JOURNAL OF INFORMATION SCIENCE AND ENGINEERING XX, XXXX-XXXX (2024)
DOI: XX.XXX.XXXX.XX

An Integrated Gaze-Based Framework for Analyzing Offensive and Defensive Awareness in FPS Games

JASON C. HUNG, HAO-SHANG MA+, AND CHENG YU HSUEH
Department of Computer Science and Information Engineering
National Taichung University of Science and Technology
No. 129, Section 3, Sanmin Road, North District, Taichung City, Taiwan
E-mail: {hsma}@mail.example.edu.tw

Offensive and defensive awareness is a critical determinant of player performance in first-person shooter (FPS) games. However, most existing studies rely on self-reported questionnaires or coarse gameplay statistics, which fail to adequately capture the fine-grained visual attention and dynamic decision-making inherent in fast-paced gameplay. To overcome these limitations, this study proposes an integrated framework to quantitatively assess player awareness by combining webcam-based eye tracking, questionnaire-derived awareness profiling, and deep learning-driven region-of-interest (ROI) detection.

To facilitate natural gameplay conditions, we developed a consumer-grade webcam eye-tracking module featuring automated calibration. Compared to traditional click-based calibration, this proposed method significantly reduces gaze estimation errors and improves data stability. Furthermore, a YOLOv7-based object detection model was trained to identify critical in-game ROIs, such as enemies, weapons, maps, and defensive items. By utilizing Optuna for hyperparameter optimization, the detection robustness was further enhanced, particularly for visually ambiguous objects.

Finally, questionnaire-derived awareness scores were integrated with gaze–ROI intersection data for behavioral analysis. Experimental results demonstrate that players with higher awareness scores exhibit more strategically balanced and focused gaze distributions toward critical ROIs, which strongly correlates with superior in-game performance. Ultimately, the proposed framework provides a quantitative, multi-modal, and interpretable approach for evaluating player awareness and analyzing behavior in FPS environments by combining questionnaire-derived awareness profiles with gaze-based ROI measurements.

Keywords: FPS Games, Eye Tracking, Region-of-Interest Detection, Visual Attention, Player Awareness

1. Introduction

First-person shooter (FPS) games have rapidly evolved from entertainment platforms into highly competitive e-sports, attracting global recognition such as their inclusion in the 2023 Asian Games. These games require players to make rapid, high-stakes decisions in real time, where victory often hinges on more than just reflexes. The core components of

+Corresponding author

FPS gameplay can be categorized into three key operational behaviors: operational awareness, tactical movement, and enemy perception. For example, players with strong visual acuity can quickly detect enemy appearances and movements and coordinate with teammates to issue warnings. These players also make split-second decisions about whether to attack or retreat, which often determines the match's outcome.

Training in FPS games typically involves both live practice and retrospective analysis, such as reviewing recorded gameplay to refine strategies and correct previous tactical errors. These practices aim to enhance a player's ability to formulate effective offensive and defensive strategies, minimize errors, and foster team synergy. Despite the centrality of these awareness-driven behaviors, there is a lack of systematic research that quantifies a player's offensive and defensive cognition, leaving competitive players without a structured reference to evaluate or improve these skills.

Eye tracking technology has emerged as a powerful tool for understanding user attention, decision making, and cognitive load in a wide range of domains, including education [1], healthcare [2], and marketing [3]. However, its application in gaming research, particularly in analyzing moment-to-moment tactical decisions in FPS environments is still in its infancy. Traditional eye-tracking devices, such as a chin rest or wearable eye trackers, pose several limitations. Wearable systems, while highly accurate, often require the camera to be positioned within approximately 30 cm of the eye. This configuration can obstruct parts of the face, such as the forehead and eyebrows, thereby limiting the ability to simultaneously analyze facial expressions, which reduces the system's extensibility for multimodal behavioral analysis. Additionally, the high cost and discomfort of such systems hinder their use in large-scale, naturalistic experiments like gaming.

Although eye-tracking has been applied to gaming and visual-attention studies, most existing works mainly analyze general gaze behavior, screen engagement, or performance-related attention patterns. These studies provide useful evidence for understanding where players look, but they often do not explicitly connect gaze behavior with FPS-specific tactical constructs such as offensive and defensive awareness. In addition, many existing approaches rely either on self-reported questionnaires or gaze measurements alone, making it difficult to jointly examine the relationship between players' perceived awareness and their actual object-level visual attention during gameplay. To address this gap, we propose an integrated gaze-based framework for analyzing offensive and defensive awareness in FPS games. The proposed framework lies in the integration of three complementary components: webcam-based eye tracking, questionnaire-derived awareness profiling, and YOLOv7-based ROI detection. By synchronizing gaze coordinates with detected in-game ROIs, the framework enables gaze-ROI intersection analysis, which links players' subjective awareness profiles with their object-level visual attention toward enemies, weapons, maps, scopes, defensive items, and other gameplay elements. This design allows offensive and defensive awareness to be examined not only as self-reported tendencies, but also as observable gaze allocation patterns during FPS gameplay.

In summary, the main contributions of this work are as follows:

- We develop a low-cost webcam-based eye-tracking module for FPS gameplay analysis. The module extracts facial and eye-related features and supports automated calibration to improve gaze stability under natural gameplay conditions.
- We define offensive and defensive awareness as FPS-specific tactical constructs and

use questionnaire-derived awareness profiling to capture players' perceived attention priorities during attack- and defense-related situations.

- We construct a YOLOv7-based ROI detection module to automatically identify semantically meaningful in-game objects and interface elements, including enemies, weapons, maps, scopes, and defensive-related cues.
- We integrate gaze tracking, questionnaire-based awareness profiling, and ROI detection through gaze-ROI intersection analysis. This integration enables the comparison between players' self-reported awareness tendencies and their actual visual attention toward task-relevant in-game objects.
- We provide an in-game validation showing that gaze allocation toward different ROI categories can offer interpretable behavioral evidence for analyzing offensive and defensive awareness in FPS gameplay.

2. Relate Works

This section reviews prior works on FPS player behavior, eye-tracking technologies, and automated ROI detection. These areas form the foundation for our proposed system and highlight key gaps in linking gaze behavior with tactical awareness in the game.

2.1 Player Behavior and Tactical Awareness in FPS Games

First-person shooter (FPS) games require players to integrate rapid decision-making, precise motor coordination, and real-time tactical judgment. The core behaviors involved in such games can be broadly categorized into operational awareness, tactical movement, and enemy perception. These factors collectively shape a player's offensive and defensive capabilities, which are often decisive in competitive settings.

The motivations for playing FPS games include achievement, social interaction, and immersive experience [4]. These motivational factors correlate with distinct play styles and decision-making patterns. For example, experienced players typically exhibit faster reaction times, higher spatial awareness, and more effective use of tools such as aiming reticles and scopes [5, 6]. Additionally, in multiplayer formats such as battle royale games, players rely on communication channels like voice or text chat to share resources and alert teammates about enemy positions, creating a social-cognitive dynamic during gameplay [7].

A key dimension in FPS games is awareness — both offensive and defensive. Offensive awareness involves proactive strategies such as pre-aiming, flanking, and target prediction, while defensive awareness includes evasion, cover usage, and risk assessment. Psychological studies have investigated aggression and its manifestations in gaming, showing that male players tend to score higher in aggression measures, which may influence their offensive behavior [8]. Emotional factors, such as frustration, can trigger displaced aggression, sometimes directed toward unintended targets or vehicles [9, 10]. These patterns underscore the complexity of player behavior and the role of emotional and situational cues in shaping in-game actions.

Conversely, defensive awareness involves protective behavior, such as utilizing armor or making evasive maneuvers to avoid detection or damage. Players may adopt avoidance tactics like crouching or hiding behind obstacles when under threat [11, 12]. Defensive behavior is often initiated unconsciously, especially when players are under high cognitive load or stress. In such cases, habitual responses based on prior gameplay experience become critical [13].

Recent studies have attempted to link these behavioral constructs with cognitive mechanisms like attention and awareness. Conscious processing is often associated with strategic decisions and team coordination, while unconscious processes are responsible for rapid, reflexive behaviors. Understanding the interaction between these systems is essential for modeling in-game cognition and player expertise [14, 15].

Despite extensive research on gaming psychology and behavior, few studies have quantitatively modeled offensive and defensive awareness or explored their connections with perceptual data such as eye movement. This gap highlights the need for integrated approaches that combine behavioral models with objective tracking methods such as eye-tracking, which will be addressed in the following sections.

2.2 Eye-Tracking Technology and Application

Eye-tracking provides real-time insights into users' cognitive and perceptual processes by monitoring their gaze behavior. It enables the measurement of key visual attention indicators such as fixation duration, saccades, and pupil dilation, which are directly associated with cognitive load and decision-making. Foundational research by Rayner established the link between fixations and attentional demand in reading and scene perception tasks [16, 17], while Posner's model emphasized the triad of attention components: orienting, alerting, and executive control [18].

Recent technological developments have led to a diversification of eye-tracking hardware and software. Traditional infrared-based systems remain the gold standard for precision but typically require controlled lab environments and specialized hardware [19, 20]. In contrast, more scalable solutions such as WebGazer utilize standard webcams to detect facial landmarks and estimate gaze vectors [21], making them well-suited for remote, large-scale data collection. Deep learning-based approaches have further advanced the field. CNNs and 3D modeling techniques enhance gaze estimation by addressing variations in lighting, occlusion, and head pose [22, 23, 24]. These models allow for flexible integration with consumer-grade devices, enabling real-time tracking without the constraints of traditional hardware setups.

Wearable devices, including head-mounted and glasses-type trackers, provide greater spatial accuracy and are effective in naturalistic settings such as VR or mobile applications [25, 26, 27]. However, their close-range placement—often within 30 cm of the user's eye—can obstruct peripheral vision and facial cues, potentially affecting user experience and ecological validity [28, 29].

