



ARTICLE

Image Segmentation-P300 Selector: A Brain–Computer Interface System for Target Selection

Hang Sun, Changsheng Li* and He Zhang

Mechanical Engineering, Nanjing University of Science and Technology, Nanjing, 210018, China

*Corresponding Author: Changsheng Li. Email: 12012128@njust.edu.cn

Received: 22 January 2024 Accepted: 26 March 2024 Published: 15 May 2024

ABSTRACT

Brain–computer interface (BCI) systems, such as the P300 speller, enable patients to express intentions without necessitating extensive training. However, the complexity of operational instructions and the slow pace of character spelling pose challenges for some patients. In this paper, an image segmentation P300 selector based on YOLOv7-mask and DeepSORT is proposed. The proposed system utilizes a camera to capture real-world objects for classification and tracking. By applying predefined stimulation rules and object-specific masks, the proposed system triggers stimuli associated with the objects displayed on the screen, inducing the generation of P300 signals in the patient’s brain. Its video processing mechanism enables the system to identify the target the patient is focusing on even if the object is partially obscured, overlapped, moving, or changing in number. The system alters the target’s color display, thereby conveying the patient’s intentions to caregivers. The data analysis revealed that the self-recognition accuracy of the subjects using this method was between 92% and 100%, and the cross-subject P300 recognition precision was 81.9%–92.1%. This means that simple instructions such as “Do not worry, just focus on what you desire” effectively discerned the patient’s intentions using the Image Segmentation-P300 selector. This approach provides cost-effective support and allows patients with communication difficulties to easily express their needs.

KEYWORDS

Brain-computer interface; deep learning; electroencephalogram; P300; YOLOv7-mask

1 Introduction

Exploring the generation of signals within the brain and the human thoughts they encapsulate constitute an important research topic. Extensive research has been conducted to decipher the information contained in electroencephalograms (EEGs). These studies include the use of EEGs for emotion recognition [1], exploration of users’ mental states [2,3], person identification [4], and assessment of epilepsy patients’ conditions in clinical medicine [5]. Researchers are now attempting to reconstruct thoughts within people’s brains to assist those struggling to express intentions directly. The objective is to facilitate nonverbal information exchange between patients and the outside world by identifying EEGs. Furthermore, these studies aimed to develop brain–computer interface (BCI) systems for the direct manipulation of external devices, allowing the execution of specific actions or



information transmission. There are several methods for implementing BCI, among which motor imagination shows the most potential due to its extensive applicability, such as achieving control of vehicles [6], enabling movement with hexapod robots [7], and assisting patients in controlling wheelchairs for autonomous mobility [8]. However, given the complexity of human brain activities, the implementation of BCI systems based on motor imagination is challenging due to the limited number of recognizable EEGs, which are typically categorized into two [9] or six [10] types. Research on classifying more than 10 types of EEGs is scarce [11]. Despite the advent of advanced network models, the accuracy rate for motor imagination in most subjects does not exceed 90% [12]. As such, practical application requires extensive patient training, which could impose an insurmountable burden on patients.

In contrast to BCI systems that rely on motor imagination for subjective control of external devices, BCI systems implemented using steady-state visual evoked potentials (SSVEP) and the P300 paradigm offer the advantage of being usable with minimal or no training. These systems have demonstrated a significant improvement in signal classification accuracy. For instance, a high-speed brain speller capable of 40 classifications can achieve an accuracy of 97.5% by using SSVEP [13]. Moreover, a calibration-free SSVEP-BCI system with 160 classifications can maintain an accuracy of 90% [14]. Furthermore, SSVEP can be employed to control UAV swarms in a virtual reality environment [15] and assist people with disabilities in recreational activities [16]. However, SSVEP poses a challenging issue because they may trigger epileptic seizures in photosensitive individuals [17], rendering them unsuitable for long-term use [18]. Another important control approach in BCI systems is the P300 paradigm, with its foremost application being the P300-based BCI speller [19]. This system allows patients to spell out specific words [20] by analyzing the correlation between P300 signals and the timing of stimuli [21]. However, the effective use of the P300 speller requires that patients comprehend its operational process and possess an educational background and adequate cognitive capacity to complete the spelling task. To enhance the usability of the P300 speller for individuals who struggle, it can be improved by replacing flashing character blocks with patterns, achieving a relatively high accuracy rate [22]. Nevertheless, the predefined nature of these images limits their applicability. Expanding the array of selectable targets is essential for direct patient interaction with their surroundings. This can be accomplished through the use of camera systems for monitoring the external environment, enabling patients to select targets and perform operations [23]. However, in this methodology, selecting a target requires temporarily stopping the video stream. A system has been developed for wheelchair control in real-world settings [24]; however, this system is constrained by a fixed number of selectable targets and the problem of overlapping targets. An alternative approach involves utilizing Mask-RCNN [25] for object detection [26], which facilitates the control of robotic arms by focusing on object centroids. However, this technique restricts subjects to making selections based on the centroids of objects rather than their entire scope. Furthermore, constraints within the network architecture pose challenges in achieving real-time image segmentation. The literature on BCI systems reveals that challenges such as obstructions in real-world interactions and limitations in the number of detectable elements are yet to be fully addressed, with real-time processing efficacy falling short of expectations.

This study aimed to minimize the operational burden on patients while achieving real-time target selection and to increase the methods for patients to accurately express their own intentions. The key to this approach is the ability to select from an unrestricted range of target types and quantities without immobilizing these targets. The aim is to simplify patients' expression of intentions without the need for complex control mechanisms to provide them with a stress-free experience by encouraging them to "do not worry, simply focus on what you desire." This approach seeks to reduce the strain and

learning curve associated with BCI usage. The contributions and innovations of this paper are as follows:

- 1) For the first time, based on the YOLOv7-mask [27] and DeepSORT [28] algorithms, we achieved target selection by using P300 in real-world environments without any restrictions on categories or quantity and by expanding the number of objects that patients can choose.
- 2) In real-time video streams, we utilized object masks to implement object-triggered P300 stimuli. The presence of overlapping objects, dynamic real-time movement, or changes in the number of objects does not compromise the effectiveness of target selection, providing the possibility of synchronous interaction between patients and the outside world.
- 3) A real-time display paradigm that does not limit the number of targets has been designed to improve the accuracy of target selection and avoid interactions between different targets.

2 Method

2.1 Participants

In the current environment of the IS-P300 selector, directly requesting patient assistance for research would be burdensome. Therefore, the study recruited participants to assist in completing the research. The participants consisted of six healthy subjects aged between 21 and 25 years, and the standard deviation of the ages of the six subjects was 1.39, all without neurological or relevant medical conditions and with normal vision (either natural or corrected). Participants agreed to participate in this research. None of the participants had prior training or knowledge of BCI systems. Before collecting the EEG data, the experimenter explained the experimental protocols and objectives to the participants. Throughout the experiment, participants were encouraged to minimize physical activity, maintain composure, and immediately halt the session if they encountered any discomfort. The study was conducted in two phases over a period of 1 week: Sample collection and testing phases. During the sample collection phase, the participants were instructed to focus their gaze solely on the object relevant to the current experimental segment. The frequency with which this item produced stimuli during the experimental segment was documented to enhance the participants' concentration on the target object. In the testing phase, participants were directed to fixate their attention exclusively on any object of interest until the testing segment concluded.

