

10-30-2023

A Precise Attention Tracking System Based on Computer Vision

Jiyuan Liu

School of Electrical Engineering, Nantong University, Nantong 226019 China, 1055869947@qq.com

Hanwen Qi

School of Electrical Engineering, Nantong University, Nantong 226019 China

Zhicheng Liu

School of Electrical Engineering, Nantong University, Nantong 226019 China

Minrui Fei

School of Mechatronic Engineering and Automation, Shanghai University, Shanghai 210053, China

See next page for additional authors

Follow this and additional works at: <https://dc-china-simulation.researchcommons.org/journal>



Part of the [Artificial Intelligence and Robotics Commons](#), [Computer Engineering Commons](#), [Numerical Analysis and Scientific Computing Commons](#), [Operations Research, Systems Engineering and Industrial Engineering Commons](#), and the [Systems Science Commons](#)

This Paper is brought to you for free and open access by Journal of System Simulation. It has been accepted for inclusion in Journal of System Simulation by an authorized editor of Journal of System Simulation. For more information, please contact xtfzxb@126.com.

A Precise Attention Tracking System Based on Computer Vision

Abstract

Abstract: A precise attention tracking system based on machine vision is designed to address the difficulty in studying students' attention allocation. The system includes an image capture device and an accurate attention tracking algorithm. The image capture device can capture the clearer ocular images. The pupil center localization algorithm replaces VGG16 with lightweight MobileNetv3 and uses twostage feature fusion and center keypoint prediction techniques to improve the speed and accuracy. The algorithm achieves a speed of up to 36 frames/s and 97.42% accuracy. The gaze tracking algorithm compensates for the head movements to achieve the meticulous gaze tracking. An interactive software for assessing cognitive abilities in school-age children is developed. The software calculates the eye movement indicators by using the collected gaze coordinates and evaluates the cognitive abilities based on psychological theory, and provides a reference for the psychology and education research.

Keywords

pupil positioning, improved single shot multibox detector(SSD) algorithm, image gradient tilt algorithm, oriented FAST and rotated brief (Eye-ORB) algorithm, reading cognition, attention tracking

Authors

Jiyuan Liu, Hanwen Qi, Zhicheng Liu, Minrui Fei, and Kun Zhang

Recommended Citation

Liu Jiyuan, Qi Hanwen, Liu Zhicheng, et al. A Precise Attention Tracking System Based on Computer Vision[J]. Journal of System Simulation, 2023, 35(10): 2087-2100.

一种基于机器视觉的精准注意力追踪系统

刘纪元¹, 祁瀚文¹, 刘志诚¹, 费敏锐², 张堃^{1,3*}

(1. 南通大学 电气工程学院, 江苏 南通 226019; 2. 上海大学 机电工程与自动化学院, 上海 210053;
3. 南通市智能计算与智能控制重点实验室, 江苏 南通 226019)

摘要: 针对学生注意力分配困难和对学习影响等问题, 提出一种基于机器视觉的精准注意力追踪系统。该系统包括图像采集装置和精准的注意力追踪算法。图像采集装置可以获得更清晰的眼部区域图像。瞳孔中心定位算法用轻量级的 MobileNet v3 替换 VGG16 (visual geometry group network), 采用两级特征融合和中心关键点预测技术, 提高了检测速度和准确率。该算法检测速度可达 36 帧/s, 准确率为 97.42%。视线追踪算法旨在解决头部偏移的影响, 实现对视线的精确追踪。研发了一款面向学龄儿童的阅读认知评价交互软件。该软件利用采集到的视线坐标计算相关眼动指标, 再通过心理学理论分析建模来评估学龄儿童的思维认知能力, 为心理学和教育学相关领域研究提供了参考和借鉴。

关键词: 瞳孔定位; 改进型 SSD (single shot multibox detector) 算法; Eye-ORB (oriented FAST and rotated brief) 算法; 阅读认知; 注意力追踪

中图分类号: TP391.7 文献标志码: A 文章编号: 1004-731X(2023)10-2087-14

DOI: 10.16182/j.issn1004731x.joss.23-FZ0829

引用格式: 刘纪元, 祁瀚文, 刘志诚, 等. 一种基于机器视觉的精准注意力追踪系统[J]. 系统仿真学报, 2023, 35(10): 2087-2100.

Reference format: Liu Jiyuan, Qi Hanwen, Liu Zhicheng, et al. A Precise Attention Tracking System Based on Computer Vision[J]. Journal of System Simulation, 2023, 35(10): 2087-2100.

A Precise Attention Tracking System Based on Computer Vision

Liu Jiyuan¹, Qi Hanwen¹, Liu Zhicheng¹, Fei Minrui², Zhang Kun^{1,3*}

(1. School of Electrical Engineering, Nantong University, Nantong 226019 China; 2. School of Mechatronic Engineering and Automation, Shanghai University, Shanghai 210053, China; 3. Nantong Key Laboratory of Intelligent Control and Intelligent Computing, Nantong 226019, China)

Abstract: A precise attention tracking system based on machine vision is designed to address the difficulty in studying students' attention allocation. The system includes an image capture device and an accurate attention tracking algorithm. The image capture device can capture the clearer ocular images. The pupil center localization algorithm replaces VGG16 with lightweight MobileNetv3 and uses two-stage feature fusion and center keypoint prediction techniques to improve the speed and accuracy. The algorithm achieves a speed of up to 36 frames/s and 97.42% accuracy. The gaze tracking algorithm compensates for the head movements to achieve the meticulous gaze tracking. An interactive software for assessing cognitive abilities in school-age children is developed. The software calculates the eye movement indicators by using the collected gaze coordinates and evaluates the cognitive abilities based on psychological theory, and provides a reference for the psychology and education research.

收稿日期: 2023-07-02 修回日期: 2023-08-07

基金项目: 江苏省产学研合作项目(BY2022224)

第一作者: 刘纪元(1999-), 男, 硕士生, 研究方向为图像处理。E-mail: 1055869947@qq.com

通讯作者: 张堃(1983-), 男, 教授, 博士, 研究方向为人工智能等。E-mail: zhangkun_nt@163.com

Keywords: pupil positioning; improved single shot multibox detector(SSD) algorithm; image gradient tilt algorithm; oriented FAST and rotated brief (Eye-ORB) algorithm; reading cognition; attention tracking

0 引言

视觉是人类获取和处理外界信息的重要渠道, 获取到的外界信息高达 80%~90%, 而眼动记录技术能够对读者的阅读过程进行实时记录, 是研究阅读过程的一种重要的研究方法。人眼运动轨迹可以直观地反映当前注视点和注视时间, 从而反应出心理变化过程, 因此, 视线追踪也被广泛应用在人类的认知和记忆的研究中。

传统的注意力追踪方法一般基于穿戴式传感设备, Dong 等制造了一种蛇形结构的软性多功能电子皮肤收集眼电信号、皮肤温度和汗液信号, 用于眼部注意力追踪^[1]。Wang 等采用一种柔性水凝胶生物传感器和小波变换支持向量机算法结合, 用于识别眼球的基础运动^[2]。穿戴式传感设备虽然可以较准确地追踪眼球运动, 但造价昂贵且穿戴不便, 只适用于小部分特殊需求的场景, 无法普及使用。随着机器视觉技术的不断发展, 部分研究人员将注意力追踪的研究方向转移到机器视觉领域, Tesfamikael 等采用图像分割的方法将人体眼部区域分割出来并对其进行跟踪, 但是该方法是基于眼眶进行检测的, 并没有准确地定位到瞳孔, 而人的注意力主要是瞳孔决定的, 因此, 对注意力判断的准确度不高^[3]。Zhu 等提出了一种在线多对象跟踪方法, 将单个对象跟踪和数据关联方法的优点集成在统一的框架中, 以处理噪声检测和目标之间的频繁交互, 并提出了具有空间和时间注意力机制的双重匹配注意网络, 可以对人体眼球注意力进行有效跟踪^[4]。

