

基于 AR-SSVEP 和 YOLOv3 的时敏目标识别方法

马留洋, 胡争争, 栗武华

(中国电子科技集团公司第二十七研究所, 河南 郑州 450047)

摘要: 针对目标跟踪过程中目标身份(ID)跳变而影响时敏目标识别的问题,提出了一种融合增强现实技术(AR)、稳态视觉诱发电位(SSVEP)和 YOLOv3 的人在回路的“检测-决策”时敏目标识别方法(AR-SSVEP-YOLOv3)。利用目标感知模块获取前端场景视频,并通过增强现实眼镜实时呈现,YOLOv3 算法完成场景视频中敏感目标检测,AR-SSVEP 脑电处理模块解析受试者的脑电数据,在 ID 变化过程中对时敏目标进行识别。对比分析时敏目标的识别正确率,AR-SSVEP-YOLOv3 时敏目标识别方法相比 YOLOv3 算法识别正确率平均提升了 40% 左右,相比 YOLOv3-Sort 算法平均提升了 15% 左右。实验结果表明:AR-SSVEP-YOLOv3 时敏目标识别方法可以降低目标 ID 跳变对时敏目标识别的影响,提升人机交互能力和时敏目标识别正确率。

关键词: 增强现实; 人工智能; 时敏目标; 目标检测; 稳态视觉诱发电位; 目标识别

中图分类号: R318.04; TP391.4

文献标志码: A

doi: 10.13705/j.issn.1671-6833.2025.01.017

现有时敏目标识别技术研究主要基于“检测-跟踪-决策”环路,重点关注有限窗口内发现、定位、识别和瞄准的目标,对视频连续帧信息进行检测、分析、评估、排序,识别出特定目标^[1]。国内外相关研究主要从算法优化角度提升检测器和跟踪器的精度。Gurkan 等^[2]为解决有遮挡目标跟踪问题,集成粒子跟踪和深度检测器优势,提出了一种基于检测的跟踪(tracking-by-detection, TBD)方法。Chen 等^[3]融合相关滤波器和粒子滤波器,提高了相机快速运动和目标部分遮挡下目标跟踪的鲁棒性,从算法优化角度提升目标检测算法和目标跟踪算法的精度,可以实现对特定目标的识别。然而,在复杂环境下,特别是目标完全遮挡后重新出现,以及相似目标交互重叠干扰等情况,目标跟踪算法可能对目标与 ID 的关联不准确,存在目标 ID 频繁变换的问题^[4],影响目标识别准确率。

目前,时敏目标识别判定并没有一个完全可靠的算法模型,对于时敏目标识别结果仍需要操作人员最终确认。如何将人的感知、判断和直觉等能力引入时敏目标识别环路,进行半结构化决策^[5]是未来时敏目标识别的一个可行方案。脑机

接口(brain-computer interface, BCI)技术是一种建立大脑和外部设备通信渠道的方式,其在大脑意图解析和人机交互方面的突破性进展能够为目标识别过程提供新的思路,特别是 SSVEP,可以用来辅助实现目标识别。Arpaia 等^[6]设计 AR-BCI 机器人移动控制模式,辅助受试者实现不依赖屏幕与机器人进行自然交互。黄章瑞^[7]设计了 AR-BCI 脑机融合交互系统,初步验证了 AR-SSVEP 技术在虚拟和真实两种情境驱动交互的可行性。蒋永玉^[8]将增强现实技术引入 SSVEP-BCI 系统,验证了 AR-SSVEP 技术在假手脑控的异步交互能力。基于 SSVEP 的 AR-BCI^[9-10]仍处于初步探索阶段,多用于控制和静态目标检测。

本文设计了人在回路的“检测-决策”时敏目标识别系统,提出了一种基于 AR-SSVEP 脑电信息和 YOLOv3 的时敏目标识别方法,可以弱化目标识别与目标跟踪的关联关系,利用 AR 眼镜叠加场景视频,呈现视觉刺激闪烁块,通过深度学习的分布式运算,提取并检测场景中的目标信息,建立刺激频率与目标 ID 的关联关系,融合受试者的稳态视觉诱发电位信号,实现指挥员直接参与时敏目标之间的识别过程。

收稿日期:2024-07-15;修订日期:2024-07-30

作者简介:马留洋(1990—),男,河南南阳人,中国电子科技集团公司第二十七研究所工程师,主要从事工程脑机接口技术、人机混合技术智能增强技术等研究,E-mail:mliuyang1119@163.com。

1 技术方案

1.1 时敏目标识别系统

人在回路的“检测-决策”时敏目标识别系统包括目标感知、目标检测和目标识别 3 个部分,主要由前端摄像头感知识别模块、网络传输设备、基于 AR 眼镜的刺激闪烁和视频叠加显示模块、基于 YOLOv3 的敏感目标检测模块、脑电采集及识别解析模块组成,如图 1 所示。受试者脑电数据获取采用博睿康公司的 32 通道无线采集设备,采集枕区 Pz、P3、P4、PO3、PO4、O1、Oz 和 O2 等 8 个导联电极^[11]的 SSVEP 脑电信号。

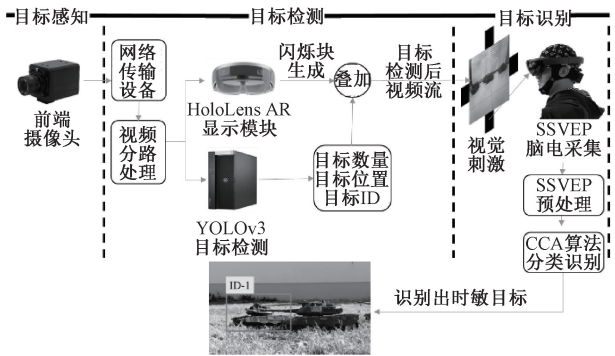


图 1 人在回路的“检测-决策”时敏目标识别系统
Figure 1 A time-sensitive target recognition system for human in the loop based on "detection-decision"