In this study, the most commonly used international standard, the 10–20 lead system, was used to collect EEG data. The electrode layout of the system is depicted in Fig. 1. P300 signals were acquired using a 10-channel configuration: Fz, C3, Cz, C4, P3, P4, P7, P8, Pz, and Oz (highlighted in red in Fig. 1). The sample collection site is shown in Fig. 2. P300 signals were acquired with participants seated comfortably in a quiet environment to minimize unnecessary triggering of EEG activity by external stimuli.

Data collection and processing were performed using an Intel 5900 hs processor and an NVIDIA RTX 3070 laptop. The camera had a resolution of 1920×1080 , and images were displayed on a 17-inch monitor with a refresh rate of 165 Hz. Python was used for image and signal processing. EEG signals were sampled at 1000 Hz, and EEG signal transmission was accomplished using the TCP/IP protocol, with eight EEG information packets transmitted per second.

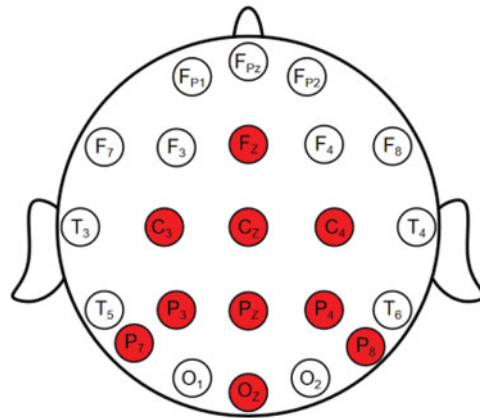


Figure 1: The employed EEG channel

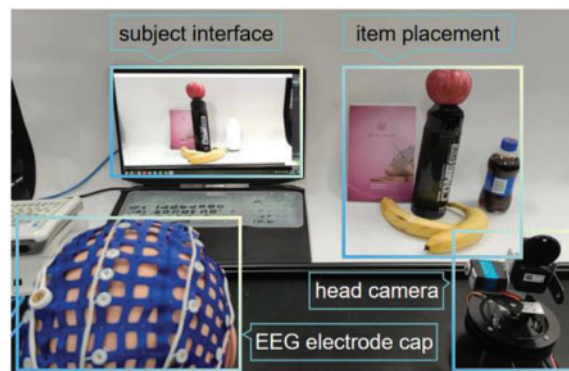


Figure 2: Testing site

2.2 Image Segmentation-P300 Selector

The implementation process of the image segmentation-P300 selector (IS-P300 selector) proposed in this study is depicted in Fig. 3.

The IS-P300 selector captures real-time environmental information by using a camera. The images are then fed into the executing program. There are three main aspects of data processing in the IS-P300 selector: Image processing, image display, and EEG processing. These are employed to acquire mask information for different objects, selectively adjust the RGB values of objects to trigger stimuli, and analyze the EEG information of subjects.

The image processing workflow is primarily based on YOLOv7 and DeepSORT for object category recognition and tracking. This approach allows real-time extraction of information about objects captured by the camera in both temporal and spatial domains, along with their respective masks. YOLOv7, a one-stage object detection algorithm, excels in accomplishing object classification and localization simultaneously and provides details such as object names and position information. Compared to other network architectures such as Fast-RCNN, YOLOv7 is superior in terms of image processing speed and delivers excellent real-time video processing performance. It includes various branches, such as the YOLO-pose branch for human pose detection and keypoint identification and the YOLO-mask branch for image segmentation. In this study, image segmentation was implemented

using YOLOv7-mask. Upon inputting an image, the program promptly returns information regarding the identified object’s category, location, and corresponding mask. The term “image mask” refers to the extraction of both the outline and internal details of an object, encompassing all pixel positions occupied by the object in the image. Even when certain parts of the target object are obscured by other objects, unobstructed sections can be extracted. This allows for the application of a single color value to cover the object in subsequent processing and achieve stimulus triggering. The image processing process is illustrated in Fig. 4.

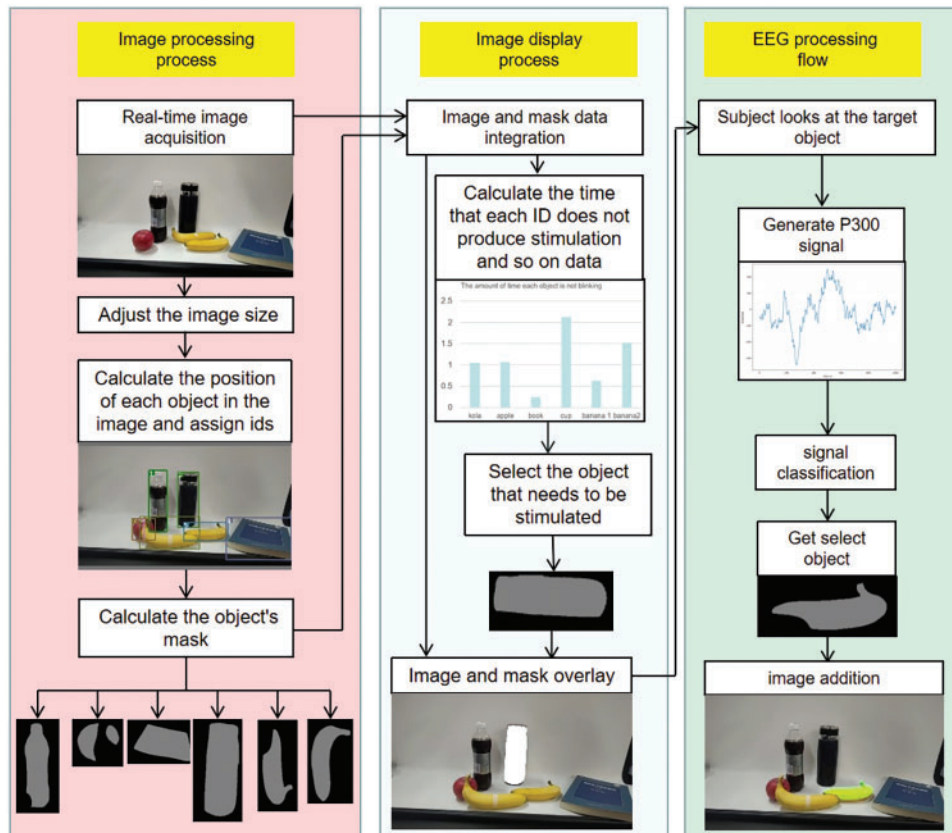


Figure 3: Implementation process of the IS-P300 selector

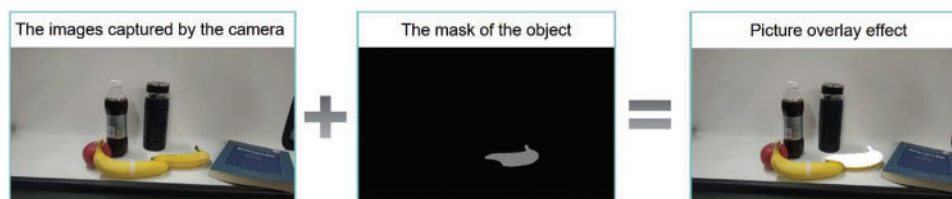


Figure 4: Image composition principle

However, due to the limitations inherent in the YOLOv7 mechanism, it is mainly used for recognizing individual images. YOLOv7 treats different images or sequential frames of a video as unrelated pictures, effectively breaking down and categorizing objects into separate points, with no

continuity between consecutive frames. This poses challenges for operations that require continuity on two images or across a video. To address this, we integrated DeepSORT for advanced object tracking in this study. DeepSORT calculates the degree of matching between objects in consecutive frames by using the Kalman filter and the Hungarian algorithm. In addition, YOLOv7 is utilized to ascertain the order in which objects appear, and each identified object is allocated an ID. This allows for efficient gathering of relevant information about the object across the time domain. Kalman filtering comprises a time update equation (Eq. (1)) employed for computing the current state variables and covariance estimates, as well as a measurement update equation (Eq. (2)) used for combining prior estimates with new measurements to construct an improved posterior estimate.

