Article

Not peer-reviewed version

A Knowledge-Augmented Two-Stage Workflow for Architectural Concept-to-Massing Generation and Evaluation

Shangci Sun and Yao Fu *

Posted Date: 26 February 2026

doi: 10.20944/preprints202602.1657.v1

Keywords: architectural concept design; massing proxy images; prototype library; knowledge graph; knowledge-augmented generation; diffusion model; large language models (LLMs); massing composition

Article

A Knowledge-Augmented Two-Stage Workflow for Architectural Concept-to-Massing Generation and Evaluation

Shangci Sun [1] and Yao Fu [2,3,*]

[1] Architectural Association School of Architecture(AA), London WC1B 3ES, UK
[2] School of Architecture and Urban Planning, Shenyang Jianzhu University, Shenyang 110168, China
[3] Key Laboratory of Health and Environmental Performance Improvement, Shenyang Jianzhu University, Shenyang 110168, China
* Correspondence: fuyao@sjzu.edu.cn

Abstract

Large language models (LLMs) and diffusion-based image generators can rapidly produce architectural ideas and imagery, yet translating conceptual narratives into massing composition is often implicit and difficult to reproduce. In this paper, we present a knowledge-augmented two-stage workflow for architectural concept-to-massing generation and evaluation. The outputs are represented as axonometric massing proxy images, which serve as 2D visual proxies for early-stage massing refinement rather than editable 3D models. The workflow integrates a prototype library and Knowledge Graph (KG) routing to map narrative cues into executable strategy and operation tokens and compile stage-specific prompts. Stage 1 produces structural concept sketches emphasizing legible composition, while Stage 2 generates axonometric massing proxy images conditioned on Stage 1 sketches to stabilize composition across candidates. Under a fixed sampling budget, candidates are ranked using a Rubric-based scoring protocol with Top-K selection, and evaluation signals can be written back to update prompt compilation iteratively. Across diverse project briefs, ablation studies demonstrate that knowledge augmentation improves constraint compliance and composition readability while maintaining controlled diversity for early exploration. We report expert ratings with inter-rater agreement and paired statistical tests to support reproducible comparisons.

Keywords: architectural concept design; massing proxy images; prototype library; knowledge graph; knowledge-augmented generation; diffusion model; large language models (LLMs); massing composition

---

1. Introduction

Generative artificial intelligence is rapidly entering the stage of architectural concept design. Large language models (LLMs) can generate a large number of conceptual narratives, strategic points and spatial intention descriptions in a short time. Visual generators such as diffusion models can further convert text into sketch-like images and massing expressions, thus significantly improving the efficiency of early exploration [1,2].

However, the current mainstream approach to generative architectural form design relies on prompt engineering to generate images using foundation models, a method that exhibits significant limitations. The correspondence between narrative intention, spatial strategy and geometric structure in the concept is usually implicit in the model, which makes it difficult to maintain stable massing composition and solid-void logic between different candidates. In addition, generated outputs struggle to maintain consistent massing compositions and void-solid logic across different candidates, and lack interpretable grounds for further refinement or convergence. In other words, large models are good at providing "inspirational outputs", but they are still not enough to convert inspiration into a comparable, selectable and iterable massing composition [3–5].

In architectural design practice, designers usually start from images or narratives, constantly scrutinize massing composition through sketches, and transform vague feelings into clear organizational logic through evaluation and correction. Without a clear intermediate expression and a unified filtering mechanism, the generation process is easy to go to two extremes: first, constantly changing prompts and random seeds to obtain "beautiful images", which makes these images difficult to reproduce and attribute; second, although the output results are diverse, they lack clear constraints and editable structures, thus failing to progress to subsequent refinement stages [6–8]. Therefore, the generation workflow of the conceptual stage not only requires diversified results, but also needs to provide an interpretable "concept-strategy-geometry" connection method, as well as a phased evaluation and filtering mechanism to fairly compare different method settings under the same budget.

In this paper, we present a knowledge-augmented two-stage workflow for architectural concept-to-massing generation and evaluation. The workflow outputs axonometric massing proxy images (hereafter, massing proxy images), which refer to axonometric 2D outputs used as intermediate representations for comparing and iterating massing composition. We introduce lightweight prototype libraries and knowledge graph (KG) as knowledge routing, and explicitly inject evaluation feedback into the subsequent generation process through the write-back mechanism. The prototype library and knowledge graph are used to organize and design knowledge units and their relationships, so that the workflow can extract goals and constraints from project requirements, and select appropriate strategies and operation sets accordingly, and finally compile them into staged prompts. The workflow adopts a two-stage generation process: Stage 1 primarily produces conceptual sketches, emphasizing foundational massing contours and creative concept expression; Stage 2 is based on the operation set extended massing description, and outputs massing proxy images for early discussion, emphasizing the massing composition, void-solid organization and key connections, enabling comparative analysis of different design logics under unified presentation conditions.

In terms of evaluation and selection, this paper adopts Rubric-based Top-K under a fixed sampling budget, so that the quality distribution and stability under different settings can be compared at the same generation cost. At the same time, the workflow will write the evaluation results of each stage back to the prompt template to improve the quality of the generated results. In order to verify the effectiveness of the workflow, this paper analyses the ablation control of multiple architectural design projects, scores them according to stage metrics, compares the performance of different settings in terms of diversity and stability, and provides representative case studies for demonstration and discussion.

The contributions of this paper can be summarized in four points:

(1) A two-stage, knowledge-augmented generation workflow that is divided into conceptual design and massing composition generation stages, enabling consistent intermediate representations for comparison and iteration.

(2) A design knowledge representation that organizes prototypes and a lightweight KG into explicit concept, strategies, and geometric operations for compiling project briefs into stage-specific prompts.

(3) A Rubric-based, fixed-budget evaluation and selection protocol that supports fair ablation comparisons and produces traceable evidence for why certain candidates are preferred.

(4) Ablation studies on several architectural design projects verify the limits of knowledge routing and write-back in early form exploration.

The rest of this paper is organized as follows: Section 2 reviews the relevant research on architectural concepts and generative artificial intelligence in massing composition, and summarizes the shortcomings of the existing process in terms of cross-stage connection and evaluation; Section 3 introduces the methodological workflow, knowledge representation and two-stage generation

 doi:10.20944/preprints202602.1657.v1

process; Section 4 introduces experimental devices, ablation comparisons and result analysis; Section 5 discusses the main discoveries, limitations and future work; Section 6 summarizes the whole paper.

2. Related Work

This section reviews the research progress of generative artificial intelligence in early architectural design from four aspects: conceptual expression, massing composition, knowledge augmentation and evaluation protocol. Each section first summarizes the representative technical paths, and then points out their shortcomings in cross-stage connection, controllability and evaluability, thus leading to the research motivation of this paper from concept to massing alignment and phased evaluation of iteration.

2.1. Application of Generative AI in the Architectural Conceptual Stage

Generative artificial intelligence is mainly used for rapid divergence and visual expression in the conceptual stage. Typical paths include: diffusion generation based on text or reference images to form diversified conceptual images [3,5,8–15]; conditional generation from sketches to images to preserve contours and supplement expression [16–19]; and adding constraints to intermediate representations such as layout or structure to enhance controllability [20,21]. These works have significantly improved the efficiency of drawing output, but most of the results remain at the image-based inspiration level. The problem is that it is difficult to interpret and reuse conceptual narratives, spatial strategies and massing. In addition, the lack of a unified candidate selection process leads to insufficient comparability and reproducibility of output results under project constraints.

2.2. Application of Generative AI in Architectural Form-Finding

In the form generation stage, relevant research includes the generation of explicit 3D geometric representations (e.g., point clouds, voxels, meshes, and implicit fields) [7,22], as well as image-based conceptual massing expressions that emphasize early-stage readability and comparability of silhouettes, solid-void structure, and key connections [23]. Technically, one line of work generates or classifies forms driven by parameters or text [24–28], while another infers forms from sketches, plans, and images while maintaining structural consistency [29–32]. Other studies adopt 3D–2D representation compromises to reduce learning difficulty, and integrate environmental and performance evaluation into the selection loop [33–37]. However, two limitations are common: (i) the interface between generated form and design intent is often implicit, yielding forms that are obtainable but difficult to interpret and edit; and (ii) the generation–selection loop is rarely compared with clear attribution under a fixed budget. Accordingly, this paper constrains the output to concept-stage axonometric massing proxy images, and supports fair comparison with a reproducible evaluation protocol.

2.3. Knowledge-Augmented Generative Architectural Design

To enhance controllability and interpretability, knowledge-augmented methods integrate structured knowledge (e.g., knowledge graphs) and semantic retrieval by vector libraries into the generation process: the former explicitly articulates design entities, strategies, and constraints, while the latter performs similarity recall across vast datasets, providing traceable context for foundation models [38–41]. Additionally, multi-agent frameworks further organize task chains through "role division" and are frequently coupled with BIM or parametric tools for implementation [42–45]. Overall, existing work predominantly focuses on knowledge injection or collaborative mechanisms within specific steps, while the formation of stable, executable interfaces across the "conceptual narrative-strategy selection-geometric organization" stages remain underdeveloped. This paper constructs a knowledge graph and prototype library to provide knowledge routing, compiling reusable conceptual/strategic/geometric operation cues into stage-specific prompts. Combined with explicit evaluation and write-back, this enhances the comparability, traceability, and reproducibility of the generation process.

2.4. Controllable Generation and Evaluation Protocol

In the conceptual generation stage, the trade-off between "diversity" and "constraint fit" is the core challenge. Methods such as diffusion model and LoRA provide strong style and condition control [1,7,10]. However, in architectural tasks, the quality of the candidate plan usually depends on the details of the prompt and the randomness of the sample, which can easily lead to form drift or local details dominate the overall organization. In order to solve this problem, some studies have introduced statistical distribution alignment, artificial preference surveys or task-specific metrics to select candidates [9,11,12]. However, there is still a lack of a unified early batch expression evaluation protocol, which makes it difficult to evaluate different settings fairly. This paper employs Rubric-based Top-K ranking to select candidates under a fixed budget, feeding evaluation results back into prompt templates. This approach enhances constraint adherence and expressive stability while preserving exploratory diversity.

In summary, existing research can generate abundant candidates at the conceptual and morphological levels, but there is still a lack of intermediate interfaces needed to stably transform narrative intentions into operable massing compositions, as well as comparison and attribution mechanisms that can be reproduced under a unified budget. At the same time, early iterations often rely on prompts for trial and error, and lacks traceable feedback updates. Therefore, this paper proposes a two-stage knowledge-augmented generation workflow, and establishes an early convergence path through Rubric-based Top-K and feedback mechanisms.

3. Method

3.1. Overview

This paper proposes a two-stage generative framework for early architectural design, as illustrated in Figure 1. It comprises a knowledge base layer, a knowledge routing layer, and a two-stage generation layer. The knowledge base comprises general design principles, a case library, a knowledge graph, and a prototype library. This supports the stable translation of early architectural concepts from narrative imagery to massing compositions. The case library, composed of existing materials from typical architectural concept design phases, trains sketch and massing LoRA models and extracts knowledge. The prototype library provides generative large models with associative imagery and analogical translation. The knowledge graph derives partially from existing design rule sets and partially from structured representations of case library knowledge. The Knowledge Routing layer searches the prototype library for entries matching design requirements, retrieves achievable strategy combinations and geometric structures from the KG, and outputs compilable stage prompts for use by the generation layer.

The two-stage generation layer comprises the Conceptual Design Stage (Stage 1) and the Massing Composition Generation Stage (Stage 2). Stage 1 consists of two steps: Concept Generation and Sketch Generation. Stage 2 comprises two steps: Massing Description Generation and Massing Proxy Image Generation. These steps form a continuous generative sequence. Each step incorporates corresponding evaluation modules. Evaluation results from the sketch and massing proxy image generation steps are fed back into the prompt compilation process, thereby establishing an iterative closed-loop optimization mechanism.

