ingegneria sismica
International Journal of Earthquake Engineering
Anno XLIII - Num. 2 - 2026

Differentiable Layout Optimization: A Visual Fidelity Enhancement Model Supporting Multi-dimensional graphical element reordering

Jiping Liu[1,*], *Xinhua Li*[1], *Hongbin You*[1], *Zheng Wei*[1] *and Xia Zhao*[1]

[1] State Grid Gansu Electric Power Company Baiyin Power Supply Company, Baiyin 730299, Gansu, China

SUMMARY: *A visual fidelity enhancement model supported by differentiable layout optimization is proposed to solve the problems of element occlusion, structural deviation, semantic break and visual fidelity degradation in multi-dimensional graphic element rearrangement. In the model, elements such as text, images, ICONS, buttons and charts are transformed into a unified parametric representation, and the spatial adjacency, alignment, hierarchy and semantic proximity relationships are depicted by relational graphs. On this basis, a joint loss function of structure preservation, occlusion penalty, boundary constraint, visual balance and semantic consistency is constructed to realize the continuous derivable update of layout parameters. Experimental results show that the structural similarity of the proposed model reaches 0.914, the occlusion rate is reduced to 2.3%, the semantic preservation rate is 94.2%, and the average optimization time is 0.42 s, which is better than the regular layout, force-directed layout, genetic algorithm and deep generative layout models. The results show that the model can improve the structural stability, visual clarity and semantic continuity of complex graphics after reordering, and provide technical support for graphics editing, interface reconstruction and automatic typesetting.*

KEYWORDS*: differentiable layout optimization; Multidimensional graphical element; Visual fidelity; Reordering elements*

1 Introduction

With the development of information visualization, intelligent poster design, interface generation and multi-modal document editing, graphic layout has shifted from simple element placement to structural coordination and visual fidelity optimization of multi-dimensional graphic elements. Text, images, ICONS, buttons, color blocks, and decorative elements not only have geometric and visual attributes such as position, size, hierarchy, and color, but also carry reading order, semantic proximity, and visual focus relationships. When the scale of canvas changes, content modules are added or deleted, or elements are adjusted, the traditional regular layout and heuristic search methods are prone to problems such as element occlusion, boundary violation, visual center of gravity shift, and semantic relationship fracture, which affect the readability and overall consistency of the rearrangement results.

In recent years, generative layout models provide a new technical foundation for graphical element reordering. Cheng et al. (2023) proposed the Play model, which uses parametric conditions and latent diffusion mechanism to realize layout generation control, and provides a reference for the continuous expression of graphical elements [1]. Inoue et al. (2023)

*Liujiping291948832@163.com
https://doi.org/10.65102/is2026915
Published: 30 April, 2026

proposed LayoutDM to study the application of discrete diffusion models in controllable layout generation, and proved that the diffusion process can learn the spatial combination law between elements [2]. Jiang et al. (2023) proposed Layoutform ++ to improve the quality of conditional graph layout generation through constraint serialization and decoding space constraints, so that category, location, and relationship constraints can participate in layout reasoning [3]. Zhang et al. (2023) proposed LayoutDiffusion to study the improvement effect of discrete diffusion probability model on the quality of graphic layout generation, which provided a method support for layout disturbance recovery and element rearrangement [5]. Li et al. (2023) proposed a relation-aware diffusion model, emphasizing the important influence of semantic and spatial relationships between poster elements on controllable layout [7]. Chen et al. (2024) studied the alignment layout generation method with aesthetic constraints, and introduced visual alignment and aesthetic constraints into the diffusion model, which provided a basis for visual fidelity loss design [9]. Shen et al. (2025) proposed LayoutRectifier, which uses an optimized post-processing method to correct the layout generation results of graphic design, indicating that further structural correction and local optimization are still needed after layout generation [19]. Sun et al. (2025) proposed LayoutVLM to study the differentiable optimization of 3D layout supported by visual language model, and showed that the differentiable mechanism can effectively coordinate spatial constraints and visual goals [20].

In general, existing researches have made progress in automatic layout generation, conditional control and relationship modeling, but still pay insufficient attention to continuously differentiable representation, visual fidelity enhancement and local conflict correction in multi-dimensional graphical element rearrangement. Based on this, this paper intends to transform element position, scale, hierarchy and semantic relationships into derivable variables, and construct a joint loss function of structure preservation, occlusion suppression, boundary constraint, visual balance and semantic consistency, and improve the visual stability after graphics rearrangement through gradient optimization. A comparison of related research and concerns in this paper is shown in Table 1.

Table 1: Comparison of research concerns between related studies and this paper

Reference	Scholar(s) and Year	Main Method	Implications for This Paper
[1]	Cheng et al. (2023)	Latent diffusion layout generation	Supports parametric layout representation
[3]	Jiang et al. (2023)	Constraint serialization	Supports layout constraint modeling
[7]	Li et al. (2023)	Relation-aware diffusion	Supports semantic relationship preservation
[9]	Chen et al. (2024)	Aesthetic-constrained diffusion	Supports visual fidelity loss design
[19]	Shen et al. (2025)	Optimization-based post-processing	Supports layout reordering correction
[20]	Sun et al. (2025)	Differentiable layout optimization	Supports differentiable parameter update mechanism

2 Modeling Multidimensional graphical element Rearrangement Problem and Visual Fidelity Constraints

2.1 Parametric representation of multidimensional graphical elements

The key of parametric representation of multi-dimensional graphic elements is to transform the heterogeneous objects in visual layout into a unified vector that can be calculated, constrained and optimized. Suppose there are N graphic elements in the layout, and the set of elements is denoted as $E = \{e_i\}_{i=1}^{N}$. The ith element can be expressed as follows.

$$e_i = [x_i, y_i, w_i, h_i, c_i, s_i, z_i, \alpha_i, t_i, r_i] \quad (1)$$

where, x_i, y_i represent the center coordinates of elements, w_i, h_i represent width and height, c_i represent color features, s_i represent shape coding, z_i represent visual hierarchy, α_i represent transparency, t_i represent semantic labels, and r_i represent interactions with other elements. When Chai et al. (2023) studied the transformer-based diffusion layout generation method, they regarded the layout object as a serializable element representation, indicating that the unified encoding of element attributes is the basis for complex layout generation [4]. Levi et al. (2023) proposed a discrete-continuous joint diffusion layout Transformer to jointly model variables such as category, coordinate and size, which provides a reference for the unified integration of geometric and semantic variables into element vectors in this paper [6]. In the concrete modeling, the continuous attributes are normalized, let $x_i, y_i, w_i, h_i, \alpha_i \in [0,1]$, and the discrete attributes are transformed into dense vectors by embedding mapping, namely:

$$v_i = \phi_g(g_i) + \phi_s(t_i) + \phi_v(c_i, z_i, \alpha_i) \quad (2)$$

Here, ϕ_g, ϕ_s, ϕ_v represent geometric, semantic, and visual attribute encoding functions, respectively. This representation can avoid the separation of different types of elements in scale and semantic space, so that text, images, ICONS, controls and decorative elements can all enter the same optimization framework, and provide a unified input for subsequent differentiable layout updates, visual fidelity constraints and element rearrangement output.

2.2 Modeling spatial and semantic relations of graphic elements

Graphic element rearrangement is not the position update of a single object, but the collaborative adjustment of spatial order and semantic dependence of multiple elements. Figure 1 shows the modeling of spatial and semantic relationships of graphic elements.

