IIETA International Information and Engineering Technology Association
Advancing the World of Information and Engineering

Traitement du Signal
Vol. 43, No. 1, February, 2026, pp. 505-517
Journal homepage: http://iieta.org/journals/ts

Automatic Image Composition Optimization and Visual Perception Simulation Evaluation for Architectural Interior Design

Shuo Zhang[1], Lidan Zhang[2*]

[1] Hebei Art & Design Academy, Baoding 071051, China
[2] Gaobeidian Vocational Education Center, Baoding 074099, China

Corresponding Author Email: zhanglidan789@126.com

https://doi.org/10.18280/ts.430136

Received: 5 September 2025
Revised: 20 November 2025
Accepted: 5 December 2025
Available online: 28 February 2026

Keywords:
architectural interior environment design, image composition optimization, generative adversarial networks, computable prior modeling, visual perception simulation evaluation, deep learning

ABSTRACT

The composition optimization of architectural interior design images has long relied on human experience, resulting in low optimization efficiency, subjective evaluation criteria, and a lack of constraints from professional design guidelines, which severely restricts the standardization and automation of interior design visualization schemes. To address the above issues, this paper proposes an automatic image composition optimization and visual perception simulation evaluation method for interior design images, integrating interior composition prior knowledge, generative deep learning, and cognitive science. This method constructs a differentiable prior module for assessing interior composition quality, quantitatively embedding abstract interior design composition criteria into the optimization process; an end-to-end sparse displacement field-guided generative network is designed to intelligently relocate movable objects and accurately restore background content, avoiding content loss and distortion of traditional methods; a multidimensional, no-reference visual perception simulation evaluation system based on cognitive science is built to achieve objective quantitative assessment of optimization results. Experimental results show that the proposed method significantly outperforms traditional cropping, image retargeting, and mainstream deep learning composition optimization algorithms in key indicators such as subjective aesthetic scores, visual saliency concentration, and eye movement path efficiency, effectively improving the compositional quality of interior design images, and realizing automated and standardized composition optimization and perceptual evaluation, providing technical support for efficient optimization of architectural interior visualization schemes.

1. INTRODUCTION

The core of architectural interior environment design is to construct a spatial experience that meets user requirements and professional aesthetic standards through reasonable spatial layout, element arrangement, and light-shadow control [1-3]. As a visualization carrier of design schemes, the composition quality of interior design images directly determines the efficiency of conveying design intentions, the accuracy of scheme evaluation, and the visual perception experience of users [4, 5]. Traditional interior design image composition optimization highly relies on designers' experience, with prominent problems such as low optimization efficiency, subjective evaluation criteria, and poor cross-designer consistency, making it difficult to meet the large-scale and high-efficiency requirements of modern architectural interior design industrialization [6, 7]. With the rapid development of computer vision [8] and deep learning technologies [9], automatic image composition optimization has become a research hotspot in the intersection of computer vision and design. However, existing automated methods mostly focus on general scene images [10, 11] and lack sufficient targeting for interior design scenarios, with two main bottlenecks: the optimization process lacks constraints from composition rules specific to interior scenes, easily generating images that violate professional design principles such as spatial perspective and visual balance [12]; the evaluation of optimization results mostly relies on manual subjective scoring, lacking an objective and quantifiable cognitive science-guided evaluation system, making it difficult to accurately represent the visual perception quality and design rationality of images [13]. Therefore, conducting research on automatic composition optimization algorithms for architectural interior design images and constructing matching objective visual perception simulation evaluation methods has important theoretical and practical value for promoting the automated generation and optimization of interior design schemes, improving design efficiency, and unifying evaluation standards, providing technical support for the intelligent development of architectural interior design visualization.

Traditional image composition optimization methods are centered on image cropping and retargeting, achieving composition adjustment through low-level visual feature extraction. Although simple to operate, these methods have obvious limitations, lacking deep understanding of scene semantics and professional composition rules, and are difficult to adapt to the complex spatial logic and aesthetic requirements of interior design scenarios [14, 15]. The rise of deep learning technologies provides a new technical path for image composition optimization. Methods based on generative adversarial networks and Transformers can generate and reorganize image content, significantly improving the flexibility of composition optimization. However, existing deep learning-based composition optimization methods mostly target general scenes [16, 17] and do not fully combine design rules specific to interior scenes such as perspective consistency and spatial balance, making optimization results prone to spatial logic confusion and unreasonable element arrangement. Some studies attempt to integrate semantic information into composition optimization, but semantic modeling only stays at the region recognition level [18, 19], without achieving computable quantification of professional design rules, limiting the constraint effect on the optimization process and failing to meet professional requirements of interior design. Corresponding to composition optimization methods, the field of image visual perception evaluation also has similar limitations. Existing evaluation methods are mainly divided into subjective evaluation and objective evaluation. Although subjective evaluation conforms to human cognitive habits, it has inherent defects such as being time-consuming, highly subjective, and inconsistent, and cannot meet the high-efficiency evaluation requirements of automated composition optimization [20]. Objective evaluation methods are mainly based on low-level visual features or aesthetic models. These methods do not incorporate visual attention mechanisms from cognitive science, making it difficult to accurately represent human perception and cognition of interior design images, and the evaluation results significantly deviate from human subjective perception [21]. Eye-tracking technology has been proven to effectively reflect human visual attention patterns. Some studies attempt to integrate eye movement features into image evaluation, but relevant research mostly focuses on general images [22, 23], without constructing a dedicated evaluation system for interior design scenarios, and lacks a no-reference automated evaluation implementation, making it difficult to form an effective match with automated composition optimization methods and failing to achieve closed-loop interaction between optimization and evaluation.

Comprehensively analyzing existing research results, there are still three core deficiencies in the research on composition optimization and visual perception evaluation of architectural interior design images: first, there is a lack of computable composition prior modeling methods specific to interior scenes, failing to transform professional design rules such as perspective consistency, visual balance, and white space ratio into quantifiable and differentiable constraint modules, making the optimization results prone to violating interior spatial design logic; second, the content generation and restoration capability of composition optimization networks is insufficient, making it difficult to achieve precise relocation of movable objects and consistent restoration of background light and shadow, prone to content holes, texture distortion, and other problems, affecting the realism and aesthetics of optimized images; third, the visual perception evaluation system lacks cognitive science support, failing to construct a no-reference objective evaluation system integrating visual attention mechanisms and eye movement behavior features, unable to accurately quantify the perception quality and composition guidance effect of interior design images, and difficult to objectively verify the effectiveness of optimization algorithms.

To address the above research deficiencies, this paper proposes an automatic image composition optimization algorithm and visual perception simulation evaluation method for architectural interior environment design. The core innovations are reflected in three aspects: proposing a computable modeling method for interior composition prior knowledge, constructing a differentiable composition aesthetic evaluation prior module including perspective consistency loss, visual balance loss, and white space ratio loss, achieving quantitative embedding and gradient constraint of professional design rules; designing an end-to-end sparse displacement field-guided composition optimization generative network to realize intelligent relocation of movable objects and accurate restoration of background content, solving the problems of content loss and distortion in traditional methods; building a multidimensional no-reference visual perception simulation evaluation system based on cognitive science, integrating visual saliency analysis, virtual eye movement simulation, and no-reference aesthetic scoring, achieving objective quantitative evaluation of optimization results and overcoming the limitations of manual subjective evaluation. The chapter arrangement of this paper is as follows: Chapter 2 describes the computable modeling method of interior composition prior knowledge; Chapter 3 details the design of the composition optimization generative network based on generative adversarial networks; Chapter 4 constructs the cognitive science-guided visual perception simulation evaluation system; Chapter 5 verifies the effectiveness and superiority of the proposed method through experiments; Chapter 6 summarizes the research results of the paper and discusses future research directions.

2. COMPUTABLE MODELING OF INTERIOR COMPOSITION PRIOR KNOWLEDGE

2.1 Fine-grained semantic parsing of interior scenes

Figure 1 shows the overall framework of automatic image composition optimization and simulation evaluation for interior design images. As shown in the figure, the computable modeling of interior composition prior knowledge requires precise semantic parsing of interior scenes as the basis. Only by clearly identifying key regions and element types in the image can abstract composition rules be quantitatively transformed and gradient-constrained. In this paper, a pre-trained DeepLabV3+ semantic segmentation network is used to perform pixel-level semantic segmentation on input interior design images, accurately identifying key semantic regions such as floors, walls, ceilings, core furniture, and decorative elements, and outputting semantic segmentation maps and class probability maps to provide basic semantic information for the subsequent design of composition prior losses. Considering the spatial layout characteristics of interior scenes, secondary optimization of the semantic segmentation results is required. By performing connected component analysis to remove noisy regions and correcting semantic categories based on spatial positional relationships, it ensures that the semantic parsing results conform to the logic of interior spatial layout, thereby providing precise and reliable regional semantic support for the computable modeling of composition prior rules.

