# GardenDesigner: Encoding Aesthetic Principles into Jiangnan Garden Construction via a Chain of Agents

Mengtian Li[1,2], Fan Yang[1], Ruixue Xiong[1], Yiyan Fan[1], Zhifeng Xie[1,2†], Zeyu Wang[3†]
[1]Shanghai University
[2]Shanghai Engineering Research Center of Motion Picture Special Effects
[3]The Hong Kong University of Science and Technology (Guangzhou)
{mtli, yangphan, xiongruixue, yiyanfan, zhifeng_xie}@shu.edu.cn, zeyuwang@ust.hk

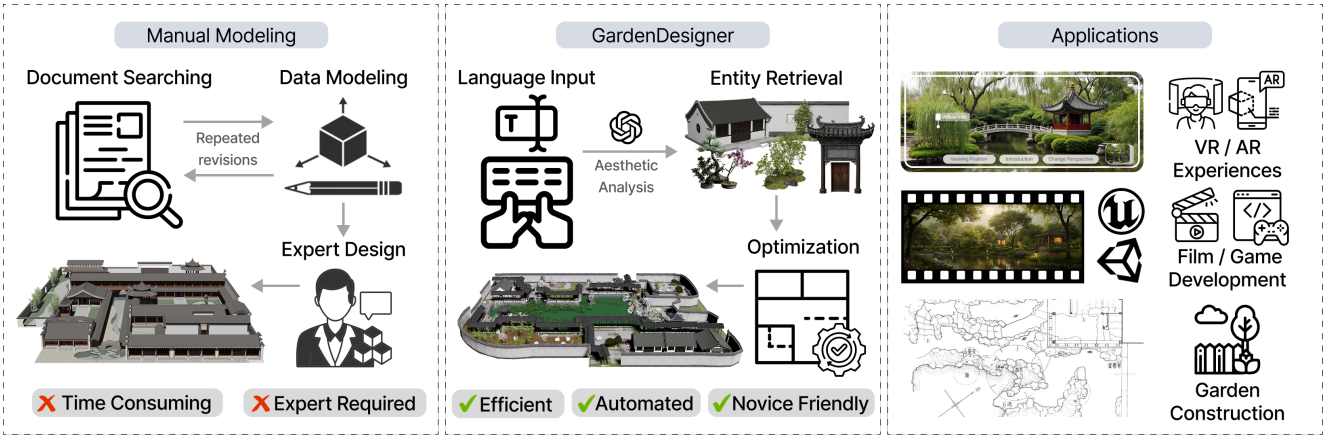


Figure 1. The motivation of GardenDesigner. Traditional manual modeling of Jiangnan gardens requires document searching, data modeling, and expert design, making it time-consuming and expertise-dependent. GardenDesigner automates Jiangnan garden construction via analyzing the user text and acquiring the assets, and then optimizes the garden layout. For applications, users can construct a Jiangnan garden through text input, which can be used for creating VR/AR experiences, film and game development, and real garden construction.

## Abstract

*Jiangnan gardens, a prominent style of Chinese classical gardens, hold great potential as digital assets for film and game production and digital tourism. However, manual modeling of Jiangnan gardens heavily relies on expert experience for layout design and asset creation, making the process time-consuming. To address this gap, we propose* ***GardenDesigner****, a novel framework that encodes aesthetic principles for Jiangnan garden construction and integrates a chain of agents based on procedural modeling. The water-centric terrain and explorative pathway rules are applied by terrain distribution and road generation agents. Selection and spatial layout of garden assets follow the aesthetic and cultural constraints. Consequently, we propose asset selection and layout optimization agents to select and arrange objects for each area in the garden. Additionally, we introduce* ***GardenVerse*** *for Jiangnan garden construction, including expert-annotated garden knowledge to enhance the asset arrangement process. To enable interaction and editing, we develop an interactive interface and tools in Unity, in which non-expert users can construct Jiangnan gardens via text input within one minute. Experiments and human evaluations demonstrate that GardenDesigner can generate diverse and aesthetically pleasing Jiangnan gardens. Project page is available at* https://monad-cube.github.io/GardenDesigner.

## 1. Introduction

As the most important genres of Chinese classical gardens, Jiangnan gardens exemplify compact urban compositions with intricate spatial configurations [6]. Unlike general landscape parks, they emphasize a balance of architecture, plants, and rocks. Typical features include winding

† Corresponding authors.

corridors, attics, pavilions that frame ever-changing views, rockeries that simulate mountains within limited space, and ponds that reflect both natural scenery and surrounding structures [30]. The traditional construction of Jiangnan gardens involves much manual effort, including three main steps: (1) document search, collecting historical documents, drawings, and photographs; (2) asset modeling, reconstructing architectural elements and plants based on these materials; (3) expert design, addressing terrain shaping and garden layout, relying on specialized knowledge. However, this process typically involves three to four designers and takes about three to four weeks to complete, making it heavily reliant on manual effort and time-consuming.

Current learning-based scene generation methods [16, 26, 50] exhibit limited generalizability due to domain constraints in training datasets. Procedural modeling methods [24, 37, 47] that incorporate large language models (LLMs) or visual language models (VLMs) focus on either spatially limited room space or unstructured natural environments. However, the construction of Jiangnan gardens remains unexplored, and three problems remain to be addressed. **(1) Complex terrain and garden layout**: Compared to general landscapes, Jiangnan gardens exhibit intricate terrain structures and spatial layouts, where terrain, water, and architecture are interwoven under implicit aesthetic logic. **(2) Aesthetic principle constraints**: Due to the abstract nature of Jiangnan gardens' design rules, encoding the aesthetic principles into a computational generation framework remains challenging. **(3) Absence of Jiangnan garden dataset**: Lacking stylistic appearance and cultural annotation, existing 3D datasets of ordinary or urban objects are not suitable for Jiangnan garden construction.

To address these challenges, we propose **GardenDesigner**, which integrates a chain of agents, procedural modeling, and aesthetic principles encoding for Jiangnan gardens construction. Specifically, GardenDesigner is composed of three modules: Hierarchical Garden Composition (Section 3.2), Knowledge-embedded Asset Arrangement (Section 3.3). First, Hierarchical Garden Composition decomposes the construction process into procedural terrain and road generation. Subsequently, Knowledge-embedded Asset Arrangement focuses on asset selection and optimizing objects according to the specified constraints for each area. The key insight is to select objects and set constraints according to area information and expert-guided garden knowledge. Consequently, we introduce **GardenVerse**, a high-quality Jiangnan garden dataset that contains typical Jiangnan garden style of digital assets with expert-annotated garden knowledge, enhancing the specific knowledge context for knowledge-embedded asset arrangement.

To support convenient designing and interaction, we develop an interface and editing tools in Unity, in which the non-expert user can construct Jiangnan gardens via text input within one minute. After construction, the system can output the 2D garden layout as a reference for the real garden creation and building. In summary, GardenDesigner opens new avenues for intangible cultural heritage preservation and creative applications in digital art and games.

Our main contributions are as follows:

- We propose **GardenDesigner**, a novel framework that encodes aesthetic principles for Jiangnan garden construction via integrating a chain of agents with an expert-annotated artistic dataset **GardenVerse**.
- We propose a hierarchical garden composition module to generate terrain and roads with aesthetic principles, and a knowledge-embedded asset arrangement mechanism for asset selection and layout optimization.
- We develop an interface and editing tools in Unity, in which non-expert users can construct Jiangnan gardens via text input. The system outputs a 2D layout for real garden construction and supports virtual tourism.

