机器学会“分步观察”｜哈工大（深圳）×复旦儿科联合攻关，破解自闭症AI早筛的隐藏瓶颈

作者：writer 来源：科学网

本文原地址：https://www.iikx.com/news/progress/43018.html

本文仅供学习交流之用，版权归原作者所有，请勿用于商业用途！

机器学会“分步观察”｜哈工大（深圳）×复旦儿科联合攻关，破解自闭症AI早筛的隐藏瓶颈。自闭症谱系障碍（ASD）的早筛查、早干预，能够显著改善儿童长期发育结局。但在真实临床场景中，早筛工作长期面临现实困境：金标准评估高度依赖资深临床医师的经验判断，主观差异难以避免，国内具备资质的专业人才缺口巨大。

近些年，计算机视觉、视频行为分析技术，为自闭症客观筛查带来新希望。不少研究聚焦识别孩子的社交手势、眼神反应等行为生物标志物，但有一个极易被忽视的隐性卡点：什么时候才是观察儿童反应的正确时间窗口？很多AI系统可以精准识别行为，却找不准该分析哪一段视频。窗口一旦偏移，再精妙的算法也如同在流沙上筑塔，分析结果失去临床可信度。

近日，哈尔滨工业大学（深圳）刘洪海教授团队与复旦大学附属儿科医院徐秀教授团队合作，在中国工程院Engineering集群分刊ENGINEERING Health and Rehabilitation Engineering上发表了题为A Framework for Behavioral Biomarker Analysis via Action Segmentation in Early Autism Screening的研究论文。该研究提出基于时间动作分割的自闭症早期筛查框架TAS²F（Temporal Action Segmentation-based Autism Early Screening Framework）。不再单纯检测孤立事件，而是完整解析医生与儿童的全部临床互动流程，自动锁定有效观测时段，为行为生物标志物的客观量化提供新方案。

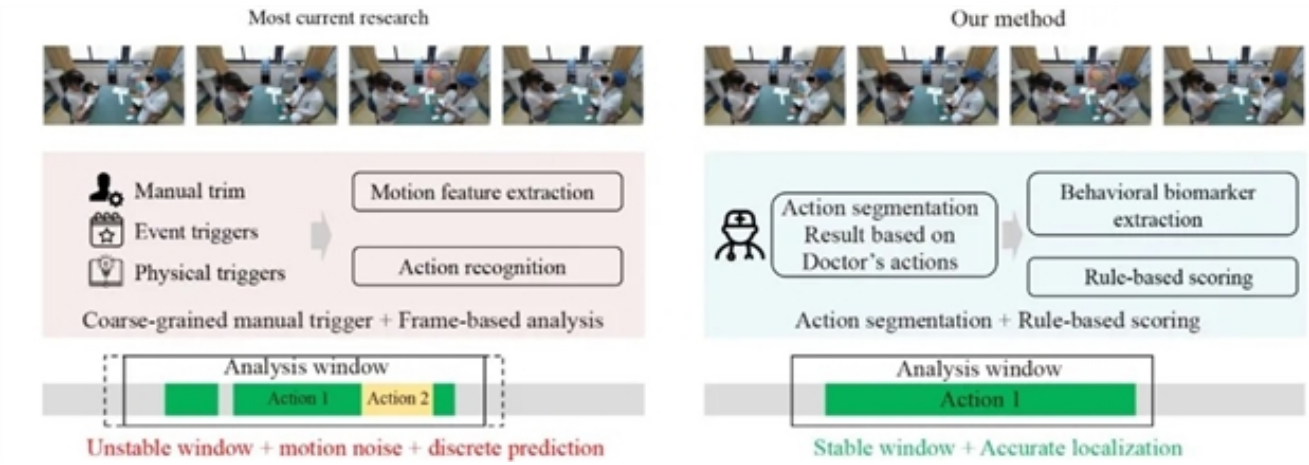


图1.传统机器辅助筛查范式与本研究TAS ^{2} F框架概念对比

旧方案的痛点：不稳定的分析窗口

传统机器辅助自闭症筛查流程，一般依靠人工剪辑视频、物理按键、语音识别等方式标记社交刺激的起始时刻。这种粗粒度触发方式存在明显短板：临床互动动作边界模糊，还存在各类环境噪声，很容易造成分析窗口偏移。系统会把无关动作、无效画面纳入计算，带来运动噪声与碎片化预测。 简单来说：标记的时间点错了，后续行为识别再准，结果也不可靠。

TAS ^{2} F框架转换了解决思路：不对单点事件做检测，而是把整套临床诊疗流程拆解为连续标准化动作片段，建立一套 时间坐标系。只有当医生完成刺激操作、进入规定的观察阶段，系统才启动对儿童行为的评估，得到稳定、精准的分析窗口。