The knowledge augmentation process in this paper proceeds as follows: The system parses project requirements into objectives and constraints, then uses knowledge routing to instantiate design knowledge as drivers and operations. Finally, it compiles these into a two-stage prompt generation to drive subsequent generation and filtering.

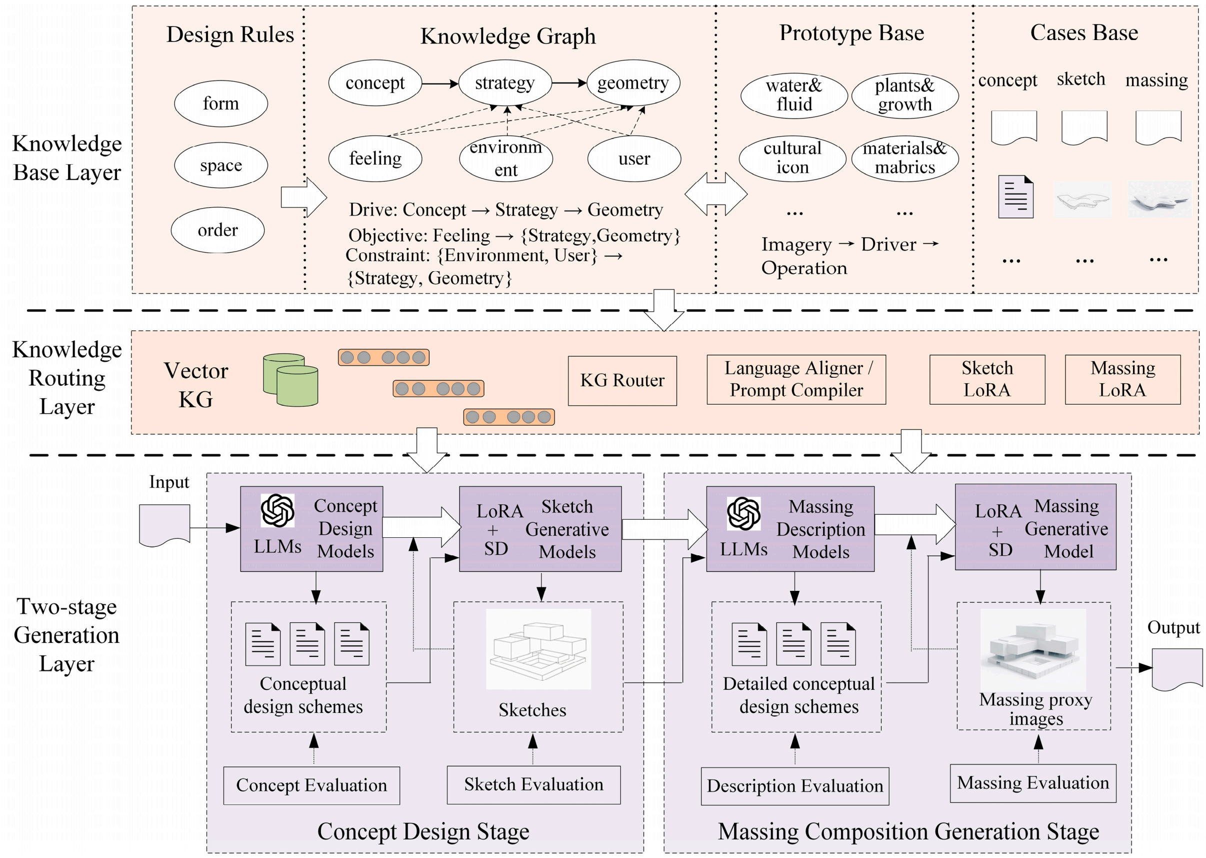


Figure 1. A knowledge-augmented two-stage workflow for architectural concept-to-massing generation and evaluation.

3.2. Data Collection and Knowledge Representation

To support interpretable generation and reproducible experimentation from narrative imagery to massing composition, this paper constructs three interrelated data sets and representations: a Case Base, a Prototype Library, and a Knowledge Graph. This section details the construction process and usage methods for these three repositories.

3.2.1. Case Base

The Case Base is utilized for LoRA style and structure tuning, alongside offline relation extraction; no case text or image snippets are retrieved at runtime. The Case Base comprises a collection of case sketches and massing proxy images, alongside associated metadata. We curated 150 projects, each with one sketch and one massing proxy image (150 each).

Case data originates from open-license media libraries and search engines (e.g., Wikimedia Commons and Openverse), with cases selected based on open licensing. Uniform screening principles ensure conceptual-stage applicability: (1) Primarily public buildings including museums, libraries, cultural centers, and complexes to guarantee representative spatial organization and public interfaces; (2) Possess legible volumetric relationships, such as clear enclosures, through-spaces, courtyards, and plinths; (3) Feature complete textual and visual information, including at least a general conceptual description and identifiable form characteristics, facilitating alignment from description to sketch and massing proxy images; (4) Avoid homogenization by deduplicating repeated projects under the same theme to enhance coverage.

Regarding massing proxy image dataset acquisition, we simplified images by removing architectural environment and material details, retaining only the general form of the building model. A conditional diffusion image-to-image translation workflow was employed. This workflow was modelled as a translation problem from rendering to massing proxy images, where the input is a perspective or axonometric rendering, and the output is a geometry-centered, textureless massing proxy image that preserves primary volumes, voids, and traffic-related openings.

In field design, the case studies aim for cross-stage alignment. Each case record comprises four fields: (1) Concept: key ideas and imagery expressed as concise summaries; (2) Description: form and spatial organization descriptions for retrieval and prompt compilation; (3) Sketch: conceptual sketches or structural representations for training the LoRA model during the sketch generation stage; (4) Massing: volumetric and massing representations for training the LoRA model during the massing composition generation stage.

3.2.2. Prototype Library

The prototype library design draws upon the common process in architectural conceptualization where imagery is translated through analogy. Architects often employ metaphors or analogies to map experiential imagery, such as "flowing water", into rule sets for organizing space and form, transforming implicit, difficult-to-reuse conceptual processes into traceable knowledge units [46]. Structured retrieval of prototype knowledge supports design reuse and adaptation. Spatial forms can be abstracted into executable rule systems, enabling semantic-to-geometric translation [47,48].

This paper organizes prototype imagery into a Prototype Base, with the objective of making the "Imagery → Driver →Operation" translation process explicit. This enables retrieval, combination, and subsequent generation as input to executable interfaces. Twelve common prototype imagery categories are employed, aiming to cover the primary semantic sources at the conceptual stage. These categories include: (1) water and fluids; (2) topography and geology; (3) plants and growth; (4) animal morphology and skeletons; (5) celestial phenomena and meteorology; (6) materials and textiles; (7) tools and machinery; (8) transportation and motion; (9) settlement and urban prototypes; (10) mathematics and abstract geometry; (11) cultural symbols and artefacts;(12) religion and ritual. Among these, categories 1-5 draw upon natural imagery from the Natural Heuristic Knowledge Base [49]; categories 6-8 originate from natural systems as sources of form and mechanism generation [50]; categories 9-10 are inspired by high-frequency spatial patterns and abstract geometry [51]; categories 11-12 correspond to symbolic and narrative sources in conceptual design [52].

To translate abstract archetypal vocabulary into legible and actionable geometric forms, this paper employs a three-tier schema to compress each archetype into a structural chain: Imagery → Drivers → Operations, as illustrated in Table 1. Specifically: (1) The Imagery layer describes the archetype in a single sentence, e.g., "Rivers form meandering pathways," serving retrieval and semantic recall; (2) The driver layer decomposes the imagery into comparable design drivers, such as "flow"; "continuity"; "loop_circulation", etc., for alignment with design objectives; (3) The operation layer maps drivers to a set of executable geometric operation verbs, such as "split_masses", "carve_void", "smooth_blend", etc., for injecting prompt tokens into the generation model during the generation stage.

Table 1. An example of prototype library entry schema.

Field	Type	Example	Role in pipeline
Prototype_id	ID	P-NAT-FLOW-001	Unique key
Title	Text	River Meander	Human-readable label
Category	ID	Nature	Filtering & coverage analysis
Imagery	Text	"Rivers form meandering paths..."	Runtime keyword extraction and retrieval
Drivers	List	"flow"; "continuity"; "loop_circulation"	Aligns concept intent to operations
Operations	List	"split_masses"; "carve_void"; "smooth_blend"	Executable geometry grammar
Prompt_tokens	List	"ribbon following a spine..."	Prompt compiler injection

 doi:10.20944/preprints202602.1657.v1

3.2.3. Knowledge Graph

To explicitly digitize the imagery associations commonly employed by architects during the conceptual phase, enabling retrieval, combination, and geometric translation, this paper constructs a knowledge graph employing a six-dimensional representation: Concept, Geometry, Strategy, User, Environment, Feeling. The graph's knowledge derives from a case library and a design rule set, which structures universal design knowledge including formal, spatial, and aesthetic principles [53]. These six dimensions serve as an accessible interface for early conceptual generation. The sequence of "Concept → Strategy → Geometry" operates as a design method, bridging the gap between initial ideas and their final geometric forms [54]. In addition to this narrative-driven process, this study incorporates two key aspects to ground the design in reality: the User and Environment dimensions. These factors encode specific requirements, usage scenarios, and site conditions, forming a set of constraints that prevent the generated concepts from losing touch with their intended context [55]. Finally, the Feeling dimension is introduced to capture desired spatial qualities and emotional impacts, which then help guide the later stages of screening and evaluation [56].

There are three types of relationships in the KG as follows: driving relationships "Concept → Strategy → Geometry" for deriving geometric operations; target relationships "Feeling → {Strategy, Geometry}" for generating optimized solutions; constraint relationships "Environment, User → {Strategy, Geometry}" for forming the constraint framework. Edge weights express relationship strength, as illustrated in Figure 1. To address multiple types of knowledge conflicts, knowledge routing prioritizes satisfying hard constraints (Environment or User), then optimizes Feeling, and finally selects Geometry operations.

3.2.4. Vector Knowledge Base and Retrieval

To support prototype-driven generation, this paper introduces a Vector Knowledge Base for semantic recall. When given project requirements, the system retrieves relevant prototype and executable operation prompts, which then feed into the knowledge graph router and prompt compiler. Each prototype entry includes textual fields, such as prototype, drivers, operations, and prompt_tokens, which capture the relationship between visual semantics and generation steps. On the other hand, stage summaries are structured abstractions of intermediate outputs (e.g., from Stage 1, Stage 2, and Top-K selections), which support subsequent stability analysis. All records are stored as structured data to ensure deterministic filtering and reproducible comparisons.

During runtime, the system first converts project requirements into semantic queries, combining intent keywords and experiential phrases with functional filters (e.g., scale, functionality, or design stage). These queries search the prototype library for semantically similar Top-K candidates, with diversity constraints applied to avoid redundant results. Retrieved prototypes then enter the knowledge routing module, which expands and scores feasible generation paths along the "Concept → Strategy → Geometry". The output is a structured representation, which the prompt compiler uses to generate phased prompts for sketch and massing composition. This routing layer is supported by two LoRA models trained on the project case library and dedicated to sketch and massing composition generation tasks respectively.

3.3. Conceptual Design Stage

The conceptual design stage comprises two steps including concept generation and sketch generation. In concept generation step, project requirements are fed into an LLM augmented with knowledge to produce executable concepts. Concepts are constrained into actionable structured elements, translating abstract narratives into geometric operations for subsequent generation. The sketch generation step employs sketch LoRA-finetuned Stable Diffusion to produce conceptual sketches. These sketches serve as structural conditions for subsequent generation, ensuring stable geometric foundations and reducing randomness in massing composition development.

 doi:10.20944/preprints202602.1657.v1

During the conceptual design stage, we employ the sketch LoRA-finetuned Stable Diffusion generator to enhance output geometric stability and readability. On the input side, prompts are generated by Prompt Compiler, which are from project requirements and operation sets. Each prompt consists of two parts: (1) the semantic section, which includes concepts, imagery, strategies, and geometry; (2) the directive section, which lists executable action verbs and structural descriptions, along with explicit constraints (e.g., "no context," "no people," "no façade details") to preserve the sketch abstraction. On the control side, the Sketch LoRA ensures a consistent linework style and abstract expression. The final output is a set of candidate sketches, from which a Top-K selection is made during the evaluation phase (detailed in Section 3.5) to proceed to the massing description generation stage.