## 2. Related Work

### 2.1. Scene Generation

**Procedural Scene Generation.** Procedurally generating scenes with rules and manual algorithms has long been a robust methodology. CityEngine [29] and Khan et al. [19] procedurally model the city. Recently, Raistrick et al. [31, 32] generates assets in scenes from shape to texture.

**Data-Driven Scene Generation.** Previous learning-based methods have explored different modalities to generate scenes, including images [48], texts [16], layouts [1], scene graphs [50] and raw room [49], while some methods [45, 46] extend to large-scale city generation.

**Scene Generation with LLMs and VLMs.** Feng et al. [13] takes the first step to utilize LLMs to generate object position, while some methods [4, 14, 47] generate the scene graph. Other methods [17, 25, 35, 52] explore the outdoor generation based on Blender [3] or Infinigen [31]. Liu et al. [24] procedurally model the landscape with LLMs. Feng et al. [13] adapt VLM to optimize indoor layout and some methods [2, 23] explored simple outdoor scene.

Previous methods have primarily focused on ordinary indoor spaces or unstructured landscapes. In contrast, generating Jiangnan gardens poses unique challenges, requiring fine-grained spatial composition, hierarchical reasoning, and the integration of aesthetic and cultural principles.

### 2.2. 3D Object Datasets

**Indoor and Ordinary Objects.** ShapeNet [5] collects 3D CAD models from public repositories and previous datasets. GSO [10] offers scans of household objects, and OmniObject3D [44] expands both quantity and diversity. Objaverse-XL [7] extends Objaverse [8] to 10.2M 3D assets. However, existing datasets lack sufficient diversity or

fidelity for cultural scenes such as Jiangnan gardens.

**Outdoor and Natural Objects.** BuildingNet [33] and CityCraft [9] mine the architectures from websites [34, 38]. Zhu et al. [53] collects a scanned 3D crops dataset and some methods [21, 52] create architectural or natural assets with Unreal [12] or Blender [3]. Procedural modeling methods [19, 29] employ parametric or L-system rules for virtual cities. Other works [15, 22] simulate vegetation, and Infinigen [31] extends to large-scale natural textured assets.

Despite extensive research on architectural and natural objects, Jiangnan gardens featuring traditional architecture, distinctive flora, and rocks remain underexplored. Existing datasets lack the stylistic coherence, cultural context, and fine-grained diversity needed for heritage-oriented scenes.

### 2.3. Cultural Heritage and Digital Tourism

Cultural Heritage (CH) encompasses tangible and intangible artifacts, traditions, and environments, in which interactive systems transform preservation from passive documentation to active participation.To foster public engagement, prior works have explored diverse cultural heritage applications, including immersive cultural tourism [20], underwater heritage exploration [51], and interactive historical storytelling and artifact preservation [18].

These applications effectively employ immersion, narrative, and interaction to represent CH. However, most focus on heritage exploration and exhibition rather than the generation of heritage-inspired content. Therefore, a generative and interactive system is essential to lower the creative threshold, translating complex garden aesthetics into tangible designs through simple text input. Such an approach not only preserves the Jiangnan garden tradition but also revitalizes it as a living, participatory form of cultural heritage.

## 3. Method

### 3.1. Problem Statement

Jiangnan garden construction involves generating terrain and roads, configuring objects in a bounded space based on the user's instructions, and following certain Jiangnan garden design rules. After integrating expert experience of garden designers and literature search [6, 30], we summarize key aesthetic principles to guide Jiangnan garden construction from four perspectives of *terrain distribution*, *road generation*, *asset selection*, and *relational constraint*:

- **Naturalistic and Water-Centric Foundation:** The terrain is designed to be an idealized, miniature microcosm of a natural landscape, and the water is considered the lifeblood and the soul of the garden to organize elements.
- **Discovery and Winding Paths:** Following the water and area border, paths are designed for exploration, creating a series of unfolding, painterly scenes rather than simple transit, deliberately avoiding straight lines and symmetry.
- **Symbolism and Miniature:** Selects assets that are symbolic miniatures of the natural world. The assets should be culturally appropriate and fit their specific location, reflecting both natural content and cultural intentionality.
- **Asymmetrical Balance:** Arranges objects harmoniously, creating a dynamic and natural balance. Positional constraints are used to hide and reveal views, making the garden feel larger and encouraging exploration.

Formally, given a user text $U$, aesthetic principles $K_{\text{global}}$ and garden assets $O_{\text{asset}} = \{o_1, ..., o_n\}$ with knowledge annotation $K_a = \{k_1, ..., k_n\}$, the objective is to create a Jiangnan garden that meets the textual input and aesthetic principles. We decompose the Jiangnan garden construction task into four steps and implement it via a chain of agents, in which the content generated by previous agents serves as the basis for subsequent agents. (1) **Terrain Distribution Agent ($\mathcal{A}_T$):** this agent generates the terrain $T$ based on the user's text and global knowledge. (2) **Road Generation Agent ($\mathcal{A}_R$):** based on terrain $T$, this agent generates the road $R$, also guided by global knowledge. (3) **Asset Selection Agent ($\mathcal{A}_S$):** with terrain and road context $(T, R)$ available, this agent selects a set of appropriate assets $O_s$ from the library. (4) **Layout Optimization Agent ($\mathcal{A}_C$):** the final agent takes the terrain, paths, and selected objects as input, arranges the selected objects, optimizing their position and rotation. Therefore, the complete garden $G$ is the composite output derived from the chain of agents:

$$G = (T, R, (O_s, P)), \tag{1}$$

where selected objects $O_s = \{o_1, ..., o_m\}$ with properties $P = ((x_i, y_i, z_i, r_i), ...)$, representing position and rotation.

### 3.2. Hierarchical Garden Composition

**Challenges.** (1) *Water-centric spatial organization.* Conventional landscape procedural algorithms fail to capture the water-centered logic of Jiangnan gardens. As a result, they often produce scattered ponds and unnatural terrain that disrupt the intended harmony between land and water. (2) *Exploratory path generation.* Existing path-generation methods focus on geometric efficiency or uniform coverage, neglecting the exploratory routing principles of Jiangnan gardens. Thus, they cannot reproduce the winding, layered paths that define the authentic visitor experience.

To address these challenges, we introduce two agents: (1) **Terrain Distribution Agent** $A_{\text{T}}$ and (2) **Road Generation Agent** $A_{\text{R}}$. These agents leverage specific garden composition prompts, integrate a water-centric loss to guide the generation and optimization of terrain, and redesign the road scoring mechanism to encourage roads to follow terrain boundaries and avoid excessive linearity.