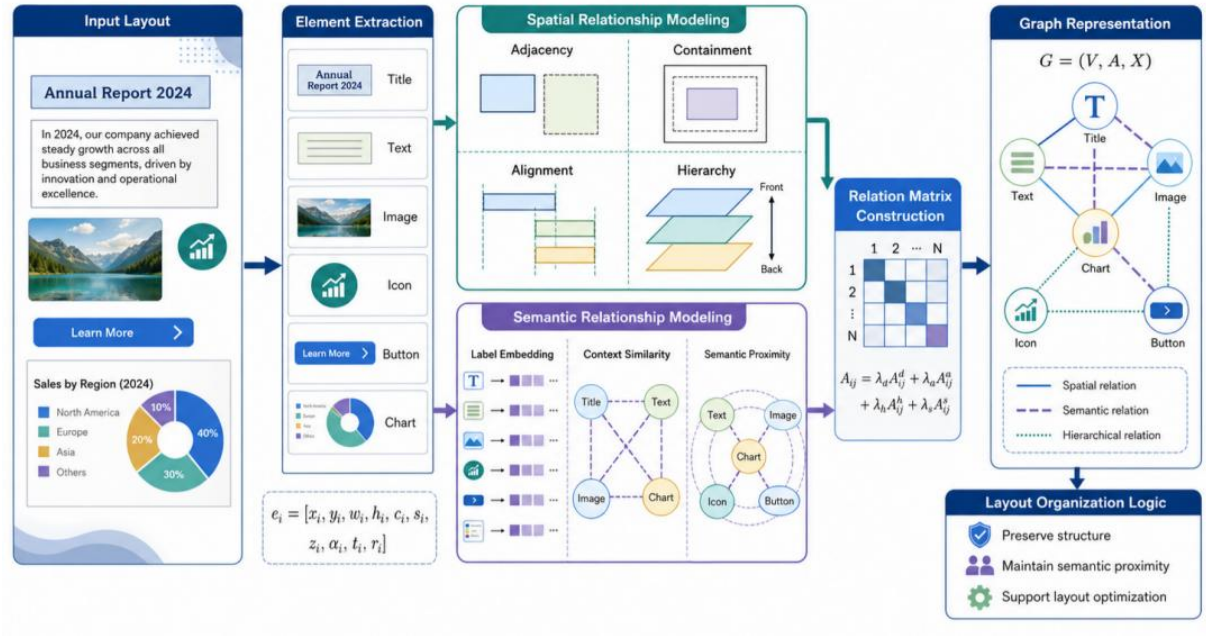

Figure 1: Modeling spatial and semantic relationships of graphical elements

In order to describe the relationship between elements, the layout is represented as A graph structure G=(V,A,X), where V is the node set of elements,X is the attribute matrix of elements, and A is the adjacency matrix of relations. The adjacency matrix should not only record geometric distances, but also contain inclusion relations, alignment relations, hierarchical relations, and semantic proximity relations, which can be defined as follows.

$$A_{ij} = \lambda_d A^d_{ij} + \lambda_a A^a_{ij} + \lambda_h A^h_{ij} + \lambda_s A^s_{ij} \tag{3}$$

where, A^d_{ij} represents distance adjacency, A^a_{ij} represents alignment relationship, A^h_{ij} represents hierarchical inclusion relationship, A^s_{ij} represents semantic correlation, and λ is the weight coefficient. When studying the flexible multimodal document model, Inoue et al. (2023) emphasize that the text, image and layout relationships between document elements need to be jointly represented to improve the ability to understand and generate complex layout [8]. Horita et al. (2024) proposed a retrieval enhanced layout Transformer, which uses content relationships in similar layouts to assist generation, indicating that content context has a significant impact on element location organization [10]. In the proposed model, spatial relationships can be calculated from element bounding boxes, as follows.

$$d_{ij} = \frac{\sqrt{(x_i - x_j)^2 + (y_i - y_j)^2}}{\sqrt{W^2 + H^2}} \tag{4}$$

An edge connection is established when d_{ij} is less than a threshold or two elements have semantic dependency. The semantic relationship is calculated by the tag embedding similarity, $A^s_{ij} = \cos(q_i, q_j)$. Through the coupling modeling of spatial relationship and semantic relationship, the model can maintain the reasonable proximity between the title and the main text, the image and the caption, the button and the functional area, and the legend and the chart during the rearrangement process, and avoid the semantic break and visual organization imbalance caused by simple geometric optimization.

2.3 Visual fidelity Enhancement Goal Definition

The visual fidelity enhancement objective is used to measure whether the structural order, reading path, visual emphasis and semantic consistency of the original layout are still maintained after the reordering of multi-dimensional graphic elements. Unlike single layout similarity, visual fidelity is defined in this paper as the combined result of structure preservation, element readability, occlusion suppression, visual balance, and semantic consistency. Shabani et al. (2024) proposed the image-vector double diffusion model, which considered visual salient regions and vector layout together, indicating that the quality of graphic layout depends not only on element coordinates, but also on background images, salient regions and visual focus [11]. Seol et al. (2024) proposed PosterLlama to introduce the design capabilities of language models into content-aware layout generation, showing that semantic content has a constraint effect on visual element organization [12]. Zhang et al. (2024) proposed Vascar to improve the quality of content-aware layout through visual perception self-correction, which provides a direct reference for constructing the fidelity enhancement goal in this paper [15]. In the model, the visual fidelity goal can be expressed as follows.

$$F = \omega_1 S_{str} + \omega_2 R_{read} + \omega_3 (1 - O_{occ}) + \omega_4 B_{vis} + \omega_5 S_{sem} \tag{5}$$

Here, S_{str} represents structure preservation, R_{read} represents element readability, O_{occ} represents occlusion rate, B_{vis} represents visual balance, and S_{sem} represents semantic consistency. Readability is calculated by text region size, contrast and occlusion state. Visual balance is evaluated by element area, visual weight and offset of canvas center. Through this goal definition, the reordering task is not only about compressing or moving elements, but also about multi-objective constraints around visual perception quality, so that the model can maintain the overall recognition and design continuity of the graphic layout while meeting the boundary and non-overlapping requirements.

2.4 Formal description of layout reordering problem with multiple constraints

Multi-constraint layout reordering needs to unify element geometry updating, visual fidelity preservation, and constraint conflict resolution into a continuously optimizable framework. Let the original layout be $L^0 = \{e_i^0\}_{i=1}^N$, the rearranged layout be $L' = \{e_i'\}_{i=1}^N$, and the optimization variables be $\Theta = \{x_i, y_i, w_i, h_i, z_i\}_{i=1}^N$. The goal is to obtain a rearrangement result with minimum visual loss under the constraints of canvas boundaries, non-overlapping elements, hierarchical order, semantic proximity and visual balance, namely:

$$\Theta^* = \arg\min_{\Theta} \mathcal{L}_{total} \tag{6}$$

$$\mathcal{L}_{total} = \beta_1 \mathcal{L}_{str} + \beta_2 \mathcal{L}_{occ} + \beta_3 \mathcal{L}_{bd} + \beta_4 \mathcal{L}_{bal} + \beta_5 \mathcal{L}_{sem} \tag{7}$$

Among them, $\mathcal{L}_{str}$ is used to preserve the original structural relations, $\mathcal{L}_{occ}$ is used to penalize element overlap, $\mathcal{L}_{bd}$ is used to limit boundary overbounds, $\mathcal{L}_{bal}$ is used to maintain visual center of gravity stability, and $\mathcal{L}_{sem}$ is used to preserve semantic proximity relations. Guerreiro et al. (2024) proposed LayoutFlow, which uses flow matching method to realize smooth state transformation during layout generation, indicating that the continuous migration of layout variables helps to reduce the structural disturbance caused by abrupt rearrangement [13]. Iwai et al. (2024) proposed laylay-Corrector to correct fixations and

incongruity in the diffusion Layout, indicating that an optimization mechanism still needs to be introduced to deal with local conflicts after layout generation [14]. Cheng et al. (2025) studied graphic design tasks supported by most modal models, emphasizing that complex design generation requires understanding both visual content and semantic intent [16]. Kikuchi et al. (2025) proposed a multi-modal marked document model for graphic design completion, which further illustrates the importance of structured document representation for layout optimization [17]. Lin et al. (2025) proposed a hierarchical graphic design combination method, which provided ideas for the joint optimization of element level, structure level and design level [18]. Therefore, this paper models the rearrangement problem as a multi-constrained optimization process driven by differentiable loss, which provides a mathematical foundation for subsequent visual fidelity enhancement models.

