

引文格式：

黄敏, 王端, 宋光辉, 等. 基于自然情绪感知下异常眼动轨迹的孤独症预测模型 [J]. 集成技术, 2023, 12(4): 64-76.
 Huang M, Wang D, Song GH, et al. Autism spectrum disorder prediction model based on gaze trajectory of natural emotional perception [J]. Journal of Integration Technology, 2023, 12(4): 64-76.

基于自然情绪感知下异常眼动轨迹的孤独症预测模型

黄 敏^{1,2#} 王 端^{3,4#} 宋光辉^{2,5} 梁 岩¹ 苏荣锋^{2,6}
 陈艳妮^{3,4*} 王 岚^{2,6*} 燕 楠^{2,6*}

¹(电子科技大学资源与环境学院 成都 611731)

²(中国科学院深圳先进技术研究院 深圳 518055)

³(西安市儿童医院(西安交通大学附属儿童医院)儿童保健中心 西安 710003)

⁴(西安交通大学医学部 西安 710049)

⁵(中国科学院大学 北京 100049)

⁶(中国科学院人机智能协同系统重点实验室 深圳 518055)

摘要 孤独症谱系障碍 (autism spectrum disorder, ASD) 是一类以社会交流、刻板行为和狭隘兴趣为主要特征的神经发育障碍性疾病, 致残率较高, 严重影响着儿童的健康成长。ASD 主观临床诊断存在耗时长、主观性强等问题。因此, 迫切需要一种快速、经济、有效的客观筛查方法。研究发现, ASD 儿童具有非典型的情绪视觉感知模式, 有望将眼动追踪技术用于 ASD 的辅助诊断。该文提出一个在自然场景下, ASD 非典型情绪视觉感知模式结合机器学习的自动筛查 ASD 患者的模型。该模型可提取自然场景下感知情绪的眼动轨迹特征, 通过机器学习模型进行建模, 以实现根据眼动轨迹自动识别 ASD 患儿。实验结果表明, 该方法的准确率为 79.71%, 有望成为一种 ASD 儿童早期筛查的辅助工具。

关键词 孤独症谱系障碍; 眼动追踪; 机器学习; 情绪感知

中图分类号 R 749.94; TP 181 **文献标志码** A doi: 10.12146/j.issn.2095-3135.20221114001

收稿日期: 2022-11-14 修回日期: 2023-02-20

基金项目: 深圳市科技计划项目 (KQTD20200820113106007, JCYJ20210324115810030, JCYJ20220818101217037); 国家自然科学基金项目 (81371900, 62271477, U1736202); 陕西省重点研发计划项目 (2020GXLYH-Y-013, 2021SF-194); 陕西中医药大学中西医结合防治脑发育障碍性疾病创新团队 (2019-YL07); 陕西省自然科学基础研究计划项目 (2022JQ-979); 国家重点研发计划项目 (2020YFC2004100); 西安市科技计划项目 (20YXYJ0006(4)); 陕西省科技研发计划项目 (2013SF2-09)

作者简介: 黄敏(共同第一作者), 硕士研究生, 研究方向为信号与信息智能处理; 王端(共同第一作者), 主治医师, 研究方向为儿童发育障碍性疾病; 宋光辉, 硕士研究生, 研究方向为说话人日志与多模态机器学习; 梁岩, 研究员, 研究方向为环境健康; 苏荣锋, 高级工程师, 研究方向为复杂声学环境下大规模鲁棒语音识别; 陈艳妮(通讯作者), 主任医师, 研究方向为儿童发育障碍性疾病的诊断与康复, E-mail: chenyannichil@163.com; 王岚(通讯作者), 研究员, 研究方向为语音信息处理以及多模式智能信息系统, E-mail: lan.wang@siat.ac.cn; 燕楠(通讯作者), 研究员, 研究方向为言语智能康复技术, E-mail: nan.yan@siat.ac.cn。

Autism Spectrum Disorder Prediction Model Based on Gaze Trajectory of Natural Emotional Perception

HUANG Min^{1,2#} WANG Duan^{3,4#} SONG Guanghui^{2,5} LIANG Yan¹ SU Rongfeng^{2,6}

CHEN Yanni^{3,4*} WANG Lan^{2,6*} YAN Nan^{2,6*}

¹(School of Resources and Environment, University of Electronic Science and Technology of China, Chengdu 611731, China)

²(Shenzhen Institute of Advanced Technology, Chinese Academy of Sciences, Shenzhen 518055, China)

³(Child Healthcare Department, Xi'an Children's Hospital (The Affiliated Children's Hospital of Xi'an Jiaotong University), Xi'an 710003, China)

⁴(Health Science Center, Xi'an Jiaotong University, Xi'an 710049, China)

⁵(University of Chinese Academy of Sciences, Beijing 100049, China)

⁶(Key Laboratory of Human-Machine Intelligence-Synergy Systems, Chinese Academy of Sciences, Shenzhen 518055, China)

*Corresponding Authors: chenyannichil@163.com; lan.wang@siat.ac.cn; nan.yan@siat.ac.cn

#Equal Contribution

Abstract Autism spectrum disorder (ASD) is a neurodevelopmental disorder characterized by impaired social communication, repetitive behaviors, and restricted interests. Signs of autism usually appear by age 3, but the clinical diagnosis is time-consuming and subjective. Therefore, a rapid and cost-effective assessment method is urgently needed. Children with ASD have atypical gaze patterns when they perceive emotional stimuli, which suggests a great potential use of eye-tracking technology as an assessment method for ASD detection. This paper proposes a model for automatically assessing children with ASD based on atypical gaze patterns. The model extracts the eye movement trajectory features of perceived emotions in natural scenes, and uses the machine learning model to learn to automatically identify ASD according to the eye movement trajectory features. Results show that the accuracy reaches 79.71%, which has potentially become an early ASD children screening approach.