基于深度学习神经网络的时敏目标检测模块和脑电分析模块在数据处理计算机(带独立显卡 GTX 1060Ti 及以上)中运行,通过 YOLOv3 目标检测算法和脑电分析算法,实现前端场景中目标的检测和受试者 SSVEP 识别解析,其中导联分布如图 2 所示。AR 眼镜叠加前端视频和刺激闪烁块,呈现 SSVEP 视觉刺激。利用感知识别模块获取前端场景视频,通过视频分流处理,将视频分别发送给 AR 眼镜和目标检测设备,YOLOv3 目标检测算法实时检测目标,通过局域网将目标框的位置信息($x, y, width, height$)和数量等信息输入 AR 眼镜,与眼镜内置视觉刺激叠加呈现。

目标感知模块采集远端目标场景的视频流信息,对视频图像帧进行下采样,并将所采集的视频流信息分别传输至 AR 眼镜和目标检测模块。目标检测模块中运行 YOLOv3 目标检测算法,对目标图像帧进行缩放处理,将图像大小调整为 416×416 ,将此图像传递给主干网络 Darknet-53 进行卷积处理,输出 13×13 个网格,通过预测框和真实框的交并比(IOU)确定检测目标,输出目标检测结果。将目标视频帧中的目标信息(编号、位置和数量)发送给

AR 眼镜,并在眼镜中对应位置绘制检测信息,同时生成 SSVEP 视觉刺激闪烁块。如果目标 ID 变化,对应编号改变会同步更新到 AR 眼镜中。受试者根据视频态势,注视时敏目标对应的刺激块,由 AR 眼镜内置的视线追踪模块判断是否注视刺激块,并通过 CCA 算法提取受试者的脑电对应频率特征。输出注视频率对应时敏目标,完成时敏目标的识别。整体流程如图 3 所示。

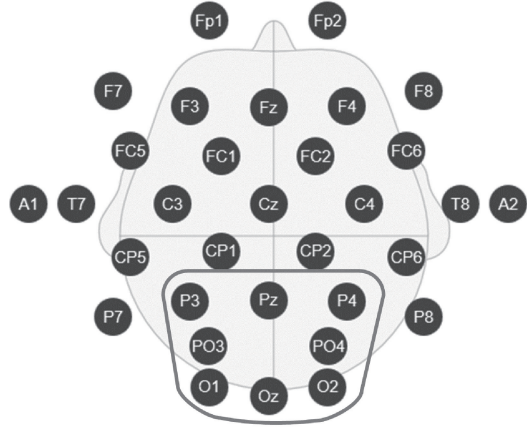


图 2 脑电采集导联位置分布图
Figure 2 Location distribution map of EEG acquisition channels

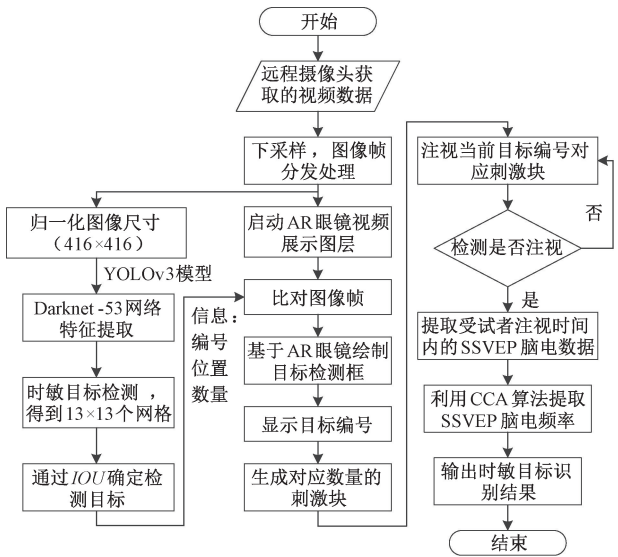


图 3 基于 AR-SSVEP-YOLOv3 的时敏目标识别选择流程图
Figure 3 Flowchart of time-sensitive target recognition selection based on AR-SSVEP-YOLOv3 method

1.2 基于 YOLOv3 算法的时敏目标检测

目前 YOLOv5、YOLOv8 等最新的稳定算法对隐藏、遮挡目标的检测速度和检测效果较好,但是在相似度较高的目标遮挡情况下仍然存在 ID 跟踪错误、ID 跳变^[12-13]。为研究 AR-SSVEP 技术在目标 ID 跳变过程中对时敏目标的识别效果,需要考虑目标在

视野中的停留时间,保证 SSVEP 脑电数据时长,因此选择 YOLOv3 算法。

YOLOv3 目标检测算法是一种端到端的回归学习方法,利用网格划分方式处理输入目标图像,如果一个网格中存在待检测目标,则在此网格中检测目标,输出相应的目标信息和目标置信度^[13-14]。每个网格输出 n 个 bounding box,每个 box 中含有以下参数:中心点坐标 (x, y) 、宽度 ($width$)、高度 ($height$) 和置信概率 ($confidence$) :

$$confidence = \Pr(\text{Object}) \times IOU_{\text{pred}}^{\text{truth}} \quad (1)$$

式中: $\Pr(\text{Object})$ 为此网格中存在目标的概率,如果网格内存在目标物体,则 $\Pr(\text{Object})$ 的值为 1,反之则为 0; $IOU_{\text{pred}}^{\text{truth}}$ 用来表征目标检测过程中预测框的精度。