国内部分学者也在注意力追踪上做了深层次的研究, 提出了一些眼动识别的新算法。靳冰凌等针对驾驶员的注意力追踪进行研究, 利用 AdaBoost 算法定位人脸和眼睛区域, 计算眼睑和瞳孔区域像素个数占眼部区域像素总个数的比值,

判断眼睛状态^[5]。陈加徐等采用优化的梯度重心法和 Dlib 特征点法进行虹膜中心和眼角点定位, 完成了自然光源下眼动的视觉测量, 但该算法的视频帧定位时间较长, 不能达到实际检测的效果^[6]。宋琴琴等首先对视频图像进行灰度化处理, 将视频图像从 RGB 色彩空间转换到灰度空间, 然后对灰度图像进行高斯滤波处理, 去除图像的噪点, 接着对图像进行二值化处理找到人眼区域并检测出瞳孔轮廓, 最后进行瞳孔中心定位实现人眼动态追踪^[7]。王晓明等采用深度学习的方式, 构建了用于眼动识别的计算模型, 通过测量预测注视序列与实际注视序列之间的相似性距离进行识别, 但是该方法对眼球追踪的准确率较低^[8]。

受高成本和固定性的限制, 传统的眼动追踪装置不允许在实验室外收集数据。便携式科研级眼动仪在检测速度和精度方面表现出色, 但由于设备价格昂贵, 无法进行大规模应用, 严重限制了眼动追踪技术的使用场景^[9]。为了克服这些限制, 设计了一种图像采集装置, 该装置在设备体积上实现了便携化, 并且自主研发的装置成本约为科研级眼动仪的几十分之一, 使眼动追踪技术能够得到更广泛的应用。

本文致力于解决学生阅读中的注意力分配等问题, 设计了一种基于机器视觉的精准注意力追踪系统。首先, 设计了一款高清图像采集装置, 该装置配备高灵敏度、高分辨率的摄像头, 并配备红外光电二极管, 以获得更清晰的瞳孔图像。对 SSD(single shot multibox detector)^[10]算法进行了改进, 将其特征提取网络从 VGG16(visual geometry group network) 替换为轻量化的 MobileNetv3, 并通过多层特征融合来获取更多深层特征, 从而实现准确定位和获得丰富的语义信息。在目标输出网络中加入了中心关键点预测, 以去除不正确的边界框, 最终实现准确定位瞳孔

位置。随后, 采用 Eye-ORB(oriented FAST and rotated brief)算法计算头部偏移量, 解决头部偏移对视线追踪的影响。最后, 引入了改进型 RBF 算法, 拟合去除头部偏移量的瞳孔坐标与屏幕坐标之间的变换矩阵, 实现对视线的追踪。此外, 还设计了一款认知阅读评估软件, 通过采集到的数据和心理学理论构造多维认知维度, 并进行数据建模, 全面评估测试者认知状态。

1 精准瞳孔定位方法

整体检测流程如图1所示。视频输入是自主研发的眼动设备采集面部图像, 然后通过改进型 SSD 算法实时精准定位出瞳孔中心坐标位置, 接着使用 Eye-ORB 算法计算头部偏移量, 解决头部偏移对视线追踪的影响, 最后引入改进型 RBF 算法拟合去除头部偏移量的瞳孔坐标到屏幕坐标的变换矩阵, 完成对人眼的视线追踪。

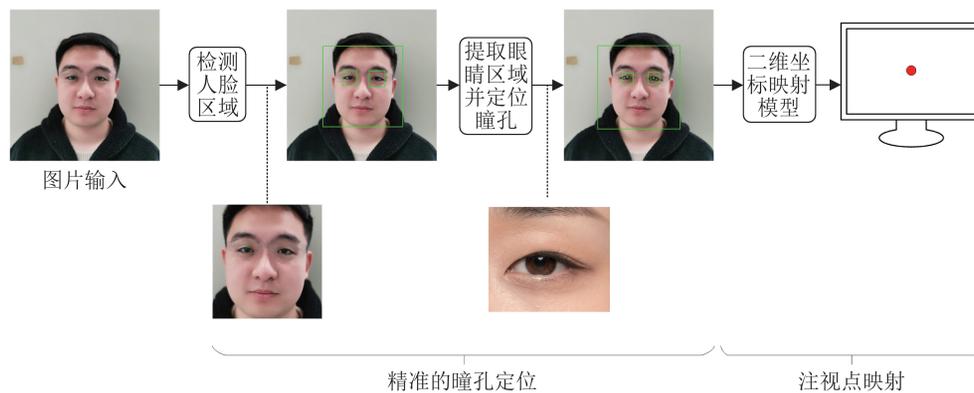


图1 注意力追踪算法流程图

Fig. 1 Flow chart of attention tracking algorithm

1.1 精准的图像采集装置

图像采集装置如图2所示, 主要由摄像头、IR 红外光源和 USB 扩展接口等模块组成。其中, 摄像头模块采用的是高灵敏度、高分辨率的摄像头, 同时对该摄像头还进行一定的改装, 将摄像头原有的红外光阻挡镜片更换成可见光阻挡镜片, 使其能够拍摄到 IR 红外光在眼睛上的反射, 从而得到更加清晰的瞳孔图像。IR 红外光源模块主要是增加额外的红外光源来照亮眼睛的红外光谱, 使拍摄到的瞳孔图像更加的突出, 方便摄像头捕捉到。USB 扩展接口可以将采集到的图像传输给电脑, 完成图像的传输。图像采集装置的采集效果如图3所示。

1.2 改进型 SSD 瞳孔检测算法

改进型 SSD 检测算法是一种基于深度学习的瞳孔定位算法, 由眼睛区域提取和瞳孔中心检测

组成。首先, 通过 Fast RCNN 网络检测出人脸, 并将人脸部分裁剪下来, 然后, 从人脸图像中提取路标, 根据得到的路标信息选中人脸中的眼睛区域, 最后, 瞳孔中心检测由 SSD 网络进行定位。算法通过大量的数据进行训练, 得到了一个很好的追踪效果, 瞳孔定位对比效果如图4所示。

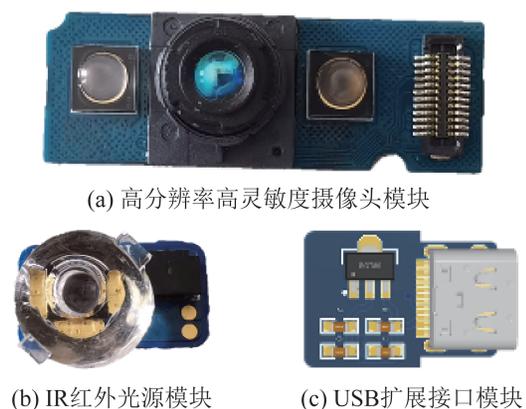


图2 图像采集装置

Fig. 2 Image acquisition device



图 3 图像采集装置采集到的图像
Fig. 3 Images captured by image acquisition device

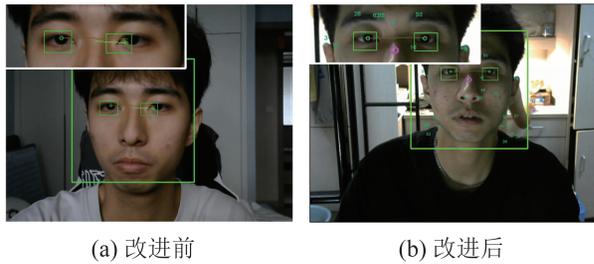


图 4 瞳孔定位改进对比效果
Fig. 4 Detection effect of precise pupil positioning method

SSD 的核心设计在于采用多尺度特征用于检

测，并设置先验框。该网络结构可以分为 3 个部分：特征检测、特征融合和目标输出。在特征检测方面，SSD 采用了 VGG16^[11] 作为基础模型，并在此基础上新增了卷积层来获取更多的特征图以用于检测；在特征融合方面，该算法利用多尺度特征图进行检测；在目标输出网络方面，SSD 采用了多个不同级别的特征图进行融合，以进行多尺度预测。

