An Innovative Experimental Method for Evaluating the Effectiveness of Visual Design Based on Eye-Tracking Technology

Xinzhen Zhuo, Zhixiong Hu, Yang Lin
School of Design, Fujian University of Technology, Fuzhou 350001, Fujian, China
Corresponding author: Xinzhen Zhuo (e-mail: fanshaing@163.com)

ABSTRACT To overcome the limitations of conventional design assessment in capturing instant cognition and visual dynamics, this paper proposes a multimodal evaluation method for visual design effectiveness based on eye-tracking technology. By integrating dynamic region-of-interest sequence analysis with retrospective cognitive attribution, the method captures both the spatial flow and temporal dynamics of visual attention. A contextual task is designed to simulate realistic viewing behavior while maintaining experimental control, guiding participants through goal-oriented browsing under standardized presentation conditions. Dynamic region-of-interest sequence analysis is then used to measure gaze retracing rate and evaluate the clarity of information hierarchy. Retrospective interviews and visual entropy analysis are combined to link objective eye-movement trajectories with subjective semantic interpretation. Experimental results show that the method effectively differentiates design effectiveness under different contrast ratios. The high-contrast group has the shortest initial fixation time (180 $\pm$ 38 ms) and the lowest gaze retracing rate (26.1 $\pm$ 5.9%). Gaze retracing rate is negatively correlated with semantic comprehension accuracy ($r = -0.81$), supporting continuous evaluation from perception to cognition.

INDEX TERMS visual communication design, eye tracking, design evaluation, first fixation time, gaze retracing rate

I. INTRODUCTION

VISUAL design effect evaluation serves as a critical link connecting design intent and user perception. The accuracy of this evaluation directly determines the scientific validity of design optimization directions. By quantifying users' visual attention patterns and cognitive responses, researchers can ensure that the final design aligns with intended communicative goals and functional requirements [1,2]. However, current mainstream evaluation methods, such as subjective questionnaires and task performance indicators, can only capture the explicit feedback after user operation, and are difficult to touch the instantaneous cognitive state and subconscious aesthetic preferences of users when browsing the design. Especially when facing complex interface layouts, the dynamic relationship between "first impression" and "deep information processing" is often separated by traditional methods, resulting in a deviation that is difficult to quantify between design intent and actual user visual behavior. Therefore, introducing technical means that can track user visual trajectory in real time has important theoretical value and practical significance for deconstructing the visual cognitive process and quantifying the efficiency of design guidance.

In recent years, scholars at home and abroad have carried out a lot of research on visual design effect evaluation. For example, Hu et al. used eye-tracking technology to compare the differences in attentional bias of college students when freely browsing cartoon-style posters and traditional graphic posters through in-subject experiments [3]. In response to the problems of urban smart advertising in terms of design and visual effect, Guo proposed an image aesthetic quantitative evaluation method based on multi-scale feature extraction network to improve the visual aesthetic effect of advertising [4]. Huang, through analysis of relevant case literature, explored the role of crowdsourced visual perception data and its analysis methods in promoting the integration of public perspective into urban design decisions in data collection, analysis, and scheme generation [5]. Zhu and Kong, focusing on environmental visual information analysis and subjective perception measurement, systematically reviewed the visual perception mechanism in motion processes and its application in people-oriented urban public space design [6]. In the process of developing visualization and visual analysis tools for epidemiologists and modeling scientists, Khan et al. proposed a technical solution to efficiently apply multiple visual designs to different datasets with limited development resources [7]. However, most of the above studies remain at the level of comparing single indicators, lacking in-depth analysis of the coupling relationship between the "spatial flow" and "time series" of visual trajectories, and even less involving cognitive attribution between gaze behavior and design semantic understanding, resulting in the evaluation results failing to fully reflect the continuous psychological process of users from perception to cognition.

In response to the shortcomings of existing methods, literature research shows that combining eye tracking with cognitive psychology methods can more comprehensively reveal the cognitive mechanism when users interact with design works. For example, Novák et al. explored the application of eye tracking in usability and user experience evaluation from a holistic perspective, focusing on automated data processing and pointing out the research direction of combining eye tracking with machine learning to achieve problem pattern recognition [8]. Liu et al., through a systematic review of previous eye tracking studies on facial expression recognition, pointed out that facial morphology, cognitive factors, and emotion recognition barriers are the three main factors affecting the eye movement patterns of expression recognition [9]. Worthy et al., through experiments with various psychological scales and behavioral tasks, found that eye tracking is not prone to producing the Hawthorne effect in most psychological experiments, but may have an impact in risk decision-making tasks involving known probabilities [10]. However, there is currently a lack of mature paradigms that systematically integrate these technologies and apply them to the evaluation of visual design effects. Based on the above ideas, this paper proposes a multimodal evaluation experimental method that integrates spatiotemporal sequence analysis and cognitive attribution. It quantifies visual trajectories through dynamic region of interest sequence technology and combines retrospective interviews to achieve the superposition analysis of subjective semantic understanding and objective eye movement data, so as to solve the problem that traditional methods are difficult to capture the dynamic process of visual cognition.