3 Visual Fidelity Enhanced Model Design supported by differentiable Layout Optimization

3.1 Multi-dimensional element feature Encoding and Relation graph construction

Multi-dimensional element feature coding establishes a unified input layer for heterogeneous graphical objects, and converts text, images, ICONS, buttons, charts and decorative elements into node features that can participate in gradient calculation. Firstly, the element bounding box is normalized, and the position and size are mapped to the relative coordinates at the canvas scale. At the same time, the attributes such as color histogram, saliency intensity, transparency, shape category, hierarchical index and semantic label are extracted. Continuous variables are normalized by Min-Max, discrete variables are converted into dense vectors by Embedding layer, concatenated with visual features and input into MLP encoder to obtain the initial node representation of elements:

$$h_i^0 = \mathrm{MLP}([g_i \| v_i \| p_i^{emb} \| t_i^{emb}]) \tag{8}$$

Here, g_i represents normalized geometric properties, v_i represents color, texture and saliency features, p_i^{emb} represents shape and hierarchical embeddings, and t_i^{emb} represents semantic label embeddings. This encoding method can map different types of elements to a unified dimension, avoiding gradient update offset caused by inconsistent feature scales in subsequent optimization.

In the construction phase of the relation graph, elements are taken as nodes, and spatial and semantic relations are taken as edges to form G=(V,A,H). The spatial edges are calculated from the center distance, the intersection over union ratio of bounding boxes, and the horizontal or vertical alignment error, while the semantic edges are determined from the label embedding cosine similarity and functional dependencies. Specifically, when the center distance of two elements is less than the set threshold, the IoU is greater than the overlap threshold, or the boundary difference of two elements meets the alignment condition, a higher weight is given in the relationship matrix. If there is a semantic dependency between the title and the main text, the image and the description, the chart and the legend, and the button and the functional area, the semantic edge weight will be enhanced. The resulting relation matrix not only describes whether elements are adjacent, but also describes "why adjacent" and "should remain adjacent", so that layout optimization is no longer limited to geometric collision avoidance.

In the high-order relationship fusion stage, the model uses graph neural network or Transformer with relationship bias to update node features. Graph neural network focuses on local neighborhood propagation, which is suitable for maintaining the internal structure of the element group. Transformer focuses on global dependency modeling and is suitable for capturing cross-region visual balance and reading path relationships. The updated node representation is denoted $Z = \{z_i\}_{i=1}^{N}$, which simultaneously contains the element's own attributes, local spatial constraints, and global semantic context. This representation is then fed into a differentiable layout optimization module that predicts element displacement, scale adjustment, and hierarchy correction parameters, thus providing a structured feature basis for visual fidelity enhancement.

3.2 Differentiable layout parameter update mechanism

The core of the differentiable layout parameter update mechanism is to transform the graphic element rearrangement process from discrete rule adjustment to a gradient optimization process driven by continuous variables. Let the optimized layout parameters of the ith element be $\theta_i = [x_i, y_i, w_i, h_i, z_i]$, corresponding to the center coordinate, width and height scale and hierarchical order, respectively. The layout parameters of all elements consist of $\Theta = \{\theta_i\}_{i=1}^{N}$. In the parameter initialization stage, based on the original layout L^0, the element coordinates and dimensions are normalized to the canvas coordinate system, and the initial relationship matrix A between elements is retained, so that the subsequent updates can be carried out in a continuous space instead of relying on manually set movement rules. To avoid abrupt changes in scale, the position and size updates are expressed as residuals, meaning that the model predicts layout adjustments instead of generating final coordinates:

$$\theta_i' = \theta_i^0 + \Delta\theta_i, \quad \Delta\theta_i = f_\psi(z_i, A) \tag{9}$$

Here, θ_i^0 represents the initial layout parameters of elements, $\Delta\theta_i$ represents the rearrangement offset to be learned, z_i is the high-dimensional node features obtained in the previous section, and $f_\psi(\cdot)$ is the parameter update network. This approach preserves the original layout structure and reduces the risk of visual distortion caused by large movements.

In the optimization process, position, scale, spacing and level are all used as derivable variables to participate in loss back propagation. For the boundary constraint, the soft penalty method is used to deal with the element out of the boundary problem to avoid hard clipping damaging the gradient propagation. For the occlusion constraint, the penalty term is calculated according to the overlapping area of the bounding boxes of the two elements, so that the overlapping area is automatically reduced in the model update process. The spacing constraint is used to keep the distance within the element group stable and prevent the title and body text, image and caption, and chart and legend from being too far apart after reordering. The hierarchical parameter zi approximates the occlusion relationship by continuous values, and is then mapped to a discrete hierarchical order in the final output stage, so as to meet the needs of differentiable optimization and layout rendering. The combined loss function can be written as follows.

$$\mathcal{L}_{\text{upd}} = \mu_1\mathcal{L}_{\text{move}} + \mu_2\mathcal{L}_{\text{scale}} + \mu_3\mathcal{L}_{\text{overlap}} + \mu_4\mathcal{L}_{\text{bound}} + \mu_5\mathcal{L}_{\text{rank}} \tag{10}$$

Among them, $\mathcal{L}_{\text{move}}$ controls the amplitude of element movement, $\mathcal{L}_{\text{scale}}$ controls the range of scale change, $\mathcal{L}_{\text{overlap}}$ suppresses element occlusion, $\mathcal{L}_{\text{bound}}$ constrains elements to be within the scope of the canvas, and $\mathcal{L}_{\text{rank}}$ maintains the stable hierarchical order. The weights $\mu_1 - \mu_5$ of each item can be adaptively adjusted according to the layout type and the

rearrangement goal.

In the parameter iteration stage, AdamW optimizer is used to update the layout variables and network parameters to improve the convergence stability under complex constraints. After each iteration, the system recalculates the element bounding box, adjacency distance, and occlusion state based on the current layout, and feeds the feedback results into the next update round. The parameter update process is expressed as follows.

$$\Theta^{k+1} = \Theta^{k} - \eta \frac{\hat{m}_k}{\sqrt{\hat{v}_k} + \varepsilon} - \eta\lambda\Theta^{k} \tag{11}$$

where η is the learning rate, $\hat{m}_k$ and $\hat{v}_k$ are the first and second moment estimates, λ is the weight decay coefficient, and ε is used to ensure numerical stability. In order to prevent gradient oscillation, the gradient clipping and learning rate attenuation strategies are introduced in the training, and the iteration is stopped when the decrease of visual fidelity loss is less than the threshold for consecutive rounds. Through the above mechanism, the model can gradually complete position correction, scale adjustment, occlusion resolution and hierarchy rearrangement on the basis of maintaining the semantic relationship of elements and the overall visual structure, which provides a guided and stable parameter update path for subsequent visual fidelity joint loss optimization.

3.3 Joint Loss Function Design for Visual Fidelity

The visual fidelity joint loss function is used to constrain the reordering of multi-dimensional graphic elements to maintain the original structural relationship, visual order and semantic organization logic. Different from layout optimization that solely aims at element collision avoidance or boundary limitation, we define visual fidelity as the comprehensive result of "structural stability, occlusion controllability, boundary compliance, barycenter balance, and semantic continuity", and transform it into multiple losses that can participate in backpropagation. Figure 2 shows the visual fidelity joint loss function design.

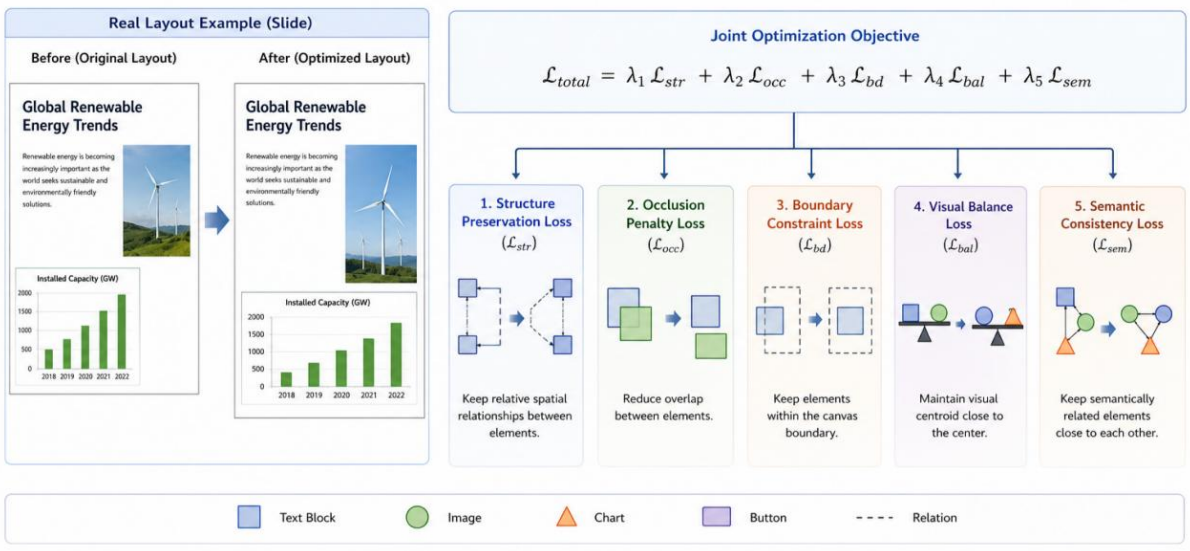

Figure 2: Joint loss function design for visual fidelity

Let the original layout be L^0 and the optimized layout be L' and the set of elements $E = \{e_i\}_{i=1}^{N}$, then the joint loss function is defined as follows.