本文对原算法进行了改进。首先，在考虑实时性检测的要求下，将 SSD 模型的 VGG16 改为轻量级的 MobileNet v3 网络。其次，提出了一种二级特征融合策略，可以获取更多的深层特征，从而提高对小目标检测的精度，进而提高瞳孔的检测精度。最后，在目标输出网络中，加入了中心关键点预测，这样可以有效地滤除不正确的包围框，进一步提高输出结果的准确性。改进型 SSD 模型如图 5 所示。

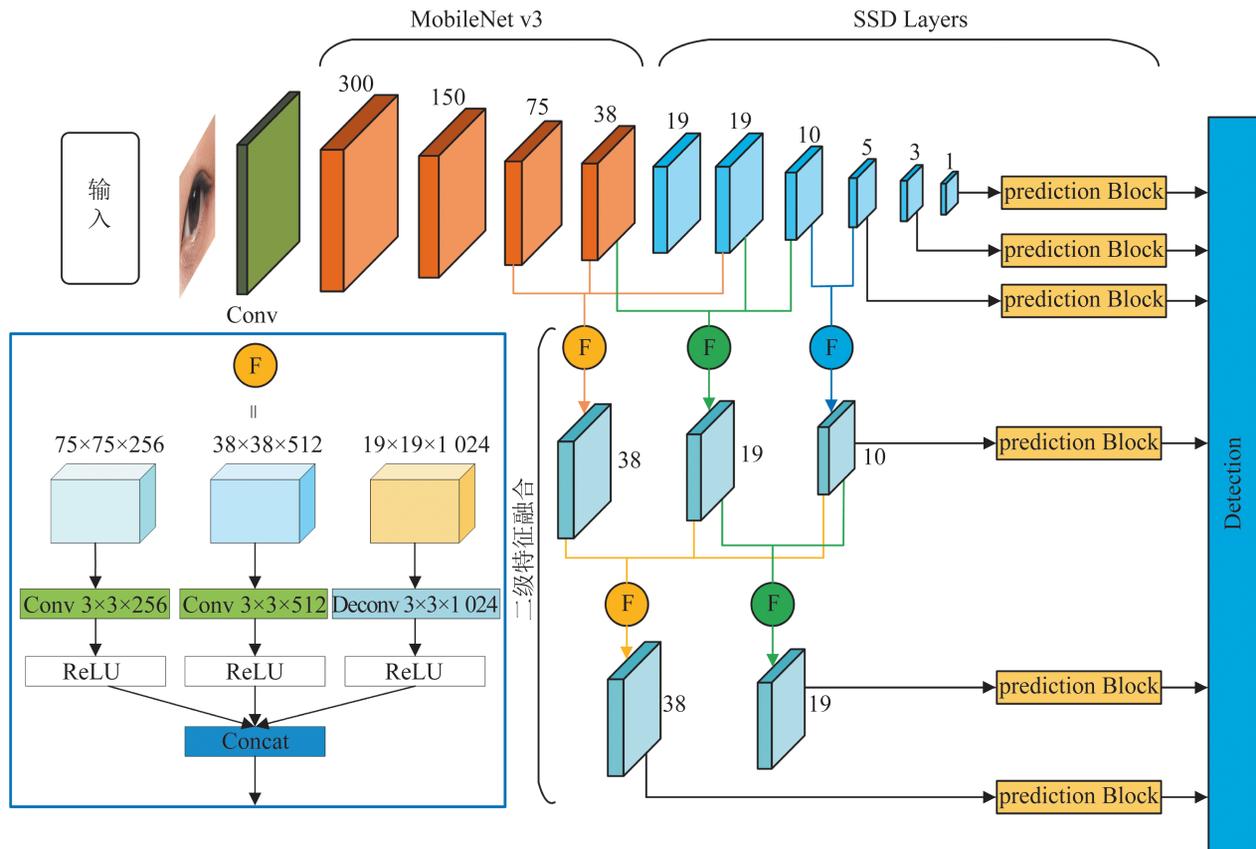


图 5 改进型 SSD 算法结构图
Fig. 5 SSD algorithm structure diagram

1.2.1 MobileNet v3

MobileNet v3^[12]是由谷歌团队在2019年提出的,其在ImageNet图像分类任务上的准确率为75.2%,明显高于VGG16的71.2%,而其参数数量仅为540万,约为VGG16的1/25。

MobileNet v3的主要特点包括:采用深度可分离的卷积设计,将每一个卷积分解为点卷积和深度卷积2个部分,以减少模型的参数数量和计算量,从而加快模型的训练和推理速度;使用了一个自适应网络结构设计,可以根据不同的任务需求,动态调整模型的宽度和深度,适应不同的场景和设备。

将MobileNet v3替换SSD模型中的特征提取网络,将大大提升SSD的检测速度,从而达到实时检测的目的。

1.2.2 二级特征融合策略

为了提高检测精度并更好地捕捉图像中的各种特征,提出了一种二级特征融合策略。该策略旨在使浅层特征图能够获取更多的语义和全局信息,而无需依赖其他复杂的功能。在一级融合过程中,同时合并了不同尺寸(75、38、19、20、5)的特征图。通过一个融合块与它们建立连接,合并为3个具有38×38、19×19、10×10特征尺寸的新一级融合层。在二级融合过程中,融合了一级融合的3个尺寸的特征层,其融合过程与一级特征融合类似,生成了2个具有38×38、19×19的特征尺度深层融合层。以第一级特征融合网络为例,特征融合网络如图5蓝色框所示。

总之,预测层分别从6个不同特征尺度进行预测,其中包括3个一级融合模块和2个深层融合模块。深度融合模块是一个增强型融合模块,它包含了精确的定位和强大的语义信息,在检测瞳孔这种小目标时,具有更好的效果。

1.2.3 预测模块

与传统的SSD一样,最终的预测层直接在所选的特征图上应用分类损失和回归损失。为了有效

地过滤错误边界框,利用检测到的边界框中心关键点进行如下操作:①根据选中边界框确定中心关键点;②为每一个边界框定义一个中心区域,并检查中心区域是否包含中心关键点;③如果在中心区域检测到中心关键点,便保留该边界框,并将该中心关键点作为瞳孔中心。如果在中心区域没有检测到中心关键点,则将预测的边界框移除。

边界框中心区域的大小影响检测结果。例如,对于小的边界框,中心区域较小,召回率较低,而对于大的边界框,中心区域较大,召回率较高。因此,将检测区域分为9个小区域,是否落在中心区域是最终的检查中心关键点。