This study extends eye-tracking technology into a multimodal evaluation method based on spatiotemporal sequence analysis and cognitive attribution. Compared with traditional approaches that rely on single indicators such as gaze duration and pupil diameter, this method incorporates the coupled analysis of the "spatial flow" and "temporal sequence" of visual trajectories. By integrating dynamic region-of-interest sequence analysis with retrospective interviews, a continuous chain of evidence from perception to cognition is constructed. This approach enables quantitative diagnosis of initial capture efficiency and information hierarchy clarity, providing a novel perspective for objective design evaluation.

II. CONTEXTUALIZED STIMULUS MATERIALS AND TASK CONSTRUCTION

To explore the relationship between contrast and design effectiveness in real application scenarios, this study developed experimental stimulus materials through designed scenarios. Firstly, we selected three typical carriers of visual communication in actual application scenarios, namely, e-commerce detail pages, app launch pages and posters, as design prototype. Then, we developed three groups of stimulus materials based on the design prototype by Figma. The contrast manipulation was realized by controlling the brightness difference and color saturation of text and background: the high-contrast group used black background and white text, and the main image with high saturation; the medium-contrast group used dark gray background and light gray text, and the main image with medium saturation; the low-contrast group used similar brightness and color level with the low-saturation image. Each group contained 5 images, totaling 15 stimulus materials. All images were uniformly sized at 1200×800 pixels, with a neutral gray background (RGB 128,128,128) to minimize extraneous visual information. To ensure comparability across stimulus types, all materials were designed with a similar number of visual elements (4–6 key elements per design) and a consistent hierarchical structure (title, main image, button, logo), thereby controlling for visual complexity and making the 8 s presentation duration sufficient for processing each design.

To ensure experimental control while maintaining ecological validity, a behavioral guidance statement was presented before each stimulus. These statements were designed to simulate the intentional nature of real-world viewing, where users typically browse with a purpose (e.g., understanding product information) rather than engaging in purely passive aesthetic evaluation. For instance, for e-commerce product detail pages, the statement read: "Please browse this product page; you will later be asked about your understanding of the product's core selling points." These instructions encouraged participants to engage in goal-oriented browsing rather than passive aesthetic judgment, thus approximating the intentional nature of real-world viewing behavior.

A single-factor within-subjects design was employed, with all stimuli presented in random order. Based on pilot testing with five independent participants (not included in the final sample), which indicated that 8 s was sufficient for participants to complete a typical viewing cycle—including initial exploration and core information extraction—across all three types of stimulus materials (e-commerce detail pages, app launch pages, and posters), each stimulus was presented for a fixed duration of 8 s, followed by a 2 s gray screen as an inter-stimulus buffer to minimize visual carryover effects. This controlled duration ensures comparability across participants and conditions while still capturing the key phases of naturalistic visual processing. The presentation of stimulus was synchronized with the eye-tracking recording, and thus the timestamp of data was aligned with the stimulus presentation. After presenting all stimulus materials, we randomly selected part of materials for retrospective interview. The method of retrospective interview was similar to that in the eye-tracking experiment. We asked participants to watch their own eye-tracking video, and then explain why they were fixated at that position, which provided subject data for the later cognitive attribution. Table 1 shows the main design characteristics of three kinds of stimulus materials developed in this study. The range of contrast manipulation, color saturation, quantity, and the division of areas of interest of stimulus materials are shown.

