buildings

MDPI

Article

AI for Garden Design Visualization: Development and Validation of the GardenDiff Model

Xiaolong Sun [1], Xi Chen [1], Chao Zhou [1], Shengsha Wu [1], Hongbo Zhao [1,*] and Kun Li [1,2,*]

[1] School of Landscape Architecture, Zhejiang Agricultural & Forestry University, Hangzhou 311300, China; sxl0215@stu.zafu.edu.cn (X.S.)
[2] School of Engineering and Design, Technical University of Munich, 80333 Munich, Germany
* Correspondence: zhaohb@zafu.edu.cn (H.Z.); kunli@zafu.edu.cn (K.L.)

Abstract

The rapid advancement of AI-driven generative design brings new opportunities, but its application in landscape garden design remains limited by two gaps: (1) semantic misalignment between generated images and the designer's intent, and (2) low-resolution outputs with insufficient details. To address these gaps, we developed GardenDiff, a domain-adapted diffusion model trained via parameter optimization and a specialized landscape garden dataset. Central to this approach is Structured Design Captioning (SDC), a hierarchical annotation system specifically designed for garden design that encodes design elements, style features, and auxiliary scene information. To develop this model, we designed a three-stage experimental framework. In Stage 1, we examined the effects of training caption systems and training resolution on generated landscape garden imagery by controlled experiments. In Stage 2, we conducted joint training across five garden styles (Chinese, Japanese, Mediterranean, Nordic, and English) based on the optimized parameter settings from Stage 1 to construct the GardenDiff model. In Stage 3, we validated the model performance through expert evaluation (N = 36) and public evaluation (N = 136) and analyzed style-specific variations in the generated outcomes. Research results showed that Structured Design Captioning (SDC) improved Spatial Rationale by 19.67–39.46% compared with generic captions, and training at 1536 × 1536 pixels improved image quality by 23.2% compared with 768 × 768-pixel training. GardenDiff trained with these optimized parameters showed notable advantages. Its overall scores (5.06) exceeded those of Stable Diffusion XL base 1.0 (SDXL 1.0) by 16.4% and DreamShaper XL by 22.4%. The model improved across four dimensions, including Design Rationale, Design Professionalism, Design Accuracy, and Design Satisfaction. Our study offers a new model to improve the perspective visualization of generative garden design and provides insights into AI-informed landscape and urban design.

Keywords: landscape garden design; LoRA fine-tuning; structured design captioning; generative design; artificial intelligence

Check for updates

Academic Editor: Xuan Ma

Received: 2 April 2026
Revised: 7 May 2026
Accepted: 26 May 2026
Published: 29 May 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 Attribution (CC BY) license.

1. Introduction

Due to rapid urbanization, urban residents have experienced a significant decline in their daily opportunities for contact with nature, leading to growing concerns about issues such as "nature deficit disorder" [1,2]. High-quality green spaces in residential areas are critical for enhancing residents' physical health and mental well-being. Among various forms of urban green spaces, gardens are the most familiar and directly experienced

green spaces for people in residential settings. They provide essential ecosystem services, including microclimate regulation, human–nature interaction, and cultural value [3,4]. In this context, as demands for diverse garden designs continue to grow, there is an increasing need to develop approaches and tools that can accurately capture differences among garden styles while efficiently generating design solutions [5].

Generative Artificial Intelligence (GenAI) technologies present an opportunity to speed up planning, design, communication, and feedback in the urban design area [6]. The conventional design workflow includes multiple technical stages from initial schematic plans to final photorealistic presentations using 2D drafting platforms, 3D modeling applications, and rendering tools. It often lacks the capacity for rapid iteration and diverse exploration of design alternatives [7]. In contrast, diffusion models are valued for their training stability and controllable generation capabilities [8,9]. It can rapidly generate diverse architecture and landscape designs, showing the potential to update conventional design workflows. In the landscape design area, research has evolved from Generative Adversarial Network (GAN)-based plan generation [10] to integrate workflows combining GAN and Stable Diffusion for green space planning and design, evaluating the environment's influence on generation results, and extending applications from remote sensing inputs to final design schemes [11–13]. Except for changes in the design workflow, diffusion models have been applied at specific design stages (e.g., concept generation, detail optimization). For instance, Ye et al. [14] demonstrated the application of text-to-image GenAI in the conceptual design stage, while other studies have highlighted the capability of these models to produce diverse architectural renderings with professional lighting and material quality [15,16].

However, integrating generative AI in landscape garden design presents greater computational challenges due to the garden's small scale, culturally nuanced expressions, and intricate spatial arrangements. These features require more refined generations to scene atmosphere creation, cultural imagery expression, and landscape spatial composition [17]. Moreover, garden designs vary significantly across cultural contexts. For example, the design of Chinese gardens emphasizes the subtle artistic conception of "enhancing landscape layering through winding path" [17], while the design of English gardens seeks open, romantic, and layered landscapes [18].

Yet current GenAI applications in landscape garden design still face challenges in processing complex semantic content and natural forms [19]. First, general-purpose models exhibit semantic bias because they are trained primarily based on large-scale generic image data. They inadequately cover the spatial organizational logic and aesthetic characteristics of specific garden styles [8], resulting in noticeable biases in the model's understanding of design semantics [20]. In a preliminary diagnostic assessment, the research team generated garden images across the five target styles using SDXL 1.0 [21] under both unconstrained and ControlNet-assisted conditions, and systematically examined the outputs based on domain expertise. The results revealed that AI-generated garden images often exhibit scale distortion, omit essential design elements, and present blurred details, thus failing to achieve professional and accurate garden design (see Appendix A).

These issues are closely related to how data is processed during training. Specifically, the caption systems determine the accuracy of the model's semantic understanding, while the training resolution determines the precision of the details the model can learn [22,23]. In massive generic training datasets, captions generated by automatic tagging tools tend to be overly generalized. For example, WD1.4 (Waifu Diffusion 1.4 Tagger) [24] categorizes diverse Asian architectural features uniformly as "east_asian_architecture", and such generalized captions fail to effectively distinguish the unique characteristics of architectural elements from different cultural traditions. In particular, representative design elements such as the upturned "flying eaves" of Chinese buildings and the "thatched roofs" of

Japanese structures differ significantly, but these distinctions cannot be effectively distinguished in current caption systems. Similarly, the resolution setting of training images also affects the model's ability to learn detailed features. Although some studies have explored the effects of parameters such as caption systems and training resolution in general image generation tasks [25,26], systematic analysis is still lacking regarding how these factors influence the model's learning of cultural semantics and spatial scale in garden design.

To address these issues, this study focuses on optimizing GenAI applications in landscape garden design through answering three key questions: (1) how captioning systems and training resolution affect generation quality of landscape garden design; (2) whether domain-specific training strategies can effectively improve both accuracy and professional standards; and (3) in what aspects the proposed training strategy improves the generation of different garden styles over general-purpose models.

This study makes three key contributions to the field of AI-assisted landscape design. First, we propose Structured Design Captioning (SDC), a hierarchical annotation framework specifically designed for garden design that bridges domain-specific terminology with diffusion model training. Unlike generic captioning tools, SDC encodes design elements, style features, and auxiliary scene information in a structured three-layer architecture. Second, we develop GardenDiff, a domain-adapted diffusion model that integrates SDC with high-resolution training to generate garden imagery across five culturally distinct styles. Third, through controlled experiments, we provide empirical evidence quantifying the independent effects of captioning systems and training resolution on generation quality, offering transferable insights for domain adaptation in other design fields.

2. Methodology

To address these research questions, this study employs a quantitative experimental approach to systematically analyze how captioning systems, training resolution, and domain-specific training strategies influence generation quality, while revealing the adaptability of different design features to these optimization methods.

2.1. Overall Framework of Methods

We designed a three-stage progressive experimental framework utilizing a newly constructed dataset of five representative garden styles (Chinese, Japanese, Mediterranean, Nordic, and English) to systematically translate traditional garden design knowledge into a generative AI model. As shown in Figure 1, the process included parameter optimization, model development, and performance validation.

Stage 1: Parameter Optimization. In this stage, we addressed the first research question (RQ1) by identifying the optimal training configurations. We focused on two key variables: captioning systems (WD1.4, BLIP, SDC) and training resolutions (768 px, 1024 px, 1536 px). To evaluate these independently, we employed a controlled-variable design with a shared baseline (BLIP + 1024 px). When comparing the three captioning systems, we fixed the training resolution at 1024 px. Conversely, when comparing the three resolution settings, we kept the captioning method fixed to BLIP and applied ControlNet constraints to evaluate scale consistency. This experimental design produced five models per style (one shared baseline plus two variants for captioning and two for resolution), yielding 25 parameter-diagnostic models. The detailed setup is described in Section 2.2.

Stage 2: Model Development. In this phase, we developed the final GardenDiff model. Based on the optimal parameter combination identified in Stage 1, we integrated the full training data of all five garden styles to train the domain-specific diffusion model. This process transformed the visual patterns and design principles of the styles into the model's latent space, as detailed in Section 2.3.

Stage 3: Performance Validation. Finally, we addressed RQ2 and RQ3 by comparing GardenDiff against two general-purpose models (SDXL 1.0 [21] and DreamShaper XL [27] under identical ControlNet conditions. The results are discussed in Sections 3.3 and 3.4.

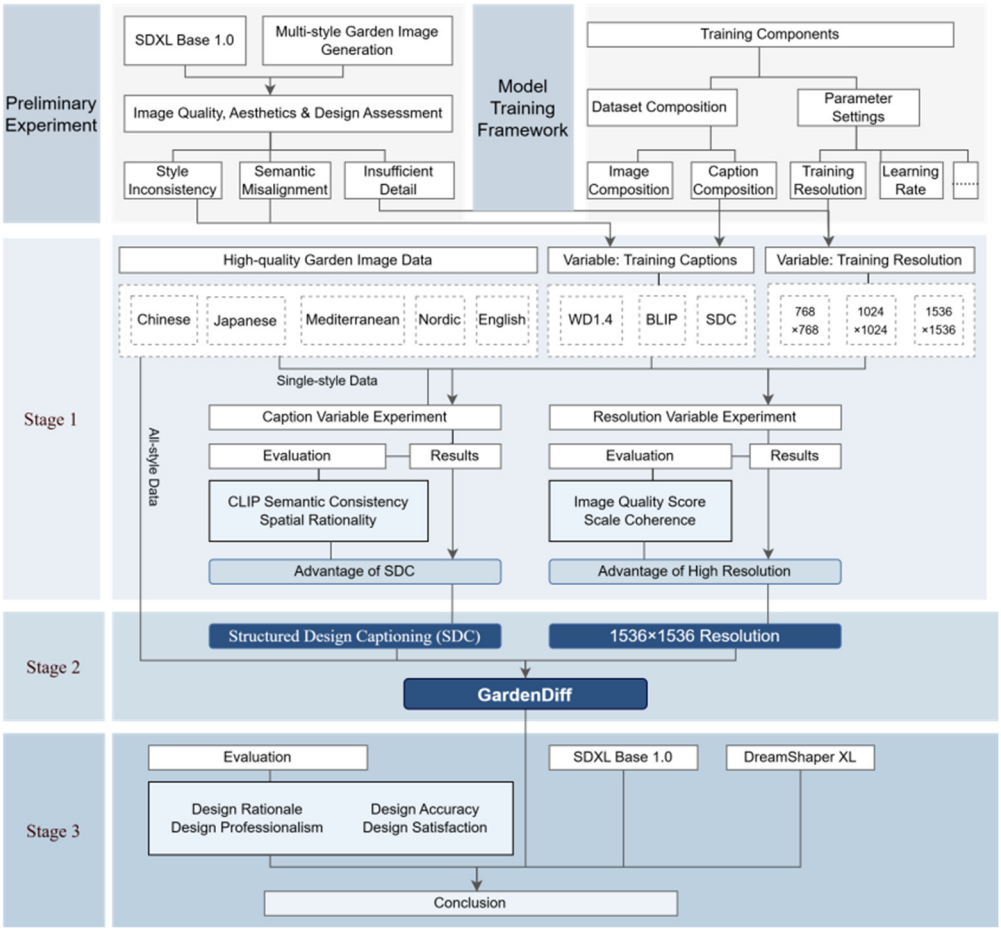


Figure 1. The methodological framework for the training and validation of the GardenDiff model.

2.2. Dataset Construction and Annotation Strategy

2.2.1. Image Collection and Garden Style Selection

To construct a specialized dataset for landscape garden design, we used images from professional platforms (Freepik, Pexels, Unsplash) under open licenses (see Appendix B.1) and selected 250 high-quality images per style, yielding 1250 images in total. The selection followed three criteria: (1) resolution $\geq 1536 \times 1536$ px for sufficient detail; (2) complete composition without noticeable noise or watermarks; and (3) accurate representation of core design elements for each style.

The dataset size of 250 images per style was determined by balancing LoRA's parameter efficiency [28] with the complexity of learning five distinct styles encompassing diverse design elements (building facades, garden structures, paving, and planting design). This size is consistent with established LoRA fine-tuning practices, where Hu et al. [28]. demonstrated effective adaptation with similarly small datasets due to LoRA's low-rank parameter constraints. Moreover, the 1250 final images were curated through a rigorous four-stage filtering pipeline from an initial pool of 8460 candidates (retention rate: 14.8%), ensuring high consistency and quality within each style category (Figure 2). This configuration was subsequently validated by the training results, as all model groups achieved stable convergence within 20 epochs (see Section 3.1).

