

# 自闭症儿童视觉特征数据库的构建与应用

冉瑞生 邓姗 刘宸佚 梁伟 樊欣

**【摘要】**目的：由于儿童自闭症的隐私性，其数据集往往难以获取，这给相关的研究造成了一定的困难。为此，构建了自闭症儿童特征数据库。方法：首先设计交互场景，采集了一个自闭症儿童面部数据集，然后利用计算机视觉中的面部关键点检测、视线跟踪、头部姿态和面部动作单元等技术，提取自闭症儿童面部视觉特征，然后在提取的面部特征的基础上对所有特征数据进行编码，从而构建了一个自闭症儿童视觉特征数据库。结果：该视觉特征数据库可供该领域的研究人员使用。此外，基于该视觉特征库，本文以面部动作单元为例对自闭症儿童进行分析和预测，实验结果表明，分类精度最高可达 93%。结论：该数据库可用于自闭症儿童的相关研究，具有较大的应用价值。

**【关键词】** 自闭症；计算机视觉；特征提取；数据库；临床辅助诊断

Doi:10.3969/j.issn.1673-7571.2024.04.018

**【中图分类号】** TP391；R319

## Construction and application of visual characteristics database for autistic children

RAN Ruisheng, DENG Shan, LIU Chenyi, LIANG Wei, FAN Xin. School of Computer and Information Science, Chongqing Normal University, Chongqing 401331, China (RAN Ruisheng, DENG Shan, LIU Chenyi, LIANG Wei); Department of Pediatrics, Chongqing Maternal and Child Health Hospital (FAN Xin)

Corresponding author: RAN Ruisheng, Email: rshran@cqnu.edu.cn

**【Abstract】 Objective** Due to the privacy of children with autism, their datasets are often difficult to obtain, which creates certain difficulties for related research. For this reason, this paper constructs a database of autistic children's characteristics. **Methods** Firstly, an interaction scene was designed and a facial dataset of children with autism was collected, and then facial visual features of children with autism were extracted using techniques of facial key point detection, line-of-sight tracking, head posture and facial action units in computer vision, and then all the feature data were encoded on the basis of the extracted facial features, so as to construct a database of visual features of children with autism. **Results** This visual feature database can be used by researchers in this field. In addition, based on this visual feature database, this paper analyzed and predicted children with autism using facial action units as an example, and the experimental results showed that the classification accuracy was up to 93%. **Conclusion** This database can be used in the research related to autistic children, and has a greater application value.

**【Keywords】** Autism; Computer vision; Feature extraction; Database; Clinical aided diagnosis

自闭症谱系障碍，简称自闭症，儿童发育早期的神经发育障碍疾病。自闭症也被称为孤独症（autism spectrum disorder, ASD），是一种常发于儿童发育早期的神经发育障碍疾病。自闭症的典型症状主要有 3 个，即社交障碍、言语障碍和行为异常，而社交障碍是自闭症的核心障碍<sup>[1]</sup>。2005 年，Zwaigenbaum 等<sup>[2]</sup>研究观察了大量自闭症儿童的数据

基金项目：人工智能驱动下的自闭症儿童社会情绪评估研究（20YJAZH084）

作者单位：401331 重庆，重庆师范大学计算机与信息科学学院（冉瑞生、邓姗、刘宸佚、梁伟）；重庆市妇幼保健院儿保科（樊欣）

通信作者：冉瑞生，Email: rshran@cqnu.edu.cn

记录, 结果表明, 基于儿童的视频, 可通过特定的面部行为特征, 如视线专注、呼叫反应、面部表情等, 区分疑似自闭症儿童与正常儿童。同时, 采用人工智能对自闭症儿童进行诊断评估的研究逐渐成为新的研究方向<sup>[3-9]</sup>, 且在近年来逐渐成为一个行之有效的诊断评估方法。

基于人工智能的自闭症儿童诊断评估需要自闭症儿童的数据库作为基础。国外在自闭症方面的研究相对国内起步较早<sup>[10-13]</sup>, 有比较完备的研究环境和数据采集条件。Hashemi 等<sup>[14]</sup>为了量化自闭症风险行为, 采集了自闭症儿童数据库。他们开发了一个独立的、封闭的移动应用程序来引发儿童的注意力和特定的行为与社交反应, 硬件采用 iPad 自带的摄像头进行视频影像记录。英国诺丁汉大学的 Jaiswal 等<sup>[15]</sup>针对注意力缺陷多动症障碍 (attention deficit hyperactivity disorder) 和自闭症, 以计算机为媒介进行问卷调查, 用现代 RGBD (颜色 + 深度) 摄像机记录受试者并制作了数据集 KOMAA (kinect data for objective measurement of ADHD and ASD)。该数据集包含 55 条数据, 所有的受试者都是 18 岁以上的成年人。国内的相关研究也都是采用自制数据集的方案, 由于隐私等问题限制, 相关数据图像都不能进行公开。陈靓影等<sup>[16]</sup>在验证人机交互游戏评价可定性和有效性的过程中通过人机游戏交互的方式采集儿童数据, 数据采集在武汉市某自闭症康复中心进行, 所有受试者为 4 ~ 6 岁, 数据集共计包

含 18 名被试儿童。该采集方案采用非计算机视觉的方法, 采集的前期准备也较复杂。

从文献报道来看, 许多团队不具备采集条件, 或者需为数据采集付出较大的代价, 虽然目前已有一些团队制备了数据库, 但是这些数据库尚存在一定缺陷。主要表现为: 有些团队在采集时需要有相应的高端仪器做支撑, 如 Kinect 深度相机, 成本较高; 有些团队在采集时需要与人有较多的交互, 对于相关研究的支持意义不大。更为主要的是, 自闭症儿童是一个敏感的群体, 由于隐私性, 采集自闭症儿童的数据较为困难。即便可以采集儿童的数据, 其采集的数据一般不予公开。这就造成每个研究团队在开展相关研究时, 都需要采集自闭症儿童的数据库, 从而产生了大量重复且非标准化的工作。因此, 迫切需要一种方式, 既能保护自闭症儿童的隐私, 又能方便研究人员开展自闭症儿童的诊断评估研究。

