

Research Article

Strategies of Applying Visual Element Combination to Improve Visual Cognitive Efficiency in the Era of Big Data Network

Qiaohe Zhang ¹ and Huijuan Lai²

¹Academy of Fine Arts, Huaibei Normal University, Huaibei, Anhui 235000, China

²School of Architecture and Design, Jiangxi University of Science and Technology, Ganzhou, Jiangxi 341000, China

Correspondence should be addressed to Qiaohe Zhang; zhangqh@chnu.edu.cn

Received 7 April 2022; Revised 3 May 2022; Accepted 13 May 2022; Published 3 June 2022

Academic Editor: Chia-Huei Wu

Copyright © 2022 Qiaohe Zhang and Huijuan Lai. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Fundamentally, visual cognition refers to a way of thinking in which the objective information of the external world is transmitted to the brain through the visual nerve and then processed by the brain. In the evolutionary history of human beings, the pressure brought by the environment has led to the rapid development of the visual system and visual cognition. As with language ability and logical ability, the understanding of visual cognition is the key to the analysis of human intelligence. Therefore, research on improving the efficiency of visual cognition is urgently needed. The purpose of this paper is to find methods and strategies to improve the efficiency of visual cognition based on the application of visual element combination. This article first gives a general introduction to visual elements and design methods. Then, the eye movement behavior and visual saliency calculation model are established. Using Gaussian analysis and channel saliency test method based on “where” and “what” principles, the cognitive effect of pictures in visual elements is specifically analyzed. Then, the questionnaire survey method was used to conduct experiments on the problem that visual elements affect the efficiency of visual cognition, and finally the results were obtained. At the macro level, dynamic pictures with color and auxiliary text can effectively improve the cognitive efficiency of vision. It enables people to grasp cognitive objects quickly and form stronger cognitive ability. However, at the micro level, due to the intuitiveness and meaning of words, the visual cognitive efficiency of the recognition effect is high. In terms of image scale, the two scale parameters $[4 \times 4]$ and $[6 \times 6]$ perform better. The human eye is relatively optimal at 4/90~6/90 on the attention scale of the image. At the same time, under the $[4 \times 4]$ scale parameter, even if the image loses some features, it can save about 95% of the cognitive time overhead.

1. Introduction

With the rapid development of Internet big data, cognitive theory has gradually formed. The current situation that the fusion and collision of various visual elements affects the efficiency of visual cognition. Humans acquire and process information from the outside world through the sensory and nervous systems. Of these acquired information, more than 80% come from the visual system [1, 2]. When the human eye processes this visual information, it automatically selects the parts of the brain of interest based on its limited capabilities. This is because human vision has an attention mechanism. In the process of visual cognition, the cognitive subject is stimulated by different external visual elements.

After processing by the brain, it is stored in the brain in the form of memory, thereby forming visual cognition. When different visual elements are presented to the eyes, the brain processes these elements differently. That is, cognitive efficiency is different. Therefore, analyzing and researching the combination of visual elements can not only effectively improve the efficiency of visual cognition and human cognitive ability, but also provide certain ideas and theoretical foundations for research in teaching, design, film and television, and other fields.

The visual system undertakes most of the human information acquisition tasks. The human visual system has efficient data filtering capabilities. It allows humans to quickly extract important information from complex and

changeable scenes and process them cognitively. Visual cognition is a relatively complex process. As the basic unit of cognitive objects, visual elements have a direct impact on visual cognition. Therefore, optimizing the combination of visual elements can effectively improve the visual cognitive efficiency and improve human cognitive ability. On this basis, the improvement of visual cognitive efficiency can also promote the long-term and stable development of related sciences and other fields, such as big data networks, computer vision, bionic technology, visual design, instructional design, and film and television art. And in the process of cognitive practice, it has effectively promoted the solution of many practical problems.

The innovation of this paper is that (1) it focuses on visual elements, analyzes the characteristics and applicability of specific visual elements, and then specifically studies the combination method of visual elements to find methods and strategies to improve visual cognitive efficiency. (2) It integrates eye movement behavior and visual saliency calculation into the analysis and learns the human visual cognitive process and its characteristics. In this way, the analysis of visual elements is promoted and the efficiency of element combination is improved. (3) It uses the Gaussian analysis and channel saliency test methods to accurately simulate the dynamic saccade behavior in the process of human visual cognition to increase the reliability of the data.

2. Related Work

At present, research on visual cognition is becoming more and more common. Among them, Akhmetshin et al. analyzed the application of visual teaching aids in modern educational settings and evaluated its improvement in students' cognitive efficiency in secondary vocational education institutions. The authors used a questionnaire to collect teaching data and students' learning data to find out which errors in audiovisual teaching aids reduced cognitive efficiency. And as these errors were eliminated, students' grades and cognitive efficiency improved by 15–20%. However, due to the limited sample, the obtained results are slightly inaccurate [3]. Pereira et al. investigated the association between eye movement indicators and visual cognitive efficiency in MCI and AD. They believed that visual cognition is critical to activities of daily living and that it is affected in Alzheimer's disease (AD). They tested 127 participants and had them complete an eye-tracking visual cognitive task. They studied its cognitive efficiency. However, due to large individual differences, the results are inevitably imprecise [4]. Wong conducted an exploratory study of a new analytical method to examine visual elements in relation to the underlying visual cognitive processes involved. He believed that the analysis of visual elements proposed by the social semiotics method perfectly complements the internal cognition of image viewers and improves the efficiency of cognition. However, because the method is relatively new and the experimental model is not perfect, the data obtained are slightly lacking [5]. Li et al. analyzed the cognitive efficiency and evaluation index of CNC interface layout design. Based on the eye-tracking data

and the traffic network accessibility model, the calculation equations for the return time rate, refixation rate, saccade duration rate, and accessibility are derived. Finally, the proposed layout evaluation index is tested by correlation analysis, and the cognitive efficiency of visual element design is analyzed by ANOVA. But this operation is more complicated and needs to be carried out in a stable environment [6].

Inspired by the Müller-Lyer illusion in the biological visual system, Zhang et al. proposed an improved LineMod (CT-LineMod) model for visual cognitive template clustering. The model uses 7D cognitive feature vectors in place of standard 3D spatial points in the clustering process of Patch-LineMod. Among them, the cognitive distance of different 3D space points will be further affected by additional 4D information related to feature orientation and size. However, this method is difficult to operate and requires higher requirements on the machine [7]. Xi et al. believed that neural oscillatory responses in different frequency bands are associated with different cognitive functions and cognitive efficiency. Therefore, they investigated the topological characteristics of time-varying networks across multiple frequency bands. This helps to elucidate the mechanism of multisensory multielement integration. They designed an event-related experiment and constructed delta-, theta-, alpha-, and beta-band networks at eight time points after stimulation. Graph theory is then used to compute global properties, node out-degrees, and their correlation to cognitive behavior. This method requires a large amount of data to support. Therefore, there are higher requirements for the accuracy of the algorithm model [8]. Huestegge and Tzsch compared 2 visualization methods: a traditional circular representation (pie chart) and a rectangular representation (constant column width dendrogram). They hypothesized that they differ in cognitive ease and cognitive efficiency in the visual comparison process. Complemented by eye movement analysis, it was concluded that facilitating the comparison process by representing key variables in a less complex visual dimension (i.e., straight line length with constant orientation rather than surface area or curved length) ultimately increases the cognitive efficiency of the graph. However, there are certain errors in this method, resulting in insufficient accuracy [9].