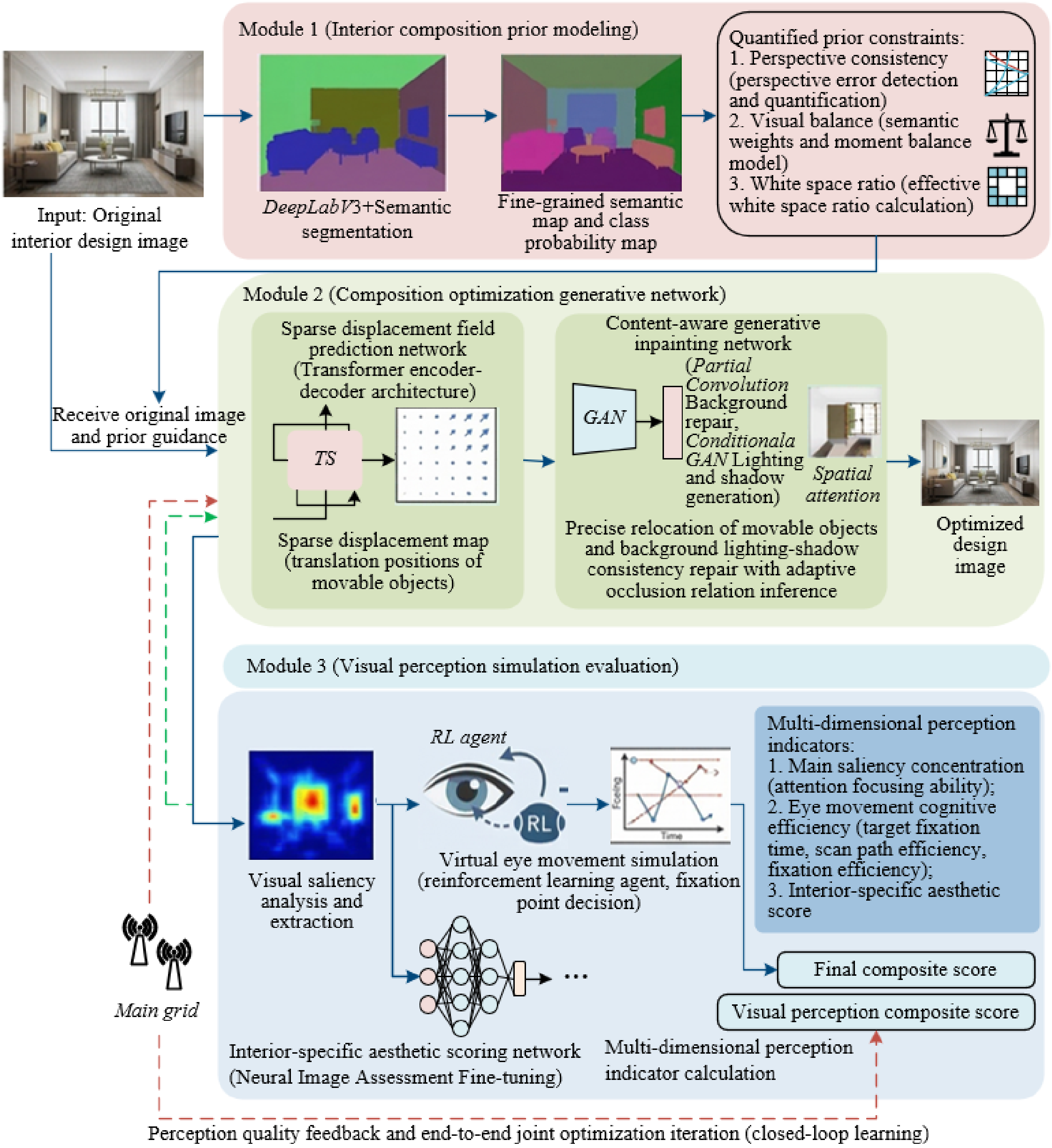

Module 1 (Interior composition prior modeling)
Input: Original interior design image
DeepLabV3+Semantic segmentation
Fine-grained semantic map and class probability map
Quantified prior constraints:
1. Perspective consistency (perspective error detection and quantification)
2. Visual balance (semantic weights and moment balance model)
3. White space ratio (effective white space ratio calculation)
Module 2 (Composition optimization generative network)
Receive original image and prior guidance
Sparse displacement field prediction network (Transformer encoder-decoder architecture)
TS
Sparse displacement map (translation positions of movable objects)
Content-aware generative inpainting network
GAN
(Partial Convolution Background repair, Conditionala GAN Lighting and shadow generation)
Spatial attention
Precise relocation of movable objects and background lighting-shadow consistency repair with adaptive occlusion relation inference
Optimized design image
Module 3 (Visual perception simulation evaluation)
RL agent
RL
Visual saliency analysis and extraction
Virtual eye movement simulation (reinforcement learning agent, fixation point decision)
Multi-dimensional perception indicators:
1. Main saliency concentration (attention focusing ability);
2. Eye movement cognitive efficiency (target fixation time, scan path efficiency, fixation efficiency);
3. Interior-specific aesthetic score
Main grid
Final composite score
Visual perception composite score
Interior-specific aesthetic scoring network (Neural Image Assessment Fine-tuning)
Multi-dimensional perception indicator calculation
Perception quality feedback and end-to-end joint optimization iteration (closed-loop learning)

Figure 1. Overall framework of automatic image composition optimization and simulation evaluation for interior design images

2.2 Design of differentiable composition prior loss

Perspective consistency is the core logical principle of interior spatial composition. Existing methods are difficult to convert it into a differentiable constraint, resulting in optimization outcomes prone to spatial logic confusion. This paper innovatively quantifies interior perspective rules into a differentiable perspective consistency loss, accurately detecting the main perspective lines of the image through deep Hough transform, and quantifying perspective error based on the arrangement direction of core furniture. Let the main perspective line direction vector be $\vec{v}_{vanish}$, and the set of core furniture regions be $F = \{f_1, f_2, \ldots, f_n\}$. By extracting the main arrangement direction vector $\vec{v}_{f_i}$of each furniture region f_i via the minimum bounding rectangle, the perspective error of a single piece of furniture is defined to quantify the deviation of furniture arrangement from the perspective line:

$$e_i=1-\frac{|\vec{v}_{f_i}\cdot\vec{v}_{vanish}|}{\|\vec{v}_{f_i}\|\cdot\|\vec{v}_{vanish}\|} \tag{1}$$

The perspective consistency loss is the mean of all core furniture perspective errors, i.e.:

$$L_{perspective}=\frac{1}{n}\sum_{i=1}^{n} e_i \tag{2}$$

This loss can directly participate in gradient backpropagation, penalizing perspective deviation to force optimized furniture arrangements to maintain consistency with interior spatial perspective rules, solving the problem that traditional methods cannot quantify perspective rules as constraints.

Visual balance is the core aesthetic requirement of interior composition. Existing semantic-driven composition optimization only achieves region recognition and does not establish a quantitative mechanism for visual weight, making it difficult to ensure composition balance. This paper innovatively proposes a visual balance loss based on semantic weights and the principle of moment equilibrium, accurately calculating balance errors by quantifying the visual weights of semantic regions. First, semantic region weights are defined according to interior design priority: core furniture, secondary furniture, decorative elements, and background regions are set as $w_{core} = 5$, $w_{secondary} = 3$, $w_{decor} = 2$, and $w_{bg} = 1$, respectively. Based on semantic segmentation results, the total visual weights of left, right, top, and bottom regions of the image W_{left}, W_{right}, W_{top}, W_{bottom} are calculated, and horizontal and vertical balance errors are defined:

$$e_{\mathrm{horiz}}=\frac{|W_{left}\text{-}W_{right}|}{W_{left}+W_{right}} \tag{3}$$

$$e_{vert}=\left|\frac{W_{top}\text{-}W_{bottom}}{W_{top}+W_{bottom}}\right| \tag{4}$$

The visual balance loss is the weighted sum of the two, i.e., $L_{balance} = \alpha_{horiz} \cdot e_{horiz} + \alpha_{vert} \cdot e_{vert}$, where $\alpha_{horiz} = \alpha_{vert} = 1$. Through gradient optimization, the visual weight distribution of the image tends to balance, overcoming the limitation of traditional methods that rely only on low-level features and cannot accurately match interior aesthetic requirements.

The white space ratio directly determines the spatial openness and visual comfort of interior images. Existing methods do not clearly define effective white space in interiors, making precise control of density difficult. This paper innovatively defines effective interior white space as the part of the background region not occluded by core and secondary furniture, excluding the influence of decorative element occlusion, ensuring that white space quantification conforms to interior design aesthetic standards. Let the total number of image pixels be P_{total} and the number of pixels in the effective white space region be P_{empty}, then the white space ratio $R_{empty} = P_{empty}/P_{total}$. Based on the optimal white space range in interior design, the target white space ratio R_{target} (range 0.4–0.6) is set, and the white space loss is constructed using mean squared error: $L_{empty} = (R_{empty} - R_{target})^2$. This loss can control the white space ratio within the optimal range through gradient optimization, avoiding compositions that are too crowded or too sparse, realizing computable and optimizable interior white space aesthetics, and solving the problem that traditional methods have coarse white space control disconnected from professional design requirements.