$$\mathcal{L}_{vf} = \lambda_1 \mathcal{L}_{str} + \lambda_2 \mathcal{L}_{occ} + \lambda_3 \mathcal{L}_{bd} + \lambda_4 \mathcal{L}_{bal} + \lambda_5 \mathcal{L}_{sem} \tag{12}$$

Here $\mathcal{L}_{str}, \mathcal{L}_{occ}, \mathcal{L}_{bd}, \mathcal{L}_{bal}, \mathcal{L}_{sem}$ represent the structure preservation loss, occlusion penalty loss, boundary constraint loss, visual balance loss, and semantic consistency loss, respectively, and $\lambda_1 - \lambda_5$ are the weight coefficients used to control the influence strength of different visual fidelity objectives in the optimization process.

The loss of structure preservation mainly restricts the relative spatial relationship between elements before and after rearrangement, and avoids the breakage of organizational relationship after the movement of elements such as titles, text, images, legends and buttons. The loss can be constructed based on the distance between the center of the elements and the direction relationship, and the constraint is implemented by comparing the difference in the relative relationship between the original layout and the rearranged layout:

$$\mathcal{L}_{str} = \frac{1}{N^2} \sum_{i=1}^{N} \sum_{j=1}^{N} A_{ij} \left\| \frac{p_i' - p_j'}{\|p_i' - p_j'\|_2 + \varepsilon} - \frac{p_i^0 - p_j^0}{\|p_i^0 - p_j^0\|_2 + \varepsilon} \right\|_2^2 \tag{13}$$

where $p_i = (x_i, y_i)$ represents the element center point, A_{ij} represents the element relationship weight, and ε is a numerical stable term. This item can reduce the layout structure drift caused by excessive movement of local elements in the rearrangement process.

The occlusion penalty loss is used to control the overlap region between elements. For any two element bounding boxes B_i' and B_j', if their intersection area is large, it indicates occlusion or decreased readability after rearrangement. The occlusion loss is defined as follows.

$$\mathcal{L}_{occ} = \frac{1}{N(N-1)} \sum_{i \neq j} \frac{|B_i' \cap B_j'|}{|B_i' \cup B_j'| + \varepsilon} \tag{14}$$

This term essentially penalizes the IoU of the rearranged elements and guides the model to reduce overlap between text and images, buttons and charts, and decorative graphics and main content, improving layout clarity.

The bounding constraint loss is used to ensure that the element does not go out of the canvas. Let the canvas width and height be W,H, and the element boundaries are x_i^r, y_i^t, y_i^b, respectively. Then the boundary loss is defined as follows through the soft penalty function.

$$\mathcal{L}_{bd} = \frac{1}{N} \sum_{i=1}^{N} [\,\mathrm{ReLU}(-x_i^l) + \mathrm{ReLU}(x_i^r - W) + \mathrm{ReLU}(-y_i^t) + \mathrm{ReLU}(y_i^b - H)] \tag{15}$$

Compared with hard cropping, this method can preserve the gradient propagation, and make the out-of-bounds elements gradually return to the valid canvas area during the optimization process without abrupt position correction.

The visual balance loss is used to control the shift of the visual center of gravity of the rearranged layout. Considering the different contributions of different elements to visual attention, we calculate the visual centroid based on the area, saliency and hierarchical weight of the element and constrain it to be close to the center of the canvas or the original visual center of gravity:

$$\mathcal{L}_{\text{bal}} = \left\| \frac{\sum_{i=1}^{N} \rho_i \, a_i' p_i'}{\sum_{i=1}^{N} \rho_i \, a_i' + \varepsilon} - c_0 \right\|_2^2 \tag{16}$$

where a_i' represents the element area, ρ_i represents the element visual weight, and c_0 is the original layout visual center of gravity or the canvas center point. This item can prevent the rearrangement of elements too concentrated on one side, and improve the overall coordination of the layout.

Semantic consistency loss is used to maintain reasonable proximity of elements with functional or semantic dependencies after rearrangement. For example, there should be no excessive separation between headings and body text, images and captions, charts and legends, buttons and action areas. Let S_{ij} be the semantic relevance matrix and d_{ij}' be the distance of elements after rearrangement, then the semantic consistency loss can be expressed as follows.

$$\mathcal{L}_{\text{sem}} = \frac{1}{N^2} \sum_{i=1}^{N} \sum_{j=1}^{N} S_{ij} \cdot d_{ij}' \tag{17}$$

This term makes the elements with higher semantic relevance more likely to remain close to each other during the optimization process, thus reducing the semantic break caused by visual rearrangement. Through the joint constraints of the above five types of losses, the model can simultaneously take into account structural stability, visual clarity, boundary legality, picture balance and semantic continuity when performing position adjustment, scale correction and hierarchical rearrangement, providing a differentiable, interpretable and visual fidelity oriented optimization goal for multi-dimensional graphic element rearrangement.

3.4 Model Training and Convergence Control Strategy

In the model training phase, the element node features, relationship matrix and original layout parameters are taken as input, the element position, scale and level offset are predicted by updating the network, and the visual fidelity joint loss is used as the main optimization objective. Since the multi-dimensional graphic element rearrangement is affected by structure preservation, occlusion resolution, boundary restriction and semantic proximity constraints at the same time, the training process is prone to gradient oscillation or strong local constraints. Therefore, this paper adopts a joint control strategy of "learning rate decay-gradient culture-weight adaption-regularization-early stopping" to improve the convergence stability of the model in complex layout scenarios.

In terms of learning rate control, cosine decay strategy is used to maintain strong search ability in the early stage of training, and gradually enter the fine optimization stage in the later stage:

$$\eta_t = \eta_{\min} + \frac{1}{2}(\eta_0 - \eta_{\min}) \left(1 + \cos \frac{\pi t}{T}\right) \tag{18}$$

Here, η_0 is the initial learning rate, $\eta_{\min}$ is the minimum learning rate, t is the current round, and T is the total training round. This method can avoid the later oscillation caused by the fixed learning rate, and make the element coordinate and scale change gradually become stable.

In terms of gradient control, the norm clipping mechanism is adopted for large gradients that may be generated by samples with severe occlusion or boundary crossing:

$$\tilde{g}_t = g_t \cdot \min\left(1, \frac{\tau}{\|g_t\|_2}\right) \tag{19}$$

Here, gt is the current gradient and τ is the clipping threshold. This mechanism can prevent local abnormal elements from dominating the overall update and ensure that the layout parameters are iterated within a controllable range.

Due to the different numerical scale and convergence speed of each loss term, this paper further introduces an adaptive adjustment mechanism of loss weight, which dynamically assigns weights according to the relative loss level of different targets:

$$\lambda_m^{t+1} = \frac{\exp(\mathcal{L}_m^t / \bar{\mathcal{L}}^t)}{\sum_{k=1}^{M} \exp(\mathcal{L}_k^t / \bar{\mathcal{L}}^t)} \tag{20}$$

Here, $\mathcal{L}_m^t$ represents the MTH term loss, $\bar{\mathcal{L}}^t$ is the average loss, and m is the number of loss terms. This strategy can avoid the model only optimizing the occlusion or boundary single target, and make the structure preservation, visual balance and semantic consistency converge synchronously.

