行为

异常行为检测和情感识别是两个相互关联但又不同的任务。情感识别旨在识别短期的情感状态，而异常行为检测则侧重于识别可能表明潜在心理或行为条件的异常或意外模式 [140]。尽管有这种区别，这两个领域紧密相连，因为许多异常行为要么是由情感变化引发，要么伴随着情感变化——例如，焦虑情绪的增加可能表现为躁动不安或回避行为。

基于 RNN 的模型。在研究的早期阶段，递归神经网络（RNNs）是解决这些任务的主要技术。在 [141]，使用了一种单目固定摄像机深度估计算法将二维骨架转换为三维表示。一个基于自注意力的时空卷积神经网络（ST-CNN）被用来提取局部时空特征，而一个增强注意力的 LSTM（ATT-LSTM）专注于关键帧以捕捉全局时间动态。Morais 等人在 [142] 中将人体骨骼运动分解为全局身体运动和局部身体姿势成分，它们是通过一种新颖的消息传递编码器-解码器递归网络（MPED-RNN）建模的。该模型由两个相互作用的 RNN 分支组成——一个用于全局特征，一个用于局部特征，通过跨分支消息传递在每个时间步交换信息。该网络利用重建和预测损失进行训练。

www.xueshuxiangzi.com

基于 GCN 的模型。随后，图卷积网络（GCNs）逐渐成为基于骨架的异常行为检测的主流范式。Markovitz 等人提出了用于异常检测的图嵌入姿势聚类（GEPC）方法，将人体姿势表示为时空图。他们的框架使用时空图卷积自编码器（STGCAE）来嵌入姿势图，应用深度嵌入聚类生成软分配向量，并利用狄利克雷过程混合模型（DPMM）计算正常性评分。

Liu 等人 [143] 提出了一个空间自注意力增强图卷积（SAA-Graph）模块，该模块结合了改进的时空图卷积和 Transformer 风格的自注意力，以捕捉局部和全局联合信息。他们的架构使用 SAA-STGCAE 进行特征提取，随后是深度嵌入聚类和 DPMM 用于异常评分。

Flaborea 等人 [144] 引入了 COSKAD，这是一种使用时空可分离图卷积网络 (STS-GCN) 编码骨骼运动的方法。该模型将骨骼嵌入投射到多个潜在空间——欧几里得空间、球面和双曲空间——并通过最小化每个空间中学习到的中心点的距离来检测异常。

Karami 等人提出了 GiCiSAD，这是一种用于基于骨架的视频异常检测的图拼图条件扩散模型。该框架由三个新颖的模块组成：(1) 一个基于图注意力的预测模块用于建模时空依赖，(2) 一个图级拼图制造器用于突出显示细微的区域级别的差异，(3) 一个基于图的条件扩散模型用于生成多样化的人体运动模式以帮助异常检测。

预训练和提示引导模型。最近，越来越多的研究开始将预训练权重和大规模模型结合起来，以提高基于骨架的异常行为检测的识别性能。佐藤等人 [145] 提出了一种提示引导的零样本框架用于异常动作识别，解决了传统基于骨架的方法的几个限制——即对目标域训练的依赖、对骨架误差的敏感性，以及正常样本注释的稀缺。他们的方法使用基于 PointNet 的置换不变提取器，以实现稀疏特征传播，并提高对骨架噪声的鲁棒性。该框架还通过对比学习将骨架特征与文本嵌入（由用户提供的异常动作提示中得出）的余弦相似度整合到异常评分中，从而间接整合了正常行为的知识。

Liu 等人引入了对比语言-骨架预训练框架（SkeletonCLSP），该框架利用大型语言模型通过三个关键机制增强基于骨架的动作识别：语义补偿、跨模态特征整合和异常校正。这个框架弥合了文本和骨架模态之间的语义差距，使得异常动作的识别更加稳健和具有广泛适应性。

6 挑战与未来研究方向

6.1 构建多样化的数据集

尽管对于从 3D 骨架数据中识别情感身体表达的兴趣日益增长，但高质量、大规模数据集的可用性仍然有限。大多数现有数据集的规模相对较小，并主要使用离散情感类别（例如：快乐、愤怒、悲伤）进行注释，如表 4 中所总结。基于这一限制，在数据集开发中需要解决几个关键方面：

- 丰富标注方案：大多数现有数据集使用离散的情感类别（例如，快乐、愤怒、悲伤）进行标注，但这些可能无法捕捉到情感肢体运动的全部复杂性。需要结合维度情感模型，例如效价-唤醒模型或 PAD，以获得更细致入微的情感表现，特别是在异常行为检测和情境感知干预等应用中 [146] 。此外，数据集开发不仅要关注规模和多样性，还应采用更丰富的、多模态的标注方案，整合分类和维度标签。加入年龄、性别和文化背景等变量可以提高识别系统在不同人群中的普适性和公平性 [57] 。
- 扩展数据收集场景：如表 5 所示，目前现有的公共情感身体表达数据集大多是在高度受控的实验室环境中收集的。虽然这样的设置确保了数据的干净和注释的一致性，但在实际应用中通常缺乏鲁棒性。将数据收集工作从高度受控的实验室环境扩展到更自然的现实世界环境中是至关重要的。这包括在工作场所、公共空间、教室或医疗环境等多样化环境中获取 3D 骨架数据，在这些地方，情感表达是在复杂的社会和环境刺激下产生的 [147] 。
- 探索生成技术：未来的研究应探索用于数据增强的生成技术，如半监督、无监督 [148] ，或合成数据生成 [131] 。这些方法可以帮助创建更丰富、更多样化的数据集，并减少对大型人工标注数据集的依赖，同时提高模型的泛化能力和可扩展性。

www.xueshuxiangzi.com

6.2 提高模型性能

基于对现有基于骨架的情感识别方法的分析，在模型设计的若干方面仍有显著的改进空间。

- 准确性：实现高识别准确率仍然是核心目标。身体动作中的情感表达可能微妙、模糊，或者因个人和上下文变化，使得模型难以可靠地捕捉区分性特征。复杂的情感或重叠的表达进一步增加了难度。
- 泛化：在特定数据集或人群上训练的模型常常难以泛化到未见的用户、多样的文化背景或不同的录制环境中。域迁移现象——由姿势质量、运动风格或相机视角的变化导致——可能会显著降低模型在现实应用中的性能。
- 可解释性：黑箱模型对为什么预测出特定的情绪提供的洞察有限。增强模型的可解释性对于调试、信任以及下游应用（如情绪推理或伦理 AI）非常重要。注意力机制、显著性映射或符号推理等方法可能有助于揭示模型输出背后的决策逻辑。