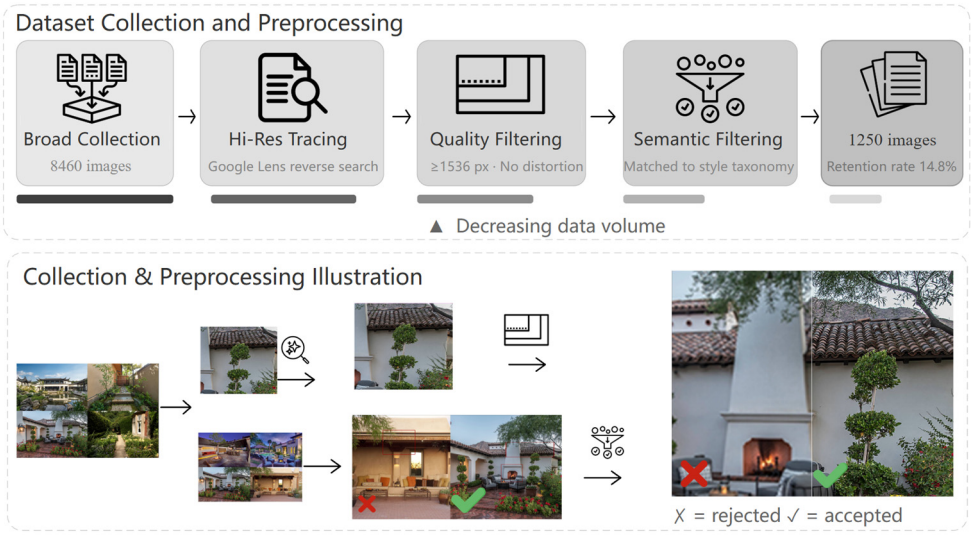


Figure 2. Dataset collection and preprocessing.

Beyond sample size and complexity, verifying the semantic distinctiveness of these garden styles is crucial. We extracted image features using the CLIP ViT-L/14 encoder [29] and applied t-SNE (perplexity = 30, PCA initialization) to visualize the dataset's distribution (Figure 3). The visualization reveals clear clustering for each of the five styles, confirming that the collected images possess distinguishable visual semantics. This separability provides a solid foundation for the model to learn specific style definitions.

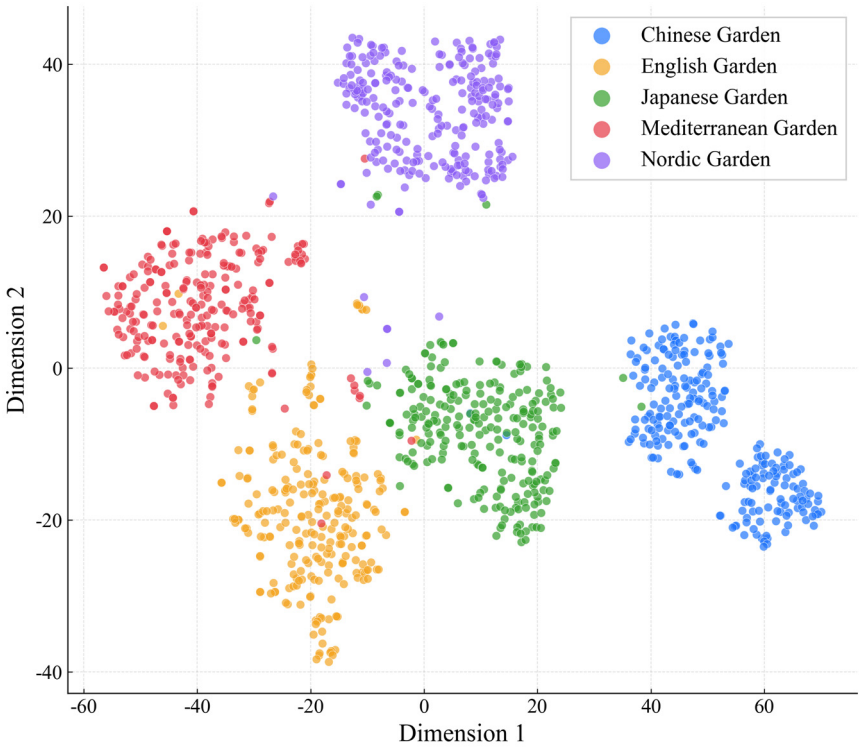


Figure 3. t-SNE Visualization of feature clustering across five garden styles.

Based on these validated clusters, the specific design definitions for the five selected styles are as follows: Chinese, Japanese, Mediterranean, Nordic, and English (Table 1). Chinese gardens follow the traditional principle of "harmony between nature and human," characterized by white walls, black tiles, symmetrical layouts, and rock and water landscapes designed to imitate natural scenery [17]. Japanese gardens embody Zen aesthetics

through sandy grounds, bamboo fences, and water features that emphasize simplicity and natural beauty [30,31]. Mediterranean gardens are characterized by openness and bright colors. They typically feature terracotta flooring, wrought-iron decorations, and arched windows, creating vivid and spacious environments [32,33]. Nordic gardens emphasize functionality and natural textures. They commonly integrate wood, stone, and minimalist planting to achieve a harmonious blend of minimalism and nature [34]. English gardens follow romantic and naturalistic principles, with diverse arrangements of lawns, flower beds, and trees that create layered landscapes and structural order [18].These five styles were selected based on their established historical significance and continued prevalence in contemporary garden design practice. While this selection does not exhaust all cultural traditions, it covers a range of distinct spatial and aesthetic characteristics sufficient to evaluate the adaptability of domain-specific training.

Table 1. Design elements of five garden styles.

Design Category	Chinese	Japanese	Mediterranean	Nordic	English
Architectural Facades	Upturned eaves, Whitewashed walls, Black tiles	Thatched roofs, Tiled roofs, Shoji windows, Bamboo walls	Terracotta roof tiles, Limestone walls, Arched doorways	Natural timber, Stone, Glass	Brick walls, Brick walls with climbing vines
Garden Structures	Pavilions, Corridors, Bridges, Moongates	Stone lanterns, Bamboo fences, Stone basins, Water basins	Terracotta planters, Wrought-iron furniture	Outdoor seating, Fire pits	Sculptures, Bird baths
Paving	Stone pavers, Brick, Ceramic tiles	Stone slabs, Gravel, Pebbles	Terracotta tiles, Colored ceramics, Natural stone	Stone, Natural wood	Regularly shaped stone paving
Rock & Water Features	Rockwork (artificial rockery), Reflection pools	Japanese rock garden (karesansui), Ponds, Streams	Swimming pools, Fountains	N/A	Fountains
Planting	Bamboo, Pine, Plum blossoms	Cherry blossoms, Maple, Moss, Ferns	Olive trees, Lemon trees, Grapevines, Pomegranates, Herbs	Pine, Cedar, Spruce	Roses, Boxwood, Lavender, Iris
Atmosphere	Serene, Harmonizing tradition	Restrained, Contemplative	Bright, Vibrant	Minimalist, Functional	Lush, Layered

2.2.2. Construction of Structured Design Captioning

To achieve precise translation of design knowledge, we constructed the Structured Design Captioning (SDC) system. To establish a comparative baseline for the subsequent experiments, we also selected two general-purpose tools: BLIP [35] and WD1.4 [24].

SDC employs word-level tags organized into a three-layer hierarchical structure. The first layer, the Design-Element Layer, establishes precise pixel-text alignment through three differentiated mapping strategies: (1) retaining culture-specific terms (e.g., “Karesansui”) for distinct features to enforce strong visual associations; (2) mapping rare concepts to hypernyms (e.g., “Ha-ha wall” to “low retaining wall”) to leverage the base model’s priors;

(3) simplifying botanical names to common English (e.g., "Pampas grass") to facilitate training convergence.

The second layer, the Style-Feature Layer, contains style identifiers (specifying garden type) and atmosphere descriptors (conveying spatial characteristics and cultural connotations). The third layer, the Auxiliary-Info Layer, records viewpoint, seasonal weather conditions, and layout characteristics to provide complete scene understanding (captioning examples are provided in Appendix B.2).

To generate SDC captions efficiently while maintaining domain accuracy, we developed a semi-automatic annotation pipeline (Figure 4). Images were first grouped by style and preprocessed for consistent input quality. A multimodal large language model (GPT-4o, OpenAI, San Francisco, CA, USA) then generated initial captions guided by SDC-structured prompts (see Appendix B.3) specifying the target style, required design elements, and feature categories. Rule-based cleaning was subsequently applied to unify terminology, remove duplicates, and standardize formatting. Finally, landscape architecture researchers reviewed all captions to verify domain accuracy and ensure consistency with established design knowledge.

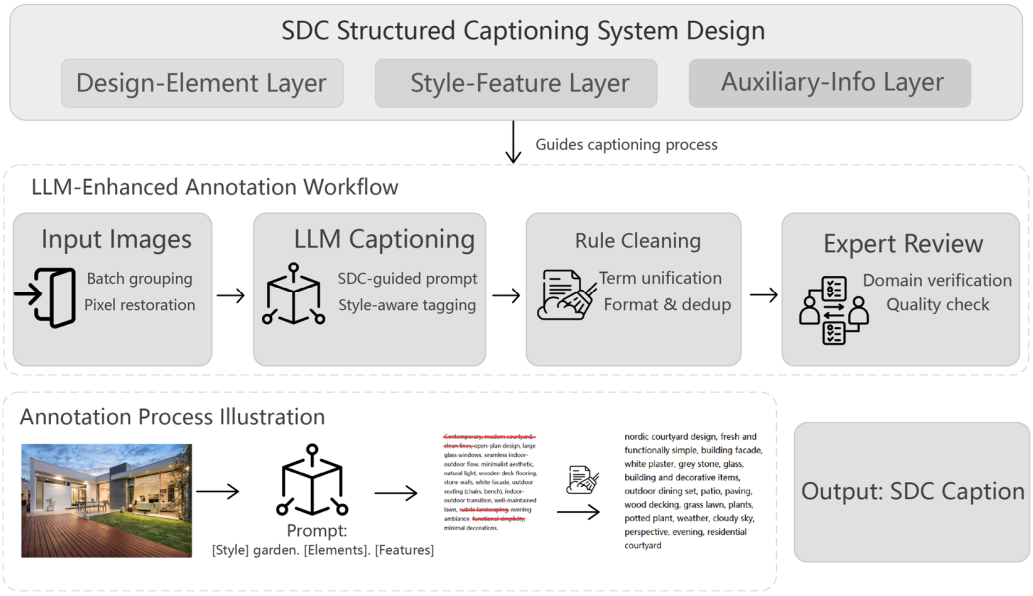


Figure 4. SDC structured captioning system design and annotation workflow.

2.3. Experimental Setup and Training Configuration

2.3.1. Base Model and Training Method Selection

For garden design tasks, we selected diffusion models [36] based on their superior detail preservation and training stability. Diffusion models are significantly less susceptible to mode collapse than GANs [37], which is particularly important when modeling the intricate spatial structures and fine details characteristic of garden designs. We adopted Stable Diffusion XL (SDXL 1.0) as the base model [21], which generates images at a native resolution of 1024 × 1024 px and has demonstrated strong performance in high-resolution image synthesis.

For the training methodology, we employed LoRA (Low-Rank Adaptation) [28] to reduce the number of trainable parameters and accelerate convergence. LoRA offers several advantages over alternative fine-tuning methods (see Table 2). Compared with DreamBooth [38]. It reduces GPU memory requirements by 90%, making training more accessible. It also retains style features more accurately than Textual Inversion [39] and scales more effectively than HyperNetwork [40] for multi-style training tasks.

Table 2. Comparison of common fine-tuning methods.

Method	Training Time	Generation Quality	Computational Cost	Storage	Primary Use Case
LoRA	Short	Medium–High	High	Low	Balanced performance–storage trade-off
DreamBooth	Long	High	Medium	High	High-fidelity, resource-intensive tasks
HyperNetwork	Medium	Medium	Medium	Medium	Preserving original model characteristics
Textual Inversion	Shortest	Medium	High	Very Low	Quick adaptation, low-resource scenarios

2.3.2. Training Parameters and Phased Implementation

We fine-tuned SDXL 1.0 [21] using the lora-scripts (v1.7.3, based on Kohya-ss sd-scripts) [41]. To balance generation quality with computational efficiency, we implemented several optimization strategies. For memory efficiency, we enabled the XFormers attention mechanism and fp16 mixed-precision training, along with gradient checkpointing to manage GPU memory usage. To preserve image quality across varying aspect ratios, we enabled bucketing with side lengths ranging from 256 to 2560 px, allowing the model to train on images at their original proportions. For optimization, we employed the AdaFactor optimizer [42] with gradient accumulation steps set to 1. AdaFactor was selected for its adaptive learning-rate schedule and memory-efficient parameter updates, which maintain stable convergence during large-scale fine-tuning. Training was conducted for 20 epochs. For each style, we selected the best-performing saved model version (the optimal checkpoint) using a strategy that combined loss curve monitoring and visual quality assessment. Detailed convergence analysis and selection results are presented in Section 3.1.

Based on the unified architecture described above, Stage 1 (Parameter Optimization) focused on the differentiated configuration of experimental models. To implement the controlled-variable design, we constructed a total of 25 experimental models: for the captioning variable group, the resolution was fixed at 1024 × 1024 px to isolate pixel-level effects; conversely, for the resolution variable group, the BLIP captioning method was used uniformly to exclude semantic interference. This configuration ensured that the subsequent evaluation could precisely attribute variations in generation quality to specific independent variables.