本文针对自闭症儿童的数据库

问题, 与重庆市妇幼保健院合作, 从数据库的采集、儿童的特征分析与提取方面开展研究, 构建了一个标准化的自闭症儿童视觉特征库, 为该领域的研究人员提供了一个公开的数据集。本文首先设计了一个社会交互场景和数据采集流程, 并以视频的方式获取儿童在社会交互场景下的面部感官信息 (无需高端设备), 获取了一个儿童面部视频数据集。社交障碍是自闭症患者的核心障碍, 主要表现为他们会避免眼神接触, 对父母的呼唤声充耳不闻, 面部表情缺失等<sup>[17]</sup>。基于此, 在采集的自闭症儿童视频数据集基础上, 本文采用了计算机视觉中头部姿态估计、视线跟踪和人脸表情分析等方法, 分析自闭症儿童面部占比率、唤名反应、头部累积移动距离、头部移动速度等特征, 构建了自闭症儿童视觉特征数据库。基于该特征库, 本文以面部动作单元为例对自闭症儿童进行分析和预测, 实验结果表明其分类精度最高可达 93%, 具有较大的应用价值。研究流程见图 1。

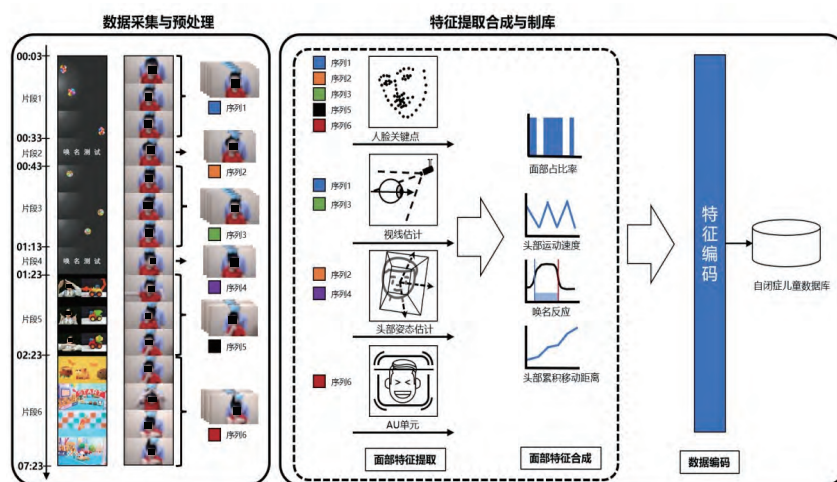


图 1 研究流程

## 1 视频数据采集

### 1.1 数据采集前期准备

**1.1.1 实验视频素材设计** 本文旨在采用计算机视觉方法对自闭症儿童的面部视觉特征进行提取,从而形成自闭症儿童面部特征数据。为获取儿童面对各种社会交互场景时的面部感官信息,首先需要设计合适的社会交互场景。为此,我们采用了有代表性的视频片段、故事情节以及对儿童姓名的呼唤作为交互场景,并制作成一个时长 7 分 23 秒的视频作为实验素材。本文采用的视频实验素材,是与自闭症领域的专业医生共同探讨,经专题研究后,设计并制作的与自闭症疾病特征相关素材。本文设计并制作的实验素材包括 6 个视频片段。

第一片段,时长为 30 秒。该片段是一组气球从屏幕底部均匀分布的 5 个位置先后上升至顶部。第二片段为唤名测试,时长为 10 秒。在屏幕提示进入唤名测试环节后,实验人员于儿童侧方呼叫儿童姓名,并记录唤名时间点,受试儿童将转头以回应实验人员的唤名。第三片段是来回跳动的皮球,时长为 30 秒。皮球从屏幕左侧移动至右侧再回到左侧,往返两次。第四片段为第二次唤名测试,时长为 10 秒。该片段内容和操作方式与第二片段相同。第五片段为左动右静的影片展示,时长为 60 秒。第六片段是两段动画展示,时长为 5 分钟。在专业医生的指导下选择合适的动画内容,旨在让儿童产生观影兴趣以及表现出高兴的情绪。设计和制作的实验素材见图 2。

**1.1.2 数据采集软硬件设计** 为了采集儿童的视频数据,本研究需要两款软件:视频音频采集软件和数据采集软件。

视频音频采集软件使用某公司的音视频串流软件 Logitech Capture。该软件界面自带白天、黑夜模式,支持调节自动对焦、自动曝光、图像设置等,该软件仅工作在后台,从而避免儿童看到自己的实时图像而产生抗拒心理。

数据采集软件独立开发,包含以下几个功能:视频素材文件的播放及预览;采集软件与播放素材的同步控制;儿童的基本信息记录及保存。该软件可以方便地管理采集到的视频数据。数据采集软件见图 3。

采集数据的视频摄像头为笔记本自带的高清摄像头。由于要采集多个儿童,因此我们需要对每一位儿童进行编号,并且记录儿童姓名或昵称,用于姓名呼叫刺激。开始

前,可以通过软件提供的文本框进行记录,并在视频采集流程结束时,自动生成相关信息的 txt 文档并保存在视频保存的同一路径目录下。

### 1.2 视频采集过程

准备一台带有 1080p 高清摄像头和麦克风的笔记本电脑,以及一套桌椅,并选定一个安静、舒适的实验环境。实验人员在测试过程中均站至受试儿童的左右两侧。在数据采集时,首先将笔记本电脑放在办公桌上,将椅子放在电脑屏幕前约 50 cm 处;再让受试儿童单独或由家长陪同坐在笔记本屏幕前;然后播放实验素材让儿童观看,同时用笔记本电脑上的高清摄像头记录儿童观看实验素材时的面部感官信息,儿童观看结束后制作成视频。

实验人员在本地妇幼保健院采集了自闭症儿童的视频。采集之前,医生为每位受试儿童进行自闭症筛查,首先让儿童接受发育诊断量表 (gesell) 评估<sup>[18]</sup>。然后,根