2.3 Construction of differentiable composition prior module

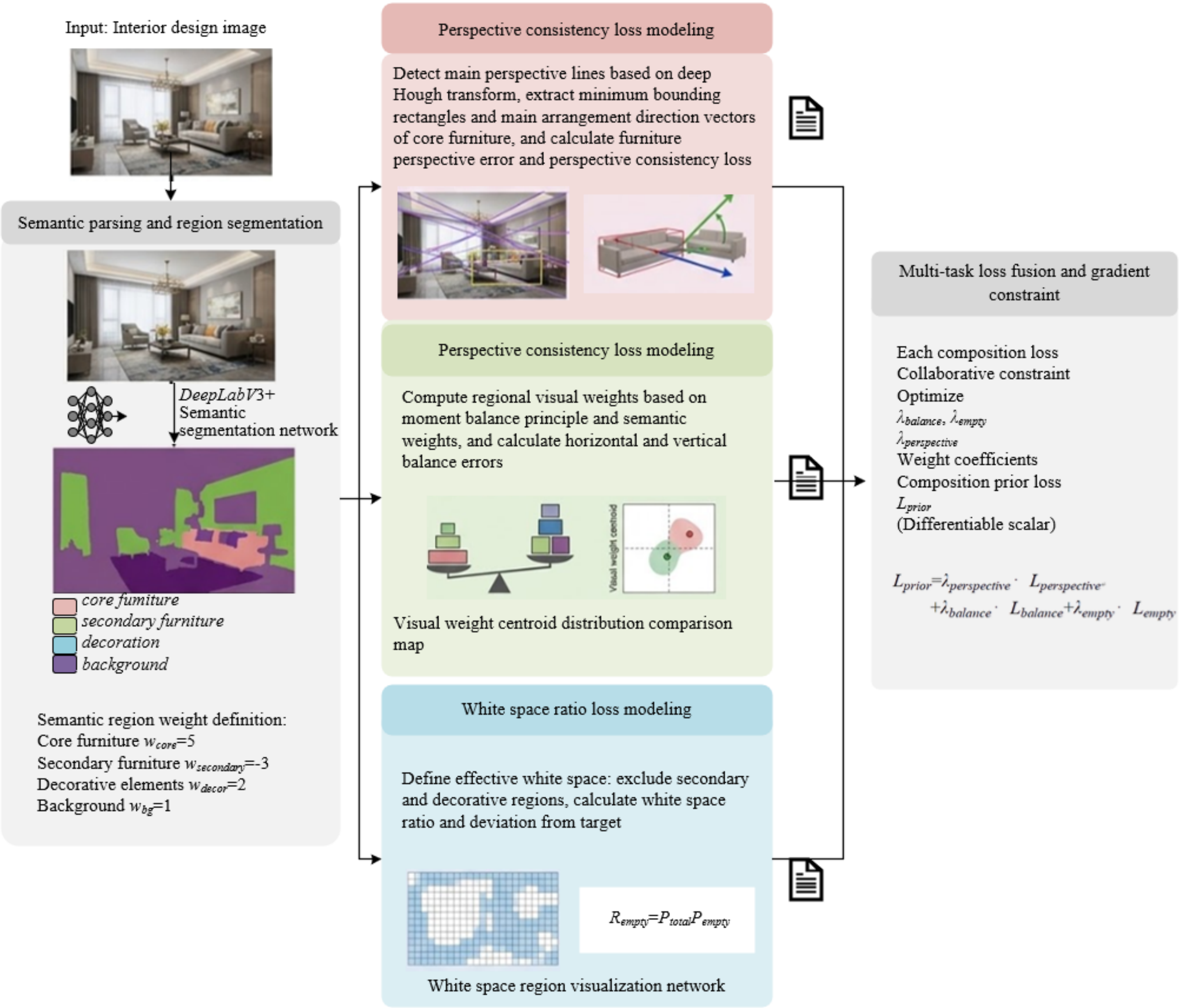

Input: Interior design image
Semantic parsing and region segmentation
DeepLabV3+
Semantic
segmentation network
🟥 core fumiture
🟩 secondary furniture
🟦 decoration
🟪 background
Semantic region weight definition:
Core furniture w_{core}=5
Secondary furniture $w_{secondary}$=-3
Decorative elements w_{decor}=2
Background w_{bg}=1
Perspective consistency loss modeling
Detect main perspective lines based on deep Hough transform, extract minimum bounding rectangles and main arrangement direction vectors of core furniture, and calculate furniture perspective error and perspective consistency loss
Perspective consistency loss modeling
Compute regional visual weights based on moment balance principle and semantic weights, and calculate horizontal and vertical balance errors
Visual weight centroid
Visual weight centroid distribution comparison map
White space ratio loss modeling
Define effective white space: exclude secondary and decorative regions, calculate white space ratio and deviation from target
R_{empty}=$P_{total}P_{empty}$
White space region visualization network
Multi-task loss fusion and gradient constraint
Each composition loss
Collaborative constraint
Optimize
$\lambda_{balance}$, λ_{empty}
$\lambda_{perspective}$
Weight coefficients
Composition prior loss
L_{prior}
(Differentiable scalar)
L_{prior}=$\lambda_{perspective}\cdot$ $L_{perspective}$
+$\lambda_{balance}\cdot$ $L_{balance}$+$\lambda_{empty}\cdot$ L_{empty}

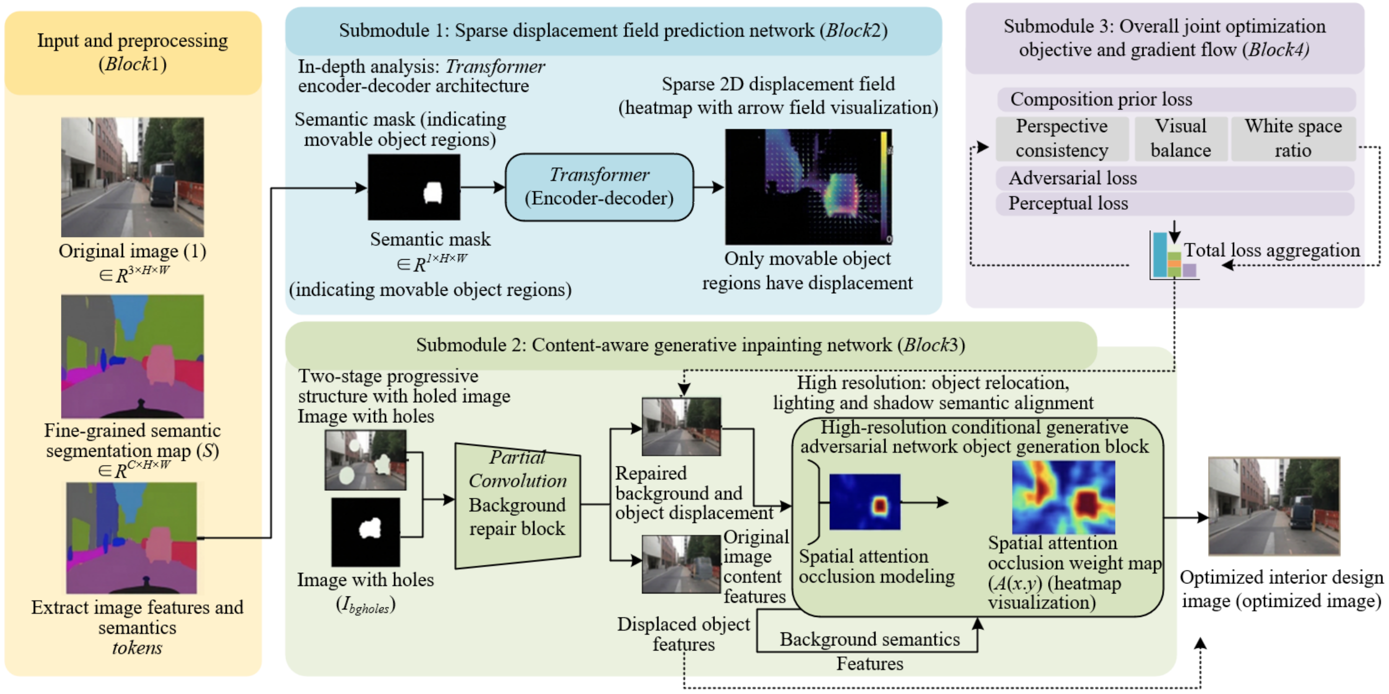
Figure 2. Fine-grained semantic parsing of interior scenes and differentiable quantification of composition prior

To achieve collaborative constraint and gradient optimizability of interior composition prior rules, this paper innovatively constructs a unified differentiable composition prior module, integrating perspective consistency loss, visual balance loss, and white space loss through weighted fusion to form the total module loss, solving the core problem that existing methods lack scene-specific and differentiable unified prior constraints. Figure 2 shows the complete schematic of fine-grained semantic parsing of interior scenes and differentiable quantification of composition prior. The total module loss L_{prior} is calculated as $L_{prior} = \lambda_{perspective} \cdot L_{perspective} + \lambda_{balance} \cdot L_{balance} + \lambda_{empty} \cdot L_{empty}$, where $\lambda_{perspective}$, $\lambda_{balance}$, and λ_{empty} are the weighting coefficients of the three losses, used to balance the constraint strength of each composition rule. Their optimal values are iteratively determined through ablation experiments, set as 0.4, 0.3, and 0.3 respectively, ensuring that perspective logic as the core of interior spatial composition has priority while coordinating visual balance and white space aesthetics. This module takes the original interior design image and fine-grained semantic segmentation map as joint inputs, outputs a quantified composition prior loss value, and can be directly embedded into the loss function of subsequent generative networks for end-to-end gradient backpropagation, allowing the generative network to be jointly constrained by perspective consistency, visual balance, and white space ratio during composition optimization. This ensures that the optimization results comply with both interior spatial layout logic and professional design aesthetic requirements, completely solving the limitations of existing methods where prior constraints are scattered, cannot efficiently interact with generative networks, and are difficult to adapt to professional interior design requirements.