In Stage 2 (Model Development), we applied the "optimal parameter combination" identified in Stage 1 to the full-dataset training. The final GardenDiff model integrated the SDC system with a high-resolution configuration of 1536 × 1536 px (see Section 3.3, the rationale for the selection). This training was conducted on the complete dataset of 1250 images to achieve a deep translation of domain-specific design knowledge.

2.4. Evaluation Methods

This study establishes a comprehensive evaluation framework that combines objective metrics with subjective evaluation to validate the effects of training parameters and the performance advantages of specialized models through two experimental stages.

2.4.1. Evaluation Methods for Parameter Optimization Experiments

The parameter optimization experiments employed both objective metrics and subjective evaluation to assess the effects of training caption systems and training resolution on generation quality.

For objective metrics, we employed the CLIP (Contrastive Language–Image Pre-training) model [29] to calculate semantic similarity between generated images and text descriptions, with scores ranging from 0 to 1 (higher scores indicating better semantic alignment). The comprehensive Image Quality Score integrates five technical dimensions: frequency domain quality [43], multi-scale gradient magnitudes [44], edge coherence [45], local band-limited contrast [46], and sharpness [47]. Each dimension is standardized, and the arithmetic meaning is calculated to derive an overall score ranging from 0 to 5. Equal weighting was applied across the five dimensions, as each captures a complementary aspect of image quality and no domain-specific basis exists for prioritizing one over another. The detailed definitions of these metrics are listed in Table 3.

Table 3. Evaluation metrics for parameter optimization experiments.

Experiment Stage	Evaluation Metric	Metric Description	Evaluation Type	Scoring Range & Direction
Caption System Experiment	CLIP Score	Semantic alignment between generated images and text descriptions	Objective metrics	0–1, higher is better
	Spatial Rationale	Rationality of spatial layout, functional zoning, and element configuration	Subjective evaluation	1–7, higher is better
Training Resolution Experiment	Comprehensive Image Quality Score	Comprehensive assessment based on frequency domain quality, multi-scale band-limited contrast, and sharpness	Objective metrics	0–5, higher is better
	Scale Coherence	Accuracy of proportional relationships among design elements such as buildings, plants, and paving	Subjective evaluation	1–7, higher is better

Subjective evaluation adopted the 7-point Scenic Beauty Estimation (SBE) framework [48], where 1 represents very dissatisfied and 7 represents very satisfied. To ensure evaluation accuracy, we distinguish between potentially confusing metrics: Scale Coherence focuses on whether the proportional relationships of design elements conform to real-world expectations (for example, tree heights should be 2–3 times that of single-story buildings), while Spatial Rationale focuses on the design logic of spatial layout and element configuration (for example, moon gates should connect different spatial layers rather than exist in isolation).

The questionnaires for this stage consisted of two groups. The first group evaluated Scale Coherence by comparing generation quality across three training resolutions (768 × 768, 1024 × 1024, and 1536 × 1536 px). The second group evaluated Spatial Rationale by comparing generation quality across three caption systems (WD1.4, BLIP, and SDC). Each questionnaire group covered 5 garden styles, with 2 image sets generated per style, totaling 60 images.

2.4.2. Evaluation Methods for Multi-Model Comparison Experiments

The multi-model comparison experiments employed subjective evaluation to comprehensively assess performance differences between the specialized model and general-purpose models. Two evaluator groups assessed the generated images from parallel perspectives using the 7-point SBE scale (Table 4).

Table 4. Evaluation metrics for multi-model comparison experiments.

Experiment Stage	Evaluation Metric	Metric Description	Evaluation Type	Scoring Range & Direction
Multi -model comparison experiment	Design Rationale	Integrated assessment of Scale Coherence and Spatial Rationale, comprehensively evaluating spatial layout, proportional relationships, and element configuration	Subjective evaluation	1–7, higher is better
	Design Professionalism	Professional competence and technical maturity of the design		
	Design Accuracy	Completeness and precision of style characteristics and design elements		
	Design Satisfaction	Overall acceptability and practical application potential of the design | Subjective evaluation		
Multi-model comparison (Public)	Scenic Beauty	Visual pleasantness, color harmony, and first-impression aesthetics		
	Recreational Appeal	Realism and appeal motivating visiting or recreational use		
	Style Recognition	Visual distinctiveness enabling identification of garden style		

Subjective evaluation used the same 7-point SBE scale as that in Section 2.4.1. Design Rationale integrates the Scale Coherence and Spatial Rationale metrics from Section 2.4.1, providing a comprehensive assessment of design quality from both proportional relationships and spatial logic perspectives. Design Professionalism evaluates technical competence, such as whether plant arrangements conform to ecological requirements and material combinations demonstrate professional standards. Design Satisfaction assesses the design from the end-user perspective, evaluating the integrated performance of functionality and aesthetics (Table 4) (see Appendix C for details).

The public group assessed three perceptual dimensions designed to reflect the same evaluation constructs as the expert metrics while reducing cognitive load for non-specialist respondents. Scenic Beauty evaluates visual pleasantness and color harmony, serving as a proxy for Design Rationale and Design Professionalism. Recreational Appeal assesses whether the scene feels realistic and inviting, corresponding to Design Satisfaction. Style Recognition evaluates whether viewers can identify the garden style, aligning with Design Accuracy.

The questionnaires for this stage compared generation quality between GardenDiff and two general-purpose models (SDXL 1.0 and DreamShaper XL), covering five garden styles with 2 image sets per style, totaling 30 images. All 36 expert evaluators assessed the full set of 30 images across three models, while the 136 public evaluators assessed only the GardenDiff outputs (10 images).

2.4.3. Questionnaire Survey and Data Analysis

This study recruited two independent evaluator groups. The expert group consisted of 36 graduate students in landscape architecture, All evaluators had 6–7 years of training in landscape architecture, had completed comprehensive graduate curricula, and possessed the capability to independently execute design projects. They were responsible for all evaluation tasks across both experimental stages. The public group comprised 136 non-specialist respondents with diverse backgrounds, recruited via an online platform, who participated only in the Stage 2 multi-model comparison.

The experiments were conducted from September 2024 to January 2025 at a university laboratory with standardized, controlled lighting conditions. All evaluators used identical 27-inch 2K resolution displays. The evaluation sample sets were generated according to the rigorous protocols defined in Section 3.2 (Table 5). These images were subsequently uniformly processed to 1000 × 1000 pixels with display sizes controlled at 18–30 cm, simulating designers' typical image review conditions.

Table 5. Experimental generation matrix and protocols.

Stage	Variable Group	Evaluation Metric	Generation Method	Control Input	Sample Composition	Total
Stage 1	A:Captioning Systems	CLIP Score	T2I	None	5 Styles × 3 Captioning Systems × 20 Prompts	300
		Spatial Rationale			5 Styles × 3 Captioning Systems × 2 Prompts	30
	B: Training Resolutions	Image Quality			5 Styles × 3 Training Resolutions × 20 Prompts	300
		Scale Coherence	T2I + ControlNet (Softedge)	Test Map (w/Scale Ref.)	5 Styles × 3 Training Resolutions × 2 Prompts	30
Stage 2	Comparison	Overall Performance		Test Map	5 Styles × 3 Models × 2 Prompts	30

We designed three independent questionnaire sets corresponding to the evaluation objectives in Sections 2.4.1 and 2.4.2. To ensure data reliability, all 36 expert evaluators completed all three questionnaire sets, with each image receiving 36 independent ratings. The two questionnaires in Section 2.4.1 were administered consecutively with a 10 min break between them, while the Section 2.4.2 questionnaire was administered separately. Each questionnaire session lasts approximately 20–40 min. All 36 expert evaluators completed all three questionnaire sets. The public group completed only the Stage 2 questionnaire for the multi-model comparison.

Before the evaluation, researchers provided uniform explanations of all evaluation metrics and illustrated the judgment criteria with example images. The questionnaires employed a blind test design where evaluation images were randomly arranged by style, without displaying model source information [49], to minimize evaluator bias and order effects. All generated images were produced under identical ControlNet conditions to ensure comparability. Inference parameters, including sampling method, step count, CFG scale, and ControlNet weight, were standardized across all generation tasks (see Appendix E for the full configuration).

We employed reliability and validity assessments to verify the robustness of the evaluation framework. Reliability analysis used Cronbach's alpha coefficient, while validity analysis used the Kaiser–Meyer–Olkin (KMO) test and Bartlett's test of sphericity. The subjective evaluation data demonstrated strong reliability and validity (Cronbach's $\alpha > 0.9$, KMO = 0.92, Bartlett's $p < 0.001$). We note that the use of graduate students rather than practicing professionals as expert evaluators may limit the generalizability of the professional assessment; this limitation is discussed in Section 4.4. Statistical analyses were performed using SPSS 26.0 (IBM Corp., Armonk, NY, USA), and data visualization was completed using Origin 2024 (OriginLab Corp., Northampton, MA, USA).

3. Results

3.1. Training Convergence and Model Selection Results

All 26 model groups were successfully trained, producing 520 total checkpoints across 20 epochs. The training loss curves (Figure 5) revealed two distinct convergence patterns corresponding to the two experimental variables. In the Resolution Group, higher-resolution training (e.g., 1536 px) involved learning from images with substantially more visual details, which naturally required more training iterations and led to slower convergence (Figure 5a). In contrast, the Captioning Group showed largely overlapping loss trajectories (Figure 5b), because the underlying image data remained identical across captioning conditions with only the textual annotations varying.

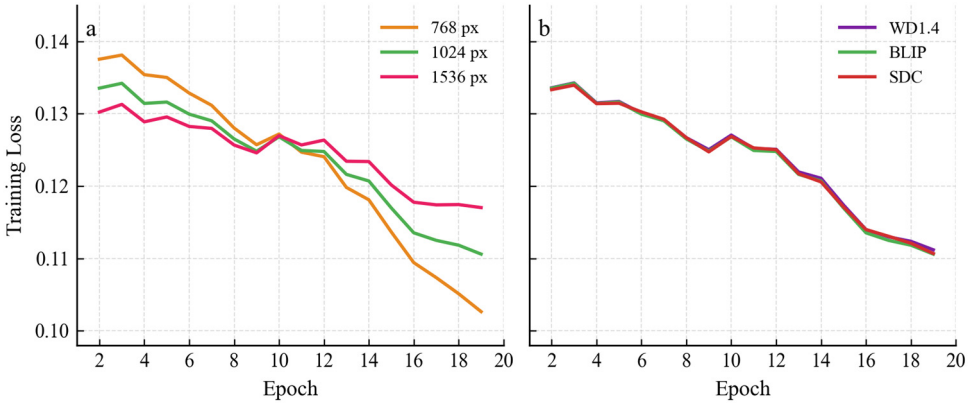


Figure 5. Training loss convergence trajectories across experimental groups. (**a**) Training Resolution Group (Chinese style, BLIP). (**b**) Captioning System Group (Chinese style, 1024 px).

Based on these convergence patterns, we selected the optimal checkpoint for each model through a two-step process. First, we identified the training stage where loss values had stabilized (typically Epochs 15–20). Then, within this stabilized range, we conducted systematic visual comparisons across candidate checkpoints (see Appendix D) to determine the model version with the best balance of semantic accuracy and image quality. The selected models served as the basis for generating evaluation samples, as described in Section 3.2.

3.2. Image Generation Strategy for Evaluation

Using the optimal models selected in Section 3.1, we generated 690 evaluation samples via Text-to-Image (T2I) generation. This section describes the prompt construction and generation strategies applied across the two experimental stages.

To evaluate model generalization rather than memorization of training data, we constructed standardized evaluation prompts for each captioning system (WD1.4, BLIP, and SDC). These prompts were newly assembled by combining high-frequency design keywords extracted from the corresponding training lexicons into novel scenario descriptions that were not present in the training data. For instance, "20 Prompts" in Table 5 indicates that 20 unique design scenarios were generated per style-variable combination. Unified inference parameters (e.g., sampling method and guidance scale) were applied across all generation tasks (see Appendix E).

In the parameter optimization experiments, distinct control strategies were applied to isolate the effects of each variable. For the Captioning System Group, standardized prompts were used without spatial constraints, allowing the model to autonomously construct scene layouts. The resulting images served both objective semantic scoring via CLIP (Figure 6a) and expert evaluation of Spatial Rationale. For the Training Resolution Group, unconstrained generation across three resolutions was used for objective image

quality assessment (Figure 6b), while ControlNet (Softedge) with a standardized test base map was additionally introduced for expert evaluation of Scale Coherence. In the multi-model comparison experiments, the same ControlNet strategy with standardized test base maps was applied to ensure a fair comparison between GardenDiff and two general-purpose models (SDXL 1.0 and DreamShaper XL), so that differences in the output reflected only the model's style rendering capabilities (Figure 6c). The complete generation matrix is summarized in Table 5.



Figure 6. Representative generation samples across evaluation scenarios. (**a**) Semantic Consistency: Comparison of samples generated by different captioning systems (WD1.4, BLIP, SDC). (**b**) Image Quality: Unconstrained samples across resolutions (768–1536 px) for IQS metric calculation. (**c**) Overall Performance: Comparative samples on standardized base maps for Stage 2 model validation.

