

Article

Exploration of Early-Stage Floor Plan Design for University Research Buildings Based on a Conditional Diffusion Model

Zimo Chen ¹, Yufei Liu ^{1,2,*}, Zhenling Wu ^{2,3} and Bing Li ^{2,3}

¹ College of Civil Engineering and Architecture, Zhejiang University, Hangzhou 310058, China; 22512072@zju.edu.cn

² Center for Balance Architecture of Zhejiang University, Hangzhou 310028, China; wuzhenling@zju.edu.cn (Z.W.); 0088162@zju.edu.cn (B.L.)

³ The Architectural Design & Research Institute of Zhejiang University Co., Ltd., Hangzhou 310028, China

* Correspondence: 12412019@zju.edu.cn

Abstract

This research proposes a conditional diffusion-based workflow for early-stage floor plan design in university research buildings, addressing complex functional organization, strict boundary constraints, and quantitative area control. The method performs denoising directly in two-dimensional grid space and coordinates building outlines and functional area proportions through dual-condition injection using boundary masks and functional area matrices. A two-stage generation mechanism first constructs horizontal circulation and then generates the complete layout, while a statistic-network-guided explicit constraint improves global area consistency. Based on 600 standard-floor samples and an independent test set of 10 real projects, the method is evaluated through model comparison, ablation, and double-blind experiments. The results show that the proposed model achieves the best overall performance, with an FID of 50.3, a building boundary IoU of 99.9%, and horizontal circulation connectivity of 89.8%. The ablation results confirm that the two-stage mechanism and explicit statistical constraint substantially improve generation success and reduce area error. The expert evaluation indicates that AI-generated floor plans approach real cases in functional spatial form and design inspiration, although spatial organization rationality still requires improvement. The generated layouts can be converted into layered DXF files, supporting subsequent editing and human–AI collaborative design.

Keywords: conditional diffusion model; university research buildings; early-stage floor plan design; architectural floor plan generation; human–AI collaborative design



Academic Editor: Xuan Ma

Received: 13 May 2026

Revised: 3 June 2026

Accepted: 9 June 2026

Published: 11 June 2026

Copyright: © 2026 by the authors.

Licensee MDPI, Basel, Switzerland.

This article is an open access article

distributed under the terms and

conditions of the [Creative Commons](https://creativecommons.org/licenses/by/4.0/)

[Attribution \(CC BY\)](https://creativecommons.org/licenses/by/4.0/) license.

1. Introduction

Architectural creation is a complex decision-making activity that integrates multi-dimensional factors. In essence, it is a process in which architects seek a dynamic balance between perceptual artistic expression and rational technical logic [1]. Unlike standardized engineering problems, architectural design is characterized by significant nonlinearity and multiple possible solutions. A design scheme is not derived from a single formula but is gradually formed through continuous generation, evaluation, and feedback. However, in current industry practice, traditional architectural creation is facing multiple challenges. On the one hand, architects are naturally limited by cognitive capacity and available effort, making it difficult to generate a large number of diverse and high-quality design schemes within a short period. As building functions become increasingly integrated and project

requirements become more complex, the amount of information that designers need to process continues to increase. When facing complex spatial arrangement problems, designers often rely on experience to develop only a small number of schemes, which may lead to the omission of potential optimal solutions [2]. On the other hand, the architectural creation process itself involves a high degree of uncertainty. Ambiguous design logic in the early stage, changes in design briefs, and shifting client requirements can all lead to repeated revisions [3]. Traditional linear design workflows lack a generation mechanism capable of rapidly responding to changes and dynamically adjusting design logic. As a result, scheme optimization often depends on individual intuition and repeated trial and error, making it difficult to meet the practical demands of high-efficiency and highly variable design tasks.

A review of the development of architectural creation tools shows that each technological iteration has promoted updates in design efficiency and methodological paradigms [4]. In the era of manual drafting, architects relied on rulers and compasses for drawing production and design refinement, resulting in long design cycles and high revision costs [5]. At the end of the twentieth century, the popularization of computer-aided design (CAD) enabled the transition from paper-based drawing to digital drafting, greatly improving drawing accuracy and revision efficiency [6]. In the twenty-first century, parametric design represented by building information modeling (BIM) and visual programming languages such as Grasshopper gradually emerged. These technologies enabled architects to control complex geometric forms through logical rules, constraints, and algorithms, marking the formal entry of architectural design into the digital stage [7]. However, parametric design still essentially relies on manually predefined rigid rules, and the flexibility and adaptability of its generated results are constrained by the algorithmic structure itself. In particular, when facing building types with complex functional organization and frequently changing requirements, the limitations of traditional parametric methods in design support have become increasingly evident [8].

In recent years, the rapid development of artificial intelligence, especially generative deep learning, has brought new possibilities to architectural design [9]. At present, AI has demonstrated potential across multiple dimensions of architectural design. In terms of visual representation, diffusion-model-based image generation tools can already produce high-quality architectural renderings and conceptual sketches from text prompts [10]. In sustainable design, AI can assist in predicting building energy consumption, daylighting, and other performance indicators, while also supporting multi-objective optimization [11]. In the field of floor plan generation, researchers have also begun to explore the use of artificial intelligence to assist in spatial figure-ground division [12]. However, existing floor plan generation studies are mostly limited to residential buildings and other building types with relatively simple spatial logic. For university research buildings, which have highly complex internal functions, existing AI models often lack the ability to accurately constrain physical boundaries and have difficulty transforming functional area proportions into floor plan layouts. Second, most current generative models still focus primarily on learning floor plan geometry and emphasize visual resemblance in the generated results, while insufficient attention is paid to functional arrangement logic and spatial organization relationships [13]. Finally, the condition injection strategies of existing mainstream models are mostly limited to text descriptions or sketch guidance. They lack an effective mechanism for transforming quantitative functional area indicators from design briefs into floor plan layout control conditions. Therefore, these models cannot adequately meet the dual requirements of area proportion control and physical boundary constraints in the early-stage floor plan design for research buildings.

To address these issues, this study proposes a conditional diffusion generation workflow for the early-stage floor plan design of university research buildings. To preserve the

spatial precision of architectural floor plans as much as possible, this study abandons the latent-space generation strategy in latent diffusion models, which compresses images into a lower-dimensional latent space. Instead, diffusion denoising is performed directly in a two-dimensional grid space. On this basis, this study constructs a dual-condition injection mechanism. On the one hand, a building footprint mask that defines the buildable range is used as the boundary condition to provide physical outline constraints. On the other hand, a functional area matrix reflecting design brief requirements is encoded and injected into the cross-attention layers of the U-Net as a hidden state, thereby enabling explicit control over functional area proportions.

Based on this framework, the main contributions of this study are as follows:

1. This study constructs a non-latent-space conditional diffusion model architecture, enabling precise constraints on the spatial topology and irregular boundaries of architectural floor plans.
2. This study proposes a condition control strategy that maps macro-level functional area indicators to micro-level pixel generation guidance, thereby establishing a connection between quantitative indicators and floor plan generation.
3. For university research buildings with complex functions, this study proposes an AI-assisted design method that combines controllability and design inspiration, providing a new technical pathway for early-stage floor plan design for university research buildings.

2. Background

2.1. Early-Stage Floor Plan Design for University Research Buildings

University research buildings are composite spaces that accommodate teaching experiments, academic seminars, and interdisciplinary collaboration. Their functional organization has moved beyond the boundaries of traditional single-discipline buildings and is characterized by a high degree of functional complexity, close spatial connections, and diversified spatial types [14]. As scientific research gradually shifts from closed experimental modes toward open collaboration, contemporary research buildings not only support experimental operations but also serve as important places for promoting knowledge exchange and interdisciplinary collaboration [15].

In terms of functional composition, university research buildings in China usually integrate multiple functional units, including research laboratory space, laboratory support space, research office space, research support space, open communication space, and public service space [16]. These spaces must maintain convenient connections while avoiding circulation conflicts and mutual interference. Therefore, the early-stage floor plan design of university research buildings usually faces two core issues. On the one hand, the spatial topological relationships are complex, and different functional zones are highly coupled; recent research on higher education spaces also shows that spatial performance emerges from the interaction between configuration, actual behavior, and user perception [17]. This significantly increases the possible solution space and decision-making difficulty in the early design stage. On the other hand, design briefs usually impose strict requirements on the area proportions of different functional rooms. Architects must devote substantial effort to calculating area indicators in the early design stage to ensure that the design results meet use requirements and construction standards [18].

In response to these complex requirements, existing design methods show clear limitations in practice. In traditional design workflows, architects usually rely on personal experience and intuitive judgment to develop only a small number of design directions within a limited time. In particular, when facing complex laboratory circulation and strict functional area constraints in research buildings, designers often perform local optimization

within existing cognitive frameworks and find it difficult to break through habitual design patterns. Although this experience-driven design mode remains effective for small-scale or single-function projects, its efficiency and exploratory breadth are clearly limited in research buildings with highly composite functions.

Parametric design tools represented by Grasshopper and BIM have improved the logical expression and constraint control capabilities of architectural design to a certain extent [19]. Architects can use predefined rules, parameter relationships, and algorithmic workflows to generate and adjust complex geometric forms, while dynamically controlling some functional indicators. However, the core advantage of parametric design is mainly concentrated on geometric constraint expression and form generation, and its support for complex functional arrangement and spatial organization logic remains limited [20]. At the same time, such methods depend heavily on manually predefined rigid rules. Once functional requirements or design forms change, the original rule system often needs to be reconstructed, resulting in insufficient model flexibility and adaptability [21]. For university research buildings, which involve diverse functions, complex circulation relationships, and frequently changing requirements, traditional parametric methods still have difficulty efficiently supporting early-stage floor plan design.

Therefore, an auxiliary design tool that can simultaneously respond to complex functional constraints, quantitative area indicators, and spatial organization relationships is urgently needed to support efficient generation and multi-scheme exploration during the early-stage floor plan design stage of university research buildings. Such a tool should not only have strong constraint control capability but also be able to produce diverse feasible floor plan schemes within a short time, thereby providing more effective technical support for subsequent scheme selection and design development by architects.

2.2. Generative Artificial Intelligence-Assisted Architectural Floor Plan Design

In recent years, the development of generative artificial intelligence has provided new technical pathways for architectural floor plan design [22]. Unlike traditional parametric design, which mainly depends on explicit rules, generative models can automatically extract spatial organization features and layout patterns from a large number of samples and generate new design schemes under given conditions. In the field of architectural floor plan design, related studies have mainly explored generative adversarial networks (GANs), graph neural networks, reinforcement learning, graph grammar, and data-driven optimization, gradually promoting the transition from rule-driven design to data-driven generation.

Among existing studies, generative adversarial networks are one of the most widely used technical frameworks. GANs realize the mapping from input conditions to target floor plan layouts through adversarial learning between a generator and a discriminator [23]. According to different generation motivations and technical pathways, GAN applications show diverse characteristics. Some studies focus on layout generation guided by specific criteria. For example, Wang et al. proposed ActFloor-GAN, which uses human activity as guidance and aims to generate floor plans with both geometric rationality and topological rationality [24]. Hu et al. developed Graph2Plan, which uses deep neural networks to model and transform user-input layout graphs into specific architectural floor plans, thereby supporting human-AI collaborative design [25]. Dong et al. proposed EdgeGAN, a framework focused on edge detection and vectorization of floor plans to improve the precision of generated results [26]. In residential design, Upadhyay et al. constructed the CIGMA platform, which can generate 2D layouts according to user constraints and provide 3D views and furniture customization functions [27]. In image-to-image translation tasks, Pix2Pix has become a benchmark model for many studies due to its low training difficulty and favorable generation performance [28]. Nauata et al. proposed House-GAN,

which introduces room relationship graphs as constraints into a generative adversarial network and realizes graph-constrained layout generation for residential floor plans [29]. With further technical development, the text-controlled InstructPix2Pix framework has expanded the interactivity of floor plans by combining GPT-3 and Stable Diffusion to enable image editing based on natural language instructions [30]. In addition, Nauata et al. further proposed House-GAN++, which improves the geometric completeness and editability of floor plan layouts through a generative adversarial layout refinement network, making generative models closer to the design assistance needs of professional architects [31]. Ye et al. proposed MasterplanGAN and used CycleGAN for the intelligent rendering of urban master plans, indicating that generative adversarial networks can be applied not only to floor plan layout generation but also to design drawing representation and visual translation [32]. Overall, GANs provide an important technical foundation for architectural floor plan generation, enabling researchers to rapidly generate visually complete floor plan results under given conditions.

In addition to GANs, other neural-network-based generative methods have also shown diverse development pathways in architectural floor plan design. Researchers have introduced mechanisms such as graph neural networks, reinforcement learning, graph grammar, data-driven optimization, multimodal generation, and large language models. In layout generation, the FLNet framework proposed by Upadhyay et al. guides floor layout design through user constraints and realizes effective control over generated results [33]. In the direction of reinforcement learning, Kakooee et al. [34] modeled space layout design as a Markov decision process and constructed the SpaceLayoutGym environment, enabling agents to automatically explore spatial layout schemes under geometric constraints and topological objectives. This demonstrates the application potential of reinforcement learning in architectural space generation. In formal methods, Wang et al. used graph grammar based on semantic-driven embedding to generate floor plan designs that conform to geometric attributes [35]. More recent graph-based work has also explored bubble-diagram generation by combining graph neural networks and variational encoding, indicating the importance of representing functional adjacency before producing detailed layouts [36]. Data-driven strategies have also received increasing attention. Recent reviews point out that the evaluation of generated floor plans should not rely solely on visual similarity but should also incorporate multiple dimensions, such as spatial organization, functional relationships, usability, and environmental performance [37]. Zeng et al. further proposed a unified residential floor plan generation framework with multimodal inputs, showing that flexible design inputs can improve the practical adaptability of floor plan generation systems [38]. Qiu et al. introduced an LLM-based framework for customized vectorized floor plan design from natural language requirements, reflecting the growing tendency to connect generative layout models with more accessible design interaction interfaces [39]. These studies have expanded the technical boundaries of intelligent architectural floor plan generation from different perspectives and also indicate that floor plan design does not have a single universally effective technical route; instead, it is more suitable to integrate the structural advantages of different models according to specific design tasks.