TABLE 1. Stimulus Material Design Attributes

Contrast Level	Text-Background Luminance Difference	Color Saturation	Quantity	Design Prototype Types	AOI Definition Basis
High contrast	$> 80\%$	High saturation	5	E-commerce detail page/App launch page/Poster	Title, Main image, Button, Logo
Medium contrast	50%–80%	Medium saturation	5	E-commerce detail page/App launch page/Poster	Title, Main image, Button, Logo
Low contrast	$< 50\%$	Low saturation	5	E-commerce detail page/App launch page/Poster	Title, Main image, Button, Logo

III. EYE TRACKING DATA ACQUISITION AND PREPROCESSING

Eye tracking data were recorded with a Tobii Pro Spectrum 120Hz eye tracker. The eye tracker is capable of head movement compensation and its sampling accuracy is sufficiently high for computation of the eye movements needed for the visual trajectory analysis. The stimuli were presented on a 24” monitor with a resolution of 1920×1080 pixels and a refresh rate of 60Hz. The eyes of the participant were situated approximately 65cm from the centre of the screen. In the experimental environment, a constant artificial light source was used to prevent variations in natural light from influencing the pupil data. This study was approved by the Ethics Committee of Fujian University of Technology, and was performed in accordance with the principles of the Declaration of Helsinki. All eligible participants signed an informed consent form.

Before the formal experiment, each participant underwent a 9-point calibration. A binocular tracking error of less than 0.5° was required to proceed to the formal experiment. After successful calibration, participants read the instructions and completed a practice run to familiarize themselves with the experimental procedure. During the formal experiment, the eye tracker simultaneously recorded raw data such as binocular fixation coordinates, fixation start and end times, pupil diameter, and saccade amplitude at a sampling rate of 120Hz. Data preprocessing was performed using Tobii Pro Lab software. First, an I-VT filter (threshold 30°/s) was used to classify the raw eye-tracking data into fixations and saccades, removing blinks and segments with a loss rate higher than 20%. Second, coordinate smoothing was performed on the fixation point data to remove abnormal drift caused by head micro-movements. Then, based on a pre-defined region of interest (AOI) segmentation scheme, the first fixation time (TFF), total fixation duration (TFD), and fixation transition matrix between AOIs were extracted for each stimulus material. Finally, the preprocessed data was exported in CSV format for subsequent statistical analysis. All participants' data were manually reviewed to ensure sampling quality met analytical requirements. Of the 30 initially recruited participants, two were excluded due to excessive data loss (>20%) or calibration failure, yielding a final valid sample of 28 participants.

IV. DYNAMIC ZONE OF INTEREST (ROI) TEMPORAL ANALYSIS

To quantify the visual scanning paths and temporal information, this study segmented multiple ROIs based on the design characteristics of stimulus materials. Each stimulus material was segmented into title area, main image area, button area, and brand logo area based on their contents. The boundary segmentation was finished by two researchers separately and the disagreement was discussed. After ROI segmentation, each stimulus was coordinate-transformed to participants' fixation points data by using AOI tool of Tobii Pro Lab, and then the fixation start time, fixation duration and fixation sequence in every ROI were extracted. Based on above results, this study further defined fixation transition matrix to analyze the visual jump situations between different ROIs. For every participant's fixation sequence on every stimulus, the number of times that eye gaze transferred from ROI A to ROI B was counted and then the transition frequency matrix was constructed (3×3 or 4×4 depends on ROI number). The diagonal elements of transition matrix reflected the number of fixations in same ROI and the other elements reflected the eye gaze transfer between different design elements. Finally, the transition probability was calculated and used to find out the high-frequency transition path to discover the guidance relationship between different design elements.

The main analysis indicator in this section is gaze retracing rate, which reflected the proportion of times that eye gaze transferred from core information area (title area or button area) to secondary area (background area or brand logo area) among all transfer times. The higher the retracing rate, the more times that eye gaze transferred between core information area and secondary area repeatedly. It might reflect the cognitive confusion caused by unclear hierarchy of information. In addition, this study further extracted the temporal order of first arrival at every area of interest to help to determine whether the visual priority met the design expectation. All temporal indicators were summarized by participants and contrast condition later to prepare for the basis of cognitive load assessment.

V. VISUAL ENTROPY CALCULATION AND RETRACING COGNITIVE ATTRIBUTION

Based on the gaze shift matrix analysis, this study further introduces visual entropy to quantify the degree of disorder in the distribution of gaze points. Visual entropy, borrowing the concept of Shannon entropy from information theory, is used to measure the dispersion of the distribution of participants' gaze points among regions of interest. Specifically, the proportion of each participant's gaze points within each region of interest on each stimulus material is first calculated to obtain the probability distribution p_i, where i is the region of interest number. The formula for calculating the visual entropy H is

$$H = -\sum_{i} (p_i \times \log_2 p_i) \tag{1}$$

A higher entropy value means that fixations were scattered more evenly among ROIs and the visual search was more dispersed; a lower entropy value means that the fixation was concentrated in a few ROIs and the visual search was more focused (had clear guidance). This index can be used as an indirect index of cognitive load; usually, there is no clear guidance in visual search or information hierarchy is very chaotic.