图 2 实验视频素材



图 3 数据采集软件示意



据量表和临床经验,如果疑是自闭症,则通过自闭症诊断观察量表(ADOS)对儿童进行进一步评估<sup>[19]</sup>。最后根据儿童发育情况和其他诊断测试方法,综合上述量表的评估结果做出最终诊断。正常儿童数据采集自本地幼儿园。经医生观察诊断,受试儿童均为没有语言障碍、行为障碍以及其他自闭症症状的正常儿童。

### 1.3 儿童面部视频库

共采集 62 名儿童的面部视频数据,其中包含 42 名自闭症儿童(ASD)、20 名正常儿童(typical development, TD)。已有研究表明,男孩的自闭症发病率是女孩的 4 倍,所以数据采集时,ASD 组中男性儿童占比较大,在数据集性别上二者存在统计学差异。视频库中自闭症的年龄在 20~60 个月之间,正常儿童的年龄在 24~60 个月之间,两组受试儿童在年龄无统计学差异。我们对录制的儿童面部视频进行了处理,包括筛除未完成的实验和采集流程不规范的视频、对视频中家长的面部进行遮挡打码处理等。用一个表格记录了儿童编号、年龄、性别、两次唤名时间、诊断结果等必要信息。构建的儿童面部视频库的分布见表 1。

## 2 特征提取及编码

自闭症患者的核心障碍是社交障碍,主要表现为他们会避免眼神接触,面部表情缺失,不能正确表达内心感受,对父母的呼唤充耳不闻等症状<sup>[20-21]</sup>。基于此,本文将从

儿童的视线跟踪、头部姿态估计、唤名反应以及面部情绪表达 4 个方面对自闭症儿童和正常儿童的面部图像进行特征提取。本文借助传统机器学习和深度学习网络模型<sup>[22]</sup>提取了人脸关键点标记、头部姿态估计、视线跟踪和人脸表情 4 个面部特征,对提取的特征数据进行编码,形成自闭症儿童视觉特征数据集。特征提取示意图见图 4。

### 2.1 人脸关键点信息

**2.1.1 人脸关键点** 人脸关键点共计 68 个,本研究采用卷积专家约束局部模型(convolutional experts constrained local model, CEN)计算响应图以及标定人脸关键点位置,结合点分布模型(point distribution models, PDM)视线形状参数更新<sup>[23-24]</sup>。由于面部器官分布位置相对稳定,故面部结构关键点的位置产生相对约束关系,通过评估各个像素位置的标记点对齐概率,对每个关键点进行准确定位。本文使用笛卡尔坐标系表示每个点,即每个点有  $X$ 、 $Y$ 、 $Z$  三个坐标值。68 个人脸关键点用  $X_0, \dots, X_{67}$ ,  $Y_0, \dots, Y_{67}$ ,  $Z_0, \dots, Z_{67}$  表示,共

204 维。

**2.1.2 面部占比率** 面部占比率是受试者在观看实验视频素材时,其面部完整出现在电脑屏幕的时长占视频总时长的比值。该行为特征反应了受试者是否全程观看实验素材,从而反应了儿童的专注度。

我们用计算机视觉中的人脸检测算法来检测每一帧视频中的儿童人脸,并用  $\text{conf}(f_i)$  表示第  $i$  帧的人脸检测可信度,取值范围在 0 到 1 之间。实验中,我们设定阈值  $\varepsilon = 0.77$ ,如果可信度  $\text{conf}(f_i)$  大于该阈值认为当前帧出现人脸,否则视为没有出现人脸。整个视频中出现人脸的帧数除以视频的总帧数即为面部出现占比率  $R_{\text{face}}$ ,其中每一帧的面部占比率是从开头到当前帧为止,公式表示如下:

$$R_{\text{face}} = \frac{\text{card}(\{\text{conf}(f_i) \geq \varepsilon, f_i \in F\})}{\text{card}(F)} \quad (1)$$

其中,  $R_{\text{face}}$  值越大说明儿童面部出现时间越长,专注度越高;  $R_{\text{face}}$  值越小说明儿童的专注度越低。实验中随机抽取了一个 ASD 儿童和一个 TD 儿童,分析其人脸检测可信度和  $R_{\text{face}}$  值,见图 5。由

表 1 儿童视频库分布情况

信息	人数 / %		年龄 / 岁	
	男	女	男	女
ASD	39 (92.8)	3 (7.2)	30	31
TD	7 (35.0)	13 (65.0)	47	42

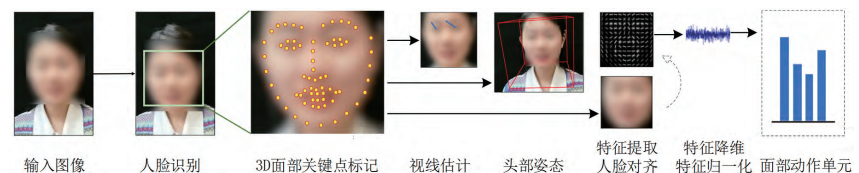


图 4 特征提取示意

图5可知, ASD儿童的 $R_{face} = 0.75$ , TD儿童的 $R_{face} = 0.99$ , 说明ASD儿童的专注度明显比TD儿童的专注度差。

## 2.2 头部姿态信息

**2.2.1 头部位置** 头部位置信息使用CE-CLM模型提取<sup>[25]</sup>得到的面部关键点, 该关键点使用三维坐标表示。以正视前方的头部姿态为基准向量, 以此得到头部姿态在三维空间中的向量坐标和偏转角度。本文使用的是头部相对于相机的位置和头部围绕X、Y、Z轴旋转的弧度。pose\_Tx, pose\_Ty, pose\_Tz表示头部相对于相机的位置, 单位是毫米(远离相机的方向是正)。pose\_Rx, pose\_Ry, pose\_Rz表示头部围绕X、Y、Z轴旋转的弧度, 分别对应俯仰角、偏航角、翻滚角, 坐标系是以相机为原点的世界坐标系。位置和弧度信息共6维。