However, although the above methods have made significant progress in architectural floor plan design, their limitations should not be ignored. First, the training process of GAN-based methods is usually unstable and prone to problems such as mode collapse. When dealing with complex geometric and topological relationships, these methods often produce graphical misalignment, unclear boundaries, or room adhesion [40]. Second, existing studies are mostly concentrated on residential buildings and other building types with relatively simple spatial logic and abundant public datasets. For university research buildings, which involve highly composite functions, complex circulation organization,

and limited data availability, the applicability of existing methods remains limited. Third, although many models can generate floor plan results that visually resemble real cases, their control strategies often rely mainly on text, sketches, boundaries, or adjacency relationships, and their ability to represent and constrain quantitative functional area proportions from design briefs remains insufficient. Therefore, although existing generative artificial intelligence methods provide an important foundation for architectural floor plan design, they still cannot simultaneously satisfy the requirements of functional logic, physical boundaries, and quantitative area control in the early-stage floor plan design of university research buildings.

2.3. Diffusion Model-Assisted Architectural Floor Plan Design

A diffusion model is a generative model inspired by nonequilibrium thermodynamics. Its core idea is to simulate the data generation process through two Markov chains in opposite directions. In the forward diffusion process, Gaussian noise is gradually added to the original data until it degrades into random noise. Then, by learning the corresponding reverse denoising process, the target data are gradually recovered from pure noise. Compared with generative adversarial networks, diffusion models have clear advantages in training stability and generation quality, and they have therefore rapidly become an important technical route in the field of generative artificial intelligence. In practical applications, diffusion models have developed into many variants to meet different needs. Among them, conditional diffusion models are the most widely used. They introduce condition information, such as class labels, text descriptions, or semantic maps, to improve the controllability of generated results. Latent diffusion models (LDMs) compress the diffusion process into latent space to improve efficiency, significantly reducing computational cost, and have been widely used in mainstream text-to-image generation models such as Stable Diffusion. In addition, there are task-specific variants, such as diffusion models for accelerated sampling [41] and multi-view diffusion models for 3D generation [42].

Diffusion models have become a research focus in architectural floor plan design in recent years. Their advantages mainly lie in their ability to generate high-quality and complex-structured images through progressive denoising and to achieve fine control over floor plan layouts under conditional constraints. In graph-conditioned vector floor plan generation, Shabani et al. proposed HouseDiffusion, which represents floor plans as one-dimensional polygon loops and uses a Transformer-based diffusion model for discrete and continuous denoising, achieving a technical breakthrough in directly generating vectorized floor plans [43]. In 3D scene understanding, Huang et al. proposed SceneDiffuser, which integrates scene-aware generation, physics-based optimization, and goal-oriented planning through the diffusion denoising process, demonstrating the potential of diffusion models in multidimensional design tasks [44]. For sparse-data reconstruction, Gueze et al. combined graph neural networks with constrained diffusion models to reconstruct consistent floor plans from sparse views and room connection graphs, effectively improving the robustness of generated results [45]. In accessible multi-occupancy floor plan generation, Zhang and Zhang integrated a transformer-based diffusion model with flexible room-level and global constraints, showing that diffusion models can support more complex building layouts and fine-grained control requirements [46].

With the emergence of large models such as Stable Diffusion, fine-tuning studies based on pretrained models have increased. Zeng et al. fine-tuned LoRA in Stable Diffusion and achieved efficient generation of complex and diverse floor plans [47,48]. Recent text- and image-conditioned systems have further decomposed residential layout generation, detailed floor plan generation, and 3D visualization into connected modules, improving the flexibility of automated layout generation and editing [49]. Furthermore, Wang and

Pajarola proposed DiffPlanner, a transformer-based conditional diffusion model that directly generates vector floor plans without rasterization, reducing information loss between raster and vector representations [50]. This progressive reconstruction mechanism enables diffusion models to show strong fidelity and diversity in image generation, making them an important technical direction in intelligent architectural floor plan design after generative adversarial networks.

However, although diffusion models have demonstrated clear advantages in generation quality and controllability, existing studies still show evident limitations when dealing with complex building types such as university research buildings. First, many current studies still focus mainly on learning floor plan geometry and emphasize the visual resemblance of generated results, such as room boundaries, door and window positions, or overall graphical structure. Their modeling of functional arrangement logic and circulation organization relationships remains weak. Second, existing studies are mostly concentrated on residential buildings and other building types with relatively simple spatial logic and abundant public datasets [50]. By contrast, university research buildings are highly specialized building types. Publicly available cases are limited, and research units of different disciplines and scales vary significantly, making it difficult to form large-scale standardized datasets. This data scarcity substantially limits the applicability of deep learning models. Finally, the condition injection strategies of existing diffusion models are still mostly limited to text descriptions, sketch guidance, or local semantic control. They lack an effective mechanism for directly transforming quantitative functional area indicators from design briefs into floor plan layout control conditions. Therefore, they have difficulty satisfying the dual requirements of area proportion control and physical boundary constraints in the early-stage floor plan design of university research buildings.

In summary, diffusion models provide a more stable, higher-quality, and more condition-controllable technical pathway for architectural floor plan design than traditional generative adversarial networks. However, existing studies still show limitations in functional logic modeling, adaptability to complex building types, and quantitative constraint representation. To address these issues, this study further constructs a conditional diffusion generation framework for the early-stage floor plan design of university research buildings, aiming to achieve coordinated control over building boundaries, spatial organization, and functional area proportions.

3. Materials and Methods

3.1. Data Preprocessing

Data preprocessing converts real university research building floor plans into structured samples for conditional diffusion training. Since ground floors often contain special entrance and public functions, this study focuses on standard floors to improve the consistency of spatial organization patterns. As shown in Figure 1, the workflow includes data annotation, data augmentation, format conversion, image scaling, and feature extraction.

In the data annotation stage, this study follows the functional color-block representation used in existing studies [51]. According to the functional composition of university research buildings, floor plans are classified into eight categories: research laboratory area, laboratory support area, research support area, open communication area, public service area, research office area, horizontal transportation area, and vertical transportation area [52]. Each category is represented by a predefined RGB value in AutoCAD 2020 (version 23.1.47.0; Autodesk, Inc., San Rafael, CA, USA), while the building outline is drawn with lightweight polylines as the boundary condition, as shown in Table 1.

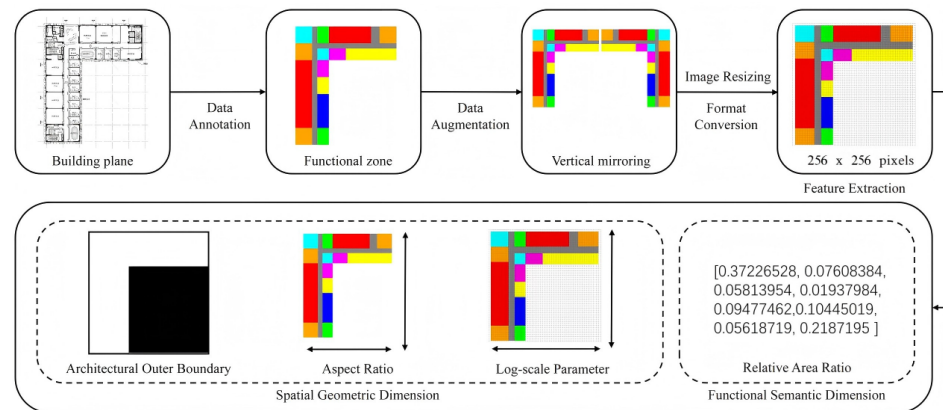


Figure 1. Experimental preprocessing workflow.

Table 1. Functional zoning of research laboratory buildings.

Functional Zone	Functional Space	RGB	Color
Research laboratory area	Specialized laboratories, general laboratories, research studios, etc.	(255, 0, 0)	Red
Laboratory support area	Preparation rooms, precision instrument rooms, cultivation rooms, laboratory animal rooms, greenhouses, darkrooms, shower rooms, disinfection rooms, laboratory equipment rooms, sample and reagent storage rooms, storage rooms, etc.	(0, 0, 255)	Blue
Research support area	Library and information rooms, academic lecture halls, meeting rooms, research exhibition spaces, etc.	(255, 0, 255)	Magenta
Open communication area	Open spaces that promote communication among researchers, such as atriums, rest platforms, lounges, cafes, etc.	(0, 255, 0)	Green
Public service area	Supporting rooms and equipment for water, electricity, gas, oil, refrigeration, air conditioning, communication, fire protection, heating systems, and restrooms, etc.	(0,255,255)	Cyan
Research office area	Research offices, administrative offices, reception rooms, administrative storage rooms, etc.	(255, 255, 0)	Yellow
Horizontal transportation area	Corridors, etc.	(130, 130, 130)	Gray
Vertical transportation area	Stairwells, elevator lobbies, etc.	(255, 165, 0)	Orange

To preserve north–south orientation semantics related to daylighting, ventilation, and functional layout, only vertical mirroring, namely reflection along the horizontal axis, is used for data augmentation. The annotated drawings are then converted into DXF format, parsed as vector geometry, and rasterized into semantic images with a fixed resolution of 256. Instead of direct image stretching, a proportional coordinate transformation is used to map vector vertices to pixel coordinates:

$$P_x = \frac{x - x_{\min}}{W_{\text{real}}} \cdot (\text{Size} - 1)$$

$$P_y = \frac{y_{\max} - y}{H_{\text{real}}} \cdot (\text{Size} - 1)$$

where P_x and P_y denote the pixel coordinates after rasterization; x and y denote the original vector coordinates; $x_{\min}$ and $y_{\max}$ denote the minimum horizontal coordinate and maximum vertical coordinate of the building bounding box, respectively; W_{real} and

H_{real} denote the real width and height of the building bounding box, respectively; and $\text{Size} = 256$ denotes the target raster image size.

Because rasterization normalizes floor plans with different physical scales into the same image size, this study further extracts the aspect ratio and Log-scale Parameter of each building outline as auxiliary geometric conditions. The aspect ratio records the east–west/north–south proportion, and the Log-scale Parameter is calculated from the maximum side length to represent absolute scale in a numerically stable form. Functional area ratios are then calculated by pixel statistics:

$$A_i = \frac{\sum P_i}{\sum P_{\text{foreground}}}$$

$\sum P_i$ denotes the number of pixels belonging to the corresponding category, and $\sum P_{\text{foreground}}$ denotes the total number of pixels of all functional areas. After preprocessing, each sample consists of a semantic layout image, a building boundary mask, geometric descriptors, and an eight-dimensional functional area proportion vector.

3.2. Conditional Diffusion Model

This study adopts a conditional diffusion model to generate university research building floor plans directly in a two-dimensional grid space. Unlike latent diffusion models, which compress images into lower-dimensional latent representations, the proposed model performs denoising directly on rasterized semantic layouts to better preserve spatial precision and boundary constraints. Given a clean layout image x_0 , Gaussian noise is progressively added during the forward process [53]:

$$q(x_t | x_{t-1}) = \mathcal{N}(\sqrt{1 - \beta_t}x_{t-1}, \beta_t \mathbf{I})$$

where β_t denotes the predefined noise schedule. The closed-form expression used during training is:

$$x_t = \sqrt{\alpha_t}x_0 + \sqrt{1 - \alpha_t}\varepsilon, \quad \varepsilon \sim \mathcal{N}(0, \mathbf{I})$$

$$\alpha_t = \prod_{i=1}^t (1 - \beta_i)$$

where α_t is the cumulative noise coefficient. The model learns the reverse denoising process under conditional guidance. To improve numerical stability, this study adopts v-prediction parameterization, in which the network $v_\theta(x_t, t, c)$ predicts the velocity variable:

$$v = \sqrt{\alpha_t}\varepsilon - \sqrt{1 - \alpha_t}x_0$$

The diffusion training objective is:

$$L_{\text{diff}} = \mathbb{E}_{x_0, \varepsilon, t} \|v_\theta(x_t, t, c) - v\|_2^2,$$

where t is a randomly sampled diffusion time step, ε is standard Gaussian noise, and c denotes the condition vector. In this study, $c = [A, W, H, S]$, where A is the functional area proportion vector, W and H are the east–west and north–south spans, and S is the Log-scale Parameter.

The reverse denoising network is a U-Net with cross-attention modules [54]. As shown in Figure 2, the network input includes the noisy image I^{noise} , the boundary mask I^{bd} , and the condition vector c . The diffusion time step is encoded as e_t , and the condition vector is

mapped to a condition embedding e_c . The encoder uses Cross-Attention Downsampling (CA-Down) modules to extract multi-scale features and inject condition information [55]:

$$h_{l+1} = \text{Down}(\text{CA}(\text{ResBlock}(h_l, e_t), e_c))$$

where $\text{ResBlock}(\cdot)$ denotes the residual block, $\text{Down}(\cdot)$ denotes downsampling, and $\text{CA}(\cdot)$ denotes cross-attention between image features and condition embedding. The bottleneck further fuses global layout features through a Cross-Attention Middle (CA-Mid) module:

$$h_{\text{mid}}^{\text{out}} = \text{CA}(\text{ResBlock}(h_{\text{mid}}, e_t), e_c)$$

The decoder adopts Cross-Attention Upsampling (CA-Up) modules to recover spatial details while continuing to incorporate the same condition information:

$$h_l^{\text{out}} = \text{CA}(\text{ResBlock}(\text{Concat}(\text{Up}(h_{l+1}), h_l^{\text{skip}}), e_t), e_c)$$

where $\text{Up}(\cdot)$ denotes upsampling, $\text{Concat}(\cdot)$ denotes feature concatenation, and h_l^{skip} denotes the skip connection feature. Through this design, time information, boundary information, and statistical condition information are jointly incorporated throughout progressive denoising, enabling controllable layout generation under physical and quantitative constraints.