6.3 建立端到端且高效的情感识别框架

实时的情感识别对于人机交互、公共安全监控以及边缘设备上的情感计算等应用至关重要。然而，目前基于骨骼的情感识别流程通常依赖于多阶段的过程——要么通过专用传感器获取骨骼数据，要么使用姿态估计算法从 RGB 视频中提取骨骼 [89, 90] 。这些中间步骤不仅增加了系统的复杂性和延迟，还会引入噪声和错误传递，最终降低识别的准确性和计算效率。

为了克服这些限制，未来的研究应该探索开发端到端框架，该框架可以将原始输入数据（例如，视频或深度帧）直接映射到情感状态，而不需要过度依赖中间骨架表示或广泛的预处理。此类方法可以显著减少计算开销，并在资源受限的环境中实现更精简的部署 [93, 149] 。

此外，设计具有强大泛化能力且内存或计算需求低的轻量级模型对于实时推断至关重要。这需要在高效模型架构上进行创新，包括 transformer 剪枝、知识蒸馏和边缘优化的神经算子，以平衡性能和部署的可行性。

6.4 扩展到多人情感识别

大多数现有方法关注的是单人情感识别；然而，许多现实应用，如公共安全、教育和社交机器人，要求具备理解群体情感的能力 [150] 。

识别集体情感不仅需要建模个人的情绪状态，还需要建模人际动态和群体背景。这带来了一系列挑战，包括拥挤场景中的遮挡、个人表达差异以及缺乏专门的多人物基于骨架的情感数据集。

未来的研究应探讨关系建模方法，例如图或超图表示，以捕捉个人线索和群体层面的交互。此类方法可以促进开发更具社交意识、上下文敏感和健壮的情感识别系统，适用于现实世界中的多代理环境。

虽然 3D 骨骼数据可以捕捉到关于身体运动的丰富信息，但情感表达本质上是多模态的，包括面部表情、声音线索、生理信号，甚至大脑活动。仅依赖骨骼数据可能会限制识别的准确性，特别是对于微妙或模糊的情感状态。

集成其他模态—例如音频、视频或 EEG—可以提供互补的线索并增强系统的鲁棒性 [151, 152] 。例如，将身体姿势与语音韵律或面部表情结合起来，可以大幅度改善自然环境中的情感识别 [153] 。EEG 信号反映了内部情感过程，在外部情绪表达微弱或故意被压抑的场景中尤其有价值 [154] 。

然而，多模态融合引入了几个关键的挑战：

- 模态对齐和同步，尤其是在输入信号具有不同时间或空间分辨率的情况下。
- 模型复杂性和计算需求的增加，这可能会阻碍实时部署。
- 大型、注释良好的多模态情感数据集的有限可用性，阻碍了训练和评估。

www.xueshuxiangzi.com

6.5 利用大型模型进行基于骨架的情感识别

近年来，大规模模型的发展——例如 ChatGPT [155] 、LLaVa [156] 和 Gemini [157] ——为基于骨架的情感识别开拓了新的途径。然而，由于该模态的稀疏性、时间性和结构性，将这些模型直接应用于 3D 骨架序列仍然具有挑战性。不同于文本或图像，骨架数据需要仔细预处理——例如转化为类似于令牌的表示或伪图像——以符合大规模模型 [94] 所期望的输入格式。

为了更好地利用这些模型的能力，未来的研究可以探索支持不仅情感识别而且理解和生成的统一架构。特别是，这样的模型可以被设计为：

- 通过结合背景和情境信息来解释情感表达的潜在原因。
- 生成：根据目标情感标签或预定义场景生成情感表现力的身体动作。
- 原因：关于在互动或多代理环境中情绪状态的时间演变。

7 结论

本综述全面回顾了基于 3D 骨骼数据的情感识别最新进展，涵盖了基于姿势和基于步态的方法。通过检查数据采集方法、公开可用的数据集以及一系列技术策略——从传统的手工特征提取到深度学习架构和大规模预训练模型——我们提供了对这一快速发展的领域的统一视角。

与基于面部表情或语音的方法相比，基于骨架的情感识别提供了明显的优势，包括对环境变化的鲁棒性和增强的隐私保护。这些特性使其特别适合于医疗监测、人机交互和公共安全等实际应用。

尽管取得了相当大的进展，仍然存在几个关键挑战。这些挑战包括需要更加多样化和生态有效的数据集，提高用户和环境之间的泛化能力，以及增强深度学习模型的可解释性。此外，整合多模态信号（例如，语音、面部表情和生理数据）并利用大规模预训练模型为未来的研究提供了有希望的途径。

我们预期，下一波研究将集中于开发统一的、端到端的框架，这些框架应轻量级、可解释，并能够适应动态的、多模态的环境。这样的系统对于在真实世界场景中实现强大、可扩展且以人为中心的情感计算至关重要。

References

[1] A. Kołakowska, A. Landowska, M. Szwoch, W. Szwoch, and M. R. Wrobel, "Emotion recognition and its applications," *Human-Computer Systems Interaction: Backgrounds and Applications 3*, pp. 51–62, 2014.

[2] R. W. Picard, *Affective computing*. MIT press, 2000.

[3] J. M. Garcia-Garcia, V. M. Penichet, and M. D. Lozano, "Emotion detection: a technology review," in *Proceedings of the XVIII international conference on human computer interaction*, 2017, pp. 1–8.

[4] S. Li and W. Deng, "Deep facial expression recognition: A survey," *IEEE Transactions on Affective Computing*, vol. 13, no. 3, pp. 1195–1215, 2022.

[5] R. A. Khalil, E. Jones, M. I. Babar, T. Jan, M. H. Zafar, and T. Alhussain, "Speech emotion recognition using deep learning techniques: A review," *IEEE access*, vol. 7, pp. 117 327–117 345, 2019.