3. COMPOSITION OPTIMIZATION GENERATIVE NETWORK BASED ON GENERATIVE ADVERSARIAL NETWORK

3.1 Overall network architecture design

To address the core deficiencies of traditional composition optimization methods relying on cropping that cause image content loss and cannot achieve element reconstruction, this paper innovatively designs an end-to-end sparse displacement field-guided composition optimization generative network, realizing high-quality composition optimization of interior design images through the collaboration of two core sub-modules. The network takes the original interior image and fine-grained semantic segmentation map as joint inputs and outputs the composition-optimized interior image. The overall architecture includes two core modules: a sparse displacement field prediction network and a content-aware generative repair network, forming a complete optimization process including semantic parsing, sparse displacement field prediction, object relocation, background reconstruction, and detail generation. To ensure that the optimized results have both compositional rationality and content authenticity, during network training, the composition prior loss, adversarial loss, and perceptual loss are jointly optimized throughout, allowing the generative network to follow professional interior composition rules while maintaining image fidelity and visual coherence, completely solving the pain points of content loss and insufficient reconstruction ability in traditional methods, achieving intelligent and lossless optimization of interior composition. Figure 3 shows the architecture of the end-to-end sparse displacement field-guided composition optimization generative network.



Input and preprocessing (*Block*1)
Original image (1) $\in R^{3\times H\times W}$
Fine-grained semantic segmentation map (S) $\in R^{C\times H\times W}$
Extract image features and semantics *tokens*
Submodule 1: Sparse displacement field prediction network (*Block*2)
In-depth analysis: *Transformer* encoder-decoder architecture
Semantic mask (indicating movable object regions)
Transformer (Encoder-decoder)
Sparse 2D displacement field (heatmap with arrow field visualization)
Semantic mask $\in R^{1\times H\times W}$
(indicating movable object regions)
Only movable object regions have displacement
Submodule 3: Overall joint optimization objective and gradient flow (*Block4*)
Composition prior loss
Perspective consistency
Visual balance
White space ratio
Adversarial loss
Perceptual loss
Total loss aggregation
Submodule 2: Content-aware generative inpainting network (*Block*3)
Two-stage progressive structure with holed image
Image with holes
Partial Convolution Background repair block
Image with holes ($I_{bgholes}$)
Repaired background and object displacement
Original image content features
Displaced object features
High resolution: object relocation, lighting and shadow semantic alignment
High-resolution conditional generative adversarial network object generation block
Spatial attention occlusion modeling
Spatial attention occlusion weight map ($A(x.y)$) (heatmap visualization)
Background semantics Features
Optimized interior design image (optimized image)

Figure 3. Architecture of end-to-end sparse displacement field-guided composition optimization generative network

3.2 Sparse displacement field prediction network

Traditional composition optimization methods use a global dense displacement field to adjust image elements, which easily causes misplacement of fixed structures, local distortion, and significant computational redundancy. To address this issue, this module innovatively proposes a sparse displacement field prediction mechanism for movable objects in interiors, abandoning the limitations of global dense displacement fields and assigning non-zero displacements only to semantic movable targets, achieving precise, controllable, and efficient composition optimization. The network takes the original interior image and fine-grained semantic segmentation map as dual inputs and adopts a Transformer encoder-decoder architecture. By capturing long-range layout dependencies in interior spaces through global self-attention, it effectively overcomes the one-sided displacement decision problem caused by the local receptive field of convolutional neural networks. Finally, it outputs a sparse 2D displacement field with the same resolution as the original image, accurately indicating the target translation positions of each movable object, providing reliable displacement guidance for subsequent content reconstruction.

To ensure the rationality and accuracy of the sparse displacement field, this module designs three constraints for fine-grained control, forming a complete displacement decision constraint system. First, hard sparsification of the displacement field is implemented through semantic masks, outputting non-zero 2D displacement vectors only in movable semantic regions such as vases, ornaments, small chairs, and decorative paintings, while enforcing zero displacement in walls, floors, ceilings, and large fixed furniture, fundamentally preventing erroneous movement of fixed structures. Second, a displacement rationality constraint is introduced, restricting the amplitude of displacement vectors based on prior knowledge of interior space scale. Let the displacement vector of a single movable object be $\vec{d_i}$, whose magnitude must satisfy $\| \vec{d_i} \| \leq D_{max}$, where $D_{max}$is determined by the spatial dimensions of the interior scene, preventing objects from exceeding reasonable spatial ranges and causing layout disorder. Finally, a second-order smoothness regularization is constructed to impose smoothness constraints on displacement vectors of adjacent pixels, calculated as:

$$L_{smooth}=\sum_{i=1}^{H\text{-}1}\sum_{j=1}^{W\text{-}1}\left((\vec{d}_{i,j}\text{-}\vec{d}_{i+1,j})^2+(\vec{d}_{i,j}\text{-}\vec{d}_{i,j+1})^2\right) \tag{5}$$

where, H and W are the image height and width, respectively. This constraint prevents edge tearing and local deformation of objects, ensuring smooth and coherent contours after displacement.

This module further designs a prior coupling constraint to achieve deep interaction with the composition prior loss in Chapter 2, making displacement decisions directly adapt to professional interior composition rules. The prediction process of the sparse displacement field is fully supervised by the composition prior loss L_{prior}, guiding displacement targets to optimal positions aligned with perspective, visual balance, and reasonable white space, ensuring that every displacement complies with the perspective logic and aesthetic requirements of interior space.

This coupling mechanism overcomes the limitations of traditional displacement prediction that relies only on low-level image features and lacks professional composition rule constraints, making the sparse displacement field both precise in displacement guidance and highly compatible with interior composition prior, laying a solid foundation for high-quality output of the subsequent content-aware generative repair network.

3.3 Content-aware generative repair network

To address three core problems after object relocation, including background holes, mismatch of lighting and shadows in new positions, and chaotic occlusion relationships, this module innovatively proposes a two-stage progressive generative repair architecture, breaking through the limitations of traditional repair algorithms that perform simple pixel filling, achieving high-quality fusion repair of background and objects, and ensuring visual coherence and authenticity of optimized images. This architecture is based on the displacement results output by the sparse displacement field prediction network, completing background repair and object detail generation in stages, and achieving deep adaptation of content repair and composition optimization through collaborative control, completely solving the problems of texture disorder and obvious synthesis traces in traditional generative repair.

The first stage is background boundary repair, focusing on the blank regions generated at the original positions after object relocation, using a Partial Convolution hole-filling network for precise repair. The core innovation is the scene-adaptive regulation during the repair process. The repair network generates pixels only for the blank regions, fully utilizing the surrounding background texture features, perspective structure, and light-shadow gradient information. Through feature extraction and texture mapping, the repaired walls, floors, ceilings, and other background regions maintain spatial continuity and perspective consistency, naturally connecting with the original background. To enhance repair accuracy, a background consistency loss is introduced: $L_{bg\_consist} = SSIM(I_{bg\_repair}, I_{bg\_ori})$, where $I_{bg_repair}$is the repaired background region, and I_{bg_ori} is the original background region. Structural similarity constraints ensure that the repaired textures are consistent with the original background, avoiding common problems in traditional repair algorithms such as texture breaks and perspective deviation.

The second stage is object lighting-shadow consistency generation and occlusion relation adaptive inference, further improving repair quality and composition rationality. In object lighting-shadow consistency generation, a high-resolution conditional generative adversarial network is innovatively used. Based on the relocated object contours, original image content, and background semantics as joint conditions, object details are generated at new target positions to completely match the surrounding environment in terms of illumination, tone, texture, and perspective. Through feature alignment and light-shadow calibration, synthesis traces after object relocation are completely eliminated. To achieve reasonable distribution of occlusion relationships, a spatial attention occlusion modeling mechanism is introduced. According to object semantic category, spatial depth relationship, and overlapping regions after displacement, the network automatically learns and assigns front-back occlusion weights, defining spatial attention occlusion weights: $w_{occ} = softmax(A(x, y))$, where $A(x, y)$ is the spatial attention map representing the occlusion priority of each object region. The priority is determined by object semantic importance and spatial depth, ensuring clear hierarchy relationships after multi-object relocation and avoiding occlusion confusion. Finally, this achieves collaborative unification of composition optimization and content repair, outputting interior design images with both rationality and authenticity.