At the same time, a regularization term is introduced to limit the model from excessively moving elements or memorizing training samples:

$$\mathcal{R} = \gamma_1 \|\Delta\Theta\|_2^2 + \gamma_2 \|\psi\|_2^2 \tag{21}$$

Here, ΔΘ is the layout offset and ψ is the model parameter. The validation set loss was monitored during training, and early stopping was triggered when the decrease amplitude was below the threshold for several consecutive rounds. Through the above strategy, the model can reduce the risk of shock while keeping the gradient derivable, so that the rearrangement results are stable in visual structure, element spacing and semantic relationship.

4 Experimental verification and result analysis

4.1 Dataset construction and experimental setup

In order to verify the effectiveness of the model in multi-dimensional graphic element rearrangement, this paper constructs a comprehensive layout data set, and the samples cover five types of scenes: information diagram, interface component diagram, mixed layout diagram, visual diagram and multi-level design diagram. For each sample, objects such as text, images, ICONS, buttons, charts, backgrounds, and decorative elements are extracted and annotated with element bounding boxes, category labels, color features, hierarchical order, transparency, semantic relationships, and spatial relationships. In the data preprocessing stage, the element coordinates and sizes are normalized to the interval [0,1], the discrete categories are encoded by embedding, and the continuous visual attributes are standardized. The relationship matrix is constructed according to the center distance, boundary overlap, alignment relationship, inclusion relationship and semantic proximity relationship. The experiment was divided into training set, validation set and test set according to 7:2:1, which were used for model parameter learning, hyperparameter adjustment and generalization ability testing respectively. In order to close to the real rearrangement task, five experimental situations are set up, including canvas scale change, element scale disturbance, partial occlusion correction, module position exchange, and content addition and deletion. The model is implemented based on PyTorch, using AdamW optimizer, initial learning rate is 2×10^{-4},

batch size is 32, maximum training rounds is 150, and gradient clipping, cosine learning rate decay and early stopping mechanism are combined to control convergence. The data set composition and experimental parameters are shown in Table 2.

Table 2: Data set composition and experimental parameter Settings

Item	Setting
Sample type	Infographic, interface component diagram, image-text mixed layout, visualization chart, multi-level design diagram
Sample size	5,000 layout samples, approximately 85,000 graphical elements
Element category	Text, image, icon, button, chart, background, decorative element
Annotation content	Bounding box, category, color, hierarchy, transparency, spatial relationship, semantic relationship
Data split	Training set 70%, validation set 20%, test set 10%
Reordering scenario	Canvas change, scale perturbation, occlusion correction, position exchange, content addition and deletion
Experimental environment	Python 3.10, PyTorch 2.1, NVIDIA RTX 3090
Training parameters	AdamW, learning rate 2×10^{-4}, batch size 32, epoch 150

4.2 Evaluation index and comparison scheme

In order to comprehensively evaluate the visual fidelity enhancement effect of the model in multi-dimensional graphic element rearrangement, we set up evaluation indicators from five levels: structural consistency, occlusion control, spatial regularity, semantic preservation and computational efficiency. Structural similarity is used to measure the degree of maintaining the relative spatial relationship between elements before and after rearrangement, which can be calculated according to the difference of element relationship matrix. The occlusion rate is used to calculate the overlap ratio between the bounding boxes of the rearranged elements, with lower values indicating better readability. Alignment error is used to measure the deviation of edge, center line and reference line of an element. Visual balance is used to evaluate the distribution deviation between the element area, saliency weight and the center of the canvas. The semantic preservation rate is used to determine whether the semantic related elements such as caption-text, image-caption, and chart-legend are still reasonably close to each other. Layout stability is measured by element displacement and scale change before and after rearrangement, and the average optimization time is used to reflect the responsiveness of the model in the actual editing scene. The core evaluation index can be expressed as follows.

$$OR = \frac{\sum_{i\neq j} |B_i \cap B_j|}{\sum_{i\neq j} |B_i \cup B_j| + \varepsilon} \tag{22}$$

where OR is the occlusion rate, Bi and Bj represent the element bounding box, and the smaller the value is, the clearer the rearrangement result is.

The comparison scheme selects the regular layout method, the force-directed layout method, the genetic algorithm layout optimization method and the deep generative layout model as the baseline. Regular layout method relies on preset grid, margin and alignment rules, which is suitable for simple scenes but has weak adaptability to complex element relationships. The force-directed method adjusts the position of elements through repulsion and attraction, which can alleviate overlap but easily destroy semantic relationships. Genetic algorithm obtains the optimal layout through population search, but the computational cost is

large. The deep generative layout model has strong global generation ability, but it is insufficient to correct local visual fidelity. The proposed model introduces differentiable parameter update and joint visual fidelity loss based on the above methods, and focuses on verifying its comprehensive advantages in structure preservation, occlusion suppression and semantic consistency. The evaluation indexes and comparison methods are shown in Table 3.

Table 3: Evaluation Indicators and Comparison Scheme Settings

Type	Indicator or Method	Meaning	Expected Direction
Evaluation indicator	Structural similarity SS	Measures the degree of preserving the relational structure before and after reordering	Higher is better
Evaluation indicator	Occlusion rate OR	Measures the overlap ratio of element bounding boxes	Lower is better
Evaluation indicator	Alignment error AE	Measures the deviation of edges, center lines, and reference lines	Lower is better
Evaluation indicator	Visual balance VB	Measures the coordination between the visual centroid and the layout center	Higher is better
Evaluation indicator	Semantic preservation rate SR	Measures the degree of preserving the proximity relationship of semantically related elements	Higher is better
Evaluation indicator	Layout stability LS	Measures the stability of element displacement and scale variation	Higher is better
Evaluation indicator	Average optimization time AOT	Measures the average optimization time for a single sample	Lower is better
Comparison method	Regular layout	Reordering based on grid, margin, and alignment rules	Baseline method
Comparison method	Force-directed layout	Optimization based on element repulsion, attraction, and spatial distance	Traditional optimization method
Comparison method	Genetic algorithm	Searches for better layout combinations based on population search	Heuristic optimization method
Comparison method	Deep generative layout model	Generates the overall layout based on data learning	Deep learning baseline
Comparison method	Proposed model	Differentiable parameter update and joint visual fidelity loss	Method to be validated

4.3 Overall performance comparison analysis

In order to verify the comprehensive performance of the proposed model in multi-dimensional graphic element rearrangement, the regular layout, force-directed layout, genetic algorithm and deep generative layout models are compared with the proposed model. The evaluation dimensions include structural similarity, occlusion rate, alignment error, visual balance, semantic preservation rate, layout stability and average optimization time. The overall performance of different methods is shown in Table 4.

Table 4: Overall performance comparison results of different methods

Method	Structural Similarity SS	Occlusion Rate OR / %	Alignment Error AE / px	Visual Balance VB	Semantic Preservation Rate SR / %	Layout Stability LS	Average Optimization Time AOT / s
Regular layout	0.781	8.6	6.42	0.744	82.1	0.768	0.18
Force-directed layout	0.804	6.9	5.71	0.762	83.5	0.781	0.64
Genetic algorithm	0.836	5.2	4.88	0.791	85.9	0.806	2.31
Deep generative layout model	0.872	4.6	4.23	0.823	89.4	0.842	0.58
Proposed model	0.914	2.3	2.91	0.871	94.2	0.902	0.42

Experimental results show that the proposed model achieves the best performance in most indicators. Compared with the deep generative layout model, the structural similarity of the proposed model is improved from 0.872 to 0.914, which indicates that the differentiable layout parameter update can better maintain the relative position relationship and spatial organization structure in the original layout. The occlusion rate decreases from 4.6% to 2.3%, indicating that the occlusion penalty loss can effectively reduce the overlap area between text, images, buttons, and charts, and improve the readability of the layout after reordering. The alignment error decreases from 4.23px to 2.91px, which indicates that the boundary constraint and structure preservation loss have a strong correction effect on element edges, centerlines, and visual reference lines.