3.3. Effects of Caption Systems and Training Resolution on Generation Quality

Structured Design Captioning (SDC) significantly outperformed standard captioning methods in both semantic alignment and spatial logic. The SDC system achieved a CLIP Score of 0.2395, marking a 6.5% improvement over BLIP (0.2249) and a 26.1% increase over WD1.4 (0.1899). In terms of Spatial Rationale, SDC captions scored 5.368, surpassing BLIP (4.486) by 19.67% (d = 0.63, 95% CI [0.59, 1.17]) and WD1.4 (3.849) by 39.46% (d = 1.06, 95% CI [1.22, 1.82]) ($p < 0.001$, Figure 7a,b). Visual comparisons of these captioning effects are shown in Figure 8.

High-resolution training demonstrated a substantial advantage in detail fidelity and Scale Coherence compared to lower-resolution configurations. The 1536 × 1536 px configuration achieved a Comprehensive Image Quality Score of 3.424, reflecting a 23.2% improvement over the 768 × 768 px configuration (2.780; d = 1.83, 95% CI [0.57, 0.72]). Regarding Scale Coherence, the 1536 × 1536 px configuration scored 4.963, exceeding the 768 × 768 px baseline (4.077) by 21.7% ($p < 0.001$, Figure 7c,d). Visual comparisons of generation quality are presented in Figure 9, with detailed examples provided in Appendix F.1.

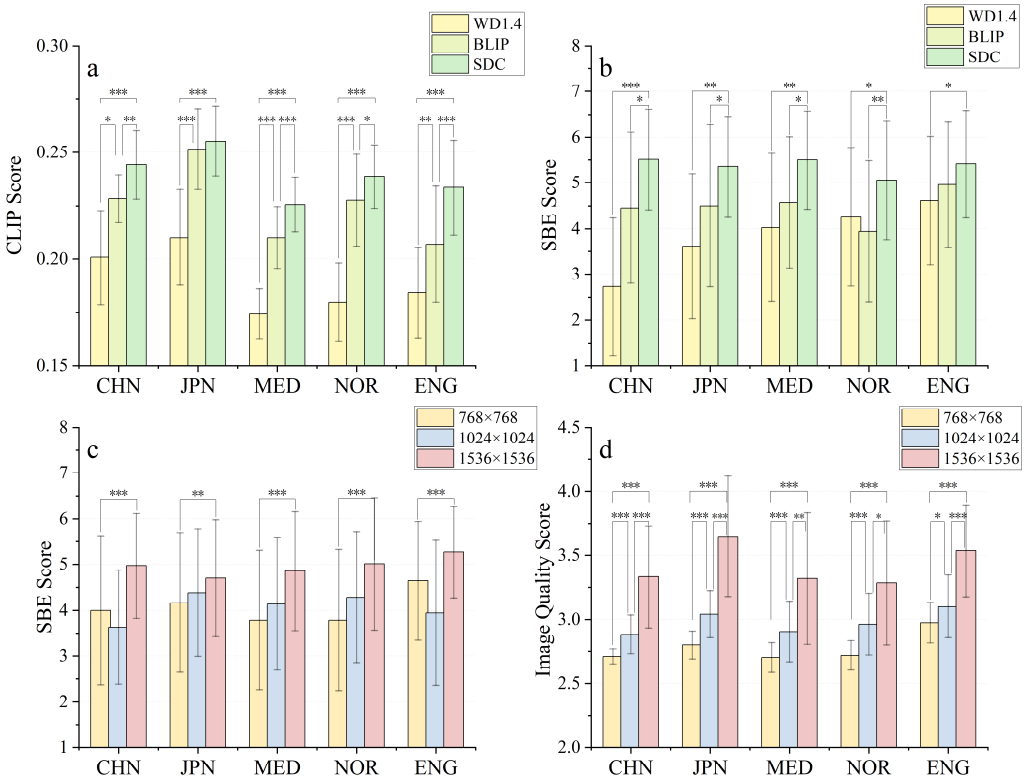


Figure 7. Effects of caption systems and training resolution on generation quality. (**a**) CLIP Score by caption system. (**b**) Spatial Rationale by caption system. (**c**) Image Quality Score by training resolution. (**d**) Scale Coherence by training resolution.* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Training Captions	WD1.4	BLIP	SDC
Generation Prompt	Chinese garden with distinct depth levels,front area with bamboo plants and stone pathway,central area with traditional moon gate and stone bridge over flowing water,rear area with classical architecture and tall trees,creating layered visual depth,		
Chinese Garden			
Evaluation	Characteristic features (e.g., moon gates) are distorted or absent. Spatial arrangement lacks coherence with discontinuous pathways.	Main elements (moon gates, pavilions) are recognizable but spatial relationships remain unclear.	Design elements (moon gates, stone pathways) are accurately rendered with coherent spatial organization.

Figure 8. Example results from caption system experiments (images generated without ControlNet to isolate caption system effects).

Based on these experimental results, we identified the optimal training configuration as SDC captions at 1536 × 1536 px resolution, which was subsequently used to train the final GardenDiff model. This optimal configuration establishes the empirical foundation for addressing research question 2. The following section evaluates the performance of GardenDiff trained with this configuration across multiple garden styles.

Training Resolution	768 × 768	1024 × 1024	1536 × 1536
Generation Prompt	chinese style,upturned eaves,black tile roofs,white walls,Rockery artificial rockwork,Hedge,shrub,bamboo,stone paving,		
Chinese Garden (Overall)			
Chinese Garden (Detail)			
Evaluation	Elements compressed below actual scale (leaves, tiles, stones); spatial density excessive.	Element scales approach actual proportions; spatial organization improved.	Accurate scale representation; appropriate spatial capacity with well-structured composition.

Figure 9. Example results from the training resolution experiment.

3.4. Performance Validation and Multi-Style Analysis of the GardenDiff Model

3.4.1. Comparison of the Performance of GardenDiff Model and Conventional Models

The domain-adapted GardenDiff model demonstrated superior overall performance compared to general-purpose baselines. Based on the optimal configuration identified in Section 3.3, GardenDiff achieved an overall score of 5.06. This represents a 16.4% improvement over SDXL (4.35; Cohen's d = 0.48, 95% CI [0.51, 0.95]) and a 22.4% improvement over DreamShaper XL (4.14; d = 0.63, 95% CI [0.75, 1.20]). This performance advantage extended to all four evaluation dimensions: Design Rationale, Design Professionalism, Design Accuracy, and Overall Satisfaction. In each dimension, GardenDiff consistently outperformed the two general-purpose models (Figure 10). Notably, Japanese gardens presented a distinct exception to this trend. In this specific case, the general-purpose DreamShaper (4.81) performed comparably to GardenDiff (4.75; d = −0.05, 95% CI [−0.61, 0.44]), maintaining consistent scores across all dimensions (4.54–4.88, Figure 11a). Representative generated images are presented in Figure 12, illustrating the visual superiority of the domain-adapted model.

3.4.2. Style-Specific Performance Analysis

GardenDiff exhibited significant performance variations across different garden styles, with Chinese gardens achieving the highest fidelity. Specifically, Chinese gardens received the highest overall score (5.44), particularly excelling in Design Accuracy (5.53). Nordic (5.12) and English (5.03) styles showed moderate performance, while Mediterranean (4.97) and Japanese (4.75) styles scored relatively lower (Figure 11b). In terms of dimensional consistency, Japanese gardens showed the most balanced performance across metrics, whereas Chinese gardens demonstrated the most significant optimization peaks.

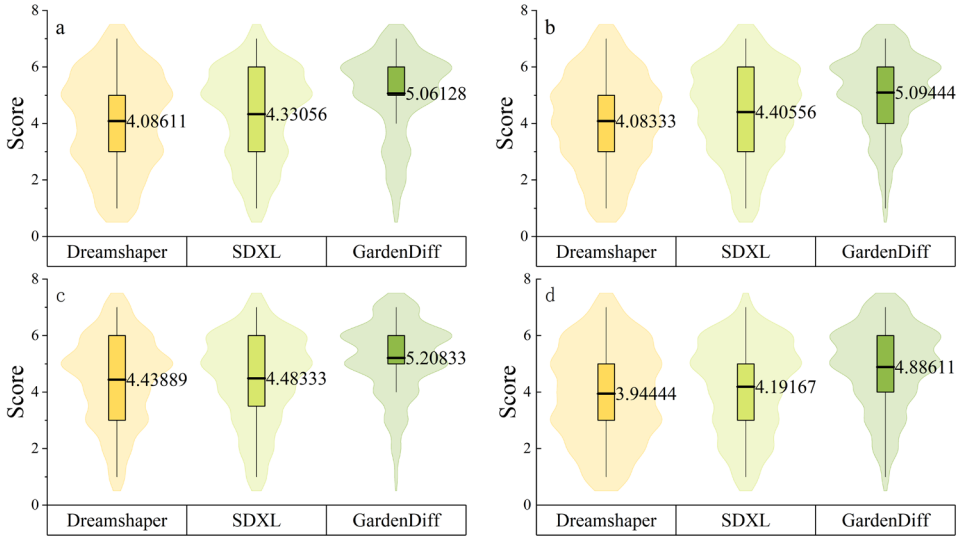


Figure 10. Performance validation of GardenDiff. (**a**) Design Rationale. (**b**) Design Professionalism. (**c**) Design Accuracy. (**d**) Design Satisfaction.

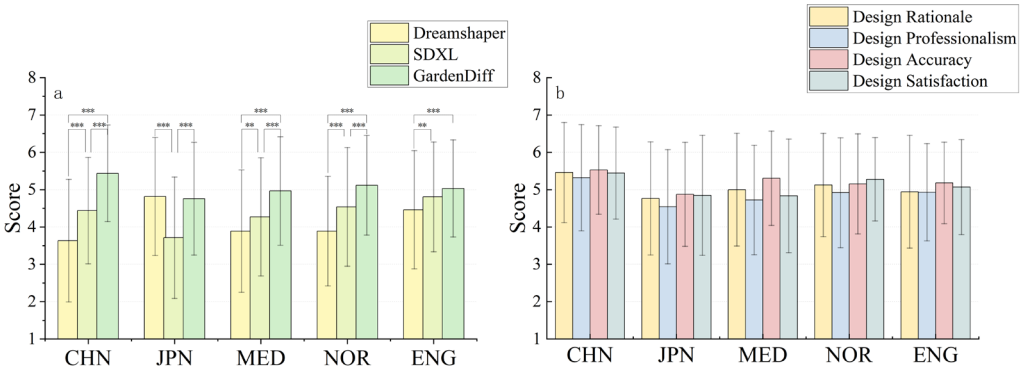


Figure 11. Performance of GardenDiff across different garden styles. (**a**) Mean scores of three models across five styles. (**b**) Four-dimensional scores of GardenDiff across five styles. ** $p < 0.01$, *** $p < 0.001$.

The magnitude of improvement over the SDXL baseline varied significantly across styles. The improvement achieved by GardenDiff ranged from a substantial 22.8% for Chinese gardens to a marginal 4.6% for English gardens. The effect sizes confirm this gradient: Chinese gardens showed the largest effect (d = 0.65, 95% CI [0.44, 1.34]) while English gardens showed a non-significant difference (d = 0.11, 95% CI [−0.33, 0.66]). This variation stems from SDXL's uneven baseline capabilities: the model exhibited strong initial performance on English gardens (4.81) but weaker performance on Chinese (4.43) and Japanese (3.71) styles (Figure 11a). This disparity reveals the effective boundaries of specialized training in different application scenarios. It indicates that domain adaptation offers the greatest marginal returns when the baseline representation is weakest (see Section 4.3 for a detailed discussion).

3.4.3. Expert–Public Evaluation Comparison

Sections 3.4.1 and 3.4.2 assessed GardenDiff from a professional perspective. This section introduces a public evaluation (N = 136) to examine whether the observed quality patterns hold across evaluator backgrounds. The public group rated GardenDiff using three perceptual dimensions: Scenic Beauty, Recreational Appeal, and Style Recognition (Figure 13). The public evaluation data demonstrated strong reliability and validity (Cronbach's α = 0.88–0.90, KMO = 0.88–0.91, Bartlett's $p < 0.001$).