**2.2.2 头部累积移动距离** 头部累积转动距离是指受试者在唤名测试阶段, 每一帧头部位置相较于前一帧头部位置的位移距离之和。如果头部转向越大, 头部累积转动距离也将越大。自闭症的一个常见表现是多动<sup>[26]</sup>, 这种多动不仅仅体现在速度快, 还常体现于运动距离长。ASD儿童往往比TD儿童表现出来更大幅度的运动行为, 因此可以用头部累积转动距离作为自闭症儿童的特征。

本文在交互场景中的唤名测试阶段计算了儿童的头部累积转动距离。首先在由CE-CLM算法标记的人脸关键点中选取内眼角的两个点及鼻子上的4个点。这些稳定的

面部关键点的位置对面部表情的变化不明显, 因此适合作为头部运动的参数即面部质心, 通过计算质心的位置变化来估计头部的运动。记每一帧视频中面部标记点的质心 $C_i$ , 将前后两帧质心 $C_i$ 的位移求和作为累计转动距离, 将累计转动距离记为特征pose\_dist, 如式(2):

$$pose\_dist = \sum_{i=1}^n \|C_i - C_{i-1}\| \quad (2)$$

其中pose\_dist是截止第*i*帧的累积头部移动距离, *n*为某个视频片段的所有帧数, *i*表示当前帧数。计算的累积头部移动距离相关参数的单位是毫米。实验中随机抽取了一个ASD儿童的在唤名反应时记录的视频, 从视频的200~2000帧为例, 其累积移动距离如图6所示。

### 2.2.3 头部运动速度

头部运动是指儿童整个头部的运动情况。头动的变化, 可以客

观地反映出儿童在整个过程中的稳定、安定情况。如果头动比较频繁、高速且幅度大, 则可以认为该儿童在采集过程中是难以专注的。为了研究儿童的头部运动, 计算了视频中每一帧的头部运动速度。头部转动速度是指受试者在唤名测试阶段, 每一帧头部转动的速度。头部运动动力学是心理障碍领域的一个研究方向, 在特定的社会条件刺激下, 自闭症儿童头部偏航和倾斜速度会更快, 与正常儿童有明显差异<sup>[27]</sup>。本文在交互场景中的唤名测试阶段计算了儿童的头部转动速度。我们先采用前面一节计算的任意帧*i*的头部转动距离, 然后计算该帧的头部移动速度pose\_speed, 如式(3)所示。

$$pose\_speed = \sum_{i=1}^{20} \|C_i - C_{i-1}\| \times \frac{f}{20} \quad (3)$$

其中*f*是录制视频的帧率

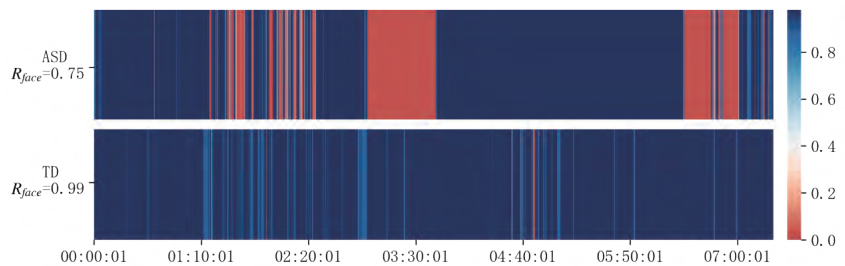


图5 人脸检测可信度变化情况

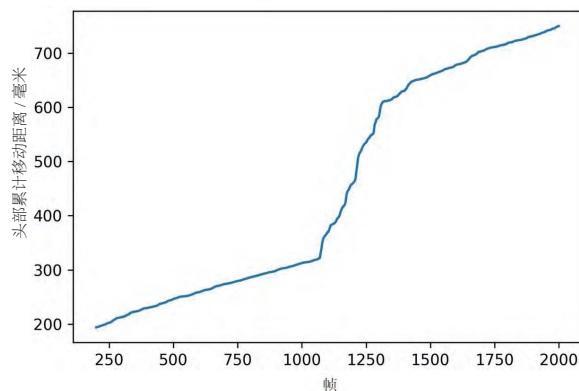


图6 ASD儿童头部累积移动距离

(1 秒内所记录的帧数量),  $\frac{1}{f}$  为单帧所耗费的时间, 为了让速度估计更可靠且不受帧率波动的影响, 通过计算 20 个连续帧的移动平均值来平滑估计速度, 以随机抽取的一个 ASD 儿童为例, 其视频数据 200 ~ 2 000 帧的头部运动速度情况如图 7 所示。

## 2.3 人眼关键点及视线信息

**2.3.1 人眼关键点** 采用约束局部神经网络模型 (constrained local neural fields, CLNF) 来检测眼部关键点<sup>[28]</sup>, 人眼关键点是分布在双眼上的共计 56 个关键点, 包含了眼睑、虹膜、瞳孔 3 个位置。本文使用 3D 坐标来标记人眼关键点, 由 eye\_lmk\_X\_0,... eye\_lmk\_X55, eye\_lmk\_Y\_0,...,eye\_lmk\_Y\_55, eye\_lmk\_Z\_0,...,eye\_lmk\_Z\_55 来表示。

**2.3.2 人眼视线** 为了提取视线信息, 本文建立以相机中心为原点的三维坐标系, 通过眼部关键点建立 3D 眼球模型, 计算眼球球体中心坐标轴和瞳孔中心坐标, 从眼球中心到瞳孔中心的三维向量即为眼部注视向量<sup>[29]</sup>。以三维空间坐标

系中平视正前方的向量作为基准向量, 得到在相机三维坐标系下的人眼视线向量。视线信息分为双眼各自的视线方向向量和眼睛在世界坐标系中的双眼平均的注视方向向量。本文使用 gaze\_0\_x, gaze\_0\_y, gaze\_0\_z 和 gaze\_1\_x, gaze\_1\_y, gaze\_1\_z 分别表示人眼视线在直接坐标系中眼睛 0 (图像中的左眼) 和眼睛 1 的视线方向向量。gaze\_angle\_x, gaze\_angle\_y 表示在坐标系中以弧度为单位的眼睛注视方向 (取双眼的平均值), 如果受试者左右看, 会导致 gaze\_angle\_x 的变化 (从正到负), 如果受试者上下看会导致 gaze\_angle\_y 的变化, 如果双眼都直直地盯着前方, 这两个数值都会接近于 0。