**Genetic Terrain Generation.** The Jiangnan garden is generally located in flat terrain areas within urban areas and occupies relatively small sites. Consequently, we adopt a

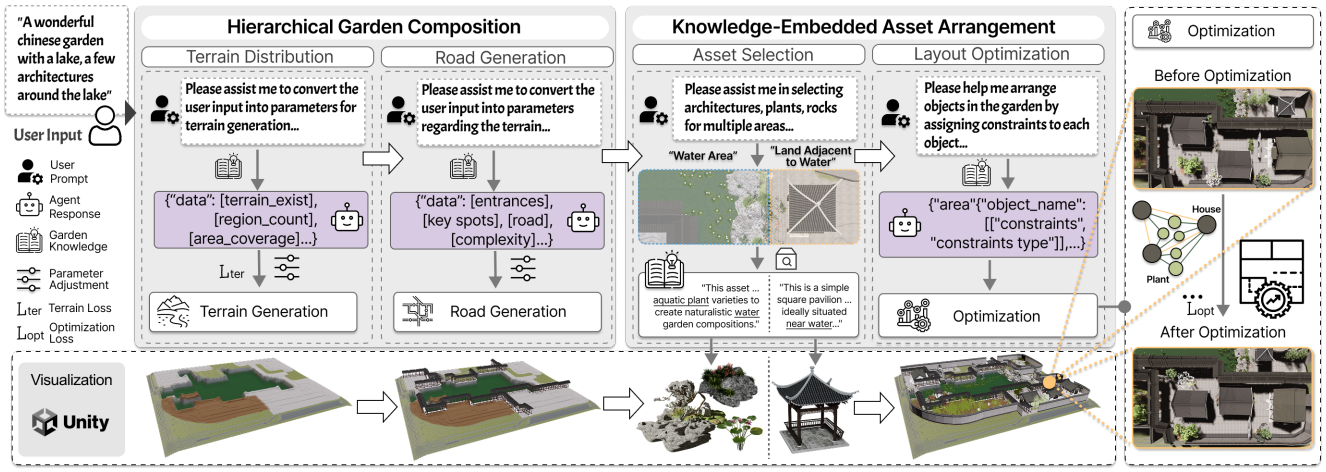


Figure 2. Overview of the GardenDesigner pipeline. GardenDesigner transforms the user input into a Jiangnan garden through Hierarchical Garden Composition and Knowledge-Embedded Asset Arrangement. First, Hierarchical Garden Composition transfers the user input into parameters for terrain and road generation with aesthetic principles. Subsequently, Knowledge-Embedded Asset Arrangement chooses the objects based on the garden knowledge and area information, and then optimization loss is used to get the feasible solution for layout.

genetic algorithm based on a 2D grid and choose four types of terrain to simulate the landform of Jiangnan garden: *Outside*, *Waterbody*, *Land*, and *Ground*, represented as integer numbers. To enable language control, $A_T$ is used to generate terrain, which transfers the text input to the parameters and then calls the genetic algorithm with the parameters:

$$T = \mathcal{A}_{\mathrm{T}}(U, K_{\mathrm{global}}), \tag{2}$$

where $U$ is the user input text, and $K_{\mathrm{global}}$ is the aesthetic principles. Specifically, we choose four types of terrain parameters: (1) existence, (2) quantity, (3) coverage, and (4) single region coverage. Based on the parameters, the genetic algorithm conducts the *Crossover*, *Mutation*, and *Evolution* operations for each iteration. Finally, the fitness function is used to select the feasible terrain solution for each iteration. Most importantly, we introduce a water-centric loss to calculate the terrain fitness as follows:

$$L_{\mathrm{terrain}} = f \cdot \max(1 - \frac{\sum_{i=0}^{n} c(T, (x_i, y_i))}{\phi}, 0), \tag{3}$$

where $T$ represents the generated terrain, $f$ is the factor, $c$ function is used to judge whether the grid is in water.

**Explorative Road Generation.** Given the discretized terrain layout, $A_R$ synthesizes roads adhering to Jiangnan aesthetic principles. We integrate cultural priors into a grid-based scoring mechanism and produces smooth spline curves for a realistic pedestrian experience and corridor arrangement. First, the agent parses the user instruction $U$ to generate the parameters, including the number of entrances and keypoints, the width of the main road, and the road complexity, which jointly determine the roads and entrances of the garden. The entrances are sampled across all directional boundaries, and then the roads are generated by scoring the grid border and selecting the best solution. Additionally, the path selection process follows the Jiangnan garden key requirements: (1) the roads can reach most of the garden area, (2) the roads prefer to follow the border, and (3) the roads should avoid excessive warping and straightening. The process is formulated as follows:

$$R = \mathcal{A}_{\mathrm{R}}(\mathcal{S}(T, e_{i,j}), U, K_{\mathrm{global}}), \tag{4}$$

where $e$ is the edge in the grid and $S$ is the scoring function according to the rules and principles.

### 3.3. Knowledge-Embedded Asset Arrangement

**Challenges.** (1) *Rule-based and aesthetic spatial logic.* Conventional retrieval or constraint methods fail to capture implicit culturally grounded relations in Jiangnan gardens, leading to aesthetically inconsistent layouts. (2) *Lack of domain-specific understanding.* General LLMs lack garden knowledge, making it hard to reason about the interplay of architectural, structural, and botanical elements, thus failing to produce layouts aligned with traditional design logic.

To tackle these challenges, we first annotate the garden asset dataset with descriptions encoding expert garden knowledge. Then, we propose a knowledge-embedded asset arrangement mechanism, consisting of knowledge-guided asset retrieval and aesthetic constraints encoding, implemented by the **Asset Selection Agent** $A_{\mathrm{S}}$ and the **Layout Optimization Agent** $A_{\mathrm{C}}$.

#### 3.3.1. Knowledge-Guided Asset Retrieval

First, we collected a Jiangnan garden dataset, GardenVerse, and then proposed a knowledge-guided agent $A_{\mathrm{S}}$ to retrieve

assets with expert-annotated garden knowledge. Specifically, we annotated the object assets with additional garden knowledge description $K_a = \{k_1, ..., k_n\}$ to provide the agents with rich knowledge about the garden objects in Section 4. We encode these annotations into a knowledge vector store and query them through a large language model to enforce culturally consistent object selection. To get appropriate objects, we provide the area information $I_{\text{area}} = \{i_1, ...i_k\}$ and garden knowledge for LLM, then agent will response a list of object $O_{\text{s}} = \{o_1, ...o_m\}$ for object arrangement of each area, as follows:

$$O_{\text{s}} = \mathcal{A}_{\text{S}}(\mathcal{Q}((\mathcal{V}(K_{\text{a}}), o_i), U), I_{\text{area}}), \tag{5}$$

where $\mathcal{Q}$ is the query operation, $\mathcal{V}$ is the process to vector store, $o_i$ refers to each object and $i \in \{0, ...m\}$.

#### 3.3.2. Aesthetic Constraints Encoding

To address the challenge of inconsistent aesthetic constraints, we set the constraints for selected objects and then optimize the layout according to the constraints. Specifically, we define eight constraint types and group them into five semantic categories according to their spatial position, direction relationship with the boundary and objects: (1) Global (edge, middle) indicates the overall placement within the entire scene; (2) Position (around, backed up) captures relative placement relationships; (3) Distance (near, far) quantifies spatial proximity; (4) Alignment (aligned) enforces consistent directional orientation among objects; and (5) Rotation (face to) specifies the facing direction of an object toward another.

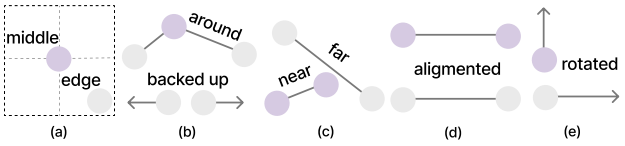


Figure 3. The five constraints categories: (a) Global, edge, and middle; (b) Position, around, and backed up; (c) Distance, near and far; (d) Alignment, aligned; And (e) Rotation, face to.