YOLOv3 利用 Darknet-53 网络作为主干网络,对图像数据进行卷积运算,提取目标特征。通过残差联接进行深层次的采样,有效解决计算过程中的梯度爆炸和梯度消失问题。通过对目标图像多次压缩提取特征,提供 3 个不同尺度的有效特征层输出,然后利用上采样和特征拼接方式^[13-14],对低层特征表征能力进行增强,检测不同尺度目标。利用阈值筛选和非极大值抑制方法,选择置信度值最高的检测框输出。

1.3 SSVEP 脑电数据处理

SSVEP 数据采集过程中会混入干扰信号。首先,利用陷波滤波器消除 50 Hz 工频干扰;其次,设计二阶的巴特沃斯带通滤波器提取信号中 1~30 Hz 信号,并对数据的漂移情况进行基线矫正,受试者 SSVEP 的 4 种频率数据的频谱如图 4 所示;最后,根据目标出现时间,按照实际情况分段截取数据,进行识别。

典型相关分析 (canonical correlation analysis, CCA) 是计算两组变量间相关性的一种多元统计学分析方法。CCA 算法将多维数据分别降维到 1 维,然后分析相关性,通过典型变量的相关系数,定量表达两组变量的相关性大小。CCA 算法可以有效提取脑电信号谐波中的有效信息,对于多通道 SSVEP 脑电数据处理具有较好的识别效果。其原理是通过构建特定频率的正弦余弦参考信号,度量 SSVEP 脑电数据与参考信号间的相关系数,根据最大相关原理,寻找与受试者 SSVEP 信号对应的视觉刺激频率^[15-16]。

在 CCA 算法处理过程中,会把原始数据进行标准化,得到均值为 0 而方差为 1 的数据 \mathbf{X} 和 \mathbf{Y} ,再将 \mathbf{X} 和 \mathbf{Y} 多维矩阵进行线性变换投影到一维,对应的

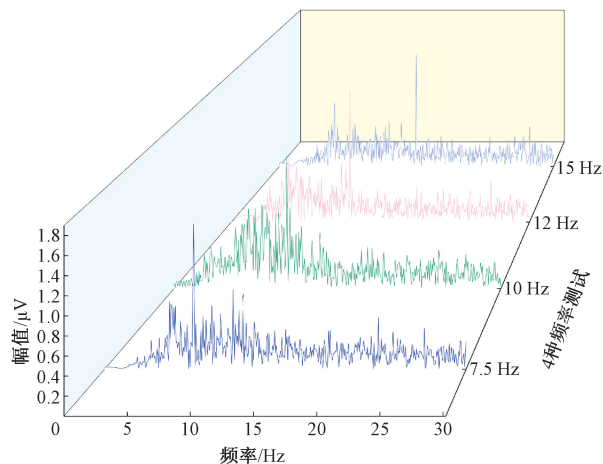


图 4 4 种刺激频率诱发的 SSVEP 平均频谱图

Figure 4 Average spectrogram of SSVEP evoked by four stimulation frequencies

线性变换向量分别为 \mathbf{a} 、 \mathbf{b} , 投影后得到的一维向量分别为 $\mathbf{X}' = \mathbf{a}^T \mathbf{X}$ 、 $\mathbf{Y}' = \mathbf{b}^T \mathbf{Y}$ 。

CCA 算法的核心就是获得最大化 \mathbf{X}' 和 \mathbf{Y}' 的相关系数^[17], 计算式为

$$\rho_{\mathbf{X}', \mathbf{Y}'} = \max \frac{\text{cov}(\mathbf{X}', \mathbf{Y}')}{\sqrt{D(\mathbf{X}')} \sqrt{D(\mathbf{Y}')}}; \quad (2)$$

$$\text{cov}(\mathbf{X}', \mathbf{Y}') = \text{cov}(\mathbf{a}^T \mathbf{X}, \mathbf{b}^T \mathbf{Y}) = \mathbf{a}^T \mathbf{E}(\mathbf{X} \mathbf{Y}^T) \mathbf{b}; \quad (3)$$

$$D(\mathbf{X}') = D(\mathbf{a}^T \mathbf{X}) = \mathbf{a}^T \mathbf{E}(\mathbf{X} \mathbf{X}^T) \mathbf{a}. \quad (4)$$

\mathbf{X} 和 \mathbf{Y} 的均值均为 0, 将 $\text{cov}(\mathbf{X}, \mathbf{Y})$ 记为 $\mathbf{S}_{\mathbf{XY}}$, 则

$$D(\mathbf{X}) = \text{cov}(\mathbf{X}, \mathbf{X}) = \mathbf{E}(\mathbf{X} \mathbf{X}^T). \quad (5)$$

可得

$$\begin{aligned} \rho(\mathbf{X}', \mathbf{Y}') &= \max \frac{\text{cov}(\mathbf{X}', \mathbf{Y}')}{\sqrt{D(\mathbf{X}')} \sqrt{D(\mathbf{Y}')}} \\ &= \max \frac{\mathbf{a}^T \mathbf{E}(\mathbf{X} \mathbf{Y}^T) \mathbf{b}}{\sqrt{\mathbf{a}^T \mathbf{E}(\mathbf{X} \mathbf{X}^T) \mathbf{a}} \sqrt{\mathbf{b}^T \mathbf{E}(\mathbf{Y} \mathbf{Y}^T) \mathbf{b}}} \\ &= \max \frac{\mathbf{a}^T \mathbf{S}_{\mathbf{XY}} \mathbf{b}}{\sqrt{\mathbf{a}^T \mathbf{S}_{\mathbf{XX}} \mathbf{a}} \sqrt{\mathbf{b}^T \mathbf{S}_{\mathbf{YY}} \mathbf{b}}}. \end{aligned} \quad (6)$$