[6] N. Alswaidan and M. E. B. Menai, "A survey of state-of-the-art approaches for emotion recognition in text," *Knowledge and Information Systems*, vol. 62, no. 8, pp. 2937–2987, 2020.

[7] L. Shu, J. Xie, M. Yang, Z. Li, Z. Li, D. Liao, X. Xu, and X. Yang, "A review of emotion recognition using physiological signals," *Sensors*, vol. 18, no. 7, p. 2074, 2018.

[8] A. Dzedzickis, A. Kaklauskas, and V. Bucinskas, "Human emotion recognition: Review of sensors and methods," *Sensors*, vol. 20, no. 3, 2020. [Online]. Available: https://www.mdpi.com/1424-8220/20/3/592

[9] A. Kleinsmith and N. Bianchi-Berthouze, "Affective body expression perception and recognition: A survey," *IEEE Transactions on Affective Computing*, vol. 4, no. 1, pp. 15–33, 2012.

[10] F. Noroozi, C. A. Corneanu, D. Kamińska, T. Sapiński, S. Escalera, and G. Anbarjafari, "Survey on emotional body gesture recognition," *IEEE transactions on affective computing*, vol. 12, no. 2, pp. 505–523, 2018.

[11] J. M. Montepare, S. B. Goldstein, and A. Clausen, "The identification of emotions from gait information," *Journal of Nonverbal Behavior*, vol. 11, no. 1, pp. 33–42, 1987.

www.xueshuxiangzi.com

[12] J. Michalak, N. F. Troje, J. Fischer, P. Vollmar, T. Heidenreich, and D. Schulte, "Embodiment of sadness and depression—gait patterns associated with dysphoric mood," *Psychosomatic medicine*, vol. 71, no. 5, pp. 580–587, 2009.
[13] F. Deligianni, Y. Guo, and G.-Z. Yang, "From emotions to mood disorders: A survey on gait analysis methodology," *IEEE Journal of Biomedical and Health Informatics*, vol. 23, no. 6, pp. 2302–2316, 2019.
[14] T. Wang, S. Liu, F. He, W. Dai, M. Du, Y. Ke, and D. Ming, "Emotion recognition from full-body motion using multiscale spatio-temporal network," *IEEE Transactions on Affective Computing*, vol. 15, no. 3, pp. 898–912, 2024.
[15] Z. Zhang, "Microsoft kinect sensor and its effect," *IEEE multimedia*, vol. 19, no. 2, pp. 4–10, 2012.
[16] Q. Peng, C. Zheng, and C. Chen, "A dual-augmentor framework for domain generalization in 3d human pose estimation," in *CVPR*, 2024, pp. 2240–2249.
[17] C. Zheng, W. Wu, C. Chen, T. Yang, S. Zhu, J. Shen, N. Kehtarnavaz, and M. Shah, "Deep learning-based human pose estimation: A survey," *ACM Computing Surveys*, 2023.
[18] E. Ghaleb, A. Mertens, S. Asteriadis, and G. Weiss, "Skeleton-based explainable bodily expressed emotion recognition through graph convolutional networks," in *FG*, 2021, pp. 1–8.
[19] H. Zhang, P. Yi, R. Liu, and D. Zhou, "Emotion recognition from body movements with as-lstm," in *2021 IEEE 7th International Conference on Virtual Reality*, 2021, pp. 26–32.
[20] C. Beyan, S. Karumuri, G. Volpe, A. Camurri, and R. Niewiadomski, "Modeling multiple temporal scales of full-body movements for emotion classification," *IEEE Transactions on Affective Computing*, vol. 14, no. 2, pp. 1070–1081, 2023.
[21] S. Xu, J. Fang, X. Hu, E. Ngai, Y. Guo, V. Leung, J. Cheng, and B. Hu, "Emotion recognition from gait analyses: Current research and future directions," *arXiv preprint arXiv:2003.11461*, 2020.
[22] M.-A. Mahfoudi, A. Meyer, T. Gaudin, A. Buendia, and S. Bouakaz, "Emotion expression in human body posture and movement: A survey on intelligible motion factors, quantification and validation," *IEEE Transactions on Affective Computing*, vol. 14, no. 4, pp. 2697–2721, 2023.
[23] T. S. Rached and A. Perkusich, "Emotion recognition based on brain-computer interface systems," *Brain-computer interface systems-Recent progress and future prospects*, pp. 253–270, 2013.
[24] T. Dalgleish, "The emotional brain," *Nature Reviews Neuroscience*, vol. 5, no. 7, pp. 583–589, 2004.
[25] A. Kołakowska, A. Landowska, M. Szwoch, W. Szwoch, and M. R. Wróbel, "Modeling emotions for affect-aware applications," *Information Systems Development and Applications*, pp. 55–69, 2015.
[26] P. Ekman, "Universals and cultural differences in facial expressions of emotion." in *Nebraska symposium on motivation*. University of Nebraska Press, 1971.
[27] H. Gunes, B. Schuller, M. Pantic, and R. Cowie, "Emotion representation, analysis and synthesis in continuous space: A survey," in *2011 IEEE International Conference on Automatic Face & Gesture Recognition (FG)*. IEEE, 2011, pp. 827–834.
[28] B. Stephens-Fripp, F. Naghdy, D. Stirling, and G. Naghdy, "Automatic affect perception based on body gait and posture: A survey," *International Journal of Social Robotics*, vol. 9, pp. 617–641, 2017.
[29] J. A. Russell and A. Mehrabian, "Evidence for a three-factor theory of emotions," *Journal of research in Personality*, vol. 11, no. 3, pp. 273–294, 1977.
[30] M. K. Greenwald, E. W. Cook, and P. J. Lang, "Affective judgment and psychophysiological response: dimensional covariation in the evaluation of pictorial stimuli." *Journal of psychophysiology*, 1989.
[31] G. F. Wilson and C. A. Russell, "Real-time assessment of mental workload using psychophysiological measures and artificial neural networks," *Human factors*, vol. 45, no. 4, pp. 635–644, 2003.
[32] A. Mehrabian, "Pleasure-arousal-dominance: A general framework for describing and measuring individual differences in temperament," *Current Psychology*, vol. 14, pp. 261–292, 1996.
[33] G. E. Kang and M. M. Gross, "The effect of emotion on movement smoothness during gait in healthy young adults," *Journal of biomechanics*, vol. 49, no. 16, pp. 4022–4027, 2016.
[34] ——, "Emotional influences on sit-to-walk in healthy young adults," *Human movement science*, vol. 40, pp. 341–351, 2015.
[35] A. Barliya, L. Omlor, M. A. Giese, A. Berthoz, and T. Flash, "Expression of emotion in the kinematics of locomotion," *Experimental brain research*, vol. 225, pp. 159–176, 2013.
[36] R. Plutchik, "The nature of emotions: Human emotions have deep evolutionary roots, a fact that may explain their complexity and provide tools for clinical practice," *American scientist*, vol. 89, no. 4, pp. 344–350, 2001.
[37] H. G. Wallbott, "Bodily expression of emotion," *European journal of social psychology*, vol. 28, no. 6, pp. 879–896, 1998.
[38] S. Dahl and A. Friberg, "Visual perception of expressiveness in musicians' body movements," *Music Perception*, vol. 24, no. 5, pp. 433–454, 2007.
[39] M. M. Gross, E. A. Crane, and B. L. Fredrickson, "Methodology for assessing bodily expression of emotion," *Journal of Nonverbal Behavior*, vol. 34, pp. 223–248, 2010.
[40] N. Dael, M. Mortillaro, and K. R. Scherer, "Emotion expression in body action and posture." *Emotion*, vol. 12, no. 5, p. 1085, 2012.