**Optimization.** To generate the garden layout, we design five types of optimization loss functions, corresponding to different categories of spatial constraints. The position and direction for each object is represented as $o_i = (x_i, y_i, z_i, \theta_i)$ and the bounding box is $b_i = (l_i, w_i, h_i)$. We formulate the optimization loss as follows.

*Global Objective* is used to decide the global position and optimize objects to the edge or middle of an area:

$$\mathcal{L}_{\text{glo}} = \begin{cases} \max\left(\frac{d(o_i, e_{\text{area}}) - d_{\text{e}}}{d_{\text{e}}}, 0\right), & \text{if edge,} \\ \max\left(\frac{\|o_i - c_{\text{area}}\| - d_{\text{m}}}{d_{\text{m}}}, 0\right), & \text{if middle,} \end{cases} \tag{6}$$

where $e_{\text{area}}$ and $c_{\text{area}}$ are the boundary and the center, $d_{\text{e}}$ and $d_{\text{m}}$ are the threshold value parameters. The $d$ is used to calculate the distance between a point and an area boundary.

*Position Objective* loss focuses on the relative position and direction between two different objects:

$$\mathcal{L}_{\text{pos}} = \begin{cases} m\left(r_{\text{l}} - d, 0\right) + m\left(d - r_{\text{h}}, 0\right), & \text{if around,} \\ f_{\text{back}} \cdot f(o_i, o_j, \theta), & \text{if backed up,} \end{cases} \tag{7}$$

where $m$ is the max function, $r_{\text{l}}$ and $r_{\text{h}}$ represent the low and high threshold. $d$ is the distance between two objects. $f_{\text{back}}$ is the parameter, $\theta$ is the front orientation of $o_j$ and $f$ is deciding if $o_i$ is backed up $o_j$.

*Distance Objective* is used to control and adjust the relative distance between different objects:

$$\mathcal{L}_{\text{dis}} = \begin{cases} \max\left(\frac{\|o_i - o_j\| - d_{\text{n}}}{d_{\text{n}}}, 0\right), & \text{if near,} \\ \max\left(\frac{d_{\text{f}} - \|o_i - o_j\|}{d_{\text{f}}}, 0\right), & \text{if far,} \end{cases} \tag{8}$$

where $d_{\text{n}}$ and $d_{\text{f}}$ are the near and far parameters from object $o_i$ and another object $o_j$. $\epsilon$ is the threshold value parameter.

*Alignment Objective* attempts to align objects of the same type for neat and regular local arrangement:

$$\mathcal{L}_{\text{ali}} = \max(\frac{|x_i - x_j| - \epsilon}{\epsilon}, 0) + \max(\frac{|y_i - y_j| - \epsilon}{\epsilon}, 0), \tag{9}$$

where $x$ and $y$ are the positions of two objects and $\epsilon$ represents the threshold value for alignment.

*Rotation Objective* is used to adjust object direction:

$$\mathcal{L}_{\text{rot}} = f_{\text{rot}} \cdot I(v_i, p(o_j, b_j)), \tag{10}$$

where $v_i$ represents the direction of $o_i$ and $p$ is the polygon of bounding box and position of $o_j$. $I$ judges whether a line and a polygon intersect. $f_{\text{rot}}$ is the scale factor parameter.

And the final optimization loss is as follows:

$$\mathcal{L}_{\text{opt}} = \lambda_1 \mathcal{L}_{\text{glo}} + \lambda_2 \mathcal{L}_{\text{pos}} + \lambda_3 \mathcal{L}_{\text{dis}} + \lambda_4 \mathcal{L}_{\text{ali}} + \lambda_5 \mathcal{L}_{\text{rot}}, \tag{11}$$

where $\lambda_{i \in \{1,..,5\}}$ are the loss balancing weight. The algorithm first identifies the main object and then explores placements for the anchor object. Subsequently, it employs Depth-First Search to find valid placements for the remaining objects according to the optimization loss. The whole layout optimization agent is formulated as follows:

$$P = \mathcal{A}_{\text{C}}(\mathcal{Q}((\mathcal{V}(K_{\text{a}}), o_i, o_j), U)), \tag{12}$$

where $o_i$ and $o_j$ are two different objects in the same area from selected objects $O_{\text{s}}$, $i! = j$ and $i, j \in \{0, ...m\}$.

## 4. The GardenVerse Dataset

GardenVerse comprises 132 high-quality artistic 3D assets across three canonical categories: Rock (33), Plant (44), and Architecture (54), including both individual elements (40.2%) and pre-composed arrangements (59.8%) of plants and rocks, enabling flexible retrieval of Jiangnan gardens.

**Collection.** We decompose four digital Jiangnan gardens into objects, including Liuyuan Garden [40], Yiyuan

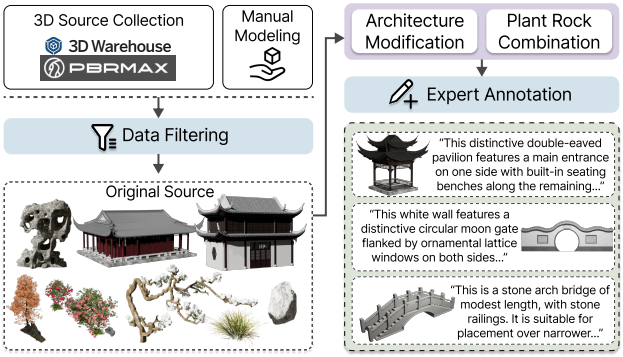


Figure 4. GardenVerse construction from Internet repositories and manual modeling. We invite experts to modify the architectures and construct object combinations. Finally, garden experts annotate the basic information and garden knowledge for assets.

Garden [42], Wangshiyuan Garden [41], and Heyuan Garden [39]. We also collect ancient architectures, plants, and rocks from the 3D Warehouse [38] and PBRMAX [11]. Then, we filter the objects with Northern garden characteristics and retain objects conforming to Jiangnan garden aesthetics. Additionally, we enforce stylistic consistency of assets through mesh optimization and material reassignment. Also, we invite professional garden designers to create a combination of plants and rocks.

**Annotation.** After obtaining the assets, we first annotate them with basic information, including object name, size, minimum and maximum position, and related file path. Recognizing the limitations of LLMs in domain-specific tasks, we engaged landscape architecture experts to comprehensively annotate assets. Each object in GardenVerse includes detailed annotations on: visual attributes of objects, spatial compositions and arrangements, suitable season, description, and contextually appropriate placements. More details can be found in the supplementary materials.

## 5. Experiments

### 5.1. Experiment Setup

**Configuration.** The garden grid is defined as $20 \times 15$ and the real garden size is defined as $200 \times 150\ m^2$. For the parameters, the weights in optimization loss are $\lambda_{i\in\{1,\dots,5\}} = \{2.0, 0.5, 1.8, 0.5, 0.5\}$, and other parameter details can be found in the supplementary materials. We chose OpenAI GPT-5 [28] as the LLM model, file search [27] for knowledge embedding, and Unity to visualize. All reported results were obtained with an Intel(R) Core i7-13620H, 16GB memory, and NVIDIA GeForce RTX 4060 Laptop GPU.