3. Methods of Improving Visual Cognitive Efficiency by Using a Combination of Visual Elements

3.1. Introduction and Design Methods of Visual Elements and Visual Cognition

3.1.1. Text Elements. Text is not only a carrier of information, but obvious visual recognition is also a major feature of text. Therefore, visualizing the text of the design object can make the information more vivid and make the information dissemination more appealing [10].

When designing, text needs to have prominent visual communication elements. By combining arrangement and image processing, the text in the interface can express

different levels of meaning. The text design in the combination of visual elements is the most basic expression method in today's information dissemination. Text can directly convey information to the audience and make the audience understand the composition framework of the interface at a glance [11]. For example, in the design of web pages, it uses text instead of icons as link buttons in the navigation bar. This is because when the audience clicks on the web page, they can intuitively find the content they want to choose without having to process and transform information in their minds, as shown in Figure 1.

Secondly, special fonts are used in the design of thematic advertisement titles, and the effect conveyed is in line with the artistic conception of the photo background, such as the theme advertisements of Halloween and Christmas in Hong Kong Disneyland, as shown in Figure 2.

Text design focuses on the handling of fonts and tones. The visual expression of text styling elements is the foundation of text design. At present, this design has become a basic and important form of artistic expression.

3.1.2. Image Elements. In interface design, designers can transform ideas into striking forms through image design and use these images to convey information. Excellent designers can show high-level image combinations and the characteristics of design objects and make the disseminated information more vivid and three-dimensional through image design [12–14]. Most designers are good at using graphic design. They argue that images have much better propagation characteristics than text. There is also a view that the visual impact of images is 85% higher than that of text. From the above point of view, it can be seen that image design can quickly express the designer's ideas. Using images to display information can bring more sensory stimulation and better emotional resonance.

3.1.3. Color Elements. When visual information is communicated, color is the main driver of emotion [15]. Excellent color matching can attract the attention of the audience and bring vitality to the design object. The sensible use of color can not only make the content more visually beautiful, but also convey emotion. This plays a very important role in the whole design process. The color configuration of the design object will directly affect the final design effect. The colors of each part need to be blended with each other and presented harmoniously. How to match colors to highlight the design theme is a problem that designers need to focus on. Excellent color matching needs to first figure out what the main idea of the design is and then make color matching around this subject, which can resonate with the audience.

If it is going to work well, it needs vibrant colors first. In the design, it is necessary to use color to highlight the subject and make the subject vivid so as to attract the audience's attention at the first time. Furthermore, it also requires unique colors. Innovation is the vitality of design. Blind imitation will make the design lose its vitality [16]. Only innovation in color configuration can impress the audience.

As shown in Figure 3, the designer used a magnificent triangular element. In the work, the designer added very interesting colorful triangles. This seemingly simple and bland element makes this work full of strong individualism. These rich colors form a sharp contrast with the deer head in the background, giving the audience a greater impact. It deepens the audience's visual cognition and leaves a deep impression on the audience. The colorful triangular color elements are matched with animal and portrait photography to create magnificent avant-garde and fashionable works.

3.1.4. Visual Element Combination Presented by Interactive Element Design. Visual element composition is an organic and systematic design arrangement for each element present in the interface [17]. That is, at the beginning of the design, the designer is required to comprehensively consider the layout of images, text, and colors according to the design requirements and goals and design with the overall coordination as the goal [18, 19]. In this way, the interactive design that conforms to the audience's cognitive model can improve the audience's cognitive ability and increase the amount of cognition and cognition accuracy per unit time. Thereby, it improves the visual cognition efficiency of the audience. The following methods of systematic integration of visual elements can help improve visual cognitive efficiency:

(1) *Unification of the Spacing between Elements.* The concept of "spacing" is also a very important element in the system integration design. The spacing between elements includes the spacing between images, between images and text, and between text and text [20]. Unifying the spacing of each element helps to form an orderly layout effect, improving the reading efficiency of content information and the overall sense of layout.

(2) *Unification of Image Forms.* Among the visual composition elements, images are the most visually expressive design elements [21]. Therefore, it is extremely important to choose the most appropriate image based on the message you want to convey and what you want to express. How to make the overall layout of the image harmonious and ornamental can start with the expression style of the image. It selects images of the same or similar styles to integrate the layout faster, removes images with different styles that destroy the overall harmony, or unifies the styles of images through image processing software to reduce the impact of differences on the layout. It creates a messy feel. Secondly, the arrangement position of the images is also a key factor. The overall arrangement of the image should be arranged in advance to form a certain design order so that the overall image is harmonious and orderly.

(3) *Unification of Text Style.* In the combination of elements, the type and arrangement of text can also have an impact on the visual perception of the audience. Therefore, when designing, designers need to choose the appropriate font style according to the content they want to convey. Black body gives a sense of fashion and eye-catching, while Arial gives a sense of elegance, classic and high-end. Different fonts bring different feelings to people. Showing the

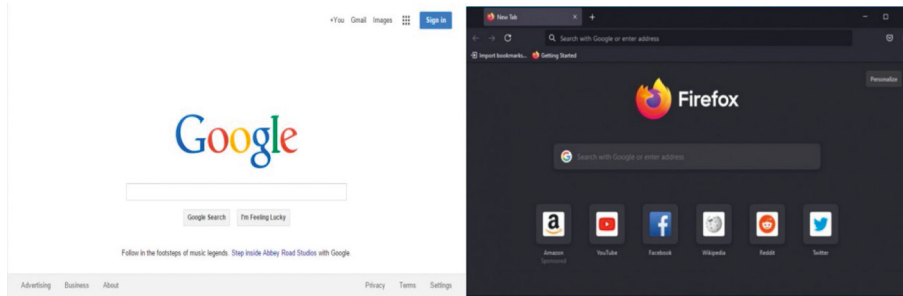


FIGURE 1: Navigation bar links in web pages.



FIGURE 2: Hong Kong Disneyland's Halloween and Christmas themed advertisements.

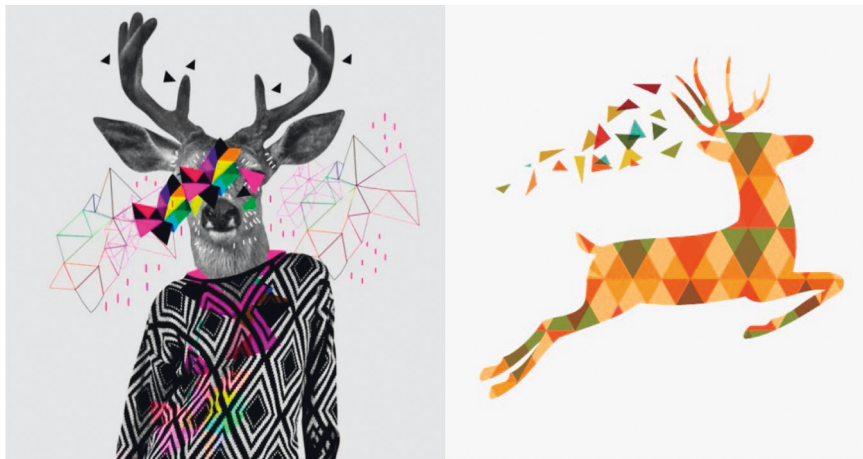


FIGURE 3: Colorful triangular color elements and animals.

personality of the font can not only reflect the obvious difference between different types of information, but also make the layout effect overall beautiful.