$$\begin{cases} \hat{x}_{\bar{k}} = A\hat{x}_{k-1} + Bu_{k-1} \\ P_{\bar{k}} = AP_{k-1}A^T + Q \end{cases} \quad (1)$$

$$\begin{cases} K_k = (P_{\bar{k}}H^T)/(HP_{\bar{k}}H^T + R) \\ \hat{x}_k = \hat{x}_{\bar{k}} + K_k(z_k - H\hat{x}_{\bar{k}}) \\ P_k = (I - K_kH)P_{\bar{k}} \end{cases} \quad (2)$$

where \hat{x}_k and \hat{x}_{k-1} represent the posterior state estimates at time $(k-1)$ and k , respectively, which are the updated results of the object; $\hat{x}_{\bar{k}}$ is the estimated value of the object's prior state. It is predicted by the state of the system and the system model at the previous time; P_{k-1} and P_k denote the posterior estimate covariance at time $(k-1)$; $P_{\bar{k}}$ is the covariance of the prior estimate at time k ; H is the state observation matrix, which can transform the state variables of the system into measurement variables; and K_k is the Kalman coefficient. When its value is close to 0, the posterior state estimator will trust the prediction of the model more; otherwise, it will favor the actual measurement. A is the state transition matrix, which can describe changes in the state of the system; Q and R represent the covariance of process excitation and the covariance of noise measurement, respectively; B serves as the transformation matrix that translates the input into a state suitable for computation; and $(z_k - H\hat{x}_{\bar{k}})$ denotes the residual value between the actual observation and the predicted observation. The fundamental objective is to predict the possible location of an object at the next moment by using a series of current motion variables of the object. The stability of tracked objects is ensured using the Hungarian algorithm, which identifies the most accurate matching effect through one-to-one comparisons between targets from the previous and current frames. This method incorporates a matching cascade and a confirmation feature to prevent the loss of IDs for obscured objects. During the image display phase, the program processes the masks of different objects according to the IDs recognized by DeepSORT to achieve the stimulation required by the IS-P300 selector.

Unlike other P300-based BCIs, such as the P300 speller, the IS-P300 selector allows the selection of nonfixed target objects. Users can freely adjust or move the arrangement order of similar or different items without the need for absolute stillness. The range of targets available for subjects to choose from includes various entities, such as persons, knives, and cats. The object recognition ability of the IS-P300 selector depends on the YOLOv7 training model. With adequate training data, the system can identify a broader array of categories. The targets that subjects can select are strictly related to the training level of the YOLO model. This significantly reduces the training demands on subjects, who merely need to focus on the specific target with whom they wish to interact. In addition, during the IS-P300 selector operation, it is possible to block certain fixed objects, preventing subjects from generating P300 signals associated with these objects.

The P300 paradigm triggers an EEG response in the parietal lobe of the brain with a latency of 300–600 ms after subjects receive low-probability anomalous stimuli. Continuous stimulation leads to a decrease in P300 wave amplitude, with the P300 waveform manifesting only at the moment of alteration. In the IS-P300 selector, the stimulation process involves covering the object with a white mask, followed by a period before restoring it to its original color, as depicted in Fig. 5.

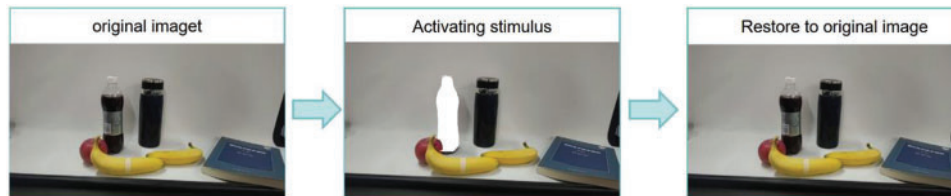


Figure 5: Stimulation activation process

The timing of when objects are covered with a mask and restored is an important research topic. This involves coordinating the frequency, timing, and sequence for stimulus generation across various objects. Although YOLOv7-mask and DeepSORT can achieve real-time processing, under the hardware conditions used in this study, each frame requires 60–120 ms for processing. The time required for image processing, coupled with data processing fluctuations, affects the effectiveness of the IS-P300 selector. Therefore, in this study, we increased the number of threads to achieve precise timing in stimulus generation. To improve program speed, we scaled down the original images before inputting them into YOLOv7-mask and DeepSORT to acquire target mask information. When triggering stimulus generation, the mask image was superimposed onto the original image to elicit stimuli.

The objects do not need to be arranged neatly, and there is no fixed number of items that can be selected. The types of objects can vary, their positions can dynamically change, and items can be added or removed. Unlike the traditional P300 speller, the timing and frequency of predetermining activation stimuli for each object are not practical. To ensure normal stimuli activation in different environments, seven rules were established in this study:

1. Upon the initial appearance of an object's ID, immediate activation occurs, and the corresponding stimulus is triggered.
2. The maximum stimulus interval for each object's ID is $T = N \times 1$ s, where N is the number of identifiable object types in the current video. In essence, if the time since the last activation stimulus for that object exceeds T , the stimulus is activated without delay to ensure that each item receives stimuli in a cyclical manner.
3. For each object ID, a minimum stimulus interval of $t = 0.8$ s is set. This means that if the time elapsed since the last activation stimulus for that object is less than 0.8 s, the object will not produce another stimulus. This is done to avoid the rapid flashing of objects, which could cause visual fatigue and unnecessary strain, while maintaining the integrity of single-stimulus data and reducing signal overlap caused by repeated stimuli.
4. If an ID activation stimulus occurs within 0.3 s, other objects will not receive stimulus activation.
5. If the last activation times for all objects are less than the maximum stimulus interval T and rule 4 is satisfied, all objects with activation times greater than t are selected. An ID is randomly selected, and the selected object is activated to complete the stimulus after introducing a random delay of t' ms (where $50 \text{ ms} < t' < 100 \text{ ms}$).

6. During each activation stimulus (duration: 80–120 ms), the object corresponding to the ID turns white, which generates the P300 signal. Following the completion of the activation stimulus, the object returns to its original color.
7. To minimize the possibility of erroneous P300 responses, only one object generated a stimulus at any given moment.