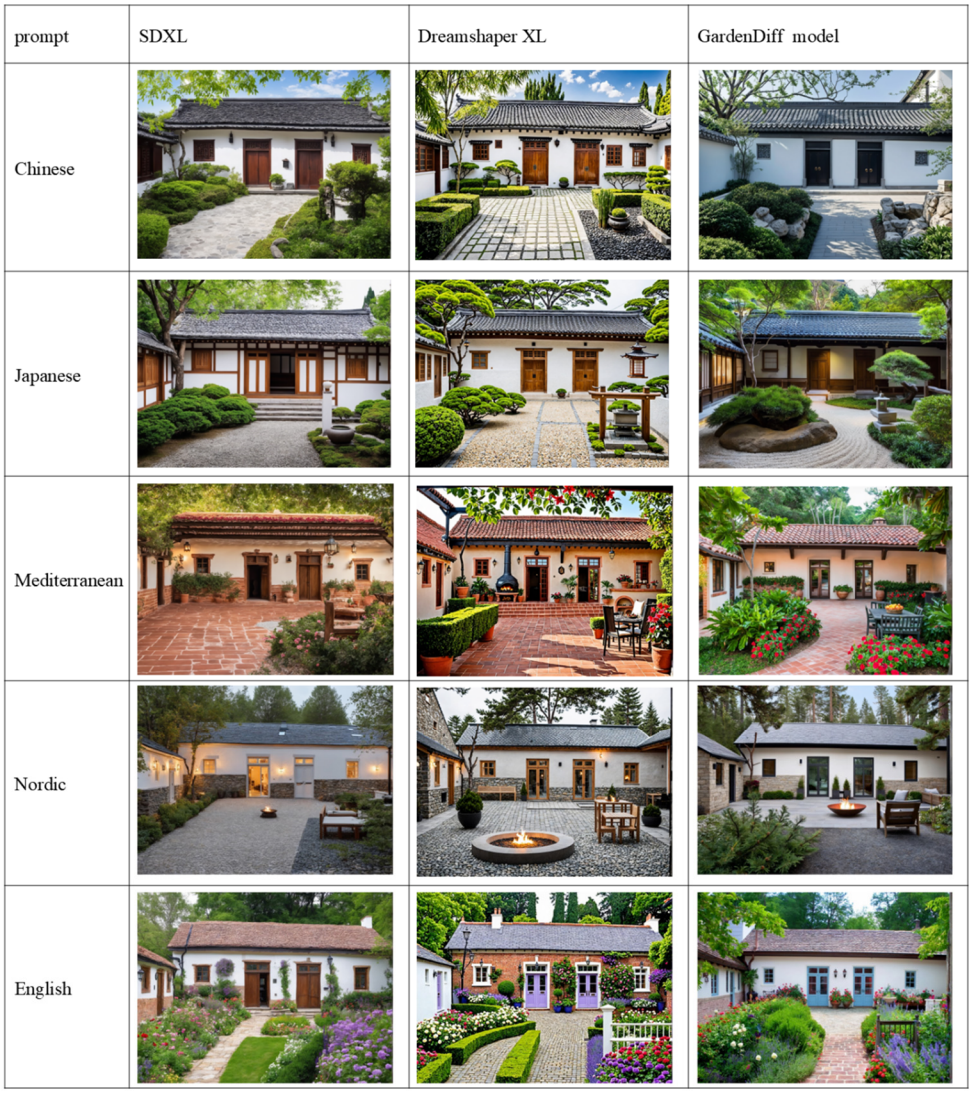


Figure 12. Visual comparison of three models across garden styles (front elevation view). Note: SDXL: Stable Diffusion XL Base 1.0; DreamShaper: community fine-tuned SDXL; GardenDiff: our proposed domain-adapted model. All images use identical prompts and ControlNet settings.

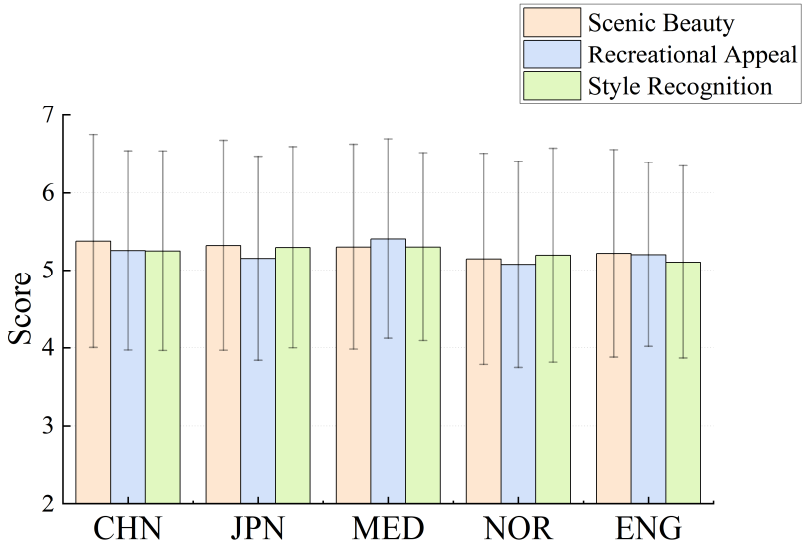


Figure 13. Public evaluation scores of GardenDiff across three perceptual dimensions by garden style.

Overall, the two groups showed convergent trends across most styles, supporting the robustness of GardenDiff's generation quality (Figure 14). Both groups rated Chinese gardens highest or near-highest and placed Nordic and English styles in the middle range. Public overall scores ranged from 5.14 (Nordic) to 5.34 (Mediterranean), with a notably narrower range (0.20) than the expert group (0.68), indicating that non-specialists perceived more uniform quality across styles. For Chinese (d = −0.12, 95% CI [−0.45, 0.21]), Nordic (d = 0.02, 95% CI [−0.35, 0.39]), and English gardens (d = 0.23, 95% CI [−0.13, 0.58]), no significant differences were found between the two groups.

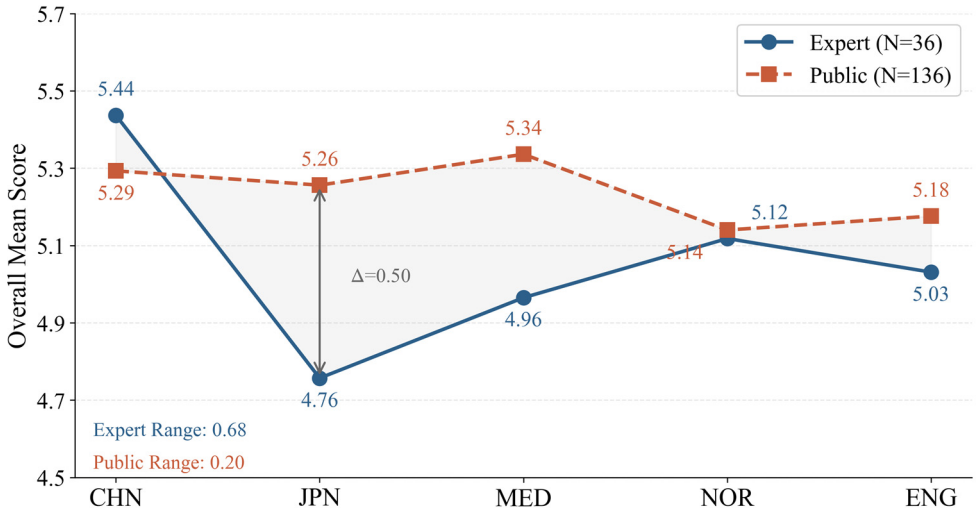


Figure 14. Overall mean score comparison between expert (N = 36) and public (N = 136) evaluations across five garden styles.

Significant divergences emerged in two styles. For Japanese gardens, the public scored significantly higher than experts (d = 0.59, 95% CI [0.20, 0.98]), likely reflecting experts' sensitivity to culturally specific elements such as Karesansui composition. For Mediterranean gardens, the public also scored higher (d = 0.49, 95% CI [0.10, 0.88]), suggesting broad visual appeal despite professional-level accuracy limitations.

These convergences and divergences highlight the complementary value of dual-perspective assessment.

4. Discussion

This study systematically addressed three research questions through a three-stage experimental design. The results demonstrate that: training parameters significantly affect generation quality (answer to research question 1); specialized training strategies offer clear advantages in garden design tasks (answer to research question 2); and the proposed training strategy improves generation quality over general-purpose models across multiple evaluation dimensions, including Design Rationale, Design Professionalism, Design Accuracy, and Design Satisfaction, with the magnitude of gains varying by garden style (answer to research question 3). The following sections analyze the underlying mechanisms of these findings.

4.1. Training Parameter Effects on Generation Quality

Specialized captions significantly enhanced the model's understanding of design elements and spatial organization, as evidenced by the substantial improvements in Spatial Rationale scores compared to generic captions.

This improvement stems from the SDC system's precise semantic representation. While text-to-image models require an accurate cultural context [50], generic annotation

tools often struggle to distinguish subtle cultural differences due to their broad descriptive approach [20]. The SDC system addresses this by bridging precise terminology (e.g., distinguishing 'upturned eaves' from 'shoji screens') with contextual descriptors (e.g., season, viewpoint) to guide spatial coherence. Building on Chen and Zhao's [51] exploration of design feature disentanglement, our system advances this logic by achieving structured encoding of design knowledge into captions, enabling diffusion models to learn cultural semantics more effectively.

Improved semantic understanding directly enhances generation quality. This study found a positive correlation between improvements in CLIP Score (6.5–26.1%) and improvements in Spatial Rationale. Saharia et al. [23] reported similar findings in their Imagen model research, where enhanced language understanding significantly improved image generation quality. This validates that semantic understanding accuracy serves as a critical prerequisite for high-quality design generation.

The effectiveness of SDC stems from the synergy among its three semantic layers [29,35]. The Style-Feature Layer anchors cultural orientation, the Design-Element Layer specifies visual content and spatial structure, and the Auxiliary-Info Layer provides contextual enrichment in a supportive role. These layers are not independent: design elements carry implicit style information (e.g., "karesansui" is exclusive to Japanese gardens), while style descriptors constrain how elements are spatially composed. This inter-layer dependency is consistent with evidence that structured captions outperform flat inputs through more precise semantic alignment [52], and is further supported by the monotonic improvements across the WD1.4–BLIP–SDC gradient in both CLIP Score (0.1899 → 0.2249 → 0.2395) and Spatial Rationale (3.849 → 4.486 → 5.368; Figure 8).

High-resolution training significantly improved detail fidelity and Scale Coherence (Section 3.3). This study found that training resolution affects generation quality by influencing how the model learns scale relationships. Low-resolution training forces the model to learn garden elements (such as plant leaves, roof tiles, and paving stones) in compressed form. During generation, the model populates the scene with excessive miniaturized elements to fill the frame, resulting in distorted spatial density. For example, areas that should accommodate 3–5 ornamental rocks instead contain over 10 small rocks. High-resolution training enables the model to learn accurate proportions of elements, ensuring generated results maintain appropriate relationships between element quantity and spatial capacity.

More importantly, high-resolution training provides the model with more complete spatial information. The model can learn more complex spatial relationships and layout logic among buildings, plants, paving, and other elements in gardens. This ability to learn multi-element spatial relationships is difficult to achieve with low-resolution training.

4.2. The Performance Advantages of GardenDiff

The multi-model comparative experiments validated the effectiveness of specialized training in garden design. Notably, GardenDiff's advantages over DreamShaper XL were particularly significant. DreamShaper XL, as a photorealistic model optimized with large-scale data, performs well in general-purpose scenarios but shows limited performance in the specialized domain of garden design. Research by Lu et al. [53]. indicates that domain-specific adaptation is typically more effective than general optimization. Our results validate this perspective.

This finding aligns with the trend toward domain specialization in landscape architecture AI research. Extending the work of Chen et al. [11], who demonstrated the efficacy of GANs in park green space design, our study applies specialized training to the more culturally nuanced domain of garden design. This progression further reveals the effectiveness of domain-adapted training in handling complex landscape semantics.

Finally, our research resonates with emerging human–AI collaboration frameworks in landscape architecture. Schroth and Maier [5] advocate iterative 'human-in-the-loop' workflows, while Huang et al. [54] highlight the necessity of local refinement. More broadly, recent work on human–AI co-creation emphasizes that AI should function as a collaborative partner that augments rather than replaces designer agency [55–57].GardenDiff supports this paradigm by generating culturally grounded visualization that supplements design-relevant knowledge, which designers can further refine through controlled generation techniques (see Appendix F.2), confirming the value of domain-adapted models in bridging automated generation and professional practice.

4.3. Style-Specific Performance Variations and Its Influencing Factors

The model's training effectiveness across different garden styles is influenced by three factors. First, the diversity of design features determines how much additional information SDC captions can provide over generic captions. Second, the baseline model's initial capabilities determine the additional gains from specialized training. Third, the compatibility of training strategies depends on the alignment between design characteristics and caption methods.

The greater the diversity of design elements, the more refined the feature–semantic correspondences that SDC captions can establish, and consequently, the greater the model's potential for performance improvement. This finding aligns with the consensus in transfer learning research that the richness of source domain data directly enhances knowledge transfer effectiveness [58]. Chinese gardens achieved the highest improvement scores (22.8%) precisely because their design encompasses multiple layers of interconnected elements, including architectural forms, rockeries, and water features [17]. The greater diversity of these elements allows SDC captions to establish more refined feature–semantic correspondences than simpler styles do. Consequently, the model not only learns generic visual features but also masters culture-specific design languages, offering potential pathways for the AI-assisted preservation of traditional design heritage.

Mediterranean and Nordic styles feature moderate quantities of design elements with relatively regularized element combinations. Specialized training achieved significant improvements in both styles (Mediterranean: 4.97; Nordic: 5.12), validating the effectiveness of SDC captions across different complexity levels.

The baseline model's initial capabilities determine the marginal benefits of specialized training. SDXL's pre-training data (LAION-5B) exhibits a structural imbalance with substantial coverage of Western imagery but limited representation of Eastern cultural aesthetics [59]. This baseline disparity results in uneven initial performance: English gardens (4.81) substantially outperformed Chinese (4.43) and Japanese (3.71) styles. This indicates that specialized training yields more pronounced gains in domains where the baseline model lacks sufficient representation. For Chinese gardens, GardenDiff achieved a substantial 22.8% improvement (4.43→5.44); in contrast, English gardens saw only a marginal 4.6% gain (4.81→5.03). This aligns with the findings of Sun and Dredze [60], who demonstrate that fine-tuning yields diminishing returns when models already possess strong capabilities for the target task.

The compatibility of training strategies depends on the alignment between design characteristics and caption methods. Japanese gardens exemplify a case where specialized training yielded limited gains (4.75 for GardenDiff vs. 4.81 for DreamShaper). Unlike the complex element combinations in Chinese gardens, Japanese design is characterized by high standardization and minimalism, typically involving a limited set of fixed elements like rock arrangements and white sand [30]. These characteristics diminish the information advantage of specialized SDC captions, as the semantic gap between precise

domain descriptions and generic tags is minimal. Since the large-scale training data of general-purpose models already covers these simplified, standardized forms, the additional benefit of domain adaptation is limited, as confirmed by DreamShaper's consistently high performance across Japanese garden metrics.