www.xueshuxiangzi.com

[41] K. H. Kim, S. W. Bang, and S. R. Kim, "Emotion recognition system using short-term monitoring of physiological signals," *Medical and biological engineering and computing*, vol. 42, pp. 419–427, 2004.
[42] M. M. Gross, E. A. Crane, and B. L. Fredrickson, "Effort-shape and kinematic assessment of bodily expression of emotion during gait," *Human movement science*, vol. 31, no. 1, pp. 202–221, 2012.
[43] Y. Ma, H. M. Paterson, and F. E. Pollick, "A motion capture library for the study of identity, gender, and emotion perception from biological motion," *Behavior research methods*, vol. 38, no. 1, pp. 134–141, 2006.
[44] "DanceMotion Capture Database (DMCD)," http://dancedb.eu/, 2021, [Online; accessed April 16, 2025].
[45] N. Fourati and C. Pelachaud, "Perception of emotions and body movement in the emilya database," *IEEE Transactions on Affective Computing*, vol. 9, no. 1, pp. 90–101, 2016.
[46] M. Zhang, L. Yu, K. Zhang, B. Du, B. Zhan, S. Chen, X. Jiang, S. Guo, J. Zhao, Y. Wang *et al.*, "Kinematic dataset of actors expressing emotions," *Scientific data*, vol. 7, no. 1, p. 292, 2020.
[47] E. Volkova, S. De La Rosa, H. H. Bülthoff, and B. Mohler, "The mpi emotional body expressions database for narrative scenarios," *PloS one*, vol. 9, no. 12, p. e113647, 2014.
[48] L. Cui, S. Li, and T. Zhu, "Emotion detection from natural walking," in *Human Centered Computing: Second International Conference, HCC 2016, Colombo, Sri Lanka, January 7-9, 2016, Revised Selected Papers 2*. Springer, 2016, pp. 23–33.
[49] S. Li, L. Cui, C. Zhu, B. Li, N. Zhao, and T. Zhu, "Emotion recognition using kinect motion capture data of human gaits," *PeerJ*, vol. 4, p. e2364, 2016.
[50] S. Yu and X. Li, "Gait-based emotion recognition using spatial temporal graph convolutional networks," in *2021 International Conference on Computer Information Science and Artificial Intelligence (CISAI)*, 2021, pp. 190–193.
[51] Z. Cao, T. Simon, S.-E. Wei, and Y. Sheikh, "Realtime multi-person 2d pose estimation using part affinity fields," in *CVPR*, 2017, pp. 7291–7299.
[52] K. Sun, B. Xiao, D. Liu, and J. Wang, "Deep high-resolution representation learning for human pose estimation," in *CVPR*, 2019, pp. 5693–5703.
[53] D. Mehta, S. Sridhar, O. Sotnychenko, H. Rhodin, M. Shafiei, H.-P. Seidel, W. Xu, and C. Theobalt, "Vnect: Real-time 3d human pose estimation with a single rgb camera," in *ACM Transactions on Graphics (TOG)*, vol. 36, no. 4. ACM New York, NY, USA, 2017, pp. 1–14.
[54] T. Randhavane, U. Bhattacharya, K. Kapsaskis, K. Gray, A. Bera, and D. Manocha, "The liar's walk: Detecting deception with gait and gesture," *arXiv preprint arXiv:1912.06874*, 2019.
[55] M. Poyo Solanas, M. J. Vaessen, and B. de Gelder, "The role of computational and subjective features in emotional body expressions," *Scientific reports*, vol. 10, no. 1, p. 6202, 2020.
[56] T. Sapiński, D. Kamińska, A. Pelikant, C. Ozcinar, E. Avots, and G. Anbarjafari, "Multimodal database of emotional speech, video and gestures," in *ICPR Workshops*. Springer, 2019, pp. 153–163.
[57] A. Kleinsmith, P. R. De Silva, and N. Bianchi-Berthouze, "Cross-cultural differences in recognizing affect from body posture," *Interacting with computers*, vol. 18, no. 6, pp. 1371–1389, 2006.
[58] U. Bhattacharya, T. Mittal, R. Chandra, T. Randhavane, A. Bera, and D. Manocha, "Step: Spatial temporal graph convolutional networks for emotion perception from gaits," in *AAAI*, 2020, p. 1342–1350.
[59] I. Habibie, D. Holden, J. Schwarz, J. Yearsley, and T. Komura, "A recurrent variational autoencoder for human motion synthesis," in *28th British Machine Vision Conference*, 2017.
[60] H. Hicheur, H. Kadone, J. Grèzes, and A. Berthoz, *The Combined Role of Motion-Related Cues and Upper Body Posture for the Expression of Emotions during Human Walking*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2013, pp. 71–85. [Online]. Available: https://doi.org/10.1007/978-3-642-36368-9_6
[61] W. Sheng and X. Li, "Multi-task learning for gait-based identity recognition and emotion recognition using attention enhanced temporal graph convolutional network," *Pattern Recognition*, vol. 114, p. 107868, 2021.
[62] N. Fourati and C. Pelachaud, "Emilya: Emotional body expression in daily actions database." in *LREC*, 2014, pp. 3486–3493.
[63] S. Narang, A. Best, A. Feng, S.-h. Kang, D. Manocha, and A. Shapiro, "Motion recognition of self and others on realistic 3d avatars," *Computer Animation and Virtual Worlds*, vol. 28, no. 3-4, p. e1762, 2017.
[64] C. M. U. G. Lab, "Cmu graphics lab motion capture database," 2003, accessed: 2025-04-21. [Online]. Available: http://mocap.cs.cmu.edu/
[65] C. Ionescu, D. Papava, V. Olaru, and C. Sminchisescu, "Human3.6m: Large scale datasets and predictive methods for 3d human sensing in natural environments," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 36, no. 7, pp. 1325–1339, 2014.
[66] A. Kapur, A. Kapur, N. Virji-Babul, G. Tzanetakis, and P. F. Driessen, "Gesture-based affective computing on motion capture data," in *Affective Computing and Intelligent Interaction: First International Conference, ACII 2005, Beijing, China, October 22-24, 2005. Proceedings 1*. Springer, 2005, pp. 1–7.