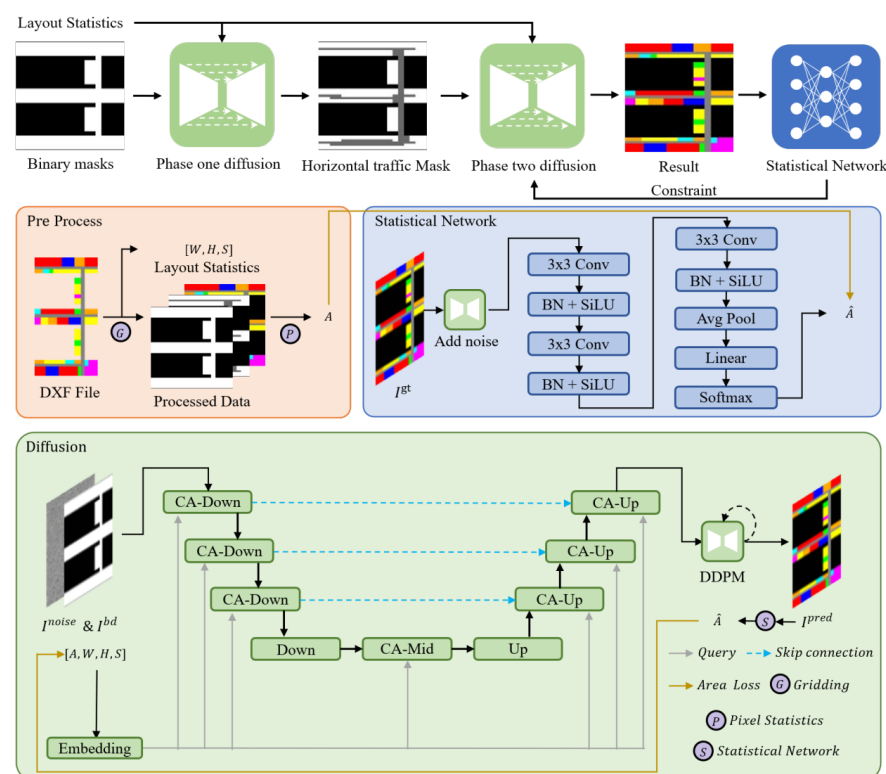


Figure 2. Overall model architecture.

3.3. Two-Stage Layout-Generation Framework

In university research building floor plan generation, directly generating a complete layout makes it difficult to maintain transportation connectivity and functional organization simultaneously. Preliminary experiments show that the horizontal transportation area often fails to connect all functional spaces when the complete layout is generated in a single step [56]. Therefore, this study decomposes the task into two stages: transportation space generation and complete layout generation [57].

In the first stage, the model generates only the horizontal transportation area. This transportation map serves as the organizational skeleton for the subsequent layout. The generation process is expressed as:

$$x_{\text{traffic}} = f_{\theta_1}(z, e_c)$$

where z denotes the initial noise and e_c denotes condition information.

In the second stage, the generated transportation map m_{traffic} is concatenated with the noisy image along the channel dimension and used as an additional condition for complete layout generation [47,50]:

$$x_{\text{layout}} = f_{\theta_2}(z, m_{\text{traffic}}, e_c)$$

This decomposition reduces generation complexity and encourages functional areas to be organized along a continuous circulation structure.

3.4. Explicit Constraint Guided by the Statistic Network

Although the functional area proportion vector is injected as a condition, pixel-level diffusion training alone may still produce deviations in the final area distribution. To improve quantitative controllability, this study introduces an explicit statistical constraint during conditional diffusion training. An independent statistic network predicts the functional area proportion vector from a generated layout:

$$\hat{A} = g_{\phi}(x)$$

where $g_{\phi}(\cdot)$ denotes the statistic network and $\hat{A}$ denotes the predicted area proportion vector. The statistic network uses a CNN-based backbone and is pretrained before diffusion training. It is then frozen during diffusion training to provide a stable supervision signal.

During diffusion training, the noise-free image estimate $\hat{x}_0$ is reconstructed from the current noisy sample x_t and the predicted velocity vector:

$$\hat{x}_0 = \sqrt{\alpha_t}x_t - \sqrt{1 - \alpha_t}v_{\theta}(x_t, t).$$

The reconstructed image is then input into the statistic network:

$$\hat{A} = g_{\phi}(\hat{x}_0).$$

The predicted and target area proportions are compared using an L1 statistical constraint loss:

$$L_{\text{layout}} = \frac{1}{K} \sum_{i=1}^K |\hat{A}_i - A_i|$$

where K denotes the number of functional categories, A_i denotes the target area proportion, and $\hat{A}_i$ denotes the predicted area proportion. The final training objective combines the diffusion denoising loss and the statistical constraint loss:

$$L = L_{\text{diff}} + \lambda L_{\text{layout}}$$

where L_{diff} denotes the standard diffusion denoising loss, L_{layout} denotes the statistical constraint loss, and λ balances image quality and statistical consistency. The statistical constraint is applied only in denoising stages where structural information is sufficiently clear. Since the statistic network is used only during training, it improves global area control without increasing inference cost.

4. Experimental Results

4.1. Model Training

4.1.1. Training Samples

The dataset contains 600 standard-floor samples from real built projects and un-built competition schemes of university research buildings. These samples cover both single-discipline and interdisciplinary research buildings, mainly including Chemical and Materials Sciences, Earth and Environmental Sciences, Physical Sciences, and Engineering and Technology. The detailed distribution of laboratory categories and university sources is provided in Appendix A.

4.1.2. Training Settings

To expand the sample size and improve model robustness, this study adopts the same data preprocessing and data augmentation strategies for all comparison methods to ensure the fairness and comparability of the experimental results.

In terms of optimization, all models are trained using the Adam optimizer. The initial learning rate is set to 1×10^{-4} and is updated every 10 epochs with a decay coefficient of 0.2. The total number of training epochs is set to 500. The batch size for both training and testing is set to 4.

To avoid performance bias caused by implementation details or hyperparameter tuning, all baseline models are trained and evaluated under a unified training framework. This includes consistent input resolution, data augmentation strategies, optimizer settings, learning rate scheduling, batch size, number of training epochs, and evaluation procedures. Unless otherwise specified, no additional independent hyperparameter tuning is conducted for different models. All experiments are implemented using PyTorch 2.0.1 with CUDA 11.8 and conducted on a single NVIDIA GeForce RTX 3090 GPU (NVIDIA Corporation, Santa Clara, CA, USA) to ensure consistency in the experimental environment.

4.2. Experimental Testing

4.2.1. Test Set

To evaluate the generation performance of the model, this study constructs an independent test set that is used consistently in subsequent model comparison experiments, ablation experiments, and the double-blind evaluation experiment. The test set includes 10 real university research building projects, none of which participate in model training, as shown in Table 2. The test set is balanced in terms of building type, stratified in building scale, and diverse in geometric conditions. It provides a reliable basis for subsequent quantitative evaluation and subjective evaluation, thereby enabling a more systematic verification of the effectiveness and applicability of the trained conditional diffusion model in the early-stage floor plan design of university research buildings.

Table 2. Test set data.

Group	East–West Span (m)	North–South Span (m)	Aspect Ratio	Building Area (m ²)	Discipline Type
1	80.0	150.0	0.53	5687.6	Multi-discipline
2	50.0	58.0	0.86	1552.7	Single-discipline
3	80.5	85.0	0.95	3634.8	Multi-discipline
4	57.6	54.0	1.07	2239.6	Single-discipline
5	45.0	51.3	0.88	1832.4	Single-discipline
6	58.0	63.4	0.91	1766.6	Single-discipline
7	84.0	90.5	0.93	5219.1	Multi-discipline
8	57.0	69.0	0.83	2461.5	Multi-discipline
9	82.8	36.0	2.30	1970.0	Multi-discipline
10	61.2	54.0	1.13	1867.8	Single-discipline

In terms of building scale, the building area of the test set ranges from 1552.7 m² to 5687.6 m². To ensure the representativeness of samples of different scales, the test set includes five samples below 2000 m² and five samples of 2000 m² or above. This grouping allows the test set to cover both small- and medium-scale research buildings and larger research buildings, which helps compare the adaptability of the model under different scale conditions.

In terms of floor plan geometry, the east–west span of the test set ranges from 45.0 m to 84.0 m, while the north–south span ranges from 36.0 m to 150.0 m. The numerical aspect ratio, calculated as east–west span divided by north–south span, ranges from approximately 0.53 to 2.30. According to the distribution characteristics, the samples in the test set can be summarized into three categories: low-aspect-ratio samples (0.53–0.88), medium-aspect-ratio samples (0.91–1.07), and high-aspect-ratio samples (1.13–2.30). The first two categories are relatively balanced in quantity, while the number of high-aspect-ratio samples is relatively small. These samples mainly reflect a small number of cases with special and elongated floor plan proportions.

In terms of building discipline type, the test set maintains a balanced distribution between single-discipline buildings and multi-discipline buildings, with five samples in each category. Compared with single-discipline buildings, multi-discipline research buildings usually have a higher degree of functional complexity and stronger demand for public communication. Their open communication spaces, shared platforms, and transportation organization are often more complex. Therefore, they can more effectively test the generation capability of the model under complex functional relationships.

In addition, to visually analyze the distribution consistency between the training set and the test set in the feature space, this study conducts principal component analysis (PCA) on the representations of the training and test sets. The samples are projected onto two-dimensional planes formed by different principal components. In the figure, blue points represent training samples, while red hollow circles represent test samples. Figure 3 shows that the test samples are generally located within or near the high-density regions of the training samples in the projected feature space. This indicates that the test set is not an out-of-distribution sample group, but still retains sufficient diversity in scale and geometric conditions. Therefore, the following experimental results can reasonably reflect the model's generalization ability within the target data distribution.

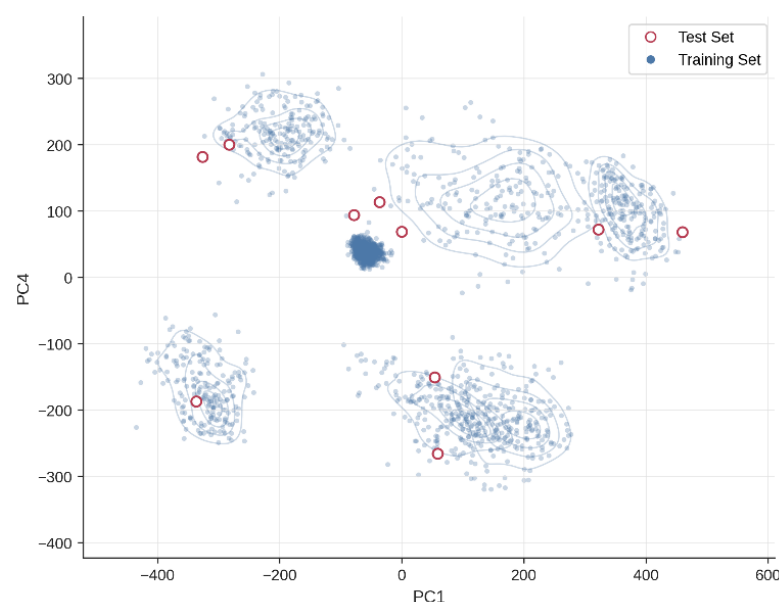


Figure 3. PCA dimensionality reduction distribution of the training set and test set.

4.2.2. Evaluation Metrics

To systematically evaluate the performance of generative models in the university research building floor-plan-generation task, this study constructs an evaluation system from four aspects: functional area control, building boundary constraints, transportation space connectivity, and overall generation feasibility. Fréchet Inception Distance (FID) is also introduced as a supplementary metric at the distribution level. These metrics correspond to the model's performance in quantitative constraint satisfaction, spatial boundary compliance, transportation organization rationality, and overall generation quality.

(1) Regional Area Error

Regional area error is used to measure the deviation between the generated result and the target area distribution. Suppose there are K functional areas. The area proportion of the k -th region in the generated result is $\hat{a}_k$, and the target area proportion is a_k . The area error is defined as follows:

$$\text{Area} = \frac{1}{K} \sum_{k=1}^K |\hat{a}_k - a_k|$$

where $\hat{a}_k$ is obtained through pixel statistics:

$$\hat{a}_k = \frac{1}{H \times W} \sum_{i=1}^H \sum_{j=1}^W \mathbf{1}(x_{ij} = k)$$

In the formula, H and W denote the height and width of the image, respectively, and $\mathbf{1}(\cdot)$ denotes the indicator function. When a pixel (i, j) belongs to the k -th category, the function takes the value of 1; otherwise, it takes the value of 0. The smaller this metric is, the more closely the generated result satisfies the target area constraint.

(2) Boundary Matching Degree

Boundary matching degree is used to measure how well the generated layout follows the given building outline. Let the input boundary mask be denoted as $B \in \{0, 1\}^{H \times W}$, and the effective region of the generated result be denoted as $\hat{B}$. The boundary matching degree is defined as follows:

$$\text{Boundary} = \frac{\sum_{(i,j)} \mathbf{1}(\hat{B}_{ij} = B_{ij})}{H \times W}$$

This metric essentially measures the pixel-level overlap between the generated layout and the input boundary. Its value ranges from 0 to 1. A value closer to 1 indicates that the generated result more accurately follows the building boundary constraint.

(3) Spatial Connectivity

Spatial connectivity is used to evaluate the overall connectivity of the transportation space. Let the extracted transportation area be denoted as a binary image $M \in \{0, 1\}^{H \times W}$. After connected-component labeling, N connected components are obtained, and their corresponding pixel numbers are denoted as $s_1, s_2, \dots, s_N$. Spatial connectivity is defined as follows:

$$\text{Conn} = \frac{\max(s_1, s_2, \dots, s_N)}{\sum_{i=1}^N s_i}$$

This metric reflects the proportion of the largest connected component to the total transportation area. When the value equals 1, the transportation space is fully connected. A higher value indicates that the transportation organization is more complete and that the spatial circulation is more continuous.

(4) Success Rate

Based on the above individual metrics, this study further introduces success rate as an overall evaluation metric to assess the model's ability to generate valid schemes under multiple constraints. Success rate measures the proportion of generated samples that simultaneously satisfy the area constraint, boundary constraint, and connectivity constraint. Suppose there are N generated samples. The area error, boundary matching degree, and spatial connectivity of the i -th sample are denoted as Area_i , Boundary_i , and Conn_i , respectively, and their corresponding thresholds are denoted as τ_a , τ_b , and τ_c . The success rate is defined as follows:

$$\text{SuccessRate} = \frac{1}{N} \sum_{i=1}^N \mathbf{1}((\text{Area}_i < \tau_a) \wedge (\text{Boundary}_i > \tau_b) \wedge (\text{Conn}_i > \tau_c))$$

where $\mathbf{1}(\cdot)$ is an indicator function. It takes the value of 1 when the sample satisfies all three constraints simultaneously; otherwise, it takes the value of 0. This metric comprehensively reflects the model's ability to generate valid solutions under multi-objective constraints. A higher value indicates stronger overall feasibility. In the following experiments, the area error threshold, boundary matching threshold, and connectivity threshold are set to $\tau_a = 6\%$, $\tau_b = 90\%$, and $\tau_c = 80\%$, respectively.