The cross-group validation (Section 3.4.3) further supports this interpretation: Japanese gardens showed the largest expert–public discrepancy ($\Delta = 0.50$, $d = 0.59$), indicating that the quality distinctions in this style operate at a perceptual level that non-specialists do not readily detect. In such cases, SDC's contribution shifts from introducing new semantic knowledge to refining the alignment between model outputs and professional design conventions, which inherently limits the magnitude of measurable improvement.

Collectively, these findings indicate that the effectiveness of domain adaptation is modulated by the interaction among style complexity, baseline coverage, and caption informativeness, and is not uniformly beneficial across all design traditions. Developing complexity-aware strategies that calibrate the degree of specialization to each target style remains an important direction (see Section 5).

4.4. Limitations

We consider the limitations below that could be enhanced in future work. Firstly, regarding dataset scale and coverage: while our curated dataset of 250 high-quality images per style proved effective for LoRA adaptation, future work could expand the dataset to capture broader design variations and mitigate potential overfitting risks. For strictly standardized styles such as Japanese gardens, more differentiated captioning strategies also remain a promising direction [53]. Moreover, the current five styles represent East Asian and European traditions, and the model's applicability to other cultural contexts and non-ideal conditions remains to be examined. Second, regarding computational resources, although high-resolution training at 1536 × 1536 px significantly improves Scale Coherence and detail fidelity, it requires an NVIDIA RTX 4090 with 24 GB VRAM for training and leads to a twofold increase in GPU memory consumption. At inference, however, generating a 1536 × 1536 image with ControlNet on an NVIDIA RTX 2060 laptop with 6 GB VRAM takes only 8 to 15 s via WebUI Forge, suggesting that future work on lightweight training architectures could make the full pipeline more accessible. Third, although the evaluation achieved strong reliability through both expert and public assessments, the expert evaluators were graduate students rather than practicing professional designers, which may limit the representativeness of professional metrics.

5. Conclusions

This study bridges the gap between generic AI capabilities and professional landscape architecture standards. Unlike previous studies that relied on general-purpose models or focused on single-style explorations, we constructed GardenDiff, a domain-adapted model, to systematically quantify how computational strategies interact with landscape design characteristics.

The research yields three principal findings for AI-aided design. First, structured domain-specific captioning serves as the foundational constraint for professional generation. Semantic alignment via SDC and Scale Coherence via high-resolution training are prerequisites that cannot be bypassed by model size alone. Second, specialized training strategies are validated as superior to generic solutions. By embedding domain knowledge into the diffusion process, GardenDiff significantly outperforms baseline models in Design Rationale and professionalism, confirming that domain adaptation is essential for professional-grade applications. Third, design complexity determines the marginal utility of AI training. Culturally complex styles with diverse design elements gain the most from

specialized training, whereas standardized styles with limited element vocabularies show smaller gains. These findings offer a transferable analytical framework: the quantitative separation of captioning and resolution effects demonstrates that semantic precision and visual resolution contribute through distinct mechanisms, a principle applicable to domain adaptation research beyond garden design.

From a practical standpoint, GardenDiff enables designers to rapidly generate culturally grounded visualization that supplements design-relevant knowledge, supporting iterative human–AI collaboration through ControlNet-assisted workflows. The SDC annotation framework itself is not model-specific and can be adapted to other specialized visual domains. The limitations discussed in Section 4.4, including evaluator composition, cultural coverage, and computational cost, suggest directions for future work such as expanding the dataset to additional cultural traditions, incorporating evaluations from practicing professionals, developing complexity-aware training strategies, and conducting ablation studies to isolate the contributions of individual SDC layers. Ultimately, this research provides a technical pathway for preserving diverse landscape heritages, demonstrating that AI can effectively reproduce culturally nuanced design principles when guided by structured domain knowledge.

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

Funding: This research was funded by the National Natural Science Foundation of China, grant number 32501724, and the Talent Research Start-up Project in ZAFU, grant number 2023LFR002. The APC was funded by the Talent Research Start-up Project in ZAFU, grant number 2023LFR002.

Data Availability Statement: The data presented in this study are available on request from the corresponding author due to privacy restrictions.

Acknowledgments: The authors would like to thank all volunteers who participated in the questionnaire survey, the anonymous reviewers for their thoughtful suggestions and comments in helping us improve our manuscript, and the editors for their support during the review process. During the preparation of this manuscript, the author(s) used Claude 4.6 Opus (Anthropic, San Francisco, CA, USA) for the purposes of language translation and linguistic refinement. The authors have reviewed and edited the output and take full responsibility for the content of this publication.

Conflicts of Interest: The authors declare no conflicts of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

Abbreviations

The following abbreviations are used in this manuscript:

AI	Artificial Intelligence
BLIP	Bootstrapping Language–Image Pre-training
CLIP	Contrastive Language–Image Pre-training
ControlNet	ControlNet (proper noun, not an abbreviation)
DreamShaper XL	DreamShaper XL, community fine-tuned diffusion model
GardenDiff	Garden Diffusion Model (proposed in this study)
LAION	Large-scale Artificial Intelligence Open Network
LoRA	Low-Rank Adaptation
SBE	Scenic Beauty Estimation
SDC	Structured Design Captioning

SDXL	Stable Diffusion XL
WD1.4	Waifu Diffusion 1.4 Tagger

Appendix A. Disadvantage Identified in Preliminary Baseline Experiments

Table A1. Identified issues in preliminary experiments.

	Semantic Misalignment	Detail Rendering Deficiency
Description	The models misinterpret design terminology, producing inconsistent styles and ambiguous element layouts.	AI models struggle to render complex design elements accurately.
Baseline Generation Overall View (No Control)		
Baseline Generation Detail View (No Control)		
ControlNet-Assisted Generation Overall View		
ControlNet-Assisted Generation Detail View		

Appendix B. Dataset and Captioning Details

Appendix B.1. Dataset Composition

Each garden style consists of 250 images distributed across five design categories. We sourced all images from professional platforms that offer CC0 licenses, so no permission is required for research use.

Table A2. Training dataset composition.

Garden Style	Design Category	No. of Images	Repeat Times	Data Source
Chinese (250 images)	Architectural Facades	50	20	https://www.freepik.com/ (CC0 License) https://www.pexels.com/zh-cn/ (CC0 License)
	Garden Structures	50	20	
	Paving	40	20	
	Planting	60	20	
	Detail Close-ups	50	15	
Japanese (250 images)	Architectural Facades	23	20	https://www.freepik.com/ (CC0 License) https://unsplash.com/ (CC0 License)
	Garden Structures	66	20	
	Paving	80	20	
	Planting	62	20	
	Detail Close-ups	20	15	
Mediterranean (250 images)	Architectural Facades	60	20	https://www.freepik.com/ (CC0 License) https://unsplash.com/ (CC0 License)) https://www.pexels.com/zh-cn/ (CC0 License)
	Garden Structures	64	20	
	Paving	40	20	
	Planting	52	20	
	Detail Close-ups	34	15	
Nordic (250 images)	Architectural Facades	65	20	https://www.freepik.com/ (CC0 License) https://unsplash.com/ (CC0 License)
	Garden Structures	55	20	
	Paving	40	20	
	Planting	50	20	
	Detail Close-ups	40	15	
English (250 images)	Architectural Facades	45	20	https://www.freepik.com/ (CC0 License) https://www.pexels.com/zh-cn/ (CC0 License)
	Garden Structures	60	20	
	Paving	30	20	
	Planting	75	20	
	Detail Close-ups	40	15	

Appendix B.2. Semantic Annotation Examples

WD1.4 outputs tag-based labels, BLIP generates natural-language captions, and SDC (our proposed Structured Design Captioning) provides hierarchical, design-specific annotations.

Appendix B.3. Annotation Prompt Template

The following template was used to guide GPT-4o in generating initial SDC captions. For each garden style, the bracketed terms were replaced with the corresponding style-specific vocabulary listed in Table 1.

Table A3. Semantic Annotation Examples.

Image Example	Wd1.4	Blip	SDC
	scenery, reflection, tree, outdoors, water, architecture, east_asian_architecture, no_humans, hood, nature, day, plant, grass, forest, hoodie, building	a pond with lily pads and a building in the background with a pond in front of it and a lily pad in the foreground, Cao Buxing, phuoc quan, a detailed matte painting, cloisonnism	chinese courtyard, peaceful, upturned eaves, white walls, black tile roofs, pavilion, corridor, reflecting pool, stones, water lilies, bamboo, lush greenery, serene atmosphere, summer season, tranquil water reflections
	no_humans, tree, scenery, outdoors, traditional_media, real_world_location, photo_background, day, sky, building, bare_tree, east_asian_architecture, architecture, realistic, road,	a small building with a tree in the middle of it and a snow covered ground in front of it, Eishōsai Chōki, kyoto studio, a tilt shift photo, mingei	japanese courtyard design, minimalism, wooden structure, tiled roof, stone paving, raked gravel, pine tree, deciduous trees, bamboo, moss, overcast weather, tranquil atmosphere
	scenery, tree, no_humans, sky, outdoors, plant, water, day, building, blue_sky, statue	a fountain in a formal garden with hedges and potted plants in front of a white building with arches, Enguerrand Quarton, arthouse, a digital rendering, heidelberg school	mediterranean courtyard, elegant style, red clay roof tiles, white facade with arched openings, central stone fountain with lion sculptures, terracotta flower pots with orange trees, trimmed hedges, clear blue sky, formal layout
	scenery, tree, stairs, outdoors, grass, day, no_humans, sunlight, plant, nature, moss, multiple_girls,	a garden with a staircase and a fountain in the middle of it and a stone staircase leading to the upper level, Enguerrand Quarton, enchanting, a flemish Baroque, arts and crafts movement	English courtyard with lush greenery, brick walls, stone stairs, ornate urns, symmetrical stone path, trimmed boxwood hedges, vine-covered walls, sculptures, bright sunny weather, perspective from ground level, strong visual hierarchy
	no_humans, scenery, tree, outdoors, house, building, road, window, ground_vehicle, autumn_leaves, door, bench, plant	a small cabin with a deck and lights on it's side in the woods near a picnic table, Dan Frazier, archdaily, a digital rendering, arts and crafts movement	nordic courtyard, minimalist design, dark wood facade, large glass sliding door, outdoor seating, string lights, wooden deck, natural paving, deciduous trees, autumn leaves, evening setting, warm lighting, open courtyard

As an AI image tagging expert, your goal is to enhance the CLIP model's understanding of images with precise, structured descriptions. This is a set of photographic images of

[STYLE] courtyard design. Please provide a detailed analysis and label the design elements involved in the image, including describing the overall style ([STYLE ATMOSPHERE EXAMPLES]) separately, and then labeling the design element categories: building facades ([FACADE EXAMPLES]), architectural and decorative items ([STRUCTURE EXAMPLES]), paving ([PAVING EXAMPLES]), mountains, rocks, water bodies ([WATER EXAMPLES]), plants ([PLANT EXAMPLES]).

Season and weather, perspective, and project type should be included only if distinctly observable. Architectural styles, materials, and design elements should be highlighted only when clearly distinguishable, avoiding repetition. For plants, describe the plant species in detail and, if there are flowers, describe the color of the flowers. Plants are mentioned only when they are a prominent part of the scene.

Ensure the description: is accurate and direct, avoiding speculative or vague language; stays within 30–75 words to maintain brevity and focus; uses clear, natural language without unnecessary prefatory phrases; seamlessly transitions from general to specific. All output vocabulary is lowercase English words separated by commas and without periods.

Appendix C. Evaluation Questionnaire Samples

Appendix C.1. Parameter Optimization Experiment Questionnaire Sample

The parameter optimization experiments were evaluated through two questionnaire groups. Group 1 assessed Scale Coherence by comparing images generated at three training resolutions (768 × 768, 1024 × 1024, and 1536 × 1536 px). Group 2 assessed Spatial Rationale by comparing images generated using three captioning systems (WD1.4, BLIP, and SDC).



Figure A1. Representative questionnaire page from the parameter optimization experiment.

Appendix C.2. Multi-Model Comparison Experiment Questionnaire Sample

The multi-model comparison experiment evaluated GardenDiff against two general-purpose models (SDXL 1.0 and DreamShaper XL) across four dimensions: Design Rationale, Design Professionalism, Design Accuracy, and Design Satisfaction.