2 视线追踪方法

上述操作已经能精准地定位瞳孔,但是由于人的头部是可以转动的,所以,当头部出现偏移,瞳孔映射到屏幕的位置也会偏移。为了解决这一问题,使用Eye-ORB算法匹配头部若干点前后帧的位置关系,通过加权处理得到最能体现头部偏移程度的参照点,基于该参考点进行帧偏移计算,得到当前头部的偏移量;引入改进型RBF算法拟合去除头部偏移量的瞳孔坐标到屏幕坐标的变换矩阵,完成对人眼的视线追踪。图6是视线追踪方法的示意图,蓝色的人代表实际的人,灰色的人代表经过算法处理去除偏移的虚拟人。

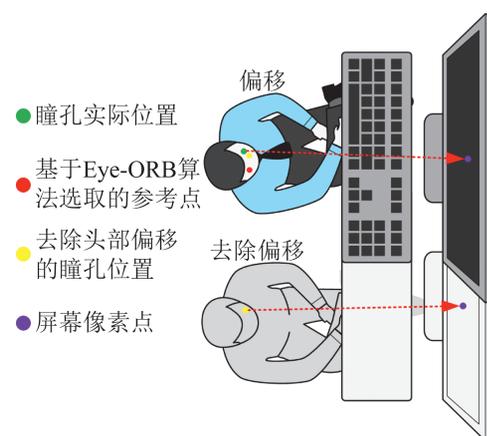


图6 视线追踪方法示意

Fig. 6 Sketch map of eye tracking method

2.1 基于 Eye-ORB 算法的头部偏移量计算

2.1.1 Eye-ORB 算法

ORB 算法是一种快速特征点提取和描述的算法，将特征点检测算法 FAST(features from accelerated segment test) 算法与特征描述算法 BRIEF(binary robust independent elementary features) 算法结合起来，并在它们原来的基础上做了改进与优化^[13]。ORB 算法最大的特点就是计算速度快。这首先得益于使用速度极快的 FAST 算法检测特征点，再次是使用 BRIEF 算法计算描述子进行特征描述，该描述子特有的二进制字符串的表现形式不仅节约了存储空间，而且大大缩短了匹配的时间^[14]。

本文提出的 Eye-ORB 算法在 ORB 的基础上进行优化，具体改进方式如下：

(1) 利用了 SURF(speededup robust features) 算法的尺度不变性和拉普拉斯算子的图像增强算法对传统 ORB 进行优化；

(2) 引入了一种基于拉普拉斯算子的图像增强算法，对用于后续特征点检测的区域进行图像增强。

ORB 算法非尺度不变问题的根本原因是 FAST 算法检测到的特征点不包含尺度不变信息。尽管通过引入特征点的方向可以获得旋转不变性，但描述符并不是尺度不变的。因此，解决 ORB 算法非尺度不变问题的方法是检测到的特征点需要具有尺度不变性，从而生成的描述符包含尺度不变的信息。SURF 算法具有尺度不变性，因此，SURF 算法利用多尺度空间提取具有尺度不变性的特征点^[15]，来代替原 ORB 算法中的 FAST，然后利用 ORB 算法中的描述符描述特征点，利用汉明距离和 Lowe's 算法匹配特征点^[16]。解决了 ORB 算法非尺度不变和亮度条件差时匹配精度差的问题，同时保留了 ORB 算法描述符匹配速度快、旋转不变性等优点。

为了解决图像亮度引起的匹配精度差的问题，

在检测特征点之前，先对图像进行增强。拉普拉斯算子是图像增强中常用的一种算子，本质是一种各向同性滤波器。拉普拉斯算子作为一种二阶线性微分算子，具有较强的边缘定位能力，其效果优于一阶微分算子。对于定义一张二维图像为 $f(x,y)$ ，拉普拉斯算子如式(1)所示：

$$\nabla^2 f(x,y) = f(x+1,y) + f(x-1,y) + f(x,y+1) + f(x,y-1) - 4f(x,y) \quad (1)$$

式中： $f(x,y)$ 为原始图像对应的二值函数。常用高通滤波模板中所有项之和为 0。当每个图像元素的灰度值相同或在模板覆盖的区域内灰度级变化缓慢时，滤镜模板输出非常小，甚至为 0。由于强调了图像中的突变，平滑灰度区域无响应，因此，模板系数之和为 0。增强方程为

$$g(x,y) = f(x,y) + c[\nabla^2 f(x,y)] \quad (2)$$

式中： $g(x,y)$ 为图像增强后的二值函数； c 为系数， c 决定增强强度， c 越大，图像的增强效果越明显。

2.1.2 Eye-ORB 算法计算头部偏移量实施历程

step 1: 基于 SURF 算法进行极值点检测。SURF 算法能够避免检测极值点时特征点重复，将高斯微分模板和图像卷积转换为框滤波，利用不同大小的盒形滤波器对图像进行滤波，形成图像金字塔，然后通过 Hessian 矩阵对金字塔的每一层进行相应的极值点检测^[17]。

step 2: 在以 SURF 算法检测的极值点为中心的 $3 \times 3 \times 3$ 立体区域进行非最大值抑制，即当一个点的极值与周围所有图像元素的值相比最大或最小时，称为候选特征点，然后对图像和尺度空间进行插值，找到所需的特征点分布和尺度值。

step 3: 利用 Harr 小波在特征点的圆形区域内的统计特征进行特征点的主方向分布。

step 4: 描述和匹配特征点，并生成具有旋转不变性的二进制描述符。此外，通过考虑汉明距离，使特征点的粗匹配更加精确。最后，利用 Lowe's 算法获得期望匹配点作为最终参考点。

step 5: 计算前后帧上最终参考点的运动向

量, 消除头部运动对计算瞳孔偏移量造成影响。

2.2 可偏移瞳孔坐标到屏幕坐标的转换

基于 Eye-ORB 算法得到了头部的偏移量, 进而得到去除头部偏移量的瞳孔坐标, 本文基于改进型 RBF 算法进行训练, 拟合去除头部偏移量的瞳孔坐标到屏幕坐标的变换矩阵, 注视方向转换为屏幕上的像素位置, 完成对人眼的视线追踪。

RBF 算法是一组加权径向基函数的线性组合, 它通常由 3 层组成一输入层、带有激活函数的隐藏层和输出层。该算法表示一个非线性映射, 通过一组基函数将输入向量转换为输出标量。当应用于回归问题时, 拟合的未知函数 $f(\mathbf{x})$ 可以直接用 RBF 进行近似。

$$f(\mathbf{x}) \approx \sum_{i=1}^N w_i \varphi(\|\mathbf{x} - \mathbf{c}_i\|) \quad (3)$$

式中: \mathbf{x} 为输入向量; w_i 为权重; φ 为基函数; \mathbf{c}_i 为中心向量; N 为中心向量的个数。但是该算法在使用过程中拟合速度较慢甚至出现拟合失败的情况^[18]。

本文在 RBF 算法的基础上进行改进, 提出了改进型 RBF 算法, 具体改进方法如下:

- (1) 引入 PROSAC(progressive sample consensus) 算法去除训练过程中的离群值;
- (2) 高斯函数具有光滑性和无限可微性等出色的性质, 选择它作为基函数;
- (3) 增加一个额外的线性项式, 这样不仅可以有效地再现函数的全局行为, 而且可以起到一定的修正作用。

在训练过程中, 当人眼处于闭合状态, 瞳孔位置定位不准确, 产生离群值, 导致拟合效果严重退化。本文引入 PROSAC 算法去除训练过程中的离群值。PROSAC 算法能够对匹配点的优劣性进行排列, 取匹配度高于一定阈值的点对作为假设生成集, 在假设生成集中进行随机采样, 即 PROSAC 算法是从不断增大的最佳对应点集合中进行采样的, 这种方法可以有效去除训练过程中