(4) *Harmony of Color Planning*. Color has the power to move people's hearts in the combination of elements, which is beneficial to convey the designer's emotion and make the audience resonate with it [22]. Through color planning, all elements in the layout can also be coordinated to form an overall, beautiful and eye-catching picture. In the specific design, it can choose a color as the keynote to connect all the elements together, or it can form a strong color contrast, leaving the audience with an intuitive sense of visual impact. The flexible use of color helps to enhance the overall visual effect of viewing, visual impact and memory, and enhance the cognitive ability of the audience.

(5) *Visual Cognitive Structure*. Visual cognition is divided into three stages of visual processing: early, middle, and late stages, as shown in Figure 4. Appearance information is obtained in the early stage, including the confirmation of the contained elements and the acquisition of general characteristics. It then performs shallow cognitive analysis. In the middle stage, a series of operations are performed on the initial impression to obtain an image that can reflect the specific characteristics of the cognitive object, which is called a two-dimensional (2.5D) sketch, including depth information and element clarity extracted by stereo vision operations. Based on this information, cognitive objects can be segmented into regions with clear meanings. From this, a higher-level description can be obtained than lines, shapes, and ranges. The later stage

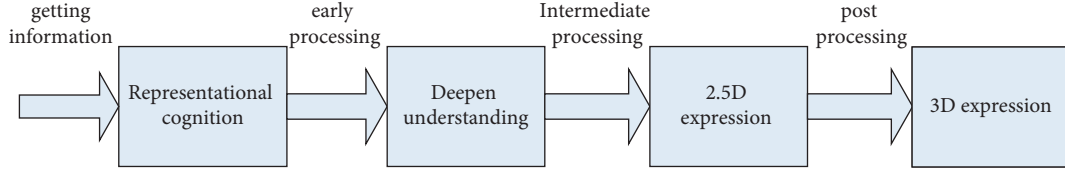


FIGURE 4: Visual cognitive structure.

builds a three-dimensional model of the process of cognitive object recovery and relies on self-cognition to identify the object.

According to the visual cognitive model in Figure 5, when the information comes into the eyes and stimulates the retina, the optic nerve will transmit the received visual information to the brain, and the brain will process it to generate cognitive behavior [23]. That is to say, visual cognition is a series of analysis and understanding processes carried out by the brain after collecting and sorting out external information. The common template theory and visual feature recognition are theories about visual cognition. It shows that after human beings acquire information, they produce a reaction process through cognitive judgment. These pieces of information composed of words, images, and symbols stimulate the brain through the optic nerve to generate cognition. These pieces of information will be stored in the brain in the form of memory to form a corresponding cognitive judgment mechanism.

3.2. Eye Movement Behavior and Visual Saliency Calculation Model. In the perceptual field of cognitive psychology, visual cognition is the main channel to obtain the vast majority of information in the external world [24]. However, human attention and energy are limited, and only certain resources are allocated to the visual system. At this time, visual selective attention will play an important role in this process. The response of visual selective attention at the human physiological level is mainly eye movement. In the research of visual cognitive efficiency, eye-tracking technology is an effective and direct research method. It can learn the human visual cognitive process and its characteristics by observing eye movements and the distribution of viewpoints [25]. This is applied to the combination of visual elements to improve the efficiency of the combination of elements and improve the efficiency of visual cognition.

3.2.1. Gaussian Analysis. As a very classic data analysis technique in statistics, projection tracking is highly similar to the dynamic saccade behavior in human visual cognition process [26]. It was originally proposed to exploit low-dimensional projection search to mine potentially meaningful data patterns or structures in high-dimensional data. This is in line with the human cognitive goal of searching and discovering important visual content in complex scenes. A linear projection from Y^d to Y^k can be defined as any linear mapping M (usually a $k \times d$ matrix of rank k):

$$\begin{aligned} W_Q &= MW, \\ W &\in Y^d, \\ W_Q &\in Y^k. \end{aligned} \quad (1)$$

Among them, W is a dimension w , a random variable obeying the distribution Q . W_Q is a random variable of dimension w that obeys distribution Q_M . Each column of W_Q can be considered an independent statistical component. The process of projection tracking is to determine the projection M of interest by maximizing or minimizing an objective function $F(Q_M)$ with respect to the distribution Q_M .

3.2.2. Extraction of Gaussian Components Based on Kurtosis Maximization. In statistics, the super-Gaussianity of a random variable is usually measured by the kurtosis function:

$$\text{kurt}(x) = E\{x^4\} - 3(E\{x^2\})^2. \quad (2)$$

Among them, x is a given random variable and $E\{\cdot\}$ is the expectation function. If x is a random variable strictly obeying a Gaussian distribution, then $\text{kurt}(x) = 0$. If the kurtosis is positive, then the random variable has a super-Gaussian property. If the kurtosis is negative, it has a sub-Gaussian shape. In order to obtain the most Gaussian projection, an initial projection M can be randomly selected. It utilizes a fixed-point iteration method such that the final converged projection has the strongest kurtosis on a given image matrix J . The objective function F_Q of Gaussian component projection tracking is

$$F_Q(M) = \text{kurt}(M^T J). \quad (3)$$

The gradient of F_Q is as follows:

$$\frac{\partial F_Q}{\partial M} = 4 \left[E\left\{ (M^T J)^3 J^T \right\} - 3M^T \|M^T\|^2 \right]. \quad (4)$$

When the projection direction is the same as the gradient direction or the deviation of the two is small enough, the search process converges. Then, the following can be derived:

$$M^T \propto \left[E\left\{ (M^T J)^3 J^T \right\} - 3M^T \|M^T\|^2 \right]. \quad (5)$$

Formula (5) can be solved using a fixed-point iterative algorithm:

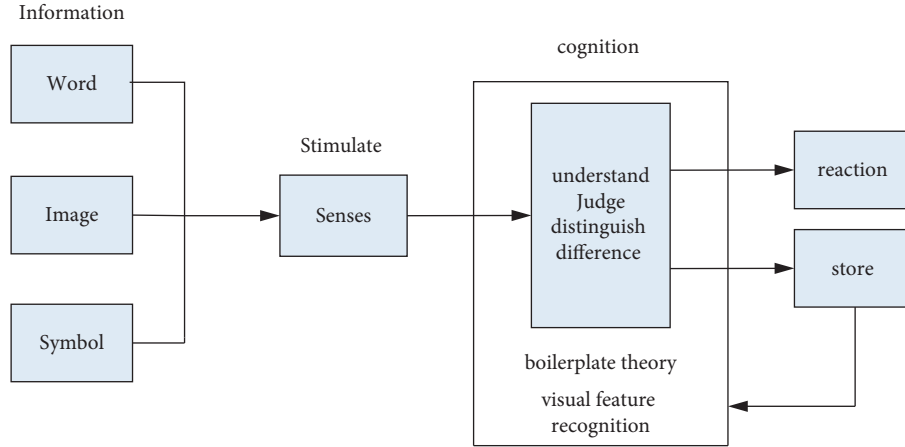


FIGURE 5: Visual cognitive model.

$$M \leftarrow E \left\{ J(J^T M)^3 \right\},$$

$$M = \frac{M}{\|M\|}. \quad (6)$$

From the above, a projection direction can be obtained so that the given data has the strongest super-Gaussianity in this direction.

