

Research on Automatic Information Hierarchy Generation Method for Graphic Design Based on Visual Cognition Mechanism

Ruoyu Wang^{1*}, Ziheng Zhao²

¹*Nanchang University, Nanchang, China*

²*Zhejiang Normal University, Jinhua, China*

**Corresponding Author. Email: rara481846778@gmail.com*

Abstract. In the digital communication environment, graphic design performs functions such as information organization, attention guidance, and cognitive load control. Posters, webpage hero sections, mobile interfaces, and infographics all need to form a clear information hierarchy within limited visual space. Traditional layout design mainly depends on designers' experience and is easily constrained by subjective judgment, high revision costs, and limited adaptability across scenarios, making it difficult to ensure that core information is consistently recognized first. To address this issue, visual cognition mechanisms and computer vision methods are introduced into the process of graphic layout generation. Based on saliency detection, structured layout element encoding, hierarchy weight modeling, and multi-objective optimization, an automatic information hierarchy generation method is constructed. The experiment selects 600 poster, webpage hero-section, and infographic samples from public platforms between 2020 and 2025. OCR and object detection are used to extract titles, body text, images, buttons, logos, and decorative elements, while a DeepGaze-type model is applied to generate visual attention heatmaps. The results show that the algorithm-optimized layouts reduce the average core information locating time from 2.84 seconds to 1.96 seconds, increase recognition accuracy from 82.3% to 91.7%, and achieve an expert rating of 4.42. The findings demonstrate that visual saliency and hierarchy optimization algorithms can improve information recognition efficiency and layout cognition consistency, providing methodological support for intelligent assisted design and computational graphic design.

Keywords: Graphic Design, Visual Cognition, Saliency Detection, Information Hierarchy, Automatic Layout Optimization

1. Introduction

Graphic design in digital communication has shifted from static visual presentation to information organization, attention guidance, and cognitive efficiency control. Posters, webpage hero sections, mobile interfaces, and infographics usually need to present titles, images, body text, buttons, and logos within limited visual space. When visual hierarchy is unclear, users may first notice secondary

content, which weakens the recognition of core information. Visual attention studies show that color contrast, spatial position, size difference, and semantic goals jointly influence the first fixation area, providing a cognitive basis for computational information hierarchy [1]. Saliency prediction research further proves that user fixation behavior can be approximately predicted through image features and learning models, making visual focus, reading path, and information weight quantifiable in layout design [2]. Saliency detection, structured element encoding, hierarchy weight modeling, and multi-objective optimization are integrated to construct an automatic information hierarchy generation method for graphic design.

2. Literature review

2.1. Visual cognition mechanism and information hierarchy theory

Visual cognition mechanisms emphasize that users do not process all layout information evenly. Instead, they form selective attention according to visual stimulus intensity, spatial organization, and content meaning. Information hierarchy design uses font size, visual weight, spacing, alignment, and focal arrangement to guide users to identify the main title and key visual first, then move to explanatory information and action cues. Research on visual preference shows that user groups differ in their perception of color, complexity, and spatial layout, so information hierarchy needs to consider both general cognitive patterns and specific design contexts [3]. Interface evaluation research also indicates that visual organization quality is closely related to comprehension efficiency, and clear hierarchy can reduce cognitive load and improve information acquisition stability [4]. Computational information hierarchy therefore needs to consider visual saliency, semantic importance, spatial centrality, and relations among elements.

2.2. Saliency detection and visual attention models

Saliency detection models provide a direct technical basis for attention prediction in graphic design. Early visual attention models mainly generate salient regions based on low-level visual features such as color, brightness, orientation, and edge, explaining why high-contrast, large-area, and high-brightness elements are more likely to be noticed first [5]. With the development of deep learning methods, saliency models have begun to integrate object semantics, scene context, and image structural relations, producing predictions closer to real fixation behavior for titles, people, product images, and buttons in complex layouts [6]. In layout design, saliency heatmaps can judge whether core information is located in high-attention areas and whether decorative elements interfere with main content. Different evaluation metrics focus on fixation prediction, hotspot overlap, and ranking consistency, so model outputs need to be interpreted together with element categories and information priority.