Eye-tracking has found wide application across disciplines including education [30], healthcare [31], e-commerce [32], and affective computing [33, 34]. In multimedia learning, gaze patterns help researchers evaluate attention distribution, identify cognitive overload, and optimize instructional design. In marketing, fixation heatmaps reveal product salience and visual engagement levels. In gaming, eye-tracking supports the analysis of screen engagement, decision-making speed, and player situational awareness. Chen and

Tsai demonstrated that expert FPS players exhibit markedly different gaze behaviors from novices, often showing shorter fixations and more anticipatory saccades during fast-paced gameplay [35]. Furthermore, eye movement metrics have been used to assess emotional responses and strategy execution under time constraints [36, 20].

While promising, several challenges persist in dynamic gaming environments. Pupil occlusion, rapid head movements, and calibration drift can impair data accuracy—especially in fast-paced genres like FPS [26, 27]. In immersive setups, such as virtual reality, mobile trackers embedded in headsets capture more naturalistic behavior [25], though they still suffer from latency and processing overhead.

Despite the growth of eye-tracking applications, relatively few studies have explored the alignment between gaze patterns and high-level FPS constructs such as offensive-defensive transitions or strategic awareness. Most existing literature focuses on basic attention or emotional analysis rather than in-depth, domain-specific cognition. This research aims to bridge that gap by incorporating gaze-based indicators into the analysis of competitive FPS behavior, examining how visual focus contributes to real-time decision-making and team coordination.

3. Methology

This section outlines the proposed system architecture designed to integrate real-time eye-tracking with deep learning models for Region of Interest (ROI) detection in First-Person Shooter (FPS) game scenarios. The framework supports data acquisition, gaze mapping, and automatic labeling of game footage, providing a foundation for understanding strategic awareness and visual attention in high-pressure game environments.

This chapter presents the architecture of the proposed system designed to assess offensive and defensive awareness in FPS games by integrating eye-tracking, survey-based analysis, object detection, and behavioral modeling. The system consists of four key modules: (1) a webcam-based eye-tracking model, (2) YOLOv7-based object detection, and (4) awareness evaluation using integrated gaze and visual stimuli data.

3.1 Webcam-based Eye-tracking Model

This section presents the eye detection and eye-feature extraction stage of the webcam-based eye-tracking module. The goal of this stage is to extract stable eye-related image features from webcam video streams, which are later used for click-based calibration and gaze mapping. All operations described in this section are performed in the camera image space, without involving any screen-space gaze estimation.

3.1.1 Overview of the Eye-Feature Extraction Pipeline

The eye-feature extraction pipeline processes each frame captured by a standard webcam and outputs a set of eye-related geometric features. As shown in Fig. 1, the pipeline consists of face detection, facial landmark localization, eye-region cropping, pupil localization, and eye-geometry feature extraction. The extracted features describe the spatial configuration of the eyes and pupils relative to the camera and face, and are computed on a frame-by-frame basis.

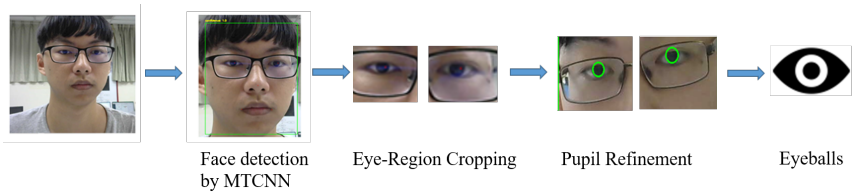


Fig. 1. The Procedure of Eye Ball Capture From Webcam.

This stage focuses exclusively on feature extraction and does not perform gaze-to-screen mapping. The extracted eye features are used as inputs to the calibration procedure described in the next section.

3.1.2 Face Detection and Facial Landmark Localization

Face detection is first applied to each input frame to locate the facial region of the user. A learning-based face detection method is employed to robustly identify the face bounding box under varying illumination conditions and head poses. After the face region is detected, facial landmark localization is performed to extract key reference points on the face.

The detected facial landmarks provide stable geometric references for eye localization. In particular, landmarks around the eye region enable accurate identification of the left and right eyes, even in the presence of minor head movements. As shown in Fig. 1, the detected facial bounding box and landmark positions serve as the basis for subsequent eye-region extraction.

3.1.3 Eye-Region Localization and Cropping

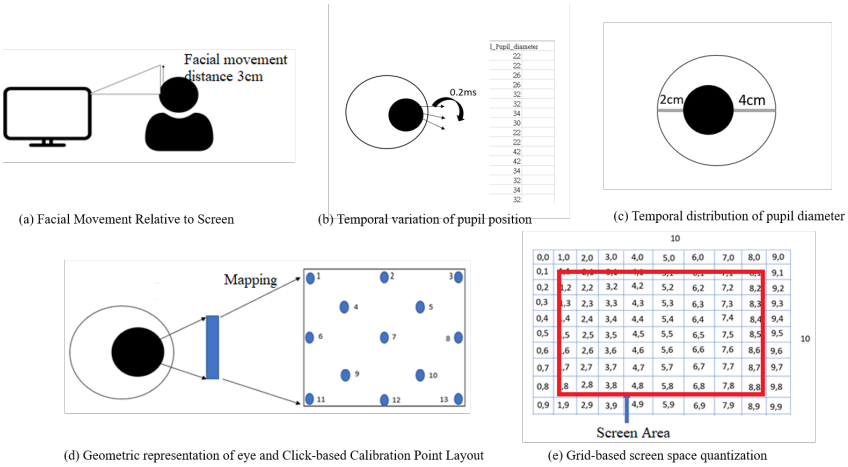
Based on the localized facial landmarks, the regions corresponding to the left and right eyes are identified. Rectangular eye regions are defined using the relative positions of eye-related landmarks to ensure consistent coverage of the eyeball and eyelid areas. Each eye region is then cropped from the facial image and processed independently.

Cropping the eye regions reduces background interference and computational complexity while preserving essential eye information. Separating the left and right eye regions also improves robustness against partial occlusion and asymmetric eye movements. As shown in Fig. 1, the eye-region cropping process extracts compact eye images from the detected face region.

3.1.4 Pupil Localization and Refinement

After eye-region cropping, pupil localization is performed within each eye region. The pupil area is identified based on its geometric and intensity characteristics, which differentiate it from surrounding eye structures. An initial pupil center is estimated and further refined to improve localization accuracy.

The refinement process reduces the influence of noise caused by illumination variations, reflections, and eyelid interference. By constraining the pupil shape to approximate



(a) Facial Movement Relative to Screen (b) Temporal variation of pupil position (c) Temporal distribution of pupil diameter

(d) Geometric representation of eye and Click-based Calibration Point Layout (e) Grid-based screen space quantization

Fig. 2. The Calibration and Gaze Mapping.

circular geometry, the stability of the pupil center estimation is enhanced across consecutive frames. Examples of pupil localization and refinement results are shown in Fig. 1.

3.1.5 Eye-Related Geometric Feature Extraction

Following pupil localization, multiple eye-related geometric features are extracted. These features include the pupil center coordinates, eye-corner positions, pupil or iris size, and the vertical distance between the upper and lower eyelids. All features are computed relative to the eye-region coordinate system and normalized to reduce sensitivity to scale variations.

Eye-corner features provide additional geometric constraints that enhance robustness under head movement. Pupil size and eyelid distance are also recorded to support blink detection and data quality assessment. These features collectively describe the eye configuration in image space.

3.1.6 Output Representation of Eye Features

For each video frame, the extracted eye features are assembled into a feature vector representing the instantaneous eye state. This feature vector includes pupil position, eye-corner geometry, and eye-opening related measurements. All features remain in camera or image coordinates and are not yet mapped to screen space. The eye-feature vectors produced in this section form the fundamental input to the click-based calibration and gaze mapping stage described in Section 3.2. By clearly separating eye-feature extraction from calibration, the proposed system maintains a modular and interpretable design.

3.2 Click-based Calibration and Gaze Mapping

This section presents the calibration and gaze mapping stage of the webcam-based eye-tracking module. Based on the eye-image features extracted in Section 3.1, this stage aims to transform camera-space eye features into screen-space gaze coordinates. The

calibration process compensates for individual differences in eye movement range, facial displacement, and non-linear variations in eye behavior, enabling gaze estimation during gameplay. The proposed gaze mapping procedure uses the eye-related feature vectors generated in the previous stage. Each feature vector contains pupil position, eye-corner geometry, and pupil size measurements. These features remain in camera or image coordinates before calibration. During calibration, the system records eye features while the participant fixates on predefined screen regions. These recorded samples are then used as reference points for estimating the relationship between eye movement and screen position. Since webcam-based eye tracking is sensitive to head movement, illumination changes, blinking, and pupil localization errors, unstable samples are filtered before gaze mapping. Samples with excessive facial displacement, abnormal pupil size variation, missing facial landmarks, or invalid eye-corner geometry are excluded from the calibration process. This filtering step improves the stability of the gaze mapping function and reduces the possibility of mapping unreliable eye samples to screen coordinates. The mapping process is based on the maximum and minimum observable eye coordinates collected during calibration. These boundary coordinates define the valid eye movement range of each participant. For valid samples, the current eye position is proportionally scaled within the calibrated eye movement range and mapped to the corresponding screen-space position. This range-based interpolation strategy normalizes individual differences in eye mobility and provides a consistent basis for estimating gaze coordinates during gameplay. If a recorded sample exceeds the calibrated coordinate range, it is regarded as a non-attentive or unreliable sample and is excluded from subsequent analysis. This calibration design is also used to reduce the influence of click-induced instability. In click-based calibration, users are required to manually click predefined calibration points. However, the clicking action may introduce involuntary saccadic eye movements and transient gaze shifts. In contrast, the automated calibration procedure records eye features continuously during the calibration process without requiring explicit user interaction. By reducing click-induced gaze fluctuations, the automated method produces more stable gaze to screen mappings and improves the reliability of the subsequent gaze-ROI analysis.