Retrospective cognitive attribution was conducted after the eye-tracking experiment. After all participants finished their stimulus viewing, researchers randomly chose 5 stimulus materials and played back the eye-tracking video of each participant on the screen (including fixation point location, fixation duration, and scanning path). During playback, researchers prompted participants to explain their own fixation behavior, such as: "Why did you look at this area repeatedly?" "What made you look at this location suddenly?" "What do you think is the main idea this image wants to tell you?" Participants' fixation behavior explanations were recorded and later transcribed into text. Overlaying objective eye-path with participants' explanations shows the cognitive motivation behind their fixation behavior: be attracted by visually salient things or confused when perceiving semantics.

Visual entropy values were ultimately combined with retrospective interview data: high entropy areas accompanied by participants' descriptions such as "I don't know where to look" or "There's too much information" confirmed the design guidance failure; low entropy areas accompanied by participants' descriptions such as "I saw the title at a glance" or "The button is very prominent" verified the effectiveness of the design guidance. This subjective and objective integrated analysis strategy provides a complete chain of evidence from visual behavior to cognitive attribution for design effectiveness evaluation.

To ensure objectivity and reproducibility, semantic comprehension accuracy was quantified through a structured coding scheme. For each of the five randomly selected stimuli per participant, the retrospective interview transcript was reviewed against the designer's intended message (e.g., for an e-commerce poster, the intended core message was "50% off promotion"). Two independent coders, blind to the experimental conditions, rated each participant's verbal response as either correct (1) or incorrect (0) based on whether the participant accurately identified the intended core message. Inter-rater reliability was assessed using Cohen's κ, yielding a value of 0.87, indicating excellent agreement. Discrepancies were resolved through discussion. For each participant, semantic comprehension accuracy was then calculated as the percentage of correct responses across the five selected stimuli, resulting in a continuous score ranging from 0% to 100% for each contrast condition.

VI. QUANTITATIVE VALIDATION OF VISUAL GUIDANCE EFFECTIVENESS

A. EXPERIMENTAL SETUP

Based on the experimental method described in Sections 2–5, this study systematically validated the effectiveness of the eye-tracking-based evaluation approach through two quantitative experiments. Both experiments employed a single-factor within-subjects design, with visual contrast (high, medium, low) as the independent variable. The dependent variables were time to first fixation (TFF) in Experiment 1, and gaze retracing rate with semantic comprehension accuracy in Experiment 2. A total of 30 participants were initially recruited; after data quality screening, 28 valid datasets were retained for analysis. The experimental protocols and results are detailed below.

B. VISUAL CAPTURE EFFICIENCY ASSESSMENT EXPERIMENT

This experiment aimed to investigate the effect of visual contrast on the efficiency of capturing design elements under goal-oriented viewing conditions. Thirty participants with normal or corrected-to-normal vision were recruited, and after data quality screening, 28 valid datasets were retained for analysis. A single-factor within-subjects design was employed. The independent variable eye movement was recorded when the subjects recognized the design works using the Tobii Pro Spectrum eye tracker. TFFs of each region of interest were extracted by Tobii Pro Lab software. The data were filtered and denoised before being used for the subsequent statistical analysis.

Figures 1 (a-b) show the data distribution characteristics and central tendency of the mean and standard deviation of time to first fixation (TFF) under different visual contrast conditions. As contrast decreased, the

TFF for the key information area increased: the average TFF was 180 ms (SD=38 ms) in the high-contrast group, increased to 242 ms (SD=41 ms) in the medium-contrast group, and further prolonged to 318 ms (SD=43 ms) in the low-contrast group. One-way ANOVA showed significant differences between groups ($F(2,87)=86.44$, $p<0.001$). This

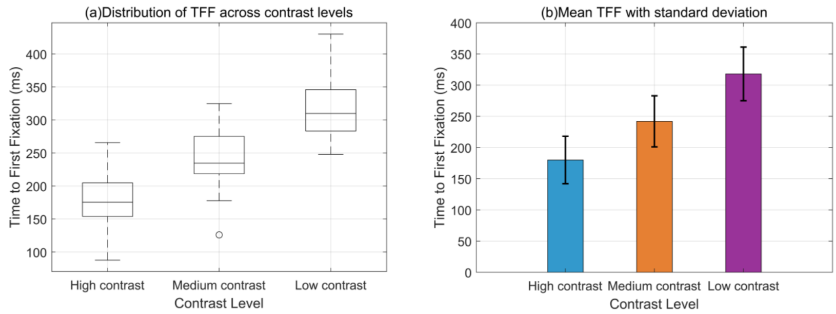

FIGURE 1. Visual Capture Efficiency Assessment(N=28 valid participants)

result validates the effectiveness of first fixation time as a sensitive indicator of visual guidance efficiency, indicating that high-contrast designs can more quickly direct user attention to core information, providing a quantitative basis for design optimization.