3.4. Massing Composition Generation Stage

The massing composition generation stage comprises two steps: massing description generation and massing proxy image generation. The massing description generation step takes the selected sketches from the conceptual design stage and their corresponding drivers and operations as input. It employs an LLM to expand the set of operations into a more explicit formal narrative, outputting a structured textual description that can be stably interpreted by the diffusion model. The objective of the massing composition generation stage is to further translate the structural sketches and operation sets from the preceding stage into generative form narratives, producing massing proxy images for early-stage discussions.

The massing description generation step structures outputs from the conceptual design stage: under operational constraints, it expands abstract intentions into more concrete geometric narratives, including: (1) mass hierarchy, such as primary masses, secondary masses, and node volumes; (2) dominant curves and skeletal types, such as ribbon, ring, or double-body fissure; (3) connection methods, such as bridging, interlocking, stacking, or elevated passage; (4) cavity types, such as courtyards, crevices, through-holes, or carved grooves.

In the massing composition generation stage, sketches are input into ControlNet, employing massing LoRA-finetuned Stable Diffusion to translate form descriptions into massing proxy images. Output comprises single-view massing proxy images, required to clearly express overall form, void-solid organization, and primary structural relationships. Cross-proposal comparability is maintained through uniform rendering and background constraints, including pure white backgrounds, absence of environmental elements, subtle shading, and no material textures.

3.5. Evaluation Protocol

To counter generative randomness while preserving feasibility and interpretability, this paper introduces evaluation and iteration loops at every stage. Once candidates are scored and the Top-K are selected, the outcomes are fed back into the Prompt Compiler, which adjusts the mix of drivers and operations for subsequent steps. This feedback, by amplifying or muting specific prompt elements, establishes traceable grounds for later filtering. Ranking is conducted under a fixed sampling budget: N=8 candidates are generated per stage, with the Top-K (e.g., K=3) proceeding based on comprehensive scores. These rankings are computed via step-wise metric evaluations, and iterative modifications to each stage's output are enabled by generating new prompt templates, ensuring refinement stays within budget.

This research divides the generation process into four verifiable gateways, aligning with industry-standard phased outputs and risk boundaries, consistent with classical architectural design practice [57].

3.5.1. Metric Determination

Architectural theory widely recognizes the value of the sketching stage in externalizing ideas for review, modification, and dialogue. The massing composition stage requires testing the translatability and integrity from concept to massing, encompassing feasibility, compliance, and construction logic [58–60].

Based on the above literature, we defined evaluation metrics for four steps across two stages, with four metrics established for each step. Each metric was defined and its scoring criteria quantified. For instance, the first metric in Stage 1 is Concept Clarity, defined as "a clear description of the dominant framework, void-solid relationships, and experiential sequence". Its scoring criteria are as follows: 1 point-"narrative is loose and fails to point towards form"; 3 points-"can describe the main framework and key voids but lacks specificity"; 5 points-"Clear primary structure/void-solid relationships/experience sequence directly translatable into operational sequences". Scores are recorded on a 1-5 Likert scale with 0.1 increments. The definitions and quantitative standards for each stage's metrics are detailed in Appendix A.

3.5.2. Weight Determination

To balance the subjectivity of design judgement with the objectivity of statistical outcomes, this paper employs a hybrid subjective-objective weighting approach to determine the weights of evaluation metrics at each stage [61]. Specifically, we derive subjective weight w_m^{AHP} through the Analytic Hierarchy Process (AHP) to reflect priorities in architectural design reviews for dimensions such as legibility, alignment with intent, convertibility, and implementability. Concurrently, we calculate objective weight w_m^{Ent} to reflect the discriminative power of different metrics within the candidate scheme set. The final metric weights are presented in Equation 1.

$$w_m^{Hybrid} = \alpha \cdot w_m^{AHP} + (1-\alpha) \cdot w_m^{Ent} \quad (1)$$

where α emphasizes the dominant role of design rationality.

To determine an appropriate value for α, we conducted sensitivity analysis of α by varying α in [0.4, 0.8] for 35 projects. Table 2 shows that α = 0.6 yields the highest Top-3 overlap (i.e., the most stable candidate ranking), providing a practical trade-off between stability and discriminability. Therefore, α is set to 0.6.

Table 2. Sensitivity of Top-3 ranking stability to $\alpha \in [0.4, 0.8]$, measured by the mean Jaccard overlap between Top-3 sets at α and at α=0.6 across 35 projects.

α	0.4	0.5	0.6	0.7	0.8
Top-3 overlap	0.51	0.71	1.0	0.73	0.56

In the two-stage, four-step evaluation framework, we further introduced stage weight W_s (Stage weight) and within-stage metric weight $w_{s,m}^{Hybrid}$ (within-stage weight), yielding the final total weight for ranking as shown in Formula 2.

$$W_{s,m} = W_s \cdot w_{s,m}^{Hybrid} \quad (2)$$

where $\sum Ws = 1$ and $\sum w_{s,m}^{Hybrid} = 1$. The final composite score $Score(i)$ for each candidate is calculated using a weighted average, as shown in Formula 3.

$$Score(i) = \sum_s \sum_m W_{s,m} \cdot r_{i,s,m} \quad (3)$$

where $r_{i,s,m}$ represents the scoring value for each stage, ranging from 1 to 5. The screening strategy employed in this paper involves retaining the top K candidates through sorting at each critical step, with a single write-back performed in the current experimental setup to mitigate iterative instability. This weighting system enables reproducible ranking for experiments and permits iterative calibration in subsequent work. Using this methodology, the calculated stage weights are: W_1= 0.20, W_2= 0.25, W_3= 0.25, W_4= 0.30. The metric weights for each stage are detailed in Appendix B.

4. Results

This section evaluates and analyses the generated results based on the assessment protocol outlined in Section 3.5.

4.1. Experimental Settings

The experiments were conducted on a desktop computer running Windows 11, equipped with an Intel Core i7 processor, 32 GB of RAM, and an NVIDIA GeForce RTX 4090 graphics card (24 GB VRAM). Workflow orchestration (requirement parsing, KG routing, prompt compilation, rubric scoring, Top-K ranking, and write-back) was implemented in Python 3.10. The diffusion generation module (SD/LoRA, supporting optional Control conditions) was implemented using PyTorch 2.10 and the Diffusers framework.

The two-stage image generation model employed in this paper is Stable Diffusion with LoRA, with the underlying diffusion model being Stable Diffusion XL (SDXL). The sampling method employed is Euler a, with 20 sampling steps. Generated images measure 1024 × 1024 pixels, with a generation budget of 8 candidate images. The prompt guidance coefficient (CFG scale) is set to 7, and the random seed list is {1, 4, 6}. We trained two LoRA models to support the Stable Diffusion generator. Device information and key parameters used in the training LoRA process are shown in Table 3. The large language model (LLM) for generating concepts is implemented via the OpenAI API.

Table 3. Device information and key parameters used in the training LoRA process.

Parameter	Value
Device	NVIDIA GeForce RTX 4090 (Ada Lovelace)
VRAM	24,576 MB
VRAM used	~16,736 MB
Memory efficient attention	On
Use xformers	On
Learning rate	1e-4
Text encoder learning rate	0.0001
Unet learning rate	0.0001
Optimizer	AdamW8bit
Source model	SDXL Base
Training steps	13050
Training time	4.5h

To validate the proposed method's efficacy, this paper compares different ablation settings, denoted as A0, A1, and A2, as shown in Table 4. To ensure the attributability of ablation comparisons, A0, A1, and A2 employ uniform generation budgets and configurations: identical base generation models, resolutions, sampling parameters, and random seed lists are used at each stage, generating Top-K results under the same candidate budget. Under this unified protocol: A0 disables knowledge graph routing and write-back, generating stage prompts and outputs volumes directly based solely on project requirements; A1 enables knowledge graph routing to compile stage prompts but omits write-back; A2 incorporates write-back on top of A1, updating prompt fragments with evaluation results. Consequently, subsequent outcome differences can be primarily attributed to knowledge routing and write-back mechanisms, rather than variations in sampling budgets or parameter configurations.

Generated solutions were scored using the evaluation metrics proposed in Section 3.5 to compare the effectiveness of different methods. Five experienced architectural designers were invited to participate in our evaluation process, with each score calculated as the average after removing the highest and lowest scores, following the approach in [11,12].

 doi:10.20944/preprints202602.1657.v1

Table 4. Ablation settings.

ID	Setting Name	Prototype + KG	Iteration
A0	Baseline (text-only)	No	No
A1	Ours (Prototype + KG)	Yes	No
A2	Ours (Prototype + KG + Iteration)	Yes	Yes

The evaluation method is as follows:

(1) Top-K Quality Distribution (Top-K scores). Generated candidates are scored on a 0-5 scale per Appendix A and weighted totals calculated using Appendix B's weighting table. Average score, variance, and improvement margin are reported to assess the contribution of knowledge routing and write-back mechanisms to quality convergence.

(2) Randomness and Stability. Under identical prompt and control conditions, repeat generation three times with different random seeds. Report: (i) Top-1 score variance; (ii) consistency of evaluator scores and significance tests.

4.2. Qualitative Comparisons

This subsection presents multiple practical generation cases, displaying results from the four generation steps across two stages using different methods to validate the proposed generation process and methodology. Due to space constraints, certain details have been omitted in case presentations, such as condensed descriptions of generation schemes, design requirements, and prompt content.

4.2.1. Case 1 Presentation

Design Requirement Input: *"Design a museum with a floor area of 7,000 square metres on a 20,000-square-metre site for glass art exhibitions. The single-storey building should feature an internal courtyard enabling interaction between the interior experience and surrounding green environment, allowing visitors to move freely and fostering a sense of exploratory strolling."*

Step 1: From Requirements to Concept Generation

Generation Requirements: Utilize LLM to generate 8 concepts. A0 and A1 output results by Name, Imagery, Strategy, and Geometric Structure; A1 additionally includes Drivers and Operations.

Table 5 presents the Top-3 conceptual solutions, with the remaining solutions for A1 detailed in Appendix A. Solutions are listed in descending order of weighted average score (WAS), alongside evaluation sub-scores r_{11}, r_{12}, r_{13}, and r_{14}. The total score is computed from the metrics in Section 3.5. The evaluation reveals that A1 consistently outperforms A0 in total scores, demonstrating greater alignment with architectural design requirements. Notably, A1 achieves higher scores for Constraint Coverage (r_{12}) and Operational Specificity (r_{13}), attributable to the structured constraints provided by the prototype library and KG spanning from conceptualization to operationalization. In contrast, the A0 scheme exhibits greater divergence. In Group 1, the A0 scheme's conceptual description lacked clarity regarding openings and its relationship with the surrounding environment, resulting in a lower r_{13} score for operational specificity. Conversely, the A1 scheme's conceptual description provided detailed and reasonable explanations of geometric relationships and organizational forms. In Group 2, the absence of descriptions regarding low-rise characteristics in the A0 scheme led to the subsequent generation of multi-storey buildings. All three groups of the A1 scheme responded to the single-storey feature specified in the design requirements.

Table 5. Concept Generation Schemes 1-3 for Case Study 1.