2.3. Automatic layout generation and computational design methods

Automatic layout generation mainly focuses on adjusting the position, size, spacing, and visual weight of design elements through algorithms to improve layout efficiency and design consistency. Existing studies show that element relations, spatial distribution, and visual proportion in single-page graphic design can be transformed into reusable layout rules through sample learning, providing a data basis for automatic generation [7]. Generative layout models further learn structural relations among elements through deep networks, enabling automatic layout for posters, interfaces,

and infographics to move from fixed templates toward flexible combinations [8]. Multi-objective optimization can also incorporate visual balance, hierarchy difference, reading order, and spatial constraints into a unified algorithmic framework. When automatic layout overemphasizes aesthetic scoring or template similarity, it may ignore whether users identify core information in the expected order, so visual cognitive paths and saliency prediction need to be embedded into optimization objectives [9].

3. Experimental methods

3.1. Dataset construction and structured encoding of layout elements

The dataset is constructed around three typical graphic design scenarios, including posters, webpage hero sections, and infographics. The samples are collected from publicly accessible online design platforms between 2020 and 2025, including Behance, Dribbble, Pinterest public design cases, Awwwards webpage hero-section cases, Canva public template resources, and ZCOOL public design works. A total of 600 layout samples are collected, with 200 samples for each category. All samples are cropped or resized to a 1920×1080 standard canvas to ensure consistent input conditions for saliency prediction, spatial feature extraction, and layout quality comparison. During data processing, PaddleOCR is first used to identify title, body text, and button text regions. YOLOv8 is then used to detect images, logos, buttons, and decorative elements. Manual calibration is further conducted to correct overlapping text boxes, inaccurate image-text boundaries, and misidentified decorative components [10]. For each element, coordinates, width, height, area, center point, color contrast, font size, text length, element category, and target hierarchy label are recorded. The hierarchy labels include primary title, secondary information, explanatory text, action button, and background decoration, providing structured input for subsequent saliency weighting and hierarchy optimization.

3.2. Visual saliency calculation and hierarchical weight modeling

After structured element annotation, each standardized layout image is input into a DeepGaze-type saliency prediction model to generate a visual attention heatmap with the same canvas size. Based on the element bounding boxes obtained in Section 3.1, the average saliency, maximum saliency, and saliency proportion are calculated for each title, image, button, body text, and decorative region. These saliency features are then combined with element area, central position, color contrast, font size, and semantic hierarchy to form element-level feature vectors. Hierarchical weight is jointly determined by visual stimulus intensity, layout position, and semantic importance [11]. Therefore, a weighted fusion function is used to calculate the information hierarchy score of each element. The process first normalizes features with different dimensions, then estimates weight parameters according to manually labeled target hierarchy in the training set, and finally outputs the predicted hierarchy value of each element, as shown in Formula (1).

$$H_i = \alpha S_i + \beta C_i + \gamma A_i + \delta T_i + \lambda P_i \quad (1)$$

Where H_i denotes the hierarchy weight of the i -th design element, S_i denotes mean saliency, C_i denotes color contrast, A_i denotes area proportion, T_i denotes textual or semantic importance, P_i denotes element centrality, and $\alpha, \beta, \gamma, \delta, \lambda$ are weight coefficients learned by the model.

3.3. Automatic hierarchy optimization algorithm and experimental process

After hierarchy weight calculation, the automatic optimization process takes the original layout as the initial state. Element position, font size, spacing, and color contrast are treated as adjustable variables, while visual focus concentration, hierarchy differentiation, reading path continuity, and layout balance are set as constraint objectives. The algorithm adopts the NSGA II multi-objective genetic optimization strategy. In each round, candidate layouts are generated and selected according to the fitness function. The fitness function penalizes insufficient saliency of primary information, excessive visual weight of auxiliary elements, excessive reading path jumps, and spatial imbalance. The overall optimization objective is shown in Formula (2).

$$F = \min (w_1 L_{\text{focus}} + w_2 L_{\text{hierarchy}} + w_3 L_{\text{path}} + w_4 L_{\text{balance}}) \quad (2)$$

Where L_{focus} denotes core information focus deviation, $L_{\text{hierarchy}}$ denotes hierarchy differentiation loss, L_{path} denotes reading path deviation loss, L_{balance} denotes layout balance loss, and w_1, w_2, w_3, w_4 are adjustment weights for different objectives.