C. EXPERIMENT ON COGNITIVE LOAD AND SEMANTIC MATCHING ACHIEVEMENT ASSESSMENT

This experiment aimed to examine the relationship between gaze retracing rate and semantic comprehension accuracy. Thirty participants with normal or corrected-to-normal vision were recruited; following data quality screening, 28 valid datasets were included in the analysis. A single-factor within-subjects design was adopted. Contrast (high, medium, low) was the independent variable. Gaze recall rate and semantic comprehension accuracy were dependent variables. We recorded participants' eye movement information with a Tobii Pro Spectrum eye tracker. We got the gaze recall rate of each region of interest through Tobii Pro Lab. Finally, we administered a questionnaire to evaluate the accuracy of subjects' understanding of design intent. Semantic comprehension accuracy was quantified from the retrospective interviews using a standardized coding rubric (see Section 5 for detailed procedure). This approach yielded a continuous accuracy score per participant per contrast condition, enabling the subsequent correlation analysis with gaze retracing rate.

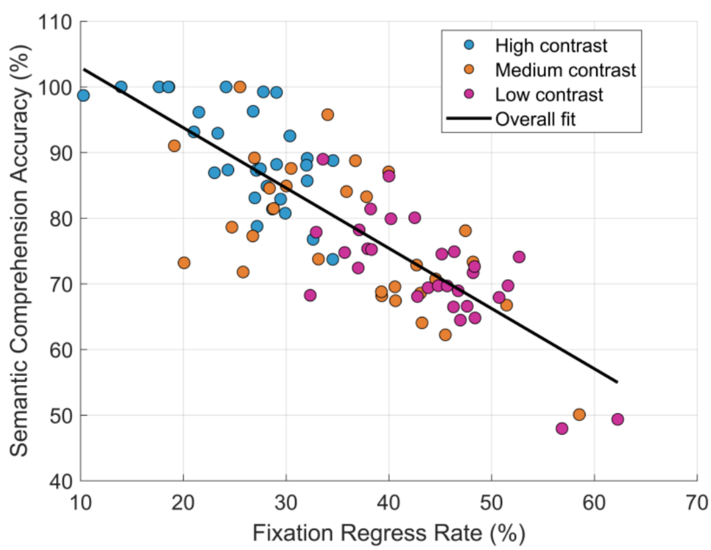

FIGURE 2. Comparison of Cognitive Load and Semantic Matching Assessment

As shown in Figure 2, a significant negative correlation was observed between gaze retracing rate and semantic comprehension accuracy ($r = -0.81$, $p < 0.001$, based on 28 valid participants across three contrast conditions, yielding 84 paired observations), indicating that higher cognitive load was associated with lower accuracy in understanding design intent. Group means revealed that the high-contrast group had the lowest retracing rate (26.1% $\pm$ 5.9%) and the highest accuracy (90.0% $\pm$ 7.8%); the medium-contrast group exhibited an increased retracing rate (36.3% $\pm$ 9.6%) and decreased accuracy (77.1% $\pm$ 11.0%); the low-contrast group showed the highest retracing rate (44.0% $\pm$ 7.1%) and the lowest accuracy (71.7% $\pm$ 8.6%). These findings indicate that visual contrast influences cognitive load, thereby modulating users' semantic alignment with the design, and provide a quantitative basis for evaluating the information communication efficiency of visual designs.