www.xueshuxiangzi.com

[67] D. Bernhardt and P. Robinson, "Detecting affect from non-stylised body motions," in *International conference on affective computing and intelligent interaction*. Springer, 2007, pp. 59–70.
[68] N. Fourati and C. Pelachaud, "Multi-level classification of emotional body expression," in *2015 11th IEEE International Conference and Workshops on Automatic Face and Gesture Recognition (FG)*, vol. 1, 2015, pp. 1–8.
[69] ——, "Relevant body cues for the classification of emotional body expression in daily actions," in *2015 International Conference on Affective Computing and Intelligent Interaction (ACII)*. IEEE, 2015, pp. 267–273.
[70] N. Fourati, C. Pelachaud, and P. Darmon, "Contribution of temporal and multi-level body cues to emotion classification," in *2019 8th International Conference on Affective Computing and Intelligent Interaction*, 2019, pp. 116–122.
[71] A. Crenn, R. A. Khan, A. Meyer, and S. Bouakaz, "Body expression recognition from animated 3d skeleton," in *International Conference on 3D Imaging*, 2016, pp. 1–7.
[72] A. Crenn, A. Meyer, R. A. Khan, H. Konik, and S. Bouakaz, "Toward an efficient body expression recognition based on the synthesis of a neutral movement," in *Proceedings of the 19th ACM International Conference on Multimodal Interaction*, ser. ICMI '17. New York, NY, USA: Association for Computing Machinery, 2017, p. 15–22. [Online]. Available: https://doi.org/10.1145/3136755.3136763
[73] A. Crenn, A. Meyer, H. Konik, R. A. Khan, and S. Bouakaz, "Generic body expression recognition based on synthesis of realistic neutral motion," *IEEE Access*, vol. 8, pp. 207 758–207 767, 2020.
[74] S. Saha, S. Datta, A. Konar, and R. Janarthanan, "A study on emotion recognition from body gestures using kinect sensor," in *2014 international conference on communication and signal processing*. IEEE, 2014, pp. 056–060.
[75] S. Piana, A. Staglianò, F. Odone, and A. Camurri, "Adaptive body gesture representation for automatic emotion recognition," *ACM Transactions on Interactive Intelligent Systems*, vol. 6, no. 1, pp. 1–31, 2016.
[76] A.-A. Samadani, R. Gorbet, and D. Kulić, "Affective movement recognition based on generative and discriminative stochastic dynamic models," *IEEE Transactions on Human-Machine Systems*, vol. 44, no. 4, pp. 454–467, 2014.
[77] A. Kleinsmith, N. Bianchi-Berthouze, and A. Steed, "Automatic recognition of non-acted affective postures," *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, vol. 41, no. 4, pp. 1027–1038, 2011.
[78] N. Savva, A. Scarinzi, and N. Bianchi-Berthouze, "Continuous recognition of player's affective body expression as dynamic quality of aesthetic experience," *IEEE Transactions on Computational Intelligence and AI in games*, vol. 4, no. 3, pp. 199–212, 2012.
[79] M. Garber-Barron and M. Si, "Using body movement and posture for emotion detection in non-acted scenarios," in *2012 IEEE International Conference on Fuzzy Systems*. IEEE, 2012, pp. 1–8.
[80] F. Ahmed, A. H. Bari, and M. L. Gavrilova, "Emotion recognition from body movement," *IEEE Access*, vol. 8, pp. 11 761–11 781, 2019.
[81] D. Glowinski, A. Camurri, G. Volpe, N. Dael, and K. Scherer, "Technique for automatic emotion recognition by body gesture analysis," in *CVPR Workshops*, 2008, pp. 1–6.
[82] Swiss Center for Affective Sciences, University of Geneva, "The geneva multimodal emotion portrayals (gemep) corpus," https://www.unige.ch/cisa/gemep, 2025, accessed: 2025-06-23.
[83] D. Glowinski, N. Dael, A. Camurri, G. Volpe, M. Mortillaro, and K. Scherer, "Toward a minimal representation of affective gestures," *IEEE Transactions on Affective Computing*, vol. 2, no. 2, pp. 106–118, 2011.
[84] D. Avola, L. Cinque, A. Fagioli, G. L. Foresti, and C. Massaroni, "Deep temporal analysis for non-acted body affect recognition," *IEEE Transactions on Affective Computing*, vol. 13, no. 3, pp. 1366–1377, 2020.
[85] A. Oğuz and Ö. F. Ertuğrul, "Emotion recognition by skeleton-based spatial and temporal analysis," *Expert Systems with Applications*, vol. 238, p. 121981, 2024.
[86] T. Wang, S. Liu, F. He, M. Du, W. Dai, Y. Ke, and D. Ming, "Affective body expression recognition framework based on temporal and spatial fusion features," *Knowledge-Based Systems*, vol. 308, p. 112744, 2025.
[87] T. Sapiński, D. Kamińska, A. Pelikant, and G. Anbarjafari, "Emotion recognition from skeletal movements," *Entropy*, vol. 21, no. 7, p. 646, 2019.
[88] S. Karumuri, R. Niewiadomski, G. Volpe, and A. Camurri, "From motions to emotions: classification of affect from dance movements using deep learning," in *Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems*, 2019, pp. 1–6.
[89] C. Mingming, F. Jiandong, and Z. Yudong, "Emotion recognition of human body's posture in open environment," in *2020 Chinese Control And Decision Conference (CCDC)*. Ieee, 2020, pp. 3294–3299.
[90] Z. Shen, J. Cheng, X. Hu, and Q. Dong, "Emotion recognition based on multi-view body gestures," in *2019 ieee international conference on image processing (icip)*. IEEE, 2019, pp. 3317–3321.
[91] J. Shi, C. Liu, C. T. Ishi, and H. Ishiguro, "Skeleton-based emotion recognition based on two-stream self-attention enhanced spatial-temporal graph convolutional network," *Sensors*, vol. 21, no. 1, 2021. [Online]. Available: https://www.mdpi.com/1424-8220/21/1/205
[92] A. Shirian, S. Tripathi, and T. Guha, "Dynamic emotion modeling with learnable graphs and graph inception network," *IEEE Transactions on Multimedia*, vol. 24, pp. 780–790, 2021.