The aforementioned rules are designed to intentionally disrupt the timing of stimulus activations for different objects on the IS-P300 selector screen. This is achieved by varying the order of stimulus generation across different cycles and thus introducing a certain degree of chaos to prevent consecutive objects from generating stimuli at identical frequencies. This, in turn, reduces the risk of inaccurate P300 judgments. The stimulus generation for various objects in distinct cycles is depicted in Fig. 6; the figure serves as an example, and the order of stimulus generation for objects in different cycles may vary.

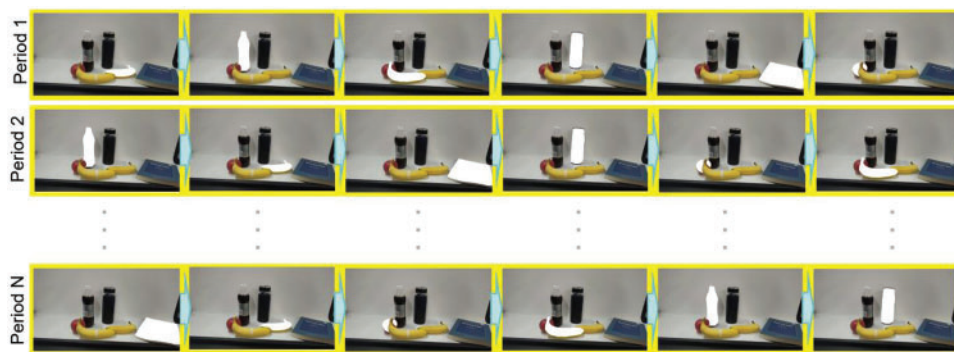


Figure 6: Sequence of object activation stimuli under different cycles

Due to blurred object boundaries and similar pixel colors in the IS-P300 selector, stimuli generated in subjects’ brains may lack clarity compared to the traditional P300 speller. To address this issue, the discrimination accuracy can be enhanced by increasing the frequency of signal overlays. The timestamp of when objects generate stimuli is used, and a 1000-ms data segment following this timestamp is extracted as a sample. This sample is then appended to the sample record list associated with the corresponding ID object. The process of adding timestamps and samples is depicted in Fig. 7.

Object ID number			
1	$t_1: 2.021s$	$t_1: 2.343s$		$t_1: 2.964s$
2	$t_2: 3.946s$	$t_2: 3.622s$		$t_2: 3.213s$
.	.	.		.
N	$t_n: 12.936s$	$t_n: 13.423s$	$t_n: 13.842s$

Figure 7: The method of recording timestamps and samples

In this study, once it was established that subjects could elicit P300 responses based on target objects, we employed the EEGNet [29] model for classifying these P300 signals. This requires acquiring samples and corresponding labels essential for model training. These labels encompass both the ID numbers of items and the timestamps of their generation. Considering that subjects may adjust their gaze and alter the placement of items during the experiment and that object IDs may not remain consistent across various phases, labels must be augmented with original image data to identify object IDs accurately. During the subjects' training, by gazing at the screen, the program superimposed the mask of the stimulated object onto the original image, resizing it to a 100×100 DPI image. As illustrated in Fig. 8, this method efficiently reduces the amount of necessary data storage and aids in post-experiment identification of objects that generate stimuli during a specified timeframe.



Figure 8: Label storage format

2.3 Experimental Phase

Before the experiment, participants chose a target to focus on. During the sample collection phase, they fixated their gaze on various objects displayed on the screen, such as a cola can, a water glass, and a banana with stickers. Throughout the experiment, staff continuously adjusted the positions of these objects or added and removed them, as depicted in Fig. 9.



Figure 9: Item sequence adjustment

The object positions were changed to simulate real-world interference and human activities. Using the operational mechanism of the IS-P300 selector, staff relocated objects; however, these objects could still trigger stimuli. This observation emphasizes that object movement has a minimal effect on the functionality of the IS-P300 selector. The stimuli produced during the movement of different objects are illustrated in Fig. 10.

During the movement of objects, the limbs of staff members may have entered the environment and served as objects of identification. In such cases, participants needed to maintain their focus on the target object to mitigate disruption caused by such movements. Although the relevant stimuli were generated during object movement, due to the program mechanism and hardware environment of the IS-P300 selector, the calculation time for the mask exceeded the video capture interval. To ensure smooth video display, the position triggering object stimuli were the location where the object existed in the previous frame, which might not coincide with the actual position of the object in the current frame.

However, even with this limitation, when the object moved at a slower speed, the placement of the mask could cover the majority of the object's area, thereby minimizing the effect on the functionality of the IS-P300 selector.

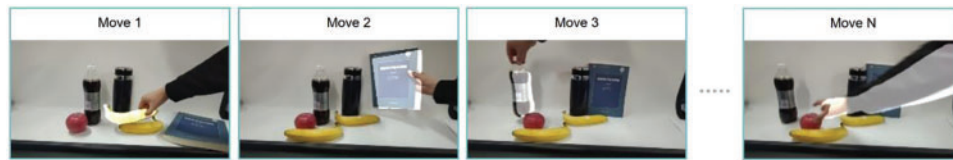


Figure 10: Dynamic movement of items

2.4 Classification Model

Based on the principles of P300 generation, a Butterworth filter (Eq. (3)) was employed to perform bandpass filtering on the EEG signals collected from various channels of the electrode cap in the frequency range of 0.1–20 Hz.

$$|H(\omega)|^2 = \frac{1}{1 + (\omega/\omega_c)^{2n}} = \frac{1}{1 + \varepsilon^2 (\omega/\omega_p)^{2n}} \quad (3)$$

where n represents the filter order, w_c denotes the cutoff frequency, w_p is the passband edge frequency, and $|H(\omega)|^2$ corresponds to the numerical value of the passband edge. Extensive research has been conducted on P300 classification methods. After a thorough comparison, the EEGNet approach was employed for sample processing in this study.

Based on the sample data obtained using the IS-P300 selector, positive samples of subjects receiving stimulation and negative samples of those not receiving stimulation were derived. In accordance with the P300 paradigm principles, the sample dataset was divided into positive and negative samples for training. A modified version of the EEGNet model, Model 1, was used to detect the presence of P300 signals; its architecture is illustrated in Fig. 11.

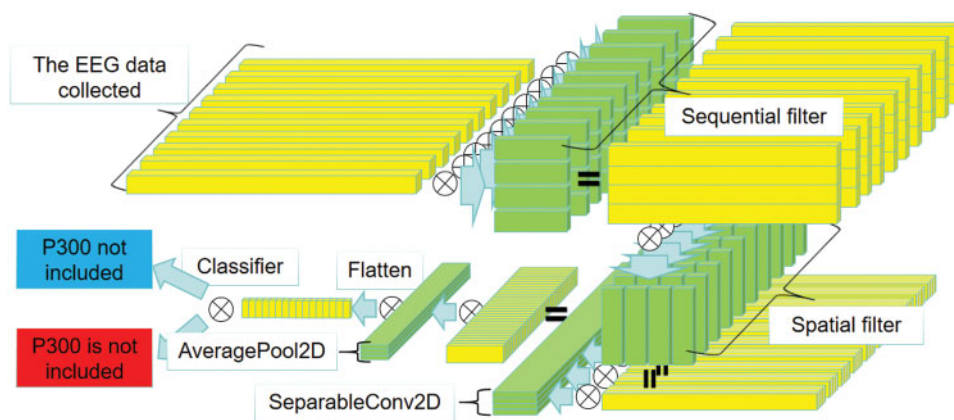


Figure 11: The EEGNet model after adaptive modifications

EEGNet is a compact, multiparadigm convolutional neural network (CNN) classification network that demonstrates excellent performance in multiclass EEG classification and maintains high accuracy for different subjects. In addition, it requires less computational power, thus avoiding excessive strain on hardware that could lead to abnormal execution during image display. Upon the activation of an object-triggered stimulus, the EEG data captured in the next 1000 ms are input into the classification network. The resulting output is used to determine whether subjects received the pertinent stimulus during that period and whether it generated a P300 signal.