的离群值, 并且能够节省计算量, 提高运行速度。

高斯函数作为基函数:

$$h_i = \exp\left(-\frac{\|\mathbf{x} - \mathbf{c}_i\|^2}{2b_i^2}\right), i = 1, 2, \dots, n \quad (4)$$

在以高斯基函数为基函数的情况下, RBF 算法的拟合表达式为

$$f(\mathbf{x}) = w_i \exp\left(-\frac{(\|\mathbf{x} - \mathbf{c}_i\|)^2}{2\sigma^2}\right) \quad (5)$$

式中: σ 为宽度。

高斯函数是一种局域函数。它从中心附近的位置迅速衰减, 表明远离马氏距离较大的中心数据点将无法激活该基函数。此外, 在中心附近, 高斯函数可能会高估函数值。为了克服这些问题, 在 RBF 算法的基础上增加一个额外的线性项式, 这样不仅可以有效地再现函数的全局行为, 而且还可以起到一定的修正作用。加入额外的线性项后, RBF 网络的形式为

$$f(\mathbf{x}) = \sum_{i=1}^N w_i \exp\left(-\frac{(\|\mathbf{x} - \mathbf{c}_i\|)^2}{2\sigma^2}\right) + a\mathbf{x} + b \quad (6)$$

式中: a 为线性项系数; b 为常数项。

定义去除头部偏移的瞳孔坐标 (x, y) , 对应的屏幕像素位置为 (x', y') , 因为眼睛偏移和屏幕位置之间的关系并不完全是线性的, 当眼睛看向某一侧时, 眼睛瞳孔的小幅度运动对应屏幕上的大幅度运动, 需要同时输入二次特征 (x_2, y_2) , 所以输入特征向量是 $[x, y, 1, x_2, y_2, xy]$, 基于上述拟合方法来训练一个 2×6 的矩阵, 该矩阵能将头部偏移的瞳孔坐标与屏幕像素位置。

3 认知阅读评价交互软件

为了更好地将设计的图像采集装置和视线追踪算法应用到更多的场景中, 自主研发了一款中小学生学习认知阅读评价交互软件。该软件通过将视线追踪算法得到的注视点坐标进行记录, 收集到测试人员的眼动轨迹数据, 通过得到的眼动轨迹数据计算相关的眼动指标, 最后, 通过对不同的

眼动指标分配不同的权重得到测试者的认知评估。

在软件设计过程中，特别注重软件界面的简洁易用和用户友好性。同时，本软件还可以根据需求自由更换测试文本。此外，在建立认知评估模型时，引入专业的心理学和教育学的认知评估理论，得到不同的认知维度与各眼动指标的相关性系数，最后，通过特征融合的方法计算出不同认知维度的评价得分。这种综合优势和特色将有助于学校更好地评估学生的认知阅读水平和思维品质，从而帮助学生取得更好的学习成果。

3.1 认知评价软件界面

认知阅读评估软件的界面如图 7 所示，最上方为软件标题栏；左下方为导航栏，分别为用户信息选择、软件功能展示和眼动轨迹数据记录三大模块；右下方为展示栏，主要用于将导航栏的信息进行展示。用户在导航栏中将自己的信息填写，便可用该软件对专注度进行测试，最终的测试结果在数据记录中展示出来，数据记录中可以同时展示眼动轨迹动图和眼动轨迹坐标。

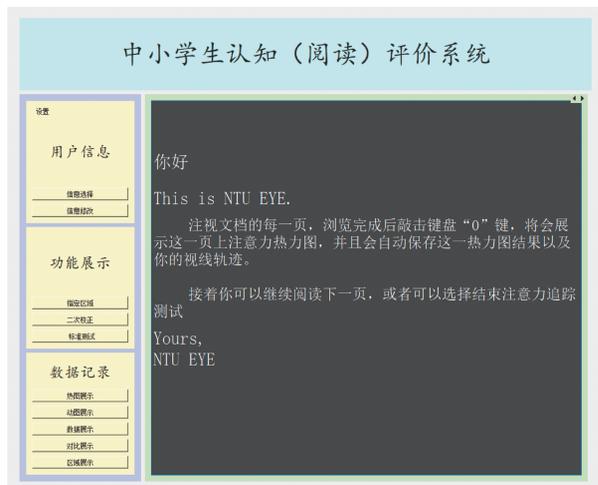


图 7 认知阅读评估软件界面
Fig. 7 Cognitive reading assessment software interface

3.2 认知阅读评估软件功能

认知阅读评估软件的功能如图 8 所示，首先，

需要测试者进行用户登录，然后，选择需要测试的信息和测试的模式，接下来，需要用户对视线坐标进行校准，使得注视点坐标更加的精准。校准完以后，测试者便可以用认知评估软件对视线进行追踪测试。测试结束后，不仅可以得到注视点坐标，还可以通过注视点坐标计算出相关的眼动指标，展示测试者阅读的热力图并还原测试者的阅读过程。同时，将所有的测试者的相关眼动指标进行存档，对比所有人在同一材料上的眼动指标的差异性。

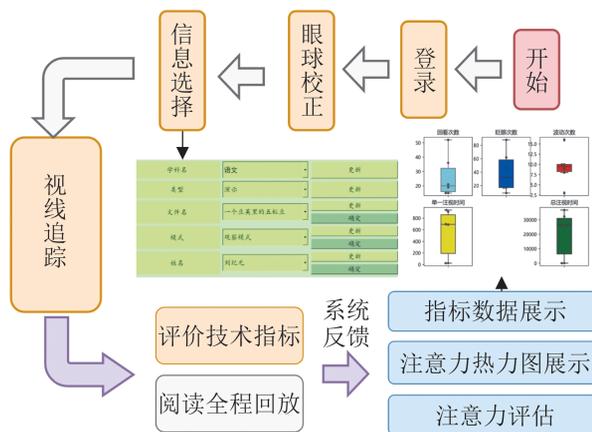


图 8 阅读认知评估使用流程图
Fig. 8 Flowchart of reading cognitive assessment

4 实验分析

4.1 实验平台

本实验所用平台来自实验室搭建的配备 2 张英伟达 RTX-3090 24 GB 显卡的深度学习计算平台。操作系统及版本：UBUNTU20.04；机器学习环境配置：TORCH1.7.0CUDA11.1；程序编译环境：PYTHON3.6.12。

4.2 数据集制作

使用 3 个公开数据集：GI4E^[19]、I2Head^[20]和 BioID^[21]数据集，如图 9 所示。

GI4E：虹膜中心和眼角检测创建的一个公共数据集。数据集中的图像是通过标准网络摄像头进行采集的，图像分辨率为 800 像素×600 像素，

该数据库共由1 339张图像组成, 其中, 804张图片作为训练集, 134张图片为验证集, 401张图片作为测试集。



图9 本实验中使用的数据集图像
Fig. 9 Images of datasets used in this experiment

I2Head: 该数据集一共收集了12个测试者结合真实地面数据的头部姿势、注视方向和简化的测试者面部模型。该数据集使用标准的网络摄像头对测试在头部固定和头部可自由移动的场景中记录8个会话, 记录不同注视点的点网格。为了使采集到的数据更加的清楚, 选取每一个固定点的最好的10帧图像, 为样本提供图像和头部姿势。共2 784张图片, 其中, 1 670张图片作为训练集, 278张图片为验证集, 836张图片作为测试集。