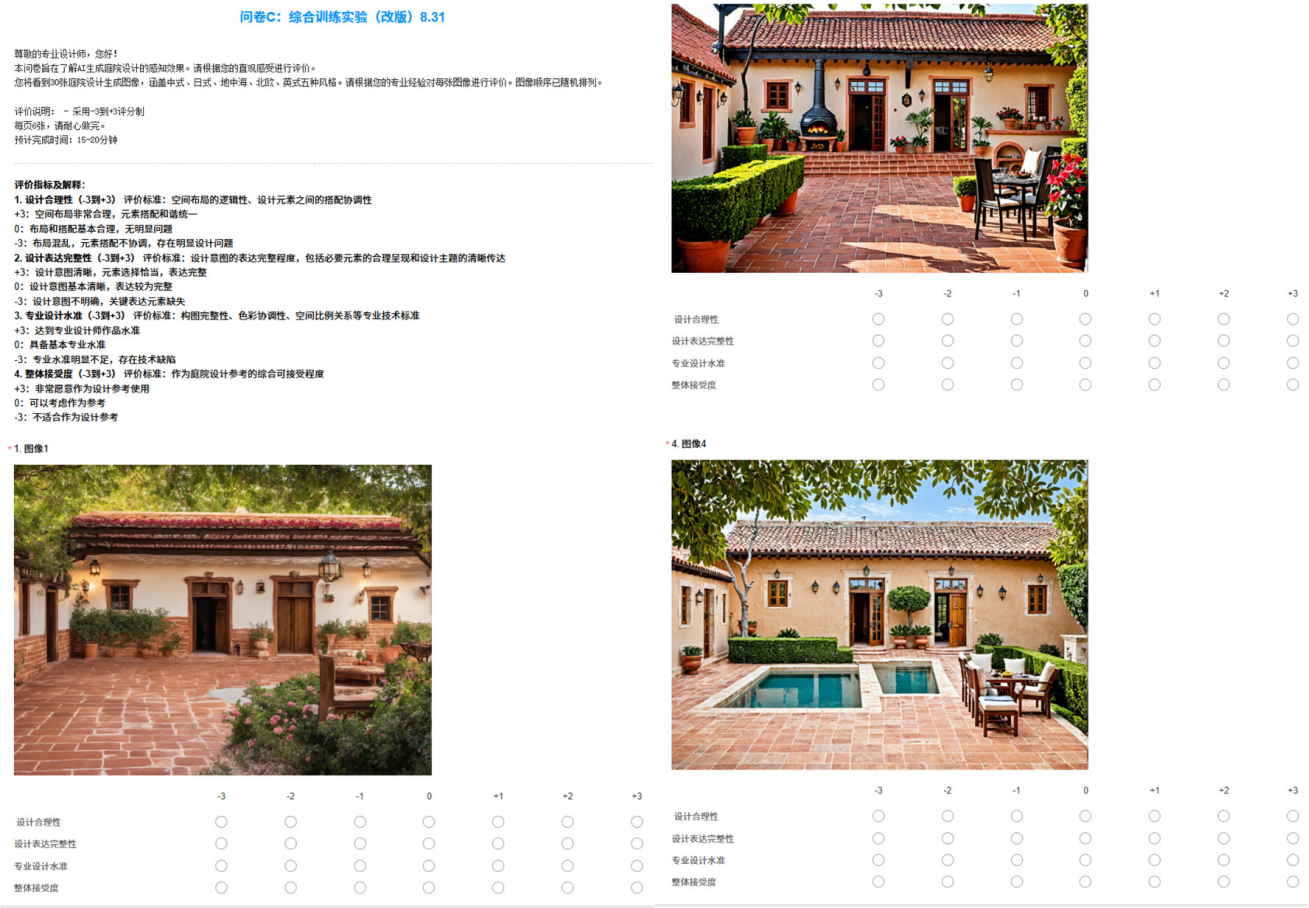


Figure A2. Representative questionnaire page from the multi-model comparison experiment.

Appendix D. Checkpoint Selection Details

To select the optimal checkpoint for each model, we conducted systematic visual comparisons across candidate checkpoints within the loss-stabilized training range (typically Epochs 15–20). Figure A3 illustrates the selection process using the Chinese garden style as an example. The top panel presents an overview grid of generation results across all styles and epochs, while the middle and bottom panels show progressive close-up comparisons of candidate checkpoints. The final selection prioritized the model version demonstrating the best balance of semantic accuracy and visual quality.

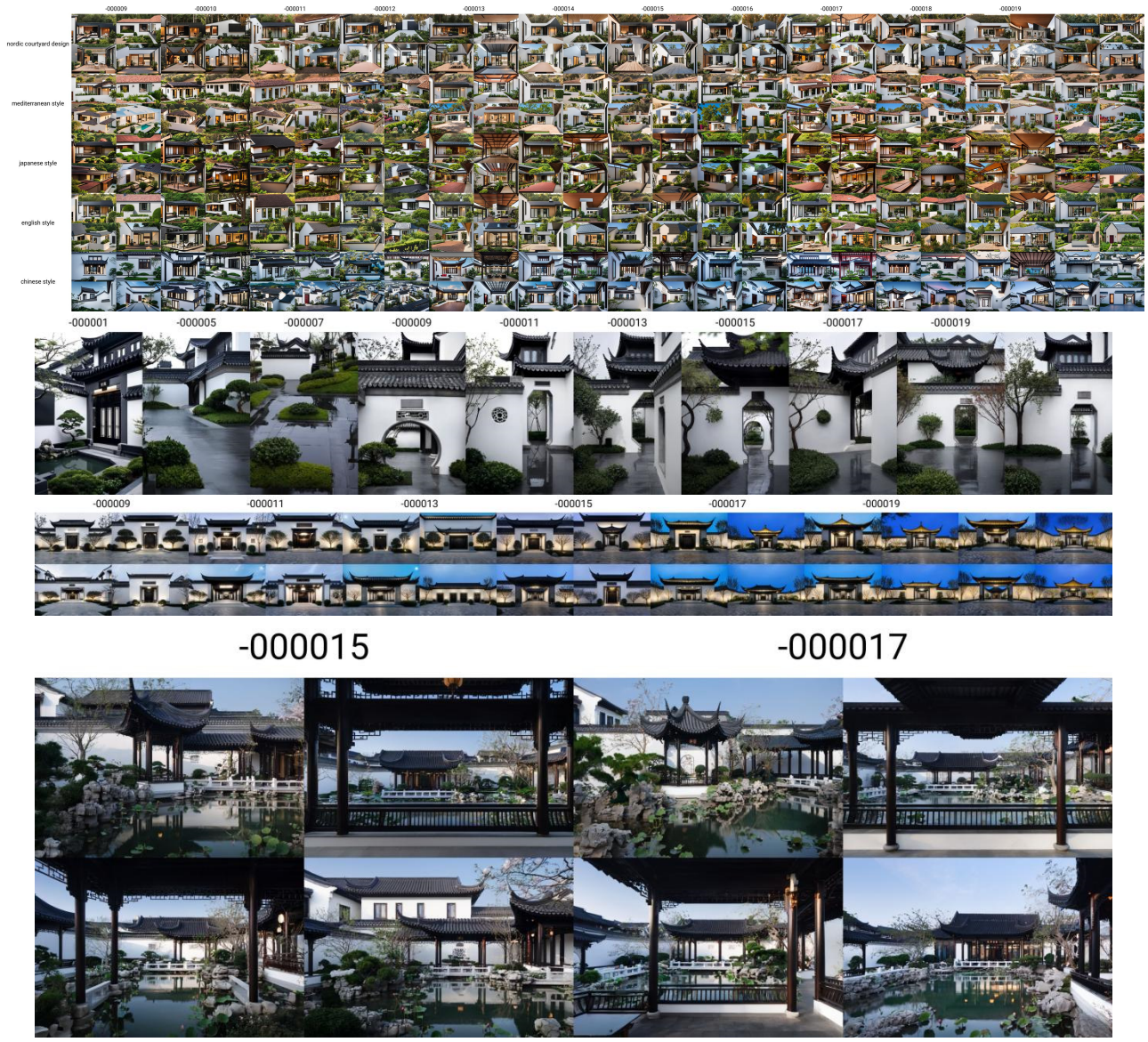


Figure A3. Visual comparison of candidate checkpoints across training epochs (Chinese garden style example).

Appendix E. Inference Parameters

This appendix summarizes the hardware, training, inference, and ControlNet configurations used in the GardenDiff experiments.

Table A4. Detailed configuration of the GardenDiff experiments.

Category	Parameter	Value/Setting	Note
Hardware	GPU	RTX4090(24 GB)/ 64 GBRAM	CUDA acceleration enabled
	Training toolkit	Kohya-ss(sd-scripts)	Fine-grained control over training parameters
	Inference UI	WebUIForge(f2.0.1)	CoreVersion:gf5330788
Model training	Base model	SDXL1.0(VAEFix)	Base model version
	Bucket resolution	1536 × 1536(Bucketing)	Covers 256–2560 px with multiple aspect ratios

Table A4. *Cont.*

Category	Parameter	Value/Setting	Note
Model training	Learning rate	1	Compatible with AdaFactor optimizer
	Optimizer	AdaFactor	Adaptive learning rate schedule
	Training epochs	20Epochs	Loss monitoring + visual review for checkpoint selection
Inference	Sampling method	DPM++SDEKarras	Suitable for generating fine-detailed textures
	Sampling steps	20 Steps	Balances generation speed and detail fidelity
	Generation resolution	1720 × 1280/ 1328 × 1760	Matches training resolution density
	CFG scale	4.0–5.0	Low values to achieve natural color rendering
	Seed	−1(Random)	Multiple images generated per prompt; manually selected
ControlNet	Preprocessor	Softedge_teed	Processor resolution: 2048 px
	Control model (example)	canny-sdxl-V2.0	Canny model used for edge-based soft control
	Control weight	0.5–0.8(dynamic)	Adjusted based on base map complexity

Appendix F. Additional Technical Analysis

Appendix F.1. Image Generation Stage Details

To address potential impacts of generation resolution differences, we provide the following supplementary analysis.

The comparison reveals that generation resolution does affect output quality within the same model—higher inference resolutions add surface detail, but the training resolution governs overall scale accuracy and element proportions.

Table A5. Generation resolution comparison across models.

Generation Resolution	768 × 768	1024 × 1024	1536 × 1536
Generation Prompt	chinese style, upturned eaves, black tile roofs, white walls, Rockery artificial rockwork, Hedge, shrub, bamboo, stone paving		
SDXL Base 1.0			
768 × 768 Chinese Mode			

Table A5. *Cont.*

Generation Resolution	768 × 768	1024 × 1024	1536 × 1536
1024 × 1024 Chinese Model			
1536 × 1536 Chinese Model			

Critically, generating at resolutions outside the training resolution range causes compositional defects and scale distortions. Therefore, increasing training resolution not only improves image quality but also enhances model controllability across higher resolution ranges during generation.

Appendix F.2. Control Plugin Impact Analysis

We validated training methodology robustness by comparing ControlNet-Canny V2 [61] and MistoLine [62] using the Softedge_teed preprocessor [63].

Both control models produced consistent results, confirming methodology generalizability.

Table A6. Control model comparison.

	Canny V2	MistoLine
Base Model—Chinese		
Chinese		

Table A6. *Cont.*

	Canny V2	MistoLine
Japanese		
Mediterranean		
Nordic		
English		

References

1. Dong, X.; Geng, L. Nature Deficit and Mental Health among Adolescents: A Perspectives of Conservation of Resources Theory. *J. Environ. Psychol.* **2023**, *87*, 101995. [CrossRef]
2. Seastedt, H.; Schuetz, J.; Perkins, A.; Gamble, M.; Sinkkonen, A. Impact of Urban Biodiversity and Climate Change on Children's Health and Well Being. *Pediatr. Res.* **2025**, *98*, 452–457. [CrossRef]
3. Fu, E.; Zhou, J.; Ren, Y.; Deng, X.; Li, L.; Li, X.; Li, X. Exploring the Influence of Residential Courtyard Space Landscape Elements on People's Emotional Health in an Immersive Virtual Environment. *Front. Public Health* **2022**, *10*, 1017993. [CrossRef] [PubMed]
4. Soflaei, F.; Shokouhian, M.; Soflaei, A. Traditional Courtyard Houses as a Model for Sustainable Design: A Case Study on BWhs Mesoclimate of Iran. *Front. Archit. Res.* **2017**, *6*, 329–345. [CrossRef]
5. Schroth, O.; Maier, A. Integrating Generative Artificial Intelligence into the Landscape Architecture Design Process. *J. Digit. Landsc. Archit.* **2025**, *10*, 665–675.
6. Wang, Q.; Liang, Y.; Zheng, Y.; Xu, K.; Zhao, J.; Wang, S. Generative AI for Urban Planning: Synthesizing Satellite Imagery via Diffusion Models. *Comput. Environ. Urban Syst.* **2025**, *122*, 102339. [CrossRef]
7. Jang, S.; Roh, H.; Lee, G. Generative AI in Architectural Design: Application, Data, and Evaluation Methods. *Autom. Constr.* **2025**, *174*, 106174. [CrossRef]
8. Yang, L.; Zhang, Z.; Song, Y.; Hong, S.; Xu, R.; Zhao, Y.; Zhang, W.; Cui, B.; Yang, M.-H. Diffusion Models: A Comprehensive Survey of Methods and Applications. *ACM Comput. Surv.* **2024**, *56*, 105. [CrossRef]