(5) Fréchet Inception Distance (FID)

In addition to constraint-satisfaction metrics, this study further introduces Fréchet Inception Distance (FID) to evaluate the distribution-level difference between generated samples and real samples. FID measures the distance between generated and real samples in the feature space, based on high-dimensional features extracted by the Inception network. Suppose the features of real samples follow a Gaussian distribution $\mathcal{N}(\mu_r, \Sigma_r)$, and the features of generated samples follow a Gaussian distribution $\mathcal{N}(\mu_g, \Sigma_g)$. FID is defined as follows:

$$\text{FID} = \|\mu_r - \mu_g\|^2 + \text{Tr}(\Sigma_r + \Sigma_g - 2(\Sigma_r \Sigma_g)^{1/2})$$

where μ_r and μ_g denote the feature means of real and generated samples, respectively; Σ_r and Σ_g denote the corresponding covariance matrices; and $\text{Tr}(\cdot)$ denotes the matrix trace operation. A lower FID indicates that the generated distribution is closer to the real distribution and that the generation quality is higher.

All statistical analyses and graph generation were conducted using Python 3.11.7, including NumPy 2.4.4, Pandas 3.0.2, SciPy 1.10.1, scikit-learn 1.3.0, and Matplotlib 3.10.8.

4.3. Model Comparison Experiment

This study selects several representative conditional generative models for comparison, covering both generative adversarial networks and diffusion models. The purpose is to systematically evaluate their capabilities in structural modeling and constraint representation from the perspective of different generation mechanisms. Specifically, the comparison models include the classic conditional adversarial generation model Pix2Pix, BicycleGAN, and Stable Diffusion v1.5 fine-tuned with ControlNet. Pix2Pix maps condition inputs to target layouts through adversarial learning and represents the deterministic image-to-image translation paradigm. BicycleGAN introduces latent variables and bidirectional consistency constraints on this basis, enabling the modeling of one-to-many mapping relationships and improving generation diversity to some extent. Stable Diffusion v1.5 fine-tuned with ControlNet injects structural information as a condition into the diffusion process, thereby guiding and controlling the generated results. It represents a mainstream technical route based on large-scale pretrained diffusion models and recent conditional diffusion floor-plan systems [55,58]. To ensure the fairness of the comparison, Pix2Pix, BicycleGAN, and Stable

Diffusion v1.5 fine-tuned with ControlNet are all trained using the same dataset, optimizer, and training settings as the proposed method. This minimizes performance bias caused by differences in training conditions.

During the experiment, the 10 university research building projects in the test set are used as unified input objects. Floor plan layouts are generated according to the building outline boundary of each project. For Stable Diffusion v1.5 and the proposed method, which have strong random sampling ability and output diversity, this study generates 200 candidate images for each test project. Some generated results are shown in Figure 4. The mean values of all evaluation metrics are then calculated based on all generated results to more comprehensively reflect the overall generation performance for each sample. By contrast, in the experimental setting of this study, common Pix2Pix and BicycleGAN models produce only one output for each input and have difficulty incorporating functional area constraints. Therefore, for these two methods, the mean values of the evaluation metrics can only be calculated based on their respective outputs for the 10 test samples. Because these methods lack a comparable mechanism for area control, regional area error is not included in the model comparison evaluation in this section. Instead, FID, building boundary IoU, and Horizontal transportation connectivity are used for comprehensive comparison. The experimental results are shown in Table 3.

The results in Table 3 show that the proposed method achieves the best overall performance among the compared models, with the lowest FID, the highest building boundary IoU, and the highest horizontal transportation connectivity. Figure 4 further provides a visual comparison of generated results, showing that the proposed method better preserves building boundary constraints and produces more continuous horizontal transportation spaces. A more detailed interpretation of the model comparison results is provided in Section 5.1.

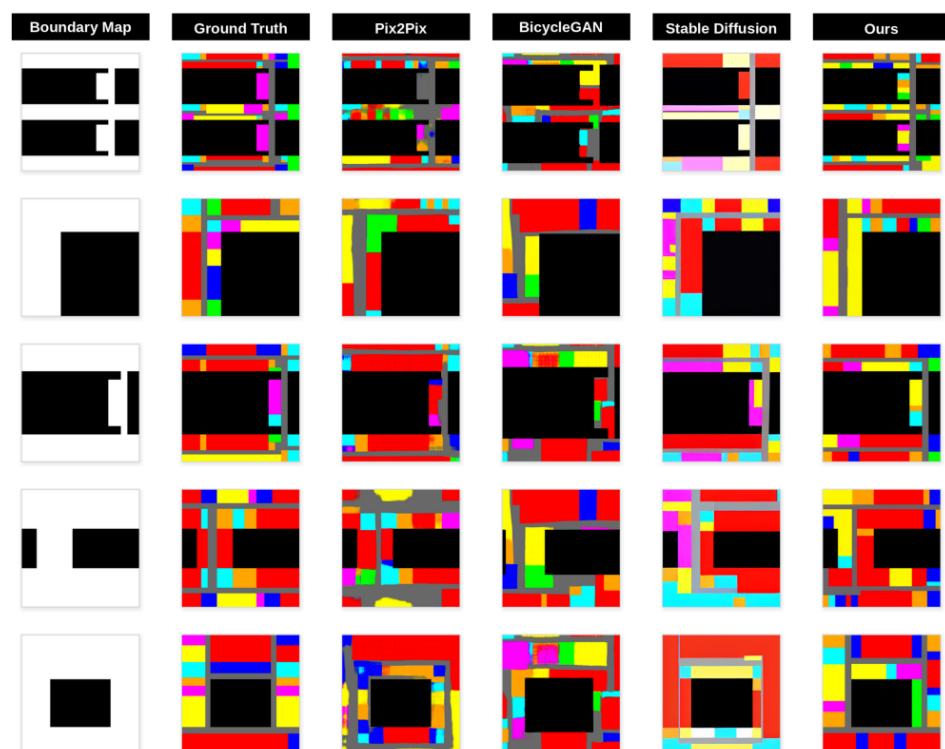


Figure 4. Comparison of generated results from different models.

Table 3. Comparison of different models.

Model	FID	Building Boundary IoU	Horizontal Transportation Connectivity
Pix2pix	137.4	97.6%	44.3%
BicycleGAN	107.2	96.2%	47.2%
Stable Diffusion	62.4	98.3%	82.5%
Our methods	50.3	99.9%	89.8%

4.4. Ablation Experiment

To verify the effects of the key strategies proposed in this study on the university research building floor-plan-generation task, this study further conducts ablation experiments. The experiments examine the influence of the two-stage layout generation mechanism and the explicit constraint guided by the statistic network on model performance. To ensure the comparability of the experimental results, all ablation models are trained using the same dataset, optimizer, and training settings as the complete model and are evaluated on the same test set. During the experiment, this study generates floor plan layouts for the 10 university research building projects in the test set according to the corresponding building boundary conditions and area requirements of each project. For each test sample, each model generates 200 candidate images. The mean values of all evaluation metrics are then calculated based on all generated results to reflect as comprehensively as possible the generation performance of the model under different constraint conditions.

4.4.1. Verification of the Two-Stage Generation Mechanism

To verify the effectiveness of the two-stage layout generation strategy, this study compares the complete model with the strategy against a single-stage baseline model without the two-stage mechanism. The experimental results are shown in Table 4.

Table 4. Comparison between the two-stage model and the single-stage model.

Model	FID	Area Error	Building Boundary IoU	Horizontal Transportation Connectivity	Success Rate
Single-stage model	52.1	6.5%	99.9%	85.1%	2.3%
Two-stage model	50.3	5.9%	99.9%	89.8%	18.2%

The results show that the two-stage model outperforms the single-stage baseline in FID, area error, horizontal transportation connectivity, and success rate. Specifically, the success rate increases from 2.3% to 18.2%, and horizontal transportation connectivity increases from 85.1% to 89.8%. Meanwhile, both models achieve a building boundary IoU of 99.9%, indicating that the two-stage mechanism improves transportation space organization without weakening boundary compliance. Figure 5 shows the probability distribution of area error between the single-stage and two-stage models. The distribution of the two-stage model is more concentrated in the lower-error range, suggesting that generating the horizontal transportation structure before the complete layout helps stabilize the subsequent functional area allocation. Further interpretation of the role of the two-stage mechanism is provided in Section 5.2.

In terms of building boundary IoU, both models reach 99.9%, with almost no difference. This indicates that the two-stage mechanism does not weaken the model's ability to satisfy boundary constraints when introducing an additional generation step. In terms of Horizontal transportation connectivity, the complete model improves from 85.1% to 89.8%, indicating that the two-stage strategy performs better in overall spatial organization. This is because the transportation skeleton generated in the first stage provides clear structural guidance for subsequent functional areas, thereby improving transportation connectivity.

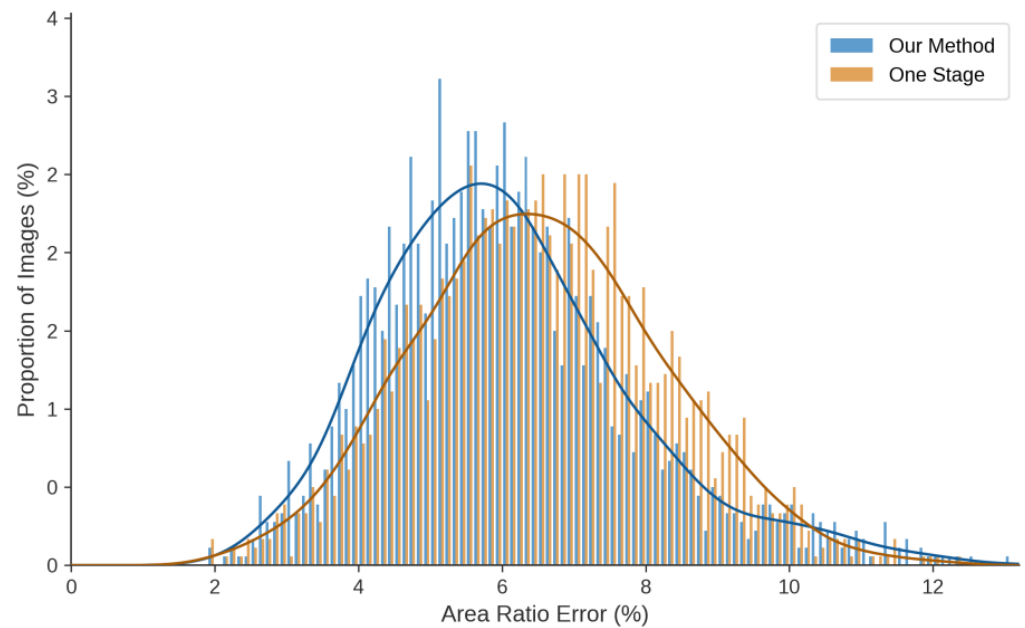


Figure 5. Probability distribution of area error between the single-stage and two-stage models.

More importantly, the success rate of the two-stage model increases significantly from 2.3% to 18.2%. This result shows that the two-stage layout generation strategy not only improves individual metrics but also significantly enhances the model's ability to generate valid schemes under multiple constraints. Overall, the two-stage layout generation strategy improves distributional consistency, area control accuracy, spatial connectivity, and overall success rate without affecting boundary accuracy, verifying its potential in research building floor plan generation.

4.4.2. Verification of the Explicit Statistical Constraint

To verify the role of the proposed statistic network in the layout-generation process, this study conducts an ablation experiment comparing the complete model with a model in which the statistical constraint is removed. The experimental results are shown in Table 5.

Table 5. Comparison with and without the statistic network constraint.

Model	FID	Area Error	Building Boundary IoU	Horizontal Transportation Connectivity	Success Rate
w/o statistic network	55.3	9.4%	99.9%	82.1%	1.6%
Proposed method	50.3	5.9%	99.9%	89.8%	18.2%

As shown in Table 5, the model with the statistic-network-guided explicit constraint achieves better performance across all evaluation metrics. Compared with the model without the statistic network, the proposed method reduces the FID from 55.3 to 50.3, decreases the area error from 9.4% to 5.9%, improves horizontal transportation connectivity from 82.1% to 89.8%, and increases the success rate from 1.6% to 18.2%. The building boundary IoU remains 99.9% in both settings, indicating that the explicit statistical constraint mainly improves global area consistency and layout feasibility without affecting boundary control. Figure 6 shows the probability distribution of area error with and without the statistic network constraint. The proposed method presents a clearer concentration in the low-error range, indicating that the additional statistical supervision helps the generated layouts better satisfy the target functional area proportions. The mechanism behind this improvement is further discussed in Section 5.2.

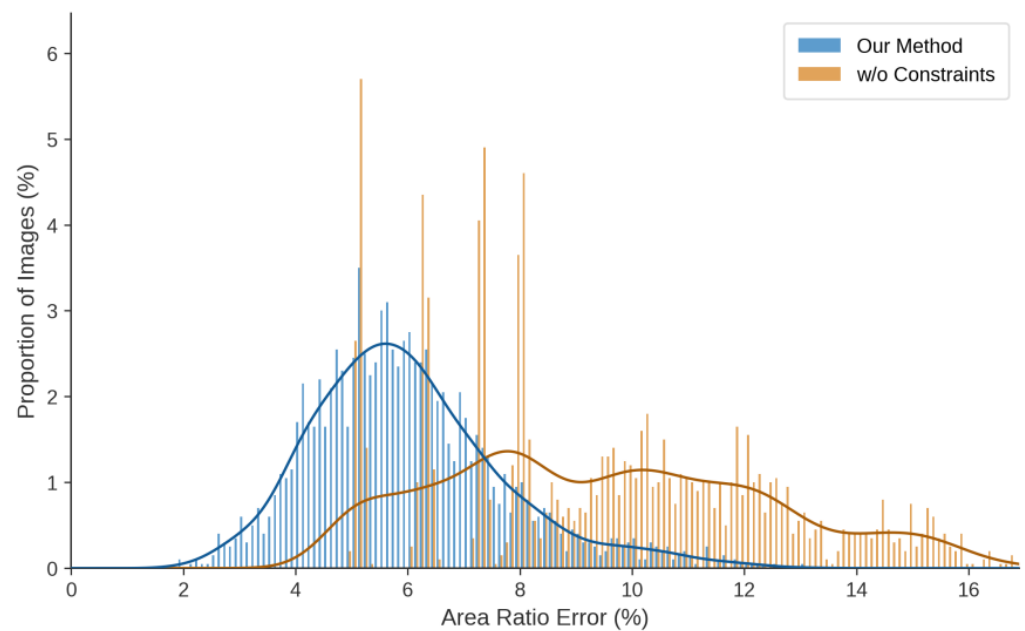


Figure 6. Probability distribution of area error with and without the statistic network constraint.