BioID: 该数据集一共收集了23位测试者的正面拍摄的照片, 共1 521张分辨率为384像素×286像素的灰度照片。其中, 913张图片作为训练集, 152张图片为验证集, 456张图片作为测试集。此外, 该数据集还包含了人工对图像中人眼的位置进行标注的文件。

4.3 评价指标

本文提出的算法分为2个部分: 定位瞳孔的方法和匹配特征点的方法。对所提出的瞳孔中心定位的精准度进行评估, 精准度的定量评估标准为

$$e = \frac{1}{d} \max(d_l, d_r) \quad (7)$$

式中: d_l 和 d_r 为估计的左右眼与中心点之间的欧几里得距离; d 为双眼之间真实的欧几里得距离。在正常光照条件下, 对眼睛瞳孔定位的准确度进

行归一化计算时, $e \leq 0.05$ 表示估计的瞳孔中心位置在真实瞳孔半径范围内, 为第一精度水平, 此时反映瞳孔中心检测的最高精度; $e \leq 0.1$ 表示估计的瞳孔中心位置在虹膜半径范围内, 为第二精度水平; $e \leq 0.25$ 表示估计的瞳孔中心位置在眼睛中心和眼角半径范围内, 为第三精度水平。因此, 在瞳孔中心检测中, 必须保证 $e \leq 0.25$ 的同时最大程度地提高性能, 若 $e \leq 0.05$ 时将取得一个良好的结果。

根据人脸数据集, 对所提出的方法和SOTA (state of the art)方法进行比较, 表1是根据GI4E数据集比较的结果, 表示的定量结果都是经过5次实验获得的, 表中的数据是5次实验的平均值。

表1 GI4E数据集比较结果
Table 1 GI4E dataset comparison results

算法	$e \leq 0.05$	$e \leq 0.10$	$e \leq 0.25$
文献[22]	99.50	NA	NA
文献[23]	99.10	100	100
文献[24]	99.60	99.84	100
本文方法	99.77	99.97	100

注: NA表示空值, 不适用。

表1显示了GI4E数据集的实验结果, 精准度误差 $e \leq 0.05$ 时, 本文方法精准度为99.77%, 比文献[24]方法的准确度高0.17%; 在精准度误差 $e \leq 0.1$ 时, 本文方法精准度为99.97%, 与文献[23]方法相比稍稍逊色; 精准度误差 $e \leq 0.25$ 时, 本文方法的精准度达到了100%。

表2显示了BioID和I2Head数据集的实验结果, 精准度误差 $e \leq 0.05$ 时, 本文算法在2个数据集中的检测精准度分别为97.63%和99.42%, 远优于其他的检测算法; 精准度误差 $e \leq 0.1$ 时, 本文算法在2个数据集中的检测精准度分别为99.3%和100%, 同样优于其他的检测算法; 精准度误差 $e \leq 0.25$ 时, 本文算法在2个数据集中的精准度为100%。综上所述, 本文算法具有很好的鲁棒性, 在多个数据集中都可以取得良好的效果。

4.4 瞳孔定位算法的消融实验

本文采用FAST RCNN网络选出人脸中的眼睛区域，然后采用改进型SSD网络完成对瞳孔的精准定位，为了证明改进型SSD网络的有效性，将原始的SSD模型作为消融实验的基线方法，将MobileNet v3模块、二级特征融合模块和预测模块按照顺序叠加在基线方法上。本文分析了统计指标的变化，消融实验的统计结果如表3所示。

如表3所示，在基线方法中加入MobileNet v3后，3项指标均有所提高，最明显的是检测速度，提高了27帧/s，表明将主干网络替换成MobileNet v3可以有效加快检测速度。在加快检测效果的同时，没有削弱SSD网络的检测精度。在Model3中加入了二级特征融合模块，检测精度(AP)得到了改善，AP提高了3.8%，AP50提高了2.5%。为了使检测精度得到进一步提升，在SSD网络中加入了预测模块，检测精度得到显著提升，AP50提升了8%。由于加入了2个模块，融合了更多的信息，导致检测精度有所下降，但已满足实际操作的需求，这表明所提出的3个模块对瞳孔定位产生了积极影响。因此，本文模型是上述模型中最优的。

4.5 图像采集装置性能分析

为了获取更清楚的眼部图像，本文采用了自主设计的图像采集装置，将安全功率的近红外光照射在人眼上。由于眼睛的不同部位对近红外光的反射率不同，从而得到更加清楚的眼部图像，完成对眼部图像的高清采集。为了验证本文图像采集装置的有效性，将装置与一般的摄像头采集到的数据做对比实现。如表4所示，采用检测速度和检测精度来评价本文图像采集装置的性能。本文图像采集装置的性能明显优于普通的摄像头采集到的数据，本文图像采集装置，在检测速度上比普通的摄像头每秒多检测11张图片，在检测精度上比普通的摄像头高4.74%。

4.6 视线追踪方法实验分析

首先针对本文提出的Eye-ORB算法匹配前后帧特征点的能力进行分析。在ORB的基础上进行优化，为了验证本文算法对前后帧脸部特征点匹配的能力，选择2种传统的特征点匹配算法ORB算法和SURF算法与本文提出的Eye-ORB算法进行5组对比实验，如表5所示。由表5可知Eye-ORB算法性能优于ORB算法和SURF算法，前后帧平均匹配成功数约为169对。

表2 I2Head和BioID数据集比较结果

Table 2 Comparison results of I2Head and BioID datasets

精准确误差	BioID			I2Head			%
	文献[22]	文献[25]	本文算法	文献[26]	文献[25]	本文算法	
$e \leq 0.05$	89.48	96.71	97.63	91.24	99.00	99.42	
$e \leq 0.10$	98.22	98.95	99.30	96.38	99.72	100	
$e \leq 0.25$	NA	100	100	97.42	100	100	

表3 瞳孔定位算法对比实验

Table 3 Comparison experiment of pupil localization algorithm

模型	MobileNet v3	二级特征融合模块	预测模块	AP/%	AP50/%	检测速度/(帧/s)
原始SSD模型				68.4	84.1	32
改进模型1	√			71.3	86.7	59
改进模型2	√	√		75.1	89.2	48
本文模型	√	√	√	83.4	97.2	43

表4 图像采集装置对比实验
Table 4 Comparison test of image acquisition device

图像采集装置	检测速度/(帧/s)	检测精度/%
自主研发采集装置	36	97.42
普通摄像头	25	92.68

表5 前后帧特征点匹配对数对比实验
Table 5 Comparison experiment of matching logarithm of feature points in front and back frames

组数	SURF	ORB	Eye-ORB
第1组	53	134	200
第2组	62	136	180
第3组	61	128	150
第4组	65	140	146
第5组	67	131	171
平均	61.6	133.8	169.4

为了验证视线追踪方法的有效性, 将本文的视线追踪方法与其他的方法进行对比, 表6为对比结果。由表6可知, 本文提出的视线追踪方法的误差最小, 精度最高。

表6 视线追踪比较结果
Table 6 Comparison results of gaze tracking mm

方法	水平平均误差	垂直平均误差
文献[27]	1.5	1.7
文献[28]	1.2	1.4
本文算法	1.1	1.2

4.7 认知阅读评估软件性能分析

为验证认知评估软件的性能, 从当地小学中抽取了38名年龄为10岁的三年级志愿者, 男女各19人, 其中, 近视者10人, 视力正常者28人。图10展示了测试期间志愿者使用认知阅读评估软件时的图片。