3.2.3. Multi-Gaussian Component Extraction Based on Schmidt Orthogonalization. In order to ensure that the optimization process will not converge to the extracted Gaussian components, an orthogonalization constraint needs to be introduced in the process of multicomponent extraction. That is, the new Gaussian component should be strictly orthogonal to the previously extracted component. It is given a set of extracted orthogonal projection vectors M_1, M_2, \dots, M_Q , and the orthogonalization constraint of multiple solving processes can be guaranteed by the processing of the following formula:

$$M_{Q+1} \leftarrow M_{Q+1} - \sum_{s=1}^Q (M_{Q+1}^T M_s M_s) M_s. \quad (7)$$

To sum up the above, the Gaussian components in the image data can be extracted one by one.

3.3. Channel Saliency Test Method Based on “Where” and “What” Principles

3.3.1. “Where” Saliency Map Generation Based on Pure Background. In the human visual cognitive system, when the vision is stimulated, the cognitive process of the nerve to the outside world mainly depends on the sparse response state received by the neurons [27]. The segmented image is sparsely represented (SR) using the pure background after low-rank matrix factorization as a dictionary, and the sparse encoding of the entire image region is

$$\alpha_j = \arg \min \|X_j - D\alpha_j\|_2^2 + \gamma \|\alpha_j\|_1. \quad (8)$$

Among them, the parameter γ is used to balance the sparse constraint term $\|\alpha_j\|_1$ and the error reconstruction term $\|X_j - D\alpha_j\|_2^2$. Then, the saliency value based on the pure background can be regarded as the reconstruction error after sparse coding (SC), which is expressed as

$$S_j^{bg} = \|X_j - D\alpha_j\|_2^2. \quad (9)$$

The “where” feature saliency map can be binarized by the adaptive threshold T_g , and the superpixel area corresponding to the “where” salient area part will be separated. To sum up, the results of time $T_g = 0.8 * t$ are the best.

3.3.2. Target-Centered “What” Saliency Map Generation. In biology, each neuron in a neural network is interconnected with surrounding neurons to form a crisscrossed neural network. These neurons can work in parallel and can quickly find the information that needs to be processed [28]. On the basis of the modified mechanism, a fully connected graph $S = \{C, G\}$ is established. Among them, C is the set of all superpixel nodes $X = \{X_1, \dots, X_1, X_{j+1}, \dots, X_n\} \in R^{m \times n}$; there are j marked points and u unmarked points, and the marked points are denoted as $(X_1, Y_1), \dots, (X_j, Y_j)$. So, $N = j + u$ is the number of superpixels after image segmentation. S represents the degree of correlation between two superpixel nodes, which is defined as

$$\omega_{j,k} = s^{-\|c_j - c_k\|/2v^2}, j, k \in V. \quad (10)$$

Among them, v is a constant that controls the size of the weight. c_j and c_k represent the color features of the CIELAB color space corresponding to different superpixel nodes, respectively.

The salient values of each node are computed using graph-based Gaussian fields and harmonic functions. Given a graph-based function $f: V \rightarrow R$ and some fixed labels, it is agreed that in the label data $j = 1, \dots, l$, the function f

takes the value $f(j) = f_l(j) = y_j$. Therefore, to mark the significant values of nodes,

$$S(F) = \infty \sum_{j \in L} (f_j - y_j)^2 + \frac{1}{2} \sum_{j,k} \omega_{j,k} (f_j - f_k)^2, \quad f_j \in R. \quad (11)$$

It converts to a graph-based Laplacian matrix:

$$S(F) = \infty \sum_{j \in L} (f_j - y_j)^2 + F^T L F = F^T L F. \quad (12)$$

Among them,

$$\begin{aligned} L &= D - A, \\ A &= [\omega_{j,k}]_{n \times n}, \\ D &= \text{diag}(d_j), \\ d_j &= \sum_k \omega_{j,k}. \end{aligned} \quad (13)$$

With a slight variation, it gets

$$f(j) = \frac{1}{d_k} \sum_{j-k} \omega_{j,k} f(j), \quad k = l+1, l+2, \dots, l+u. \quad (14)$$

For the convenience of calculation, take the l th row of the weight matrix A as the benchmark, divide it into four parts, and get

$$A = \begin{bmatrix} A_{ll} & A_{lu} \\ A_{ul} & A_{uu} \end{bmatrix}. \quad (15)$$

In the same way, it will be divided into two parts to get

$$F = \begin{bmatrix} F_l \\ F_u \end{bmatrix}. \quad (16)$$

Among them, F_u represents the value of the unlabeled node; then, the solution of the optimization problem is

$$F_u = (D_{uu} - A_{uu})^{-1} A_{ul} F_l. \quad (17)$$

Therefore, the object-centric saliency based on the "what" feature is

$$Y^{\text{goal}}(j) = F(j), \quad j = 1, 2, \dots, n. \quad (18)$$

4. Visual Element Model and Sample Experiment under the Guidance of SGP

4.1. Performance of the Model. Based on the SGP framework, it selects four single samples and compares them in the human eye gaze prediction experiment. First, stretch the image so that the longest side has 90 pixels. Then, four single-scale block sampling strategies of $[2 \times 2]$, $[4 \times 4]$, $[6 \times 6]$, and $[8 \times 8]$ are used to calculate the saliency of the image. Finally, quantitative evaluation is carried out based on the YORK-120 database.

Looking at the absolute value of performance, the multiscale fusion method achieves the best performance on

all four samples. $[4 \times 4]$, $[6 \times 6]$ have similar performance and have a small gap with multiscale fusion strategies. The gap between $[2 \times 2]$ and $[8 \times 8]$ is larger, and $[2 \times 2]$ is better than $[8 \times 8]$. In terms of relative performance, the multiscale fusion method achieves better results than other strategies on 40% of the images. The proportion of $[4 \times 4]$ is 25.36%, and the images of $[2 \times 2]$, $[6 \times 6]$, and $[8 \times 8]$ are 16.26%, 16.26%, and 17%, respectively.

The optimal processing effects obtained by different scaling strategies are shown in Table 1; it can be seen from the table that the test results of the IL-KL model are ideal, with a higher mean value, while the test results of the SL-KL model are not ideal, with the lowest mean value compared to other methods.