3.2.1 Facial Movement Distance Estimation

As shown in Fig. 2(a), facial movement is quantified by estimating the distance between the face and the camera based on geometric principles. Let f_s denote the detected face size in the image, a_s the focal length of the camera, and f the real-world face size. Based on the similarity of triangles and basic optical geometry, the distance between the face and the camera D is computed as:

$$D(f_s, f, a_s) = \frac{a_s \cdot f}{f_s} \tag{1}$$

This formulation allows the system to estimate facial displacement relative to the camera. By monitoring changes in D, excessive facial movement can be detected. When the displacement exceeds a predefined threshold, the corresponding samples are excluded to prevent instability in gaze calibration and to ensure that only reliable eye coordinates are retained.

3.2.2 Pupil Size Collection and Pruning

During eye-tracking, pupil center diameter data are continuously collected. Pupil size may vary depending on the viewed screen region, illumination conditions, and physiological responses. Additionally, blinking can cause abrupt changes in detected pupil size, leading to jump-like artifacts in the collected gaze data.

To address this issue, pupil size variations are monitored and used as a criterion for data filtering. Samples associated with sudden pupil diameter changes—commonly caused by blinks or partial occlusions—are identified and excluded from subsequent analysis. As shown in Fig. 2(b), it illustrates examples of pupil size collection and the detection of unstable samples. This filtering process helps maintain the consistency and reliability of eye-tracking measurements.

3.2.3 Eye-Corner Feature Collection

In addition to pupil-related features, eye-corner distances are collected to verify the correctness of eye orientation and movement direction. Eye images are also recorded to ensure that eye alignment is correctly captured by the eye-tracking method.

Eye-corner features are obtained by measuring the distance between the pupil center and the eye corners. By analyzing the variation of this distance over time, the direction of eye movement can be inferred. If the distance continuously decreases, it indicates that the eyeball is moving toward the camera. This information is used to estimate the maximum range of eye movement, which is essential for accurate gaze mapping.

As shown in Fig. 2(c), eye-corner feature collection provides an additional geometric constraint that improves the reliability of eye position estimation. Only when the eye orientation is verified as valid are the corresponding gaze data considered accurate and included in further processing.

3.2.4 Gaze Mapping Based on Maximum and Minimum Coordinates

Gaze mapping is performed by collecting fixation coordinates at calibration points corresponding to the maximum and minimum observable eye positions. During calibration, gaze coordinates are recorded while the participant fixates on predefined screen locations.

If a recorded fixation point exceeds the established coordinate range, it is classified as a non-attentive sample and excluded from the experimental analysis. For valid samples, gaze coordinates are proportionally scaled according to the maximum eye movement range to normalize individual differences in eye mobility. This range-based scaling ensures the accuracy and reliability of the mapped gaze data and serves as the basis for calibration evaluation. The gaze mapping process is illustrated in Fig. 2(d).

3.2.5 Grid-based Gaze Coordinate Transformation

To enable integration with the YOLOv7 object detection results, continuous gaze coordinates are transformed into grid-based coordinates. YOLOv7 outputs object locations as bounding boxes, while gaze estimation produces point-based screen coordinates. Therefore, both gaze points and ROI bounding boxes are converted into a shared grid-based representation before gaze–ROI intersection analysis. In this study, the screen

is divided into a fixed 10×10 grid. This setting is adopted as a practical balance between spatial resolution and robustness to webcam-based gaze estimation errors. Since consumer-grade webcam eye tracking is more sensitive to calibration error, facial displacement, and pupil localization noise than specialized eye-tracking devices, directly using pixel-level gaze coordinates may lead to unstable gaze–ROI matching. The grid-based representation reduces the influence of small gaze deviations and provides a more stable spatial unit for matching gaze points with detected objects. The 10×10 grid provides sufficient spatial granularity to distinguish major gameplay regions, such as the central combat area, minimap region, health-related interface area, and object regions appearing in the player's field of view. At the same time, it avoids excessive sensitivity to small gaze errors. A finer grid, such as a 20×20 grid, may provide higher spatial resolution, but it may also increase false mismatches when the estimated gaze point falls slightly outside the corresponding ROI due to webcam tracking noise. Conversely, a coarser grid may improve robustness but may lose useful spatial distinctions between nearby gameplay objects. Therefore, the 10×10 grid was selected as a compromise between interpretability, spatial precision, and error tolerance in the current experimental setting. When the gaze is located near the center of the screen, the corresponding grid coordinate is recorded as $(5,5)$. Gaze directed toward other regions is mapped to the corresponding grid cell based on its relative position. Each detected ROI bounding box is also mapped to the set of grid cells it occupies. A gaze-ROI intersection is then determined by checking whether the gaze grid cell overlaps with the grid cells covered by the detected ROI. This conversion enables consistent comparison between gaze information and detected object regions and facilitates quantitative gaze-object association in subsequent analysis.

Fig. 2(e) illustrates the transformation of continuous gaze coordinates into a 10×10 grid representation. This conversion enables consistent data representation between gaze information and detected object regions, facilitating quantitative gaze-object association in subsequent analysis.

3.3 Questionnaire-Based Awareness Profiling

In first-person shooter (FPS) games, offensive and defensive awareness is not only reflected in motor execution and visual attention but also in players' situational judgment regarding *what objects or events should be attended to* under different gameplay conditions. While eye-tracking provides objective measurements of gaze behavior, it does not directly reveal the player's internal awareness structure or strategic intent. Therefore, a questionnaire-based profiling approach is adopted to complement gaze-based analysis by quantifying players' awareness priorities in FPS gameplay.

FPS games present players with dense and rapidly changing visual stimuli, including enemies, teammates, weapons, environmental cover, user interface elements, and dynamic events such as gunfire or explosions. Under such conditions, effective performance depends on the player's ability to selectively allocate attention to task-relevant objects and events while suppressing irrelevant information.

The questionnaire in this study is designed to identify:

- Which in-game objects players consciously consider critical during offensive situations?

- Which environmental cues or events players prioritize when adopting defensive strategies?
- How these priorities vary with gameplay experience and training background?

Rather than measuring personality traits or emotional states, the questionnaire serves as a structured tool to externalize players' internal awareness models, providing a reference framework for interpreting gaze-object interactions derived from eye-tracking data.

3.3.1 Operational Definition of Offensive and Defensive Awareness

Based on the gameplay characteristics of FPS games, awareness is divided into two complementary dimensions:

Offensive awareness refers to the player's tendency to actively seek opportunities to engage enemies, initiate attacks, or gain positional advantage. This includes attention to enemy appearance, enemy movement trajectories, weapon readiness, aiming reticles, and tactical openings.

Defensive awareness refers to the player's tendency to prioritize survival, risk avoidance, and positional safety. This includes attention to cover objects, escape routes, teammate positions, health indicators, ammunition status, and incoming threats.

The questionnaire items are constructed to reflect these two awareness dimensions by asking players to report the relative importance of different in-game objects and events during offensive and defensive decision-making phases.

3.3.2 Questionnaire Design and Content Structure

The questionnaire consists of multiple items grouped into thematic categories corresponding to common FPS gameplay elements, including:

Enemy-related cues: enemy presence, movement direction, distance, and visibility.

Environmental objects: walls, cover structures, terrain features, and chokepoints.

Interface elements: minimap, crosshair, health bar, ammunition indicators, and cooldown timers.

Event-based cues: gunfire sounds, explosion effects, teammate communication, and alert notifications.

Participants rate the importance or frequency of attention allocation to each item using a Likert-scale format. This design enables subjective awareness priorities to be converted into quantitative scores suitable for statistical analysis and comparison across player groups.

3.3.3 Integration with Eye-Tracking and ROI Analysis

The questionnaire-based awareness profiles serve as a semantic bridge between subjective cognition and objective eye-tracking data. By identifying which objects or events players *claim* to attend to during offensive or defensive situations, the questionnaire provides a reference against which actual gaze distributions and ROI overlaps can be evaluated. Specifically, questionnaire responses are incorporated into the analytical framework to inform the selection of target regions of interest (ROIs) used in gaze–object intersection analysis. By identifying which in-game objects and events players report as being critical during offensive and defensive situations, the questionnaire provides a semantic reference

for interpreting gaze allocation patterns. This allows discrepancies between self-reported awareness and actual visual attention behavior to be systematically examined. Through the integration of subjective questionnaire data and objective eye-tracking measurements, the proposed approach supports a more comprehensive evaluation of offensive and defensive awareness in FPS gameplay. Through this integration, the study aims to establish a more comprehensive and interpretable framework for quantifying situational awareness in FPS games.

3.4 YOLOv7-Based ROI Detection in FPS Gameplay

In first-person shooter (FPS) games, visual scenes are highly dynamic and densely populated with task-relevant objects, including enemies, weapons, interface elements, vehicles, and environmental cues. Accurately identifying these objects is necessary for interpreting gaze behavior and evaluating offensive and defensive awareness. Therefore, this study employs a YOLOv7-based object detection module to automatically identify and localize regions of interest (ROIs) in FPS gameplay frames. The ROI categories were selected according to three criteria. First, each ROI must be relevant to offensive or defensive decision-making in FPS gameplay. Objects such as enemies, weapons, and scopes are closely associated with threat identification, target engagement, and attack preparation. In contrast, maps, blood bars, defensive weapons, vehicles, and survival-related interface elements provide information related to risk monitoring, positioning, and defensive decision-making. Second, each ROI must be visually observable in gameplay frames. Since the YOLOv7 model learns from image-based evidence, the selected categories must have identifiable visual appearances that can be annotated using bounding boxes. Objects that cannot be consistently observed in gameplay footage are not suitable for object detection training. Third, the selected ROI categories must be consistent with the questionnaire-based awareness framework. The questionnaire captures the objects and situations that players consider important during offensive and defensive decision-making. Therefore, the ROI categories are used as object-level representations of these awareness-related visual cues. This design ensures that the detected ROIs are not arbitrarily selected, but are connected to the offensive and defensive awareness dimensions examined in this study. Based on these criteria, the ROI categories include blood bars, maps, weapons, enemies, scopes, defensive weapons, death packs, vehicles, explosions, number of survivors, and airdrop boxes. These categories represent observable in-game objects or interface elements that may influence offensive or defensive visual attention. For example, enemies, weapons, and scopes are mainly related to offensive awareness, while maps, blood bars, defensive weapons, and survival-related cues are mainly related to defensive awareness. Some ROI categories may support both dimensions depending on the gameplay context. For instance, the map may support offensive positioning as well as defensive movement planning.