**Metrics.** We evaluate generated Gardens from physical plausibility, structure complexity, semantic coherence, and aesthetic quality. *1. Pathway Score (Path-S).* Path-S is used to determine whether significant plants and buildings can be reached or viewed along the road:

$$S_p = \sum_{i=0}^{n} \min(\frac{d_i}{N} - \phi), \tag{13}$$

where $d_i$ is the distance between each key spot architecture and each edge, and the $\phi$ is the threshold. *2. Class Diversity (Class-Div).* Additionally, we also use Class Diversity to measure the object categories' diversity:

$$D_c = \frac{\sum_{i=0}^{n} c_i}{N}, \tag{14}$$

where $c_i$ is the class number of the generated garden and $N$ is the whole asset number. *3. Fractal Dimension (FD).* We calculate structure complexity [36] as follows:

$$D_f = -\lim_{r\to 0} \frac{\ln N_r}{\ln r}, \tag{15}$$

where $N_r$ is the number of self-similar pieces needed to cover the set at scale $r$. *4. CLIP-Score* is used to measure the consistency between the generated garden and the instruction. *5. CLIP-Aesthetic* is used to evaluate the aesthetic score. *6. VLM-Score.* We prompt the VLMs to rate the rendered garden image. *7. QA-Quality.* We also used VLM-based visual scorer Q-Align [43] to evaluate results.

Table 1. Quantitative comparison. We evaluate our method with the baseline method from four metrics: (1) the pathway rationality (Path-S), (2) the diversity of objects (Class-Div), (3) the structural complexity (FD), and (4) text and scene similarity (CLIP-S).

| Method | Path-S ↑ | Class-Div | FD | CLIP-S↑ |
|---|---|---|---|---|
| Liu et al. [24] | 0 | $21.8 \pm 1.6$ | $1.42 \pm 0.1$ | $27.4 \pm 0.1$ |
| Ours | **8.1** $\pm$ **2.5** | **68.3** $\pm$ **5.6** | **1.36** $\pm$ **0.1** | **27.6** $\pm$ **0.1** |

Table 2. VLMs-based comparison. We render garden images and utilize VLMs to evaluate them from rationality, aesthetic quality, and atmosphere via CLIP-A, VLM-S, and QA-Quality.

| Method | CLIP-A ↑ | VLM-S ↑ | QA-Quality ↑ |
|---|---|---|---|
| Liu et al. [24] | $52.9 \pm 1.0$ | $24.9 \pm 1.2$ | $43.8 \pm 2.5$ |
| Ours | **54.2** $\pm$ **2.0** | **32.5** $\pm$ **2.3** | **53.8** $\pm$ **3.1** |

Table 3. Ablation study for object layout optimization. We evaluate the Knowledge-Embedded Asset Arrangement module by removing it, based on three metric perspetives.

| Method | FD | CLIP-S ↑ | VLM-S ↑ |
|---|---|---|---|
| Ours w/o Arrange. | **1.27** $\pm$ **0.1** | $27.4 \pm 0.1$ | $31.6 \pm 1.1$ |
| Ours | $1.36 \pm 0.1$ | **27.6** $\pm$ **0.1** | **32.5** $\pm$ **2.3** |

### 5.2. Qualitative Analysis

**Comparison with the Baseline Method.** In Figure 5(a), we input the same prompts for the baseline method and GardenDesigner to generate different Jiangnan gardens. The

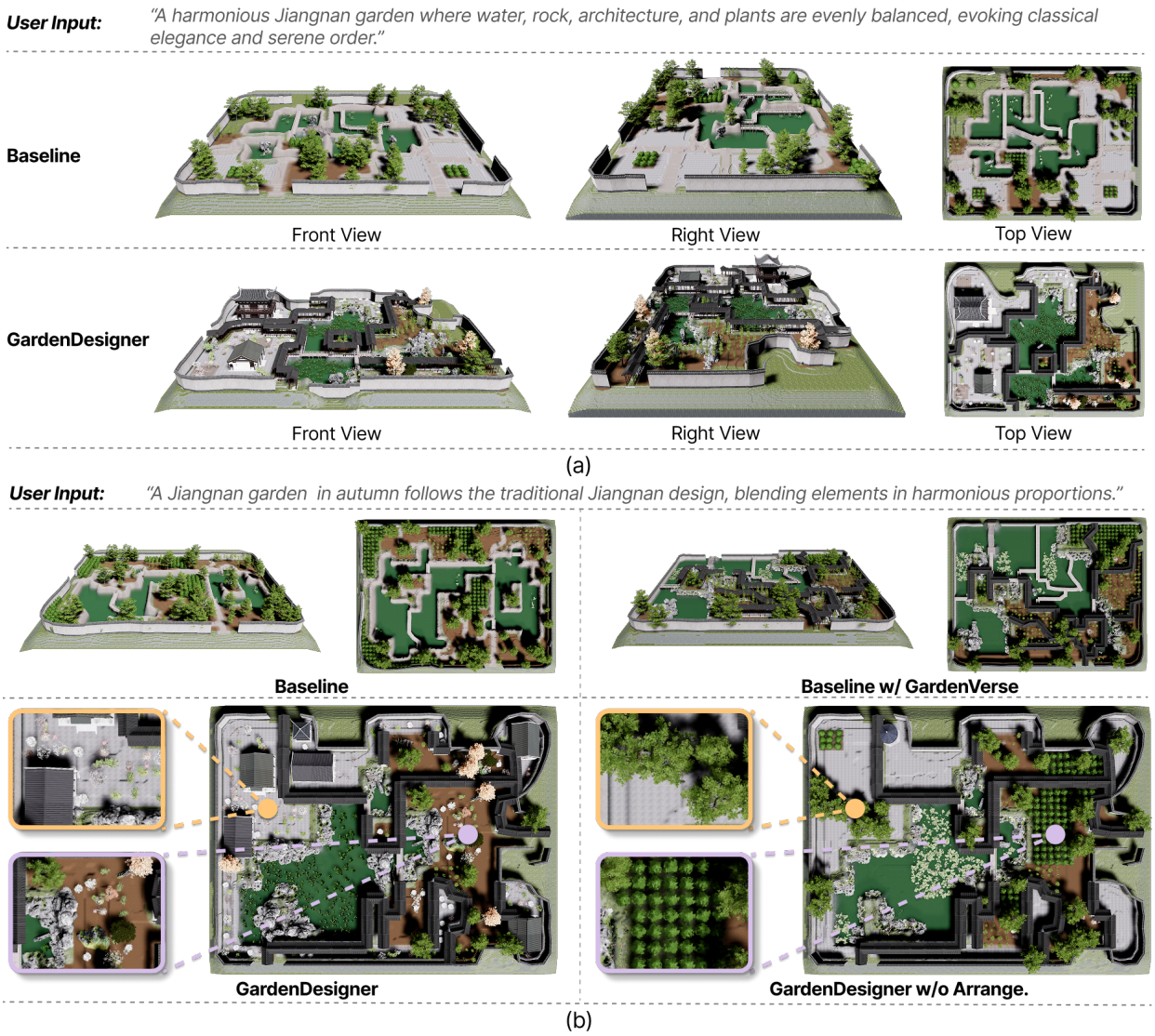


Figure 5. Qualitative analysis. In (a), we input the same prompt to GardenDesigner and the baseline [24] to evaluate the generated garden quality with three different views for each garden. In (b), we compare four methods: (1) Baseline [24], (2) Baseline with GardenVerse assets, (3) GardenDesigner, and (4) GardenDesigner without Knowledge-Embedded Asset Arrangement to conduct the ablation experiment.

views from left to right are front, right, and top. The garden generated from the baseline method has large areas of vacancies, regular plant distribution, and no necessary architecture. In contrast, GardenDesigner gets water-centric terrain distribution, explorative road covering most garden area, and the natural garden layout.