ID	A0	Score	A1	Score
1	**Title**: Transparent Corridor Loop + Green Courtyard **Imagery:** A semi-transparent corridor forms a soft connection between the exhibition hall and	r_{11}: 4.3 r_{12}: 4.2 r_{13}: 3.6 r_{14}: 4.0	**Title**: Glass Stream, Meandering Through the Forest.	r_{11}: 4.6 r_{12}: 4.5 r_{13}: 4.3 r_{14}: 4.3

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 26 February 2026 doi:10.20944/preprints202602.1657.v1

	the woodland, creating a naturally looping visitor route that prevents disorientation. Transparency serves not as a technical flourish, but as a means to continuously draw greenery and human flow into the building. **Strategy and Geometry**: A single 'low ring-shaped structure' encloses a large courtyard; along the outer perimeter of the ring, pocket courtyards with localized 'concave-convex' forms are integrated. Visitors meander along the circular colonnade, alternately drawing near the courtyard and the woodland, establishing a continuous rhythm of 'viewing-walking-pausing'.	WAS: 4.05	**Imagery:** A 'glass stream' meanders through the woodland, converging at its heart to form a tranquil, transparent courtyard space. **Strategy and Geometry**: A looping main ridge organizes the layout, with ribbon-like single-storey volumes segmented to enclose the central courtyard. Multiple openings and lateral incisions along the periphery create traversable thresholds, while locally elevated sections form daylighting nodes. This allows circulation paths to alternate between "courtyard proximity" and "forest proximity". **Drivers**: flow | continuity_loop||porous_edges **Operations**: streamline_spine | sweep_loft_ribbon | carve_gaps_thresholds	WAS: 4.45
2	**Title**: Disrupted Enclosure, Four Segmented Curved Volumes + Multiple Entrance Slits. **Imagery**: A strong central core (green courtyard) anchors directional orientation, while pocket courtyards function as 'branching nodes', transforming free exploration into a controlled network of strolls. **Strategy and Geometry**: Four curved (or zigzag) volumes enclose courtyards while deliberately leaving 6–8 "slit entrances" to form traversable loops. Enclosure derives from volumetric continuity; freedom of movement stems from the permeable nature of the "slits," akin to forest paths.	r_{11}: 4.5 r_{12}: 3.4 r_{13}: 3.8 r_{14}: 3.6 WAS: 3.86	**Title**:River Bend Courtyard, traversable yet not bewildering. **Imagery**: River bends form accessible recessed pockets where water swirls within boundaries and flows naturally. **Strategy and Geometry**: Centered on a "shallow cove" courtyard as the primary directional anchor, single-storey volumes encircle the bay line while reserving two to three traversable openings. Gradually expanding and contracting corridor widths, heights, and framed vistas establish a clear "enter-circulate-re-emerge" pathway. **Drivers:** central void|directional navigation|enclosure **Operations:** carve_shallow_courts|carve_canyon_void|frame_openings_to_views	r_{11}: 4.6 r_{12}: 4.1 r_{13}: 4.3 r_{14}: 3.8 WAS: 4.26
3	**Title**: Garden in a Box, Restrained Exterior + Highly Transparent Interior **Imagery:** Rather than forming a complete, closed 'circle', the enclosure is segmented into legible arcs (or broken lines); multiple 'slit entrances' facilitate permeation and traversal akin to forest paths. **Strategy and Geometry**: An outer ring of more substantial, low-profile shell (protecting exhibits, controlling light) encloses a fully transparent courtyard and colonnade. Visitors transition from the woodland into the "serene outer shell," then abruptly enter the "core of light and greenery"—transforming contradiction into a sequential narrative.	r_{11}: 4.6 r_{12}: 3.6 r_{13}: 3.4 r_{14}: 3.5 WAS: 3.84	**Title**: "Floating Roof Beneath the Canopy" Enveloping a Circular Arcade **Imagery**: The canopy resembles a delicate, translucent canopy, beneath which a shaded colonnade forms a space for strolling. **Strategy and Geometry**: A continuous, thin roof covers the circular colonnade and encloses the courtyard. The core thickens slightly to form structural and daylighting nodes, while the edges thin out and extend towards the woodland. The colonnade links exhibition spaces and resting points, with light filtering evenly through linear skylights and softened openings, creating a light, continuous sequence for wandering. **Drivers**: promenade_sequence | diffuse_light | lightweight **Operations**: loft_canopy_continuity | thicken_at_core_thin_at_edges | perforate_linear_skylight	r_{11}: 4.3 r_{12}: 4.0 r_{13}: 4.2 r_{14}: 4.3 WAS: 4.19

Step 2: Massing Concept Description to Sketch

Building upon Step 1, select the generated conceptual description and utilize the sketch LoRA-finetuned Stable Diffusion model for sketch generation. During generation, merge the input and output of Step 1 to form the prompt. Only the newly added prompt content is provided below. The specific additions to the A1 prompt template are as follows:

Added prompt template content: *ARCH_SKETCH_STYLE, axonometric architectural line drawing, conceptual massing sketch, A single-storey museum with a central courtyard. Segmented curved ribbon volumes enclosing an inward-facing courtyard, gaps serving as side entrances, continuous loop circulation, porous edges, clean black ink contour lines, uniform thin strokes, pure white background, no shading, no shadows, no gradients, minimal details, no texture.*

Figure 2 compares three schemes without knowledge enhancement (A0 (a)-(c)) and three with it (A1 (e)-(g)), all of which have been assessed and presented here. Figure 3 demonstrates that schemes of A1 generally achieve higher scores overall, particularly excelling in Legibility and Design Intent Match metrics. This advantage stems from the prototype library and case studies providing rational spatial syntax and structural organization principles. Scheme 1 was selected for iteration. As the generated scheme scored low on Legibility, this metric was enhanced by modifying the prompt with operations incorporating specific refinements: circular rings, promenades, courtyards, and distinct entrances. The revised prompt included:

ARCH_SKETCH_STYLE, architectural concept sketch, axonometric view, segmented circular ring building massing, thick outer ring split into several arc segments, concentric inner ring promenade, large central oval courtyard void (water court), carve a clear portal opening / entry.

Figure 2(j) illustrates the sketch proposal following iterative modifications, demonstrating a more rationalized approach. Evaluation results in Figure 3 indicate that at this stage, A2 outperforms both A0 and A1.

Step 3: From Sketch to Detailed Massing Description

We selected the highest-scoring scheme from Stage 1 of A0 for detailed elaboration. For the first row of the A0 schemes in Table 5, the results were refined in terms of strategy and geometric content: *"Generate a low, single-storey free-form elliptical ring volume within the site as the building's outer contour. Hollow out the ring's center to form a large-scale green core courtyard, while creating pocket courtyards along the ring's outer edge as pause points; The ring's thickness is divided into two continuous bands: a fully transparent glass colonnade band (alternating in plan between proximity to the courtyard and proximity to the woodland to create rhythm), with the entrance accessed via a cut-out recess into the colonnade, allowing multiple short-cut connections."*

Regarding A2, based on the above sketches, architectural requirements, and Stage 1 concept prompts, the strategy and geometry are elaborated as follows: *Employing a "single-storey ribbon ring with eccentric elliptical courtyard void" as the matrix: Enclose the courtyard with a continuous curved ribbon ring, designed with an outer thicker, inner thinner profile and variable width (more substantial and solid externally, lighter and more transparent internally); extend the roofline to form a ramp to ground level. The façade shall be transparent glass.*

Compared to the initial conceptual description, this refinement provides a more concrete architectural proposal. Leveraging prototype libraries and knowledge graphs, the LLM offers explicit solutions for the positioning, detailing, and solid-void relationships of each architectural component. This provides clearer guidance for generating massing proxy image in subsequent steps. The A2 refinement demonstrates significantly greater rationality than the A0 scenario.

Step 4: Generate Massing Proxy Images

For A1, the description generated in Step 3, combined with the sketch from Step 2, was input into ControlNet. Using massing LoRA-finetuned Stable Diffusion for massing proxy image generation, four proposals were produced. The optimal result is presented in Figure 3(h). The ControlNet mechanism maintains stability in generated forms while limiting randomness, resulting in multiple similar outputs at this stage. The trained massing LoRA model enhances clarity in formal relationships within the proposals. For A2, iterative refinement further improves Feasibility (r_{42}) and Editability (r_{44}). The specific prompt is as follows: *"Semi-translucent internal corridor facades, light-permeable walkways beneath thin floating roofs, intricate cantilevers and layered ring structures, continuous circular walkways, minimalist site contours and platform bases."* This enhances spatial hierarchy, improves structural rationality, and facilitates further refinement of the massing. The iterative result for A2 is superior, as shown in Figure 2(l).

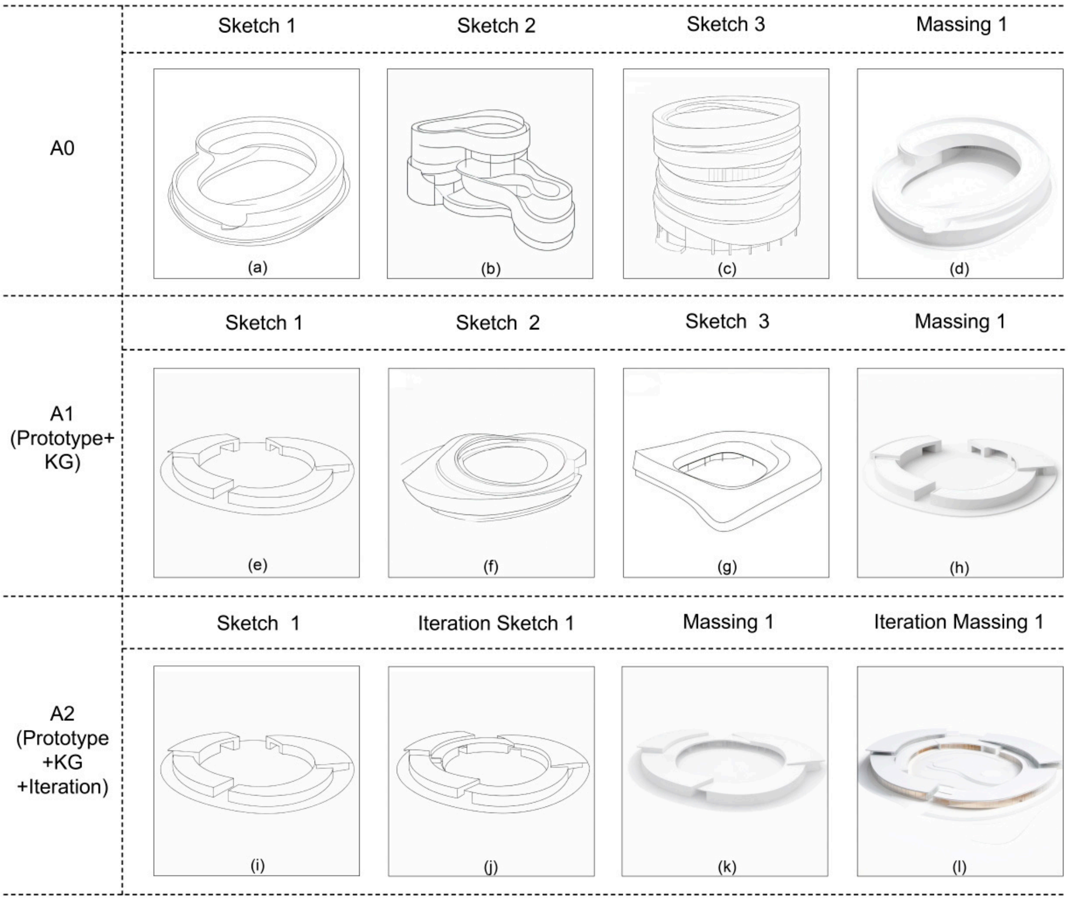


Figure 2. Results of case 1. (a)-(c) are sketches of A0, (d) is the massing of (a); (e)-(g) are sketches of A1, (j)(A2) is an iteration of (i), (k) is the massing of (j), (l)(A2) is an iteration of (k), the final massing.