Keywords autism spectrum disorder; eye tracking; machine learning; emotional perception

Funding This work is supported by Shenzhen Science and Technology Program (KQTD20200820113106007, JCYJ20210324115810030, JCYJ20220818101217037), National Natural Science Foundation of China (81371900, 62271477, U1736202), Shaanxi Province Key R&D Program (2020GXH-Y-013, 2021SF-194), Shaanxi University of Traditional Chinese Medicine Innovative Team Program (2019-YL07), Natural Science Basic Research Program of Shaanxi (2022JQ-979), National Key R&D Program of China (2020YFC2004100), Xi'an Science and Technology Planning Project (20YXYJ0006(4)) and Shaanxi Province Science and Technology Research and Development Plan (2013SF2-09)

Ethical Statement The study was approved by the Institutional Review Board of the Ethics Committee of the Affiliated Children's Hospital of Xi'an Jiaotong University (ID number: 20200409-4) and the Ethics Committee of the Shenzhen Institute of Advanced Technology, Chinese Academy of Sciences (ID number: YSB-2020-Y0802), and written informed consent was obtained from each participant

1 引言

孤独症谱系障碍 (autism spectrum disorder, ASD) 是一类以不同程度的社会交流和互动缺陷, 以及刻板行为、狭隘兴趣为主要特征的神经发育障碍性疾病, 其症状在儿童早期出现, 会限制或损害儿童的日常功能, 导致儿童在后续的人生发展中受到限制。孤独症发病率逐年提高, 据美国疾病控制与预防中心统计, 平均每 44 名儿童中就有一人患有 ASD^[1]。然而, 当前 ASD 临床诊断以症状描述为主, 基于临床中医生的观察和与儿童及儿童父母的访谈, 根据评估标准(如 ASD 诊断观察时间表第 2 版 (diagnostic observation schedule for ASD-second edition, ADOS-2)、精神障碍诊断与统计手册第 5 版 (diagnostic and statistical manual of mental disorders-5, DSM-5)^[2-3] 等) 进行诊断。该诊断过程繁复, 往往需要一个专业的诊疗团队耗费大量时间, 且结果具有一定主观性。特别是在经济欠发达的地区, 由于缺乏重测信度高的客观诊断方法, 可能导致对 ASD 儿童的误诊, 使其错过最佳干预时间。因此, 对 ASD 儿童的诊断急需一种经济、方便快捷、可信度高的客观辅助诊断方法。

构建客观辅助诊断系统, 需要从 ASD 的核心症状入手, 从获取的生理信号中提取能够反映 ASD 儿童核心症状的特征进行建模, 进而辅助医生诊断 ASD 儿童。由于行为和脑电生理信号较易获取, 两者一直是研究者探索的重点方向, 如姿势动作^[4-5]、语音^[6]、脑电信号^[7-8]与眼动轨迹^[9-12]等, 已有研究尝试将其用于辅助诊断 ASD。鉴于难以理解别人的情绪是 ASD 的核心特征, 也是社会交流和互动障碍的重要原因, 且具有跨文化普遍性^[2,13]。因此, 利用眼动追踪技术结合机器学习方法, 揭示情绪识别过程中的识别缺陷模式及严重程度, 对 ASD 儿童辅助诊断具有极大潜力^[14]。研究发现, ASD 在识别情绪

时具有非典型的视觉模式^[15-18]。与正常发育儿童相比, ASD 儿童非典型注视模式表现为对情绪面孔的注视更少, 缺乏兴趣^[15]; 正常儿童会动态地从一个区域转向另一个区域, 而大多数 ASD 儿童倾向专注于一个特定的区域^[16]; ASD 儿童对面部情绪表达的视觉注意力较低, 尤其是眼部区域^[17]。特别地, 在识别不同情绪时, 正常儿童眼神注视模式是不同的, 如在负面情绪中更多关注眼睛, 而在正面情绪中更多关注嘴巴^[18]; ASD 儿童则没有表现出这种情绪相关偏好, 仅视线出现下移^[17]。眼动跟踪是一种方便快捷且对于阐明潜在的情绪加工异常策略有价值的客观定量技术, 因此, 利用自然情绪感知下异常的眼动轨迹特点对 ASD 儿童进行识别, 有望成为 ASD 客观辅助诊断指标。

近年来, 大量研究将眼动追踪技术用于 ASD 辅助诊断。Cilia 等^[9]通过采集 59 名平均年龄为 8 岁的儿童观看视频的眼动数据, 对其扫视路径的动力学信息进行可视化信息提取, 并利用卷积神经网络进行分类, ASD 分类准确率达到了 71%。Wan 等^[10]通过采集 74 名 4~6 岁儿童观看一个 10 s 无声视频的眼动数据, 提取 10 个感兴趣区域 (areas-of-interest, AoI) 的注视相关特征, 并利用支持向量机 (support vector machines, SVM) 对 ASD 儿童和典型发育 (typically developing, TD) 儿童的注视模式进行分类, 其准确率达到了 85.1%。Tao 等^[19]基于 28 名 5~12 岁儿童观看 300 张图像的眼动数据集, 应用卷积神经网络与长短期记忆网络进行学习, 以判断观看一幅图像的眼动数据是否来自 ASD 儿童, 其准确率达到了 57.90%。

此外, 任务刺激的类型对 ASD 患儿的情绪识别研究也起着决定性的作用^[20]。如 de Wit 等^[21]发现, 在看到高兴、愤怒和恐惧的静态表情时, 与正常对照儿童相比, ASD 患儿眼睛注视并未减少; 但当刺激物为动态刺激时, ASD 患儿呈现出异常注视模式。韩俊霞等^[22]将儿童老师互动视频

作为刺激物的研究发现, 与正常对照儿童相比, ASD 儿童在脸部有显著的注视缺失, 其注意力更多集中于视频材料中的背景和身体部分。以上研究结果暗示, 由于静态刺激较外界刺激误差较少, ASD 儿童可以刻板地学习注视模式, 因而无法在真实复杂环境中表现出注视异常行为^[23]。电影场景刺激已被证实是量化情绪识别技能的有效刺激方式^[24], 通过对动态社会场景片段刺激时眼动追踪研究, 可以揭示儿童对现实情感认知过程中的行为特征。

基于上述讨论, 本文提出了一种基于自然情绪感知状态下, ASD 非典型注视特征的 ASD 儿童自动筛查方法。该方法通过提取自然情绪感知状态下儿童的眼动轨迹特征和可视化信息, 利用机器学习方法对 ASD 儿童与 TD 儿童的眼动轨迹进行分类, 进而实现对 ASD 儿童的辅助诊断。