人眼关键点共计 168 维数据, 加上人眼视线的 8 维, 本部分共计 176 维数据。

## 2.4 面部动作单元

面部表情可以表达出一个人的情绪和心理状态。人的情绪变化伴随着表情的变化, 比如开心的时候伴随着微笑, 生气的时候伴随着面部的狰狞等, 这种表情的变化在生

理结构上是由面部肌肉的运动组合而成的, 这些面部肌肉运动就是对应的面部动作单元<sup>[30]</sup>。因此, 有学者提出了面部动作编码系统 (facial action coding system, FACS), 在该系统中为每一个面部肌肉群标注了标签, 定义了 44 个面部动作单元 (facial action unit, AU)<sup>[31]</sup>, 简称 AU 单元 (action unit)。AU 单元将面部图像分成若干子区域并根据预先设定的图像进行尺寸对齐, 提取面部定向梯度直方图 (histogram of oriented gradient, HOG) 特征, 采用主成分分析法对 HOG 特征向量进行维度提取, 最终通过分类模型得到 AU 单元分类<sup>[32]</sup>。面部动作单元反应了儿童的表情和情绪情况。

本文提取了自闭症儿童相关的 AU 单元, 包括 AU01、AU02、AU04、AU05、AU06、AU07、AU07、AU10、AU12、AU14、AU15、AU17、AU20、AU23、AU25、AU26、AU28、AU45 共计 18 个 AU。分别从 AU 单元是否出现和 AU 单元强度来进行编码描述。AU 单元强度使用 AU01\_r, AU02\_r, AU04\_r, AU05\_r, AU06\_r, AU07\_r, AU09\_r, AU10\_r, AU12\_r, AU14\_r, AU15\_r, AU17\_r, AU20\_r, AU23\_r, AU25\_r, AU26\_r, AU45\_r 来表示, 范围是 0 ~ 1。AU 单元是否出现使用 AU01\_c, AU02\_c, AU04\_c, AU05\_c, AU06\_c, AU07\_c, AU09\_c, AU10\_c, AU12\_c, AU14\_c, AU15\_c, AU17\_c, AU20\_c, AU23\_c, AU25\_c, AU26\_c, AU28\_c, AU45\_c 表示, 本部分共计 35 维特征。

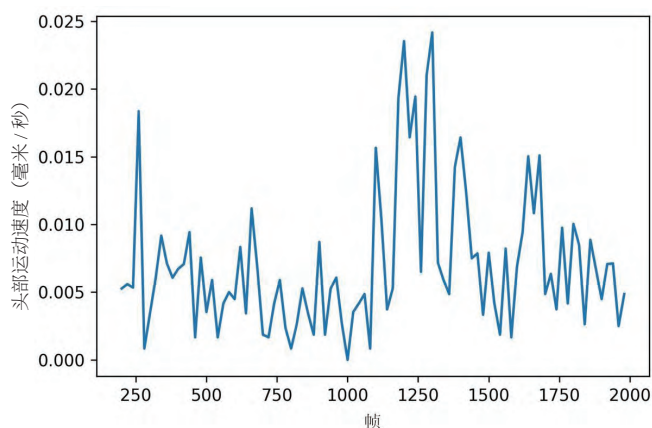


图 7 ASD 儿童头部运动情况



## 2.5 唤名反应时间

根据 Hashemi 等<sup>[33]</sup>的研究表明,自闭症儿童和普通儿童在听到他人唤名后的反应是不一样的。有的儿童能够在听到自己的名字后做出正常反应,如转头等,但更多的自闭症儿童对于唤名的反应感知并不强烈,有些甚至直接忽视他人的唤名。唤名反应时间是指在唤名测试阶段当实验人员呼唤受试者姓名时,受试者从听到姓名呼唤到转头以表回应所需的反应时间。该特征主要测试儿童对姓名的呼唤是否有反应,以及反应的时间长短。

实验人员在测试过程中均站至受试者左右两侧呼叫受试者姓名,同时记录下唤名时间点。从唤名时间点开始,录制 10 秒的视频片段。我们提取视频每一帧表示头部水平转动的头部姿态偏航角组成头部偏航角序列。计算头部偏航角序列的一阶导函数的第一个峰值对应的时间点,即为反应时间点。唤名开始时间点与反应时间点之间的差值即为儿童的唤名反应时间。

本文根据两段唤名测试视频片段计算了儿童的唤名反应时间,记为特征 name\_call,有两维。

## 2.6 自闭症儿童视觉特征数据库

将人脸关键点信息、头部姿态信息、人眼关键点及视线信息、面部动作单元、唤名反应时间特征编码进行汇总,我们构建了一个自闭症儿童视觉特征数据库,见表 2。由于自闭症儿童的隐私性,往往不容易获取自闭症儿童的真实数据,这给相关研究带来一定的困难。我们采用计算机视觉技术对自闭症儿

童的特征进行量化,从而抽取了儿童的视觉特征。这些视觉特征一方面能反应儿童的自闭特征,另一方面不涉及隐私性,可作为公开数据集,提供给广大研究者,具有较大的实用价值。

## 3 基于面部动作单元的应用

在构建完成自闭症儿童视觉特征数据库后,可利用其中的视觉特征对自闭症儿童进行分析或预测。本节以面部动作单元为例来对自闭症儿童的情绪进行分析,并使用机器学习模型实现了自闭症儿童的预测。

### 3.1 基于面部动作单元的情绪分析

因为自闭症儿童情绪表达与正常儿童的情绪表达有明显的差异,对自闭症儿童数据集的表情进行分析,能够获得自闭症儿童的情绪波动情况。为了更好地对自闭症儿童的情绪进行分析,我们首先基于面部动作单元(AU),建立了表情与情绪的映射关系,然后对比分析了自闭症儿童和正常儿童的情绪表达。