3.4 Overall joint optimization objective

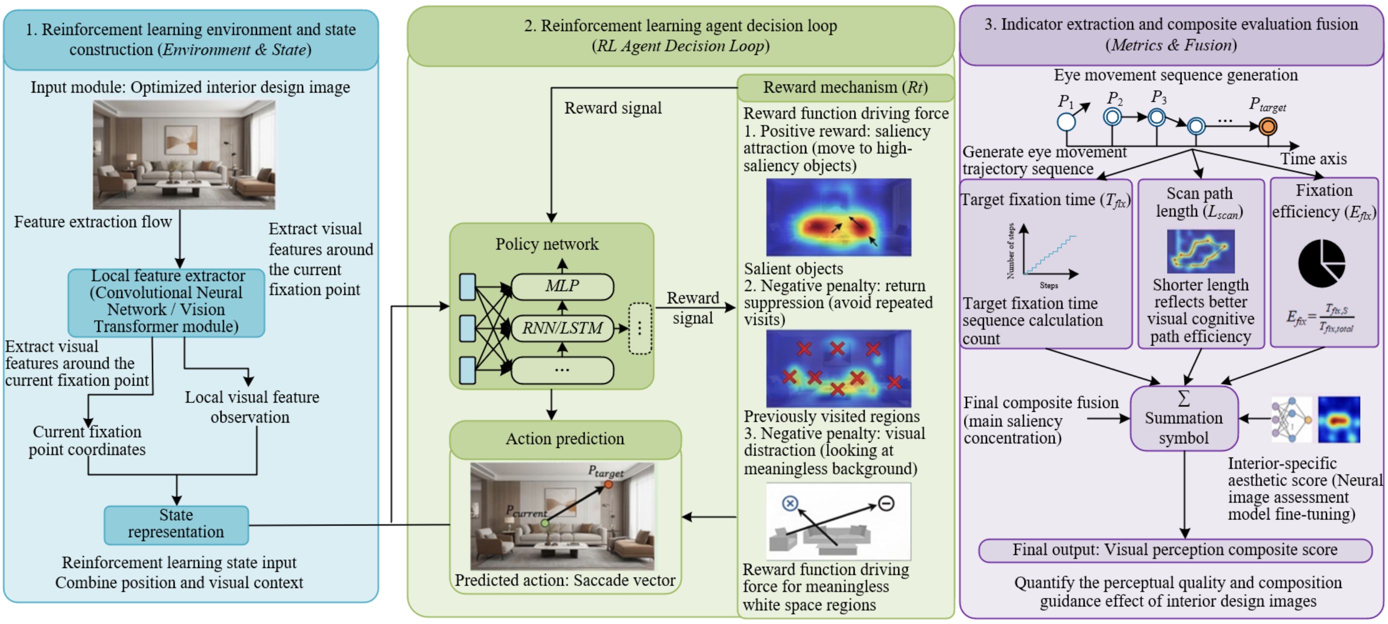
To achieve end-to-end efficient training and optimal performance of the composition optimization generative network, this paper innovatively designs a multi-task weighted fusion overall joint optimization objective. The core breakthrough is directly embedding the interior composition prior loss into the generative network loss function, breaking the limitation in traditional methods where composition aesthetic rules and generative image quality are disconnected, achieving collaborative regulation and joint optimization of both. The total joint optimization loss is calculated as: $L_{total} = \lambda_{adv}L_{adv} + \lambda_{perc}L_{LPIPS} + \lambda_{prior}L_{prior}$, where each loss term has a clear role and collaborates to ensure the composition rationality and visual authenticity of the optimized result. The adversarial loss constrains the authenticity of generated images, making the optimized images highly consistent with real interior design images in visual features and avoiding generation traces; the perceptual loss preserves the semantic information and detail fidelity of the image, preventing detail distortion and texture breaks during object relocation and background repair; the interior composition prior loss serves as the core constraint, directly converting professional interior composition rules into gradient signals, forcing the generative network to output compositions that satisfy perspective consistency, visual balance, and white space aesthetics, ensuring that optimization results conform to professional interior design requirements.

The weighting coefficients of each loss are iteratively optimized through ablation experiments to balance the optimization priority of each task, ensuring both the visual quality of generated images and the constraint effect of professional interior composition rules, forming an integrated whole. This joint optimization objective allows the generative network to be supervised simultaneously by generation quality and composition aesthetics during training, and during gradient backpropagation, displacement prediction accuracy, content repair quality, and composition rationality can be optimized synchronously. This completely solves the core problem in traditional methods where generation quality and composition rules are disconnected, and optimization results do not meet professional design requirements, ensuring that the final output interior design images possess both high-quality visual coherence and authenticity while strictly following professional interior composition rules, achieving dual optimization of aesthetic value and generative quality.

4. COGNITIVE SCIENCE-BASED VISUAL PERCEPTION SIMULATION EVALUATION SYSTEM

The core innovation of this chapter is to construct a reference-free, cognition-driven, multi-dimensional coupled visual perception simulation evaluation system, breaking through the limitations of traditional manual subjective evaluation and single-objective metrics, achieving automated evaluation highly aligned with human visual perception. Figure 4 shows the visual perception simulation evaluation and eye-tracking comparison visualization. First, a comparative evaluation based on visual saliency heatmaps is conducted. A pre-trained visual saliency model is used to extract the saliency distribution of the original and optimized images. Three types of core quantitative indicators are innovatively designed to achieve precise characterization of attention guidance effects. The main saliency concentration is used to measure the proportion of saliency within the design subject region, calculated as:

$$C_{sal}=\frac{\sum_{x\in S}M_{opt}(x)}{\sum_{x\in I}M_{opt}(x)} \tag{6}$$



1. Reinforcement learning environment and state construction (Environment & State)
Input module: Optimized interior design image
Feature extraction flow
Extract visual features around the current fixation point
Local feature extractor (Convolutional Neural Network / Vision Transformer module)
Extract visual features around the current fixation point
Local visual feature observation
Current fixation point coordinates
State representation
Reinforcement learning state input Combine position and visual context
2. Reinforcement learning agent decision loop (RL Agent Decision Loop)
Reward signal
Policy network
MLP
RNN/LSTM
...
Reward signal
Action prediction
P_{target}
$P_{current}$
Predicted action: Saccade vector
Reward mechanism (Rt)
Reward function driving force
1. Positive reward: saliency attraction (move to high-saliency objects)
Salient objects
2. Negative penalty: return suppression (avoid repeated visits)
Previously visited regions
3. Negative penalty: visual distraction (looking at meaningless background)
Reward function driving force for meaningless white space regions
3. Indicator extraction and composite evaluation fusion (Metrics & Fusion)
Eye movement sequence generation
P_1
P_2
P_3
...
P_{target}
Generate eye movement trajectory sequence
Time axis
Target fixation time (T_{fix})
Number of steps
Steps
Target fixation time sequence calculation count
Scan path length (L_{scan})
Shorter length reflects better visual cognitive path efficiency
Fixation efficiency (E_{fix})
$E_{fix}=\frac{T_{fix,S}}{T_{fix,total}}$
Final composite fusion (main saliency concentration)
Σ Summation symbol
Interior-specific aesthetic score (Neural image assessment model fine-tuning)
Final output: Visual perception composite score
Quantify the perceptual quality and composition guidance effect of interior design images

Figure 4. Visual perception simulation evaluation and eye-tracking comparative visualization analysis

where, S is the design subject region such as sofas, beds, and core furniture, and $M_{opt}(x)$ is the saliency value of the optimized image at pixel x. This metric directly reflects the composition's ability to focus on core visual elements. Saliency distribution similarity quantifies distribution differences by calculating the structural similarity between pre- and post-optimization saliency heatmaps: $S_{sim} = SSIM(M_{ori}, M_{opt})$, where M_{ori} and M_{opt} are the saliency heatmaps of the original and optimized images, respectively. Interference suppression rate characterizes the decrease of saliency in background redundant regions, defined as:

$$R_{inh}=1-\frac{\sum_{x\in B}M_{opt}(x)}{\sum_{x\in B}M_{ori}(x)} \quad (7)$$

where, B is the background redundant region. A higher value indicates more significant suppression of visual interference by the composition.

To further simulate the dynamic process of human visual cognition, this chapter innovatively introduces a virtual eye-movement agent based on reinforcement learning, applied for the first time to the evaluation of interior composition optimization, achieving digital replication of real human free-viewing behavior. This agent adopts a sequential decision reinforcement learning architecture, using local image features as observation input, and outputs the coordinates of the next fixation point through a policy network. The agent is pre-trained on a large-scale interior image dataset to ensure behavior closely approximates real users' eye movement patterns. Based on the agent simulation results, three types of cognitive efficiency indicators are designed: Target fixation time T_{fix} is the number of steps required for the agent to first fixate on the design subject; fewer steps indicate higher attention guidance efficiency. Scan path length L_{scan} is the total scanning path length before reaching the subject; shorter length reflects better path efficiency of visual cognition. Fixation efficiency E_{fix} is the proportion of total fixation time spent on the subject region, calculated as:

$$E_{fix}=\frac{T_{fix,S}}{T_{fix,total}} \quad (8)$$

where, $T_{fix,S}$ is the fixation duration on the subject region, and $T_{fix,total}$ is the total fixation duration. This indicator directly reflects the rationality of visual attention allocation in the composition.