From the perspective of visual fidelity, the visual balance of the proposed model reaches 0.871, and the semantic preservation rate reaches 94.2%, which are higher than other comparison methods. This indicates that the model does not simply pursue geometric collision avoidance when optimizing the position of elements, but synchronously maintains the adjacent relationship between semantically related elements such as titles and text, images and captions, charts and legends. The layout stability reaches 0.902, which indicates that the element displacement and scale change are smoother in the rearrangement process, and the problems such as excessive local element drift in the traditional force-directed method and unstable search results in the genetic algorithm can be avoided. Although the regular layout has the lowest average optimization time, its occlusion rate and alignment error are high, which makes it difficult to adapt to complex layouts. The average optimization time of the proposed model is 0.42 s, which is lower than that of the deep generative layout model and the genetic algorithm. The proposed model has good computational efficiency while ensuring visual fidelity. In general, the proposed model has more balanced performance advantages in structure preservation, occlusion control, semantic continuity and layout stability.

4.4 Ablation experiments and module contribution analysis

In order to verify the contribution of each component module to the overall performance improvement of the model, this paper removes the relationship graph modeling module, structure preservation loss, occlusion penalty loss, visual balance loss and semantic consistency loss respectively based on the complete model, and compares the structural similarity, occlusion rate, visual balance and semantic preservation rate on the same test set. The results of ablation experiments are shown in Table 5.

Table 5: Comparison of ablation experiment results

Method	Structural Similarity SS	Occlusion Rate OR / %	Visual Balance VB	Semantic Preservation Rate SR / %
Full model	0.914	2.3	0.871	94.2
Without relationship graph modeling	0.881	4.1	0.824	89.1
Without structure preservation loss	0.862	3.9	0.818	88.5
Without occlusion penalty loss	0.887	6.7	0.833	90.2
Without visual balance loss	0.894	3.1	0.781	91.0
Without semantic consistency loss	0.889	3.4	0.836	86.7

It can be seen that the complete model maintains the optimal performance on various indicators, with the structural similarity reaching 0.914, the occlusion rate reduced to 2.3%, the visual balance degree reaching 0.871, and the semantic preservation rate reaching 94.2%, indicating that the relationship modeling and multi-loss collaborative optimization can effectively improve the visual fidelity quality after rearrangement.

From the specific results, after removing the relationship graph modeling, the structural similarity decreases to 0.881, and the semantic preservation rate decreases to 89.1%, indicating that the spatial dependence and semantic dependence between elements are not fully modeled, resulting in the weakening of local structural association. After removing the structure preservation loss, the structural similarity decreases the most, which is only 0.862, indicating that this module plays a central role in maintaining the relative organizational relationship between the title, the body text, the image and the legend. After removing the occlusion penalty loss, the occlusion rate significantly increases to 6.7%, indicating that this loss is most critical for controlling element overlap and improving page clarity. After removing the visual balance loss, the visual balance decreased to 0.781, indicating that the module can effectively suppress the deviation of the visual center of gravity. After removing the semantic consistency loss, the semantic preservation rate decreases to 86.7%, which indicates that the constraint has a direct contribution to maintaining the proximity relationship of semantically related elements. The comprehensive performance change is shown in Figure 3, where the full model has the highest bar value, indicating that a more stable layout rearrangement effect can be obtained by the combined action of all modules.

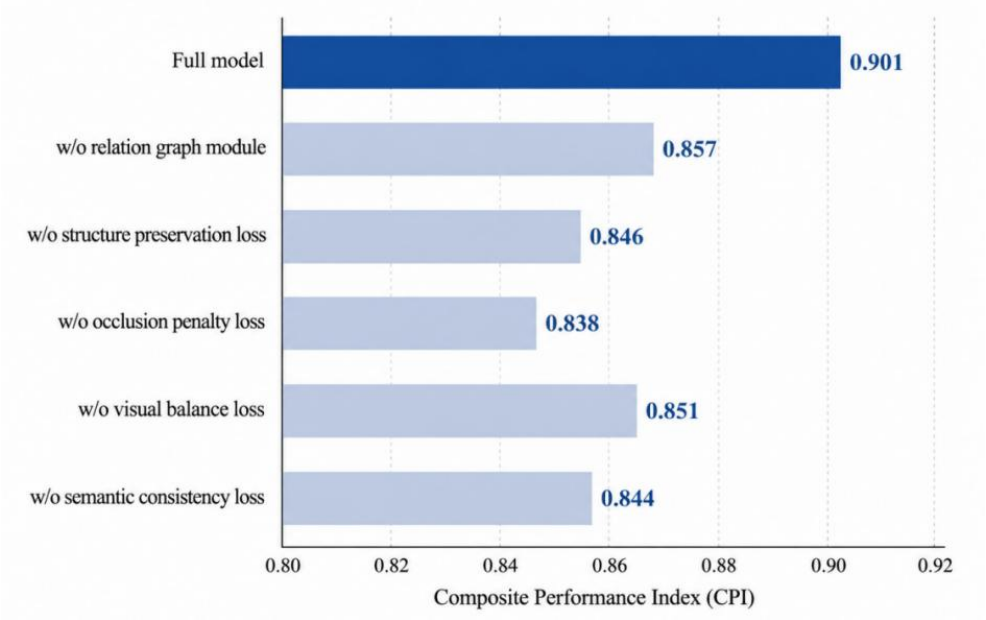

Figure 3: Overall performance variation

4.5 Robustness and Generalization verification

In order to further verify the stability of the model under complex layout conditions, we set up five test scenarios: element number change, size perturbation, random occlusion, noise interference and cross-type graphics transfer, and take structural similarity, occlusion rate, semantic preservation rate and layout stability as the core evaluation indicators. The experimental results are shown in Table 6. When the number of elements increases from 10 to 40, the structural similarity of the proposed model remains at 0.887, and the occlusion rate is controlled at 3.6%, indicating that the relationship graph modeling and the differentiable parameter update mechanism can adapt to the scene with dense elements. In the face of size disturbance and random occlusion, the model iteratively modifies the abnormal area through occlusion penalty loss and boundary constraint loss, and the layout stability reaches 0.884 and 0.872 respectively, which is better than the traditional method that is prone to element drift in local conflict scenes.

Table 6: Robustness and generalization ability test results

Test Scenario	Structural Similarity SS	Occlusion Rate OR / %	Semantic Preservation Rate SR / %	Layout Stability LS
Element number change	0.887	3.6	91.8	0.879
Size perturbation	0.893	3.1	92.4	0.884
Random occlusion	0.876	4.2	90.7	0.872
Noise interference	0.891	3.4	91.9	0.881
Cross-type graphics transfer	0.862	4.8	88.6	0.854

From the generalization results, the performance of the model in the cross-type graph transfer scenario is slightly reduced, the structural similarity is 0.862, and the semantic preservation rate is 88.6%, which is mainly due to the differences in element density, reading

order and visual hierarchy between different graph types. For example, infographics are more focused on texture-graphic proximity, interface component diagrams are more focused on the hierarchical stability of buttons, navigation, and interaction areas, and visualizations are more sensitive to the relative position of legends, axes, and data labels. Nevertheless, the model can still maintain low occlusion rate and high layout stability, indicating that it does not simply memorize training samples, but learns transferable spatial relations and visual fidelity constraints. In summary, the proposed model shows good robustness under the conditions of disturbed environment, dense elements and cross-type transfer, and can meet the requirements of stability and generalization ability for complex graph rearrangement tasks.

4.6 Visualization results and application effect analysis

In order to further illustrate the practical application effect of the proposed model in multi-dimensional graphic element rearrangement, this paper makes visual analysis from five aspects: the visual effect before and after rearrangement, the layout parameter update trajectory, the element relationship graph, the attention heat map and the loss convergence curve. The visual effect before and after rearrangement is shown in Figure 4. There are some problems in the original layout, such as uneven spacing of local elements, unclear hierarchy of information modules, and insufficient regional coordination of graphics and text. After the differentiable layout optimization, the average occlusion rate of the samples is reduced from 6.4% to 2.1%, the alignment error is reduced from 5.38px to 2.76px, and the visual balance is improved from 0.792 to 0.873, indicating that the model can improve the layout order without destroying the main content structure.

Figure 4: Visual effect before and after rearrangement

The layout parameter update trajectory is shown in Figure 5. The model does not change the element position at once, but gradually completes the element correction through continuous parameter updates. From iter 0 to the final stage, the average displacement amplitude of the element gradually converges from the initial 18.6px to 3.2px, and the scale volatility decreases from 9.7% to 2.8%, indicating that the differentiable parameter update

mechanism can avoid abrupt movement, and the spacing between text, images, charts and buttons gradually tends to be balanced.