The Top-1 massing proxy images A0(d), A1(h) and A2(l) were evaluated under the three settings A0, A1 and A2 in Figure 2, with results shown in Figure 3. It can be observed that, compared to A0, the improvement in A1 during the massing composition stage is primarily reflected in metrics related to implementability. Feasibility (r_{42}), Compliance (r_{43}), and Editability (r_{44}) all show significant enhancements, while the increase in Aesthetics (r_{41}) is relatively modest. Knowledge augmentation provides constraints for alignment and operational pathways, thereby enhancing the fundamental feasibility and consistency of massing proxy images. Compared to A1, A2 demonstrates notable improvements in Feasibility (r_{42}) and Editability (r_{44}), attributable to its iterative feedback mechanism which elevates the maturity and overall quality of the massing.

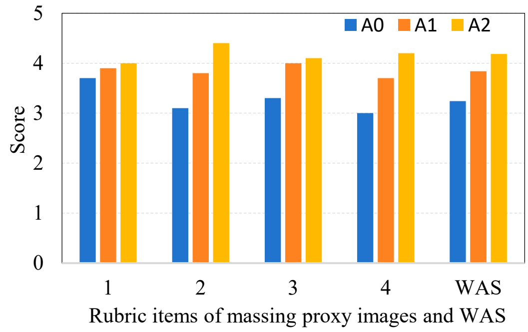


Figure 3. Rubric scores for the massing proxy images of Case 1 in A0, A1 and A2.

To evaluate the system's capacity for generating multiple high-quality solutions addressing identical design requirements, the eight conceptual scheme outputs for Case 1 are presented in Figure 4, numbered 1-8. These outputs demonstrate that under identical design briefs and constraints, the system can produce architecturally coherent yet morphologically distinct solutions, exhibiting strong formal diversity and compositional variation.

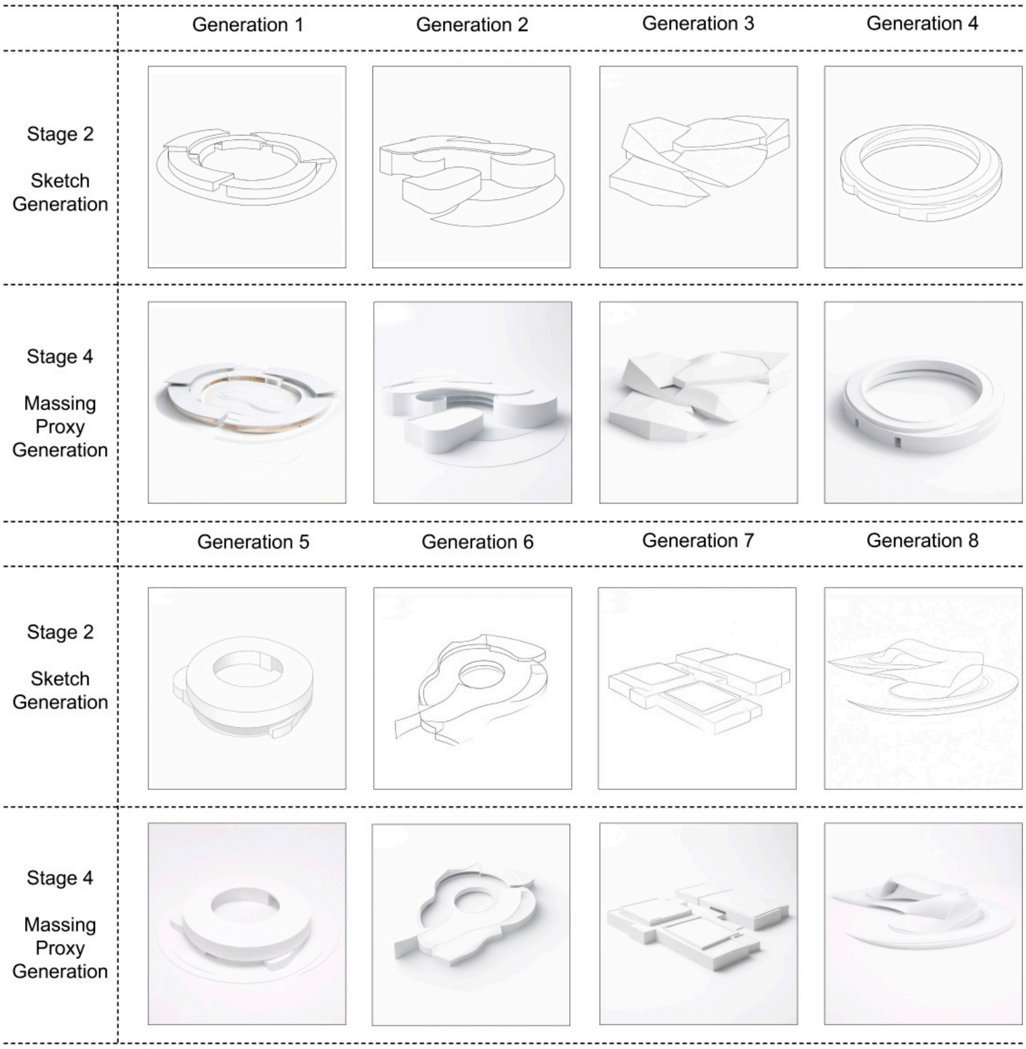


Figure 4. Generated sketches and massing proxy images of eight conceptual schemes for Case 1.

From a geometric perspective, the eight schemes exhibit distinct differences in overall silhouette, volumetric composition, and treatment of void-solid relationships. Schemes 1, 4, 5, and 6 adopt continuous ring or enclosed organizational forms, emphasizing inward-facing courtyards and meandering circulation paths; Schemes 2 and 3 create flexible spatial layouts through offset volumes or dual-body confrontations; Schemes 7 and 8 express fluidity and directionality through undulating roofs or layered volumes. These variations represent not mere morphological transformations, but distinct spatial prototype translations generated within the driving and operational framework.

From a spatial structural perspective, the schemes exhibit distinctiveness in massing organization, void typologies, volumetric connection logic, and boundary treatment. Notably, despite pronounced formal variations, all schemes maintain responsiveness to core project intentions such as circulation potential, the coexistence of enclosure and permeability, and internal courtyard interaction, while adhering to project constraints including building typology, area, storey count, and fundamental requirements. This demonstrates that the system can expand the formal search space while ensuring conceptual consistency, thereby avoiding convergence towards a single solution.

Therefore, Figure 4 not only demonstrates the diversity of results but also reflects the balance achieved in concept generation under knowledge augmentation between “structural stability” and “morphological diversity“.

4.2.2. Evaluation of Generation Outcomes Across Different Project Types

To evaluate the generalization capability of the proposed method, this section presents the performance of models generated for eight project requirements tests across different project types. Projects correspond to eight categories including museum, library, mixed-use complex, etc., with parameter settings for model generation as described in Section 4.1. Due to space constraints, Table 6 displays only the typologies and imagery from the conceptual generation stage. Figure 5 illustrates the generated sketches and massing proxy images.

Table 6. Project Categories and Imagery for Eight Projects.

ID	Category	Imagery
1	Museum	A 'glass stream' winds through the woodland, converging at its heart to form a tranquil, transparent courtyard space.
2	Library	Like a monastic cloister, the outer ring is substantial while the inner circle is serene; the central atrium is a protected light well. People circulate within the cloister.
3	Mixed-use complex	Urban rock face and oasis rift: 'carving a canyon' within the thick plinth.
4	Waterfront Museum	"White terraces folded along the waterfront + vortex courtyard". The outer Artterraces resemble waterfront platforms, their edges continuously eroded into layered tiers. The interior is hollowed out by a "backwater/vortex" to form a tranquil, inward-facing public valley.
5	Community Centre	A four-directional permeable "community circulation loop" features a ring of low arcades enclosing a central plaza, with four openings serving as the community's "gates".
6	Cultural Complex	A repeatedly folded urban tapestry: several homogeneous volumes, like "panels" on a woven rug, are cut and connected by a series of slender streets and alleys; the junctions of these streets and alleys link entrances, activities and resting points like beads on a string, forming a navigable public network.
7	Opera	A membrane of water surface, 'lifted and lowered' by sea breezes and tides, guides visitors from lower levels to wander beneath the shell, then ascends them towards light and vistas.
8	Art Gallery	Viewing proceeds along a 'band of light': linear skylights uniformly illuminate the gallery, with shallow light wells at key junctures. The flow of visitors is naturally organized, paused, and redirected by the light.

Generation results in Figure 5 show distinct geometric configurations across proposals, including ring volumes, dual-volume confrontations, fluid continuums, and layered compositions. As illustrated in Figure 6, both the generated sketches and their corresponding massing proxy images are ranked by sketch score, following the same order established in Figure 5. The scores exhibit strong stability throughout, from the sketch generation stage r_{21}-r_{24} through to massing composition r_{41}-r_{44}, indicating that the proposed framework consistently delivers within a fixed budget.

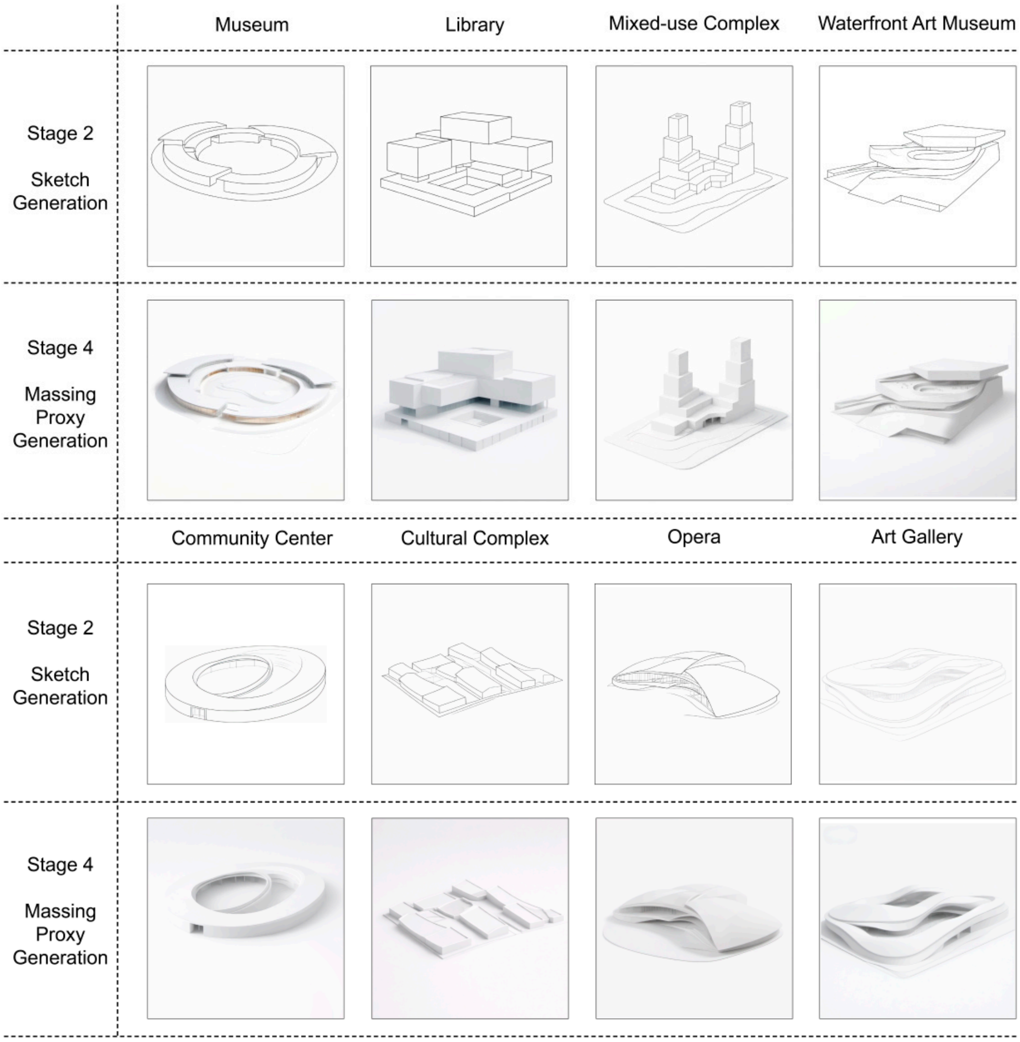


Figure 5. Sketches and massing proxy images generated for eight project types under Method A2.