The IS-P300 selector operates in a real-world environment. The mask engendered by the stimulus is pure white; as such, chromatic variations during the stimulus mechanism are less pronounced compared to the conventional P300 speller. Detecting subject concentration on a target with only a single stimulus signal is challenging and affects accuracy. To address this issue, P300 responses induced by stimuli at multiple time points for a given object are superimposed. The effects of EEG superposition for different numbers of instances are shown in [Fig. 12](#).

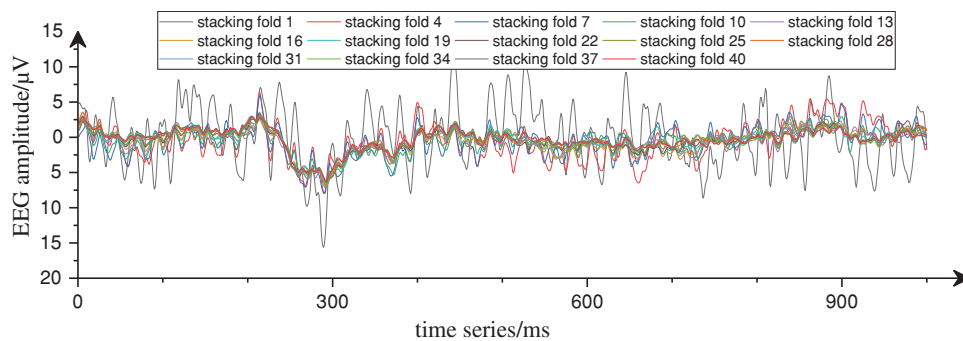


Figure 12: EEG superposition of signals for different numbers of instances

As P300 accumulates, differences in signals between positive and negative samples become increasingly apparent. However, distinguishing between positive and negative EEG samples may lead to the identification of two or more positive samples simultaneously, contrary to the principles of the P300 paradigm. This discrepancy indicates a misjudgment, wherein data from objects that did not evoke a P300 response in subjects are incorrectly classified as positive samples. In the traditional P300 speller with fixed target stimuli, statistical analysis can be performed on samples across one or several cycles. In this process, the letters or shapes fixated upon by subjects are assigned a value of 1, and the rest are assigned a value of 0; this results in the formation of a tensor that serves as the label. However, the IS-P300 selector, with its fluctuating target count, cannot guarantee that the objects the subjects focus on are indeed selectable during practical use. Therefore, only data deemed positive samples undergo binary classification training, and negative samples are obtained from the aberrant classification results of EEGNet. The structure of Model 2 used for training is illustrated in [Fig. 13](#).

Instead of simply classifying positive and negative samples, EEGNet processes input signals by concatenating 1000 ms of EEG data following stimuli from different targets and outputs a pointer to the target object most likely to generate the P300 signal. According to the EEGNet classification results, if P300 signals are detected in three or more targets within a specific time frame, Model 2 is cyclically employed. Targets are compared in pairs, and scores are assigned to the object IDs identified as targets. Following the binary classification of all relevant objects, the object with the highest score is selected as the indicated target, that is, the object with the most distinct P300 signal. This result is then communicated to the image display program, which retrieves the object based on the queried ID

number and highlights the identified target in semi-transparent green to allow others to determine the object that the subject is fixating upon. Alternatively, it can employ recorded audio to read out the name of the object category that the subject is fixating upon to assist others.

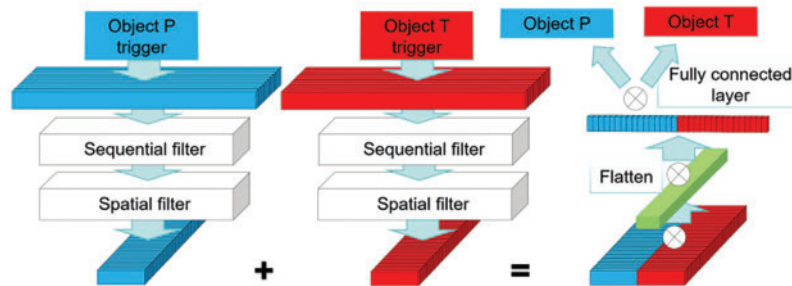


Figure 13: The binary classification model

3 Results

In the experimental setup, subjects were presented with six distinct items: Two bananas, a book, a water glass, a coke, and an apple. We meticulously counted the occurrences of each item within 10-s intervals to confirm the stimulus frequency for each specific object. The outcomes are systematically illustrated in Fig. 14.

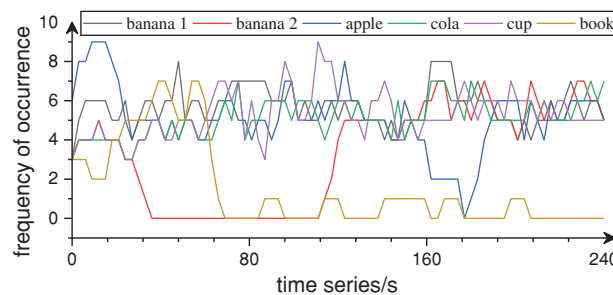


Figure 14: Frequency of dynamic appearances for each object

To more clearly indicate the position of the object and its movement over time, the position information of the object in one of the experiments is visualized, as shown in Fig. 15. The horizontal and vertical coordinates in the figure represent the position of the object's centroid on the screen, the depth coordinate represents the progression of time, different colors correspond to different objects, the displacement of the centroid represents the trajectory of the object, and the appearance and disappearance of the polyline represent the addition or removal of the object on the display screen.

Due to these constraints, objects generate stimuli randomly, and the frequency is not fixed. This variability can serve to differentiate the P300 information of subjects. During data processing, we applied bandpass filtering to the EEG signals. To capture P300 information contained within the samples, we opted for a single-sample duration of 1000 ms. Subsequently, we superposed the EEG data generated by subjects and conducted a comparative analysis between EEG datasets containing P300 signals and those without. The IS-P300 selector exhibited a more pronounced P300 response in subjects, as validated by the EEG comparison shown in Fig. 16.