**3.1.1 表情与情绪的映射** 首先,根据 CK+ 数据集建立了 AU 单元与面部表情的关系。CK+ 数据集是一个人脸表情数据集,此数据集标注了人脸图像序列的表情和面部动作单元。因为人的基本表情分为生气、轻蔑、厌弃、害怕、开心、悲伤和惊讶共 7 种表情,CK+ 数据集参考面部动作编码系统(FACS),给出了 7 种基本表情与 AU 单元的关系<sup>[34]</sup>,如表 3 所示。

将上述 7 种基本表情加上中性表情,共计 8 种面部表情。为了从

面部表情得到人的情绪表达,将 8 种面部表情与 4 种主流情绪进行映射,见表 4。

**3.1.2 自闭症儿童情绪分析** 基于第 3 节所建立的儿童面部视频库,本文对儿童的视频每两秒统计一次面部动作单元,再根据表 3 将面部动作单元映射为面部表情,然后根据表 4 将面部表情映射成情绪。为了表达自闭症儿童和正常儿童的情绪差异,本文给出了两个情绪波动折线图,见图 8 和图 9,分别是自闭症儿童的情绪变化和正常儿童的情绪变化的折线图。

从图 8 和图 9 可看出:①自闭症儿童表情的变化频率很高,对应的情绪波动很明显;正常儿童的情绪变化相对稳定。②在能激发儿童高兴的表情的时间段上,正常儿童的表情反应明显很高兴,从而识别的情绪是开心,在最后讲故事的视频段上,情绪一直很平静,说明儿童在认真地看故事。③从儿童观看实验素材的视频中看到,自闭症儿童在采集数据的过程中,注意力不集中,经常产生各种表情;而正常儿童在整个过程中,注意力比较专注,对于视频中能激发儿童产生情绪的片段,也有相应的情绪反馈。这也印证了本文方法的正确性。

### 3.2 基于面部动作单元的自闭症儿童预测

通过对自闭症儿童和正常儿童的情绪波动实验结果的分析,两者之间存在着明显的情绪波动差异,其根本原因在于自闭症儿童的面部动作单元变化与正常儿童的面部动作单元变化存在差异。由此,本文

表 2 儿童视觉特征编码汇总

类别 (维度, 共 428 维)	字段	含义	备注
基本信息 (2)	frame	帧	从 0 开始的自然数
	success	是否检测到人脸	有为 1, 无为 0
人脸关键点 信息 (205)	X_0, ..., X_67, Y_0, ..., Y_67, Z_0,...,Z_67	人脸关键点坐标	3D 坐标
	R <sub>face</sub>	面部占比率	0 ~ 100, 单位: 百分比
头部姿态信息 (8)	pose_Tx, pose_Ty, pose_Tz	头部位置 (相对于相机)	远离相机的方向是正; 单位: 毫米
	pose_Rx, pose_Ry, pose_Rz	头部位置 (相对于坐标轴)	以相机为原点的世界坐标系, 分别对应俯仰角、偏航角、 翻滚角; 单位: 弧度
	pose_dist	头部累积移动距离	单位: 毫米
	pose_speed	头部移动速度	单位: 毫米 / 秒
	eye_lmk_X_0, eye_lmk_X_1,... eye_lmk_X55, eye_lmk_Y_0,... eye_lmk_Z_55	人眼关键点坐标	单位: 毫米
	gaze_0_x,gaze_0_y,gaze_0_z, gaze_1_x,gaze_1_y,gaze_1_z gaze_angle_x, gaze_angle_y	人眼视线方向向量  眼睛注视方向	字段中的 0 表示图像中的左眼, 1 表示右眼  双眼的平均值, 左右看导致前 者变化 (从), 上下看导致 后者变化; 单位: 弧度
面部动作单元 (35)	AU01_r, AU02_r, AU04_r, AU05_r, AU06_r, AU07_r, AU09_r, AU10_r, AU12_r, AU14_r, AU15_r, AU17_r, AU20_r, AU23_r, AU25_r, AU26_r, AU45_r	面部动作单元 (强度)	1 表示存在, 0 表示不存在
	AU01_c, AU02_c, AU04_c, AU05_c, AU06_c, AU07_c, AU09_c, AU10_c, AU12_c, AU14_c, AU15_c, AU17_c, AU20_c, AU23_c, AU25_c, AU26_c, AU28_c, AU45_c	面部动作单元 (是否 存在)	范围 0 ~ 1
唤名反应 (2)	name_call	唤名反应时间	有两次唤名测试

提取面部动作单元强度在时域和频域上的变化信息作为特征, 通过传统的支持向量机 (SVM)、随机森林及极端梯度提升法 (XGBoost) 进行分类, 实现在自闭症儿童数据集上进行分类, 通过训练得到预测模型。将原始的 AU 强度信息通过在时域和频域上提取特征之后, 这些特征在模型训练时并非都具有正

表 3 面部基本表情与 AU 单元的关系

表情类别	标准
厌恶	AU9 出现, 或者 AU10 出现
生气	AU23 和 AU24 必须同时出现
高兴	AU12 必须出现
悲伤	AU1+AU4+AU15 出现, 或者 AU6+AU15 出现
害怕	AU1+AU2+AU4 必须出现或者 AU5 的强度为 E 并且 AU4 不出现
惊讶	AU1+AU2 出现, 或者 AU5 出现
蔑视	AU14 必须出现