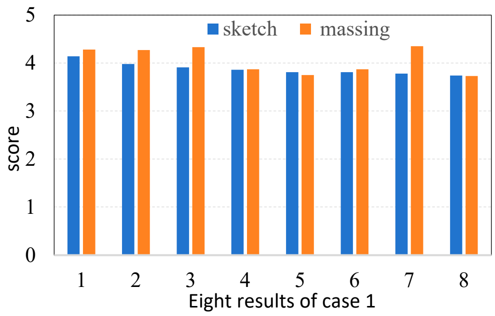


Figure 6. Sketches and massing proxy images generated for eight project types under Method A2, with scoring results.

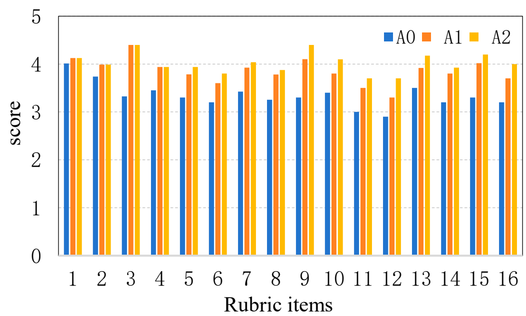
For all schemes, metric Alignment and Geometry Completeness preserved the core logic and spatial framework passed down from the concept generation stage. Analysis indicates that the mechanisms of the prototype library and knowledge graph functioned as "design space boundaries" during generation. Furthermore, the massing proxy image generation quality aligns with the sketch ranking, also displaying scores from highest to lowest, demonstrating that sketch outcomes significantly influence the generated massing proxy images.

Scheme 7 exhibited relatively low Legibility and Editability scores during the sketching stage due to insufficient clarity in its dominant framework and solid-void relationships. However, during the massing composition stage, its continuous surfaces and overall contour features achieved greater visual unity and formal tension, significantly enhancing Aesthetics (r_{41}) and Feasibility (r_{42}) scores. This phenomenon indicates that continuous geometry may reduce structural legibility in two-dimensional line drawings, yet can be reinforced during the massing composition stage through light-shadow interplay and volumetric continuity. Consequently, structural clarity in Stage 1 does not exhibit a purely linear correlation with visual maturity in Stage 2. Thus, the proposed methodology does not merely enhance the visual quality of individual schemes, but rather constructs a structurally stable, explorable design space to achieve the generative characteristic of "controlled diversity".

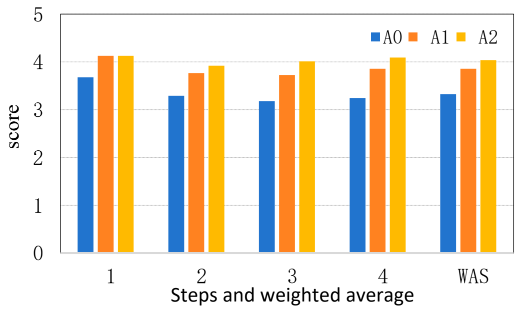
4.2.3. Quality Distribution and Stability Testing of Generated Results

We evaluate the workflow on 35 projects across 8 categories (matching those in Figure 5) with approximately balanced distribution, chosen to balance coverage and expert-evaluation cost under a fixed sampling budget. We evaluated the average scores and weighted average improvement rates for each metric under settings A0, A1, and A2. Variance in scoring was calculated to test proposal stability, alongside consistency and significance tests for juror evaluations.

The assessment comprises four steps and sixteen rubric items. As shown in Figure 7(a), the A2 configuration consistently outperformed both A0 and A1, particularly achieving higher scores on items 3, 9, 13, and 15. Figure 7(a) presents the average scores generated across steps for all items. The average scores for each step and the overall average score (WAS) under the A2 configuration hovered around 4 points, outperforming the other two configurations. Compared to A0, the improvement rate was 21.39%, and compared to A1, it was 4.68%. Calculating the variance of all evaluation values across 16 metrics for the 35 projects under settings A0, A1, and A2 yields values of 0.35, 0.22, and 0.18 respectively. These figures indicate that the A0 generation scheme exhibits greater randomness in quality, while the A2 scheme demonstrates the highest stability. These results demonstrate that the proposed generation scheme is applicable to most items and possesses good generalizability.



(a)



(b)

Figure 7. Scoring results for 35 projects across 8 categories under three settings (A0, A1, A2) for two stages and four steps. (a) Comparison of scores across 4 steps and 16 metrics (b) Comparison of average scores per step versus total weighted average.

We employ the Shapiro-Wilk test (with a significance level of 0.05) to choose between a paired *t*-test and the Wilcoxon signed-rank test for paired comparisons, and report two-tailed p-values together with paired Cohen's d (0.2/0.5/0.8 for small/medium/large effects). Based on Table 7, A1 significantly outperforms A0 (moderate effect). A2 further improves over A1 (small-to-moderate effect), and achieves the largest gain over A0 (moderate-to-large effect).

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 26 February 2026 doi:10.20944/preprints202602.1657.v1

Table 7. Paired comparisons on seed-averaged project-level WAS (35 projects).

Comparison	ΔWAS	p-value	Cohen's d
A1 – A0	+0.53	0.015	0.57
A2 – A1	+0.18	0.043	0.34
A2 – A0	+0.71	0.006	0.75

4.3. Massing Proxy Image-Conditioned Rendering

This section compares direct text-generated renderings with massing proxy image-conditioned renderings, demonstrating the necessity of refining the massing proxy before rendering. By rendering the massing proxy image-conditioned outputs generated for the first three projects in Figure 5, we contrast these with results directly inputting architectural design requirements into DALL·E 3. As shown in Figure 8, the first column presents the Baseline-A (Text-only) results, following the workflow "LLM → prompt → DALL·E 3"; Column 2 presents Baseline-B (massing proxy image-conditioned but without knowledge), following the workflow "massing proxy image → DALL·E3", where the massing proxy image was directly generated by DALL·E3 based on the design requirements; Column 3 displays our method Ours (Knowledge-Augmented), following the workflow "prototype library and KG → massing proxy image → DALL·E3".

Figure 8. Rendered effect of Scheme 1 Top-3 generated massing proxy images.

Representative results in Figure 8 reveal that while Baseline-A produces visually appealing images, its spatial organization and volumetric logic often exhibit pronounced expression-driven characteristics. This tendency leads to deviation from volumetric logic when influenced by materials and lighting effects. Baseline-B's outcomes depend on the stability of the generated massing proxy image. However, lacking knowledge constraints and prototype logic support, its volumes often appear templated, with weak void-solid relationships and connection strategies. This results in rendered outputs that are formally thin and lacking in design tension. In contrast, when our knowledge-augmented method uses the generated massing proxy image as conditional input, the final rendering tends to inherit the volumetric relationships and void-solid organization defined by the massing, embodying a "Imagery → Driver →Operation" process. Comparative results demonstrate that this workflow enhances controllability and consistency from concept to visual expression through structural sketches and massing intermediate representations.

5. Discussion

Under a unified budget and a consistent comparison agreement, our knowledge-driven process shows a more stable advantage in terms of overall scoring. Compared with the baseline setting that directly generates prompts, solutions based on knowledge routing are easier to meet project constraints, and the massing compositions and their relationship with abstract concepts are also clearer and easier to understand. This improvement does not depend on a more powerful generation model, but comes from the structuring of conceptual content before generation. Even if only massing proxy images are output in the early stage without pursuing accuracy at the structural level, this improvement is still significant because it directly affects the possibility of subsequent refinement and comparison of different candidates. Experimental results show that through knowledge-augmented prompt vocabulary compilation, the narrative intention can be more reliably transformed into a discussable massing composition.

From the perspective of mechanism, these improvements mainly come from the collaborative division of labor between "prototype library, knowledge graph and prompt compilation". The prototype library provides reusable image entries, their drivers and operational a priori information, so that the concept is no longer completely dependent on the improvisation of the language model. The knowledge graph routes these prompts to more specific strategies and geometry, such as massing types, gap patterns, opening logic and operation priorities, thus converting descriptive narratives into executable structures. This routing process also reduces the possibility of conceptual drift, indicating that the same project target can maintain recognizable main massing and key space prototypes in different candidate schemes. In other words, the role of knowledge graph is not to replace the generation model, but to provide a more traceable and reusable intermediate channel from Concept to Massing, thus improving the interpretability and controllability of generation.

Due to the significant randomness of diffusion sampling and prompt interaction, this paper adopts a Rubric-based Top-K strategy to control uncertainty within a fixed budget, so as to obtain comparable choices. Compared with the gate-through of "pass or fail", Top-K is more aligns with the exploration logic of the conceptual stage: it permits diversity, but ranks the candidates into a few feasible solutions through a unified scoring criterion, thus avoiding low pass rates and sparse samples that are difficult to attribute due to high thresholds. The scoring standard plays an important role in early repeatability evaluation. It clarifies the dimensions of constraint fit, structural readability and editability, so that the differences between different ablation settings can be discussed under the same standard.

Regarding the write-back mechanism, this paper designs it as a single iteration, mainly out of the trade-off between practical feasibility and reproducibility: at this stage, it is difficult for the model to stably understand and strictly implement "fine-grained modification based on existing output", and multi-round editing is easy to introduce new deviations, thus weakening the attributability of control experiments. Therefore, this paper positions the rewriting as the "minimum available closed loop": the main questions of the scoring standard are fed back to the prompt template to achieve constraint and massing-oriented correction. At the same time, it is necessary to clarify the representation boundary of this paper: the output is the massing proxy image in the conceptual stage, which aims to present the massing organization and spatial prototype, which is not equivalent to the precise 3D geometry that can be directly used in CAD or BIM. This choice is consistent with the research goal of this paper, which prioritize addressing issues of "readability, comparability, and controllability" at an early stage.

We observed several failure patterns during training. In Stage 1, unclear cues lead to area or storey mismatches, and enough negative cues need to be added to avoid generating distractions such as environments and stray colors that are not related to the building massing. In Stage 2, deviations in massing stories, openings and relationships reduce editability and constraint compliance. Sampling hyper-parameters (Sampler method, CFG, Sampling steps) together with LoRA weights define the system boundaries: too high results in detail domination or pattern collapse, too low results in insufficient constraint responsiveness. Adjustments to the ControlNet sampling process can also have an impact on the generated results, e.g., an overly high Ending Control Step will make the constraints too strong and weaken the prompt effect. Tighter operational constraints and adaptive condition strengths, more stable sketch quality control, and parameter interval normalization will be used to reduce drift and improve cross-task generalization and controllability.

In summary, although the workflow has made steady improvements in early massing characterization and reproducible screening, there are still three limitations. Firstly, expert scores are inevitably subjective: although the consistency metrics in this report quantify the stability of the score, the evaluation results may still be affected by the background and preferences of the reviewers. Secondly, the current index system focuses on the readability and demand consistency at the "output quality" level, and has not been included in the automatic verification of downstream hard constraints (such as constructability, functional areas and standards), thus limiting the conclusion to the project feasibility and the promotion and application of the subsequent development stage. Third, its universality still has limitations: the existing prototype library and knowledge graph cover a limited range of types and projects, and need to be expanded and verified in more types and regional contexts. In order to solve the above limitations, the future work will include:

(1) Expanding data and project coverage under the premise of maintaining budget reproducibility, and introducing richer review configurations and more objective automatic metrics to supplement subjective scoring;

(2) Adding hard constraint verification modules for geometry and performance to form from concept to closed loop of massing composition;

Continuously expand the coverage and representation ability of the prototype library and knowledge graph to improve the generalization performance of cross-project types and reduce the drift in the iteration process.