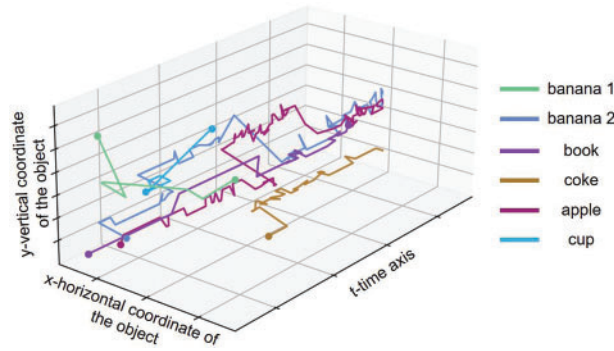


Figure 15: The trajectory of each object in the experiment

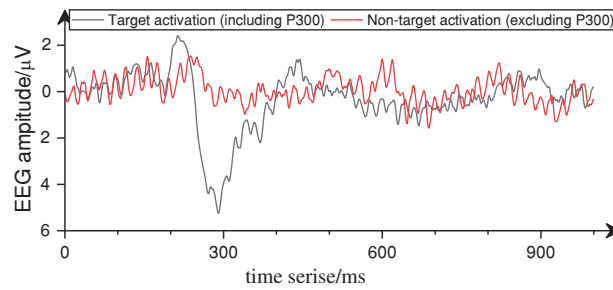


Figure 16: Comparison of superimposed P300 signals in subjects

After confirming that the IS-P300 selector distinctly exhibited a significant P300 component, we utilized Model 1 to discriminate between the EEG signals generated by subjects at different times. This process relies on object ID numbers to verify whether the object is the focal point of the subject’s gaze, which, in turn, generates the corresponding P300 signal. First, EEG samples resulting from a single stimulus event were analyzed. Subsequently, individual samples from each subject were utilized for training. The achieved accuracy rates are depicted in [Table 1](#).

Table 1: Classification accuracy for different subjects

Subject number	Subject 1	Subject 2	Subject 3	Subject 4	Subject 5	Subject 6
Recognition accuracy	83.02%	77.11%	83.33%	86.36%	80.11%	80.00%

To measure the impact of accuracy and recognition speed on BCIs, the commonly used evaluation metric in BCI, the information transfer rate (ITR), is used to evaluate the model. This allows for a comprehensive quantification of the amount of information transmitted per unit of time by the HCI system. The formula is as follows:

$$ITR = (\log_2 N_f + A_c \log_2 A_c + (1 - A_c) \log_2 ((1 - A_c)/(N_f - 1))) \times f_d \tag{4}$$

where N_f is the number of target recognition types, A_c is the target recognition accuracy, and f_d is the target selection time. Since the number of target recognition types in the study is not fixed and is affected by the types of targets appearing on the screen, objects are classified only as targets and nontargets, i.e., $N_f = 2$. According to the design rules of the paradigm, a flash is completed every

475 ms on average, so $f_d = (60/(0.475 \times N_f))$. The ITR of different subjects was compared, and the results are shown in [Table 2](#).

Table 2: The ITR of different subjects

Subject number	Subject 1	Subject 2	Subject 3	Subject 4	Subject 5	Subject 6	Average
ITR (bits/min)	21.65	14.14	22.09	26.85	17.70	17.56	20.00

For the majority of participants, Model 1 successfully distinguished between positive and negative samples generated by the IS-P300 selector. Although faster temporal resolution was found to be beneficial when using EEG data generated from a single stimulus for target identification, it was unable to consistently maintain high classification accuracy across all participants. In this study, the single-trial recognition accuracy ranged from 77.1% to 86.4%, which is less than satisfactory. To address this, we superposed 1000-ms EEG signals generated in response to stimuli for objects with consistent ID numbers and obtained the classification accuracy for n superimposed instances by averaging the EEG signals corresponding to a particular item over N consecutive trials. The variations in accuracy are illustrated in [Fig. 17](#).

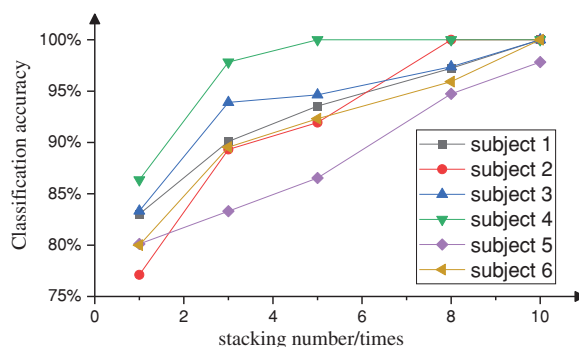


Figure 17: Relationship between signal superposition and accuracy

Increasing the value of N yielded a notable enhancement in classification accuracy but reduced the temporal resolution of EEG. The highest accuracy rate reached 100%. Thus, determining the optimal balance between the superposition count n and accuracy necessitates careful consideration of the specific context. Using Model 2, a binary classification assessment of EEGs from different objects at proximate time intervals ($n = 5$) was performed. The results are shown in [Table 3](#).

Table 3: Binary classification accuracy for different objects at similar time intervals

Subject number	Subject 1	Subject 2	Subject 3	Subject 4	Subject 5	Subject 6
Recognition accuracy	91.53%	96.15%	92.77%	98.36%	91.67%	94.44%

As can be observed from [Table 2](#), although the selected objects generated stimuli at similar times, there were differences in the timing of signal superposition. This generally makes the recognition accuracy rate above 91.5%. This variation in the information contained in EEG corresponding to different objects demonstrates that Model 2 can accurately distinguish objects of interest at close time intervals.

For those who are using the IS-P300 selector for the first time, they can use the data from other subjects as samples. For example, we assessed the accuracy of S6 by using samples from S1 to S5, that is, cross-subject training. As the subject's usage time increases and data gradually accumulate, their EEG can be mixed with the samples of other subjects and gradually completely replaced. The results of cross-subject training, full data training, and self-training are shown in [Table 4](#).

Table 4: Cross-subject target recognition accuracy

Subject number	Subject 1	Subject 2	Subject 3	Subject 4	Subject 5	Subject 6
Cross-Subject accuracy	85.34%	83.65%	92.11%	91.27%	83.65%	87.91%
Mixed-Subject accuracy	87.93%	86.15%	92.81%	98.12%	83.75%	87.95%
Self-Subject accuracy	93.55%	91.94%	94.94%	100.00%	86.54%	92.31%

The research results show that during the initial cross-subject training, the recognition accuracy rate is not as good as the self-training results, but the average accuracy rate can still be maintained above 86.6%. With the introduction of the subject's own data, the recognition accuracy rate improved until it was completely replaced by self-training data. This shows that the IS-P300 selector has good reference value for the recognition results of the target, proving the feasibility of this method.

4 Discussion

The use of P300 as a control input in BCI has reached a relatively advanced stage; P300 can be used for tasks such as text spelling with the P300 speller or as a button trigger for subjects' P300 signals. With the past limitations where static images were required to assist subjects in generating P300 signals, real-time image segmentation for dynamic objects presented a challenge. Consequently, dynamic objects were used to generate P300 signals. With progress in artificial intelligence, real-time image processing has become achievable. Compared to the majority of past P300 paradigms that could only recognize specific targets and integrate real-time image segmentation into P300 applications, the proposed system relieves patients from the burden of mastering complex control methods. This allows patients to dynamically select targets. Advising patients with the phrase "Do not worry, simply focus on what you desire" enables them to express their goals relatively accurately without causing fatigue or unnecessary stress to the patients. For caregivers, the placement of items in front of the camera is needed without the need for elaborate guidance or procedures, thus reducing learning pressure for both patients and caregivers and reducing system usage prerequisites. The IS-P300 system designed in this study is not limited to EEGNet; any P300 classification-capable method can be employed, thus expanding the applicability of P300 and promoting the development of more practical BCI applications.