www.xueshuxiangzi.com

[93] P. V. Paiva, J. J. Ramos, M. Gavrilova, and M. A. Carvalho, "Skelett-skeleton-to-emotion transfer transformer," *IEEE Access*, 2025.
[94] H. Lu, J. Chen, F. Liang, M. Tan, R. Zeng, and X. Hu, "Understanding emotional body expressions via large language models," *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 39, no. 2, pp. 1447–1455, Apr. 2025. [Online]. Available: https://ojs.aaai.org/index.php/AAAI/article/view/32135
[95] L. Omlor and M. A. Giese, "Extraction of spatio-temporal primitives of emotional body expressions," *Neurocomputing*, vol. 70, no. 10, pp. 1938–1942, 2007, computational Neuroscience: Trends in Research 2007. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0925231206004309
[96] G. Venture, "Human characterization and emotion characterization from gait," in *2010 Annual International Conference of the IEEE Engineering in Medicine and Biology*, 2010, pp. 1292–1295.
[97] G. Venture, H. Kadone, T. Zhang, J. Grèzes, A. Berthoz, and H. Hicheur, "Recognizing emotions conveyed by human gait," *International Journal of Social Robotics*, vol. 6, pp. 621–632, 2014.
[98] M. Karg, K. Kühnlenz, and M. Buss, "Recognition of affect based on gait patterns," *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, vol. 40, no. 4, pp. 1050–1061, 2010.
[99] M. Karg, R. Jenke, W. Seiberl, K. Kühnlenz, A. Schwirtz, and M. Buss, "A comparison of pca, kpca and lda for feature extraction to recognize affect in gait kinematics," in *2009 3rd International Conference on Affective Computing and Intelligent Interaction and Workshops*. IEEE, 2009, pp. 1–6.
[100] M. Karg, R. Jenke, K. Kühnlenz, and M. Buss, "A two-fold pca-approach for inter-individual recognition of emotions in natural walking." in *MLDM posters*, 2009, pp. 51–61.
[101] M. Daoudi, S. Berretti, P. Pala, Y. Delevoye, and A. D. Bimbo, "Emotion recognition by body movement representation on the manifold of symmetric positive definite matrices," in *International Conference on Image Analysis and Processing*, 2017, pp. 550–560.
[102] B. Li, C. Zhu, S. Li, and T. Zhu, "Identifying emotions from non-contact gaits information based on microsoft kinects," *IEEE Transactions on Affective Computing*, vol. 9, no. 4, pp. 585–591, 2016.
[103] F. Ahmed, B. Sieu, and M. L. Gavrilova, "Score and rank-level fusion for emotion recognition using genetic algorithm," in *2018 IEEE 17th International Conference on Cognitive Informatics & Cognitive Computing (ICCI*CC)*, 2018, pp. 46–53.
[104] Z. Zhang, Y. Song, L. Cui, X. Liu, and T. Zhu, "Emotion recognition based on customized smart bracelet with built-in accelerometer," *PeerJ*, vol. 4, p. e2258, Jul. 2016. [Online]. Available: https://doi.org/10.7717/peerj.2258
[105] J. C. Quiroz, E. Geangu, and M. H. Yong, "Emotion recognition using smart watch sensor data: Mixed-design study," *JMIR Ment Health*, vol. 5, no. 3, p. e10153, Aug 2018. [Online]. Available: http://mental.jmir.org/2018/3/e10153/
[106] J. C. Quiroz, M. H. Yong, and E. Geangu, "Emotion-recognition using smart watch accelerometer data: Preliminary findings," in *Proceedings of the 2017 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2017 ACM International Symposium on Wearable Computers*, 2017, pp. 805–812.
[107] M. Chiu, J. Shu, and P. Hui, "Emotion recognition through gait on mobile devices," in *2018 IEEE International Conference on Pervasive Computing and Communications Workshops (PerCom Workshops)*, 2018, pp. 800–805.
[108] T. Randhavane, U. Bhattacharya, K. Kapsaskis, K. Gray, A. Bera, and D. Manocha, "Identifying emotions from walking using affective and deep features," *arXiv preprint arXiv:1906.11884*, 2019.
[109] Y. Bhatia, A. H. Bari, G.-S. J. Hsu, and M. Gavrilova, "Motion capture sensor-based emotion recognition using a bi-modular sequential neural network," *Sensors*, vol. 22, no. 1, p. 403, 2022.
[110] S. Zhang, J. Zhang, W. Song, L. Yang, and X. Zhao, "Hierarchical-attention-based neural network for gait emotion recognition," *Physica A: Statistical Mechanics and its Applications*, vol. 637, p. 129600, 2024. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0378437124001080
[111] U. Bhattacharya, C. Roncal, T. Mittal, R. Chandra, K. Kapsaskis, K. Gray, A. Bera, and D. Manocha, "Take an emotion walk: Perceiving emotions from gaits using hierarchical attention pooling and affective mapping," in *ECCV*. Springer, 2020, pp. 145–163.
[112] X. Sun, K. Su, and C. Fan, "Vfl—a deep learning-based framework for classifying walking gaits into emotions," *Neurocomputing*, vol. 473, pp. 1–13, 2022.
[113] C. Hu, W. Sheng, B. Dong, and X. Li, "Tntc: two-stream network with transformer-based complementarity for gait-based emotion recognition," in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2022, pp. 3229–3233.
[114] Z. Zhu, C. P. Chen, H. Liu, and T. Zhang, "Temporal group attention network with affective complementary learning for gait emotion recognition," in *2024 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*. IEEE, 2024, pp. 3026–3033.
[115] T. Zhang, Y. Chen, S. Li, X. Hu, and C. L. P. Chen, "Tt-gcn: Temporal-tightly graph convolutional network for emotion recognition from gaits," *IEEE Transactions on Computational Social Systems*, vol. 11, no. 3, pp. 4300–4314, 2024.