In the experimental validation stage, three comparison conditions are established, including original human-designed layouts, rule-based optimized layouts, and algorithm-optimized layouts. Sixty participants with normal vision are invited to complete the recognition tasks, including 30 participants with design-related backgrounds and 30 without such backgrounds. Each participant views 30 layouts from each condition and completes core information locating and recognition tasks. First information locating time, recognition accuracy, and visual path deviation are recorded.

4. Results

4.1. Results of information hierarchy recognition efficiency

The results of information hierarchy recognition efficiency show that algorithm-optimized layouts outperform original human-designed layouts and rule-based optimized layouts in core information locating speed, recognition accuracy, and saliency hotspot overlap. The mean locating time of the original layouts is 2.84 seconds, while that of the rule-based optimized layouts decreases to 2.31 seconds. The algorithm-optimized layouts further reduce the time to 1.96 seconds, indicating that primary information becomes easier to find after optimization. Recognition accuracy increases from 82.3% in the original layouts to 87.6% in the rule-based optimized layouts and reaches 91.7% in the algorithm-optimized layouts. Meanwhile, the overlap between saliency hotspots and predefined primary information regions increases from 0.58 to 0.76, while the visual dispersion index decreases from 0.41 to 0.27. This indicates that the interference caused by decorative elements and low-priority text is clearly reduced. The detailed results are shown in Table 1.

Table 1. Information hierarchy recognition efficiency across three layout conditions

Layout condition	Mean locating time/s	Recognition accuracy/%	Saliency overlap	Visual dispersion index
Original human-designed layout	2.84	82.3	0.58	0.41
Rule-based optimized layout	2.31	87.6	0.67	0.34
Algorithm-optimized layout	1.96	91.7	0.76	0.27

4.2. Results of layout quality and visual cognition consistency

The results of layout quality and visual cognition consistency show that algorithm-optimized layouts achieve stable improvements in hierarchy clarity, visual balance, reading path rationality, usability, and saliency consistency. The overall mean score of original human-designed layouts is 3.78, while the rule-based optimized layouts reach 4.05, and the algorithm-optimized layouts reach 4.42. Hierarchy clarity increases from 3.65 to 4.51, and reading path rationality increases from 3.71 to 4.46, indicating that the optimization algorithm improves visual focus strength and information reading order. Visual path data further show that algorithm-optimized layouts are closer to the expected sequence of "title, main image, explanatory text, and action button." The path deviation rate decreases from 31.4% to 18.7%. Figure 1 uses a radar chart to present the differences among the three layout conditions across five evaluation dimensions, clearly showing the overall improvement of the algorithm-optimized layouts in cognitive and design indicators.