3.4.1 Regions of Interest Detection and Automatic Identification

In order to enable automatic identification of regions of interest (ROIs) in FPS gameplay, a task-specific object detection dataset is constructed based on recorded in-game video streams. The primary goal of dataset construction is to provide frame-level object annotations that can be directly aligned with eye-tracking data for gaze–object intersec-

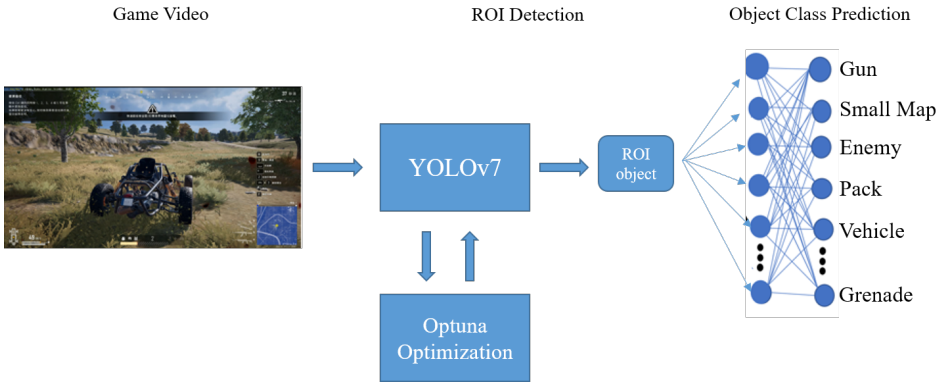


Fig. 3. Overview of the YOLOv7-based ROI detection framework with Optuna-based hyperparameter optimization for FPS gameplay analysis.

tion analysis.

As illustrated in Fig. 3, gameplay videos are first recorded under natural playing conditions to preserve realistic visual complexity, including variations in player viewpoint, map layout, lighting, and combat intensity. The recorded videos are then sampled at fixed temporal intervals to extract representative frames, ensuring that both static objects (e.g., environmental structures) and dynamic objects (e.g., enemies in motion) are sufficiently captured.

Each extracted frame is manually annotated using bounding boxes following the YOLO annotation format. As shown in Fig. 4, bounding boxes are assigned to visually observable in-game objects that may contribute to offensive or defensive awareness. The annotation focuses on object appearance rather than inferred player intent, ensuring that the detection model learns from image-based evidence only.

The annotated dataset is divided into training, validation, and testing subsets to support model training and performance evaluation. This separation ensures that the detection model is evaluated on unseen gameplay scenes, reducing overfitting and improving generalization.

Through this dataset construction and annotation process, YOLOv7 is trained to automatically detect ROIs in FPS gameplay frames. The resulting bounding boxes provide structured spatial information that can be temporally synchronized with gaze coordinates, forming the basis for quantitative awareness analysis in subsequent stages.

To optimize the performance of the YOLOv7-based ROI detection model, Optuna is employed for automated hyperparameter tuning. The optimization process is integrated into the training pipeline, where each trial corresponds to a distinct set of hyperparameter configurations. During optimization, YOLOv7 is trained using parameters sampled by Optuna, and model performance is evaluated on a validation set. The objective function is defined based on detection performance metrics obtained from the validation data, guiding the search toward parameter settings that yield stable and accurate object detection. The hyperparameters optimized include learning rate, batch size, momentum, weight decay, and confidence threshold. These parameters directly influence convergence behavior,



Fig. 4. An ROI example of the scope

detection sensitivity, and robustness under dynamic FPS gameplay conditions.

A tree-structured Parzen estimator (TPE)–based sampling strategy [37] is adopted to efficiently explore the hyperparameter space. Through iterative trials, Optuna identifies the parameter configuration that achieves the best validation performance. The optimized hyperparameter set is then used to train the final YOLOv7 model, which is applied to detect ROIs across the gameplay dataset. This optimized detection output provides reliable bounding box information for subsequent gaze–ROI intersection analysis.

3.5 Gaze-ROI Intersection and Awareness Evaluation

This section describes the technical procedure used to integrate webcam-based gaze tracking data with YOLOv7-based region-of-interest (ROI) detection for evaluating offensive and defensive awareness in FPS gameplay. The proposed method quantitatively determines whether a player's visual attention is directed toward semantically meaningful in-game objects at each frame and aggregates this information to characterize awareness patterns.

3.5.1 Gaze-ROI Alignment and Intersection Computation

Eye-tracking data and ROI detection results are synchronized using timestamps to ensure frame-level correspondence. For each gameplay frame t, the calibrated gaze position is represented as a two-dimensional screen coordinate

$$G_t = (x_t, y_t),$$

while each detected ROI is represented by a bounding box

$$B_t^{(k)} = (x_{min}^{(k)}, y_{min}^{(k)}, x_{max}^{(k)}, y_{max}^{(k)}),$$

where k denotes the ROI index.

To reduce sensitivity to small gaze estimation errors and to unify representations across modules, both gaze coordinates and ROI bounding boxes are mapped to a grid-based coordinate system. The screen is divided into a fixed 10×10 grid. The gaze point

G_t is mapped to a grid cell C_t, and each ROI bounding box is mapped to the set of grid cells it occupies.

A gaze-ROI intersection is defined as

$$I_t^{(k)} = \begin{cases} 1, & \text{if } C_t \cap \mathscr{B}_t^{(k)} \neq \emptyset, \\ 0, & \text{otherwise,} \end{cases}$$

where $\mathscr{B}_t^{(k)}$ denotes the grid cells covered by the k-th ROI. This binary indicator determines whether the player is visually attending to a specific ROI at frame t.

3.5.2 Attention Aggregation and Awareness Quantification

Over a gameplay segment consisting of T frames, gaze-ROI intersections are accumulated to quantify attention allocation. For each ROI class c, the total attention score is computed as

$$A_c = \sum_{t=1}^{T} \sum_{k \in c} I_t^{(k)},$$

where $k \in c$ indicates ROIs belonging to class c.

To account for differences in gameplay duration, attention scores are normalized by the total number of valid gaze samples, yielding a relative attention proportion for each ROI class. These normalized values represent how frequently a player visually attends to different categories of in-game objects. Based on questionnaire-based profiling, ROI classes are grouped into offensive-related and defensive-related categories. Offensive awareness is characterized by higher attention proportions toward enemy-related and attack-relevant ROIs, while defensive awareness is characterized by increased attention toward cover objects, escape routes, and safety-related interface elements.

3.5.3 Optuna-Based Hyperparameter Optimization

To improve the robustness of the YOLOv7-based ROI detection model, Optuna was used for automated hyperparameter tuning. The optimized parameters include learning rate, batch size, momentum, weight decay, and confidence threshold. These parameters directly influence convergence behavior, detection sensitivity, and classification stability under dynamic FPS gameplay conditions. The optimization process evaluates different hyperparameter configurations based on validation performance. After each trial, the configuration with better validation results is used to guide the search toward more suitable parameter settings. The final optimized model is then applied to detect ROIs across the gameplay dataset. This optimized detection output provides more reliable bounding box information for subsequent gaze-ROI intersection analysis.

3.5.4 Integration with Questionnaire-Based Profiling

Questionnaire responses provide a subjective reference for determining which in-game objects and events players consider important during offensive and defensive decision-making. These responses are used to define the mapping between ROI classes and awareness dimensions. By comparing normalized gaze-based attention scores with

Table 1. Overview of participants' FPS gaming background based on questionnaire responses ($N = 425$).

Category	Level	Number of Participants
FPS experience	None	34
	$<$ 1 year	131
	1–3 years	109
	3–5 years	55
	$>$ 5 years	96
Daily playtime	None	34
	10–30 minutes	89
	30 minutes–1 hour	98
	1–3 hours	147
	$>$ 3 hours	58

questionnaire-derived awareness priorities, the proposed framework evaluates the consistency between subjective awareness and observed visual behavior. This integration enables a multi-modal assessment of situational awareness that combines objective gaze measurements with self-reported strategic intent.

4. Questionnaire-Based Awareness Analysis

4.1 Participants of the Questionnaire Survey

The questionnaire survey was conducted to collect subjective perceptions of offensive and defensive awareness in FPS games. A total of $N = 425$ valid responses were obtained from participants recruited through university populations and online gaming communities. Respondents covered a wide range of FPS gaming backgrounds in terms of experience level and daily playtime, ensuring diversity in gameplay familiarity and decision-making styles.

The questionnaire dataset serves two purposes in this study. First, it provides a quantitative characterization of players' perceived offensive and defensive awareness. Second, it supports the identification of in-game objects and events that players consider important during offensive and defensive situations, which are later used to define awareness-relevant regions of interest (ROIs) for experimental validation.

4.2 Quantification of Offensive and Defensive Awareness

To quantify players' offensive and defensive awareness in FPS gameplay, a structured questionnaire was designed to capture self-reported attentional focus, decision priorities, and situational judgment. Rather than measuring general gaming motivation or personality traits, the questionnaire focuses on awareness-related behaviors that are directly connected to gameplay situations. The questionnaire was designed using a Likert-scale format. Participants rated the perceived strength or importance of each awareness-related behavior during FPS gameplay. The numerical responses were then used to rep resent players' subjective awareness tendencies. These questionnaire-derived scores provide a semantic reference for interpreting gaze–ROI interactions obtained from eye-tracking and

object detection. The questionnaire contains two major dimensions: offensive awareness and defensive awareness. Offensive awareness refers to the player's tendency to actively identify attack opportunities, monitor enemy-related cues, prepare for engagement, and make attack-oriented decisions. Defensive awareness refers to the player's tendency to monitor survival-related information, avoid risk, use defensive resources, maintain positional safety, and respond to incoming threats. The offensive-awareness items focus on attack-related behaviors, such as actively searching for enemies, observing the map before engagement, using scopes or weapons for attack, responding to enemy damage, attacking from vehicles or buildings, and evaluating attack behavior under different equipment conditions. The defensive-awareness items focus on survival-related behaviors, such as monitoring the map and blood bar, avoiding unnecessary exposure, responding to enemy attacks, using cover or buildings, avoiding vehicle-related threats, and searching for defensive items. To improve transparency, the complete questionnaire items used in this study are provided in Appendix A. The appendix allows readers to examine the actual items used to construct the offensive- and defensive-awareness dimensions. This also helps clarify how the questionnaire content is connected to the ROI categories used in the gaze-based analysis.