可以看出, 分子分母成比例增加时, $\rho_{\mathbf{X}', \mathbf{Y}'}$ 不变, 因此最大化 $\rho_{\mathbf{X}', \mathbf{Y}'}$ 可以等价于最大化分子, 将上述问题转化为

$$\begin{cases} \text{argmax} & \mathbf{a}^T \mathbf{S}_{\mathbf{XY}} \mathbf{b}; \\ \text{s. t.} & \mathbf{a}^T \mathbf{S}_{\mathbf{XX}} \mathbf{a} = 1; \\ & \mathbf{b}^T \mathbf{S}_{\mathbf{YY}} \mathbf{b} = 1. \end{cases} \quad (7)$$

利用拉格朗日函数, 将优化目标最大化:

$$\begin{aligned} J(\mathbf{a}, \mathbf{b}) &= \mathbf{a}^T \mathbf{S}_{\mathbf{XY}} \mathbf{b} - \frac{\lambda}{2} (\mathbf{a}^T \mathbf{S}_{\mathbf{XX}} \mathbf{a} - 1) - \\ &\quad \frac{\theta}{2} (\mathbf{b}^T \mathbf{S}_{\mathbf{YY}} \mathbf{b} - 1). \end{aligned} \quad (8)$$

分别对参数 \mathbf{a} 、 \mathbf{b} 求导, 利用限制条件, 可得 $\lambda =$

$$\theta = \mathbf{a}^T \mathbf{S}_{XY} \mathbf{b}。$$

由此可得需要优化的目标即为求解拉格朗日系数 λ, θ , 整合后得

$$\mathbf{S}_{XX}^{-1} \mathbf{S}_{XY} \mathbf{S}_{YY}^{-1} \mathbf{S}_{YX} \mathbf{a} = \lambda^2 \mathbf{a}。 \quad (9)$$

基于式(9)可求解最大相关系数 λ , 只需要对矩阵 $\mathbf{S}_{XX}^{-1} \mathbf{S}_{XY} \mathbf{S}_{YY}^{-1} \mathbf{S}_{YX}$ 进行特征分解, 找出最大的特征值取平方根即可^[18]。

利用 CCA 算法处理 SSVEP 脑电数据时, 截取单个 trail 数据为 $\mathbf{X}(N \times T)$, 其中, N 为采用的电极通道数, 取值为 8; $T = d \times f_s$ 为待处理数据采样点数, d 为数据时长; f_s 为采样频率。数据 \mathbf{X} 每一行是一个通道的数据。基于诱发范式选用的刺激频率构建参考信号 \mathbf{Y} 。利用刺激频率的基波频率和谐波频率的正余弦信号构建参考信号 \mathbf{Y} , 参考信号 \mathbf{Y} 的表达式^[10, 18-19]如下所示:

$$\mathbf{Y} = \begin{bmatrix} \sin & 2\pi f_i T \\ \cos & 2\pi f_i T \\ \vdots & \\ \sin & 2\pi n f_i T \\ \cos & 2\pi n f_i T \end{bmatrix}。 \quad (10)$$

式中: n 为谐波信号, 本文取 $n=3$; f_i 表示刺激频率, $i=1, 2, 3, 4$, 用于构建与 7.5、10、12、15 Hz 这 4 个刺激频率对应的参考信号 \mathbf{Y} 。

基于 CCA 算法原理, 首先, 计算 \mathbf{X} 和 \mathbf{Y} 的均值分别为 $\bar{\mathbf{X}}(N, 1)$, $\bar{\mathbf{Y}}(2n, 1)$; 然后, 计算 \mathbf{X} 的协方差 \mathbf{S}_{XX} 、 \mathbf{Y} 的协方差 \mathbf{S}_{YY} 、 \mathbf{X} 和 \mathbf{Y} 的协方差 \mathbf{S}_{XY} 、 \mathbf{Y} 和 \mathbf{X} 的协方差 \mathbf{S}_{YX} 。最后, 依次求得单个 trail 的 SSVEP 数据 $\mathbf{X}(N \times T)$ 与 4 个参考信号的相关系数, 最大的相关系数对应的参考信号频率即为 SSVEP 刺激频率, 至此完成 SSVEP 刺激频率的注视检测^[20]。

2 实验与结果分析

2.1 实验过程

本实验中采用微软 HoloLens AR 眼镜呈现诱发 SSVEP 的视觉刺激闪烁块^[21], 闪烁块的数量根据 YOLOv3 算法识别到的目标数量自动调整, 本实验设置最大目标数量为 4 个, 选取 YOLOv3 中 score 最大的 4 个目标, 若数量少于 4 个, 则减少刺激块的个数。由于 AR 设备的处理能力有限, 无法直接运行 YOLOv3 算法, 将摄像头获取的前端场景视频流分别传送给 AR 设备和 YOLOv3 目标检测系统。YOLOv3 算法实现目标检测, 并向 AR 设备上传目标位置信息包括中心点坐标 (x, y) 、宽度 $(width)$ 、高度 $(height)$ 和置信概率 $(confidence)$ 以及目标 ID 和数

量等信息。AR 设备显示视频流, 并根据位置和数量绘制目标框和显示 SSVEP 刺激闪烁块, 基于 AR 眼镜眼动追踪功能, 捕捉受试者的视觉注视情况, 通过视觉注视启动 SSVEP 数据解析程序, 将闪烁块与场景中的目标 ID 建立关联, 解析受试者脑电信号, 完成时敏目标识别。刺激界面如图 5 所示。