6. Conclusions

This paper proposes a knowledge-augmented generation process for the architectural concept design stage, which is used to stably transform the narrative intention in the project requirements into an easy-to-understand and comparative massing expression. Unlike the generation method directly based on prompts, this method provides reusable drivers and operations a priori information through the prototype library, and completes the path construction and compilation of "concept-strategy-geometry" through the knowledge graph, so as to form more structured clue information before generation. By combining rubric scoring and Top-K ranking under a fixed budget, the system can better meet key constraints under the background of random diffusion generation and improve the readability of massing composition, spatial physics relationships and organizational hierarchy.

In terms of experimental design, this paper adopts a unified generation budget and control protocol to compare the performance of the tasks in the next two stages of different knowledge enhancement settings. By ablation settings, the performance improvement is attributed to "knowledge routing and structured compilation" rather than the generation model itself. The results show that knowledge-augmented processes can generate more stable massing compositions and more challenging spatial prototypes, demonstrating the practical value of prototype libraries and knowledge routing in the early design stage. At the same time, this paper implements the write-back mechanism as a single iteration, emphasizing its role as the smallest available closed loop: the evaluation results are used to update the path and prompt template to reduce concept drift and improve the practicality of the candidate scheme.

Future work will focus on expanding the coverage of the prototype library and knowledge graph, introducing automatic verification of specifications and performance, and enhancing the executability and feedback control of intermediate representation to further improve the cross-project generalization ability and iterative stability.

Author Contributions: Conceptualization, S.S. and Y.F.; methodology, S.S. and Y.F.; Software, S.S.; validation, S.S. and Y.F.; Formal analysis, S.S. and Y.F.; Investigation, S.S. and Y.F.; Resources Data curation, S.S.; writing—original draft preparation, S.S. and Y.F.; writing—review and editing, S.S. and Y.F.; visualization, S.S.; supervision, Y.F.; funding acquisition, Y.F. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported and funded by the National Natural Science Foundation of China (grant number 52278031).

Data Availability Statement: The data that support the findings of this study are available from the corresponding author upon reasonable request.

Conflicts of Interest: The authors declare no conflicts of interest.

Appendix A

Table A1. Rubric item definitions and evaluation criteria.

Step	Rubric Item	Definition
1	Concept Clarity (r_{11})	**Definition**: Clear description of the dominant massing structure, void-solid relationships, and experiential sequence. **Rubric** (1-5): 1 = Narrative is loose and fails to point towards massing; 3 = Can describe main framework and key voids but lacks specificity; 5 = Main framework/void-solid relationships/experience sequence are clear and directly translatable into an operational chain.
	Constraint Coverage (r_{12})	**Definition**: Degree of coverage for key constraints, such as scale/entrance organization/traversability/courtyard interaction/lighting and circulation routes. **Rubric** (1-5): 1 = Multiple hard constraints ignored; 3 = Primary hard constraints addressed but lacking detail; 5 = Hard constraints addressed with clear spatial response.
	Operational Specificity (r_{13})	**Definition**: Whether operations are specific, combinable, and controllable, avoiding uncontrolled statements. **Rubric** (1-5): 1 = Adjectives without operational steps; 3 = Operational steps with ambiguous boundaries; 5 = Clear operational sequence mappable to geometric steps.
	Novelty under Control (r_{14})	**Definition**: Demonstrates differentiation without compromising readability or translatability. **Rubric** (1-5): 1 = heavily derivative; 3 = innovative but conventionally expressed; 5 = distinctly recognizable while maintaining structural clarity.
2	Aesthetics (r_{21})	**Definition**: The sketch's overall contour and hierarchical relationships are the main focus, along with its proportion, rhythm, and continuity. **Rubric** (1–5): 1 = Fragmented or proportionally unbalanced; 3 = Ordinary but clear hierarchy; 5 = Elegant, continuous outline with a clear rhythm.
	Editability (r_{22})	**Definition**: Whether the sketch supports additional development and has a solid structural foundation that permits details to be added or removed without affecting the overall design. **Rubric** (1–5): 1 = unstable structure that prevents refinement; 3 = Refinement is feasible but necessitates extensive redrawing; 5 = Sturdy framework that permits methodical detail addition.
	Legibility (r_{23})	**Definition:** Whether the sketch clearly illustrates the relationship between drivers and key spaces.

		Rubric (1-5): 1 = Relationships between volumes and circulation paths are indistinguishable; 3 = Legible but information-dense; 5 = Primary framework, voids, and circulation paths are immediately recognizable.
	Design Intent Match (r_{24})	**Definition**: Whether the sketch aligns with the project's intent/narrative. **Rubric** (1-5): 1 = Significant deviation from intent; 3 = Partially aligned; 5 = Primary intentions are consistently visualized and coherent.
3	Alignment (r_{31})	**Definition**: Whether the form description aligns with the Stage-1 conceptual skeleton, such as no deviation in principal axes, nodes, void types, or connection methods. **Rubric** (1-5): 1 = Description deviates from sketch; 3 = Primary relationships align but details drift; 5 = Massing, voids/solids, and connections align with operation chains.
	Rationality (r_{32})	**Definition**: Whether the volumetric organization and spatial logic are reasonable, with fundamental interpretability of entrances/circulation/public and service zones. **Rubric** (1-5): 1 = Logical contradictions; 3 = Fundamentally coherent but lacking explanatory constraints; 5 = Logically consistent with clearly articulated constraints.
	Translatability (r_{33})	**Definition**: Describes whether it can directly drive massing proxy image generation, clearly articulating volumetric relationships, dimensional hierarchies and connections. **Rubric** (1-5): 1 = Unable to form geometry; 3 = Generable but highly ambiguous; 5 = Minimal ambiguity, stable generation of consistent forms achievable.
	Geometry Completeness (r_{34})	**Definition**: Describes coverage of key geometric elements required for generation, including dominant curves/node volumes/connecting components/opening types. **Rubric** (1-5): 1 = Critical elements missing; 3 = Elements present but lacking parameter ranges; 5 = Elements present with controllable ranges/constraints provided.
4	Aesthetics (r_{41})	**Definition**: The aesthetic quality of the massing images in terms of proportion, rhythm, and continuity, focusing on the overall silhouette and hierarchical relationships. Rubric (1-5): 1 = Fragmented/proportionally unbalanced; 3 = Clear hierarchy but ordinary; 5 = Elegant, continuous outline with distinct rhythm.
	Feasibility (r_{42})	**Definition:** The massing proxy image's basic constructability and structural integrity, avoiding major impossibilities like unrealistically thin shells or cantilevers. **Rubric** (1–5): 1 = Clearly unbuildable; 3 = Buildable but dubious; 5 = Overall feasible with believable structural logic.
	Compliance (r_{43})	**Definition**: Complying with strict requirements during the massing composition phase, such as massing control, scale, number of stories, and courtyard/through-access. **Rubric** (1–5): 1 = Violates key constraints;3 = Generally compliant; 5 = Strictly compliant and exhibits optimization.
	Editability (r_{44})	**Definition**: Whether there is room for improvement in the massing proxy image, with volumes that can be broken down and detailing strategies integrated. **Rubric** (1–5): 1 = Cannot be developed further or requires complete overhaul; 3 = Can be developed but needs restructuring; 5 = Clear structure, ready for in-depth refinement.

Appendix B

 doi:10.20944/preprints202602.1657.v1

Table A2. Rubric item weightings.

Stage	Stage weight	Rubric Item	Weight
1	W_1 =0.20	Concept Clarity	W_{11} =0.30
		Constraint Coverage	W_{12} =0.30
		Operational Specificity	W_{13} =0.25
		Novelty under Control	W_{14} =0.15
2	W_2 =0.25	Aesthetics	W_{21} =0.20
		Editability	W_{22} =0.30
		Legibility	W_{23} =0.25
		Design Intent Match	W_{24} =0.25
3	W_3 =0.25	Alignment	W_{31}= 0.30
		Rationality	W_{32} =0.25
		Translatability	W_{33} =0.30
		Geometric Completeness	W_{34} =0.15
4	W_4 =0.30	Aesthetics	W_{41} =0.15
		Feasibility	W_{42} =0.35
		Compliance	W_{43} =0.25
		Editability	W_{44}= 0.25

Appendix C

Table A3. Concept Generation Schemes 4-8 of A1 setting for Case Study 1.

ID	Schemes of A1
4	**Title:** Ring of Still Water, where thick boundaries enclose a tranquil light court **Imagery**: Like a ring of still water forming a circle, the center is a protected void of light, with circulation meandering gently along the ring. **Strategy and Geometry**: A single-storey, thick ring-shaped volume encloses the central courtyard. A continuous ring corridor on the inner side organizes exhibition routes and resting areas. The thickness of the ring varies gradually from outer to inner (outer thick, inner light). Continuous or segmented light slots are cut into the ring to introduce uniform daylight, while a few incisions create multiple entrances and traversable passages. **Drivers**: ring_enclosure ǀ continuous_loop ǀ diffuse_light **Operations**: generate_thick_ring_mass ǀ carve_central_void ǀ perforate_ring_skylight
5	**Title:** Glass Bubbles, Embedded in Thick Walls – A juxtaposition of lightness and weight **Imagery**: A series of transparent bubble cavities embedded within a solid rock wall, forming accessible exhibition pockets. **Strategy and Geometry**: Employing thick boundary volumes as 'container walls', hollowed out internally to embed multiple bubble-like exhibition halls/light courts; these bubbles are strung together within the thick walls to create a continuous, meandering exhibition sequence. Simultaneously, a few concealed entrances are opened along the periphery to maintain the overall contrast between the substantial exterior and the light, permeable interior. **Drivers**: light_vs_mass ǀ porosity ǀ programme_pockets **Operations**: extrude_plinth ǀ pocket_carving ǀ boolean_subtract
6	**Title:** The Transparent Translation of the Monastery Cloister, Where Boundaries Become Glass and Order Becomes Circulation **Imagery**: The cloister encircles the courtyard, yet its boundary transforms into a transparent membrane. **Strategy and Geometry**: A single-storey cloister encircles an inward-facing courtyard to establish clear order. The outer perimeter maintains a tranquil boundary through restrained openings, while the interior enhances courtyard interaction via continuous transparent interfaces. Partial

	voids at ground level create passageways, with the entrance threshold at employing double-glazing and blurring techniques to soften the sense of boundary. **Drivers**: continuity_loop | threshold_blur | calm_core **Operations**: loop_circulation | wrap_double_skin | carve_passage
7	**Title:** Sponge Boundary, where architecture and forest interpenetrate **Imagery**: Like a sponge absorbing water, allowing forest paths, light and wind to permeate the architecture. **Strategy and Geometry**: The boundary is uniformly perforated to create multiple entry points and permeable seams. Pathways between interior and exterior spaces interpenetrate, converging and pausing at several pocket courtyards. The whole is composed of homogeneous modules, softening primary and secondary axes to enable visitors to enter, traverse, and circulate freely from multiple directions, exploring at will. **Drivers**: porous_edges | multi_entry_permeability | isotropic_field **Operations**: punch_entries | carve_multiple_courtyards | equalise_modules
8	**Name:** Wind-Patterned/Dune Roof, Directional Field-Driven Continuous Undulations **Imagery**: Wind etches parallel, gentle ripples across the dune surface, directionality shaping spatial rhythm. **Strategy and Geometry**: Organizing the roof and profile along a dominant directional field, the single-storey volume undulates continuously in the same direction, creating a gradient of elevation and light/shadow. Slender skylights cut into the wind-patterned direction, allowing light and circulation to jointly reinforce the experiential rhythm of "moving with the flow, gradually turning, then unfolding". **Drivers**: directional_flow | field_deformation | gradient_experience **Operations**: align_flow_direction | apply_field_deform | carve_light_slots