www.xueshuxiangzi.com

[116] Y. Zhai, G. Jia, Y.-K. Lai, J. Zhang, J. Yang, and D. Tao, "Looking into gait for perceiving emotions via bilateral posture and movement graph convolutional networks," *IEEE Transactions on Affective Computing*, vol. 15, no. 3, pp. 1634–1648, 2024.
[117] Y. Zhuang, L. Lin, R. Tong, J. Liu, Y. Iwamoto, and Y.-W. Chen, "G-gcsn: Global graph convolution shrinkage network for emotion perception from gait," in *ACCV Workshops*, 2021, pp. 46–57.
[118] H. Lu, S. Xu, S. Zhao, X. Hu, R. Ma, and B. Hu, "Epic: Emotion perception by spatio-temporal interaction context of gait," *IEEE Journal of Biomedical and Health Informatics*, vol. 28, no. 5, pp. 2592–2601, 2024.
[119] Y. Yin, L. Jing, F. Huang, G. Yang, and Z. Wang, "Msa-gcn: Multiscale adaptive graph convolution network for gait emotion recognition," *Pattern Recognition*, vol. 147, p. 110117, 2024. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0031320323008142
[120] C. Chen and X. Sun, "Sta-gcn:spatial temporal adaptive graph convolutional network for gait emotion recognition," in *2023 IEEE International Conference on Multimedia and Expo (ICME)*, 2023, pp. 1385–1390.
[121] Q. Zeng and L. Shang, "Gaitcycformer: Leveraging gait cycles and transformers for gait emotion recognition." in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 39, no. 9, 2025, pp. 9815–9823.
[122] H. Lu, X. Hu, and B. Hu, "See your emotion from gait using unlabeled skeleton data," in *AAAI*, vol. 37, no. 2, jun 2023, pp. 1826–1834. [Online]. Available: https://doi.org/10.1609%2Faaai.v37i2.25272
[123] V. Narayanan, B. M. Manoghar, V. S. Dorbala, D. Manocha, and A. Bera, "Proxemo: Gait-based emotion learning and multi-view proxemic fusion for socially-aware robot navigation," in *IROS*, 2020, pp. 8200–8207.
[124] N. Zhao, Z. Zhang, Y. Wang, J. Wang, B. Li, T. Zhu, and Y. Xiang, "See your mental state from your walk: Recognizing anxiety and depression through kinect-recorded gait data," *PloS one*, vol. 14, no. 5, 2019.
[125] Y. Yuan, B. Li, N. Wang, Q. Ye, Y. Liu, and T. Zhu, "Depression identification from gait spectrum features based on hilbert-huang transform," in *Human Centered Computing: 4th International Conference, HCC 2018, Mexico, December, 5–7, 2018, Revised Selected Papers 4*. Springer, 2019, pp. 503–515.
[126] Y. Wang, J. Wang, X. Liu, and T. Zhu, "Detecting depression through gait data: examining the contribution of gait features in recognizing depression," *Frontiers in psychiatry*, vol. 12, p. 661213, 2021.
[127] H. Lu, W. Shao, E. Ngai, X. Hu, and B. Hu, "A new skeletal representation based on gait for depression detection," in *2020 IEEE International Conference on E-health Networking, Application Services (HEALTHCOM)*, 2021, pp. 1–6.
[128] H. Lu, S. Xu, X. Hu, E. Ngai, Y. Guo, W. Wang, and B. Hu, "Postgraduate student depression assessment by multimedia gait analysis," *IEEE MultiMedia*, vol. 29, no. 2, pp. 56–65, 2022.
[129] J. Fang, T. Wang, C. Li, X. Hu, E. Ngai, B.-C. Seet, J. Cheng, Y. Guo, and X. Jiang, "Depression prevalence in postgraduate students and its association with gait abnormality," *IEEE Access*, vol. 7, pp. 174 425–174 437, 2019.
[130] T. Wang, C. Li, C. Wu, C. Zhao, J. Sun, H. Peng, X. Hu, and B. Hu, "A gait assessment framework for depression detection using kinect sensors," *IEEE Sensors Journal*, vol. 21, no. 3, pp. 3260–3270, 2021.
[131] J. Yang, H. Lu, C. Li, X. Hu, and B. Hu, "Data augmentation for depression detection using skeleton-based gait information," *Medical & Biological Engineering & Computing*, p. 2665–2679, jul 2022.
[132] W. Shao, Z. You, L. Liang, X. Hu, C. Li, W. Wang, and B. Hu, "A multi-modal gait analysis-based detection system of the risk of depression," *IEEE Journal of Biomedical and Health Informatics*, vol. 26, no. 10, pp. 4859–4868, 2021.
[133] X. Liu, Q. Li, S. Hou, M. Ren, X. Hu, and Y. Huang, "Depression risk recognition based on gait: A benchmark," *Neurocomputing*, vol. 596, p. 128045, 2024.
[134] Q. Li, M. Ren, X. Hu, X. Liu, L. Yao, and Y. Huang, "Spatio-temporal multi-granularity for skeleton-based depression risk recognition," *IEEE Journal of Biomedical and Health Informatics*, pp. 1–12, 2025.
[135] C. A. Mazefsky, J. Herrington, M. Siegel, A. Scarpa, B. B. Maddox, L. Scahill, and S. W. White, "The role of emotion regulation in autism spectrum disorder," *Journal of the American Academy of Child & Adolescent Psychiatry*, vol. 52, no. 7, pp. 679–688, 2013.
[136] A. A. Al-Jubouri, I. H. Ali, and Y. Rajihy, "Generating 3d dataset of gait and full body movement of children with autism spectrum disorders collected by kinect v2 camera," *Compusoft*, vol. 9, no. 8, pp. 3791–3797, 2020.
[137] Y. Zhang, Y. Tian, P. Wu, and D. Chen, "Application of skeleton data and long short-term memory in action recognition of children with autism spectrum disorder," *Sensors*, vol. 21, no. 2, p. 411, 2021.
[138] S. Zahan, Z. Gilani, G. M. Hassan, and A. Mian, "Human gesture and gait analysis for autism detection," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2023, pp. 3328–3337.
[139] M. Yang, M. Ni, T. Su, W. Zhou, Y. She, W. Zheng, and B. Hu, "Body posture based detection of autism spectrum disorder in children," *IEEE Sensors Journal*, 2025.
[140] P. K. Mishra, A. Mihailidis, and S. S. Khan, "Skeletal video anomaly detection using deep learning: Survey, challenges, and future directions," *IEEE Transactions on Emerging Topics in Computational Intelligence*, vol. 8, no. 2, pp. 1073–1085, 2024.
[141] K. Zhou, T. Wu, C. Wang, J. Wang, and C. Li, "Skeleton based abnormal behavior recognition using spatio-temporal convolution and attention-based lstm," *Procedia Computer Science*, vol. 174, pp. 424–432, 2020.