Although the IS-P300 can help patients express their own intentions, there are certain limitations. For example, when there are too many selectable objects in the video, the stimulus generation cycle for the same object will significantly lengthen, meaning that the system needs more time to understand the patient's intention; YOLOv7 can identify objects with a high accuracy rate, but there may be situations where it cannot recognize, affecting the use effect of the IS-P300 selector; DeepSORT, as an object tracking method, however, rapid object deformation may also cause loss of tracking and reassignment of ID, which has a negative impact on the IS-P300 selector. These existing problems will, to some extent, reduce the use effect of the IS-P300 selector. In further research, we can try to use more advanced networks to improve object recognition accuracy and tracking effects, perhaps breaking

through the use limitations of the IS-P300 selector and better helping patients express their intentions. In addition, at the patient personal information level, there is a lack of direct use of the IS-P300 selector. Since patients cannot accurately express their intentions in the initial state, biological information leakage about EEG signals may occur when collecting patient EEG [30], and it is also possible that in the final stage, the intrusion of other people's EEG samples may lead to being exploited by hackers, causing serious consequences. These AI security issues need to be considered in the application of the IS-P300 selector, and methods such as the Friend-Guard adversarial noise design [31] can be used to effectively identify the patient's own EEG, prevent it from being used by others, effectively maintain the patient's biological information security, and provide security for the widespread application of BCI. These security precautions are also issues that the IS-P300 selector should consider during its future application.

5 Conclusion

The IS-P300 selector developed in this study employs YOLOv7-mask for the recognition and real-time segmentation of video images. It employs the DeepSORT algorithm to track and assign matching IDs to various objects. Following the research-defined rules, stimulus signals are promptly generated to cover targeted objects, evoking P300 responses when subjects fixate on them. Using deep learning networks and timestamps, the proposed system accurately identifies targets of interest and overlays them with a semi-transparent green shade. The data analysis revealed that it can yield a recognition precision of 92%–100% and a cross-subject P300 recognition precision of 81.9%–92.1%. In addition, we demonstrated that overlay-based P300 stimulation of images can be effectively recognized, irrespective of the number or type of target, motion status, or appearance and disappearance. Through this study, the application scope of P300 has been expanded, providing new research directions for future research.

Acknowledgement: This research is sponsored by Nanjing University of Science and Technology.

Funding Statement: The authors received no specific funding for this study.

Author Contributions: The authors confirm contribution to the paper as follows: Hang Sun, and Changsheng Li designed the IS-P300 selector. Hang Sun and He Zhang analyzed the data and drafted the manuscript. Changsheng Li and He Zhang critically revised the manuscript and contributed important. Hang Sun and Changsheng Li designed the experimental scheme and data collection. Changsheng Li and He Zhang made critical revisions to the original manuscript and made important contributions.

Availability of Data and Materials: The data collection code for this study can be obtained at SunAoMi/IS-P300-selector (<https://github.com/SunAoMi/IS-P300-selector>).

Conflicts of Interest: The authors declare that they have no conflicts of interest to report regarding the present study.

References

- [1] Y. DaŞDemİR and R. Özakar, "Affective states classification performance of audio-visual stimuli from EEG signals with multiple-instance learning," *Turk J. Electr. Eng. Comput. Sci.*, vol. 30, no. 7, pp. 2707–2724, 2022. doi: [10.55730/1300-0632.3964](https://doi.org/10.55730/1300-0632.3964).

- [2] A. Muraja-Murro *et al.*, “Forehead EEG electrode set versus full-head scalp EEG in 100 patients with altered mental state,” *Epilepsy Behav.*, vol. 49, pp. 245–249, 2015. doi: [10.1016/j.yebeh.2015.04.041](https://doi.org/10.1016/j.yebeh.2015.04.041).
- [3] H. Zeng, C. Yang, G. Dai, F. Qin, J. Zhang and W. Kong, “EEG classification of driver mental states by deep learning,” *Cogn. Neurodyn.*, vol. 12, no. 6, pp. 597–606, 2018. doi: [10.1007/s11571-018-9496-y](https://doi.org/10.1007/s11571-018-9496-y).
- [4] Z. A. A. Alyasseri *et al.*, “EEG feature fusion for person identification using efficient machine learning approach,” in *Proc. 2021 Palestinian Int. Conf. Inform. Commun. Technol. (PICICT)*, Gaza, Palestine, Sep. 28–29, 2021, pp. 97–102.
- [5] H. Chen and M. Z. Koubeissi, “Electroencephalography in epilepsy evaluation,” *Continuum*, vol. 25, no. 2, pp. 431–453, 2019. doi: [10.1212/CON.0000000000000705](https://doi.org/10.1212/CON.0000000000000705).
- [6] Y. Zhi and J. Xu, “Research on brain-computer interface system for vehicle control based on motion imagination,” in *Proc. 2020 2nd Int. Conf. Big Data Artif. Intell.*, Johannesburg, South Africa, Apr. 28–30, 2020, pp. 516–520.
- [7] T. Y. Mwata-Velu, J. Ruiz-Pinales, H. Rostro-Gonzalez, M. A. Ibarra-Manzano, J. M. Cruz-Duarte and J. G. Avina-Cervantes, “Motor imagery classification based on a recurrent-convolutional architecture to control a hexapod robot,” *Mathematics*, vol. 9, no. 6, pp. 606, 2021. doi: [10.3390/math9060606](https://doi.org/10.3390/math9060606).
- [8] T. Saichoo, P. Boonbrahm, and Y. Punsawad, “Investigating user proficiency of motor imagery for EEG-based BCI system to control simulated wheelchair,” *Sens.*, vol. 22, no. 24, pp. 9788, 2022. doi: [10.3390/s22249788](https://doi.org/10.3390/s22249788).
- [9] P. Arpaia *et al.*, “Multimodal feedback in assisting a wearable brain-computer interface based on motor imagery,” in *Proc. 2022 IEEE Int. Conf. Metrol. Extend. Real. Artif. Intell. Neural Eng. (MetroXRINE)*, Rome, Italy, Dec. 05, 2022, pp. 691–696.
- [10] P. Ofner, A. Schwarz, J. Pereira, and G. R. Müller-Putz, “Upper limb movements can be decoded from the time-domain of low-frequency EEG,” *PLoS One*, vol. 12, no. 8, pp. e0182578, 2017. doi: [10.1371/journal.pone.0182578](https://doi.org/10.1371/journal.pone.0182578).
- [11] J. H. Jeong *et al.*, “Multimodal signal dataset for 11 intuitive movement tasks from single upper extremity during multiple recording sessions,” *GigaScience*, vol. 9, no. 10, pp. giaa098, 2020. doi: [10.1093/gigascience/giaa098](https://doi.org/10.1093/gigascience/giaa098).
- [12] J. Shin and W. Chung, “Multi-band CNN with band-dependent kernels and amalgamated cross entropy loss for motor imagery classification,” *IEEE J. Biomed. Health Inform.*, vol. 27, no. 9, pp. 4466–4477, 2023. doi: [10.1109/JBHI.2023.3292909](https://doi.org/10.1109/JBHI.2023.3292909).
- [13] M. Nakanishi, Y. Wang, X. Chen, Y. T. Wang, X. Gao and T. P. Jung, “Enhancing detection of SSVEPs for a high-speed brain speller using task-related component analysis,” *IEEE Trans. Biomed. Eng.*, vol. 65, no. 1, pp. 104–112, 2017. doi: [10.1109/TBME.2017.2694818](https://doi.org/10.1109/TBME.2017.2694818).
- [14] Y. Chen, C. Yang, X. Ye, X. Chen, Y. Wang and X. Gao, “Implementing a calibration-free SSVEP-based BCI system with 160 targets,” *J. Neural Eng.*, vol. 18, no. 4, pp. 046094, 2021. doi: [10.1088/1741-2552/ac0bfa](https://doi.org/10.1088/1741-2552/ac0bfa).
- [15] T. Deng *et al.*, “A VR-based BCI interactive system for UAV swarm control,” *Biomed. Signal Process. Control*, vol. 85, no. 4, pp. 104944, 2023. doi: [10.1016/j.bspc.2023.104944](https://doi.org/10.1016/j.bspc.2023.104944).
- [16] G. A. M. Vasiljevic and L. C. de Miranda, “The CoDIS taxonomy for brain-computer interface games controlled by electroencephalography,” *Int. J. Human-Comput. Interact.*, vol. 2014, no. 4, pp. 1–28, 2023. doi: [10.1080/10447318.2023.2203006](https://doi.org/10.1080/10447318.2023.2203006).
- [17] R. S. Fisher, G. Harding, G. Erba, G. L. Barkley, and A. Wilkins, “Photic-and pattern-induced seizures: A review for the epilepsy foundation of America working group,” *Epilepsia*, vol. 46, no. 9, pp. 1426–1441, 2005. doi: [10.1111/j.1528-1167.2005.31405.x](https://doi.org/10.1111/j.1528-1167.2005.31405.x).
- [18] S. Ladouce, L. Darmet, J. J. Torre Tresols, S. Velut, G. Ferraro and F. Dehais, “Improving user experience of SSVEP BCI through low amplitude depth and high frequency stimuli design,” *Sci. Rep.*, vol. 12, no. 1, pp. 8865, 2022. doi: [10.1038/s41598-022-12733-0](https://doi.org/10.1038/s41598-022-12733-0).
- [19] O. Maslova *et al.*, “Non-invasive EEG-based BCI spellers from the beginning to today: A mini-review,” *Front. Hum. Neurosci.*, vol. 17, pp. 1216648, 2023. doi: [10.3389/fnhum.2023.1216648](https://doi.org/10.3389/fnhum.2023.1216648).