2 材料与方法

2.1 参与者

本文研究对象为 2020 年 12 月—2021 年 4 月在西安交通大学附属儿童医院门诊招募的 ASD 儿童, 以及从西安市不同托儿机构招募的 TD 儿童。本研究已获得西安交通大学附属儿童医院伦理委员会和中国科学院深圳先进技术研究院伦理委员会批准。实验前, 提供实验说明, 并获得所有参与儿童监护人的同意。

参与本研究的儿童均排除: (1)严重颅脑外伤病史; (2)躯体重大疾病病史; (3)近期使用

神经精神类药物情况; (4)其他神经精神疾病病史; (5)斜视、散光等眼科疾病。ASD 组纳入标准: 要求 2 名以上的临床儿童神经精神科医生诊断儿童症状符合 DSM-5 和 ADOS-2^[2-3] 评分高于 ASD 诊断分值。TD 组纳入标准: ADOS-2 评分低于 ASD 诊断分值且格里菲斯发育评估量表中文版(Griffiths Development Scales-Chinese Edition, GDS-C) 分值处于正常范围内^[25]。本研究共纳入 ASD 儿童 48 例(轻度 8 例、中度 35 例、重度 5 例), TD 儿童 34 例, 研究对象的基本信息如表 1 所示。

2.2 数据采集

本研究使用的刺激源均来自 CASIA 中国自然情感视听数据库^[26-27], 将数据库中的视频标记为特定的情感类别。本研究共选取 14 个动态的中国电视剧视频, 时长为 2~10 s, 每个视频均展示了一个成年人表达自然情感的画面。14 个刺激视频包括 7 个快乐情绪视频(3 个快乐程度强, 4 个快乐程度弱)和 7 个悲伤情绪视频(3 个悲伤程度强, 4 个悲伤程度弱)。为防止夸张的装饰影响研究结果, 刺激中的人物都没有戴眼镜、可见的耳环或其他修饰。图 1 为情绪类型和级别示例, 所示图片为 CASIA 中国自然情感视听数据库中视频的截图^[26]。

实验在正常照明的房间中进行, 刺激材料在一个分辨率为 1 440×900 像素的 19.5 英寸液晶显示屏上播放, 儿童坐在椅子上, 距离显示器 60 cm, 眼睛与显示器平齐。监护人坐在距离儿童左侧约 80 cm 的椅子上, 避免儿童紧张和恐

表 1 研究对象基本信息

Table 1 The basic characteristics of participants

	男/女人数	年龄(月)	年龄均值(标准差)	ADOS-2 得分	GDS-C 得分	运动得分	语言得分	手眼协调得分	表现得分	实际推理得分	个人社交得分
ASD 组	40/8	37.73~73.10	51.36(8.58)	7~23	1.00~60.00	1.00~95.00	1.00~55.00	1.00~97.50	1.00~70.00	0.00~90.00	1.00~60.00
TD 组	25/9	38.07~79.53	56.74(8.83)	0~3	20.00~99.00	30.00~99.00	17.50~99.00	2.50~99.00	7.50~99.00	12.50~99.00	10.00~99.00



图 1 情绪类别与级别示例

Fig. 1 Example of the types of emotions

惧，实验全程保持安静。为捕捉受试者在自然状态下的眼球运动轨迹，实验时未固定受试者的头部，仅在整个实验过程中(持续 4~5 min)将儿童的背部固定在椅背上，如图 2 所示。



图 2 数据采集示意图

Fig. 2 Schematic diagram of data collection

受试者被要求观看所有视频 3 次，视频以 4 种顺序播放。序列 1 包含 14 个随机呈现的视频；序列 2 将快乐情绪视频和悲伤情绪视频分开，且相同类型的情绪视频进行随机播放；序列 3 和序列 4 根据情感的强弱，分别呈现两种情感视频，序列 3 由强到弱，序列 4 由弱到强。序列 1 和序列 2 为每个孩子提供，序列 3 和序列 4 随机提供一个。这样的随机化避免了儿童在实验过

程中可能出现的顺序效应和疲劳。为消除声音干扰，所有视频均为静音播放，每两个视频之间有 2 s 的空白间隔。眼动数据使用 Tobii Pro Fusion 非接触式眼球追踪装置进行采集，采样频率为 120 Hz。在视频播放开始前，采用标准的五点校准方法对数据进行校准。

2.3 感兴趣区域眼动特征提取方法

2.3.1 原始眼动数据获取

Tobii 眼动仪可记录当前时刻眼动行为产生的一系列数据，包括眼睛坐标、角速度、瞳孔直径等，并将眼动行为分为 4 种类型：注视、扫视、未知类型和没有发现眼睛。注视的次数和时间可反映主体对注视目标的兴趣程度。本文主要提取注视相关的特征，定义注视为眼睛对一个小范围目标 60 ms 以上较为稳定的锁定。通过 Tobii 眼动仪的数据处理软件 Tobii Pro Lab version 1.162 将儿童注视相关的信息进行提取并保存，以供后续试验分析。

2.3.2 感兴趣区域划分标准

由于本研究主要关注儿童情绪感知的眼动模式，因此将视频材料逐帧划分为 5 个 AoI 区域：眼睛、鼻子、嘴巴、面部其他区域和视频中除面部以外的区域。并在此基础上，分析两类儿童在不同 AoI 区域的注视分布特征。

AoI 区域划分的具体方法如下：(1) 采用 OpenFace 开源工具箱^[28]检测视频每一帧的人脸 68 个关键点 $P_i, i \in (0, 1, \dots, 67)$ 的坐标。OpenFace 是一个基于神经网络的人脸识别工具包，常用于检测视频中的人脸。自动检测人脸后，手动复核人脸坐标的准确性，若出现关键点坐标偏移，则手动对该关键点进行标注，以确保整体 AoI 区域划分准确。(2) 根据关键点坐标确定 5 个 AoI 区域。具体实现方法如图 3 所示。