图10 志愿者测试软件图

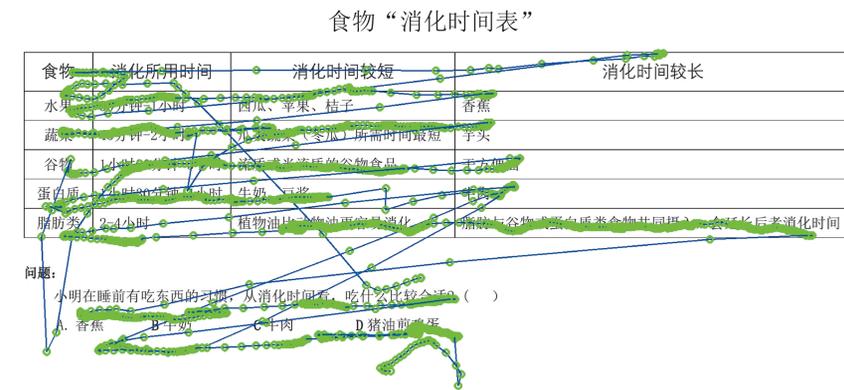
Fig. 10 Volunteer testing software diagram

在本次测试中, 测试文本选择的是一个图表阅读, 测试者需要在阅读后回答设置好的问题。在本次实验中, 通过眼动轨迹计算的眼动指标有眨眼次数(BC)、阅读总时长(RT)、首次看问题的时间(FT)、题目与图表的交替次数(AT)、眼跳次数(NOS)等指标。如表7所示, 对比了0号测试者的指标数据与均值数据。同时, 图11展示了0号测试者的眼动轨迹图和热力图。从表7中可以看出, 0号测试者的交替次数为0, 眼跳次数比均值少3.27次, 看到问题的时间较晚, 这表明0号测试者的阅读相对稳定, 没有出现较大的波动, 且在图表与问题之间没有进行交替加工, 这一点在图11中也得到了很好的体现。此外, 表7还显示, 0号测试者的眨眼次数较多, 比均值多23.53次, 说明0号测试者在阅读过程中进行了一定程度的加工。阅读时长比平均时长多出9.69s, 这表明0号测试者在阅读时非常认真。

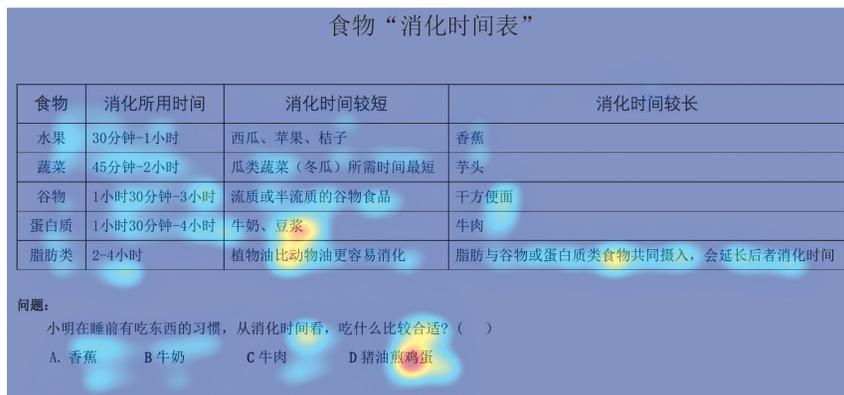
表7 眼动指标数据
Table 7 Eye movement index data

对比项	BC	RT/s	FT/s	AT	NOS
均值	54.47	52.51	22.17	7.92	6.27
0号	78.00	62.20	45.78	0	3.00

将得到的眼动指标通过特征融合策略将上述的指标抽象为阅读的稳定性、速度、理解能力、阅读技巧和元认知能力^[29]5个维度对所有测试者进行认知评价。如图12所示, 同样是抽取0号测试者5个认知维度的雷达图, 红色代表测试所有人的认知维度均值, 蓝色表示0号测试中的真实认知维度得分值, 可以明显看出, 0号测试者在稳定性、速度和阅读技巧上均低于平均水平, 说明0号测试者在阅读时容易受到外界的干扰, 导致稳定性差、阅读慢, 说明阅读较仔细。但是在理解能力和元认知能力上均高于平均值, 说明0号测试者对题目的理解又是较为出色的, 有较好的元认知能力。



(a) 眼动轨迹图



(b) 热力图

图 11 眼动轨迹图和热力图

Fig. 11 Eye tracking and heat mapping

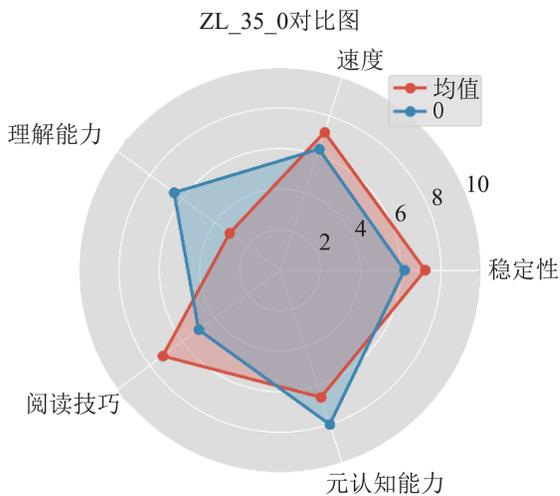


图 12 认知维度雷达图

Fig. 12 Cognitive dimension radar chart

5 结论

本文提出了一种基于机器视觉的高精度注意

力追踪系统, 包括低成本的图像采集装置和注意力追踪算法。通过引入低成本的图像采集装置, 克服了昂贵和庞大的科研级眼动仪限制, 拓展了眼动追踪应用场景。该系统采用精准的算法: 通过 FAST RCNN 模型定位眼部区域; 使用改进的 SSD 模型准确定位瞳孔位置; 采用 Eye-ORB 算法计算头部偏移量, 解决头部偏移对视线追踪的影响; 引入改进的 RBF 算法实现对人眼视线的精确追踪。除了视线追踪功能, 还设计了认知阅读评价交互软件, 利用视线坐标计算眼动指标, 通过心理学理论分析建模来评估思维认知能力。

总体而言, 该系统的设计与实现为阅读认知评估提供了一种全新的方法, 并为心理学和教育学领域的相关研究提供了参考和借鉴。尽管取得了一些成果, 但系统未在当前研究中充分考虑受试者视线溢出屏幕的扰动因素。然而, 这一问题

将成为未来研究的重点。开发创新的方法, 以过滤和处理这些干扰信息。通过深入研究并应用多模态融合的视线追踪技术, 提高眼动跟踪的精确性, 并确保该系统在各种复杂应用场景下得到有效应用。

参考文献:

- [1] Dong Wentao, Yang Lin, Gravina R, et al. ANFIS Fusion Algorithm for Eye Movement Recognition via Soft Multi-functional Electronic Skin[J]. *Information Fusion*, 2021, 71: 99-108.
- [2] Wang Xiaoming, Xiao Yineng, Deng Fangming, et al. Eye-movement-controlled Wheelchair Based on Flexible Hydrogel Biosensor and WT-SVM[J]. *Biosensors*, 2021, 11(6): 198.
- [3] Tesfamikael H H, Fray A, Mengsteab I, et al. Simulation of Eye Tracking Control Based Electric Wheelchair Construction by Image Segmentation Algorithm[J]. *Journal of Innovative Image Processing*, 2021, 3(1): 21-35.
- [4] Zhu Ji, Yang Hua, Liu Nian, et al. Online Multi-object Tracking With Dual Matching Attention Networks[C]// *Computer Vision- ECCV 2018*. Cham: Springer International Publishing, 2018: 379-396.
- [5] 靳冰凌, 张震, 张子耀. 基于视觉的驾驶员疲劳特征提取方法[J]. *计算机技术与发展*, 2018, 28(11): 193-197.
Jin Bingling, Zhang Zhen, Zhang Ziyao. A Driving Fatigue Feature Extraction Method Based on Vision[J]. *Computer Technology and Development*, 2018, 28(11): 193-197.
- [6] 陈加徐, 陆永华, 赵采仪, 等. 一种优化的自然光源下眼动视觉测量方法[J]. *应用科技*, 2021, 48(4): 61-67.
Chen Jiayu, Lu Yonghua, Zhao Caiyi, et al. An Optimized Visual Measurement Method for Eye Movement Under Natural Light Source[J]. *Applied Science and Technology*, 2021, 48(4): 61-67.
- [7] 宋琴琴, 杨国平. 基于视频图像的人眼动态追踪[J]. *计算机与数字工程*, 2021, 49(7): 1457-1459, 1474.
Song Qinqin, Yang Guoping. Human Eye Dynamic Tracking Based on Video Images[J]. *Computer & Digital Engineering*, 2021, 49(7): 1457-1459, 1474.
- [8] Wang Xiaoming, Zhao Xinbo, Zhang Yanning. Deep-learning-based Reading Eye-movement Analysis for Aiding Biometric Recognition[J]. *Neurocomputing*, 2021, 444: 390-398.
- [9] Brand J, Diamond S G, Thomas N, et al. Evaluating the Data Quality of the Gazeport GP3 Low-cost Eye Tracker When Used Independently by Study Participants [J]. *Behavior Research Methods*, 2021, 53(4): 1502-1514.
- [10] Liu Wei, Anguelov D, Erhan D, et al. SSD: Single Shot MultiBox Detector[C]// *Computer Vision-ECCV 2016*. Cham: Springer International Publishing, 2016: 21-37.
- [11] Qassim H, Verma A, Feinzimer D. Compressed Residual-VGG16 CNN Model for Big Data Places Image Recognition[C]// *2018 IEEE 8th Annual Computing and Communication Workshop and Conference (CCWC)*. Piscataway, NJ, USA: IEEE, 2018: 169-175.
- [12] Howard A, Sandler M, Chen Bo, et al. Searching for MobileNetV3[C]// *2019 IEEE/CVF International Conference on Computer Vision (ICCV)*. Piscataway, NJ, USA: IEEE, 2019: 1314-1324.
- [13] Ma Dan, Lai Huicheng. Remote Sensing Image Matching Based Improved ORB in NSCT Domain[J]. *Journal of the Indian Society of Remote Sensing*, 2019, 47(5): 801-807.
- [14] Zhu Juntao, Gong Chaofei, Zhao Miaoxing, et al. Image Mosaic Algorithm Based on PCA-ORB Feature Matching [J]. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 2020, XLII-3/W10: 83-89.
- [15] 王素琴, 施文豪, 李兆歆, 等. 无人机航拍视频中的车辆检测方法[J]. *系统仿真学报*, 2018, 30(7): 2776-2786.
Wang Suqin, Shi Wenhao, Li Zhaoxin, et al. Vehicle Detection Method in UAV Aerial Video[J]. *Journal of System Simulation*, 2018, 30(7): 2776-2786.
- [16] 柳有权, 裴雪, 李婉, 等. 基于邻近目标置信度评估的视觉目标跟踪与定位[J]. *系统仿真学报*, 2020, 32(7): 1294-1300.
Liu Youquan, Pei Xue, Li Wan, et al. Visual Tracking and Localization Based on Confidence Evaluation of Adjacent Targets[J]. *Journal of System Simulation*, 2020, 32(7): 1294-1300.
- [17] Li Xianguang, Ren Chao, Zhang Tengxu, et al. Unmanned Aerial Vehicle Image Matching Based on Improved RANSAC Algorithm and Surf Algorithm[J]. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 2020, XLII-3/W10: 67-70.
- [18] He Qingfeng, Shahabi H, Shirzadi A, et al. Landslide Spatial Modelling Using Novel Bivariate Statistical Based Naïve Bayes, RBF Classifier, and RBF Network Machine Learning Algorithms[J]. *Science of the Total Environment*, 2019, 663: 1-15.
- [19] Villanueva A, Ponz V, Sesma-Sanchez L, et al. Hybrid Method Based on Topography for Robust Detection of Iris Center and Eye Corners[J]. *ACM Transactions on*

- Multimedia Computing, Communications, and Applications, 2013, 9(4): 25.
- [20] Martinikorena I, Cabeza R, Villanueva A, et al. Introducing I2head Database[C]//Proceedings of the 7th Workshop on Pervasive Eye Tracking and Mobile Eye-Based Interaction. New York, NY, USA: Association for Computing Machinery, 2018: 1.
- [21] Jesorsky O, Kirchberg K J, Frischholz R W. Robust Face Detection Using the Hausdorff Distance[C]//Audio-and Video-based Biometric Person Authentication. Berlin, Heidelberg: Springer Berlin Heidelberg, 2001: 90-95.
- [22] Cai Haibin, Liu Bangli, Ju Zhaojie, et al. Accurate Eye Center Localization via Hierarchical Adaptive Convolution [C]//Proceedings of the 29th British Machine Vision Conference. Durham, UK: BMVA Press, 2018: 11504341.
- [23] Xia Yifan, Yu Hui, Wang Feiyue. Accurate and Robust Eye Center Localization via Fully Convolutional Networks[J]. IEEE/CAA Journal of Automatica Sinica, 2019, 6(5): 1127-1138.
- [24] Choi J H, Il Lee K, Kim Y C, et al. Accurate Eye Pupil Localization Using Heterogeneous CNN Models[C]//2019 IEEE International Conference on Image Processing (ICIP). Piscataway, NJ, USA: IEEE, 2019: 2179-2183.
- [25] Lee K I, Jeon J H, Song B C. Deep Learning-based Pupil Center Detection for Fast and Accurate Eye Tracking System[C]//Computer Vision-ECCV 2020. Cham: Springer International Publishing, 2020: 36-52.
- [26] Xiao Feng, Huang Kejie, Qiu Yue, et al. Accurate Iris Center Localization Method Using Facial Landmark, Snakuscle, Circle Fitting and Binary Connected Component[J]. Multimedia Tools and Applications, 2018, 77(19): 25333-25353.
- [27] 胡艳红, 魏江, 梅少辉. 基于瞳孔角膜反射技术的视线估计方法[J]. 计算机工程与应用, 2018, 54(14): 7-10, 18.
Hu Yanhong, Wei Jiang, Mei Shaohui. Gaze Estimation Method Based on Pupils and Corneal Reflection Technique[J]. Computer Engineering and Applications, 2018, 54(14): 7-10, 18.
- [28] 李擎, 胡京尧, 迟健男, 等. 视线追踪中一种新的由粗及精的瞳孔定位方法[J]. 工程科学学报, 2019, 41(11): 1484-1492.
Li Qing, Hu Jingyao, Chi Jiannan, et al. A New Pupil Localization Method From Rough to Precise in Gaze Tracking[J]. Chinese Journal of Engineering, 2019, 41(11): 1484-1492.
- [29] Akturk A O, Sahin I. Literature Review on Metacognition and Its Measurement[J]. Procedia-social and Behavioral Sciences, 2011, 15: 3731-3736.