4.5. Double-Blind Evaluation Experiment

4.5.1. Experimental Settings

In this research, the “correctness” of an AI-generated floor plan is not defined as equivalence to a complete architectural design or construction drawing. Instead, it is defined as the degree to which the generated result satisfies the requirements of the early-stage floor plan design stage. Therefore, design correctness is verified from two complementary perspectives. First, quantitative metrics are used to evaluate whether the generated layouts satisfy explicit constraints, including functional area control, building boundary matching, and horizontal transportation connectivity. Second, expert review is used to evaluate qualitative design attributes that cannot be fully captured by pixel-based metrics.

To evaluate, from the perspective of architects, the difference between university research building floor plans generated by the conditional diffusion model constructed in this study and real design cases, this study further conducts a double-blind evaluation experiment. Unlike the previous model comparison and ablation experiments, which focus on objective metrics, this section focuses on the performance of AI-generated floor plans at the level of architects’ subjective perception. Particular attention is paid to the degree to which AI-generated floor plans approach real cases in terms of spatial organization, spatial form, and design inspiration.

This study selects the 10 real university research building projects in the test set as evaluation objects. For each project, based on the building outline and functional indicators corresponding to the real floor plan, the trained conditional diffusion model generates 200 candidate floor plans. Four floor plans with the best functional organization and generation quality are selected from these candidates as AI-group samples. Therefore, each experimental group contains one real floor plan and four AI-generated floor plans, forming 10 groups and 50 floor plans to be evaluated. The floor plans to be evaluated are shown in Figure 7. To prevent reviewers from forming prior judgments based on information about drawing sources, all samples are anonymized before evaluation. This includes removing source information, unifying the drawing representation style, and randomly renumbering the drawings, so that reviewers cannot distinguish real floor plans from AI-generated floor plans.

Considering the differences in representation between functional color-block diagrams and conventional architectural floor plans, this study organizes preparatory training before formal scoring to reduce reviewers' misinterpretation of the drawing representation. The training includes comparative learning using 20 groups of functional color-block diagrams and their corresponding architectural floor plans. This helps reviewers understand the spatial logic of different functional zones in color-block diagrams and their correspondence with actual architectural floor plans, thereby improving the consistency and validity of subsequent scoring.

This study invites five architects from the Architectural Design and Research Institute of Zhejiang University, all with experience in research building design, to form the expert review panel. All reviewers independently complete scoring without knowing the source of each drawing. The questionnaire criteria are defined according to the main tasks of the early-stage floor plan design of university research buildings. At this stage, the evaluation focuses on whether the generated layout can provide a feasible functional zoning structure and a useful design starting point, rather than whether it satisfies all technical requirements of later design stages. Therefore, the questionnaire is organized into three primary dimensions. Spatial organization rationality evaluates the basic functional logic and circulation organization of the floor plan. Functional spatial form evaluates the morphological quality, scale appropriateness, regularity, and environmental potential of the generated spaces. Innovation and inspiration evaluates whether the generated result can provide alternative design possibilities and support subsequent design development. A five-point scale is adopted, with higher scores indicating higher reviewer evaluations of the corresponding item. The specific evaluation dimensions and scoring criteria are shown in Table 6.

The evaluation dimensions also respond to several qualitative aspects of architectural design. Physical environmental quality is preliminarily assessed through the daylighting and ventilation potential of functional spaces. Humanistic and user-related considerations are reflected in the evaluation of circulation rationality, functional layout logic, open communication spaces, and the potential for subsequent design development. The formal connotation of architectural design is mainly examined through spatial regularity, layout novelty, and design inspiration. However, these evaluations remain at the level of initial floor plan assessment and are based on expert judgment of functional color-block diagrams. They do not replace detailed environmental simulation, structural design, fire-safety verification, or post-occupancy user evaluation.

In terms of data processing, this study takes the reviewer-experimental group combination as the basic observation unit for primary analysis. For the real group, each reviewer corresponds to one real floor plan in each experimental group. The score for a primary dimension is obtained by taking the arithmetic mean of all sub-indicator scores under that dimension. For the AI group, each reviewer corresponds to four AI-generated floor plans in each experimental group. This study first calculates the score of each AI-generated floor plan under the corresponding primary dimension and then averages the dimension scores of the four AI-generated floor plans in the same group. This value represents the reviewer's overall evaluation of the AI-generated results in that experimental group under the corresponding primary dimension. Based on this processing method, each reviewer forms a paired observation of "real floor plan-AI-generated floor plan" under each experimental group. Because this study includes five reviewers and 10 experimental groups, a total of 50 paired samples are formed for primary-dimension analysis. This data structure maintains consistency in the statistical units of the real group and the AI group while avoiding sample-structure imbalance caused by the fact that the number of AI-generated floor plans within each group is larger than the number of real floor plans. It therefore provides a reliable basis for subsequent statistical testing.

Considering that the scoring data in the double-blind experiment are ordered categorical variables based on a five-point scale and that the sample size is relatively limited, this study mainly adopts nonparametric statistical methods for analysis.

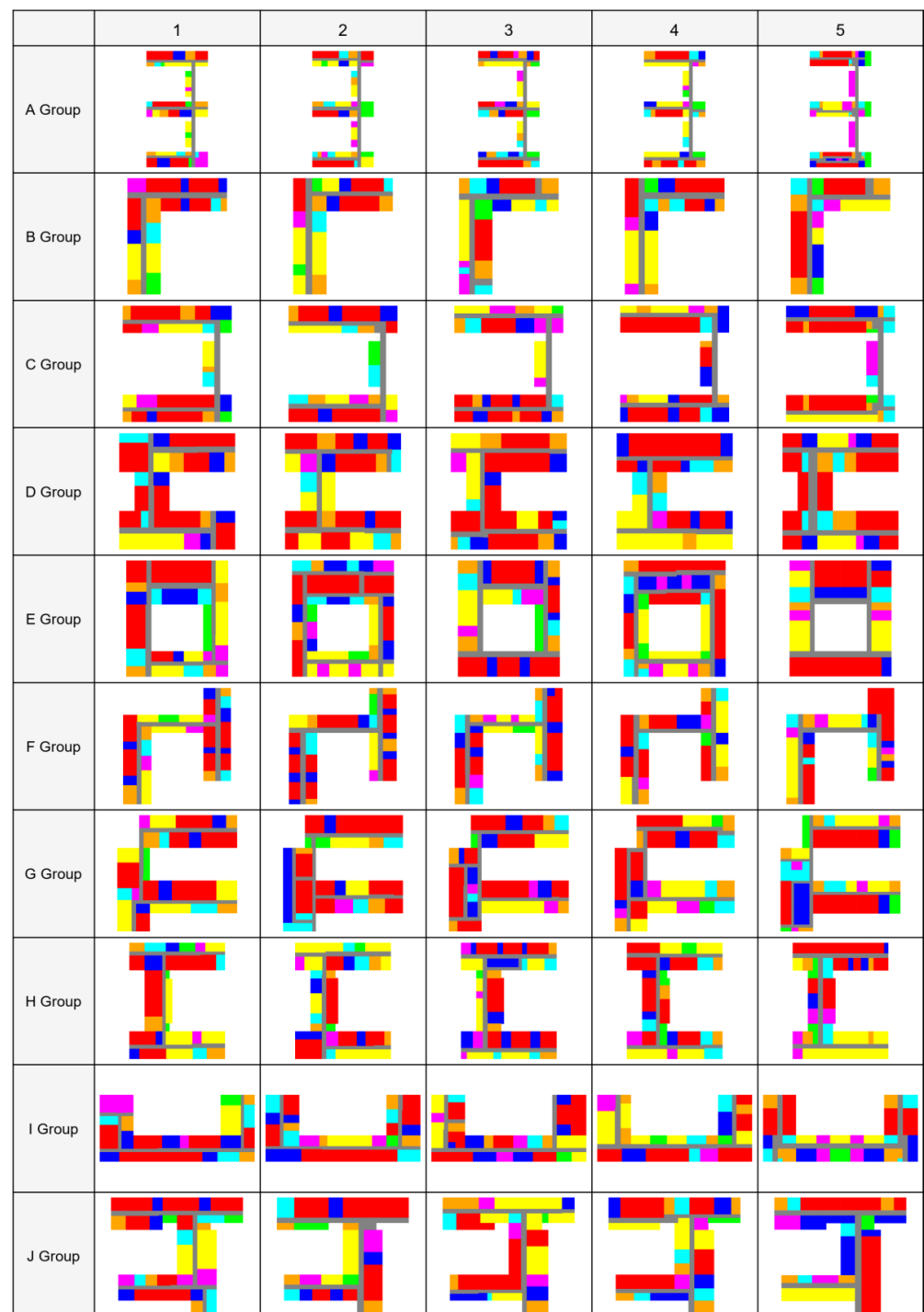


Figure 7. Floor plans to be evaluated in the double-blind evaluation experiment.

Table 6. Evaluation dimensions and scoring criteria.

Evaluation Dimension	Sub-Indicator	Evaluation Criterion (Five-Point Scale)
Spatial organization rationality	Functional layout logic	Evaluates whether the layout positions of the eight functional zones conform to the research workflow.
	Functional-zone boundary smoothness	Evaluates whether the edges of functional color blocks are clear and free from fragmentation and whether there is adhesion between rooms.
	Circulation connectivity	Evaluates whether the horizontal transportation space effectively connects all functional spaces and whether the vertical transportation space satisfies evacuation requirements.
	Circulation rationality	Evaluates whether the research circulation, including people, materials, and experimental paths, is concise and smooth.
Functional spatial form	Spatial scale appropriateness	Evaluates whether the scale of each room conforms to its disciplinary attributes and whether the spatial scale has real physical meaning.
	Spatial regularity	Evaluates whether room shapes are regular and whether deformed spaces or overly fragmented invalid pixel regions exist.
	Daylighting and ventilation potential	Examines whether spaces with environmental requirements, such as office and laboratory spaces, are reasonably arranged along external walls, atriums, or other interfaces with daylighting and ventilation conditions.
Innovation and inspiration	Layout novelty	Evaluates whether the scheme provides a floor plan arrangement different from conventional routine designs and whether open communication areas show distinctive spatial organization.
	Design inspiration	Evaluates whether the generated result can break designers' habitual thinking and provide a new inspirational perspective or design starting point in the scheme-conception stage.
	Potential for design development	Evaluates the professional maturity of the initial floor plan and determines whether it has potential for further design development.

4.5.2. Comprehensive Analysis of Primary Dimensions

To compare the subjective evaluation differences between real floor plans and AI-generated floor plans at the overall level, this study first conducts a paired comparison between the two groups at the primary-dimension level, as shown in Table 7. The mean values reflect the overall scoring levels of the real floor plans and AI-generated floor plans under each primary evaluation dimension, while the standard deviations reflect the dispersion of the scoring results. The comprehensive comparison between the real group and the AI group is further visualized in Figure 8.

Table 7. Comprehensive comparison results of primary dimensions.

Dimension	Real Group Mean (SD)	AI Group Mean (SD)	Mean Difference	<i>p</i> -Value
Spatial organization rationality	4.155 (0.614)	3.836 (0.449)	0.319	0.0005
Functional spatial form	3.880 (0.770)	3.735 (0.440)	0.145	0.1110
Innovation and inspiration	3.507 (0.584)	3.442 (0.468)	0.065	0.5828
Overall score	3.878 (0.549)	3.688 (0.354)	0.191	0.0170

Note: *p*-values were calculated using the Wilcoxon signed-rank test.

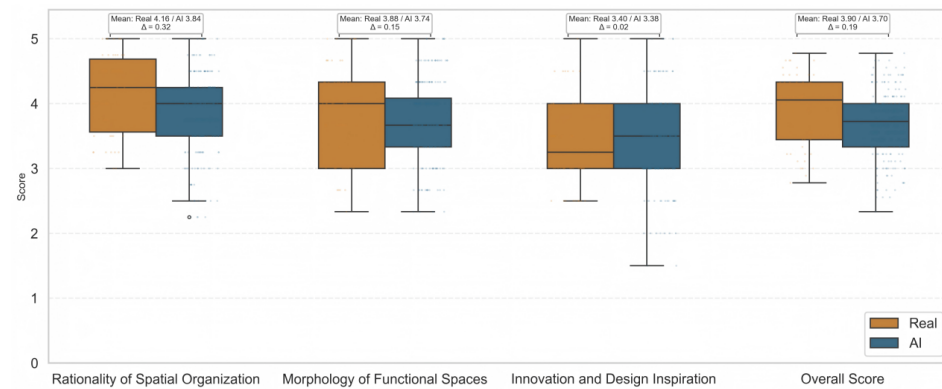


Figure 8. Boxplots of the overall comparison between real and AI groups across first-level evaluation dimensions.

Comprehensive comparison between real and AI-generated floor plans at the primary-dimension level. The figure shows the score distributions of spatial organization rationality, functional spatial form, innovation and inspiration, and overall evaluation. The results show that real floor plans obtain higher scores than AI-generated floor plans in spatial organization rationality. In contrast, the score differences between the two groups are smaller in functional spatial form and innovation and inspiration. This indicates that AI-generated layouts can approach real design cases in terms of spatial form and design inspiration, while still showing limitations in the rational organization of complex functional relationships.

4.5.3. Sub-Indicator Analysis

To further identify the specific sources of differences at the primary-dimension level, this study conducts a more detailed paired comparison between real floor plans and AI-generated floor plans at the sub-indicator level, as shown in Table 8. The comparison of sub-indicators between the real group and the AI group is shown in Figure 9.

Table 8. Comparison results at the sub-indicator level.