眼部区域：首先，定义一个偏移量 $\Delta = 0.1 \times |P_0 P_{16}|$ 。将 $P_{17} P_{26}$ 分别向左、向右各延伸 Δ ，得到 $P'_{17} P'_{26}$ ，然后分别向上、向下平移 Δ ，得到

$\mathbf{U}_{17}\mathbf{U}_{26}$ 和 $\mathbf{D}_{17}\mathbf{D}_{26}$ 。 $\mathbf{U}_{17}\mathbf{U}_{26}\mathbf{D}_{17}\mathbf{D}_{26}$ 所框定的矩形范围定义为眼睛区域。

鼻子区域: 将 $\mathbf{P}_{31}\mathbf{P}_{35}$ 分别向左、向右各延伸 Δ , 得到 $\mathbf{P}'_{31}\mathbf{P}'_{35}$, 然后向上平移 Δ , 得到 $\mathbf{U}_{31}\mathbf{U}_{35}$, 向下平移 $0.5 \times \Delta$, 得到 $\mathbf{D}_{31}\mathbf{D}_{35}$, $\mathbf{U}_{31}\mathbf{U}_{35}\mathbf{D}_{31}\mathbf{D}_{35}$ 所框定的矩形范围即为鼻子区域。

嘴巴区域: 将 $\mathbf{P}_{48}\mathbf{P}_{54}$ 分别向左、向右各延伸 Δ , 得到 $\mathbf{P}'_{48}\mathbf{P}'_{54}$, 然后分别向上、向下平移 Δ , 得到 $\mathbf{U}_{48}\mathbf{U}_{54}$ 和 $\mathbf{D}_{48}\mathbf{D}_{54}$ 。 $\mathbf{U}_{48}\mathbf{U}_{54}\mathbf{D}_{48}\mathbf{D}_{54}$ 所框定的矩形范围即为嘴巴区域。

脸部区域: 定义为以 P_{29} 为圆心, $\mathbf{P}_8\mathbf{P}_{29}$ 为长轴, 短轴长为 $0.5 \times (|\mathbf{P}_0\mathbf{P}_{29}| + |\mathbf{P}_{16}\mathbf{P}_{29}|)$ 。

其他区域: 定义为脸部区域以外的其他视频区域。

2.3.3 眼动特征参数提取方法

本研究提取了 18 维 AoI 眼动特征, 包括 5 个感兴趣区域内注视点的个数、注视的时长, 视频播放期间总的注视点个数、总的注视时长、各个注视点时长的均值和方差、视频中心距离(注视点与视频中心距离)的均值及方差、第一次注视的时间占视频时长的比重以及回看次数(重新注视之前注视过的感兴趣区域的次数), 所有注视次数、注视时长和回看次数特征均在各个视频上进行最大最小归一化处理, 以避免由于视频时长不一致所带来的影响。

2.4 眼动追踪数据可视化建模方法

由于传统 AoI 区域眼动特征提取方法需要大量手动校正, 难以实现自动化, 因此, 本文提出一种新型的眼动追踪数据可视化特征提取方法, 可实现自动提取感兴趣区域, 并将注意力的眼动数据分布与感兴趣区域相融合。

机器视觉研究中的视觉显著性检测算法能够通过模拟人的视觉特点, 提取视频或图像中的显著区域, 即人类感兴趣的区域^[29]。因此, 本文采用一种视觉显著性检测算法^[30]实现对感兴趣区域的划分。该算法联合考虑了时间信息与空间信息, 将空间信息表示成单个帧中的帧内显著性, 时间信息则提供帧间约束和运动线索, 从而实现视频序列中提取与运动相关的显著对象。视频显著性检测模型可自动生成刺激视频每一帧的显著性图像, 这种视觉显著性图像能够表示人的视觉注意力机制, 如图 4 所示。本研究由于视频感兴趣区域分布较为集中, 所以直接将视频每一帧显著性图像的像素值进行相加, 再除以帧数, 得到的图像即为视频最终的显著性图像。

在本研究中, 注意力分布由一个与刺激材料分辨率相同、像素值为刺激材料视频相应位置处的注视点个数的图像表示, 并进行归一化处理。由于矩阵较为稀疏, 本研究利用最大值滤波器进行滤波。在滤波器核尺寸选取方面, 本文探索