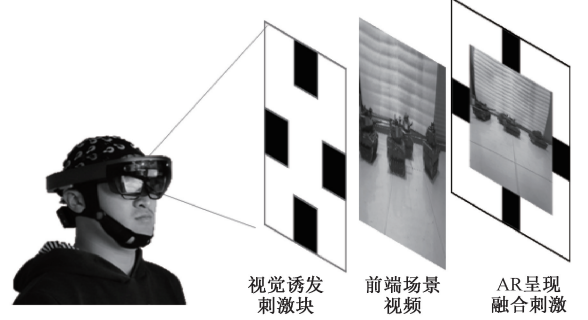


图 5 基于 AR 的 SSVEP 融合视觉刺激界面

Figure 5 AR based SSVEP visual stimulation interface

视野中央位置呈现远程场景视频信息, 在视频区域上、右、下、左分布黑白色交替闪烁块, 依次编号为 1、2、3 和 4, 视觉刺激频率分别为 7.5、10、12 和 15 Hz。本文共 14 名受试者参与, 其中 9 名男性、5 名女性, 受试者年龄 22~28 周岁, 均为身体健康且视力正常或者校正之后正常的成年人。这些受试者中, 受试者 2、受试者 9~13 等 6 人之前有参加基于 SSVEP 的实验, 其余受试者是初次参加。

实验包括离线测试和在线测试。离线测试过程中, 增强现实眼镜中不叠加视频数据, 随机提示目标 ID 编号, 显示时间为 4 s, 受试者根据提示注视相应的刺激块, 一次提示结束后, 界面呈现“0”, 受试者休息 2 s。一轮实验中每个 ID 编号出现 5 次, 每轮实验结束后, 受试者休息 10 min。

在线测试中, 选择 4 辆型号相同、颜色相同的坦克模型作为测试目标。4 辆坦克在外观上没有差异 (根据实验需要人为标记为 A、B、C、D), 利用摄像头捕捉远程目标场景, 通过无线传输模块回传场景视频。实验过程中, 4 辆坦克模型按一定路线运动, 设定目标均出现在摄像头视野内, 一辆坦克运动过程中将炮筒瞄向无人平台, 此坦克目标作为时敏目标。例如, 坦克 A 和 B 在相同方向上运动, 运动过程中出现重叠, 重叠时坦克 B 被 A 遮挡, 此时调整坦克 B 炮口朝向无人平台, 也即实验场景中目标 B 作为时敏目标, 然后坦克 B 超过 A。设定坦克 A 和 B 重叠时间 2 s, 坦克 B 超过 A 时间 2 s, 如图 6 所示, 重复上述过程, 一轮实验完成 20 次重叠分离过程, 每名受试者完成 5 轮实验。当观测 ID 丢失时, 受试者选择其他目标为时敏目标, 进行识别, 完成实验。

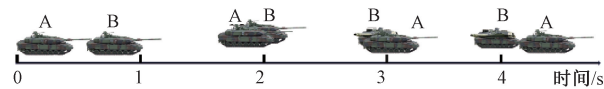


图 6 基于 AR-SSVEP 和 YOLOv3 方法的
时敏目标识别时序图

Figure 6 Temporal sequence diagram of time-sensitive
object recognition based on AR-SSVEP-YOLOv3 method

受试者实验过程中发现时敏目标,通过注视对应编号的刺激块,标记脑电数据有效起始时刻,启动 SSVEP 数据解析程序,分别记录受试者注视 0.5~4 s 的数据,进行特征提取和分类。实验过程中受试者只需要关注时敏目标在重叠分离时的目标 ID,利用 AR-SSVEP-YOLOv3 系统识别出时敏目标。

2.2 数据处理长度对目标选择的影响分析

分析 14 名受试者离线实验脑电数据,分别截取 0.5、1.0、1.5、2.0、⋯、4.0 s 数据,利用 CCA 算法进行特征提取和识别,统计不同受试者不同 SSVEP 数据长度时的分类识别率结果,结果如图 7 所示。

从图 7 可以看出,截取数据长度在 1 s 以下时,系统识别正确率会降低,漏检严重。数据长度在 2 s 时受试者的脑电数据分类识别率大部分在 85% 以上,数据长度在 3~4 s 时识别正确率更高,实验结果与文献[19]的结果一致,但相比于 2 s 时并未有明显的提升,时长增加会影响识别的时效性。6 名受试者前期参与过 SSVEP 实验,进行过 AR-SSVEP-YOLOv3 时敏目标识别实验测试,更加适应实验过程,注释视觉刺激时视线稳定,脑电数据识别正确率高于其他受试者。因此,根据实验结果,截取 2 s 的脑电数据已经满足实验应用需求。本结果为受试者进行时敏目标在线识别提供支撑。

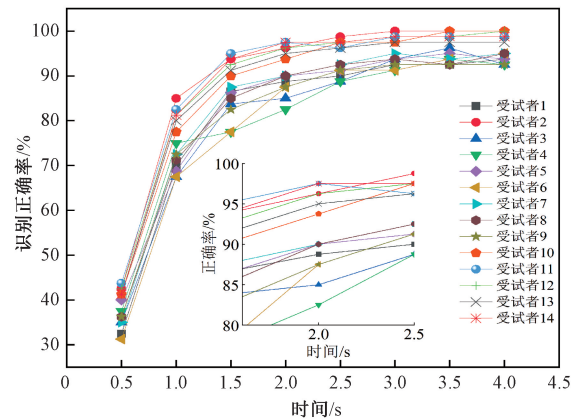


图 7 不同受试者不同数据长度下的识别正确率
Figure 7 Recognition accuracy for different subjects
and data lengths

2.3 受试者对不同刺激频率的敏感性分析

针对受试者离线训练实验数据,进一步分析了

14 名受试者对目标频率的敏感性。表 1 统计了 14 名受试者针对 4 种频率的识别正确率情况,其中,第一列 1~14 代表受试者的编号信息,第一行为目标频率,其他值代表了受试者编号在某个目标频率下的识别正确率,最后一行为各目标频率下的平均识别正确率。