Dimension	Sub-Indicator	Real Group Mean (SD)	AI Group Mean (SD)	Mean Difference	<i>p</i> -Value
Spatial organization rationality	Functional layout logic	4.160 (0.766)	3.780 (0.377)	0.380	0.0024
	Functional-zone boundary smoothness	4.280 (0.809)	3.885 (0.661)	0.395	0.0016
	Circulation connectivity	4.020 (0.958)	3.925 (0.665)	0.095	0.5596
	Circulation rationality	4.160 (0.792)	3.755 (0.561)	0.405	0.0029
Functional spatial form	Spatial scale appropriateness	3.640 (1.139)	3.630 (0.521)	0.010	0.9638
	Spatial regularity	3.920 (1.027)	4.100 (0.680)	−0.180	0.2471
	Daylighting and ventilation potential	4.080 (0.804)	3.475 (0.635)	0.605	0.0001
Innovation and inspiration	Layout novelty	3.360 (0.663)	3.435 (0.575)	−0.075	0.6217
	Design inspiration	3.440 (0.951)	3.325 (0.523)	0.115	0.3861
	Potential for design development	3.720 (0.834)	3.565 (0.544)	0.155	0.2025

Note: *p*-values were calculated using the Wilcoxon signed-rank test.

Figure 9 Comparison of sub-indicator scores between real and AI-generated floor plans. The figure illustrates the detailed evaluation differences in functional layout logic, circulation rationality, zoning relationship, spatial scale appropriateness, spatial regularity, daylighting and ventilation potential, layout novelty, design inspiration, and potential for design development. The sub-indicator results further show that the gap between real and AI-generated floor plans is mainly concentrated in indicators related to spatial

organization and circulation logic. By contrast, the AI-generated layouts perform relatively closer to real cases in indicators related to spatial form, novelty, and design inspiration. This result provides a more detailed basis for the subsequent discussion of the strengths and limitations of the proposed method.

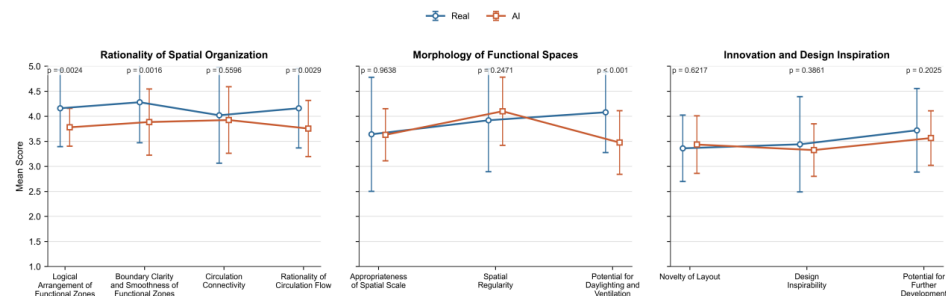


Figure 9. Comparison of sub-indicator scores between the real and AI groups.

4.5.4. Analysis of Influencing Factors Related to Building Attributes

In the analysis of building attribute factors, this study takes the scores of AI-generated floor plans as the analysis object and uses the reviewer–experimental group combination as the basic observation unit. The analysis explores the influence of building aspect ratio, building area, and building discipline type on the subjective evaluation results of AI-generated floor plans. Aspect ratio and building area are continuous variables. Therefore, Spearman rank correlation analysis is used to examine their influence on scores [59]. Building discipline type is a categorical variable. Therefore, the Mann–Whitney U test is used to compare the influence of building discipline type on scores [60]. The analysis focuses on three primary dimensions, namely spatial organization rationality, functional spatial form, and innovation and inspiration, as well as the overall score.

According to the correlation analysis results for aspect ratio in Table 9, building aspect ratio shows a significant negative correlation with the spatial organization rationality score of AI-generated floor plans. The relationship between aspect ratio and spatial organization rationality is shown in Figure 10.

The correlation analysis between building area and AI-generated floor plan scores is shown in Table 10, and the corresponding relationship is visualized in Figure 11. The results indicate whether building scale affects the subjective evaluation of AI-generated layouts under different design dimensions.

Table 9. Correlation analysis between aspect ratio and AI-generated floor plan scores.

Dimension	Spearman ρ	p -Value
Spatial organization rationality	−0.281	0.0481
Functional spatial form	−0.246	0.0853
Innovation and inspiration	−0.154	0.2855
Overall score	−0.253	0.0762

Note: p -values were calculated using Spearman rank correlation analysis.

Table 10. Correlation analysis between building area and AI-generated floor plan scores.

Dimension	Spearman ρ	p -Value
Spatial organization rationality	−0.263	0.0650
Functional spatial form	−0.319	0.0240
Innovation and inspiration	−0.088	0.5437
Overall score	−0.298	0.0354

Note: p -values were calculated using Spearman rank correlation analysis.

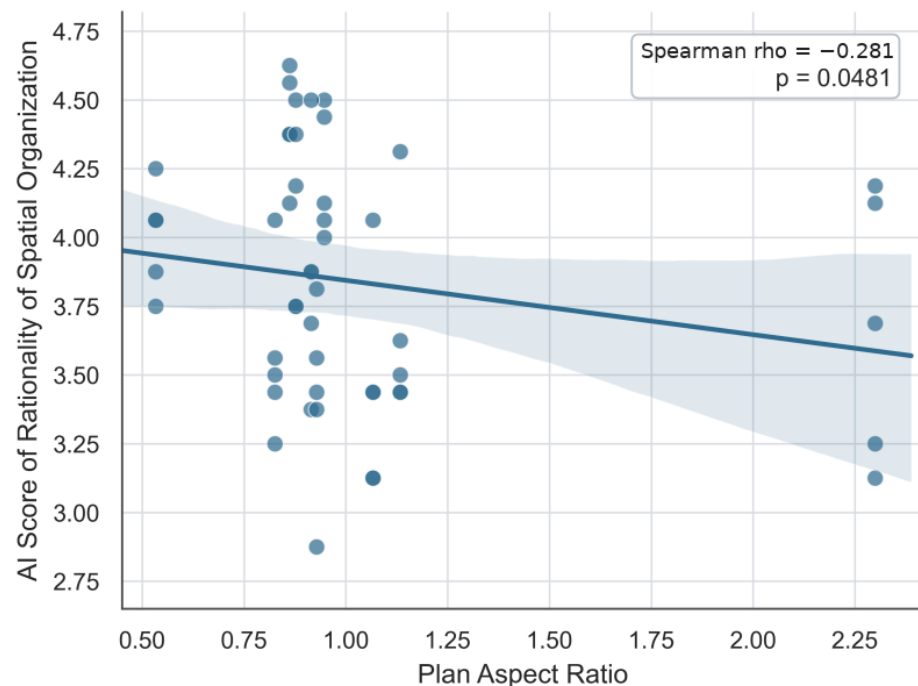


Figure 10. Relationship between plan aspect ratio and the AI score of spatial organization rationality.

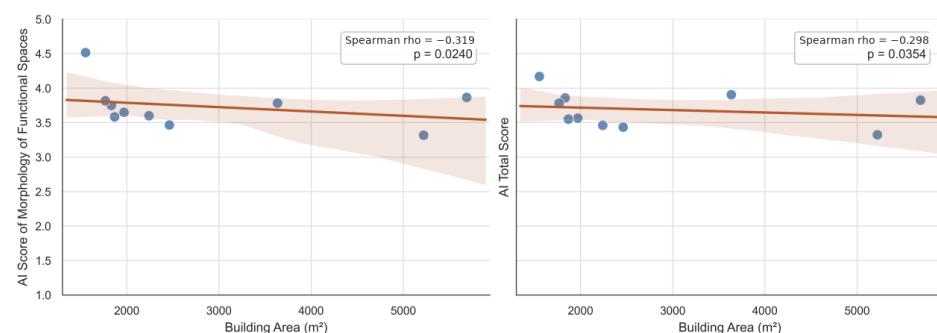


Figure 11. Relationship between building area and AI scores.

The influence of building discipline type on AI-generated floor plan scores is analyzed using the Mann–Whitney U test, as shown in Table 11. The score distributions of single-discipline and multi-discipline buildings are further shown in Figure 12.

Overall, the above results indicate that geometric and functional attributes of buildings may influence the subjective evaluation of AI-generated layouts. In particular, more elongated floor plan proportions, larger spatial scales, or more complex disciplinary compositions may increase the difficulty of generating layouts with high spatial organization rationality. A further interpretation of these results is provided in Section 5.3.

Table 11. Comparison of AI-generated floor plan scores under different building discipline types.

Dimension	Single-Discipline AI Group (SD)	Multi-Discipline AI Group (SD)	U Value	<i>p</i> -Value
Spatial organization rationality	3.898 (0.472)	3.775 (0.425)	359.0	0.3711
Functional spatial form	3.853 (0.428)	3.617 (0.428)	396.0	0.1067
Innovation and inspiration	3.497 (0.475)	3.387 (0.465)	342.5	0.5655
Overall score	3.764 (0.350)	3.611 (0.348)	385.0	0.1620

Note: *p*-values were calculated using the Mann–Whitney U test.

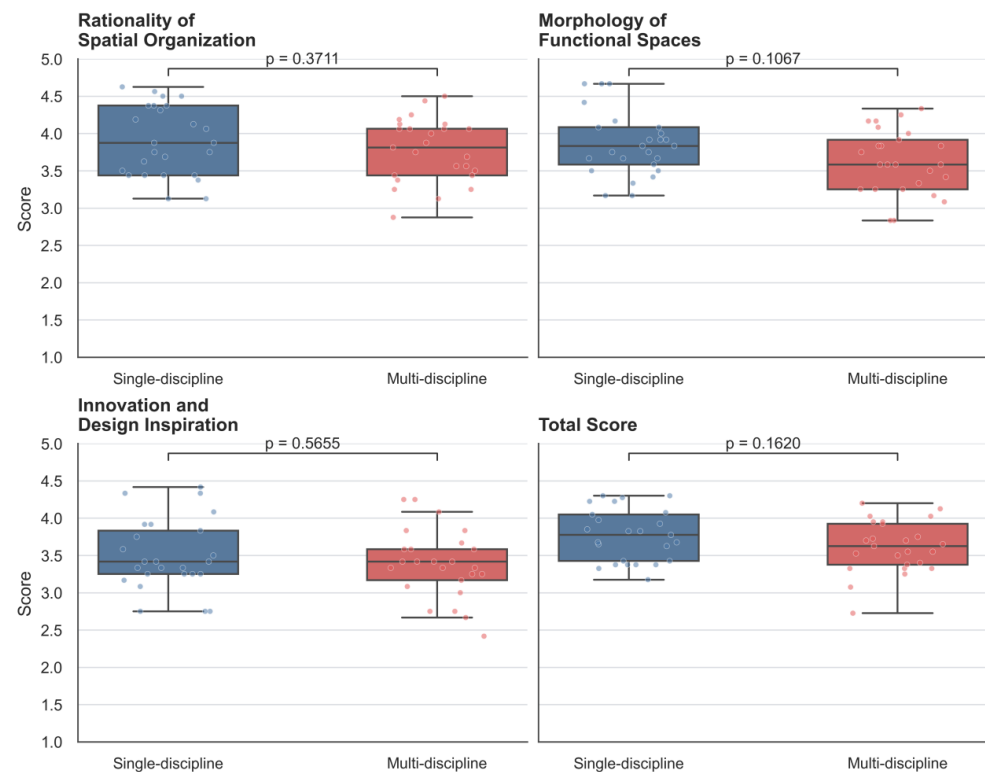


Figure 12. Comparison of AI scores between single-discipline and multi-discipline buildings.

5. Discussion

5.1. Interpretation of Model Comparison Results

The model comparison experiment shows that the proposed method achieves better overall performance than the selected baseline models. This result is closely related to the specific requirements of university research building floor plan generation. Unlike general image generation tasks, this task requires the model not only to generate visually plausible layouts, but also to respond simultaneously to irregular building boundaries, quantitative functional area requirements, and transportation space organization.

The GAN-based methods, including Pix2Pix and BicycleGAN, show limitations in this task. Although these methods can learn image-to-image mapping relationships, they tend to focus more on local visual correspondence and have difficulty maintaining global spatial organization under complex functional and boundary constraints. This is reflected in their weaker horizontal transportation connectivity and less stable functional zoning structures. For university research buildings, where laboratories, research offices, support spaces, and public communication areas must be organized through a clear circulation system, this limitation directly affects the rationality of the generated layouts.

Stable Diffusion v1.5 fine-tuned with ControlNet performs better than the GAN-based methods, indicating that diffusion models have advantages in structural consistency and conditional generation. However, the latent-space generation mechanism of Stable Diffusion is originally designed for natural image generation. When applied to architectural functional color-block layouts, the compression and reconstruction process may weaken pixel-level precision and reduce the stability of boundary and area control.

By contrast, the proposed method performs diffusion denoising directly in two-dimensional grid space and introduces both building boundary masks and functional area matrices as generation conditions. This design is more consistent with the representation characteristics of architectural floor plans, because functional zoning layouts require clear semantic boundaries, stable area proportions, and accurate response to the

building outline. Therefore, the improvement of the proposed method is not only reflected in image distribution similarity, but also in spatial constraint satisfaction and transportation space continuity.

It should also be noted that FID is used in this study mainly as a horizontal comparison metric. Architectural functional floor plans are composed of discrete semantic color blocks, and their data distribution differs substantially from natural images. Therefore, the absolute value of FID should not be interpreted in the same way as in natural image generation tasks. Under the same evaluation protocol, however, FID can still provide useful supplementary evidence when combined with building boundary IoU and horizontal transportation connectivity.

5.2. Interpretation of Ablation Results

The ablation experiments further clarify the roles of the two main components proposed in this study: the two-stage layout generation mechanism and the statistic-network-guided explicit constraint. These two components address different difficulties in early-stage floor plan generation for university research buildings.

The two-stage generation mechanism mainly improves the organization of horizontal transportation space. In university research buildings, horizontal circulation is not only a traffic component but also the primary spatial skeleton that organizes laboratories, research offices, support spaces, and open communication areas. If the complete layout is generated directly in a single stage, the model needs to determine circulation organization and functional area distribution simultaneously, which increases generation difficulty and may lead to fragmented or weakly connected transportation spaces. By first generating the horizontal transportation structure and then using it as a condition for complete layout generation, the model obtains a clearer organizational reference for functional area allocation. This explains why the two-stage model improves horizontal transportation connectivity and overall success rate.

The statistic-network-guided explicit constraint mainly improves global functional area control. Although the functional area matrix is injected into the diffusion model as a condition, condition embedding alone does not necessarily guarantee that the final pixel distribution will strictly match the target area proportions. The statistic network provides an additional global supervision signal during training. By comparing the predicted area proportions of generated layouts with the target area conditions, the model is encouraged to generate layouts that better satisfy quantitative functional requirements.