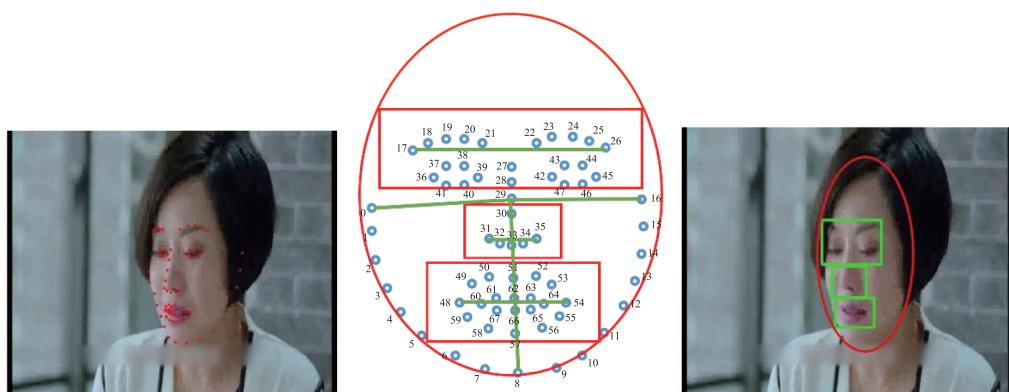


图 3 划分感兴趣区域

Fig. 3 Divide the area of interest

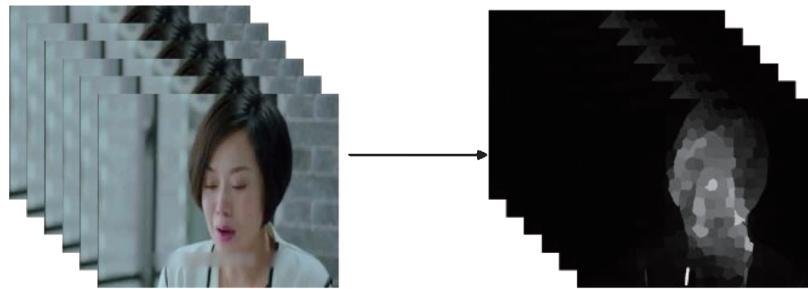


图 4 视频显著性生成

Fig. 4 Generate video saliency

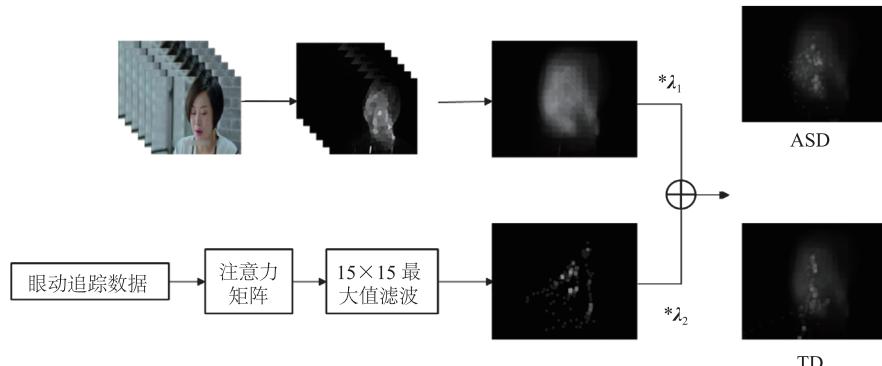


图 5 眼动追踪数据可视化建模

Fig. 5 Eye tracking data visualization

了多种不同的核尺寸，最终选用 15×15 的最大值滤波器进行滤波。最后，对视频显著性图像与注意力分布图像预设置一定的权重，进行叠加融合，整体流程如图 5 所示。

2.5 分类模型

针对不同类型的特征，本文采用不同的分类模型。对于 AoI 区域相关眼动特征，本文基于 ASD 组与 TD 组的数据分布差异分析进行特征选择，使用具有显著性差异的特征作为最终的眼动轨迹特征。应用 SVM、逻辑线性回归、 K 近邻、随机森林等机器学习算法进行分类建模。

对于眼动追踪数据可视化图像，本文采用分步训练的方式。使用预训练的 ResNet 18^[31] 进行迁移学习，模型权重初始化为 ImageNet^[32] 数据集上预训练的权重。本文模型将 ResNet18 的最后一个全连接层替换为 3 个全连接层：Fc(512,256)、Fc(256,128)、Fc(128,2)，以实现

二分类任务。为解决过拟合问题，采用 pytorch 中的 Dropout 函数对神经元激活分布，权重设为 0.5。将 ResNet 18 最后一个卷积层的输出作为特征，通过一个自适应平均池化层将其降维至 (512,1)，再进行主成分分析，降维至 (60,1)，最后使用决策树模型进行分类。

两种模型输出的结果均为视频级眼动轨迹分类的结果，即每个儿童观看一次视频所产生的眼动分类标签。对儿童所有视频级结果进行决策融合，采用软投票方式：由分类器输出儿童为 ASD 的概率，并对分类器输出结果求均值，选取 0.6 作为阈值，若均值大于 0.6，那么判定该儿童为 ASD，否则判定为 TD。

3 结 果

删除无效数据和异常数据后，最终获得的数

据集共有 2 182 个样本。其中, ASD 组共 1 125 个样本, TD 组共 1 057 个样本。

3.1 AoI 区域眼动特征的统计分析结果

在 ASD 组和 TD 组儿童中, 本研究对提取的 18 维特征分布进行了差异分析, 由于数据不满足正态分布, 故采用非参数 Mann-Whitney U 检验分析两组儿童间的数据分布是否有显著性差异, 显著性水平设置为 0.05(双尾), 结果如表 2 所示。

由表 2 可知, 与 TD 组相比, ASD 组的注视次数、注视时长、注视时长均值、注视时长方差、注视眼睛的次数和时长以及回看次数均较低, 而注视脸部其他区域的次数和时长、注视视频其他区域的时长、视频中心距均值和方差均显著较高。在 18 维眼动特征中, 有 13 维特征在两组儿童间的总体分布具有显著性差异。

3.2 分类结果

由于数据集较小, 本文进行了五折交叉验证, 以 8:2 的比例划分训练集与测试集, 模型最终的评价指标取 5 次预测的平均值。令 ASD 为正例, TD 为负例, 对模型分类结果的 4 个指标进行评估, 指标分别为准确率(Accuracy)、精

度(Precision)、召回率(Recall)和 F1。

对于 AoI 特征, 选取表 2 中具有显著性差异的 13 维特征, 利用多个机器学习模型进行学习, 结果如表 3 所示。由表 3 可知, SVM 模型的效果最好, 准确率为 75.94%, F1 为 81.07%。

对于眼动追踪可视化图像分类模型, 特征提取器的学习率为 10^{-6} , Batch 为 16, 采用 Adam 优化器, 损失函数为交叉熵损失函数, 训练轮数为 80, 采用决策树模型进行分类。对注意力分布图像与视频显著性图像融合的权重进行多次实验, 以确定最优权重, 实验结果如表 4 所示。比较发现当权重比为 6:4 时, 模型取得最优秀果, 决策融合的准确率为 79.71%。

最后, 将本文方法与当前常用的基于视频刺激的眼动轨迹 ASD 筛查方法进行对比, 并在本文数据集上实现 Cilia 等^[9]提出的眼动表征方法。该方法通过计算眼动轨迹的 3 个动力学物理量: 速度、加速度和变加速度, 并编码成红绿蓝色彩, 从而将眼动轨迹的动态转移表示成图像。采用 ResNet 18 进行分类, 3 种眼动特征工程方法的模型分类结果如表 5 所示。与 Cilia 等^[9]的