- [20] O. E. Korkmaz, O. Aydemir, E. A. Oral, and I. Y. Ozbek, "A novel probabilistic and 3D column P300 stimulus presentation paradigm for EEG-based spelling systems," *Neural Comput. Appl.*, vol. 35, no. 16, pp. 11901–11915, 2023. doi: [10.1007/s00521-023-08329-y](https://doi.org/10.1007/s00521-023-08329-y).
- [21] R. Ron-Angevin, Á. Fernández-Rodríguez, F. Velasco-Álvarez, V. Lespinet-Najib, and J. M. André, "Usability of three software platforms for modifying graphical layout in visual P300-based brain-computer interface," *Biomed. Signal Process. Control*, vol. 86, no. 4, pp. 105326, 2023. doi: [10.1016/j.bspc.2023.105326](https://doi.org/10.1016/j.bspc.2023.105326).
- [22] N. Rathi, R. Singla, and S. Tiwari, "A comparative study of classification methods for designing a pictorial P300-based authentication system," *Med. Biol. Eng. Comput.*, vol. 60, no. 10, pp. 2899–2916, Oct. 2022. doi: [10.1007/s11517-022-02626-9](https://doi.org/10.1007/s11517-022-02626-9).
- [23] J. Tang, Y. Liu, D. Hu, and Z. Zhou, "Towards BCI-actuated smart wheelchair system," *BioMed. Eng. OnLine*, vol. 17, no. 1, pp. 111, Aug. 20 2018. doi: [10.1186/s12938-018-0545-x](https://doi.org/10.1186/s12938-018-0545-x).
- [24] R. Pereira, A. Cruz, L. Garrote, G. Pires, A. Lopes and U. J. Nunes, "Dynamic environment-based visual user interface system for intuitive navigation target selection for brain-actuated wheelchairs," in *Proc. 2022 31st IEEE Int. Conf. Robot Hum. Interact. Commun. (RO-MAN)*, Napoli, Italy, Aug. 29–Sep. 02, 2022, pp. 198–204.
- [25] Z. Zhou, M. M. R. Siddiquee, N. Tajbakhsh, and J. Liang, "Unet++: Redesigning skip connections to exploit multiscale features in image segmentation," *IEEE Trans. Med. Imaging*, vol. 39, no. 6, pp. 1856–1867, 2019. doi: [10.1109/TMI.2019.2959609](https://doi.org/10.1109/TMI.2019.2959609).
- [26] A. Rakshit, S. Pramanick, A. Bagchi, and S. Bhattacharyya, "Autonomous grasping of 3-D objects by a vision-actuated robot arm using Brain-Computer Interface," *Biomed. Signal Process. Control*, vol. 84, no. 5, pp. 104765, 2023. doi: [10.1016/j.bspc.2023.104765](https://doi.org/10.1016/j.bspc.2023.104765).
- [27] C. Y. Wang, A. Bochkovskiy, and H. Y. M. Liao, "YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recogn. (CVPR)*, Seattle WA, USA, Jun. 18–22, 2023, pp. 7464–7475.
- [28] N. Wojke, A. Bewley, and D. Paulus, "Simple online and realtime tracking with a deep association metric," in *Proc. 2017 IEEE Int. Conf. Image Process. (ICIP)*, Beijing, China, Sep. 17–20, 2017, pp. 3645–3649.
- [29] V. J. Lawhern, A. J. Solon, N. R. Waytowich, S. M. Gordon, C. P. Hung and B. J. Lance, "EEGNet: A compact convolutional neural network for EEG-based brain-computer interfaces," *J. Neural Eng.*, vol. 15, no. 5, pp. 056013, Oct. 2018. doi: [10.1088/1741-2552/aace8c](https://doi.org/10.1088/1741-2552/aace8c).
- [30] K. Xia, L. Deng, W. Duch, and D. Wu, "Privacy-preserving domain adaptation for motor imagery-based brain-computer interfaces," *IEEE Trans. Biomed. Eng.*, vol. 69, no. 11, pp. 3365–3376, Nov. 2022. doi: [10.1109/TBME.2022.3168570](https://doi.org/10.1109/TBME.2022.3168570).
- [31] H. Kwon and S. Lee, "Friend-guard adversarial noise designed for electroencephalogram-based brain-computer interface spellers," *Neurocomputing*, vol. 506, no. 05, pp. 184–195, 2022. doi: [10.1016/j.neucom.2022.06.089](https://doi.org/10.1016/j.neucom.2022.06.089).