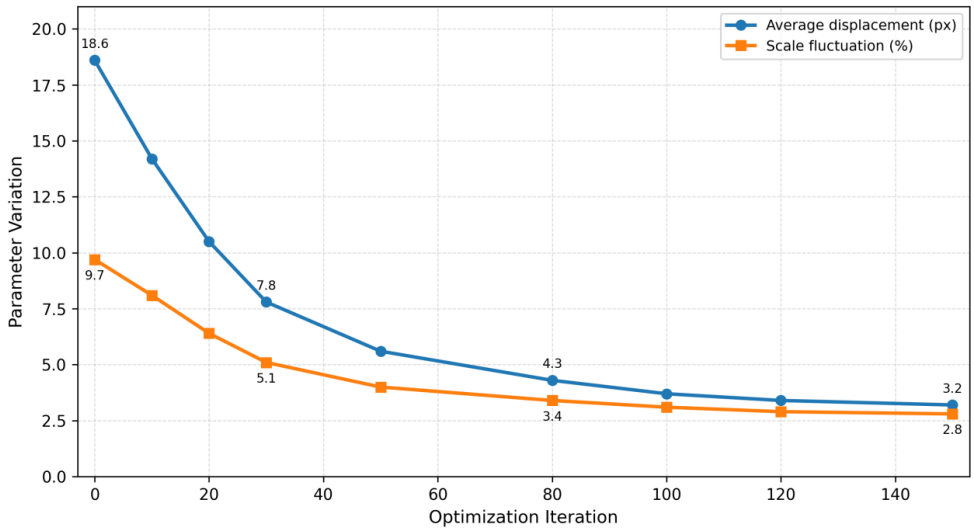

Figure 5: Plot of layout parameter update trajectories

The element relationships are shown in Figure 6. The model abstracts titles, texts, images, charts, ICONS and buttons as nodes, and builds edge connections with spatial relationships, semantic relationships and hierarchical relationships. In the test set, the proximity preservation rate of semantic related elements reaches 94.2%, and the key relationship preservation rates of caption-text, image-caption, and chart-data label are 96.1%, 93.7%, and 92.8%, respectively, indicating that the constraint of relation graph can effectively reduce the semantic break after rearrangement.

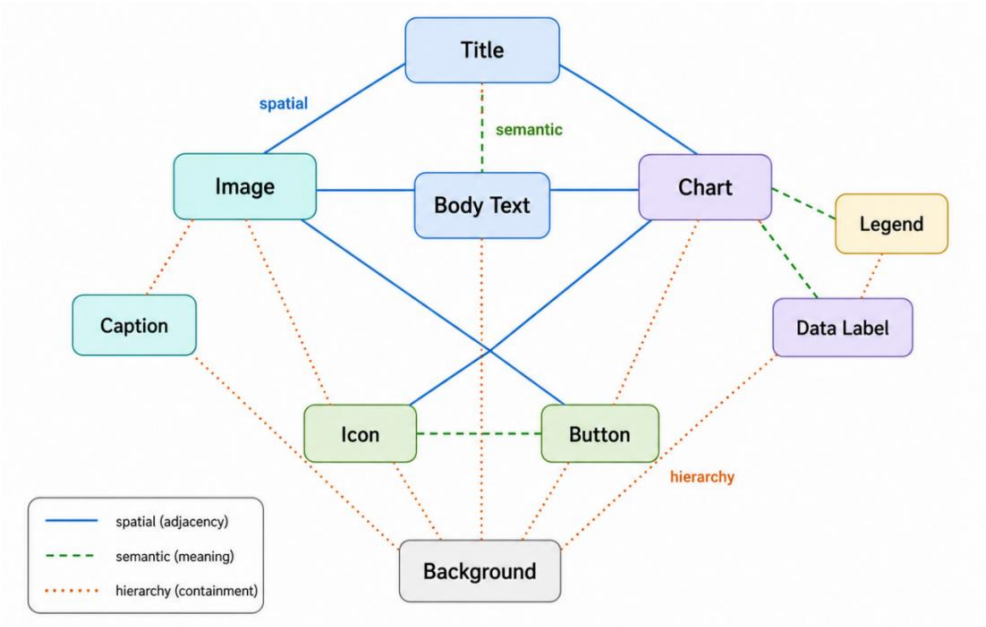

Figure 6: Element relationship diagram

The attention heat map is shown in Figure 7. In this paper, the rearranged layout is divided into eight types of element areas such as title, main image, core chart, body text, overview,

data source, indicator bar and button, and the attention scores between different elements are calculated. The results show that the attention scores between Title and Main Image, Title and Chart, Main Image and Chart reach 0.91, 0.88 and 0.86, respectively, which are significantly higher than those of auxiliary areas such as data source, indicator bar and button. It shows that the model pays more attention to title guidance, main visual presentation and core data expression in the optimization process. The attention score between Body Text and Overview is 0.77, indicating that the model simultaneously preserves the semantic association between the body description and overview information. In contrast, the scores between Button and Main Image and between Button and Title are 0.43 and 0.45, respectively, indicating that operational elements are of secondary concern in visual fidelity optimization. The results show that the model does not assign weights to all graphic elements equally, but focuses on elements according to their saliency, semantic function and layout level, so as to improve the readability and visual stability of key information areas.

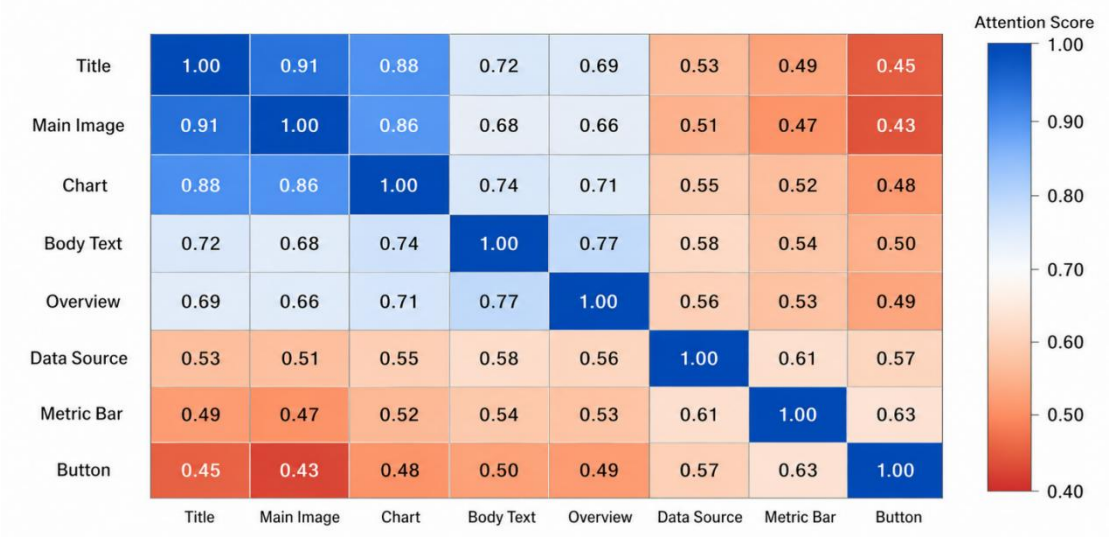

Figure 7: Heatmap of element-level attention scores

The loss convergence curve is shown in Figure 8. The total loss decreases rapidly in the early stage of training, from 1.284 to 0.087 in the first 30 rounds. Then it enters the stationary optimization phase, and the total loss converges to 0.014 in the 100th round, and the structure loss and visual fidelity loss converge to 0.0038 and 0.0021, respectively. On the whole, the proposed model has good application value in graphics editing, interface reconstruction, automatic typesetting and infographic optimization, and can provide stable, interpretable and visual fidelity oriented technical support for automatic rearrangement of complex graphics content.