表 2 眼动特征及差异分析

Table 2 The significance of eye movement characteristics

	注视次数	注视时长	注视时长均值 (ms)	注视时长方差 (ms^2)	眼注视次数	鼻注视次数
ASD 组	0.211±0.180	0.058±0.131	406±273	159 374±576 156	0.103±0.174	0.192±0.203
TD 组	0.248±0.189	0.074±0.144	459±267	181 514±400 363	0.144±0.195	0.181±0.198
显著性	0.000***	0.000***	0.000***	0.000***	0.000***	0.151
	嘴注视次数	脸部其他区域注视次数	视频其他区域注视次数	眼注视时长	鼻注视时长	嘴注视时长
ASD 组	0.285±0.220	0.191±0.225	0.126±0.184	0.079±0.159	0.154±0.187	0.281±0.234
TD 组	0.277±0.215	0.189±0.241	0.189±0.228	0.123±0.202	0.168±0.211	0.282±0.236
显著性	0.588	0.045*	0.000***	0.000***	0.733	0.958
	脸部其他区域注视时长	视频其他区域注视时长	视频中心距均值 (像素)	视频中心距方差 (像素 ²)	回看次数	第一次注视时间/视频时长
ASD 组	0.382±0.175	0.153±0.175	118±43	3495±4090	0.425±0.189	0.309±0.121
TD 组	0.338±0.148	0.118±0.143	107±34	2679±3320	0.449±0.195	0.304±0.128
显著性	0.000***	0.000***	0.000***	0.000***	0.001**	0.752

注: *显著性<0.05, **显著性<0.01, ***显著性<0.001

表3 AoI 相关特征分类结果

Table 3 Classification results for AoI-related features

模型		Accuracy (%)	Precision (%)	Recall (%)	F1 (%)
SVM	视频级	62.01±3.45	61.98±3.76	68.26±4.45	64.86±3.15
	决策融合	75.94±7.72	75.66±7.28	87.78±7.44	81.07±5.94
逻辑线性回归	视频级	60.84±3.12	62.68±3.99	59.66±5.85	60.94±3.68
	决策融合	74.53±3.60	78.23±4.95	79.56±10.56	78.20±4.57
K 近邻	视频级	58.95±0.48	60.12±1.18	60.27±4.45	60.08±1.84
	决策融合	71.07±8.65	72.88±4.52	79.78±13.99	75.85±8.51
随机森林	视频级	60.93±1.74	61.75±2.65	63.60±3.19	62.58±1.86
	决策融合	75.94±7.72	77.32±6.65	83.78±10.08	80.16±6.72

表4 眼动追踪图像分类结果

Table 4 Eye Tracking image classification results

注意力：显著性		Accuracy (%)	Precision (%)	Recall (%)	F1 (%)
5:5	视频级	60.24±1.54	61.61±1.90	60.97±3.66	61.20±1.83
	决策融合	77.20±8.40	85.84±7.74	73.33±11.50	78.62±8.62
6:4	视频级	62.51±2.04	63.13±2.52	66.07±3.69	64.46±1.85
	决策融合	79.71±11.63	85.83±8.89	77.56±13.23	81.29±10.93
7:3	视频级	60.21±1.52	61.13±1.24	62.92±5.07	61.88±2.32
	决策融合	78.13±4.26	84.51±3.06	77.33±11.39	80.13±5.46
8:2	视频级	57.54±1.27	59.57±1.83	55.75±7.80	57.22±3.92
	决策融合	70.05±11.68	97.78±4.44	51.11±23.17	63.67±17.66

表5 3种眼动特征工程方法的模型分类结果

Table 5 Model classification results of three eye movement feature engineering methods

特征工程		Accuracy (%)	Precision (%)	Recall (%)	F1 (%)
AoI 特征	视频级	62.01±3.45	61.98±3.76	68.26±4.45	64.86±3.15
	决策融合	75.94±7.72	75.66±7.28	87.78±7.44	81.07±5.94
Cilia 等 ^[9]	视频级	62.19±2.01	65.27±2.94	57.19±5.14	60.81±3.04
	决策融合	78.13±6.77	86.28±8.52	75.33±9.75	79.96±6.70
眼动追踪图像	视频级	62.51±2.04	63.13±2.52	66.07±3.69	64.46±1.85
	决策融合	79.71±11.63	85.83±8.89	77.56±13.23	81.29±10.93

算法相比，本文方法的准确率提高了 1.58%，F1 值提高了 1.33%。

3 种方法的受试者工作特征 (receiver operating characteristic, ROC) 曲线如图 6 所示，本文方法的 ROC 曲线下方面积最大，表明本文方案更适用于 ASD 儿童早期筛查。

本实验进一步计算了本文模型在不同年龄儿

童中的准确率，并进行比较分析，对比结果如表 6 所示。由表 6 可知，本文模型在 4~6 岁儿童中表现良好，而在 3 岁儿童中表现较差，可能由于 4 岁以下儿童对情绪的感知不够强烈，儿童解读面部情绪的能力随着年龄的增长而提高^[33]，因此，3 岁的 ASD 儿童与 TD 儿童的情绪感知眼动模式差异性较小。

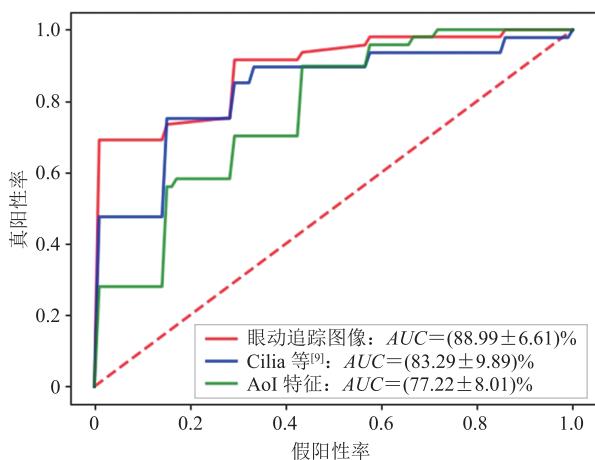


图 6 3 种眼动特征工程方法的 ROC 曲线图

Fig. 6 ROC curves of three eye movement feature engineering methods

表 6 所提出的模型在不同年龄儿童中的准确率

Table 6 Accuracy of the proposed model on children of different ages

年龄(岁)	人数	准确率(%)
3	10	63.33 ± 33.88
4	33	86.35 ± 11.72
5	32	82.43 ± 9.14
6	7	75.00 ± 25.00