Figure 6 shows the optimal processing effect of different scale strategies. The red box is the sampling window size of the corresponding scaling strategy. To sum up, it can be concluded that $[2 \times 2]$ has better effect in single-color images because it does not encode structural information. $[4 \times 4]$ to $[6 \times 6]$ include most common salient objects, such as mid-range faces, distant people, and signs. This also shows the superior performance of $[4 \times 4]$ and $[6 \times 6]$ in Table 1 under a single-scale evaluation. $[8 \times 8]$ includes close-range faces, mid-range advertisements, and long-range people. It can be seen that the multiscale fusion strategy is more effective for scene processing of salient objects at multiple scales.

Combined with the above analysis, under a single-scale evaluation, $[4 \times 4]$ and $[6 \times 6]$ have better performance. Therefore, in the attention scale of the image, 4/90~6/90 is relatively optimal, and the proportion of the area of the field of view is roughly 0.21%~0.44%. On top of this, the human visual cognition efficiency is better.

4.2. Quantization under the Influence of Feature Dimension.

Traditional saliency models extract feature vectors of specific dimensions from images with a fixed algorithm [29]. High-dimensional features do not always bring performance improvements, and complete feature representations are not optimal. Under the scale parameter of $[6 \times 6]$, the complete feature dimension is 80. The best performance of the model is under 45 dimensions, which is 3.87% higher than using 1-dimensional features. With a scale parameter of $[4 \times 4]$, the complete feature dimension is 30. When the performance of the model is optimal at dimension 25, it is improved by 4.53% compared to using 1-dimensional features. In the case of only extracting 1-dimensional features, even if some performance is lost, about 95% of the time overhead can be saved. Figure 7 shows the impact on the time overhead and SL-AUC performance metrics of the SGP model based on different numbers of Gaussian components.

The SGP model test results of $[4 \times 4]$ and $[6 \times 6]$ are obtained as shown in Figure 8.

As can be seen from Figures 7 and 8, in the scale parameters of $[4 \times 4]$ and $[6 \times 6]$, the complete feature dimension and the performance of the model generally show a positive growth mode. While the feature dimension is increasing, the cumulative time is also increasing, and the performance of the model is constantly being optimized. The

TABLE 1: Evaluation results obtained by different scale strategies.

| | $[2 \times 2]$ | $[4 \times 4]$ | $[6 \times 6]$ | $[8 \times 8]$ | Ours—MS |
|--------|----------------|----------------|----------------|----------------|---------|
| SL-AUC | 0.7795 | 0.8223 | 0.7800 | 0.7514 | 0.7905 |
| SL-KL | 0.2894 | 0.3352 | 0.3180 | 0.3322 | 0.3927 |
| IL-AUC | 0.7799 | 0.7664 | 0.7905 | 0.7438 | 0.8042 |
| IL-KL | 1.0760 | 1.4323 | 1.5230 | 1.3022 | 1.6176 |

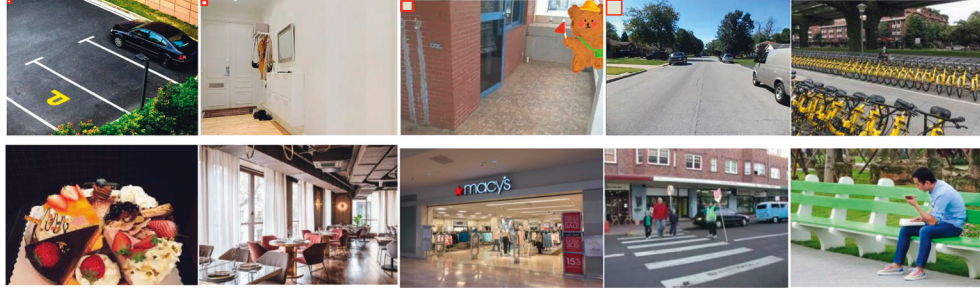


FIGURE 6: Comparison of sampling strategies at different scales.

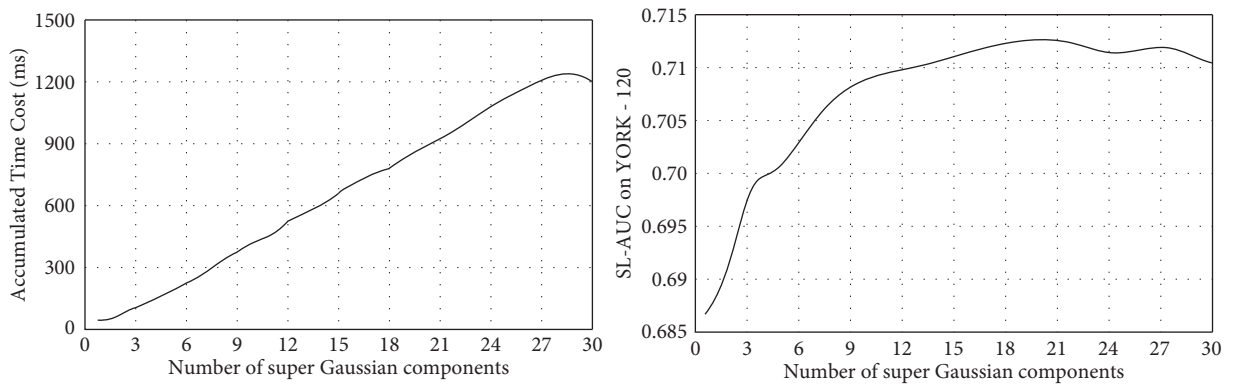
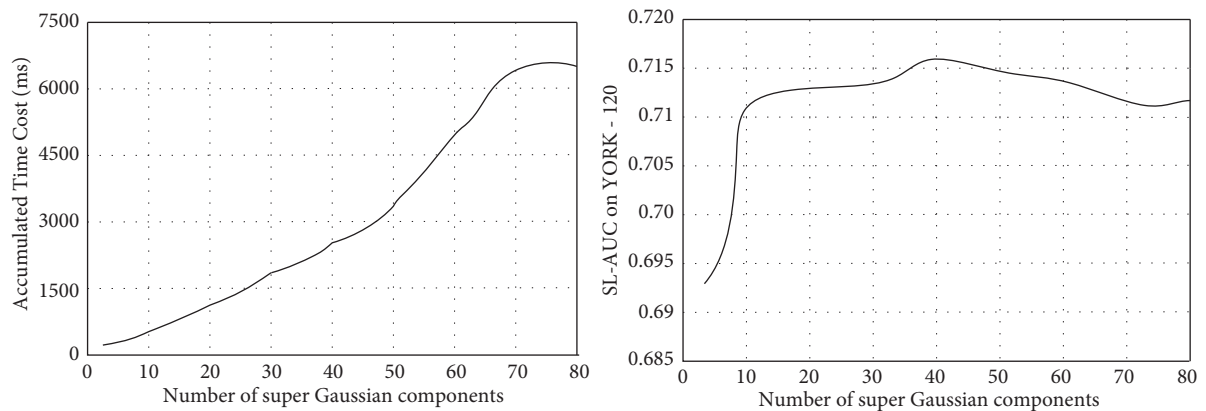


FIGURE 7: Cumulative time overhead and impact of SL-AUC performance indicators under different feature dimensions.

FIGURE 8: SGP model test results for $[4 \times 4]$ and $[6 \times 6]$.