**Ablation Study.** In Figure 5(b), we conduct the ablation study and select two representative areas. Compared to baseline [24], GardenVerse enhances the visual quality of the other three methods, containing abundant objects and natural configurations. After removing the Knowledge-Embedded Asset Arrangement, GardenDesigner achieves a regular layout and limited objects, demonstrating the effectiveness of aesthetic principles integration.

### 5.3. Quantitative Analysis

We compare the performance of GardenDesigner with the baseline [24], as summarized in Table 1. GardenDesigner achieves a higher Path-S of 8.1, reflecting more coherent relationships between architectures and roads, whereas Liu et al. [24] produces unreasonable layouts with no valid score. In terms of asset diversity, GardenDesigner generates a wider range of garden object classes from 26 to 71 types, demonstrating greater dynamism. For structural complexity, GardenDesigner attains a Fractal-dim of 1.36—closer to real Jiangnan gardens from 1.123 to 1.329 [36], indicat-

ing a more natural spatial structure. In addition, GardenDesigner achieves a slightly higher CLIP-S of 27.6. Finally, we prompt VLMs with the rendered garden images and ask to rate the gardens. In Table 2, GardenDesigner greatly exceeds baseline [24] with all three aesthetic metrics.

**Ablation Study.** Removing the Knowledge-embedded Asset Arrangement, GardenDesigner achieves a lower garden structure complexity with 1.27, caused by fewer architectures. In addition, GardenDesigner gets more visual coherence with a CLIP-S of 27.6, , achieved a higher VLM-S of 32.5, validating the aesthetic quality and effectiveness. More experiments are included in supplementary materials.

Table 4. Selection ratio (↑) of different methods for five garden types: (1) Baseline [24], (2) Baseline* (Baseline with GardenVerse), (3) Ours* (GardenDesigner without Knowledge-Embedded Asset Arrangement, (4) Ours (GardenDesigner).

| | Normal | Hydric | Floral | Arch-dense | Mazy |
|---|---|---|---|---|---|
| Baseline | 7.58 | 7.58 | 4.54 | 3.03 | 7.58 |
| Baseline* | 12.12 | 9.09 | 18.18 | 10.61 | 22.72 |
| Ours* | 18.18 | 40.91 | 10.61 | 12.12 | 19.70 |
| Ours | 62.12 | 42.42 | 66.67 | 74.24 | 50.00 |

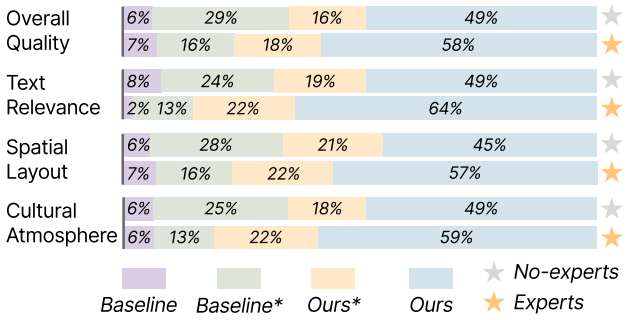


Figure 6. Comparing the selection ratio of four methods in the experiment from four perspectives: (1) Overall Quality, (2) Text Relevance, (3) Spatial Layout, (4) Cultural Atmosphere.

### 5.4. Human Evaluation

We invited 11 garden experts and 32 non-expert volunteers to evaluate the aesthetic quality of the generated Jiangnan gardens. Jiangnan gardens for human evaluation comprise five types in the Table 4. With the chain of agents and knowledge integration, **humans prefer GardenDesigner over baseline methods** from all perspectives. The baseline method receives fewer selections (all under 10%) compared to other methods adopted with GardenVerse. On the contrary, the baseline method gets more preference using the GardenVerse datasets, indicating that **GardenVerse promotes the whole garden quality**. We also removed the Knowledge-Embedded Asset Arrangement module to conduct ablation study. Although two scenes have the same terrain and structure layout, **layout with aesthetic rules gets more preference**, indicating that aesthetic principles play a significant role in determining scene quality.

### 5.5. Discussion

The chain of agents has the potential to generalize to other artistic scene generation tasks. First, by encoding new scene rules and knowledge in textual form, the knowledge-Embedded context mechanism can be directly reused by vectorizing them into semantic memory space. Second, terrain and path generation agents can be adapted to various landscape typologies by modifying the procedural loss terms and path-scoring rules. For example, European royal gardens can be generated by imposing symmetry-aware optimization loss and balanced path scoring.

## 6. Applications

We developed an interface to allow users to input text and construct a Jiangnan garden in Unity. We also provide a terrain adjustment tool to modify the terrain and output the structure map to assist engineers in the construction of physical gardens. Furthermore, users can input instructions to navigate to a spot of interest in the garden using VLMs, as shown in Figure 7. Our GardenDesigner system can support Jiangnan garden design, virtual tourism, interactive entertainment, and virtual reality experiences.

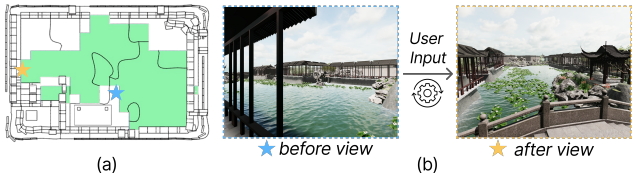


Figure 7. Two applications: (a) Generating a 2D garden construction layout, which can be used to build the garden; (b) Navigating to a spot of interest following the user's instructions.

## 7. Conclusion

This paper has proposed **GardenDesigner**, a novel framework that encodes aesthetic principles for Jiangnan garden construction and integrates a chain of agents procedurally. By structuring the generation process into hierarchical garden composition and knowledge-embedded asset arrangement, GardenDesigner ensures spatial rationality and aesthetic coherence with Jiangnan garden design principles. Looking forward, GardenDesigner can be extended to support interactive educational tools, virtual heritage reconstruction, and personalized landscape design, opening new avenues for cultural heritage preservation and creative applications in digital art and games.

## 8. Acknowledgments

This work was supported by the National Natural Science Foundation of China (Grant No. 62402306), the Natural Science Foundation of Shanghai (Grant No. 24ZR1422400, Grant No. 25ZR1401130), the Open Research Project of the State Key Laboratory of Industrial Control Technology, China (Grant No. ICT2024B72), and the Guangdong Basic and Applied Basic Research Foundation (No. 2026A1515011138).