4.2.1 Transformation of Questionnaire Responses

The questionnaire responses were transformed into quantitative awareness indicators according to the semantic meaning of each item. Items designed to represent stronger offensive behavior were grouped as high-offensive items, whereas items representing weaker or less active attack behavior were grouped as low-offensive items. Similarly, defensive-awareness items were grouped according to whether they represented stronger or weaker defensive tendencies. For each participant group, the consistency between the intended awareness level of each item and the participant's perceived rating was calculated. Responses that matched the intended awareness direction were treated as positive contributions, while responses that did not match the intended direction were treated as negative contributions. This consistency-based scoring design allows the questionnaire results to reflect whether players can correctly distinguish high- and low-awareness gameplay situations. The questionnaire-derived awareness profiles are then compared with gaze-ROI intersection results. For example, players with stronger offensive awareness are expected to show greater attention toward attack-related ROIs, such as enemies, weapons, or scopes. Players with stronger defensive awareness are expected to show greater attention toward defensive or survival-related ROIs, such as maps, blood bars, defensive weapons, and other safety-related cues. Through this comparison, the proposed framework examines whether self-reported awareness tendencies are reflected in actual visual attention behavior during gameplay.

4.2.2 Offensive Awareness Scoring

The offensive awareness score is calculated based on the consistency between the intended awareness level of questionnaire items and the participants' perceived awareness. Items designed to reflect high offensive awareness and low offensive awareness are evaluated separately. Let $ATTACK\alpha_{HIGH}$ denote the number of high-offensive items perceived as high, and let $ATTACK\alpha_{LOW}$ denote the number of low-offensive items perceived

as low. These two terms represent consistent responses and are treated as positive contributions. Conversely, $ATTACK\beta_{HIGH}$ denotes high-offensive items perceived as low, and $ATTACK\beta_{LOW}$ denotes low-offensive items perceived as high. These two terms represent inconsistent responses and are treated as negative contributions. The overall offensive awareness score is computed as:

$$\begin{aligned} OffensiveScore = {} & ATTACK\alpha_{HIGH} + ATTACK\alpha_{LOW} \\ & - ATTACK\beta_{HIGH} - ATTACK\beta_{LOW}. \end{aligned} \tag{2}$$

A positive score indicates stronger consistency in offensive awareness, whereas a lower or negative score suggests weaker or inconsistent offensive judgment.

4.2.3 Defensive Awareness Scoring

Defensive awareness is quantified using the same consistency-based scoring strategy. Items designed to reflect high defensive awareness and low defensive awareness are evaluated separately. Let $DEFENSE\alpha_{HIGH}$ denote the number of high-defensive items perceived as high, and let $DEFENSE\alpha_{LOW}$ denote the number of low-defensive items perceived as low. These two terms represent consistent responses and are treated as positive contributions. Conversely, $DEFENSE\beta_{HIGH}$ denotes high-defensive items perceived as low, and $DEFENSE\beta_{LOW}$ denotes low-defensive items perceived as high. These two terms represent inconsistent responses and are treated as negative contributions. The overall defensive awareness score is computed as:

$$\begin{aligned} DefensiveScore = {} & DEFENSE\alpha_{HIGH} + DEFENSE\alpha_{LOW} \\ & - DEFENSE\beta_{HIGH} - DEFENSE\beta_{LOW}. \end{aligned} \tag{3}$$

A positive score indicates stronger consistency in defensive awareness, whereas a lower or negative score suggests weaker or inconsistent defensive judgment.

4.3 Questionnaire Analysis Results

Table 2. Integrated Analysis of Offensive Awareness across Player Experience Levels

Experience Level	High-Offensive		Low-Offensive		Score
	H→H	H→L	L→H	L→L	
Low Experience	5	3	4	1	−1
Moderate Experience	0	5	0	4	−1
High Experience	5	2	3	2	2

Table 2 summarizes the recognition distribution and aggregated offensive awareness scores across different player experience levels.

For high-offensive awareness items, players with high experience show a balanced pattern, with five correctly recognized items (High→High) and two incorrectly recognized items (High→Low). This indicates that experienced players generally possess stable offensive awareness, despite occasional misjudgments.

In contrast, moderately experienced players fail to correctly recognize any high-offensive items, exhibiting the highest number of incorrect recognitions. This suggests

that offensive awareness in this group is unstable and prone to misinterpretation of engagement opportunities.

Low-experience players demonstrate a mixed pattern, showing both correct and incorrect recognitions for high-offensive items. This implies that novice players may rely on intuitive judgment rather than consistent offensive strategies.

For low-offensive awareness items, low-experience players show more incorrect recognitions (Low→High) than correct ones (Low→Low), indicating difficulty in identifying situations where restraint is more appropriate than engagement. Moderately experienced players again exhibit the weakest performance, with all low-offensive items being incorrectly recognized. Players with high experience perform comparatively better, demonstrating a more balanced recognition pattern.

The aggregated offensive awareness scores further consolidate these observations. Only the high-experience group achieves a positive total score, whereas both low- and moderate-experience groups obtain negative scores. These results suggest that offensive awareness does not increase linearly with gameplay exposure. Instead, a transitional phase appears among moderately experienced players, where partial familiarity with game mechanics may lead to inconsistent offensive judgment.

5. In-game Experimental Validation

The in-game experimental validation aims to examine whether subjective awareness profiles derived from the questionnaire are reflected in players' actual visual attention behavior during FPS gameplay. Specifically, the experiment tests the hypothesis that players with higher self-reported offensive or defensive awareness exhibit corresponding differences in gaze allocation toward awareness-relevant objects.

To test this hypothesis, webcam-based eye tracking and object-level ROI detection are integrated to capture where and on what objects players allocate their visual attention under dynamic gameplay conditions. The validation focuses on the consistency and discrepancy between subjective awareness scores and objective gaze-based indicators.

Table 3. Comparison of Eye-Tracking Calibration Methods

Calibration Method	Click-based	Automated
RMSE	221.66	51.62

5.1 Evaluation of Eye-Tracking Calibration Methods

To evaluate the effectiveness of the proposed webcam-based eye-tracking model, an experimental comparison was conducted between two calibration strategies: click-based calibration and automated calibration. The evaluation focuses on how different calibration mechanisms influence gaze stability and estimation accuracy.

In click-based calibration, users are required to manually click on predefined calibration targets. However, the clicking action itself introduces involuntary saccadic eye movements and transient shifts in gaze position. These abrupt movements result in large coordinate fluctuations during data recording, which significantly degrade calibration accuracy. As a consequence, the root-mean-square error (RMSE) obtained from click-based

calibration reaches 221.66, indicating substantial gaze estimation deviation. In contrast, the automated calibration approach continuously records eye features throughout the calibration process without requiring explicit user interaction. This design enables smoother gaze trajectories and preserves the temporal continuity of eye movements. By eliminating click-induced saccades, the automated method produces more stable gaze–screen mappings. The comparison results, summarized in Table 3, confirm that automated calibration provides superior performance for webcam-based eye tracking. As a result, the RMSE is significantly reduced to 51.62, demonstrating a marked improvement in calibration accuracy.

The improvement in calibration accuracy directly affects the reliability of subsequent gaze-ROI intersection analysis. Lower RMSE values indicate reduced spatial deviation between estimated gaze points and actual screen locations, which is critical when mapping gaze data onto object-level ROIs detected in gameplay frames. With high calibration error, gaze points are more likely to fall outside the correct ROI boundaries, leading to misclassification of attentional focus. By contrast, the automated calibration approach provides more precise and stable gaze coordinates, enabling accurate determination of whether gaze samples intersect with specific in-game objects. This improvement ensures that gaze-based awareness indicators derived from ROI dwell time and scanpath analysis more faithfully reflect players' true attentional allocation during gameplay.

5.2 ROI Detection and Classification Accuracy Evaluation

To evaluate the effectiveness of the proposed ROI detection module, object detection experiments were conducted on gameplay footage collected from PUBG. All game frames were manually annotated using the CVAT open-source image annotation tool, where bounding boxes were labeled for multiple in-game objects, including blood bars, minimaps, weapons, enemies, scopes, defensive weapons, death packs, vehicles, explosions, number of survivors, and airdrop boxes.

The YOLOv7 model was adopted as the core detection framework and trained using GPU acceleration, which significantly improved training efficiency compared to CPU-based training. Model performance was evaluated using precision, recall, mAP@0.5, and mAP@0.5:0.95, providing a comprehensive assessment of both classification and localization accuracy.

Evaluation Metrics and Confusion Matrix. Detection performance was analyzed based on confusion matrix statistics, where true positives (TP), false positives (FP), false negatives (FN), and true negatives (TN) were defined according to prediction and ground truth correspondence. Precision and recall were computed as $\text{Precision} = \text{TP}/(\text{TP}+\text{FP})$ and $\text{Recall} = \text{TP}/(\text{TP}+\text{FN})$, while mAP@0.5 and mAP@0.5:0.95 measured average precision under different intersection-over-union (IoU) thresholds.