SL-AUC index has a large increase rate in the early stage, a slow growth trend in the middle stage, and a partial decline trend in the later stage.

4.3. *Experiment on the Effect of Visual Elements on Visual Cognitive Efficiency.* Studies have shown that pictures with the aid of words can produce better cognitive effects. The

cognitive effect of words with color-coded key content is better than that of text without color, and pictures with colors are better than pictures without color. In the following experiments, we will specifically study the impact of text and pictures (including static and dynamic) on people's visual cognitive efficiency based on the influence of color and auxiliary text so as to find ways to improve visual cognitive efficiency. This test is mainly conducted from three aspects: self-cognitive evaluation, recognition, and reasoning test.

(1) *Experimental Steps*. This experiment was mainly assisted by junior high school students from Huaihai Middle School. The subjects of the experiment were 102 students (44 boys and 58 girls) in two classes. The level of learning of the students is about the same. The test includes three parts: personal self-perception evaluation, recognition questions, and reasoning questions. The time to complete the test is 45 minutes.

(2) *Experimental Results*. In this experiment, the average score and standard deviation of the visual element combination presentation method were calculated. The three aspects of self-cognitive evaluation, recognition, and reasoning are analyzed.

4.3.1. Differences in Self-Perception Assessment Test Results. The average score and average variance of self-cognitive evaluation of text, static pictures, and dynamic pictures were analyzed, respectively, and the results are shown in Table 2. Compared with static pictures, the F text test value is 195.405, the significance level is less than 0.05, and the visual cognitive effect is better. Compared with dynamic pictures, the F -test value of static pictures is 62.181, the significance level is less than 0.05, and the visual cognitive effect is better. The average score and average variance of the self-perception evaluation of text are the lowest, followed by static pictures, and dynamic pictures are the best. Therefore, it can be concluded that in terms of self-perception evaluation, dynamic pictures (including color and auxiliary text) > static pictures (including color and auxiliary text) > text (with color). The results of the self-perception assessment are as shown in Figure 9.

4.3.2. Recognition Test Results' Discrepancies. The results of the average score and average variance of the recognition test are shown in Table 3. Compared with static pictures, the F -test value of text is 74.873, the significance level is less than 0.05, and the visual cognitive effect is better. Compared with dynamic pictures, the F -test value of static pictures is 22.648, the significance level is less than 0.05, and the visual cognitive effect is better. It can also be found from the data that the average recognition score of the three visual elements has reached more than 3 points; that is, they have reached the passing level. However, the average score for recognition of text is higher than that of pictures with color and auxiliary text, followed by dynamic pictures, and the average score of static pictures is the lowest. It is concluded from this that in the cognition of new knowledge, learners' recognition of words has the highest effect. The recognition efficiency of static pictures is lower than that of text. It shows that even

though the picture can contain rich information, it cannot significantly represent each knowledge point. This affects the cognitive efficiency of learners. The recognition efficiency of dynamic pictures is the lowest, indicating that although dynamic pictures can attract learners' attention and express knowledge points, they cannot present all the information, which greatly reduces cognitive efficiency. The results of the recognition test are as shown in Figure 10.

4.3.3. Differences in Reasoning Test Results. The results of the average score and average variance of the recognition test are shown in Table 4. The above experiments show that, compared with static pictures, the F -test value of text is 92.367, the significance level is less than 0.05, and the visual cognitive effect is better. Compared with dynamic pictures, the F -test value of static pictures is 49.893, the significance level is less than 0.05, and the visual cognitive effect is better. It can also be found from the data that the average score of text reasoning is the lowest, and the effect is the least ideal, followed by static pictures, and the reasoning effect of dynamic pictures is the best. It is concluded from this that reasoning can examine the ability of learners to grasp knowledge points. The text has a good effect in recognition, but it is slightly lacking in the reasoning test. Static pictures can better grasp the knowledge system at a macro level, allowing learners to better understand and expand the material. Therefore, the inference effect of static pictures is better. Dynamic pictures can enhance the presentation of key knowledge. This is helpful for learners to grasp the material and improve cognitive efficiency so that learners can quickly build a cognitive framework of knowledge points and avoid the interference of other factors. The results of the reasoning test are as shown in Figure 11.

5. Discussion

This paper is devoted to studying strategies for applying visual element combinations to improve visual cognitive efficiency. It is not only an elaboration and analysis of visual elements, eye movement behavior and visual saliency calculation model, using eye movement behavior and visual saliency calculation model, Gaussian analysis method, SGP calculation model, and other methods to analyze the combination of visual elements, but also a new attempt to improve the research method of visual cognitive efficiency strategy. Through the analysis and construction of the visual saliency calculation model and the investigation and research of students, it looks for ways to improve visual cognitive efficiency and improve cognitive performance.

Through the analysis of this case, it is shown that $[4 \times 4]$ and $[6 \times 6]$ have better performance under a single-scale evaluation. On the attention scale of the image, $4/90 \sim 6/90$ is relatively optimal. In terms of visual elements, in the self-perception evaluation, the self-perception evaluation of text is positively correlated with the reasoning effect. That is, the impact of words on self-cognitive evaluation and reasoning effect is small, and the visual cognitive efficiency is low. And pictures have a greater impact on my self-cognitive

TABLE 2: Mean scores and mean variances of self-perception evaluations for the presentation of various visual elements.

| | Self-cognitive evaluation | |
|---------------|---------------------------|---------------|
| | Average score | Mean variance |
| Text | 9.8 | 0.713 |
| Static image | 13.5 | 0.945 |
| Dynamic image | 14.2 | 0.801 |

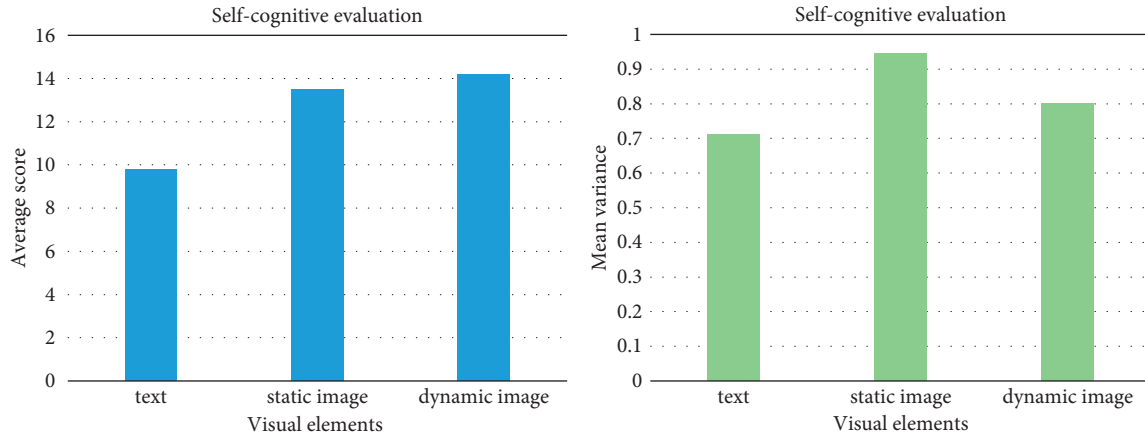


FIGURE 9: The self-perception evaluation test results of the presentation methods of each visual element.