9. Zhang, L.; Rao, A.; Agrawala, M. Adding Conditional Control to Text-to-Image Diffusion Models. In *2023 IEEE/CVF International Conference on Computer Vision (ICCV)*; IEEE: New York, NY, USA, 2023; pp. 3813–3824.
10. Zhou, H.; Xiang, S. Applicability Evaluation and Reflection on Artificial Intelligence-Based "Image to Image" Generation of Landscape Architecture Masterplans. *Landsc. Archit. Front.* **2024**, *12*, 58–67. [CrossRef]
11. Chen, R.; Zhao, J.; Yao, X.; Jiang, S.; He, Y.; Bao, B.; Luo, X.; Xu, S.; Wang, C. Generative Design of Outdoor Green Spaces Based on Generative Adversarial Networks. *Buildings* **2023**, *13*, 1083. [CrossRef]
12. Chen, R.; Zhao, J.; Yao, X.; He, Y.; Li, Y.; Lian, Z.; Han, Z.; Yi, X.; Li, H. Enhancing Urban Landscape Design: A GAN-Based Approach for Rapid Color Rendering of Park Sketches. *Land* **2024**, *13*, 254. [CrossRef]
13. Chen, R.; Yi, X.; Zhao, J.; He, Y.; Chen, B.; Liu, F.; Yao, X.; Jiang, X.; Lian, Z.; Li, H. AI for Landscape Planning: Assessing Surrounding Contextual Impact on GAN-Generated Green Land Layouts. *Cities* **2025**, *166*, 106181. [CrossRef]
14. Ye, X.; Huang, T.; Song, Y.; Li, X.; Newman, G.; Wu, D.J.; Zeng, Y. Generating Conceptual Landscape Design via Text-to-Image Generative AI Model. *Environ. Plan. B Urban Anal. City Sci.* **2025**, *52*, 1903–1919. [CrossRef]
15. Chen, F.; Mai, M.; Huang, X.; Li, Y. Enhancing the Sustainability of AI Technology in Architectural Design: Improving the Matching Accuracy of Chinese-Style Buildings. *Sustainability* **2024**, *16*, 8414. [CrossRef]
16. Liang, J. The Application of Artificial Intelligence-Assisted Technology in Cultural and Creative Product Design. *Sci. Rep.* **2024**, *14*, 31069. [CrossRef]
17. Lu, L.; Liu, M. Exploring a Spatial-Experiential Structure within the Chinese Literati Garden: The Master of the Nets Garden as a Case Study. *Front. Archit. Res.* **2023**, *12*, 923–946. [CrossRef]
18. Hoyle, H.E. Climate-Adapted, Traditional or Cottage-Garden Planting? Public Perceptions, Values and Socio-Cultural Drivers in a Designed Garden Setting. *Urban For. Urban Green.* **2021**, *65*, 127362. [CrossRef]
19. Zador, A.; Escola, S.; Richards, B.; Ölveczky, B.; Bengio, Y.; Boahen, K.; Botvinick, M.; Chklovskii, D.; Churchland, A.; Clopath, C.; et al. Catalyzing Next-Generation Artificial Intelligence through NeuroAI. *Nat. Commun.* **2023**, *14*, 1597. [CrossRef] [PubMed]
20. Ananthram, A.; Stengel-Eskin, E.; Bansal, M.; McKeown, K. See It from My Perspective: How Language Affects Cultural Bias in Image Understanding. In Proceedings of the International Conference on Learning Representations 2025 (ICLR 2025), Singapore, 24–28 April 2024.
21. Podell, D.; English, Z.; Lacey, K.; Blattmann, A.; Dockhorn, T.; Müller, J.; Penna, J.; Rombach, R. SDXL: Improving Latent Diffusion Models for High-Resolution Image Synthesis. In Proceedings of the International Conference on Learning Representations 2024 (ICLR 2024), Vienna, Austria, 7–11 May 2024.
22. Rombach, R.; Blattmann, A.; Lorenz, D.; Esser, P.; Ommer, B. High-Resolution Image Synthesis with Latent Diffusion Models. In *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*; IEEE: New York, NY, USA, 2022; pp. 10674–10685.
23. Saharia, C.; Chan, W.; Saxena, S.; Li, L.; Whang, J.; Denton, E.L.; Ghasemipour, K.; Gontijo Lopes, R.; Karagol Ayan, B.; Salimans, T.; et al. Photorealistic Text-to-Image Diffusion Models with Deep Language Understanding. In *Advances in Neural Information Processing Systems 35*; Koyejo, S., Mohamed, S., Agarwal, A., Belgrave, D., Cho, K., Oh, A., Eds.; Curran Associates, Inc.: New York, NY, USA, 2022; pp. 36479–36494.
24. SmilingWolf Wd-v1-4-Vit-Tagger-V2. Available online: https://huggingface.co/SmilingWolf/wd-v1-4-vit-tagger-v2 (accessed on 16 November 2025).
25. Lyu, M.; Yang, Y.; Hong, H.; Chen, H.; Jin, X.; He, Y.; Xue, H.; Han, J.; Ding, G. One-Dimensional Adapter to Rule Them All: Concepts, Diffusion Models and Erasing Applications. In *2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*; IEEE: New York, NY, USA, 2024; pp. 7559–7568.
26. Wu, W.; Zhao, Y.; Chen, H.; Gu, Y.; Zhao, R.; He, Y.; Zhou, H.; Shou, M.Z.; Shen, C. DatasetDM: Synthesizing Data with Perception Annotations Using Diffusion Models. In Proceedings of the Advances in Neural Information Processing Systems 36 (NeurIPS 2023), New Orleans, LA, USA, 10–16 December 2023.
27. Lykon DreamShaper XL. Available online: https://huggingface.co/Lykon/dreamshaper-xl-1-0 (accessed on 17 November 2025).
28. 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* **2022**, arXiv:2106.09685.
29. Radford, A.; Kim, J.W.; Hallacy, C.; Ramesh, A.; Goh, G.; Agarwal, S.; Sastry, G.; Askell, A.; Mishkin, P.; Clark, J.; et al. Learning Transferable Visual Models from Natural Language Supervision. In Proceedings of the 38th International Conference on Machine Learning, Virtual Event, 18–24 July 2021.
30. Chen, S.; Dewancker, B.J. The Influence of Zen Buddhism and Ink Wash Painting on Japanese Gardens during the Medieval Japan. *J. Asian Archit. Build. Eng.* **2025**, *24*, 5024–5036. [CrossRef]
31. Van Tonder, G.J.; Lyons, M.J.; Ejima, Y. Visual Structure of a Japanese Zen Garden. *Nature* **2002**, *419*, 359–360. [CrossRef]
32. Diz-Mellado, E.; López-Cabeza, V.P.; Rivera-Gómez, C.; Galán-Marín, C. Seasonal Analysis of Thermal Comfort in Mediterranean Social Courtyards: A Comparative Study. *J. Build. Eng.* **2023**, *78*, 107756. [CrossRef]
33. Vogiatzakis, I.N.; Terkenli, T.S.; Trovato, M.G.; Abu-Jaber, N. Landscapes in the Eastern Mediterranean between the Future and the Past. *Land* **2018**, *7*, 160. [CrossRef]

34. Zetterman, A. *New Nordic Gardens: Scandinavian Landscape Design*; Thames & Hudson: London, UK, 2017.
35. Li, J.; Li, D.; Xiong, C.; Hoi, S. BLIP: Bootstrapping Language-Image Pre-Training for Unified Vision-Language Understanding and Generation. *arXiv* **2022**, arXiv:2201.12086.
36. Sohl-Dickstein, J.; Weiss, E.; Maheswaranathan, N.; Ganguli, S. Deep Unsupervised Learning Using Nonequilibrium Thermodynamics. In Proceedings of the 32nd International Conference on Machine Learning, Lille, France, 6–11 July 2015.
37. Goodfellow, I.J.; Pouget-Abadie, J.; Mirza, M.; Xu, B.; Warde-Farley, D.; Ozair, S.; Courville, A.; Bengio, Y. Generative Adversarial Nets. In *Advances in Neural Information Processing Systems 27*; Ghahramani, Z., Welling, M., Cortes, C., Lawrence, N., Weinberger, K.Q., Eds.; Curran Associates, Inc.: New York, NY, USA, 2014.
38. Ruiz, N.; Li, Y.; Jampani, V.; Pritch, Y.; Rubinstein, M.; Aberman, K. DreamBooth: Fine Tuning Text-to-Image Diffusion Models for Subject-Driven Generation. In *2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*; IEEE: New York, NY, USA, 2023; pp. 22500–22510.
39. Gal, R.; Alaluf, Y.; Atzmon, Y.; Patashnik, O.; Bermano, A.H.; Chechik, G.; Cohen-Or, D. An Image Is Worth One Word: Personalizing Text-to-Image Generation Using Textual Inversion. *arXiv* **2022**, arXiv:2208.01618.
40. Ruiz, N.; Li, Y.; Jampani, V.; Wei, W.; Hou, T.; Pritch, Y.; Wadhwa, N.; Rubinstein, M.; Aberman, K. HyperDreamBooth: HyperNetworks for Fast Personalization of Text-to-Image Models. In *2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*; IEEE: New York, NY, USA, 2024; pp. 6527–6536.
41. Kohya-ss Sd-Scripts. Available online: https://github.com/kohya-ss/sd-scripts (accessed on 16 November 2025).
42. Shazeer, N.; Stern, M. Adafactor: Adaptive Learning Rates with Sublinear Memory Cost. In Proceedings of the 35th International Conference on Machine Learning, Stockholm, Sweden, 10–15 July 2018.
43. Ahmed, N.; Natarajan, T.; Rao, K.R. Discrete Cosine Transform. *IEEE Trans. Comput.* **1974**, *100*, 90–93. [CrossRef]
44. Duda, R.O.; Hart, P.E. *Pattern Classification and Scene Analysis*; Wiley: New York, NY, USA, 1973.
45. Canny, J. A Computational Approach to Edge Detection. *IEEE Trans. Pattern Anal. Mach. Intell.* **1986**, *PAMI-8*, 679–698. [CrossRef]
46. Peli, E. Contrast in Complex Images. *J. Opt. Soc. Am. A* **1990**, *7*, 2032–2040. [CrossRef]
47. Pertuz, S.; Puig, D.; Garcia, M.A. Analysis of Focus Measure Operators for Shape-from-Focus. *Pattern Recognit.* **2013**, *46*, 1415–1432. [CrossRef]
48. Daniel, T.C. *Measuring Landscape Esthetics: The Scenic Beauty Estimation Method*; USDA Forest Service, Rocky Mountain Forest and Range Experiment Station: Fort Collins, CO, USA, 1976.
49. Lothian, A. Landscape and the Philosophy of Aesthetics: Is Landscape Quality Inherent in the Landscape or in the Eye of the Beholder? *Landsc. Urban Plan.* **1999**, *44*, 177–198. [CrossRef]
50. Jeong, S.; Choi, I.; Yun, Y.; Kim, J. Culture-TRIP: Culturally-Aware Text-to-Image Generation with Iterative Prompt Refinement. *arXiv* **2025**, arXiv:2502.16902.
51. Chen, R.; Luo, X.; Zhao, J. Research on the Adaptability of Generative Algorithm in Generative Landscape Design. *Landsc. Archit.* **2024**, *31*, 12–23. (In Chinese) [CrossRef]
52. Dee, C. *Form and Fabric in Landscape Architecture: A Visual Introduction*; Spon Press: Abingdon, UK, 2001.
53. Lu, W.; Luu, R.K.; Buehler, M.J. Fine-Tuning Large Language Models for Domain Adaptation: Exploration of Training Strategies, Scaling, Model Merging and Synergistic Capabilities. *npj Comput. Mater.* **2025**, *11*, 84. [CrossRef]
54. Huang, R.; Lin, H.; Chen, C.; Zhang, K.; Zeng, W. PlantoGraphy: Incorporating Iterative Design Process into Generative Artificial Intelligence for Landscape Rendering. In *CHI '24: Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems*; ACM: New York, NY, USA, 2024; pp. 1–19.
55. Koch, J.; Lucero, A.; Hegemann, L.; Oulasvirta, A. May AI? Design Ideation with Cooperative Contextual Bandits. In *CHI '19: Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*; Association for Computing Machinery: New York, NY, USA, 2019; pp. 1–12.
56. Weisz, J.D.; Muller, M.; He, J.; Houde, S. Toward General Design Principles for Generative AI Applications. *arXiv* **2023**, arXiv:2301.05578. [CrossRef]
57. De Peuter, S.; Oulasvirta, A.; Kaski, S. Toward AI Assistants That Let Designers Design. *AI Mag.* **2023**, *44*, 85–96. [CrossRef]
58. Zhuang, F.; Qi, Z.; Duan, K.; Xi, D.; Zhu, Y.; Zhu, H.; Xiong, H.; He, Q. A Comprehensive Survey on Transfer Learning. *Proc. IEEE* **2021**, *109*, 43–76. [CrossRef]
59. Schuhmann, C.; Beaumont, R.; Vencu, R.; Gordon, C.; Wightman, R.; Cherti, M.; Coombes, T.; Katta, A.; Mullis, C.; Wortsman, M.; et al. LAION-5B: An Open Large-Scale Dataset for Training next Generation Image-Text Models. In *Advances in Neural Information Processing Systems 35*; Koyejo, S., Mohamed, S., Agarwal, A., Belgrave, D., Cho, K., Oh, A., Eds.; Curran Associates, Inc.: New York, NY, USA, 2022; pp. 25278–25294.
60. Sun, K.; Dredze, M. Amuro & Char: Analyzing the Relationship between Pre-Training and Fine-Tuning of Large Language Models. In *Proceedings of the 10th Workshop on Representation Learning for NLP (RepL4NLP-2025)*; Association for Computational Linguistics: Albuquerque, NM, USA, 2025; pp. 131–151.

61. Xinsir Xinsir/Controlnet-Canny-Sdxl-1.0. 2024. Available online: https://huggingface.co/xinsir/controlnet-canny-sdxl-1.0 (accessed on 25 May 2026).
62. TheMistoAI/MistoLine. 2023. Available online: https://huggingface.co/TheMistoAI/MistoLine (accessed on 25 May 2026).
63. Soria, X.; Li, Y.; Rouhani, M.; Sappa, A.D. Tiny and Efficient Model for the Edge Detection Generalization. In *2023 IEEE/CVF International Conference on Computer Vision Workshops (ICCVW)*; IEEE: New York, NY, USA, 2023.

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.