Baseline YOLOv7 Detection Performance. Table 4 summarizes the ROI detection performance of the proposed YOLOv7-based model under different optimization stages. The baseline model demonstrates strong overall detection capability, achieving high precision and mAP values across most in-game object categories. However, category-level analysis reveals that detection performance is not uniform across all ROIs. The confusion

Table 4. YOLOv7 ROI detection performance under different optimization stages

Model Stage	Precision	Recall	mAP@0.5	mAP@0.5:0.95	Defensive Weapon Acc.
Baseline YOLOv7	0.8927	0.7837	0.8294	0.6411	0.39
Optuna (1st cycle)	0.8914	0.7856	0.8340	0.6334	0.51
Optuna (Full)	0.8668	0.8032	0.8495	0.6406	0.60

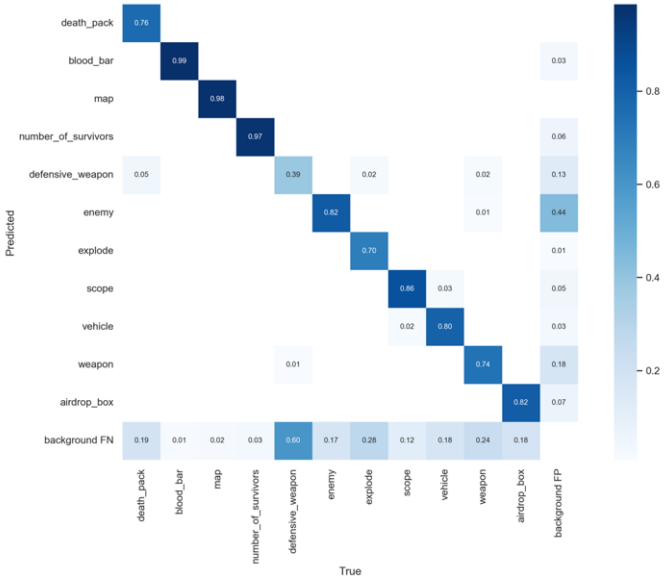


Fig. 5. Confusion matrix of the baseline YOLOv7 model for FPS in-game ROI detection.

matrix of the baseline YOLOv7 model, shown in Fig. 5, indicates that most object categories are recognized with high accuracy. Several ROIs, such as blood bars, minimaps, enemies, scopes, vehicles, and airdrop boxes, exhibit strong diagonal responses, reflecting stable and consistent classification. These objects either possess distinctive visual features or appear frequently in fixed screen locations, making them easier to identify. However, the defensive weapon category exhibits substantially lower accuracy, with a correct classification rate of approximately 0.39, reflecting higher visual ambiguity and inter-class similarity. This confusion can be attributed to visual similarity between defensive and offensive weapons, small ROI size, and rapid appearance within dynamic gameplay scenes. Such misclassifications highlight a key limitation of the baseline model and motivate further optimization.

Fully Optimized YOLOv7 Detection Performance. After applying full Optuna-based hyperparameter optimization, a clear improvement in category-level robustness is observed. The confusion matrix of the optimized model, presented in Fig. 6 shows that the diagonal response for the defensive weapon category increases to approximately 0.60. At the same time, misclassifications between defensive weapons and other visually simi-

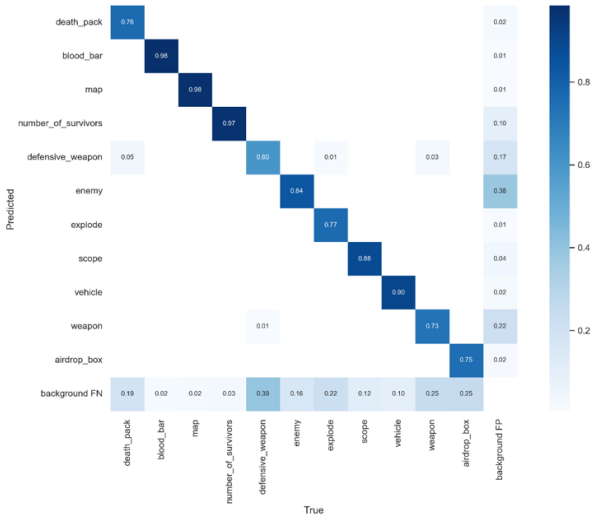


Fig. 6. Confusion matrix of the YOLOv7 model after full Optuna-based hyperparameter optimization.

lar categories are significantly reduced.

Importantly, the detection performance of other ROIs remains stable after optimization, with most categories maintaining accuracies above 0.73. This indicates that hyperparameter optimization selectively enhances performance for visually ambiguous ROIs without degrading detection accuracy for well-defined object categories.

Overall, the confusion matrix comparison demonstrates that Optuna-based optimization effectively improves category-level robustness, particularly for defensive weapons. This improvement is critical for subsequent gaze-ROI intersection analysis, as reduced false negatives and misclassifications lead to more reliable estimation of players' visual attention and defensive awareness during gameplay.

The improvement in the defensive weapon category further illustrates the importance of hyperparameter optimization. Defensive weapons are more difficult to detect than several other ROI categories because they may be visually similar to offensive weapons, appear in relatively small regions, and occur only briefly during fast gameplay. In the baseline model, the defensive weapon category shows a lower correct classification rate, indicating that this category is more sensitive to visual ambiguity and inter-class similarity. Among the optimized parameters, the confidence threshold and learning rate are especially influential for this category. The confidence threshold affects the trade-off between false positives and false negatives. If the threshold is too high, small or ambiguous defensive weapons may be missed; if it is too low, the model may produce more incorrect detections. The learning rate affects the stability of model convergence and may influence how well the model separates visually similar categories. Momentum and weight decay further support training stability and reduce unstable classification results. After Optuna-based optimization, the defensive weapon category shows improved category-level robustness, while the performance of other ROI categories remains stable. This im-

Table 5. Gaze distribution across ROIs for six players in a single FPS match.

ROI Category	P161	P040	P011	P411	P953	P001
Death Pack	0	0	0	0	0	8
Blood Bar	4	16	41	45	15	19
Map	29	57	28	59	97	103
Number of Survivors	0	0	0	0	0	0
Defensive Weapon	0	5	0	0	0	0
Enemy	49	213	57	144	36	91
Explosion	0	0	2	0	2	0
Scope	6	0	23	0	83	8
Vehicle	0	0	1	1	0	75
Weapon	0	2	0	0	0	1
Airdrop Box	0	0	0	0	0	0
Final Rank	5	1	3	6	2	4
Survival Time (mm:ss)	04:00	11:38	08:27	02:57	09:33	06:29

provement is important for the gaze–ROI intersection analysis because inaccurate detection of defensive weapons may lead to incorrect estimation of defensive visual attention. Therefore, the optimized YOLOv7 model provides a more reliable basis for evaluating defensive awareness during gameplay.

5.3 Awareness Evaluation via gaze-ROI Integration

To further validate the relationship between gaze behavior and offensive–defensive awareness, this study analyzes a representative match as a case study. Six players from a single FPS match were selected to illustrate how gaze distribution across ROIs corresponds to in-game performance and awareness characteristics.

5.3.1 Gaze Distribution across ROIs among Players

Table 5 summarizes the gaze fixation counts of six players across different ROI categories during a single match, together with their final ranking and survival time. The fixation count represents the cumulative number of gaze samples falling within each detected ROI. The results reveal distinct gaze strategies among players with different match outcomes. The first-ranked player (P040) exhibits exceptionally high fixation counts on enemy ROIs, indicating strong situational awareness and rapid threat detection, while maintaining moderate attention to the map and scope. This suggests a balanced offensive strategy supported by spatial awareness.

The second-ranked player (P953) allocates substantial attention to the map and scope, reflecting a navigation- and positioning-oriented playstyle. Both offensive and defensive ROIs are consistently monitored, indicating deliberate tactical planning. Players ranked third (P011) and fifth (P161) show relatively low fixation counts across most ROIs, with no category exceeding 50 fixations. This dispersed and limited attention suggests weaker focus on critical in-game objects, which may contribute to suboptimal per-

formance. The fourth-ranked player (P001) demonstrates diversified gaze distribution across multiple ROIs without a dominant focus. Such evenly spread attention may indicate indecision or reactive gameplay behavior. The sixth-ranked player (P411) shows high fixation on enemy ROIs, but relatively lower attention to supportive information such as the map and blood bar. This imbalance suggests aggressive engagement without sufficient defensive monitoring, which may explain the shorter survival time.

Overall, this case study demonstrates that gaze-ROI patterns provide interpretable signals of offensive and defensive awareness. Players with higher rankings tend to exhibit either focused attention on enemies supported by situational awareness, or balanced monitoring of tactical ROIs such as maps and scopes. In contrast, players with lower rankings often display fragmented or insufficient gaze allocation toward critical objects.

These findings support the validity of integrating gaze-ROI intersection analysis with awareness scoring mechanisms, offering interpretable behavioral evidence for quantifying offensive and defensive awareness in FPS gameplay. Rather than serving as a purely objective evaluation, the proposed analysis combines questionnaire-derived awareness profiles with gaze-based behavioral measurements to provide a more comprehensive interpretation of player awareness.