www.xueshuxiangzi.com

[142] R. Morais, V. Le, T. Tran, B. Saha, M. Mansour, and S. Venkatesh, "Learning regularity in skeleton trajectories for anomaly detection in videos," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2019, pp. 11 996–12 004.
[143] C. Liu, R. Fu, Y. Li, Y. Gao, L. Shi, and W. Li, "A self-attention augmented graph convolutional clustering networks for skeleton-based video anomaly behavior detection," *Applied Sciences*, vol. 12, no. 1, p. 4, 2021.
[144] A. Flaborea, G. M. D. di Melendugno, S. D' Arrigo, M. A. Sterpa, A. Sampieri, and F. Galasso, "Contracting skeletal kinematics for human-related video anomaly detection," *Pattern Recognition*, vol. 156, p. 110817, 2024.
[145] F. Sato, R. Hachiuma, and T. Sekii, "Prompt-guided zero-shot anomaly action recognition using pretrained deep skeleton features," in *CVPR*, 2023, pp. 6471–6480.
[146] M. Dahmane, J. Alam, P.-L. St-Charles, M. Lalonde, K. Heffner, and S. Foucher, "A multimodal non-intrusive stress monitoring from the pleasure-arousal emotional dimensions," *IEEE Transactions on Affective Computing*, vol. 13, no. 2, pp. 1044–1056, 2020.
[147] R. Guo, H. Guo, L. Wang, M. Chen, D. Yang, and B. Li, "Development and application of emotion recognition technology—a systematic literature review," *BMC psychology*, vol. 12, no. 1, p. 95, 2024.
[148] Z. Zhang, F. Liang, W. Wang, R. Zeng, V. C. Leung, and X. Hu, "Skeleton-based pre-training with discrete labels for emotion recognition in iot environments," *IEEE Internet of Things Journal*, 2025.
[149] Y. Luo, J. Ye, R. B. Adams, J. Li, M. G. Newman, and J. Z. Wang, "Arbee: Towards automated recognition of bodily expression of emotion in the wild," *International Journal of Computer Vision*, vol. 128, no. 1, pp. 1–25, 2020.
[150] Q. Li, Z. Liu, Z. Zhang, Q. Wang, and M. Ma, "Decoding group emotional dynamics in a web-based collaborative environment: A novel framework utilizing multi-person facial expression recognition," *International Journal of Human–Computer Interaction*, vol. 41, no. 5, pp. 3455–3473, 2025.
[151] S. Zhang, Y. Yang, C. Chen, X. Zhang, Q. Leng, and X. Zhao, "Deep learning-based multimodal emotion recognition from audio, visual, and text modalities: A systematic review of recent advancements and future prospects," *Expert Systems with Applications*, vol. 237, p. 121692, 2024.
[152] X. Li, Y. Zhang, P. Tiwari, D. Song, B. Hu, M. Yang, Z. Zhao, N. Kumar, and P. Marttinen, "Eeg based emotion recognition: A tutorial and review," *ACM Computing Surveys (CSUR)*, 2022.
[153] J. Yan, P. Li, C. Du, K. Zhu, X. Zhou, Y. Liu, and J. Wei, "Multimodal emotion recognition based on facial expressions, speech, and body gestures." *Electronics (2079-9292)*, vol. 13, no. 18, 2024.
[154] H. Kim and T. Hong, "Enhancing emotion recognition using multimodal fusion of physiological, environmental, personal data," *Expert Systems with Applications*, vol. 249, p. 123723, 2024.
[155] OpenAI, "Chatgpt: An ai language model," https://chat.openai.com/, 2024.
[156] H. Liu, C. Li, Q. Wu, and Y. J. Lee, "Visual instruction tuning," *NeurIPS*, vol. 36, 2024.
[157] G. Research, "Gemini 1.5: Unlocking multimodal understanding across millions of tokens of context," *arXiv*, vol. abs/2403.05530, 2024, technical report detailing the Gemini 1.5 Pro architecture and performance benchmarks. [Online]. Available: https://ar5iv.labs.arxiv.org/html/2403.05530

Received 20 February 2007; revised 12 March 2009; accepted 5 June 2009

www.xueshuxiangzi.com