The two components are therefore complementary. The two-stage mechanism strengthens the spatial organization logic of the generated layouts, especially the continuity and guiding role of horizontal transportation. The explicit statistical constraint strengthens the consistency between generated functional areas and target area proportions. Their combination improves not only individual indicators such as area error and transportation connectivity, but also the overall success rate under multiple constraints. This indicates that early-stage floor plan generation for university research buildings requires both spatial-structural guidance and global statistical control.

5.3. Analysis of Double-Blind Evaluation Results

The double-blind evaluation experiment provides a complementary assessment of the generated layouts from the perspective of architectural design practice. Unlike quantitative metrics, which mainly measure explicit constraints such as area control, boundary compliance, and transportation connectivity, expert evaluation reflects whether the generated layouts can be perceived as meaningful early-stage design proposals by architects.

The results show that the AI-generated layouts approach real design cases in the dimensions of functional spatial form and innovation and inspiration. This indicates that the proposed method can generate functional zoning patterns with a certain degree of morphological diversity and design potential, rather than merely reproducing rigid or repetitive layouts. In early-stage floor plan design, such results can provide architects with alternative spatial organization possibilities and support subsequent comparison, selection, and manual refinement.

However, the evaluation also shows that gaps remain between AI-generated layouts and real design cases in terms of spatial organization rationality. This suggests that although the model can learn statistical patterns of functional zoning and circulation distribution, it still has limitations in understanding deeper architectural logic. For example, the hierarchical relationship among functional spaces, the priority of research workflows, the coordination between public and private areas, and the detailed organization of laboratory-related circulation cannot be fully captured by pixel-level generation alone.

The analysis of influencing factors further indicates that the performance of AI-generated layouts is affected by geometric and scale conditions. Layouts with special aspect ratios, larger building areas, or higher functional complexity tend to impose greater challenges on the model. This result is consistent with architectural design practice, in which elongated sites, large-scale buildings, and multi-discipline research programs usually require more complex spatial coordination. Therefore, the double-blind evaluation not only confirms the design potential of the proposed method, but also reveals the conditions under which the current model is more likely to face limitations.

These findings clarify the role of AI in the proposed workflow. The model is not intended to replace architects' professional judgment or to generate complete architectural design schemes. Instead, it is more suitable for producing early-stage functional zoning layouts under given boundary and area constraints. Expert review remains necessary to evaluate spatial rationality, functional appropriateness, and the feasibility of further design development.

5.4. Scope and Limitations of the Evaluation Framework

The evaluation framework of this study is limited to the early-stage floor plan design of university research buildings. In this context, early-stage floor plan design refers to the initial exploratory phase of schematic design, in which architects develop preliminary spatial layouts based on design briefs, site boundary conditions, functional requirements, and quantitative area indicators. It does not refer to the formal preliminary design stage, nor does it aim to generate complete architectural design documents. Therefore, the generated results in this study should be understood as functional zoning layouts that support scheme exploration, comparison, and subsequent manual refinement by architects.

Based on this research scope, the purpose of the evaluation is not to verify the correctness of a complete architectural design scheme. Instead, it examines whether the generated layouts satisfy the main requirements of early-stage floor plan generation. The quantitative evaluation mainly focuses on functional area control, building boundary compliance, horizontal transportation connectivity, and distribution-level image quality. These indicators can measure whether the generated layouts respond to explicit geometric and statistical constraints. For example, area error reflects the consistency between generated functional area proportions and input conditions; building boundary IoU evaluates whether the generated layout conforms to the given building outline; horizontal transportation connectivity measures whether the circulation area forms a relatively continuous organizational structure; and FID provides a supplementary evaluation of the distributional similarity between generated and real floor plan samples.

However, these quantitative metrics cannot fully represent the overall quality of architectural design. Architectural floor plan design involves not only geometric boundaries and functional area proportions, but also structural systems, fire safety, environmental performance, equipment organization, construction feasibility, and user experience. These aspects often require detailed architectural drawings, engineering information, simulation models, or post-occupancy feedback, which are beyond the scope of the functional color-block layouts generated in this study.

The expert-based double-blind evaluation complements the quantitative metrics by introducing professional judgment from architects. It evaluates whether the generated layouts show reasonable spatial organization tendencies, appropriate functional spatial forms, and potential for design inspiration. Nevertheless, this evaluation is still based on simplified functional zoning diagrams rather than complete architectural drawings. Therefore, it can only provide an initial assessment of design potential at the early scheme exploration stage. It cannot replace professional review in later schematic design, formal preliminary design, or construction document stages.

Specifically, the current evaluation framework cannot fully assess several important aspects of architectural design. First, it cannot verify the rationality of structural grids, column arrangements, material systems, or construction details. Second, it cannot conduct code-based checks related to fire evacuation distance, emergency exits, laboratory safety zoning, clean and contaminated circulation, or hazardous material management. Third, it cannot evaluate the detailed organization of mechanical, electrical, plumbing, and laboratory equipment systems. Fourth, although daylighting and ventilation potential can be preliminarily inferred from spatial layout, accurate environmental performance still requires daylight simulation, ventilation simulation, energy modeling, and other quantitative environmental analyses. Fifth, user experience, behavioral adaptability, and long-term operational performance cannot be fully evaluated without user surveys, behavioral observation, or post-occupancy evaluation.

Therefore, the proposed evaluation framework should be regarded as an early-stage assessment method for AI-generated functional zoning layouts rather than a comprehensive architectural design evaluation system. The generated results are intended to assist architects by expanding the range of possible layout configurations under given boundary and area constraints. Future research should further integrate environmental simulation, rule-based code checking, structural and equipment constraints, user-behavior simulation, and post-occupancy evaluation to construct a more comprehensive evaluation framework for AI-assisted architectural design.

6. Conclusions

This study proposes a generation workflow based on a conditional diffusion model for the early-stage floor plan design of university research buildings under complex functional constraints and spatial organization requirements. The proposed method performs diffusion denoising directly in a two-dimensional grid space and achieves coordinated control over building outlines and functional area proportions through the dual-condition injection of building boundary conditions and a functional area matrix. On this basis, this study further introduces a two-stage layout generation mechanism and an explicit constraint guided by a statistic network to improve the model's control over transportation organization and global area distribution.

The experimental results show that the proposed method outperforms the selected comparison models in quantitative evaluation and achieves a favorable level of subjective evaluation in the double-blind evaluation experiment. The model comparison experiment shows that the proposed method obtains the best results in terms of FID, building

boundary IoU, and Horizontal transportation connectivity. The ablation experiment shows that both the two-stage generation mechanism and the explicit statistical constraint can significantly improve the model's ability to generate feasible schemes under multiple constraints. The double-blind evaluation results further indicate that AI-generated floor plans have approached real cases in the dimensions of functional spatial form and innovation and inspiration. However, there remains a gap between AI-generated floor plans and real designs in terms of spatial organization rationality, indicating that the current model still has room for improvement in organizing complex functional relationships and coordinating circulation.

An important advantage of the proposed method is that its final output is not limited to raster image results, but can be further converted into DXF files organized into layers by different functional areas. This format allows the generated results to be directly imported into a CAD environment, where architects can continue editing, modifying, and developing the schemes. In this way, a human–AI collaborative workflow from AI generation to manual adjustment is established. Compared with generative methods that only provide conceptual image results, this editable structured output is closer to architectural design practice and is more conducive to the application of the model in actual early-stage floor plan design.

At the same time, this study still has several limitations. First, the spatial forms in the current generated results are still dominated by relatively regular rectangular spaces. In real research buildings, open communication spaces, lecture halls, and other functions often have more complex and freer geometric forms. The existing model remains insufficient in representing irregular spaces. Second, university research buildings have clear disciplinary differences, and different disciplinary types correspond to different functional compositions and spatial organization patterns. Although this study excludes some special disciplinary types during data screening to improve the consistency of the research objects, it is still necessary to further introduce disciplinary semantic information in future work and explore categorized generation mechanisms for different types of research buildings. Third, the current research object mainly remains at the level of functional color-block diagrams. Although the generated results are suitable for layout analysis and scheme exploration, they still have certain limitations in terms of architects' intuitive understanding and design representation. Future research should further train a generative model from functional color-block diagrams to architectural floor plans, so as to obtain output results that are more intuitive and closer to real design representation.

Overall, the conditional diffusion model in this study can generate feasible and design-inspiring floor plan schemes for university research buildings under complex functional and boundary constraints. It also supports architects' subsequent editing and design development through structured DXF output. This indicates that artificial intelligence can not only participate in early-stage floor plan design as a scheme generation tool, but also has the potential to further develop into a human–AI collaborative assistance system embedded in real architectural design workflows.

Author Contributions: Conceptualization, Y.L.; methodology, Z.C.; software, Z.C.; validation, Y.L.; formal analysis, Z.C.; resources, Z.W., B.L. and Y.L.; writing—original draft preparation, Z.C.; writing—review and editing, B.L., Y.L. and Z.W.; visualization, Z.C.; supervision, B.L., Y.L., and Z.W.; project administration, B.L., Y.L. and Z.W.; funding acquisition, Y.L. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by Center for Balance Architecture of Zhejiang University, grant number K-20223284.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The original contributions presented in this study are included in the article. Further inquiries can be directed to the corresponding author.

Acknowledgments: The authors would like to thank the editor and reviewers for their detailed comments.

Conflicts of Interest: Authors Zhenling Wu and Bing Li were employed by the company The Architectural Design & Research Institute of Zhejiang University Co., Ltd. The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Appendix A

Research laboratory buildings refer to buildings that provide spatial environments for scientific research, experiments, testing, and related activities. According to disciplinary attributes, they can be divided into natural sciences and social sciences. Among them, laboratory buildings in the social sciences are highly similar to ordinary office buildings in terms of spatial pattern and lack the complexity specific to laboratory buildings. Therefore, they are not included in this study. Natural sciences include Mathematical Sciences, Physical Sciences, Chemical and Materials Sciences, Astronomical Sciences, Earth and Environmental Sciences, Biological and Medical Sciences, and Engineering and Technology. Some disciplines are excluded because of their extreme spatial forms or strict isolation requirements. Specifically, biological and medical laboratories often involve cleanliness-level control and animal room requirements, and their floor plan circulation organization, including people, materials, animals, and waste, is highly complex and exclusive. Laboratories in mechanical engineering, civil engineering, power engineering, and related fields within Engineering and Technology usually require very large spaces, heavy loads, crane beams, wind tunnels, or other special facilities. Special physics laboratories must accommodate large precision instruments and have strict requirements for vibration isolation and electromagnetic shielding, making them highly customized spaces. Modern mathematics and astronomy research mainly relies on computing platforms or general office environments. Unless precision instrument development is involved, their building requirements are closer to ordinary office buildings or supercomputing centers and therefore do not constitute typical objects for laboratory-building floor plan research.

Based on the above screening logic, this study focuses on disciplinary types that can reflect the common characteristics of laboratory-building spatial organization, mainly including Chemical and Materials Sciences, Earth and Environmental Sciences, basic physics laboratories, and some Engineering and Technology laboratories. The specific composition is shown in Table A1.

The university research building floor plan samples used in this study include 600 cases. They are derived from real built projects and unbuilt competition schemes from architectural design institutes, including the Architectural Design and Research Institute of Zhejiang University. The dataset covers multiple types, including single-discipline research buildings and interdisciplinary research buildings. As shown in Table A2, the training samples include 269 cases of Chemical and Materials Sciences laboratory buildings, 59 cases of Earth and Environmental Sciences laboratory buildings, 105 cases of Physical Sciences laboratory buildings, and 167 cases of Engineering and Technology laboratories. The samples mainly come from Zhejiang University and its affiliated units, including 248 cases from Zhejiang University and its affiliated units, as shown in Table A3. Other universities also provide a considerable number of cases, including Zhejiang Sci-Tech University, Hangzhou Dianzi University, Anhui University, Zhejiang Chinese Medical University, Northeastern Univer-

sity, Westlake University, Zhejiang Normal University, Guangdong Medical University, Shandong Normal University, Zhejiang A&F University, and Ningbo University.

Table A1. Laboratory types with common spatial organization characteristics of laboratory buildings.

Laboratory Building Category	Laboratory Type	Basic Laboratory Configuration
Chemical & Materials Sciences	Basic chemistry laboratory	Inorganic chemistry laboratory; organic chemistry laboratory; analytical chemistry laboratory; physical chemistry laboratory.
	Materials science and engineering laboratory	Materials synthesis and preparation laboratory; materials characterization and testing laboratory; polymer materials laboratory; nanomaterials laboratory.
Earth & Environmental Sciences	Geomaterial analysis laboratory	Rock and mineral sample preparation and analysis laboratory; isotope geochemistry laboratory; micro-area analysis laboratory.
	Environmental science laboratory	Environmental chemistry analysis laboratory; environmental biology and ecology laboratory; environmental monitoring and simulation laboratory.
	Geology and geography laboratory	General geology and mineralogy laboratory; structural geology laboratory; remote sensing and geographic information systems laboratory.
Physical Sciences	Basic physics laboratory	Mechanics laboratory; thermodynamics laboratory; electromagnetics laboratory; optics laboratory.
Engineering & Technology	Electronics and communication laboratory	Microelectronics and integrated circuits laboratory; embedded systems and circuit design laboratory; radio-frequency and microwave laboratory; communication networks laboratory.
	Optoelectronics and precision instrument laboratory	Laser technology laboratory; precision measurement laboratory; optical fiber communication laboratory.
	Control and automation laboratory	Automatic control principles laboratory; sensor laboratory.

Table A2. Laboratory building categories and sample quantities.

Laboratory Building Category	Number of Samples
Chemical & Materials Sciences	269
Earth & Environmental Sciences	59
Physical Sciences	105
Engineering & Technology	167

Table A3. Universities represented in the laboratory building cases.

University or Campus	Number of Samples
Zhejiang University Zijingang Campus	52
Zhejiang University Yuquan Campus	28
Affiliated units of Zhejiang University, including Zhejiang University International School of Medicine, Ningbo Campus, Haining International Campus, and Zhejiang University International Innovation Institute	168
Zhejiang Sci-Tech University	34
Hangzhou Dianzi University	31
Zhejiang Chinese Medical University	27
Westlake University	21
Zhejiang Normal University	19
Zhejiang A&F University	17
Ningbo University	13
Anhui University	27

Table A3. Cont.