## References

[1] Sherwin Bahmani, Jeong Joon Park, Despoina Paschalidou, Xingguang Yan, Gordon Wetzstein, Leonidas Guibas, and Andrea Tagliasacchi. CC3D: Layout-Conditioned Generation of Compositional 3D Scenes. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 7171–7181, 2023. 2

[2] Zixuan Bian, Ruohan Ren, Yue Yang, and Chris Callison-Burch. HOLODECK 2.0: Vision-Language-Guided 3D World Generation with Editing. *arXiv preprint arXiv:2508.05899*, 2025. 2

[3] Blender Foundation. Blender. https://www.blender.org, 2025. Accessed Nov 10, 2025. 2, 3

[4] Ata Çelen, Guo Han, Konrad Schindler, Luc Van Gool, Iro Armeni, Anton Obukhov, and Xi Wang. I-design: Personalized llm interior designer. In *Computer Vision – ECCV 2024 Workshops: Milan, Italy, September 29–October 4, 2024, Proceedings, Part II*, page 217–234, Berlin, Heidelberg, 2025. Springer-Verlag. 2

[5] Angel X Chang, Thomas Funkhouser, Leonidas Guibas, Pat Hanrahan, Qixing Huang, Zimo Li, Silvio Savarese, Manolis Savva, Shuran Song, Hao Su, et al. Shapenet: An Information-Rich 3D Model Repository. *arXiv preprint arXiv:1512.03012*, 2015. 2

[6] J. Cheng, A. Hardie, Z. Ming, and M. Keswick. *The Craft of Gardens: The Classic Chinese Text on Garden Design*. Shanghai Press, 2012. 1, 3

[7] Matt Deitke, Ruoshi Liu, Matthew Wallingford, Huong Ngo, Oscar Michel, Aditya Kusupati, Alan Fan, Christian Laforte, Vikram Voleti, Samir Yitzhak Gadre, Eli VanderBilt, Aniruddha Kembhavi, Carl Vondrick, Georgia Gkioxari, Kiana Ehsani, Ludwig Schmidt, and Ali Farhadi. Objaverse-XL: A Universe of 10M+ 3D Objects. In *Proceedings of the 37th International Conference on Neural Information Processing Systems*, Red Hook, NY, USA, 2023. Curran Associates Inc. 2

[8] Matt Deitke, Dustin Schwenk, Jordi Salvador, Luca Weihs, Oscar Michel, Eli VanderBilt, Ludwig Schmidt, Kiana Ehsani, Aniruddha Kembhavi, and Ali Farhadi. Objaverse: A Universe of Annotated 3D Objects. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 13142–13153, 2023. 2

[9] Jie Deng, Wenhao Chai, Junsheng Huang, Zhonghan Zhao, Qixuan Huang, Mingyan Gao, Jianshu Guo, Shengyu Hao, Wenhao Hu, Jenq-Neng Hwang, et al. Citycraft: A Real Crafter for 3D City Generation. *arXiv preprint arXiv:2406.04983*, 2024. 3

[10] Laura Downs, Anthony Francis, Nate Koenig, Brandon Kinman, Ryan Hickman, Krista Reymann, Thomas B. McHugh, and Vincent Vanhoucke. Google Scanned Objects: A High-Quality Dataset of 3D Scanned Household Items. In *2022 International Conference on Robotics and Automation (ICRA)*, page 2553–2560. IEEE Press, 2022. 2

[11] EcoPlants. PBRMAX. https://pbrmax.com/, 2025. Accessed Nov 10, 2025. 6

[12] Epic Games. Unreal. https://www.unrealengine.com/, 2025. Accessed Nov 10, 2025. 3

[13] Weixi Feng, Wanrong Zhu, Tsu-jui Fu, Varun Jampani, Arjun Akula, Xuehai He, Sugato Basu, Xin Eric Wang, and William Yang Wang. Layoutgpt: Compositional Visual Planning and Generation With Large Language Models. In *Proceedings of the 37th International Conference on Neural Information Processing Systems*, Red Hook, NY, USA, 2023. Curran Associates Inc. 2

[14] Rao Fu, Zehao Wen, Zichen Liu, and Srinath Sridhar. AnyHome: Open-Vocabulary Generation of Structured and Textured 3D Homes. In *Computer Vision – ECCV 2024: 18th European Conference, Milan, Italy, September 29–October 4, 2024, Proceedings, Part XXXIX*, page 52–70, Berlin, Heidelberg, 2024. Springer-Verlag. 2

[15] Torsten Hädrich, Bedrich Benes, Oliver Deussen, and Sören Pirk. Interactive Modeling and Authoring of Climbing Plants. *Comput. Graph. Forum*, 36(2):49–61, 2017. 3

[16] Lukas Höllein, Ang Cao, Andrew Owens, Justin Johnson, and Matthias Nießner. Text2Room: Extracting Textured 3D Meshes from 2D Text-to-Image Models. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 7909–7920, 2023. 2

[17] Ziniu Hu, Ahmet Iscen, Aashi Jain, Thomas Kipf, Yisong Yue, David A Ross, Cordelia Schmid, and Alireza Fathi. Scenecraft: An Llm Agent for Synthesizing 3D Scenes as Blender Code. In *Proceedings of the 41st International Conference on Machine Learning*. JMLR.org, 2024. 2

[18] Madiha Jamil. Augmented Reality for Historic Storytelling and Preserving Artifacts in Pakistan. *International E-Journal of Advances in Social Sciences*, 5(14):998–1004, 2019. 3

[19] Samin Khan, Buu Phan, Rick Salay, and Krzysztof Czarnecki. Procsy: Procedural synthetic dataset generation towards influence factor studies of semantic segmentation networks. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, 2019. 2, 3

[20] Kangsoo Kim, Byung-Kuk Seo, Jae-Hyek Han, and Jong-Il Park. Augmented Reality Tour System for Immersive Experience of Cultural Heritage. In *Proceedings of the 8th International Conference on Virtual Reality Continuum and Its Applications in Industry*, page 323–324, New York, NY, USA, 2009. Association for Computing Machinery. 3

[21] Hoang-An Le, Thomas Mensink, Partha Das, Sezer Karaoglu, and Theo Gevers. EDEN: Multimodal Synthetic Dataset of Enclosed GarDEN Scenes. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, pages 1579–1589, 2021. 3

[22] Bosheng Li, Nikolas Alexander Schwarz, Wojtek Pałubicki, Sören Pirk, and Bedrich Benes. Interactive Invigoration: Volumetric Modeling of Trees with Strands. *ACM Trans. Graph.*, 43(4), 2024. 3

[23] Lu Ling, Chen-Hsuan Lin, Tsung-Yi Lin, Yifan Ding, Yu Zeng, Yichen Sheng, Yunhao Ge, Ming-Yu Liu, Aniket Bera, and Zhaoshuo Li. Scenethesis: A Language and Vision Agentic Framework for 3D Scene Generation. *arXiv preprint arXiv:2505.02836*, 2025. 2

[24] Jia-Hong Liu, Shao-Kui Zhang, Chuyue Zhang, and Song-Hai Zhang. Controllable Procedural Generation of Landscapes. In *Proceedings of the 32nd ACM International Conference on Multimedia*, page 6394–6403, New York, NY, USA, 2024. Association for Computing Machinery. 2, 6, 7, 8

[25] Xinhang Liu, Chi-Keung Tang, and Yu-Wing Tai. WorldCraft: Photo-Realistic 3D World Creation and Customization via LLM Agents. *arXiv preprint arXiv:2502.15601*, 2025. 2

[26] Quan Meng, Lei Li, Matthias Nießner, and Angela Dai. Lt3sd: Latent trees for 3d scene diffusion. *arXiv preprint arXiv:2409.08215*, 2024. 2

[27] OpenAI. File Search. https://platform.openai.com/docs/guides/tools-file-search, 2025. Accessed Nov 10, 2025. 6

[28] OpenAI. GPT-5. https://platform.openai.com/docs/models/gpt-5, 2025. Accessed Nov 10, 2025. 6