研究选取自闭症筛查中两个核心协议开展验证：

表达性需求指向ENP：观察儿童是否使用手指指向物品表达需求，是早期沟通能力的关键标志；

回应性共同注意RJA：考察儿童能否跟随医生手势看向目标物体，共同注意缺陷是自闭症核心特征之一。

整套框架分为四大环节：

1.身份识别与特征提取：从视频中分离医生与儿童，提取医生人体骨架序列；

2.医生动作流解析、定位观察窗口：通过动作分割模型，自动切分出 互动建立-施加刺激-观察反应 完整流程，锁定合法观测时间段；

3.儿童行为计算：在划定窗口内，完成手势识别、头部姿态、视线方向、目标物体检测；

4.基于规则打分：依据临床协议规则，输出量化评分，为医生提供客观参考。

LAPS-AS：面向临床场景的骨架动作分割模型

框架的技术核心为语言辅助部分级分割模型LAPS-AS（Language-Assisted Part-level Segmentation），专门解决临床细粒度动作区分难题。临床场景里很多动作外观高度相似，例如放置泡泡瓶 和 放置发光小球，普通模型极易混淆。

模型包含两大核心模块：

1.解耦姿态编码器DPE不把人体当作一个整体处理，独立建模头部、手部等局部身体部件运动，同时兼顾全身运动上下文，捕捉手指、转头这类对临床评估至关重要的细微动作，避免关键细节被平滑丢失。

2.语言辅助分布对齐LDA + 语义偏移适配器SOA引入大语言模型生成临床动作的文本语义描述，引导骨架运动特征学习；通过损失函数约束，让语义相近的动作特征在空间上聚拢，拉开不同动作的决策边界。考虑文本描述与真实人体运动天然存在差距，SOA适配器进一步弥合文本语义和

真实运动之间的鸿沟。

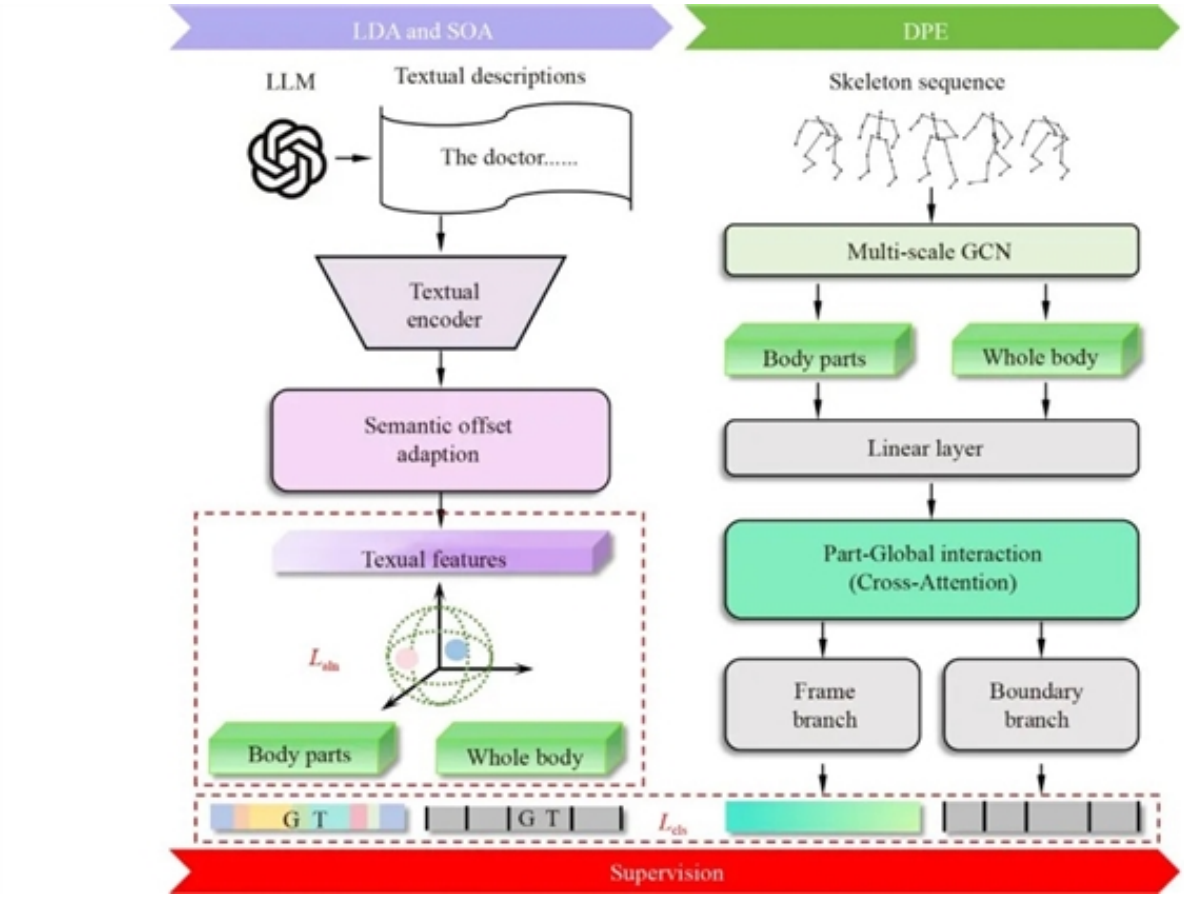


图2.LAPS-AS模型架构示意图，融合骨架运动特征与大模型语义引导

临床数据集与实验结果

研究依托复旦大学附属儿科医院，搭建真实临床数据集：包含48名受试儿童（25名ASD确诊，23名典型发育），共计96段临床筛查录像，总时长超230分钟，全部实验通过医院伦理委员会审批。