TABLE 3: The average score and average variance of the recognition test for each visual element presentation method.

| | Re-recognize | |
|---------------|---------------|---------------|
| | Average score | Mean variance |
| Text | 4.7 | 0.879 |
| Static image | 4.1 | 0.414 |
| Dynamic image | 4.3 | 0.921 |

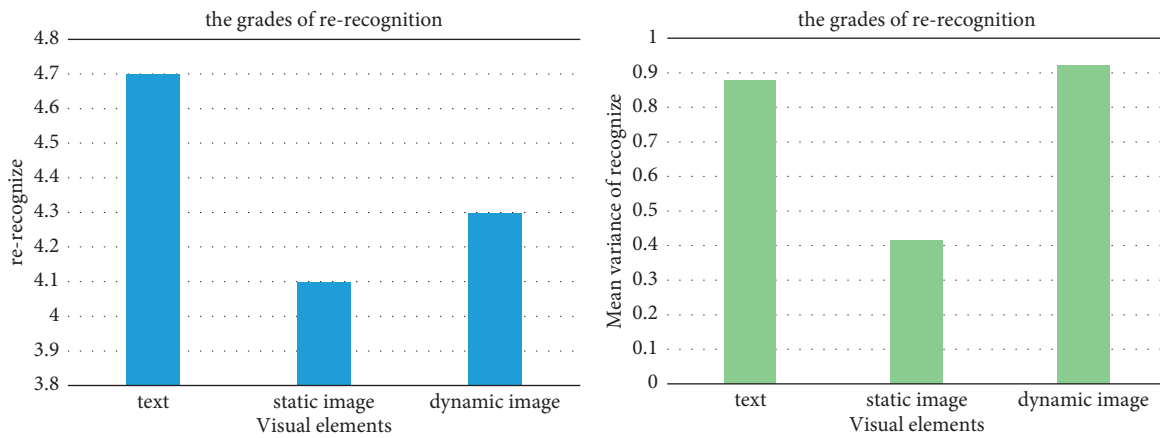


FIGURE 10: Recognition test results of the presentation methods of each visual element.

evaluation and reasoning effect, and the visual cognitive efficiency is higher. Among them, dynamic pictures are better than static pictures, and the visual cognition efficiency

of dynamic pictures is the best. That is to say, at the macro level, dynamic pictures with colors and auxiliary text can effectively improve the cognitive efficiency of vision so that

TABLE 4: Mean scores and mean variances of the reasoning test for each visual element presentation method.

| | Reasoning | |
|---------------|---------------|---------------|
| | Average score | Mean variance |
| Text | 2.6 | 0.698 |
| Static image | 3.1 | 0.703 |
| Dynamic image | 3.3 | 1.114 |

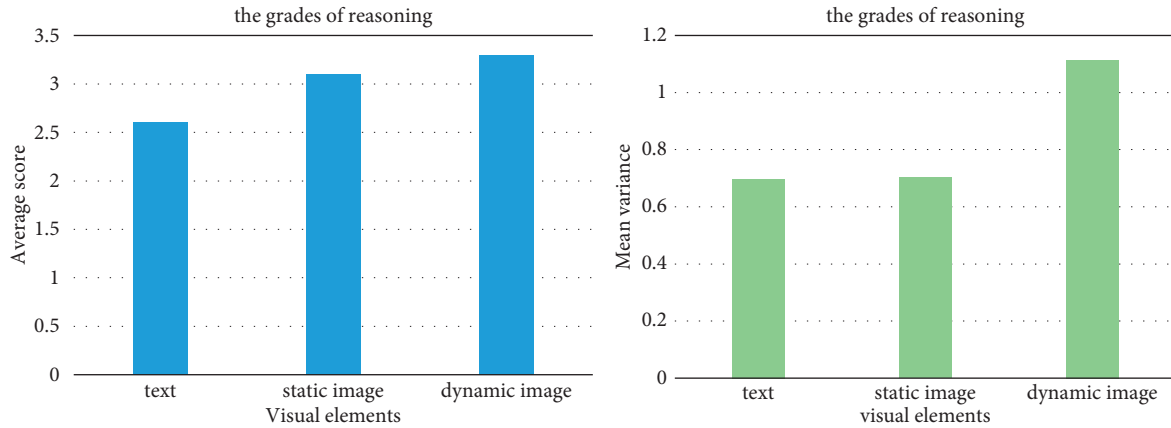


FIGURE 11: Reasoning test results for the presentation of various visual elements.

people can grasp cognitive objects quickly and form strong cognitive ability. However, at the micro level, due to the intuitiveness and meaning of words, the visual cognitive efficiency of the recognition effect is high. Words can better help people to reunderstand things and effectively consolidate their cognition of things.

6. Conclusions

Through the analysis of this paper, the following conclusions are drawn: (1) Color can have a good auxiliary effect on the cognitive level of text and pictures. That is, the presentation method with color has a good effect on improving the visual cognitive efficiency. The increase in the use of color can play a better effect. (2) The aid of words can effectively deepen the understanding of the recognizing memory of the recognized object. Text rendering is the best. (3) The visual cognitive efficiency of pictures is the best in self-cognitive evaluation. And the performance of $[4 \times 4]$ is relatively optimal. The corresponding image area is about the size of a face at a medium distance or a sign at a distance. (4) Through eye movement behavior simulation and experiments, a significant Gaussian prior is found, which proposes an SGP framework model to simulate eye movement behavior. First, four samples were selected, and four different scale blocks were used as examples to calculate the saliency. Then, based on the YORK-120 database, data analysis and evaluation were performed to compare the visual cognitive efficiency of the four scale blocks. Experiments show that the SGP model has strong robustness and rationality. In the case of only extracting 1-dimensional features, even if part of the performance of the image is lost, it can save about 95% of the time overhead, and

the optimal field of view ratio is calculated to be about 0.21%~0.44%. In the method, the channel significance test method based on the “where” and “what” principles is adopted. This method effectively improves the accuracy of the experimental results. And even if the background of the experimental subject is more complex, more accurate results can be obtained. (5) Although these methods and experiments can improve the detection accuracy, there are also some shortcomings: in the saliency test method using low-rank matrix factorization, the accuracy of the saliency map is very dependent on the results of superpixel segmentation. If the processing result of superpixel segmentation is not ideal, it will affect the accuracy of the saliency map. Therefore, the preprocessing step for the saliency map is where attention and improvement are needed.

Data Availability

The data that support the findings of this study are available from the corresponding author upon reasonable request.

Conflicts of Interest

The authors declare no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Acknowledgments

This work was supported by Huaibei Normal University Top Curriculum Construction Foundation (Grant no. 2021ZLGC071).