To address the insufficient adaptability of general aesthetic scoring models to interior design rules, this chapter innovatively constructs a reference-free aesthetic scoring network dedicated to interior scenes. Based on the Neural Image Assessment (NIMA) architecture, it is fine-tuned using a self-built interior composition quality-labeled dataset to achieve precise quantification of interior image aesthetics. The network takes an interior design image as a single input and outputs a continuous aesthetic score in the range of 0 to 5, with higher scores indicating better aesthetic quality of the composition. This score can directly quantify the aesthetic improvement brought by composition optimization. Compared with general aesthetic models, the fine-tuned network fully incorporates professional design rules such as interior spatial perspective, visual balance, and white space ratio, effectively eliminating adaptation bias of general models to interior scenes and providing a reliable objective basis for aesthetic quality evaluation.

To achieve unified quantification of multi-dimensional perception indicators, this chapter designs a comprehensive perception evaluation score fusion mechanism. Visual saliency heatmap comparison metrics, virtual eye-movement agent metrics, and interior-specific aesthetic scores are normalized and weighted to obtain the final visual perception composite score. First, the three types of indicators are min-max normalized to eliminate dimensional differences, calculated as:

$$x_{norm}=\frac{x-x_{min}}{x_{max}-x_{min}} \quad (9)$$

where, x is the original indicator value, and $x_{min}$and $x_{max}$are the minimum and maximum values of the indicator. Subsequently, weighted fusion is used to obtain the composite score: $S_{comp} = \omega_1 \cdot C_{sal,norm} + \omega_2 \cdot E_{fix,norm} + \omega_3 \cdot S_{score}$, where $\omega_1, \omega_2, \omega_3$ are the weight coefficients of the three types of indicators, optimized through ablation experiments; $C_{sal,norm}$ and $E_{fix,norm}$ are the normalized main saliency concentration and fixation efficiency; S_{score} is the interior-specific aesthetic score. This composite score achieves a comprehensive and objective quantitative evaluation of optimization results from three dimensions: attention guidance, cognitive efficiency, and aesthetic quality, providing a unified and comparable core metric for performance verification of composition optimization algorithms.

5. EXPERIMENTS AND RESULTS ANALYSIS

5.1 Dataset construction

The experimental dataset is composed of public datasets and a self-built 3D interior rendering dataset, ensuring data diversity and annotation accuracy, providing reliable support for experimental validation. Public datasets include LSUN Bedrooms and SceneNet RGB-D, containing images of various interior styles, layouts, and lighting conditions, used for model basic training and generalization verification. The core innovation is the self-built 3D interior rendering dataset, which is generated through 3D modeling tools, containing controllable spatial layout, perspective angles, furniture positions, and lighting intensity. Each sample provides precise composition quality labels, semantic segmentation annotations, and perspective parameter annotations, effectively addressing issues of missing annotations and insufficient controllable variables in real datasets, providing a standardized verification basis for ablation experiments and quantitative evaluation. The overall dataset scale is 12,000 images, with 8,000 for training, 2,000 for validation, and 2,000 for testing, covering three mainstream interior scenes: living room, bedroom, and study, ensuring representativeness and generality of experimental results.

5.2 Experimental setup

The experiments are implemented based on the PyTorch framework. The hardware environment includes NVIDIA RTX 4090 GPU and 128GB memory, and the software environment is Python 3.8 and CUDA 11.8. Training adopts a strategy combining module-wise pretraining and end-to-end joint fine-tuning: first, the semantic segmentation module and perspective detection module are pretrained to ensure precision in semantic parsing and perspective recognition; then, the sparse displacement field prediction network and content-aware generative inpainting network are pretrained; finally, all modules are jointly fine-tuned to optimize overall performance. Comparison methods include traditional composition optimization methods and mainstream deep learning methods, including saliency-based automatic cropping, SeamCarving image retargeting, and Grid Anchor based Image Cropping (GAIC) deep learning composition optimization methods. All comparison methods use the official optimal parameter settings. Evaluation metrics include subjective and objective metrics. Subjective metrics are aesthetic scores (1~5) given by 10 professional interior designers. Objective metrics are the multi-dimensional perception indicators proposed in this paper: main saliency concentration, saliency distribution similarity, interference suppression rate, target fixation time, scan path length, fixation efficiency, aesthetic score, and comprehensive perception score, with the addition of traditional image quality metrics Peak Signal-to-Noise Ratio (PSNR) and Structural Similarity Index (SSIM) to ensure evaluation comprehensiveness.

5.3 Ablation experiments

The ablation experiments aim to verify the necessity of each innovative module proposed in this paper. By removing core innovative modules one by one, model performance changes are compared. Experimental results are shown in Table 1. Four ablation comparison groups are set: base model (removing all innovative modules), ablation of composition prior loss, ablation of sparse displacement field, ablation of two-stage inpainting. All comparison groups maintain other parameters unchanged to ensure fairness.

Table 1. Ablation experiment results comparison

Model Configuration	Main Saliency Concentration	Saliency Distribution Similarity	Interference Suppression Rate	Target Fixation Time (steps)	Scan Path Length (pixels)	Fixation Efficiency	Aesthetic Score	Comprehensive Perception Score	Peak Signal-to-Noise Ratio (dB)	Structural Similarity Index
Base Model	0.521	0.683	0.412	8.7	1286	0.453	2.87	0.512	28.35	0.821
Ablation Composition Prior Loss	0.589	0.715	0.487	7.9	1154	0.512	3.21	0.587	29.12	0.843
Ablation Sparse Displacement Field	0.613	0.732	0.513	7.5	1089	0.538	3.35	0.615	29.57	0.856
Ablation Two-Stage Inpainting	0.658	0.769	0.578	6.8	976	0.596	3.62	0.678	30.24	0.879
Proposed Method	0.786	0.857	0.724	4.3	652	0.765	4.28	0.823	32.68	0.917

The ablation experiment results indicate that each innovative module significantly improves model performance. After removing the composition prior loss, main saliency concentration and aesthetic score decrease by 0.197 and 0.97, respectively, indicating that this module effectively constrains composition professionalism, ensuring optimization results comply with interior design aesthetic guidelines. After removing the sparse displacement field, target fixation time increases by 2.2 steps and scan path length increases by 326 pixels, demonstrating that the sparse displacement field enables precise relocation of movable objects, improving visual guidance efficiency. After removing the two-stage inpainting, PSNR and SSIM decrease by 2.44 dB and 0.038, respectively, showing that this module effectively improves image inpainting quality and reduces synthetic artifacts. The proposed method, integrating all innovative modules, achieves optimal values for all metrics, fully validating the necessity and collaborative effect of each innovative module.

5.4 Quantitative result analysis

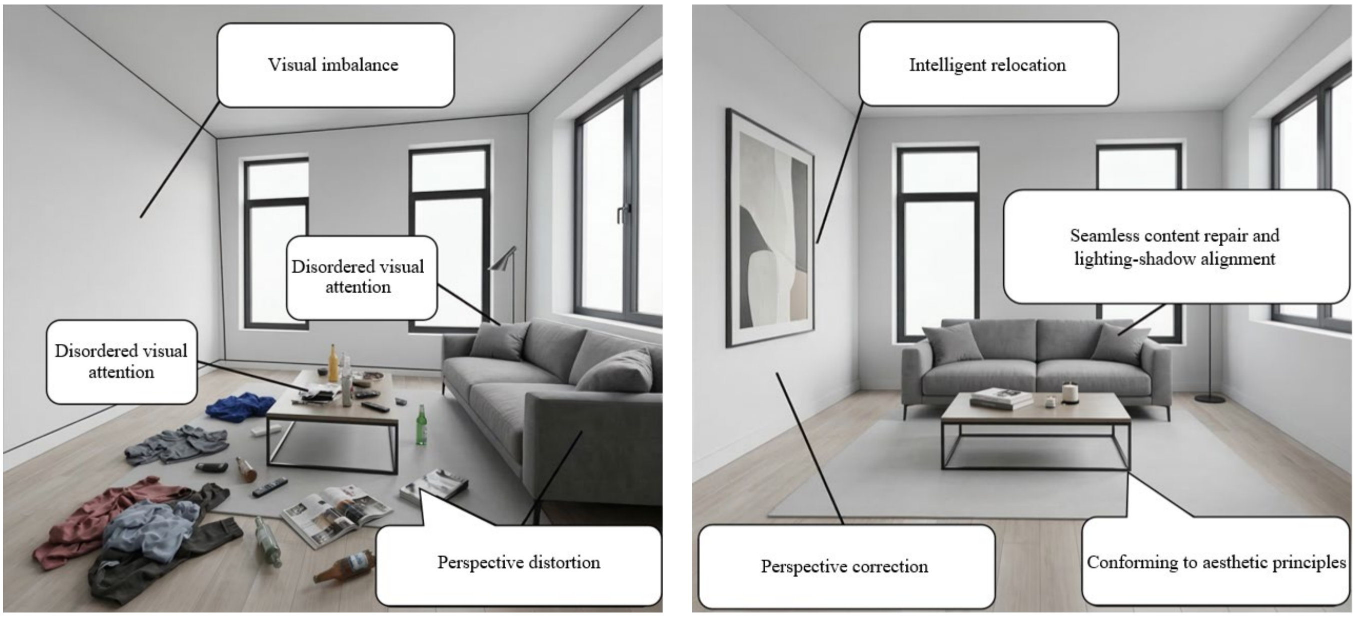
The quantitative experiments compare the performance of the proposed method with various comparison methods to verify the superiority of the proposed method. Experimental results are shown in Table 2. All methods are validated on the test set. For the metric values, higher values of main saliency concentration, saliency distribution similarity, interference suppression rate, fixation efficiency, aesthetic score, comprehensive perception score, PSNR, and SSIM indicate better performance, while lower values of target fixation time and scan path length indicate better performance.