表 1 不同目标频率特征识别正确率

| 受试者 | 正确率 | | | |
|-----|-------------------|------------------|------------------|------------------|
| | $f=7.5\text{ Hz}$ | $f=10\text{ Hz}$ | $f=12\text{ Hz}$ | $f=15\text{ Hz}$ |
| 1 | 0.85 | 0.90 | 0.95 | 0.80 |
| 2 | 1.00 | 0.95 | 0.90 | 1.00 |
| 3 | 0.80 | 0.90 | 0.85 | 0.85 |
| 4 | 0.95 | 0.90 | 0.90 | 0.85 |
| 5 | 0.90 | 0.90 | 0.95 | 1.00 |
| 6 | 0.90 | 0.85 | 1.00 | 0.95 |
| 7 | 0.85 | 0.90 | 0.95 | 0.90 |
| 8 | 0.95 | 0.90 | 1.00 | 0.85 |
| 9 | 0.95 | 0.95 | 1.00 | 0.90 |
| 10 | 0.90 | 0.90 | 0.95 | 0.95 |
| 11 | 0.95 | 1.00 | 0.95 | 1.00 |
| 12 | 1.00 | 0.95 | 0.90 | 1.00 |
| 13 | 1.00 | 0.9 | 1.00 | 0.95 |
| 14 | 0.85 | 0.85 | 0.95 | 0.90 |
| 均值 | 0.92 | 0.91 | 0.95 | 0.92 |

从表 1 可以看出,AR-SSVEP-YOLOv3 实验中平均识别正确率达到 91% 以上,符合 SSVEP 实验结果。同一测试者对不同频率的闪烁刺激产生的响应,以及同一刺激频率对不同的测试者产生的响应,与文献[19,21]中结论一致,符合个体之间的差异性特征,与生物学人体各异的观点相符,验证了基于 AR-SSVEP-YOLOv3 目标识别技术原理的可行性。

2.4 实验结果及分析

本实验重点关注目标重叠过程中目标检测和 ID 变化情况,将目标遮挡分为无遮挡、1/4 遮挡、半遮挡、3/4 遮挡、全遮挡 5 种情况,统计目标检测情况和 ID 变化情况。为了研究基于 AR-SSVEP-YOLOv3 目标识别技术的有效性,采用 YOLOv3 和 YOLOv3-Sort 多目标跟踪算法,实时在线检测和跟踪场景中的目标。选取文献[22]提出的部分常用评价指标进行评估,表 2 所示为实验过程中目标检测结果,其中 $MOTA$ 为多目标跟踪准确率, $ID-S$ 为目标身份切换次数, MT 为成功跟踪目标百分比, ML 为丢失目标百分比。

在实验过程中,目标 ID 变化主要集中在目标 A 和目标 B 重叠面积大于 3/4 的时间段内,特别是在

目标完全重叠过程中;由于摄像头角度和视野影响,目标检测算法在目标重叠过程中存在误判和漏检现象,会发生短暂的 ID 丢失和变化现象,如图 8(b)~(e)所示。根据表 2 结果,本文选择 YOLOv3 目标检测算法可有效检测出场景中的目标,利用 Sort 算法可以实现目标 ID 跟踪,在一定程度上减少 ID 切换次数,确保输出目标的 ID 具有一定的稳定性,可以用来关联 SSVEP 刺激块,保证 AR-SSVEP-YOLOv3

表 2 目标检测跟踪结果

| Table 2 Object detection and tracking results | | | | | |
|---|------|--------|------|------|-----------------------------|
| 算法 | ID-S | MOTA/% | ML/% | MT/% | 帧率/ (帧·s ⁻¹) |
| YOLOv3 | 542 | 95.7 | 30 | 78.7 | 15 |
| YOLOv3-Sort | 250 | 96.3 | 15 | 82.5 | 13 |

目标识别系统刺激有效。对比 YOLOv3 算法、YOLOv3-Sort 算法以及本文方法,分析时敏目标在重叠分离过程的识别结果。如图 8 所示,目标 ID 变化主要集中在重叠面积大于 3/4 的时间段内,受试者确定敏感目标的时间为坦克转动炮口的时刻。据此,选择目标重叠至目标分离的 2 s 时间的实验数据进行分析。基于 YOLOv3 算法和 YOLOv3-Sort 算法的目标跟踪过程,主要以目标的 ID 作为识别的结果,如果分离前后时敏目标的 ID 一致,则识别正确,反之错误。AR-SSVEP-YOLOv3 方法不受目标 ID 变化影响,所以 SSVEP 识别结果和时敏目标的 ID 一致时,则为识别正确。分别统计每位受试者 100 次目标识别的识别正确率,结果如图 9 所示。