University or Campus	Number of Samples
Northeastern University	26
China University of Mining and Technology	19
Guangdong Medical University	19
Shandong Normal University	16
Dalian University of Technology	7
Xidian University	6
Other universities with fewer than five cases	70

References

1. Lawson, B. *How Designers Think: The Design Process Demystified*, 4th ed.; Routledge: London, UK, 2005.
2. Mitchell, W.J. *The Logic of Architecture: Design, Computation, and Cognition*; MIT Press: Cambridge, MA, USA, 1990.
3. Oxman, R. Theory and design in the first digital age. *Des. Stud.* **2006**, *27*, 229–265. [\[CrossRef\]](#)
4. Licklider, J.C.R. Man-computer symbiosis. *IRE Trans. Hum. Factors Electron.* **1960**, *HFE-1*, 4–11. [\[CrossRef\]](#)
5. Eastman, C.; Teicholz, P.; Sacks, R.; Liston, K. *BIM Handbook: A Guide to Building Information Modeling for Owners, Managers, Designers, Engineers and Contractors*, 2nd ed.; John Wiley & Sons: Hoboken, NJ, USA, 2011.
6. Terzidis, K. *Algorithmic Architecture*; Architectural Press: Oxford, UK, 2006.
7. Schumacher, P. Parametricism: A new global style for architecture and urban design. *Archit. Des.* **2009**, *79*, 14–23. [\[CrossRef\]](#)
8. Schumacher, P. *The Autopoiesis of Architecture: A New Framework for Architecture*; John Wiley & Sons: Chichester, UK, 2011.
9. Albukhari, I.N. The role of artificial intelligence (AI) in architectural design: A systematic review of emerging technologies and applications. *J. Umm Al-Qura Univ. Eng. Archit.* **2025**, *16*, 1457–1476. [\[CrossRef\]](#)
10. Chaillou, S. *Artificial Intelligence and Architecture: From Research to Practice*; De Gruyter: Berlin, Germany, 2022. [\[CrossRef\]](#)
11. Yuan, F.; Xu, X.; Wang, Y. Towards the era of generative artificial intelligence augmented design. *Archit. J.* **2023**, *10*, 14–20.
12. Dhariwal, P.; Nichol, A. Diffusion models beat GANs on image synthesis. *Proc. Adv. Neural Inf. Process. Syst.* **2021**, *34*, 8780–8794.
13. Huang, W.; Zheng, H. Architectural drawings recognition and generation through machine learning. In Proceedings of the 38th Annual Conference of the Association for Computer Aided Design in Architecture (ACADIA): Re/Calibration: On Imprecision and Infidelity, Mexico City, Mexico, 18–20 October 2018; pp. 156–165. [\[CrossRef\]](#)
14. Goldstein, R.N. Architectural design and the collaborative research environment. *Cell* **2006**, *127*, 243–246. [\[CrossRef\]](#)
15. Blackwell, B. The architectures of secrecy: Negotiating openness and privacy in buildings of science and technology. *Sci. Technol. Hum. Values* **2025**, *50*, 1072–1103. [\[CrossRef\]](#)
16. Sanni-Anibire, M.O.; Hassanain, M.A.; Mahmoud, A.S.; Al-Hammad, A.-M. An evaluation of the functional performance of research and academic laboratories using the space syntax approach. *Int. J. Build. Pathol. Adapt.* **2018**, *36*, 516–528. [\[CrossRef\]](#)
17. Wang, G.; Zhang, Y.; Wang, W. Evaluating the performance of informal learning spaces in higher education: An integrated methodological framework combining Space Syntax and Post-Occupancy Evaluation. *J. Asian Archit. Build. Eng.* **2026**, *25*, 1–20. [\[CrossRef\]](#)
18. Zhang, X.; Cui, T. Evolution process of scientific space: Spatial analysis of three groups of laboratories in history (16th–20th century). *Buildings* **2022**, *12*, 1909. [\[CrossRef\]](#)
19. Hammadamin, A.B.; Nordin, J.; Mustafa, F.A. Interpretation of space syntax in higher education: A study of functional efficiency in architecture schools in Erbil. *Sustainability* **2024**, *16*, 11237. [\[CrossRef\]](#)
20. Ajtayné Károlyfi, K.; Szép, J. A parametric BIM framework to conceptual structural design for assessing the embodied environmental impact. *Sustainability* **2023**, *15*, 11990. [\[CrossRef\]](#)
21. Hong, S.W.; Lee, J.; Lee, J.K. Human behaviour simulation for promoting usefulness and user-centric values in parametric design. *Autom. Constr.* **2024**, *162*, 105386. [\[CrossRef\]](#)
22. Li, C.; Zhang, T.; Du, X.; Zhang, Y.; Xie, H. Generative AI models for different steps in architectural design: A literature review. *Front. Archit. Res.* **2025**, *14*, 759–783. [\[CrossRef\]](#)
23. Iqbal, K.; Rafique, A.; Qaisar, S.; Tabassum, M. Advancements and challenges in the development of generative adversarial network (GANs) for deep learning. *Discov. Netw.* **2025**, *1*, 11. [\[CrossRef\]](#)
24. Wang, S.; Zeng, W.; Chen, X.; Ye, Y.; Qiao, Y.; Fu, C.-W. ActFloor-GAN: Activity-guided adversarial networks for human-centric floorplan design. *IEEE Trans. Vis. Comput. Graph.* **2023**, *29*, 1610–1624. [\[CrossRef\]](#)
25. Hu, R.; Huang, Z.; Tang, Y.; van Kaick, O.; Zhang, H.; Huang, H. Graph2Plan: Learning floorplan generation from layout graphs. *ACM Trans. Graph.* **2020**, *39*, 118. [\[CrossRef\]](#)
26. Dong, S.; Wang, W.; Li, W.; Lin, K.; Gao, J. Vectorization of floor plans based on edge GAN. *Information* **2021**, *12*, 206. [\[CrossRef\]](#)

27. Upadhyay, A.; Dubey, A.; Bhardwaj, N.; Kuriakose, S.M. CIGMA: Automated 3D house layout generation through generative models. In Proceedings of the 7th Joint International Conference on Data Science & Management of Data (11th ACM IKDD CODS and 29th COMAD), Bangalore, India, 4–7 January 2024; pp. 542–546. [\[CrossRef\]](#)
28. Isola, P.; Zhu, J.-Y.; Zhou, T.; Efros, A.A. Image-to-image translation with conditional adversarial networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 1125–1134. [\[CrossRef\]](#)
29. Nauata, N.; Chang, K.-H.; Cheng, C.-Y.; Mori, G.; Furukawa, Y. House-GAN: Relational generative adversarial networks for graph-constrained house layout generation. In Proceedings of the European Conference on Computer Vision, Glasgow, UK, 23–28 August 2020; pp. 162–177. [\[CrossRef\]](#)
30. Brooks, T.; Holynski, A.; Efros, A.A. InstructPix2Pix: Learning to follow image editing instructions. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Vancouver, BC, Canada, 18–22 June 2023; pp. 18392–18402. [\[CrossRef\]](#)
31. Nauata, N.; Hosseini, S.; Chang, K.-H.; Chu, H.; Cheng, C.-Y.; Furukawa, Y. House-GAN++: Generative adversarial layout refinement network towards intelligent computational agent for professional architects. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Virtual, 19–25 June 2021; pp. 13632–13641.
32. Ye, X.; Du, J.; Ye, Y. MasterplanGAN: Facilitating the smart rendering of urban master plans via generative adversarial networks. *Environ. Plan. B Urban Anal. City Sci.* **2022**, *49*, 794–814. [\[CrossRef\]](#)
33. Upadhyay, A.; Dubey, A.; Arora, V.; Kuriakose, S.M. FLNet: Graph constrained floor layout generation. In Proceedings of the 2022 IEEE International Conference on Multimedia and Expo Workshops, Taipei, Taiwan, 18–22 July 2022; pp. 1–6. [\[CrossRef\]](#)
34. Kakooee, R.; Dillenburger, B. Reimagining space layout design through deep reinforcement learning. *J. Comput. Des. Eng.* **2024**, *11*, 43–55. [\[CrossRef\]](#)
35. Wang, X.-Y.; Liu, Y.; Zhang, K. A graph grammar approach to the design and validation of floor plans. *Comput. J.* **2020**, *63*, 137–150. [\[CrossRef\]](#)
36. Luo, G.; Zhou, X.; Liao, Y.; Ding, Y.; Liu, J.; Xia, Y.; Qi, H. Automated residential bubble diagram generation based on dual-branch graph neural network and variational encoding. *Appl. Sci.* **2025**, *15*, 4490. [\[CrossRef\]](#)
37. Meselhy, A.; Almalkawi, A. A review of artificial intelligence methodologies in computational automated generation of high performance floorplans. *npj Clean. Energy* **2025**, *1*, 2. [\[CrossRef\]](#)
38. Zeng, P.; Yin, J.; Zhang, M.; Li, J.; Zhang, Y.; Lu, S. Unified residential floor plan generation with multimodal inputs. *Autom. Constr.* **2025**, *178*, 106408. [\[CrossRef\]](#)
39. Qiu, Z.; Liu, J.; Wu, Y.; Liu, P.; Qi, H.; Liang, H.; Xia, Y. LLM-based framework for automated and customized floor plan design. *Autom. Constr.* **2025**, *180*, 106512. [\[CrossRef\]](#)
40. Barsha, F.L.; Eberle, W. An in-depth review and analysis of mode collapse in generative adversarial networks. *Mach. Learn.* **2025**, *114*, 141. [\[CrossRef\]](#)
41. Nichol, A.Q.; Dhariwal, P. Improved denoising diffusion probabilistic models. In Proceedings of the 38th International Conference on Machine Learning, Virtual, 18–24 July 2021; pp. 8162–8171.
42. Liu, R.; Wu, R.; Van Hoorick, B.; Tokmakov, P.; Zakharov, S.; Vondrick, C. Zero-1-to-3: Zero-shot one image to 3D object. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Paris, France, 1–6 October 2023; pp. 9264–9275. [\[CrossRef\]](#)
43. Shabani, M.A.; Hosseini, S.; Furukawa, Y. HouseDiffusion: Vector floorplan generation via a diffusion model with discrete and continuous denoising. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Vancouver, BC, Canada, 18–22 June 2023; pp. 5466–5475. [\[CrossRef\]](#)
44. Huang, S.; Wang, Z.; Li, P.; Jia, B.; Liu, T.; Zhu, Y.; Liang, W.; Zhu, S.-C. Diffusion-based generation, optimization, and planning in 3D scenes. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Vancouver, BC, Canada, 18–22 June 2023; pp. 16750–16761. [\[CrossRef\]](#)
45. Gueze, A.; Ospici, M.; Rohmer, D.; Cani, M.-P. Floor plan reconstruction from sparse views: Combining graph neural network with constrained diffusion. In Proceedings of the IEEE/CVF International Conference on Computer Vision Workshops, Paris, France, 2–6 October 2023; pp. 1583–1592. [\[CrossRef\]](#)
46. Zhang, H.; Zhang, R. Generating accessible multi-occupancy floor plans with fine-grained control using a diffusion model. *Autom. Constr.* **2025**, *177*, 106332. [\[CrossRef\]](#)
47. Zeng, P.; Gao, W.; Yin, J.; Xu, P.; Lu, S. Residential floor plans: Multi-conditional automatic generation using diffusion models. *Autom. Constr.* **2024**, *162*, 105374. [\[CrossRef\]](#)
48. Hu, E.J.; Shen, Y.; Wallis, P.; Allen-Zhu, Z.; Li, Y.; Wang, S.; Wang, L.; Chen, W. LoRA: Low-rank adaptation of large language models. *arXiv* **2021**, arXiv:2106.09685.
49. Zeng, P.; Gao, W.; Li, J.; Yin, J.; Chen, J.; Lu, S. Automated residential layout generation and editing using natural language and images. *Autom. Constr.* **2025**, *174*, 106133. [\[CrossRef\]](#)

50. Wang, S.; Pajarola, R. Eliminating rasterization: Direct vector floor plan generation with DiffPlanner. *IEEE Trans. Vis. Comput. Graph.* **2025**, *31*, 7906–7922. [[CrossRef](#)]
51. Knechtel, J.; Rottmann, P.; Haunert, J.-H.; Dehbi, Y. Semantic floorplan segmentation using self-constructing graph networks. *Autom. Constr.* **2024**, *166*, 105649. [[CrossRef](#)]
52. Hassanain, M.A.; Sanni-Anibire, M.O.; Mahmoud, A.S.; Ahmed, W. Design guidelines for the functional efficiency of laboratory facilities. *Archit. Eng. Des. Manag.* **2020**, *16*, 115–130. [[CrossRef](#)]
53. Ho, J.; Jain, A.; Abbeel, P. Denoising diffusion probabilistic models. In Proceedings of the Advances in Neural Information Processing Systems, Virtual, 6–12 December 2020; Volume 33, pp. 6840–6851.
54. Ronneberger, O.; Fischer, P.; Brox, T. U-Net: Convolutional networks for biomedical image segmentation. In Proceedings of the Medical Image Computing and Computer-Assisted Intervention, Munich, Germany, 5–9 October 2015; pp. 234–241. [[CrossRef](#)]
55. Rombach, R.; Blattmann, A.; Lorenz, D.; Esser, P.; Ommer, B. High-resolution image synthesis with latent diffusion models. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, New Orleans, LA, USA, 18–24 June 2022; pp. 10684–10695. [[CrossRef](#)]
56. Para, W.; Guerrero, P.; Kelly, T.; Guibas, L.J.; Wonka, P. Generative layout modeling using constraint graphs. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Montreal, QC, Canada, 10–17 October 2021; pp. 6690–6700. [[CrossRef](#)]
57. He, F.; Huang, Y.; Wang, H. iPLAN: Interactive and procedural layout planning. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, New Orleans, LA, USA, 18–24 June 2022; pp. 7793–7802.
58. Li, L.; Su, X.; Lin, H.; Han, H.; Fan, C.; Zhang, Z.; Yue, H. ChatAssistDesign: A language-interactive framework for iterative vector floorplan generation via conditional diffusion. *Inf. Fusion* **2026**, *130*, 104091. [[CrossRef](#)]
59. Spearman, C. The proof and measurement of association between two things. *Am. J. Psychol.* **1904**, *15*, 72–101. [[CrossRef](#)]
60. Mann, H.B.; Whitney, D.R. On a test of whether one of two random variables is stochastically larger than the other. *Ann. Math. Stat.* **1947**, *18*, 50–60. [[CrossRef](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.