5.4 Discussion

The experimental results demonstrate the feasibility of integrating webcam-based eye tracking, questionnaire-based awareness profiling, and YOLOv7-based ROI detection for analyzing offensive and defensive awareness in FPS gameplay. The proposed framework connects players' self-reported awareness priorities with their gaze allocation toward object-level ROIs. In this way, the framework provides a quantitative and interpretable approach for examining how players visually attend to attack-related and defense-related objects during gameplay. The calibration results show that stable gaze estimation is essential for gaze-ROI analysis. Since FPS gameplay involves rapid visual shifts and continuous changes in screen content, inaccurate calibration may cause gaze points to deviate from the correct ROI regions. The automated calibration method reduces gaze estimation error and provides more stable gaze coordinates, which improves the reliability of determining whether a gaze point intersects with a detected ROI. The ROI detection results also show that object-level detection can provide meaningful semantic information for interpreting gaze behavior. However, not all ROI categories are equally easy to detect. Defensive weapons are more difficult to recognize because they may be visually similar to other weapon-related objects, appear in smaller regions, and be visible only for short periods during gameplay. The Optuna-based optimization improves the robustness of this category, indicating that hyperparameter tuning is useful for visually ambiguous ROIs. Although the proposed framework provides useful evidence for gaze-based awareness analysis, several limitations should be acknowledged. First, gameplay performance is influenced by factors beyond visual attention, including player experience, reaction speed, familiarity with the map, gameplay strategy, device settings, team communication, and short-term combat conditions. Therefore, gaze-ROI behavior should not be interpreted as the only determinant of gameplay performance. Instead, it should be considered one behavioral indicator that complements questionnaire responses and gameplay outcomes. Second, questionnaire responses provide subjective information about players' perceived awareness. Although these responses are useful

for understanding players' internal awareness priorities, they may be influenced by self-report bias. For this reason, the questionnaire results are not treated as absolute ground truth. Instead, they are used as subjective awareness profiles and are compared with gaze-based measurements. Third, the current in-game validation focuses on selected gameplay data and representative player cases. While the results show meaningful relationships between gaze distribution and player outcomes, larger-scale experiments across more players, matches, maps, and game modes are needed to further examine the generalizability of the proposed framework. Finally, this study adopts a fixed 10×10 grid to reduce the effect of webcam-based gaze estimation errors. Although this setting improves robustness and interpretability, different grid resolutions may influence gaze–ROI matching results. Future work may compare fixed grids with finer grids or adaptive grid designs to further improve gaze-ROI association under different experimental conditions. Overall, the proposed framework should be understood as a multi-modal and quantitative evaluation approach rather than a purely objective assessment system. By combining questionnaire-based awareness profiling, webcam-based gaze tracking, and object-level ROI detection, the framework provides a scalable method for analyzing offensive and defensive visual attention in FPS games.

6. Conclusion

This study presents an integrated framework for quantifying offensive and defensive awareness in first-person shooter (FPS) games by combining webcam-based eye tracking, questionnaire-based awareness profiling, and deep learning-based ROI detection. Unlike prior work that relies solely on self-reported behavior or coarse gameplay statistics, the proposed approach bridges subjective cognition and objective visual attention through gaze-ROI interaction analysis.

First, a webcam-based eye-tracking module was developed to support natural gameplay conditions. By incorporating automated calibration and eye-feature stability assessment, the proposed method significantly reduces gaze estimation error compared to click-based calibration, thereby improving the reliability of gaze data used in downstream analysis. The reduction in calibration error directly enhances the accuracy of gaze-ROI intersection, which is critical for object-level attention modeling in dynamic game environments.

Second, a questionnaire-based awareness analysis was conducted to quantify offensive and defensive awareness from a cognitive perspective. The proposed scoring formulation captures discrepancies between perceived awareness and actual gameplay understanding, enabling awareness levels to be represented as numerical indices. These questionnaire-derived awareness scores serve as a reference for validating gaze-based observations.

Third, a YOLOv7-based ROI detection module was employed to automatically identify key in-game objects, including enemies, weapons, maps, and defensive items. Experimental results demonstrate that hyperparameter optimization using Optuna substantially improves detection robustness, particularly for visually ambiguous object categories. The final optimized model provides stable and accurate ROI localization, forming a reliable foundation for gaze-based attention analysis.

Finally, an in-game experimental validation was conducted by integrating gaze trajectories, ROI detection results, and awareness scoring. A case study involving six players illustrates how different gaze allocation strategies correspond to varying levels of offensive and defensive awareness and match outcomes. Players with higher rankings tend to exhibit focused or balanced attention toward critical ROIs, whereas lower-ranked players show fragmented or insufficient gaze distribution. These findings indicate that gaze–ROI interaction patterns can provide interpretable behavioral indicators of player awareness during gameplay. However, these indicators should be interpreted together with questionnaire-derived awareness profiles, because the proposed framework combines subjective self-reported awareness with gaze-based behavioral measurements.

Overall, this work demonstrates the feasibility of using consumer-grade eye tracking combined with deep learning-based object detection to quantitatively assess offensive and defensive awareness in FPS games The proposed framework provides a scalable, multimodal, and interpretable approach for player behavior analysis, training assessment, and performance evaluation. Since the framework incorporates both questionnaire-derived awareness profiles and gaze-based measurements, it should be understood as a combined subjective–behavioral analysis method rather than a purely objective assessment system.

Acknowledgement

This study was supported by a grant from the National Science and Technology Council, R.O.C (No. 113-2221-E-025-006-). We thank the National Science and Technology Council for funding this study.

REFERENCES

1. M. Schindler, A. Shvarts, and A. J. Lilienthal, "Introduction to eye tracking in mathematics education: interpretation, potential, and challenges," *Educational Studies in Mathematics*, Vol. 118, 2025, pp. 309–321.
2. K. Harezlak and P. Kasprowski, "Application of eye tracking in medicine: A survey, research issues and challenges," *Computerized Medical Imaging and Graphics*, Vol. 65, 2018, pp. 176–190, advances in Biomedical Image Processing. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0895611117300435
3. M. Wedel and R. Pieters, *A Review of Eye-Tracking Research in Marketing*. Emerald Group Publishing Limited, 2008, Vol. 4.
4. N. Yee, "Motivations for play in online games," *CyberPsychology and behavior*, Vol. 9, no. 6, 2006, pp. 772–775.
5. C. Gutwin, R. Vicencio-Moreira, and R. L. Mandryk, "Does helping hurt? aiming assistance and skill development in a first-person shooter game," in *Proceedings of the 2016 Annual Symposium on Computer-Human Interaction in Play*, October 2016, pp. 338–349.
6. R. P. Heitz, "The speed-accuracy tradeoff: history, physiology, methodology, and behavior," *Frontiers in neuroscience*, Vol. 8, 2014, p. 150.

7. J. A. de Mesquita Neto and K. Becker, "Relating conversational topics and toxic behavior effects in a moba game," *Entertainment computing*, Vol. 26, 2018, pp. 10–29.
8. A. H. Buss and M. Perry, "The aggression questionnaire," *Journal of personality and social psychology*, Vol. 63, no. 3, 1992, p. 452.
9. N. Miller, W. C. Pedersen, M. Earleywine, and V. E. Pollock, "A theoretical model of triggered displaced aggression," *Personality and Social Psychology Review*, Vol. 7, no. 1, 2003, pp. 75–97.
10. T. Lajunen, D. Parker, and S. G. Stradling, "Dimensions of driver anger, aggressive and highway code violations and their mediation by safety orientation in uk drivers," *Transportation Research Part F: Traffic Psychology and Behaviour*, Vol. 1, no. 2, 1998, pp. 107–121.
11. C. A. Hendrie, S. M. Weiss, and D. Eilam, "Behavioural response of wild rodents to the calls of an owl: a comparative study," *Journal of Zoology*, Vol. 245, no. 4, 1998, pp. 439–446.
12. K. Peleg, A. Rivkind, L. Aharonson-Daniel, and I. T. Group, "Does body armor protect from firearm injuries?" *Journal of the American College of Surgeons*, Vol. 202, no. 4, 2006, pp. 643–648.
13. D. Mobbs, C. C. Hagan, T. Dalgleish, B. Silston, and C. Prévost, "The ecology of human fear: survival optimization and the nervous system," *Frontiers in Neuroscience*, Vol. 9, 2015, p. 55.
14. D. J. Chalmers, "Facing up to the problem of consciousness," *Journal of consciousness studies*, Vol. 2, no. 3, 1995, pp. 200–219.
15. D. Sattin, F. G. Magnani, L. Bartesaghi, M. Caputo, A. V. Fittipaldo, M. Cacciatore, and M. Leonardi, "Theoretical models of consciousness: a scoping review," *Brain Sciences*, Vol. 11, no. 5, 2021, p. 535.
16. K. Rayner, "Eye movements and attention in reading, scene perception, and visual search," *The Quarterly Journal of Experimental Psychology*, Vol. 62, no. 8, 2009, pp. 1457–1506.
17. K. Rayner and E. M. Reingold, "Evidence for direct cognitive control of fixation durations during reading," *Current Opinion in Behavioral Sciences*, Vol. 1, 2015, pp. 107–112, cognitive control. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S235215461400028X
18. M. I. Posner and S. J. Boies, "Components of attention," *Psychological Review*, Vol. 78, no. 5, 1971, p. 391.
19. R. Andersson, M. Nyström, and K. Holmqvist, "Sampling frequency and eye-tracking measures: how speed affects durations, latencies, and more," *Journal of Eye Movement Research*, Vol. 3, no. 3, 2010.
20. P. Blignaut and D. Wium, "Eye-tracking data quality as affected by ethnicity and experimental design," *Behavior research methods*, Vol. 46, 2014, pp. 67–80.
21. A. Papoutsaki, P. Sangkloy, J. Laskey, N. Daskalova, J. Huang, and J. Hays, "Webgazer: Scalable webcam eye tracking using user interactions," in *Proceedings of the Twenty-Fifth International Joint Conference on Artificial Intelligence*, 2016, p. 3839–3845.