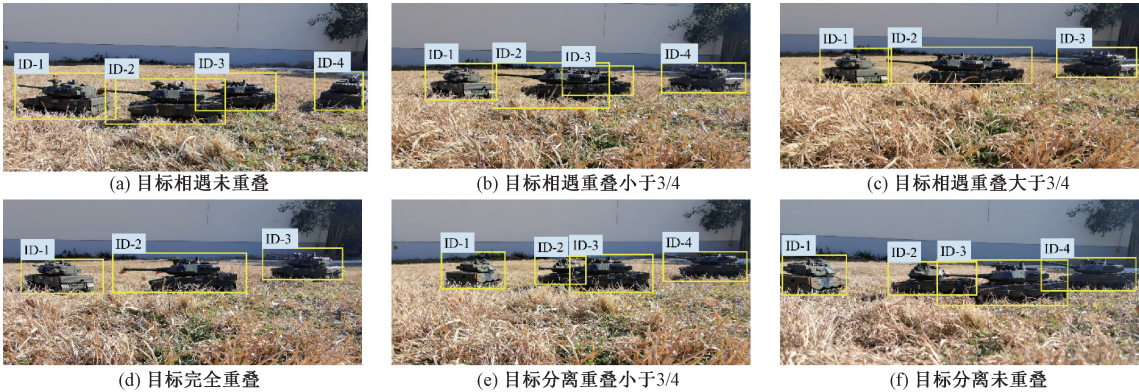


图 8 重叠过程中目标检测结果图

Figure 8 Result image of object detection during overlapping process

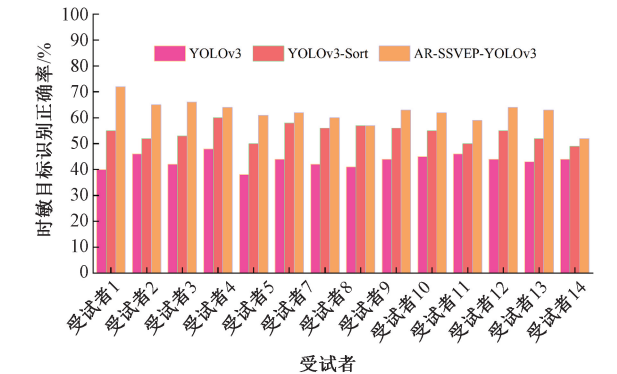


图 9 不同受试者目标识别正确率结果

Figure 9 Results of target recognition accuracy for different subjects

单独使用 YOLOv3 目标检测算法在目标重叠分离过程中目标 ID 变化次数多,直接影响目标识别正确率,加入 Sort 目标跟踪算法优化后,ID 变化现象明显改善。利用本文提出的方法,相比 YOLOv3 目标检测识别过程,其对时敏目标识别正确率平均提升了 40% 左右,相比 YOLOv3-Sort 方法平均提升了 15% 左右。在目标重叠分离过程中,目标 ID 跳变相

对频繁,受试者没有足够的时间注视刺激块,使得目标识别存在难度。此外,受试者在 AR 视觉刺激下对部分频率敏感性存在差异,因此整体目标识别正确率低于 75%,但从整体识别结果上可以看出,AR-SSVEP-YOLOv3 时敏目标识别方法在一定程度上可以很好地补充目标跟踪过程,为目标识别提供有力的技术支持。

3 结论

本文重点探究一种基于 AR-SSVEP-YOLOv3 技术的时敏目标识别方法,借助 AR-SSVEP 技术和 YOLO 目标检测技术,搭建人在回路的“检测-决策”时敏目标识别系统,完成 SSVEP 敏感目标识别测试。设计 4 种频率的刺激诱发实验,采集 14 名受试者的脑电数据,通过 CCA 算法对采集的脑电信号进行特征提取和分类识别,验证了本系统的可行性和合理性;通过 4 个相同目标的实时在线测试,利用 AR-SSVEP-YOLOv3 目标识别方法,实现相同目标重叠分离过程中时敏目标的识别。实验结果表明,

本文所提方法可以有效应对目标 ID 变化造成的目标识别正确率低的问题,为复杂环境下时敏目标识别提供了一种新的技术途径。随着脑-机接口技术、目标检测技术的进步,后续可以提高 SSVEP 脑电处理速度和识别正确率,融合 YOLOv5、YOLOv8 等最新稳定的目标检测算法,提升 AR-SSVEP 的目标识别效率,进一步提高人机混合效率和目标识别正确率,拓展本技术在地面人群及车辆识别、空中无人机标定、海上特定船只跟踪等不同场景中的应用。

参考文献:

[1] 吴晗, 张志龙, 李楚为, 等. 时敏目标的类型与瞄准点识别算法[J]. 航空兵器, 2022, 29(2): 24-29.
WU H, ZHANG Z L, LI C W, et al. Recognition algorithm for types and aiming points of the time-sensitive target[J]. Aero Weaponry, 2022, 29(2): 24-29.

[2] GURKAN F, GUNSEL B, OZER C. Robust object tracking via integration of particle filtering with deep detection[J]. Digital Signal Processing, 2019, 87: 112-124.

[3] CHEN J, LIN Y M, HUANG D T, et al. Robust tracking algorithm for infrared target via correlation filter and particle filter [J]. Infrared Physics & Technology, 2020, 111: 103516.

[4] 周孟然, 李学松, 朱梓伟, 等. 井下矿工多目标检测与跟踪联合算法[J]. 工矿自动化, 2022, 48(10): 40-47.
ZHOU M R, LI X S, ZHU Z W, et al. A joint algorithm of multi-target detection and tracking for underground miners[J]. Journal of Mine Automation, 2022, 48(10): 40-47.

[5] 黄培荣, 宋剑, 李楠. 打击时敏目标的决策任务层次分解[J]. 计算机与数字工程, 2014, 42(10): 1905-1907.
HUANG P R, SONG J, LI N. Decision-making task hierarchical decomposition of blowing time sensitive target [J]. Computer & Digital Engineering, 2014, 42(10): 1905-1907.

[6] ARPAIA P, DURACCIO L, MOCCALDI N, et al. Wearable brain - computer interface instrumentation for robot-based rehabilitation by augmented reality [J]. IEEE Transactions on Instrumentation and Measurement, 2020, 69(9): 6362-6371.

[7] 黄章瑞. 情境驱动的 AR-BCI 脑机交互方法研究[D]. 郑州: 郑州大学, 2021.
HUANG Z R. Research on situations drive brain-computer interaction based on AR-BCI [D]. Zhengzhou: Zhengzhou University, 2021.

[8] 蒋永玉. 假手脑控的增强现实视觉刺激范式及其异步应用系统研究[D]. 西安: 西安交通大学, 2022.
JIANG Y Y. Research on augmented-reality visually stimulus paradigm and its asynchronous BCI system for prosthetic hands[D]. Xi'an: Xi'an Jiaotong University, 2022.