VII. CONCLUSION

This paper presents a multimodal approach to assessment by integrating eye tracking and temporal ocular analysis with fiber-optic tactile feedback and cognitive attribution. This framework correlates visual attention with the sensory interaction of high-performance surfaces, allowing for a comprehensive evaluation of user engagement and material perception. It approaches quantitative diagnosis of visual design by initial capture up to deep processing through dynamic region of interest (ROI) sequence technology and retrospective interview. It has been found through experimentation that this approach can successfully discriminate variations in first fixation time and fixation retrospection rate across designs of varying contrast levels and indicates that there is a strong negative relationship between retrospection rate and semantic understanding accuracy, and confirms its usefulness in dissecting the dynamic process of visual cognition. This research has some limitations. First, the sample was predominantly composed of university students; future work should include more diverse user populations. Second, the stimulus materials were limited to static graphic designs; future studies could extend to dynamic interactive interfaces. Third, while a fixed 8s presentation duration was used to ensure experimental control and comparability, this uniform timing may not fully reflect the natural variation in viewing time across different design complexities (e.g., users may naturally spend longer on e-commerce detail pages than on simple app launch pages). Future studies could adopt self-paced viewing paradigms or implement duration adjustments based on design type to enhance ecological validity. Future research will focus on integrating eye-tracking and EEG data with fiber-integrated physiological sensors and machine learning algorithms to automate the identification of visual and tactile interaction tracks. This multimodal approach ensures that the conductive properties and mechanical deformation of smart textiles are precisely correlated with cognitive states, providing a robust framework for real-time performance evaluation.

ETHICS APPROVAL AND CONSENT TO PARTICIPATE

This study was approved by the Ethics Committee of Fujian University of Technology, and was performed in accordance with the principles of the Declaration of Helsinki. All eligible participants signed an informed consent form.

AUTHOR CONTRIBUTIONS

Conceptualization –Xinzhen Zhuo, Zhixiong Hu and Yang Lin; methodology – Xinzhen Zhuo, Zhixiong Hu and Yang Lin; investigation – Xinzhen Zhuo, Zhixiong Hu and Yang Lin; writing-original draft preparation – Xinzhen Zhuo, Zhixiong Hu and Yang Lin. All authors have read and agreed to the published version of the manuscript.

CONFLICTS OF INTEREST

The authors declare no conflict of interest.

FUNDING

This research received no external funding.

ACKNOWLEDGEMENTS

Not applicable.

REFERENCES

[1] Y. Xu, "Research on visual evaluation of waterfront greenway based on semantic analysis method - taking Xuanwu Lake Park in Nanjing as an example," Design, vol. 09, no. 06, pp. 219-230, 2024, doi: 10.12677/design.2024.96686.
[2] Z. Yu and F. Li, "Research on interface optimization of 3D design software based on eye-tracking test," Design, vol. 8, no. 2, pp. 313-324, 2023, doi: 10.12677/design.2023.82044.
[3] H. Hu, M. Zhang, B. Wu, et al., "Eye-tracking study on visual attention bias of health science popularization posters," China Health Education, vol. 39, no. 6, pp. 523-527, 2023, doi: 10.16168/j.cnki.issn.1002-9982.2023.06.009.
[4] Y. Guo, "Smart advertising design: a visual aesthetic effect improvement based on image data analysis," Evolutionary Intelligence, vol. 16, no. 5, pp. 1699-1705, 2023, doi: 10.1007/s12065-023-00831-5.
[5] L. Huang, "Using crowdsourced visual perception methods to promote the integration of public perspective into urban design decision-making," Landscape Architecture, vol. 32, no. 5, pp. 22-28, 2025, doi: 10.3724/j.fjyl.LA20240101.
[6] H. Zhu and Y. Kong, "Research progress on environmental visual information analysis and perception measurement methods for motion vision," Landscape Architecture, vol. 32, no. 5, pp. 12-21, 2025, doi: 10.3724/j.fjyl.LA20240083.
[7] S. Khan, H. Nguyen P, A. Abdul-Rahman, et al., "Propagating visual designs to numerous plots and dashboards," IEEE Transactions on Visualization and Computer Graphics, vol. 28, no. 1, pp. 86-95, 2021, doi: 10.1109/TVCG.2021.3114828.
[8] Š. Novák J, J. Masner, P. Benda, et al., "Eye tracking, usability, and user experience: A systematic review," International Journal of Human-Computer Interaction, vol. 40, no. 17, pp. 4484-4500, 2024, doi: 10.1080/10447318.2023.2221600.
[9] X. Liu, Q. Xu, T. Liang, et al., "Influencing factors of eye-tracking patterns in facial expression recognition," Applied Psychology, vol. 28, no. 6, pp. 530-541, 2022, doi: 10.3969/j.issn.1006-6020.2022.06.006.
[10] A. Worthy D, N. Lahey J, L. Priestley S, et al., "An examination of the effects of eye-tracking on behavior in psychology experiments," Behavior Research Methods, vol. 56, no. 7, pp. 6812-6825, 2024, doi: 10.3758/s13428-024-02393-5.