表 4 表情与情绪的映射

表情	情绪类别
中性	平静
高兴	开心
惊讶、恐惧	惊恐
悲伤、愤怒、蔑视、厌恶	消极

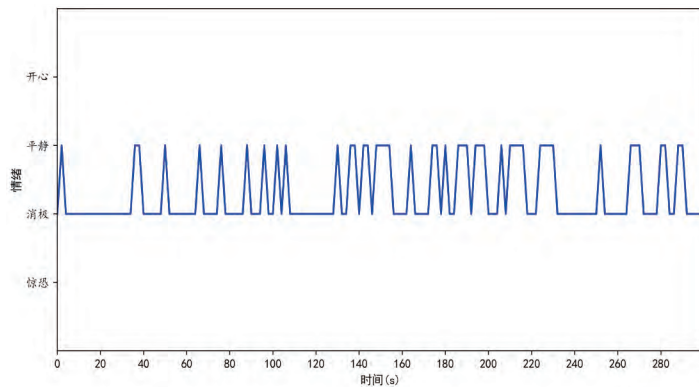


图 8 自闭症儿童情绪变化

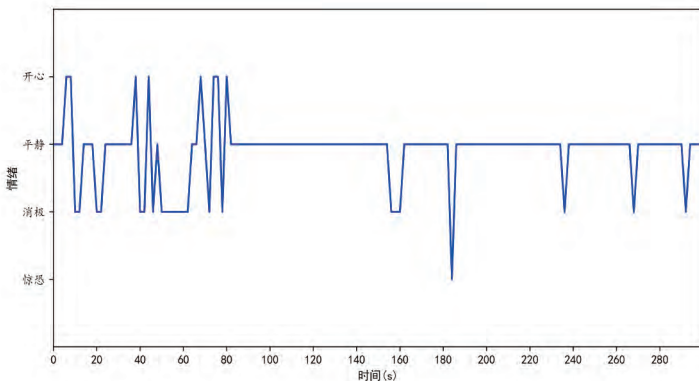


图 9 正常儿童情绪变化

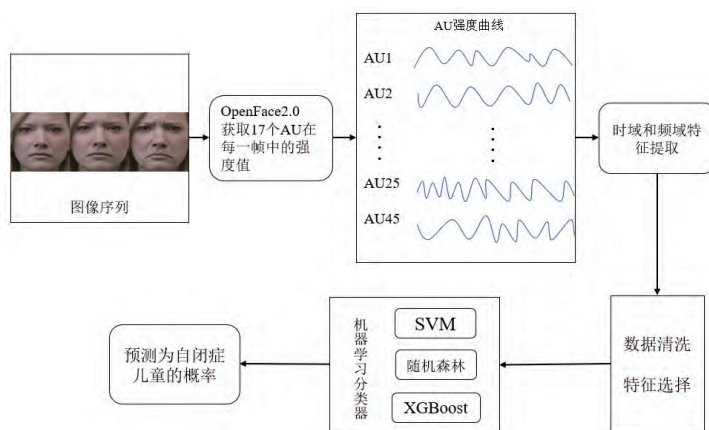


图 10 预测模型

向作用, 因此需要筛选有效的特征。特征选择的方法有很多, 常用的特征选择方式主要分为 3 类: 过滤法、嵌入法和包装法<sup>[35]</sup>, 本实验分别采用 3 种特征选择的方法进行实验。实验的预测模型如图 10 所示。

实验中, 对过滤法、嵌入法和包装法三种特征选择方式, 分别采用随机森林、SVM 及 XGBoost 进行分类, 实验结果见表 5。在 3 种特征选择方式中, XGBoost 均能保持良好的准确率。当使用嵌入法作为特征选择方式的时候, XGBoost 模型的平均准确率达到了最高值 93%。

## 4 结论

本文基于面部特征提取算法提取了自闭症儿童的人脸关键点、头部姿态、人眼视线、面部动作单元等信息, 进一步研究了儿童的头部运动速度、头部累积移动距离、面部占比率、唤名反应等特征, 最后对特征进行编码, 构建了自闭症儿童视觉特征数据库, 为自闭症儿童的相关研究提供了公开数据集。与现有的自闭症数据库相比, 该数据集的特征完整, 易于其他研究者使用。最重要的是该自闭症儿童数据库不会暴露儿童的真实面部, 全部为脱敏数据, 保护了自闭症儿童的隐私。此外, 本文选取面部动作单元特征, 进行自闭症儿童的预测, 其分类精度最高可达 93%, 说明了所提起特征的有效性。

目前本文还存在着人工标记唤名时间点及部分特征的精度不够高、特征不够全面等缺点, 未来可

表 5 SVM、随机森林及 XGBoost 法基于 AU 单元的分类结果

特征选择方法	分类模型	平均准确率	AUC
过滤法	随机森林	0.83	0.95
	SVM	0.80	0.92
	XGBoost	0.84	0.95
嵌入法	随机森林	0.90	0.97
	SVM	0.76	0.86
	XGBoost	0.93	0.99
包装法	随机森林	0.83	0.96
	XGBoost	0.90	0.99