Table 2. Quantitative results comparison of different methods

Method	Main Saliency Concentration	Saliency Distribution Similarity	Interference Suppression Rate	Target Fixation Time (steps)	Scan Path Length (pixels)	Fixation Efficiency	Aesthetic Score	Comprehensive Perception Score	Peak Signal-to-Noise Ratio (dB)	Structural Similarity Index
Saliency-based Automatic Cropping	0.532	0.678	0.405	9.2	1352	0.447	2.79	0.503	27.89	0.812
SeamCarving	0.556	0.692	0.438	8.5	1278	0.472	2.93	0.531	28.15	0.825
Grid Anchor based Image Cropping	0.643	0.758	0.562	6.9	987	0.583	3.57	0.665	30.11	0.872
Proposed Method	0.786	0.857	0.724	4.3	652	0.765	4.28	0.823	32.68	0.917

The quantitative result analysis shows that the proposed method significantly outperforms all comparison methods in all evaluation metrics. Compared with the best comparison method GAIC, the proposed method improves main saliency concentration by 0.143, interference suppression rate by 0.162, reduces target fixation time by 2.6 steps, shortens scan path length by 335 pixels, improves aesthetic score by 0.71, and improves comprehensive perception score by 0.158. PSNR and SSIM are improved by 2.57 dB and 0.045, respectively. The improvement in main saliency concentration and fixation efficiency indicates that the proposed method effectively guides visual attention to focus on design subjects; the reduction in target fixation time and scan path length validates the advantage of the proposed method in enhancing visual cognitive efficiency; the leading aesthetic score and comprehensive perception score demonstrate that the proposed method significantly improves the compositional aesthetic quality of interior design images. The above results fully verify the superiority of the proposed method in composition optimization and perception evaluation. Its performance improvement mainly comes from the collaborative effect of the composition prior module, sparse displacement field prediction, and two-stage generative inpainting.

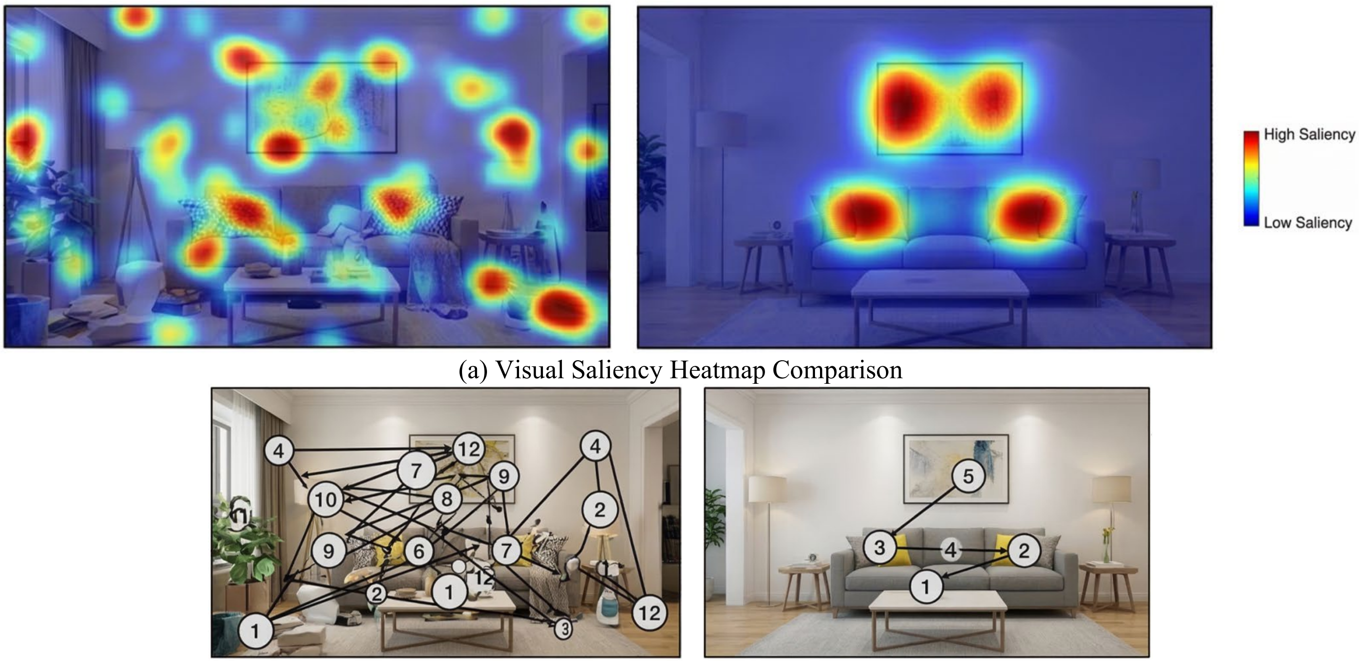
5.5 Qualitative result analysis


Visual imbalance
Disordered visual attention
Disordered visual attention
Perspective distortion
Intelligent relocation
Seamless content repair and lighting-shadow alignment
Perspective correction
Conforming to aesthetic principles

(a) Original Input Image

(b) Optimized Image by Proposed Method

Figure 5. Comparison of spatial layout and aesthetic features before and after automatic composition optimization of interior design images



(a) Visual Saliency Heatmap Comparison

(b) Virtual Eye Movement Agent Scan Path Comparison

Figure 6. Multi-dimensional simulation evaluation of visual saliency and virtual eye movement cognitive trajectories for interior composition

To qualitatively verify the practical effectiveness of the proposed image composition automatic optimization algorithm in correcting spatial layout defects in interior scenes and reconstructing visual aesthetic features, relevant experiments were conducted. The comparative analysis in Figure 5 shows that the original input images have significant spatial perspective distortion and visual center imbalance, with severe foreground visual interference, which violates professional aesthetic principles of architectural interior design. After processing by the proposed algorithm, core furniture is intelligently relocated to the visual center area conforming to aesthetic proportions, achieving high-fidelity content repair and light-shadow consistency alignment. Meanwhile, the global perspective relationships are precisely corrected, effectively eliminating the original clutter and structural distortion. This result directly demonstrates that the proposed generative optimization network can effectively transform abstract interior composition prior knowledge into visual spatial reconstruction capability, providing reliable technical support for automated output of standardized and high-aesthetic-quality design visualization schemes.

To quantitatively verify from the objective physical dimension of cognitive science the effect of spatial composition adjustment on guiding human visual attention and reducing cognitive load, further experiments were conducted. Visual saliency heatmap data in Figure 6 show that high-saliency regions in the pre-optimization image are severely dispersed, causing extreme attention diffusion, while the post-optimization heatmap demonstrates that visual focus is accurately and highly concentrated on core design subject areas, effectively suppressing background environmental interference. Additionally, virtual eye movement agent scan path analysis further reveals that before optimization, unordered composition causes the agent to exhibit high gaze entropy and complex detour paths while searching for the target, resulting in very low cognitive information decoding efficiency. In contrast, eye movement trajectories for post-optimization images show an extremely streamlined and direct visual navigation pattern, where the agent can quickly fixate on core design elements with minimal search steps. Combining the above multi-dimensional data conclusively demonstrates that the proposed optimization algorithm can significantly improve the visual ergonomics performance of interior design images and successfully realize a scientific closed-loop between composition optimization and objective perception evaluation.

5.6 Generalization and robustness experiments

To verify the generalization and robustness of the proposed method, experiments were conducted under different lighting conditions, viewing angles, and interior styles. Four extreme conditions were selected: low light, strong light, side view, and top view, as well as three mainstream interior styles: living room, bedroom, and study room. The performance of the proposed method was compared with the GAIC method. Experimental results are shown in Table 3.