4 讨论与分析

ASD 作为儿童严重的发育行为障碍, 逐渐升高的患病率使得其已成为社会公共卫生问题, 也成为医学和社会的一个焦点话题。充分的研究证据表明, 早期发现、早期诊断、早期干预可显著改善 ASD 不良预后, 减轻家庭及社会负担。因此, 精准、智能、便捷的孤独症早期诊断方法, 已成为国计民生拟解决的重大问题。基于认知心理范式和眼动轨迹特征的分析, 可帮助人们更好地、更精确地洞察 ASD 情绪识别过程中特定的注视行为调节缺陷机制, 为诊断 ASD、评估治疗结果、预测情绪识别或社交技能的临床指标提供新方法、新工具。本研究提出了一种基于自然情绪感知的动态眼动轨迹建模方法, 在 ASD 分类

任务中准确率达 79.71%, 成为一种潜在的 ASD 筛查方法。

前期研究发现, 采用自然、动态场景刺激, 可有效诱发 ASD 儿童的注视异常模式^[17], 本文研究结果与该观点相符。通过 5 个 AoI 区域的 18 维眼动特征的差异分析发现, ASD 组的总注视次数、注视时长显著低于 TD 组, 这表明与 TD 组相比, ASD 组儿童在自然情绪感知状态下有明显的注视缺失。此外, ASD 组注视眼睛的次数和时长显著低于 TD 组, 而注视脸部其他区域的次数和时长、注视视频其他区域的时长均显著高于 TD 组, 这表明在自然情绪感知状态下, ASD 组对眼睛区域有明显的回避, 更多地关注人脸与情绪表达无关的区域以及非人脸的背景区域。ASD 与 TD 儿童在自然情绪感知状态下的眼动模式具有明显差异, 这提示我们可以利用该差异对两类儿童进行区分。

本文使用自然情绪表达作为刺激材料, 证明了感知自然情绪眼动轨迹特征训练分类模型的有效性。本文提出一种眼动追踪可视化建模的方法, 可自动提取视频感兴趣区域, 并与注意力分布进行融合。首先, 利用显著性检测技术生成视频显著性, 并将视频每一帧的显著性图像进行压缩叠加, 从而为每个视频生成一张显著性图像, 以自动提取感兴趣区域; 然后, 计算注意力分布矩阵, 由于矩阵较为稀疏, 使用最大值滤波器进行滤波, 生成注意力分布图像; 最后, 将显著性图像与注意力分布图像进行加权叠加融合, 对比多组权重后发现, 当注意力分布图像与显著性图像权重的比值为 6 : 4 时, 效果最佳, 这表明显著性检测技术自动检测到的兴趣区域可以作为眼动轨迹的重要补充信息。本文提出的眼动追踪可视化建模方法, 可运用于不同的刺激场景, 包括图像以及其他类型的视频, 对于时长较长、场景较为复杂的视频, 可以将视频进行分段处理, 相应的眼动数据也可进行分段处理。利用在图像

分类问题中表现较好的经典残差网络进行迁移学习,从而训练一个特征提取器,提取特征后利用主成分分析方法进行特征选择,最后输入决策树模型进行二分类,得到眼动轨迹的判别结果。将本文提出的特征工程方法与传统的 AoI 特征提取方法以及 Cilia 等^[9]的方法进行比较,对比结果如表 5 所示。由表 5 可知,本文方法的准确率较高,在视频级的准确率达到了 62.51%,决策融合后的准确率达到了 79.71%,但本文模型的准确率低于 Wan 等^[10]的方法,可能是实验范式和 AoI 区域划分标准差异所致。

尽管本研究的预测结果较好,但仍然存在一些局限性:(1)研究对象相对较少,数据集规模相对较小,后续可扩大样本量,采用多中心收集数据,进一步测试模型的鲁棒性和泛化性。(2)ASD 儿童眼动轨迹的特征可从多个角度进行表征,如扫视路径、注意动态变化等,将眼动追踪数据和刺激材料融合可获取更多的信息。因此,后期可重点探讨特征表征方式和最优融合方式,将眼动轨迹各个维度的信息进行融合,尽可能获取眼动轨迹的高级特征,以有效提升模型筛查的准确率。(3)刺激材料的视频时长不一致,后续进行建模时需考虑该因素,以模糊视频时长不一致所导致的噪声影响。

5 结 论

本文提出一种基于自然情绪感知眼动轨迹的孤独症儿童筛查模型,对自然情绪感知场景下的眼动追踪进行可视化处理,自动提取感兴趣区域后与注意力分布进行融合,利用卷积神经网络提取图像特征,主成分分析方法进行特征选择,使用决策树模型进行分类,并基于分类结果进行决策融合,以实现对 ASD 儿童的自动筛查,准确率为 79.71%。实验结果表明,在自然情绪感知状态下的眼动跟踪,有望成为辅助 ASD 诊断的

客观工具。

参 考 文 献

- [1] Maenner MJ, Shaw KA, Bakian AV, et al. Prevalence and characteristics of autism spectrum disorder among children aged 8 years—autism and developmental disabilities monitoring network, 11 sites, United States, 2018 [J]. MMWR Surveillance Summaries, 2021, 70(11): 1-16.
- [2] American Psychiatric Association. Diagnostic and statistical manual of mental disorders: DSM-5 [M]. Frisco Center: American Psychiatric Association, 2013: 50-59.
- [3] Kamp-Becker I, Ghahreman M, Heinzel-Gutenbrunner M, et al. Evaluation of the revised algorithm of autism diagnostic observation schedule (ADOS) in the diagnostic investigation of high-functioning children and adolescents with autism spectrum disorders [J]. Autism, 2013, 17(1): 87-102.
- [4] Tariq Q, Daniels J, Schwartz JN, et al. Mobile detection of autism through machine learning on home video: a development and prospective validation study [J]. PLoS Medicine, 2018, 15(11): e1002705.
- [5] Vyas K, Ma R, Rezaei B, et al. Recognition of atypical behavior in autism diagnosis from video using pose estimation over time [C] // Proceedings of the 2019 IEEE 29th International Workshop on Machine Learning for Signal Processing, 2019: 1-6.
- [6] Eni M, Dinstein I, Ilan M, et al. Estimating autism severity in young children from speech signals using a deep neural network [J]. IEEE Access, 2020, 8: 139489-139500.
- [7] Baygin M, Dogan S, Tuncer T, et al. Automated ASD detection using hybrid deep lightweight features extracted from EEG signals [J]. Computers in Biology and Medicine, 2021, 134: 104548.
- [8] Haputhanthri D, Brihadiswaran G, Gunathilaka S,