[9] FANG B, DING W L, SUN F C, et al. Brain-computer interface integrated with augmented reality for human-robot interaction[J]. IEEE Transactions on Cognitive and Developmental Systems, 2023, 15(4): 1702-1711.

[10] ZHANG R, CAO L J, XU Z X, et al. Improving AR-SSVEP recognition accuracy under high ambient brightness through iterative learning[J]. IEEE Transactions on Neural Systems and Rehabilitation Engineering, 2023, 31: 1796-1806.

[11] 邹霞. 基于 SSVEP 的脑机接口系统设计与研究[D]. 南昌: 南昌大学, 2014.
ZOU X. Design and research of brain-computer interface system based on SSVEP [D]. Nanchang: Nanchang University, 2014.

[12] 刘辉, 刘鑫满, 刘大东. 面向复杂道路目标检测的 YOLOv5 算法优化研究 [J]. 计算机工程与应用, 2023, 59(18): 207-217.

LIU H, LIU X M, LIU D D. Research on optimization of YOLOv5 detection algorithm for object in complex road [J]. Computer Engineering and Applications, 2023, 59(18): 207-217.

[13] 院老虎, 常玉坤, 刘家夫. 基于改进 YOLOv5s 的雾天场景车辆检测方法 [J]. 郑州大学学报(工学版), 2023, 44(3): 35-41.
YUAN L H, CHANG Y K, LIU J F. Vehicle detection method based on improved YOLOv5s in foggy scene[J]. Journal of Zhengzhou University (Engineering Science), 2023, 44(3): 35-41.

[14] 李北明, 金荣璐, 徐召飞, 等. 基于特征蒸馏的改进 Ghost-YOLOv5 红外目标检测算法 [J]. 郑州大学学报(工学版), 2022, 43(1): 20-26.
LI B M, JIN R L, XU Z F, et al. An improved Ghost-YOLOv5 infrared target detection algorithm based on feature distillation [J]. Journal of Zhengzhou University (Engineering Science), 2022, 43(1): 20-26.

[15] HORII S, NAKAUCHI S, KITAZAKI M. AR-SSVEP for brain-machine interface: estimating user's gaze in head-mounted display with USB camera[C]//2015 IEEE Virtual Reality (VR). Piscataway: IEEE, 2015: 193-194.

[16] SI-MOHAMMED H, PETIT J, JEUNET C, et al. Towards BCI-based interfaces for augmented reality: feasibility, design and evaluation [J]. IEEE Transactions on Visualization and Computer Graphics, 2020, 26(3):

1608–1621.

[17] SHAO X H, LIN M X. Filter bank temporally local canonical correlation analysis for short time window SSVEPs classification[J]. Cognitive Neurodynamics, 2020, 14 (5) : 689–696.

[18] YIN X G, LIN M X. Multi-information improves the performance of CCA-based SSVEP classification[J]. Cognitive Neurodynamics, 2024, 18(1) : 165–172.

[19] ZHAO X C, LIU C Y, XU Z X, et al. SSVEP stimulus layout effect on accuracy of brain-computer interfaces in augmented reality glasses[J]. IEEE Access, 2020, 8 : 5990–5998.

[20] WANG K, ZHAI D H, XIONG Y H, et al. An MVMD-CCA recognition algorithm in SSVEP-based BCI and its application in robot control[J]. IEEE Transactions on Neural Networks and Learning Systems, 2022, 33 (5) : 2159–2167.

[21] 张力新, 张裕坤, 柯余峰, 等. 基于 HoloLens 的增强现实脑-机接口研究[J]. 中国生物医学工程学报, 2019, 38(1) : 51–58.

ZHANG L X, ZHANG Y K, KE Y F, et al. A study of argument reality based brain-computer interface (AR-BCI) in HoloLens[J]. Chinese Journal of Biomedical Engineering, 2019, 38(1) : 51–58.

[22] 何丽. 基于碎片化视频信息的关键目标人物检测和跟踪[D]. 南京: 南京邮电大学, 2023.

HE L. Key target person detection and tracking based on fragmented video information[D]. Nanjing: Nanjing University of Posts and Telecommunications, 2023.

Time-sensitive Target Recognition Method Based on AR-SSVEP and YOLOv3

MA Liuyang, HU Zhengzheng, LI Wuhua

(The 27th Research Institute of China Electronics Technology Group Corporation, Zhengzhou 450047, China)

Abstract: To address the problem of target identity (ID) fluctuation during target tracking, affecting the time-sensitive target recognition, an " detection-decision " time-sensitive target recognition method (AR-SSVEP-YOLOv3) was proposed which integrated augmented reality (AR) technology, steady state visual evoked potential (SSVEP) , and YOLOv3. The target perception module obtained the front-end scene video and presented it in real-time through an AR headset. The YOLOv3 algorithm completed the detection of sensitive targets in the scene video, and the AR-SSVEP EEG processing module decoded the EEG data of the subject during ID changes to identify time-sensitive targets. The correct recognition rate of time-sensitive targets was compared and analyzed, and the average improvement in the recognition accuracy of AR-SSVEP-YOLOv3 time-sensitive target recognition method compared to the YOLOv3 algorithm was about 40% , and the average improvement compared to the YOLOv3-Sort algorithm was about 15%. The experimental results showed that the AR-SSVEP-YOLOv3 time-sensitive target recognition method could reduce the influence of target ID fluctuation on time-sensitive target recognition and improve the human-computer interaction ability and the correct recognition rate of time-sensitive targets.

Keywords: augmented reality; artificial intelligence; time-sensitive target; object detection; steady state visual evoked potential; target recognition