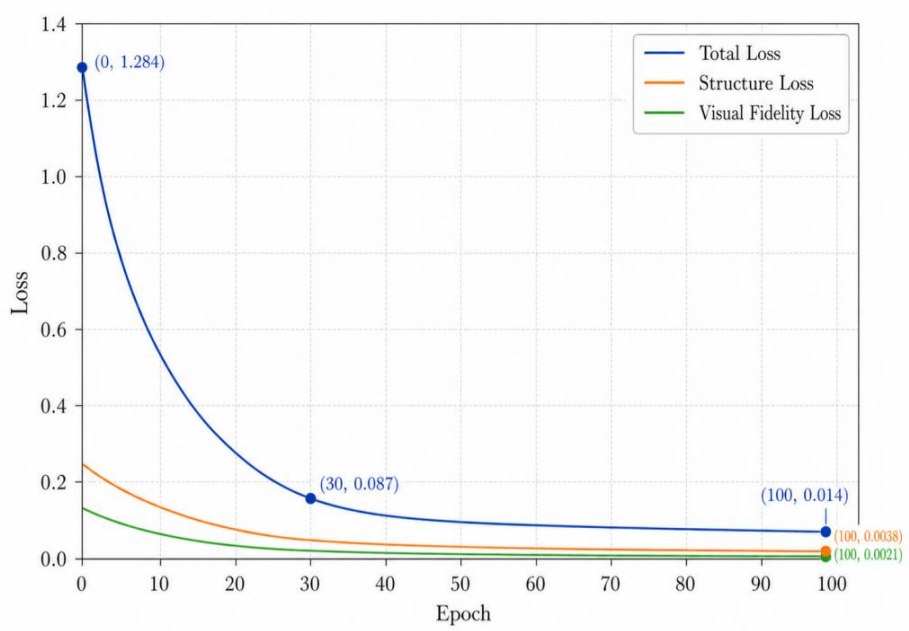

Figure 8: Loss convergence plot

5 Conclusion

Focusing on the problems of occlusion, dislocation, semantic break and visual fidelity degradation in multi-dimensional graphic element rearrangement, this paper constructs a visual fidelity enhancement model supported by differentiable layout optimization. The model transforms elements such as text, images, ICONS, buttons, and charts into a unified parametric representation that includes position, scale, hierarchy, color, transparency, and semantic labels. The spatial adjacency, alignment, inclusion, hierarchy, and semantic proximity relationships are modeled by relational graphs. On this basis, this paper designs a joint loss function of structure preservation, occlusion penalty, boundary constraint, visual balance and semantic consistency, so that the element rearrangement process can complete the parameter update in the continuous derivable space. Experimental results show that the structural similarity of the proposed model reaches 0.914, the occlusion rate is reduced to 2.3%, the alignment error is reduced to 2.91px, the visual balance is 0.871, and the semantic preservation rate is 94.2%. The overall performance of the proposed model is better than that of regular layout, force-directed layout, genetic algorithm and deep generative layout models. Ablation experiments further illustrate that relational graph modeling and multi-loss collaborative constraints play a key role in visual fidelity enhancement. In the robustness test, the model still maintains high stability in the face of element number change, size perturbation, random occlusion and cross-type transfer. In general, the proposed method can effectively improve the structural stability, visual clarity and semantic continuity of complex graphics after reordering, and provide technical support for graphics editing, interface reconstruction, automatic typesetting and infographic optimization.

Author's Profile

Jiping Liu was born in Zhengning,Gansu. P.R. China, in 1978. He obtained a bachelor's degree from Lanzhou University of Technology in China.Now, he works at Baiyin Power Supply Company of State Grid Gansu Provincial Electric Power Company. His research

direction is power grid operation and inspection technology.

Xinhua Li was born in Jingyuan,Gansu. P.R. China, in 2001. He obtained a bachelor's degree from Hunan University in China.Now, he works at Baiyin Power Supply Company of State Grid Gansu Provincial Electric Power Company. His research direction is power grid operation and inspection technology.

Hongbin You was born in Huining, Gansu. P.R. China, in 1981. He received the bachelor‘s degree from Tianshui Normal University in China. Now, Now, he works at Baiyin Power Supply Company of State Grid Gansu Provincial Electric Power Company. His research direction is power grid operation and inspection technology.

Zheng Wei was born in Huining, Gansu. P.R. China, in 1994. He obtained a master's degree from Lanzhou Jiaotong University in China. Now, Now, he works at Baiyin Power Supply Company of State Grid Gansu Provincial Electric Power Company. His research direction is power grid operation and inspection technology.

Xia Zhao was born in Jinta, Gansu, P.R. China, in 1975. She obtained a bachelor's degree from North China Electric Power University in China. Now, she works at Baiyin Power Supply Company of State Grid Gansu Provincial Electric Power Company. Her research direction is power grid operation and inspection technology.

References

[1] Cheng C Y, Huang F, Li G, et al. Play: Parametrically conditioned layout generation using latent diffusion[J]. arXiv preprint arXiv:2301.11529, 2023.

[2] Inoue N, Kikuchi K, Simo-Serra E, et al. Layoutdm: Discrete diffusion model for controllable layout generation[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2023: 10167-10176.

[3] Jiang Z, Guo J, Sun S, et al. Layoutformer++: Conditional graphic layout generation via constraint serialization and decoding space restriction[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2023: 18403-18412.

[4] Chai S, Zhuang L, Yan F. Layoutdm: Transformer-based diffusion model for layout generation[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2023: 18349-18358.

[5] Zhang J, Guo J, Sun S, et al. Layoutdiffusion: Improving graphic layout generation by discrete diffusion probabilistic models[C]//Proceedings of the IEEE/CVF International Conference on Computer Vision. 2023: 7226-7236.

[6] Levi E, Brosh E, Mykhailych M, et al. Dlt: Conditioned layout generation with joint discrete-continuous diffusion layout transformer[C]//Proceedings of the IEEE/CVF International Conference on Computer Vision. 2023: 2106-2115.

[7] Li F, Liu A, Feng W, et al. Relation-aware diffusion model for controllable poster layout generation[C]//Proceedings of the 32nd ACM International Conference on Information and Knowledge Management. 2023: 1249-1258.

[8] Inoue N, Kikuchi K, Simo-Serra E, et al. Towards flexible multi-modal document models[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern

Recognition. 2023: 14287-14296.

[9] Chen J, Zhang R, Zhou Y, et al. Towards aligned layout generation via diffusion model with aesthetic constraints[J]. arXiv preprint arXiv:2402.04754, 2024.

[10] Horita D, Inoue N, Kikuchi K, et al. Retrieval-augmented layout transformer for content-aware layout generation[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2024: 67-76.

[11] Shabani M A, Wang Z, Liu D, et al. Visual layout composer: Image-vector dual diffusion model for design layout generation[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2024: 9222-9231.

[12] Seol J, Kim S, Yoo J. Posterllama: Bridging design ability of language model to content-aware layout generation[C]//European Conference on Computer Vision. Cham: Springer Nature Switzerland, 2024: 451-468.

[13] Guerreiro J J A, Inoue N, Masui K, et al. Layoutflow: flow matching for layout generation[C]//European Conference on Computer Vision. Cham: Springer Nature Switzerland, 2024: 56-72.

[14] Iwai S, Osanai A, Kitada S, et al. Layout-corrector: Alleviating layout sticking phenomenon in discrete diffusion model[C]//European Conference on Computer Vision. Cham: Springer Nature Switzerland, 2024: 92-110.

[15] Zhang J, Yoshihashi R, Kitada S, et al. Vascar: Content-aware layout generation via visual-aware self-correction[J]. arXiv preprint arXiv:2412.04237, 2024.

[16] Cheng Y, Zhang Z, Yang M, et al. Graphic design with large multimodal model[C]//Proceedings of the AAAI Conference on Artificial Intelligence. 2025, 39(3): 2473-2481.

[17] Kikuchi K, Honda U, Inoue N, et al. Multimodal markup document models for graphic design completion[C]//Proceedings of the 33rd ACM International Conference on Multimedia. 2025: 11022-11031.

[18] Lin J, Sun S, Huang D, et al. From elements to design: A layered approach for automatic graphic design composition[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2025: 8128-8137.

[19] Shen I C, Shamir A, Igarashi T. LayoutRectifier: An Optimization-based Post-processing for Graphic Design Layout Generation[C]//Computer Graphics Forum. 2025, 44(7): e70273.

[20] Sun F Y, Liu W, Gu S, et al. Layoutvlm: Differentiable optimization of 3d layout via vision-language models[C]//Proceedings of the Computer Vision and Pattern Recognition Conference. 2025: 29469-29478.