- et al. An EEG based channel optimized classification approach for autism spectrum disorder [C] // Proceedings of the 2019 Moratuwa Engineering Research Conference, 2019: 123-128.
- [9] Cilia F, Carette R, Elbattah M, et al. Computer-aided screening of autism spectrum disorder: eye-tracking study using data visualization and deep learning [J]. JMIR Human Factors, 2021, 8(4): e27706.
- [10] Wan GB, Kong XJ, Sun BB, et al. Applying eye tracking to identify autism spectrum disorder in children [J]. Journal of Autism and Developmental Disorders, 2019, 49(1): 209-215.
- [11] Xia C, Chen KX, Li K, et al. Identification of autism spectrum disorder via an eye-tracking based representation learning model [C] // Proceedings of the 2020 7th International Conference on Bioinformatics Research and Applications, 2020: 59-65.
- [12] Xie J, Wang LF, Webster P, et al. A two-stream end-to-end deep learning network for recognizing atypical visual attention in autism spectrum disorder [Z/OL]. arXiv Preprint, arXiv: 1911.11393, 2019.
- [13] Fridenson-Hayo S, Berggren S, Lassalle A, et al. Basic and complex emotion recognition in children with autism: cross-cultural findings [J]. Molecular Autism, 2016, 7(1): 1-11.
- [14] Jiang M, Francis SM, Srishyla D, et al. Classifying individuals with ASD through facial emotion recognition and eye-tracking [C] // Proceedings of the 2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society, 2019: 6063-6068.
- [15] Bochet A, Franchini M, Kojovic N, et al. Emotional vs. neutral face exploration and habituation: an eye-tracking study of preschoolers with autism spectrum disorders [J]. Frontiers in Psychiatry, 2021, 11: 568997.
- [16] Su Q, Chen F, Li HF, et al. Multimodal emotion perception in children with autism spectrum disorder by eye tracking study [C] // Proceedings of the 2018 IEEE-EMBS Conference on Biomedical Engineering and Sciences, 2018: 382-387.
- [17] He YY, Su Q, Wang L, et al. The characteristics of intelligence profile and eye gaze in facial emotion recognition in mild and moderate preschoolers with autism spectrum disorder [J]. Frontiers in Psychiatry, 2019, 10: 402.
- [18] Eisenbarth H, Alpers GW. Happy mouth and sad eyes: scanning emotional facial expressions [J]. Emotion, 2011, 11(4): 860-865.
- [19] Tao YD, Shyu ML. SP-ASDNet: CNN-LSTM based ASD classification model using observer scanpaths [C] // Proceedings of the 2019 IEEE International Conference on Multimedia & Expo Workshops, 2019: 641-646.
- [20] Spezio ML, Adolphs R, Hurley RSE, et al. Abnormal use of facial information in high-functioning autism [J]. Journal of Autism and Developmental Disorders, 2007, 37(5): 929-939.
- [21] de Wit TC, Falck-Ytter T, von Hofsten C. Young children with autism spectrum disorder look differently at positive versus negative emotional faces [J]. Research in Autism Spectrum Disorders, 2008, 2(4): 651-659.
- [22] 韩俊霞, 康健楠, 欧阳高翔, 等. 孤独症儿童脑电与眼动追踪研究 [J]. 科学通报, 2018, 63(15): 52-61.
- Han JX, Kang JN, Ouyang GX, et al. A study of EEG and eye tracking in children with autism [J]. Chinese Science Bulletin, 2018, 63(15): 52-61.
- [23] Lawson RP, Mathys C, Rees G. Adults with autism overestimate the volatility of the sensory environment [J]. Nature Neuroscience, 2017, 20(9): 1293-1299.
- [24] Loth E, Garrido L, Ahmad J, et al. Facial expression recognition as a candidate marker for autism spectrum disorder: how frequent and severe are deficits? [J]. Molecular Autism, 2018, 9(1): 1-11.

- [25] Tso WWY, Wong VCN, Xia X, et al. The griffiths development scales-Chinese (GDS-C): a cross-cultural comparison of developmental trajectories between Chinese and British children [J]. *Child: Care, Health and Development*, 2018, 44(3): 378-383.
- [26] Bao W, Li Y, Gu ML, et al. Building a Chinese natural emotional audio-visual database [C] // Proceedings of the 2014 12th International Conference on Signal Processing, 2014: 583-587.
- [27] Li Y, Tao JH, Schuller B, et al. MEC 2016: the multimodal emotion recognition challenge of CCPR 2016 [C] // Proceedings of the Chinese Conference on Pattern Recognition, 2016: 667-678.
- [28] Baltrušaitis T, Zadeh A, Lim YC, et al. OpenFace 2.0: facial behavior analysis toolkit [C] // Proceedings of the 2018 13th IEEE International Conference on Automatic Face & Gesture Recognition, 2018: 59-66.
- [29] Cong RM, Lei JJ, Fu HZ, et al. Review of visual saliency detection with comprehensive information [J]. *IEEE Transactions on Circuits and Systems for Video Technology*, 2018, 29(10): 2941-2959.
- [30] Cong RM, Lei JJ, Fu HZ, et al. Video saliency detection via sparsity-based reconstruction and propagation [J]. *IEEE Transactions on Image Processing*, 2019, 28(10): 4819-4831.
- [31] He KM, Zhang XY, Ren SQ, et al. Deep residual learning for image recognition [C] // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016: 770-778.
- [32] Deng J, Dong W, Socher R, et al. ImageNet: a large-scale hierarchical image database [C] // Proceedings of the 2009 IEEE Conference on Computer Vision and Pattern Recognition, 2009: 248-255.
- [33] Boyatzis CJ, Chazan E, Ting CZ. Preschool children's decoding of facial emotions [J]. *The Journal of Genetic Psychology*, 1993, 154(3): 375-382.