22. X. Zhang, Y. Sugano, M. Fritz, and A. Bulling, "It's written all over your face: Full-face appearance-based gaze estimation," in *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*, 2017, pp. 51–60.
23. P. Pathirana, S. Senarath, D. Meedeniya, and S. Jayarathna, "Eye gaze estimation: A survey on deep learning-based approaches," *Expert Systems with Applications*, Vol. 199, 2022, p. 116894.
24. S. Park, A. Spurr, and O. Hilliges, "Deep pictorial gaze estimation," in *Proceedings of the European conference on computer vision (ECCV)*, 2018, pp. 721–738.
25. M. Martin Meißner, J. Pfeiffer, T. Pfeiffer, and H. Oppewal, "Combining virtual reality and mobile eye tracking to provide a naturalistic experimental environment for shopper research," *Journal of Business Research*, Vol. 100, 2019, pp. 445–458.
26. Z. H. Wan, C. H. Xiong, W. B. Chen, and H. Y. Zhang, "Robust and accurate pupil detection for head-mounted eye tracking," *Computers and Electrical Engineering*, Vol. 93, 2021, p. 107193.
27. H. Wang, J. Pi, T. Qin, S. Shen, and B. E. Shi, "Slam-based localization of 3d gaze using a mobile eye tracker," in *Proceedings of the 2018 ACM symposium on eye tracking research and applications*, June 2018, pp. 1–5.
28. Z. Al-Kassim and Q. A. Memon, "Designing a low-cost eyeball tracking keyboard for paralyzed people," *Computers and Electrical Engineering*, Vol. 58, 2017, pp. 20–29.
29. I. Schmitz and W. Einhauser, "Gaze estimation in videoconferencing settings," *Computers in Human Behavior*, Vol. 139, 2023, p. 107517.
30. B. Park, L. Knörzer, J. L. Plass, and R. Brünken, "Emotional design and positive emotions in multimedia learning: An eyetracking study on the use of anthropomorphisms," *Computers and Education*, Vol. 86, 2015, pp. 30–42.
31. M. Meo, J. A. Del Punta, I. Sánchez, R. de Luis García, G. Gasaneo, and R. Martin, "A dynamical method to objectively assess infantile nystagmus based on eye tracking. a pilot study," *Journal of Optometry*, Vol. 16, no. 3, 2023, pp. 221–228.
32. P. Mikalef, K. Sharma, S. Chatterjee, R. Chaudhuri, V. Parida, and S. Gupta, "All eyes on me: Predicting consumer intentions on social commerce platforms using eye-tracking data and ensemble learning," *Decision Support Systems*, 2023, p. 114039.
33. J. Przybyło, E. Kańtoch, and P. Augustyniak, "Eyetracking-based assessment of affect-related decay of human performance in visual tasks," *Future Generation Computer Systems*, Vol. 92, 2019, pp. 504–515.
34. G. G. Scott, Z. Pinkosova, E. Jardine, and C. J. Hand, ""thinstagram": Image content and observer body satisfaction influence the when and where of eye movements during instagram image viewing," *Computers in Human Behavior*, Vol. 138, 2023, p. 107464.
35. Y. Chen and M. J. Tsai, "Eye-hand coordination strategies during active video game playing: An eye-tracking study," *Computers in Human Behavior*, Vol. 51, 2015, pp. 8–14.
36. B. García-Carrión, S. D. Barrio-García, F. Muñoz-Leiva, and L. Porcu, "Effect of social-media message congruence and generational cohort on visual attention and information-processing in culinary tourism: An eye-tracking study," *Journal of Hospitality and Tourism Management*, Vol. 55, 2023, pp. 78–90.
37. S. Watanabe, "Tree-structured Parzen estimator: Understanding its algorithm components and their roles for better empirical performance," in *arXiv:2304.11127*, 2023.

A. Questionnaire Items for Offensive and Defensive Awareness

This appendix provides the complete questionnaire items used in this study. The questionnaire was designed to collect players' perceived offensive and defensive awareness in FPS gameplay. The original questionnaire was written in Chinese and is translated into English here for manuscript presentation.

All offensive-awareness and defensive-awareness items were rated using a five-point Likert scale:

1	2	3	4	5
Weak	Slightly weak	Moderate	Slightly strong	Strong

Participants were instructed to answer the questions based on their gameplay experience, observation, peration behavior, or intuitive judgment if they had no direct FPS gameplay experience.

A.1 Background Information

The first part of the questionnaire collected participants' basic background and FPS gameplay experience.

1. Gender:
   - Male
   - Female
2. Education level:
   - Junior high school
   - Senior high school
   - Five-year junior college
   - University
   - Graduate school
   - Doctoral degree
   - Other
3. Do you agree that your questionnaire responses may be used for this research?
   - Yes
   - No
4. Have you ever played shooting games? The question is not limited to PUBG.
   - Yes
   - No
5. FPS gameplay experience:

- Less than 1 year
- 1–3 years
- 3–5 years
- More than 5 years

6. Daily or per-session gameplay time:
   - 10–30 minutes
   - 30 minutes–1 hour
   - 1–3 hours
   - More than 3 hours

A.2 Offensive-Awareness Items

The following items measure perceived offensive awareness in FPS gameplay. Participants rated each item from 1 (weak) to 5 (strong).

1. In combat, how strong is the offensive awareness of jumping while shooting?
2. How strong is the offensive awareness of observing the enemy's position on the minimap and then moving toward that position?
3. How strong is the offensive awareness of chasing toward the enemy's direction and shooting?
4. How strong is the offensive awareness of shooting with a scope or aiming sight?
5. When being damaged by enemy gunfire, how strong is the offensive awareness of immediately counterattacking?
6. How strong is the offensive awareness of throwing a grenade or explosive toward the enemy's direction?
7. When being damaged by an enemy grenade or explosive, how strong is the offensive awareness of immediately counterattacking?
8. How strong is the offensive awareness of using a car or vehicle, such as a motorcycle or boat, to hit an enemy?
9. When riding a vehicle, such as a motorcycle or boat, how strong is the offensive awareness of immediately attacking after seeing an enemy?
10. When wearing level-three armor and a level-three helmet, how strong is the offensive awareness of attacking after seeing an enemy?
11. How strong is the offensive awareness of shooting at the enemy while remaining in the original position?
12. How strong is the offensive awareness of hiding inside a building and shooting after seeing an enemy?

13. How strong is the offensive awareness of crouching or lying prone and shooting at the enemy?
14. After being attacked by an enemy, how strong is the offensive awareness of crouching or lying prone and shooting at the enemy?
15. When being damaged by enemy gunfire, how strong is the offensive awareness of observing the situation first and then counterattacking?
16. When being damaged by an enemy grenade or explosive, how strong is the offensive awareness of observing the situation first and then counterattacking?
17. When riding a vehicle, such as a motorcycle or boat, how strong is the offensive awareness of deciding to attack based on the number of enemies observed?
18. When wearing level-two armor and a level-two helmet, how strong is the offensive awareness of attacking after seeing an enemy?
19. How strong is the offensive awareness of shooting at an enemy?
20. How strong is the offensive awareness of not shooting unless the enemy is directly in front of the player?
21. When being attacked by an enemy, how strong is the offensive awareness of standing and searching for the enemy to attack?
22. When being damaged by an enemy grenade or explosive, how strong is the offensive awareness of observing the situation and not counterattacking?
23. How strong is the offensive awareness of riding a vehicle, such as a motorcycle or boat, without attacking?
24. When wearing level-one armor and a level-one helmet, how strong is the offensive awareness of attacking after seeing an enemy?
25. When wearing no armor and no helmet, how strong is the offensive awareness of attacking after seeing an enemy?

A.3 Weapon Preference Items Related to Offensive Awareness

The questionnaire also asked participants to select the weapon types they associated with different levels of offensive awareness. These items were multiple-choice questions.

26. Which weapon types do you think players with strong offensive awareness tend to prefer? Multiple selections were allowed.
    - Sniper rifle
    - Rifle
    - Shotgun
    - Submachine gun

- Other

27. Which weapon types do you think players with moderate offensive awareness tend to prefer? Multiple selections were allowed.
    - Sniper rifle
    - Rifle
    - Shotgun
    - Submachine gun
    - Other
28. Which weapon types do you think players with weak offensive awareness tend to prefer? Multiple selections were allowed.
    - Sniper rifle
    - Rifle
    - Shotgun
    - Submachine gun
    - Other

A.4 Open-Ended Offensive-Awareness Item

Participants were also invited to add other offensive-awareness situations not covered by the listed items. The questionnaire provided the following example: riding a vehicle, such as a motorcycle or boat, without attacking may be rated as offensive-awareness strength 1.

> Please add any other offensive-awareness situation in shooting games that was not covered above, and rate its offensive-awareness strength from 1 to 5.

A.5 Defensive-Awareness Items

The following items measure perceived defensive awareness in FPS gameplay. Participants rated each item from 1 (weak) to 5 (strong).

1. How strong is the defensive awareness of jumping while walking?
2. How strong is the defensive awareness of occasionally checking the minimap for gunfire indicators?
3. How strong is the defensive awareness of not shooting randomly to attract enemies when enemies have not detected the player?
4. How strong is the defensive awareness of lying prone to recover health?
5. When being shot, how strong is the defensive awareness of immediately lying prone?

6. When being attacked by a grenade or explosive, how strong is the defensive awareness of immediately lying prone?
7. How strong is the defensive awareness of hiding inside a building or in grass for a long time?
8. How strong is the defensive awareness of staying in or hiding around a vehicle, such as a motorcycle or boat, for a long time?
9. How strong is the defensive awareness of frequently collecting medical kits or healing supplies?
10. How strong is the defensive awareness of spending a long time looking for defensive equipment, such as helmets and armor?
11. How strong is the defensive awareness of occasionally jumping while walking?
12. How strong is the defensive awareness of lying prone after seeing an enemy?
13. When being attacked by an enemy, how strong is the defensive awareness of observing for a while before lying prone?
14. How strong is the defensive awareness of occasionally hiding behind trees or inside buildings?
15. How strong is the defensive awareness of occasionally staying in or hiding around a vehicle, such as a motorcycle or boat?
16. How strong is the defensive awareness of collecting a medical kit or healing supply only after seeing it?
17. How strong is the defensive awareness of occasionally looking for defensive equipment, such as helmets and armor?
18. How strong is the defensive awareness of walking normally without intentionally jumping when being attacked?
19. How strong is the defensive awareness of not observing the minimap?
20. How strong is the defensive awareness of standing while recovering health?
21. When being attacked by an enemy, how strong is the defensive awareness of rushing toward the enemy?
22. How strong is the defensive awareness of wandering for a long time outside buildings or vehicles, such as motorcycles or boats?
23. How strong is the defensive awareness of rarely collecting medical kits or healing supplies?
24. How strong is the defensive awareness of collecting defensive equipment, such as helmets and armor, only after seeing it?

A.6 Open-Ended Defensive-Awareness Item

Participants were also invited to add other defensive-awareness situations not covered by the listed items. The questionnaire provided the following example: collecting defensive equipment everywhere may be rated as defensive-awareness strength 5.

> Please add any other defensive-awareness situation in shooting games that was not covered above, and rate its defensive-awareness strength from 1 to 5.