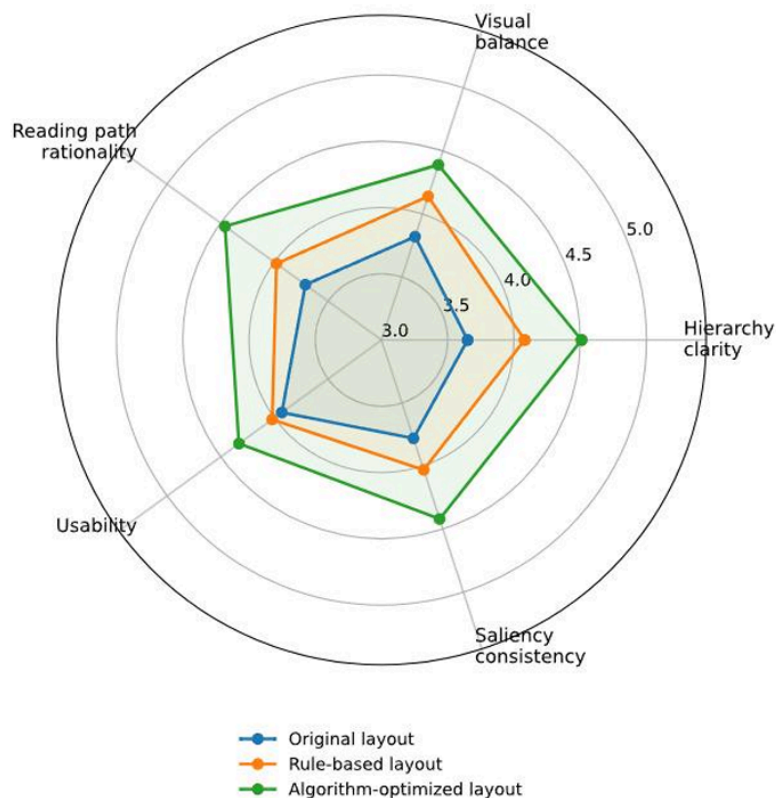


Figure 1. Layout quality and visual cognition consistency radar chart

5. Discussion

The experimental results show that visual saliency models can effectively participate in information hierarchy generation for graphic design, transforming experience-based layout adjustment into a computable, comparable, and optimizable process. Algorithm-optimized layouts outperform original human-designed layouts and rule-based optimized layouts in locating time, recognition accuracy, saliency overlap, and expert evaluation, indicating a strong coordination among saliency prediction, semantic weight, and spatial layout. The improvement in hierarchy clarity and reading path rationality further shows that optimization objectives should not focus only on whether visual focus is prominent, but also on whether users understand titles, main images, explanatory text, and action

buttons in the expected order. This method is suitable for rapid-recognition scenarios such as advertising posters, e-commerce hero sections, public infographics, and mobile communication interfaces. However, differences in public sample styles, semantic recognition errors, and simulated eye-tracking data may still affect model generalization. Future research can integrate real eye-tracking experiments, multimodal large models, and dynamic interaction data to improve adaptability in complex design scenarios.

6. Conclusion

Focusing on automatic information hierarchy generation in graphic design, a computational method framework integrating visual cognition mechanisms, saliency detection, and multi-objective optimization is constructed. The experiment collects 600 poster, webpage hero-section, and infographic samples from public design platforms. OCR, object detection, and manual calibration are used to complete structured encoding of layout elements. A DeepGaze-type model is then used to generate attention heatmaps, while element area, position, color contrast, font size, and semantic importance are combined to calculate hierarchy weights. The NSGA II multi-objective optimization algorithm further adjusts element position, scale, spacing, and visual intensity, enabling core information to obtain a more stable attention priority. The results show that algorithm-optimized layouts shorten core information locating time, improve recognition accuracy, enhance the overlap between saliency hotspots and primary information areas, and improve expert evaluations of hierarchy clarity and reading path rationality. This method provides an operational path for transforming graphic design from experience-driven practice to data-driven computation, and offers technical support for intelligent design systems, automatic poster generation, and visual interface optimization.

Author contribution

Ruoyu Wang and Ziheng Zhao contributed equally to this paper.

References

- [1] Chakraborty, S., et al. (2022). Predicting visual attention in graphic design documents. *IEEE Transactions on Multimedia*, 25, 4478–4493.
- [2] Hassanin, M., et al. (2024). Visual attention methods in deep learning: An in-depth survey. *Information Fusion*, 108, 102417.
- [3] Zhang, Y., et al. (2024). Gazefusion: Saliency-guided image generation. *ACM Transactions on Applied Perception*, 21(4), 1–19.
- [4] González, P. V., et al. (2025). A gender-aware saliency prediction system for web interfaces using deep learning and eye-tracking data. *Brain Informatics*, 12(1), 25.
- [5] Yang, M., et al. (2025). Dashboard vision: Using eye-tracking to understand and predict dashboard viewing behaviors. *IEEE Transactions on Visualization and Computer Graphics*, 31(10), 6930–6945.
- [6] Srinivasan, A., et al. (2024). Attention-aware visualization: Tracking and responding to user perception over time. *IEEE Transactions on Visualization and Computer Graphics*, 31(1), 1017–1027.
- [7] Jianu, R., et al. (2025). Gaze-aware visualisation: Design considerations and research agenda. *Computer Graphics Forum*, 44(6).
- [8] Abraham, S. E., & Kooor, B. C. (2023). Visual saliency modeling with deep learning: A comprehensive review. *Journal of Information & Knowledge Management*, 22(2), 2250066.
- [9] Cheng, S., et al. (2023). Poster graphic design with your eyes: An approach to automatic textual layout design based on visual perception. *Displays*, 79, 102458.
- [10] Xu, C., et al. (2025). Dual discriminator GANs with multi-focus label matching for image-aware layout generation. *Displays*, 87, 102970.

- [11] Zhang, N., et al. (2024). Moderating effects of visual order in graphical symbol complexity: The practical implications for design. *Applied Sciences*, 14(17), 7592.