Table 3. Generalization and robustness experimental results

Experimental Condition	Method	Main Saliency Concentration	Fixation Efficiency	Aesthetic Score	Comprehensive Perception Score	Structural Similarity Index
Low Light	Grid Anchor based Image Cropping (GAIC)	0.598	0.532	3.21	0.612	0.835
	Proposed Method	0.721	0.703	3.95	0.768	0.889
Strong Light	GAIC	0.605	0.541	3.27	0.625	0.841
	Proposed Method	0.732	0.715	4.02	0.779	0.896
Side View	GAIC	0.587	0.523	3.15	0.601	0.828
	Proposed Method	0.715	0.698	3.89	0.756	0.882
Top View	GAIC	0.579	0.517	3.11	0.593	0.821
	Proposed Method	0.708	0.689	3.85	0.748	0.877
Bedroom Style	GAIC	0.638	0.576	3.52	0.658	0.867
	Proposed Method	0.779	0.758	4.23	0.817	0.912
Study Room Style	GAIC	0.641	0.579	3.55	0.662	0.869
	Proposed Method	0.782	0.761	4.25	0.820	0.914

The experimental results show that under different extreme conditions and interior styles, all metrics of the proposed method remain stable and significantly outperform the GAIC method. Under low light and strong light conditions, the comprehensive perception scores of the proposed method are 0.768 and 0.779, only decreasing by 0.055 and 0.044 compared with normal conditions; under side and top view angles, the comprehensive perception scores are 0.756 and 0.748, with decreases controlled within 0.075; for different interior styles, the comprehensive perception scores of the proposed method remain above 0.817, demonstrating good stability. In contrast, the performance of the GAIC method decreases significantly under extreme conditions, with the comprehensive perception score dropping to 0.612 under low light and 0.601 under side view, indicating insufficient generalization and robustness. The stability of the proposed method benefits from the precise constraint of interior spatial logic by the composition prior module and the content-adaptive repair capability of the generative network, which can adapt to changes in different scene conditions, verifying the reliability of the proposed method in practical engineering applications.

6. CONCLUSION AND OUTLOOK

This paper addressed the core issues of interior design image composition optimization relying on human experience, subjective evaluation standards, and lack of professional prior constraints, and conducts research on automatic interior composition optimization and visual perception simulation evaluation. Through three core innovations, a complete technical system was constructed and experimentally validated. The paper innovatively proposed a computable modeling method for interior composition prior knowledge, converting abstract design principles such as perspective consistency, visual balance, and whitespace ratio into differentiable loss functions, forming a unified composition prior module to realize quantitative embedding and gradient constraint of professional principles. An end-to-end sparse displacement field guided generative optimization network was designed, which can precisely control movable object relocation via sparse displacement fields, combined with a two-stage generative inpainting architecture to solve issues such as background voids and lighting mismatch, achieving content-preserving intelligent composition adjustment. A cognitive science-driven multi-dimensional no-reference visual perception simulation evaluation system was built, integrating visual saliency analysis, virtual eye movement simulation, and interior-specific aesthetic scoring to achieve automated evaluation highly aligned with human visual perception. Experimental results show that the proposed method significantly outperformed traditional methods and mainstream deep learning methods on all evaluation metrics, effectively improving the composition rationality and visual aesthetics of interior design images, providing reliable technical support for automated optimization of interior design visualization schemes, and has important theoretical research value and engineering application prospects.

Based on the research results of this paper, future research can be further expanded and deepened in three directions. First, extend static image composition optimization to dynamic video interior composition optimization, optimizing displacement prediction and content repair mechanisms to meet inter-frame coherence requirements in interior videos, achieving real-time composition optimization in dynamic scenes. Second, implement adaptive composition optimization based on user personalized preferences, mining different users' aesthetic preferences and design requirements to construct personalized composition prior models, making the optimization results more aligned with user needs. Third, promote deep integration with 3D interior design software, extending 2D image composition optimization to 3D spatial layout adjustment, achieving collaborative linkage between 2D image optimization and 3D model adjustment, further improving the intelligence level and design efficiency of interior design, and promoting the digitalization and intelligence development of the interior design industry.

REFERENCES

[1] Li, Y. (2024). Application of high performance computing based on big data in architectural interior design. Intelligent Decision Technologies, 18(4): 2933-2944. https://doi.org/10.3233/idt-230171

[2] Dai, Y.W., Wang, P. (2025). Architecture support based on IoT and VR technology in architectural interior design. Journal of Information Science and Engineering, 41(5): 1121-1136. https://doi.org/10.6688/JISE.202509_41(5).0004

[3] Celadyn, M. (2020). Integrative design classes for environmental sustainability of interior architectural design. Sustainability, 12(18): 7383. https://doi.org/10.3390/su12187383.

[4] Li, T. (2022). Denoising method of interior design image based on median filtering algorithm. International Journal of Advanced Computer Science and Applications, 13(12): 1021-1029. https://doi.org/10.14569/ijacsa.2022.01312117

[5] Kim, J., Lee, J. (2020). Stochastic detection of interior design styles using a deep-learning model for reference images. Applied Sciences, 10(20): 7299. https://doi.org/10.3390/app10207299

[6] Peng, Y., Hu, Q., Xu, J., KinTak, U., Chen, J. (2025). A novel deep learning zero-watermark method for interior design protection based on image fusion. Mathematics, 13(6): 947. https://doi.org/10.3390/math13060947

[7] Lee, J., Kim, Y., Shin, E., Choo, S., Cha, S.H. (2024). An AI-assisted approach for creating and archiving interior design references using 360-degree panoramic images. Architectural Science Review, 68(6): 506-518. https://doi.org/10.1080/00038628.2024.2445049

[8] Kurz, T.L., Jayasuriya, S., Swisher, K., Mativo, J., Pidaparti, R., Robinson, D.T. (2026). Computer vision versus human vision: Analyzing middle school teachers' construct restructuring following computer vision professional development. Educational Technology Research and Development, 1-21. https://doi.org/10.1007/s11423-026-10594-2

[9] Chatterjee, S., Chaudhuri, R., Vrontis, D., Papadopoulos, T. (2022). Examining the impact of deep learning technology capability on manufacturing firms: Moderating roles of technology turbulence and top management support. Annals of Operations Research, 339(1-2): 163-183. https://doi.org/10.1007/s10479-021-04505-2

[10] Wu, H., Miao, Z., Zhang, Q., Xu, W. (2014). Image composition optimization based on feature match and detail preserved. Optik, 125(16): 4370-4373. https://doi.org/10.1016/j.ijleo.2014.03.020

[11] Li, Z.Y., Wang, H. (2002). Parameter optimization of recognition model of landmark spectrum composition based on genetic algorithm. Journal of Infrared and Millimeter Waves, 21(3): 205-208.

[12] Wang, Y., Ke, Y., Wang, K., Guo, J., Yang, S. (2023). Spatial-invariant convolutional neural network for photographic composition prediction and automatic correction. Journal of Visual Communication and Image Representation, 90: 103751. https://doi.org/10.1016/j.jvcir.2023.103751

[13] Luo, X. (2021). Three-dimensional image quality evaluation and optimization based on convolutional neural network. Traitement du Signal, 38(4): 1041-1049. https://doi.org/10.18280/ts.380414

[14] Zhang, F., Wang, M., Hu, S. (2013). Aesthetic image enhancement by dependence-aware object recomposition. IEEE Transactions on Multimedia, 15(7): 1480-1490. https://doi.org/10.1109/tmm.2013.2268051

[15] Jin, Y., Wu, Q., Liu, L. (2012). Aesthetic photo composition by optimal crop-and-warp. Computers & Graphics, 36(8): 955-965. https://doi.org/10.1016/j.cag.2012.07.007

[16] Liu, S., Zhao, C., Gao, Y., Wang, J., Tang, M. (2019). Adversarial image generation by combining content and style. IET Image Processing, 13(14): 2716-2723. https://doi.org/10.1049/iet-ipr.2019.0103

[17] Chu, K., Shang, Y., Zhang, L., Yuan, H. (2026). Content style decoupling for multi style image generation using latent diffusion architecture. Scientific Reports, 16(1): 6642. https://doi.org/10.1038/s41598-026-36407-3

[18] Hosonuma, E., Yamazaki, T., Miyoshi, T., Taya, A.,

Nishiyama, Y., Sezaki, K. (2025). Image generative semantic communication with multi-modal similarity estimation for resource-limited networks. IEICE Transactions on Communications, E108-B(3): 260-273. https://doi.org/10.23919/transcom.2024ebp3056

[19] Shakir, H.S., Nagao, M. (1995). Hierarchy-based networked organization, modeling, and prototyping of semantic, statistic, and numeric image-information. IEICE Transactions on Information and Systems, 78(8): 1003-1020.

[20] Xu, Y., Chen, C.Y., Hu, X.J., Yu, H.Y., Tian, Y. (2023). Visual perception evaluation method of stereo images based on CNN-SVR. Laser & Optoelectronics Progress, 60(8): 0811027-1-0811027-9. https://doi.org/10.3788/LOP230893

[21] Wang, Y., Durmus, D. (2022). Image quality metrics, personality traits, and subjective evaluation of indoor environment images. Buildings, 12(12): 2086. https://doi.org/10.3390/buildings12122086

[22] Oyekoya, O.K., Stentiford, F.W.M. (2006). Eye tracking —A new interface for visual exploration. BT Technology Journal, 24(3): 57-66. https://doi.org/10.1007/s10550-006-0076-z

[23] Jia, L., Tu, Y., Wang, L., Zhong, X., Wang, Y. (2018). Study of image quality using event-related potentials and eye tracking measurement. Journal of the Society for Information Display, 26(6): 339-351. https://doi.org/10.1002/jsid.653