分割性能

· ENP协议（9类动作）帧级准确率：86.6%

· RJA协议（2类动作）帧级准确率：99.0%

定性可视化结果可以直观看到，对比MS-GCN、ASRF、DeST等主流骨架分割基线，LAPS-AS输出的动作片段边界更加贴合人工标注真值，减少过度分割、碎片化预测问题。

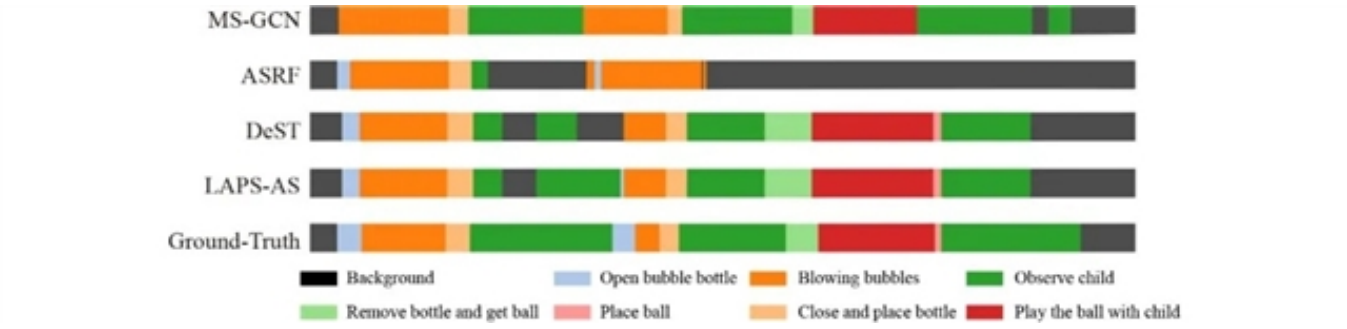


图3.ENP协议下不同模型时间动作分割结果定性对比

最关键：评分一致性

把模型自动分割得到的评分，和专家人工精准分割标注的评分做比对：

· ENP协议评分一致性：82.1%

· RJA协议评分一致性：93.6%

结果解读：RJA任务仅判断头部视线跟随，视觉信号简单稳定，一致性更高；ENP需要同时识别手势形态、指向方位、眼神接触，容易受手部遮挡、儿童快速动作干扰，是系统主要误差来源。消融实验证明，LDA语言辅助模块对整体性能提升贡献最大，证明把临床先验语义知识注入模型，对医疗场景算法至关重要。

临床意义

1.替代人工剪辑，释放科研人力。过去处理临床视频需要专家逐段剪辑标记，成本极高。TAS ^{2} F可以自动完成时间窗口定位，支持大批量视频的批量分析，加速自闭症行为表型相关研究。

2.作为临床辅助决策工具，而非替代医生诊断。该框架输出量化评分，给临床医师提供客观行为参考，缓解专业人员不足的压力，适配常规筛查工作流。

3.延伸应用潜力广阔。除辅助筛查外，这套技术还可用于：医生标准化筛查操作的质控培训；长期随访中客观追踪干预康复效果；挖掘人眼难以捕捉的细微行为生物标志物。

很多技术研发将重心放在如何识别孩子的异常行为，但这项工作提醒我们：在追求更聪明的识别模型之前，先要保证我们用于分析的时间窗口本身是干净、准确的。

TAS ^{2} F并非一套可以直接落地的最终产品，但它给出了一条务实路径：从临床流程本身出发，用时间动作分割搭建稳定的分析基座，弥合视频AI技术与真实自闭症早筛临床之间的鸿沟，为下一代客观计算机辅助筛查工具打下基础。（来源：EngineeringJournals微信公众号）

相关论文信息：https://journal.hep.com.cn/ehre/EN/10.2738/ENGHRE.2026.0007

特别声明：本文转载仅仅是出于传播信息的需要，并不意味着代表本网站观点或证实其内容的真实性；如其他媒体、网站或个人从本网站转载使用，须保留本网站注明的“ 来源 ”，并自负版权

等法律责任；作者如果不希望被转载或者联系转载稿费等事宜，请与我们接洽。
作者：刘洪海等 来源：《工程·健康与康复工程》

更多 科学进展 请访问 https://www.iikx.com/news/progress/

本文版权归原作者所有，请勿用于商业用途，爱科学iikx.com转发