以从提高特征精度、增加更多的特征方面开展研究,来提高数据库的全面性和可靠性。➡

#### 参考文献

- [1] AMERICAN PSYCHIATRIC ASSOCIATION D,American psychiatric association. diagnostic and statistical manual of mental disorders:DSM-5[M]. Washington,DC:American Psychiatric Association,2013.
- [2] ZWAIGENBAUM L,BRYSON S, ROGERS T,et al.Behavioral manifestations of autism in the first year of life[J].International Journal of Developmental Neuroscience, 2005,23(2-3):143-152.
- [3] VALSTAR M.Automatic behaviour understanding in medicine[C]// Proceedings of the 2014 Workshop on Roadmapping the Future of Multimodal Interaction Research including Business Opportunities and Challenges.2014:57-60.
- [4] ALVARI G,COVIELLO L, FURLANELLO C.EYE-C:eye-contact robust detection and analysis during unconstrained child-therapist interactions in the clinical setting of autism spectrum disorders[J].Brain Sciences,2021,11(12):1555.
- [5] ZHAO Z,ZHU Z,ZHANG X, et al.Identifying autism with head movement features by implementing machine learning algorithms[J].Journal of Autism and Developmental Disorders,2021:1-12.
- [6] ROBLES M,NAMDARIAN N, OTTO J,et al.A virtual reality based system for the screening and classification of autism[J]. IEEE Transactions on Visualization and Computer Graphics,2022,28(5):2168-2178.
- [7] ALKAHTANI H,AHMED Z A T, ALDHYANI T H H,et al.Deep learning algorithms for behavioral analysis in diagnosing neurodevelopmental disorders[J]. Mathematics,2023,11(19):4208.
- [8] LI J,CHEN Z,LI G,et al. Automatic classification of ASD children using appearance-based features from videos[J]. Neurocomputing,2022,470:40-50.
- [9] 邱晓露,吴金星,刘晓光,等."互联网+医疗联盟平台"模式在孤独症谱系障碍诊疗中的应用[J].中国数字医学,2021,16(11):34-38.
- [10] CHAWARSKA K,MACARI S,SHIC F. Decreased spontaneous attention to social scenes in 6-month-old infants later diagnosed with autism spectrum disorders[J].Biological Psychiatry,2013,74(3):195-203.
- [11] ELSABBAGH M,FERNANDES J, WEBB S J,et al.Disengagement of visual attention in infancy is associated with emerging autism in toddlerhood[J].Biological Psychiatry,2013,74(3):189-194.
- [12] NADIGAS,OZONOFF S,YOUNG GS, et al.A prospective study of response to name in infants at risk for autism[J].Archives of Pediatrics & Adolescent Medicine,2007,161(4):378-383.
- [13] OZONOFF S,IOSIF A M,BAGUIO F, et al.A prospective study of the emergence of early behavioral signs of autism[J].Journal of the American Academy of Child & Adolescent Psychiatry,2010,49(3):256-266.
- [14] HASHEMI J,DAWSON G, CARPENTER K L H,et al.Computer vision analysis for quantification of autism risk behaviors[J]. IEEE Transactions on Affective Computing,2018,12(1):215-226.
- [15] JAISWAL S,VALSTAR M F, GILLOTT A,et al.Automatic

- detection of ADHD and ASD from expressive behaviour in RGBD data[C]//2017 12th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2017). IEEE,2017:762-769.
- [16] 陈靓影,刘小迪,张坤.人机交互技术支持的自闭症儿童评估:基于修订版 C-PEP 量表与人机交互游戏的评估结果对比[J].现代教育技术,2020,30(1):7.
- [17] HIROTA T,KING B H.Autism spectrum disorder:a review[J].Jama,2023,329(2):157-168.
- [18] BALL R S.The gesell developmental schedules:arnold gesell (1880-1961) [J].Journal of Abnormal Child Psychology,1977,5:233-239.
- [19] LORD C,RISI S,LAMBRECHT L, et al.The autism diagnostic observation schedule—generic:a standard measure of social and communication deficits associated with the spectrum of autism[J].Journal of Autism and Developmental Disorders,2000,30:205-223.
- [20] LUBETSKY M J,HANDEN B L. Medication treatment in autism spectrum disorder[J].Speaker's Journal,2008,8(2003):97-107.
- [21] HUS Y,SEGAL O.Challenges surrounding the diagnosis of autism in children[J].Neuropsychiatric Disease and Treatment,2021: 3509-3529.
- [22] BALTRUSAITIS T,ZADEH A, LIM Y C,et al.Open-face 2.0:facial behavior analysis toolkit[C]// International Conference on Automatic Face & Gesture Recognition.IEEE,2018.
- [23] BALTRUSAITIS T,ROBINSON P, MORENCY L P.Constrained local neural fields for robust facial landmark detection in the wild[C]//Proceedings of the IEEE international conference on computer vision workshops,2013:354-361.
- [24] COOTES T F,EDWARDS G J, TAYLOR C J.Active appearance models[J].IEEE Transactions on Pattern Analysis and Machine Intelligence,2001,23(6):681-685.
- [25] ZADEH A,CHONG LIM Y, BALTRUSAITIS T,et al. Convolutional experts constrained local model for 3d facial landmark detection[C]//Proceedings of the IEEE International Conference on Computer Vision Workshops,2017:2519-2528.
- [26] HONG J S,SINGH V,KALB L. Attention deficit hyperactivity disorder symptoms in young children with autism spectrum disorder[J].Autism Research,2021,14(1):182-192.
- [27] HUSSEIN Q M,FADHEL Z M. Children's autism spectrum diagnostic techniques using machine learning[J].Journal of Madenat Alelem University College,2022,14(3):24-35.
- [28] BALTRUSAITIS T,ROBINSON P, MORENCY L P.Constrained local neural fields for robust facial landmark detection in the wild[C]//Proceedings of the IEEE International Conference on Computer Vision Workshops,2013,354-361.
- [29] WOOD E,BALTRUSAITIS T, ZHANG X,et al.Rendering of eyes for eye-shape registration and gaze estimation[C]//Proceedings of the IEEE International Conference on Computer Vision,2015:3756-3764.
- [30] DONG Z,WANG G,LU S,et al. Spontaneous facial expressions and micro-expressions coding:from brain to face[J].Frontiers in Psychology,2022,12:784834.
- [31] EKMAN P,FRIESEN W V. Facial action coding system[J]. Environmental Psychology & Nonverbal Behavior,1978.
- [32] BALTRUSAITIS T,MAHMOUD M, ROBINSON P.Cross-dataset learning and person-specific normalisation for automatic action unit detection[C]//2015 11th IEEE International Conference and Workshops on Automatic Face and Gesture Recognition (FG). IEEE,2015,6:1-6.
- [33] HASHEMI J,DAWSON G, CARPENTER K L H,et al. Computer vision analysis for quantification of autism risk behaviors[J].IEEE Transactions on Affective Computing,2018,12(1):215-226.
- [34] LUCEY P,COHN J F,KANADE T, et al.The extended cohn-kanade dataset (ck+):a complete dataset for action unit and emotion-specified expression[C].2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition-Workshops.IEEE,2010:94-101.
- [35] TANG J,ALELYANI S,LIU H. Feature selection for classification: a review[J].Data Classification: Algorithms and Applications, 2014:37.

【收稿日期: 2023-12-07】

【修回日期: 2024-02-03】

(责任编辑: 朱那新)