References

- [1] H. Wei, A. Sehgal, and N. Kehtarnavaz, "A deep learning-based smartphone app for real-time detection of retinal abnormalities in fundus images," in *Proceedings of the real-time image processing and deep learning*, Baltimore, MD, USA, May 2019.
- [2] K. Shankar, Y. Zhang, Y. Liu, L. Wu, and C. H. Chen, "Hyperparameter tuning deep learning for diabetic retinopathy fundus image classification," *IEEE Access*, vol. 1, no. 99, 2020.
- [3] E. M. Akhmetshin, R. R. Ibatullin, A. R. Gapsalamov, V. L. Vasilev, and S. Y. Bakhvalov, "Audiovisual aids application in the secondary-level vocational education establishments: Efficiency analysis and assessment," *The International Journal of Educational Management*, vol. 33, no. 2, pp. 374–392, 2019.
- [4] M. Pereira, M. Camargo, A. Bellan, A. Tahira, and O. Forlenza, "Visual search efficiency in mild cognitive impairment and alzheimer's disease: an eye movement study," *Journal of Alzheimer's Disease: JAD*, vol. 75, no. 1, pp. 1–15, 2020.
- [5] L.-Y. M. Wong, "Analysing visual imagery: connecting semiotic and cognitive perspectives," *Cognitive Semantics*, vol. 4, no. 1, pp. 39–75, 2018.
- [6] J. Li, S. Yu, and W. Liu, "Cognitive characteristic evaluation of CNC interface layout based on eye-tracking," *Jisuanji Fuzhu Sheji Yu Tuxingxue Xuebao/Journal of Computer-Aided Design and Computer Graphics*, vol. 29, no. 7, pp. 1334–1342, 2017.
- [7] T. Zhang, Y. Yang, Y. Zeng, and Y. Zhao, "Cognitive template-clustering improved LineMod for efficient multi-object pose estimation," *Cognitive Computation*, vol. 12, no. 4, pp. 834–843, 2020.
- [8] Y. Xi, Q. Li, M. Zhang, L. Liu, and J. Wu, "Characterizing the time-varying brain networks of audiovisual integration across frequency bands," *Cognitive Computation*, vol. 12, no. 6, pp. 1154–1169, 2020.
- [9] L. Huestegge and T. H. Tzsch, "Integration processes during frequency graph comprehension: performance and eye movements while processing tree maps versus pie charts," *Applied Cognitive Psychology*, vol. 32, no. 2, pp. 200–216, 2018.
- [10] F. Bidel, M. J. Jafari, S. Khodakarim, and A. S. S. Abadi, "Evaluation of the effect of heat stress on cognitive performance and physiological parameters of the students," *Iran Occupational Health*, vol. 17, no. 1, pp. 1–15, 2021.
- [11] M. Belk, C. Fidas, P. Germanakos, and G. Samaras, "The interplay between humans, technology and user authentication: a cognitive processing perspective," *Computers in Human Behavior*, vol. 76, no. nov, pp. 184–200, 2017.
- [12] J. Hojin, K. Jinhee, and K. Eunjoo, "Behavioral approach system (BAS) sensitivity and functional brain networks during rest: graph-theory analysis," *The Korean Journal Of Cognitive And Biological Psychology*, vol. 29, no. 2, pp. 105–125, 2017.
- [13] O. I. Khalaf, C. A. T. Romero, A. A. J. Pazhani, and G. Vinuja, "VLSI implementation of a high-performance nonlinear image scaling algorithm," *Journal of Healthcare Engineering*, vol. 2021, Article ID 6297856, 10 pages, 2021.
- [14] S. Ding, S. Qu, Y. Xi, and S. Wan, "Stimulus-driven and concept-driven analysis for image caption generation," *Neurocomputing*, vol. 398, pp. 520–530, 2019.
- [15] S. Chen, Z. Jian, Y. Huang, Y. Chen, Z. Zhou, and N. Zheng, "Autonomous driving: cognitive construction and situation understanding," *Science China Information Sciences*, vol. 62, no. 008, pp. 1–27, 2019.
- [16] S. Yalin, T. Sohraab, Y. Erpek, Y. Shi, and J. Li, "A unified solution to cognitive radio programming, test and evaluation for tactical communications," *IEEE Communications Magazine*, vol. 55, no. 10, pp. 12–20, 2017.
- [17] X. Xu, Y. Li, and Q. Wu, "A multiscale hierarchical threshold-based completed local entropy binary pattern for texture classification," *Cognitive Computation*, vol. 12, no. 1, pp. 224–237, 2020.
- [18] S. Zhou, M. Ke, and P. Luo, "Multi-camera transfer GAN for person re-identification," *Journal of Visual Communication and Image Representation*, vol. 59, pp. 393–400, 2019.
- [19] C. Grange and H. Barki, "The nature and role of user beliefs regarding a website's design quality," *Journal of Organizational and End User Computing*, vol. 32, no. 1, pp. 75–96, 2020.
- [20] H. Jia, H. Li, and D. Yu, "The relationship between ERP components and EEG spatial complexity in a visual Go/Nogo task," *Journal of Neurophysiology*, vol. 117, no. 1, pp. 275–283, 2017.
- [21] P. Klobošáková, R. Mareček, J. Fousek, E. Výtvarová, and I. Rektorová, "Connectivity between brain networks dynamically reflects cognitive status of parkinson's disease," *A Longitudinal Study. Journal of Alzheimer's disease: JAD*, vol. 67, no. 3, pp. 971–984, 2019.
- [22] J. S. Anni and A. K. Sangaiah, "Wireless integrated sensor network: boundary intellect system for elephant detection via cognitive theory and fuzzy cognitive maps," *Future Generation Computer Systems*, vol. 83, no. JUN, pp. 522–534, 2017.
- [23] J. Lees, P. G. Michalopoulou, S. W. Lewis, S. Preston, C. Bamford, and T. Collier, "Modafinil and cognitive enhancement in schizophrenia and healthy volunteers," *the effects of test battery in a randomised controlled trial. Psychological medicine*, vol. 47, no. 13, pp. 2358–2368, 2017.
- [24] A. L. Moore, D. M. Carpenter, T. Miller, and C. Ledbetter, "ThinkRx cognitive training for adults over age 50: clinician-caregiver partners in delivery as effective as clinician-only delivery," *Psychology & Neuroscience*, vol. 12, no. 2, pp. 291–306, 2019.
- [25] M. Ding, "Application of visual elements in product paper packaging design: an example of the "squirrel" pattern," *Journal of Intelligent Systems*, vol. 31, no. 1, pp. 104–112, 2022.
- [26] A. Grinstein, H. Hagtvedt, and A. Kronrod, "Aesthetically (dis)pleasing visuals: a dual pathway to empathy and prosocial behavior," *International Journal of Research in Marketing*, vol. 36, no. 1, pp. 83–99, 2019.
- [27] Y. Rukiah, S. Sarwanto, S. Sutarno, and S. Sunardi, "Visual elements of semar calligraphy on Cirebon glass painting of kusdono's work," *Dewa Ruci Jurnal Pengkajian dan Penciptaan Seni*, vol. 15, no. 2, pp. 71–79, 2020.
- [28] L. Yang, X. Xie, and J. Lai, "Learning discriminative visual elements using part-based convolutional neural network," *Neurocomputing*, vol. 316, no. 17, pp. 135–143, 2018.
- [29] S. Kuchina, "Cohesion specificity IN electronic literary texts," *Semantic, Visual And Graphical Aspects*, 2018.