[29] Yoav I. H. Parish and Pascal Müller. Procedural Modeling of Cities. In *Proceedings of the 28th Annual Conference on Computer Graphics and Interactive Techniques*, page 301–308, New York, NY, USA, 2001. Association for Computing Machinery. 2, 3

[30] Yigang Peng, Huaiyun Kou, and Ron Henderson. *Analysis of the Traditional Chinese Garden*. Springer, 1986. 2, 3

[31] Alexander Raistrick, Lahav Lipson, Zeyu Ma, Lingjie Mei, Mingzhe Wang, Yiming Zuo, Karhan Kayan, Hongyu Wen, Beining Han, Yihan Wang, Alejandro Newell, Hei Law, Ankit Goyal, Kaiyu Yang, and Jia Deng. Infinite Photorealistic Worlds Using Procedural Generation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 12630–12641, 2023. 2, 3

[32] Alexander Raistrick, Lingjie Mei, Karhan Kayan, David Yan, Yiming Zuo, Beining Han, Hongyu Wen, Meenal Parakh, Stamatis Alexandropoulos, Lahav Lipson, Zeyu Ma, and Jia Deng. Infinigen Indoors: Photorealistic Indoor Scenes using Procedural Generation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 21783–21794, 2024. 2

[33] Pratheba Selvaraju, Mohamed Nabail, Marios Loizou, Maria Maslioukova, Melinos Averkiou, Andreas Andreou, Siddhartha Chaudhuri, and Evangelos Kalogerakis. BuildingNet: Learning To Label 3D Buildings. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 10397–10407, 2021. 3

[34] Sketchfab, Inc. Sketchfab. https://sketchfab.com, 2025. Accessed Nov 10, 2025. 3

[35] Chunyi Sun, Junlin Han, Weijian Deng, Xinlong Wang, Zishan Qin, and Stephen Gould. 3D-GPT: Procedural 3D Modeling With Large Language Models. *arXiv preprint arXiv:2310.12945*, 2023. 2

[36] Ce Sun, Zhenyu Jiang, and Bingqin Yu. How to Interpret Jiangnan Gardens: A Study of the Spatial Layout of Jiangnan Gardens From the Perspective of Fractal Geometry. *Heritage Science*, 12(1):353, 2024. 6, 7

[37] Fan-Yun Sun, Weiyu Liu, Siyi Gu, Dylan Lim, Goutam Bhat, Federico Tombari, Manling Li, Nick Haber, and Jiajun Wu. LayoutVLM: Differentiable Optimization of 3D Layout via Vision-Language Models. In *Proceedings of the Computer Vision and Pattern Recognition Conference (CVPR)*, pages 29469–29478, 2025. 2

[38] Trimble Inc. 3D Warehouse. https://3dwarehouse.sketchup.com, 2025. Accessed Nov 10, 2025. 3, 6

[39] Wikipedia. He Garden. https://en.wikipedia.org/wiki/He_Garden, 2025. Accessed Nov 10, 2025. 6

[40] Wikipedia. Lingering Garden. https://en.wikipedia.org/wiki/Lingering_Garden, 2025. Accessed Nov 10, 2025. 5

[41] Wikipedia. Master of the Nets Garden. https://en.wikipedia.org/wiki/Master_of_the_Nets_Garden, 2025. Accessed, Nov 10, 2025. 6

[42] Wikipedia. Garden of Pleasance. https://en.wikipedia.org/wiki/Garden_of_Pleasance, 2025. Accessed Nov 10, 2025. 6

[43] Haoning Wu, Zicheng Zhang, Weixia Zhang, Chaofeng Chen, Liang Liao, Chunyi Li, Yixuan Gao, Annan Wang, Erli Zhang, Wenxiu Sun, Qiong Yan, Xiongkuo Min, Guangtao Zhai, and Weisi Lin. Q-Align: Teaching Lmms for Visual Scoring via Discrete Text-Defined Levels. In *Proceedings of the 41st International Conference on Machine Learning*. JMLR.org, 2024. 6

[44] Tong Wu, Jiarui Zhang, Xiao Fu, Yuxin Wang, Jiawei Ren, Liang Pan, Wayne Wu, Lei Yang, Jiaqi Wang, Chen Qian, Dahua Lin, and Ziwei Liu. OmniObject3D: Large-Vocabulary 3D Object Dataset for Realistic Perception, Reconstruction and Generation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 803–814, 2023. 2

[45] Haozhe Xie, Zhaoxi Chen, Fangzhou Hong, and Ziwei Liu. CityDreamer: Compositional Generative Model of Unbounded 3D Cities. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 9666–9675, 2024. 2

[46] Haozhe Xie, Zhaoxi Chen, Fangzhou Hong, and Ziwei Liu. Generative Gaussian Splatting for Unbounded 3D City Generation. In *Proceedings of the Computer Vision and Pattern Recognition Conference (CVPR)*, pages 6111–6120, 2025. 2

[47] Yue Yang, Fan-Yun Sun, Luca Weihs, Eli VanderBilt, Alvaro Herrasti, Winson Han, Jiajun Wu, Nick Haber, Ranjay Krishna, Lingjie Liu, Chris Callison-Burch, Mark Yatskar, Aniruddha Kembhavi, and Christopher Clark. Holodeck: Language Guided Generation of 3D Embodied AI Environments. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 16227–16237, 2024. 2

[48] Hong-Xing Yu, Haoyi Duan, Charles Herrmann, William T. Freeman, and Jiajun Wu. WonderWorld: Interactive 3D Scene Generation from a Single Image. In *Proceedings of the Computer Vision and Pattern Recognition Conference (CVPR)*, pages 5916–5926, 2025. 2

[49] Liang Yue, Shao-Kui Zhang, Lin Yuan, Yi-Tao Chen, Zirui Zhou, and Song-Hai Zhang. Synthesizing 3d scenes via diffusion model that incorporates indoor scene characteristics. page 9385–9394, New York, NY, USA, 2025. Association for Computing Machinery. 2

[50] Guangyao Zhai, Evin Pinar Örnek, Shun-Cheng Wu, Yan Di, Federico Tombari, Nassir Navab, and Benjamin Busam. CommonScenes: generating commonsense 3D indoor scenes with scene graph diffusion. In *Proceedings of the 37th International Conference on Neural Information Processing Systems*, Red Hook, NY, USA, 2023. Curran Associates Inc. 2

[51] Zaiwei Zhang, Zhenpei Yang, Chongyang Ma, Linjie Luo, Alexander Huth, Etienne Vouga, and Qixing Huang. Deep generative modeling for scene synthesis via hybrid representations. *ACM Transactions on Graphics (TOG)*, 39(2):1–21, 2020. 3

[52] Mengqi Zhou, Jun Hou, Chuanchen Luo, Yuxi Wang, Zhaoxiang Zhang, and Junran Peng. SceneX: Procedural Controllable Large-Scale Scene Generation via Large-Language Models. *arXiv e-prints*, pages arXiv–2403, 2024. 2, 3

[53] Jianzhong Zhu, Ruifang Zhai, He Ren, Kai Xie, Aobo Du, Xinwei He, Chenxi Cui, Yinghua Wang, Junli Ye, Jiashi Wang, et al. Crops3d: A Diverse 3D Crop Dataset for Realistic Perception and Segmentation Toward Agricultural Applications. *Scientific Data*, 11(1):1438, 2024. 3