References

1. Li, C.; Zhang, T.; Du, X.; et al. Generative AI models for different steps in architectural design: A literature review. Frontiers of Architectural Research 2025, 14(3). DOI: 10.1016/j.foar.2024.10.001.
2. Zhuang, X.; Zhu, P.; Yang, A.; Caldas, L. Machine learning for generative architectural design: Advancements, opportunities, and challenges. Automation in Construction 2025, 174, 106129. https://doi.org/10.1016/j.autcon.2025.106129.
3. Chen, J.; Wang, D.; Shao, Z.; Zhang, X.; Ruan, M.; Li, H.; Li, J. Using artificial intelligence to generate master-quality architectural designs from text descriptions. Buildings 2023, 13(9), 2285.
4. Cheng, K.; Neisch, P.; Cui, T. From concept to space: a new perspective on AIGC-involved attribute translation. Digital Creativity 2023, 34(3), 211–229.
5. Bauscher, E.; Dai, A.; Elshani, D.; et al. Learning and Generating Spatial Concepts of Modernist Architecture via Graph Machine Learning. In Proceedings of the 29th International Conference on Computer-Aided Architectural Design Research in Asia (CAADRIA), 2024; Volume 1, pp. 159–168. DOI: 10.52842/conf.caadria.2024.1.159.
6. Chen, R.; Yao, X.; Jiang, X. LLM4DESIGN: An Automated Multi-Modal System for Architectural and Environmental Design. arXiv 2024, arXiv:2407.12025. https://doi.org/10.48550/arXiv.2407.12025.
7. Zheng, H. A diffusion-based machine learning method for 3D architectural form-finding. Frontiers of Architectural Research 2025. DOI: 10.1016/j.foar.2024.12.002.
8. Horvath, A.-S.; Pouliou, P. AI for conceptual architecture: reflections on designing with text-to-text, text-to-image, and image-to-image generators. Frontiers of Architectural Research 2024, 13(3), 593–612.
9. Seneviratne, S., Senanayake, D., Rasnayaka, S., Vidanaarachchi, R.,Thompson, J., 2022. Dalle-urban: capturing the urban design expertise of large text to image transformers. In: 2022 International Conference on Digital Image Computing: Techniques and Applications (DICTA). IEEE, pp. 1-9.
10. Zhang, L.; Agrawala, M. Adding conditional control to text-to-image diffusion models. arXiv 2023, arXiv:2302.05543.

11. Tan, L.; Luhrs, M. Employing generative AI Midjourney to enhance divergent and convergent thinking within an architect's creative design process. The Design Journal 2024, 677–699. https://doi.org/10.1080/14606925.2024.2372159.
12. Sukkar, A.W.; Fareed, M.W.; Yahia, M.W.; et al. Analytical Evaluation of Midjourney Architectural Virtual Lab: Defining Major Current Limits in AI-Generated Representations of Islamic Architectural Heritage. Buildings 2024, 14(3), 786. DOI: 10.3390/buildings14030786.
13. Chen, J.; Shao, Z.; Hu, B. Generating Interior Design from Text: A New Diffusion Model-Based Method for Efficient Creative Design. Buildings 2023, 13, 1861. https://doi.org/10.3390/buildings13071861.
14. Zhong, C.; Shi, Y.; Wang, L. AI-enhanced performative building design optimisation and exploration. In Proceedings of the 29th International Conference on Computer-Aided Architectural Design Research in Asia (CAADRIA), 2024; Volume 1, pp. 59–68.
15. Tong, H.; Türel, A.; Şenkal, H.; Ergun, S.; Güzelci, O.Z.; Alaçam, S. Can AI function as a new mode of sketching: a teaching experiment with freshmen. International Journal of Emerging Technologies in Learning 2023, 18(18).
16. Ye, X.; Du, J.; Ye, Y. Masterplangan: facilitating the smart rendering of urban master plans via generative adversarial networks. Environment and Planning B: Urban Analytics and City Science 2022, 49(3), 794–814.
17. Qian, W.; Xu, Y.; Li, H. A self-sparse generative adversarial network for autonomous early-stage design of architectural sketches. Computer-Aided Civil and Infrastructure Engineering 2022, 37(5), 612–628.
18. Xu, Y.L.W. Research on architectural sketch to scheme image based on context encoder. In Proceedings of the 28th International Conference of the Association for Computer-Aided Architectural Design Research in Asia (CAADRIA), 2023; Volume 1, pp. 69–78.
19. Wang, Z.; Jin, H.; Du, X.; et al. From Architectural Sketch to Conceptual Representation: Using Structure-Aware Diffusion Model to Generate Renderings of School Buildings. 2025.
20. Kim, D.; Guida, G.; García, J.L.; López, D.C.Y. Participatory urban design with generative adversarial networks. In Proceedings of the 27th International Conference of the Association for Computer-Aided Architectural Design Research in Asia (CAADRIA), 2022; Volume 2, pp. 486–494.
21. Han, S.; Jiang, Y.; Huang, Y.; Wang, M.; Bai, Y.; Spool-White, A. Scan2Drawing: use of deep learning for as-built model landscape architecture. Journal of Construction Engineering and Management 2023, 149(5).
22. Kim, S.; Kim, D.; Choi, S. CityCraft: 3D virtual city creation from a single image. The Visual Computer 2020, 36, 911–924.
23. Zhuang, J.; Li, G.; Xu, H.; Xu, J.; Tian, R. TEXT-TO-CITY: controllable 3D urban block generation with latent diffusion model. In Proceedings of the 29th International Conference of the Association for Computer-Aided Architectural Design Research in Asia (CAADRIA), 2024; pp. 196–178.
24. de Miguel Rodríguez, J.; Villafañe, M.E.; Piskorec, L.; Caparrini, F.S. Generation of geometric interpolations of building types with deep variational autoencoders. Design Science 2020, 6, e34.
25. Zheng, H.; Yuan, P.F. A generative architectural and urban design method through artificial neural networks. Building and Environment 2021, 205, 108178.
26. Kim, F.C.; Huang, J. Towards a machine understanding of architectural form. In Proceedings of the 27th International Conference of the Association for Computer-Aided Architectural Design Research in Asia (CAADRIA), 2022; Volume 1, pp. 727–736.
27. Liu, Y.; Li, H.; Deng, Q.; Hu, K. Diffusion probabilistic model assisted 3D form finding and design latent space exploration: a case study for Taihu stone spatial transformation. In Proceedings of the 2023 Digital FUTURES: International Conference on Computational Design and Robotic Fabrication (CDRF 2023); Springer, 2023; pp. 11–23.
28. Geng, T.; Yang, Y. DiffDesign: Controllable diffusion with meta prior for efficient interior design generation. PLoS ONE 2025, 20(9), e0331240. DOI: 10.1371/journal.pone.0331240.
29. Tono, A.; Huang, H.; Agrawal, A.; Fischer, M. Vitruvio: 3D building meshes via single perspective sketches. arXiv 2022, arXiv:2210.13634.
30. Navarro-Mateu, D.; Carrasco, O.; Cortes Nieves, P. Colour patterns to architecture conversion through conditional generative adversarial networks. Biomimetics 2021, 6(1), 16.

31. Asmar, K.; Sareen, H. Machinic interpolations: a GAN pipeline for integrating lateral thinking in computational tools of architecture. In Proceedings of the 24th Conference of the Iberoamerican Society of Digital Graphics, Online, 18–20 November 2020; pp. 18–20.
32. Zhang, H.; Blasetti, E. 3D architectural form style transfer through machine learning. In Proceedings of the 25th International Conference of the Association for Computer-Aided Architectural Design Research in Asia (CAADRIA), 2020; pp. 659–668.
33. Zhang, H.; Huang, Y. Machine learning aided 2D-3D architectural form finding at high resolution. In Proceedings of the 2020 Digital FUTURES: the 2nd International Conference on Computational Design and Robotic Fabrication (CDRF 2020); Springer, 2020; pp. 159–168.
34. Quan, S.J. Urban-GAN: an artificial intelligence-aided computation system for plural urban design. Environment and Planning B: Urban Analytics and City Science 2022, 49(9), 2500–2515.
35. Li, J.; Guo, F.; Chen, H. A study on urban block design strategies for improving pedestrian-level wind conditions: CFD-based optimisation and generative adversarial networks. Energy and Buildings 2023, 304, 113863.
36. Ampanavos, S.; Malkawi, A. Early-phase performance-driven design using generative models. In International Conference on Computer-Aided Architectural Design Futures; Springer, 2021; pp. 87–106.
37. Huang, C.; Zhang, G.; Yao, J.; Wang, X.; Calautit, J.K.; Zhao, C.; An, N.; Peng, X. Accelerated environmental performance-driven urban design with generative adversarial network. Building and Environment 2022, 224, 109575.
38. Li, J.; Luo, Y.; Lu, S.; Zhang, J.; Wang, J.; Guo, R.; Wang, S. ChatDesign: Bootstrapping Generative Floor Plan Design with Pre-Trained Large Language Models. In Proceedings of the 29th International Conference of the Association for Computer-Aided Architectural Design Research in Asia (CAADRIA), 2024; pp. 99–108.
39. Wang, S.; Chen, C.; Le, X.; et al. CAD-GPT: Synthesising CAD Construction Sequence with Spatial Reasoning-Enhanced Multimodal LLMs. In Proceedings of the AAAI Conference on Artificial Intelligence 2024, 39(8). DOI: 10.1609/AAAI.v39i8.32849.
40. Pan, S.; Luo, L.; Wang, Y.; et al. Unifying Large Language Models and Knowledge Graphs: A Roadmap. IEEE Transactions on Knowledge and Data Engineering 2024, 36(7), 20. DOI: 10.1109/TKDE.2024.3352100.
41. Zhu, Y.; Wang, X.; Chen, J.; Qiao, S.; Ou, Y.; Yao, Y.; Deng, S.; Chen, H.; Zhang, N. LLMs for knowledge graph construction and reasoning: Recent capabilities and future opportunities. World Wide Web 2024, 27, 58.
42. Hong, S.; Zhuge, M.; Chen, J.; Zheng, X.; Cheng, Y.; Zhang, C.; Wang, J.; Wang, Z.; Yau, S.K.S.; Lin, Z.; Zhou, L.; Ran, C.; Xiao, L.; Wu, C.; Schmidhuber, J. MetaGPT: Meta programming for a multi-agent collaborative framework. In Proceedings of the International Conference on Learning Representations (ICLR), 2024.
43. Qian, C.; Liu, W.; Liu, H.; et al. ChatDev: Communicative Agents for Software Development. arXiv 2023, arXiv:2307.07924.
44. Park, J.S.; O'Brien, J.C.; Cai, C.J.; Morris, M.R.; Liang, P.; Bernstein, M.S. Generative Agents: Interactive Simulacra of Human Behaviour. In Proceedings of the 36th Annual ACM Symposium on User Interface Software and Technology (UIST '23), 2023; 22 pages. https://doi.org/10.48550/arXiv.2304.03442.
45. Tang, K.M.M.; Tam, C.M.H.; Herr, C.M. Archi-Agents: Approximating the architectural design process with collaborative specialised multi-agent system LLM and building information modelling. In Proceedings of the 30th International Conference on Computer-Aided Architectural Design Research in Asia (CAADRIA), 2025; Volume 3, pp. 141–150.
46. Lakoff, G.; Johnson, M. Metaphors We Live By; University of Chicago Press: Chicago, IL, USA, 1980.
47. Oxman, R. Digital architecture as a challenge for design pedagogy: Theory, knowledge, models and medium. Design Studies 2008, 29(2), 99–120.
48. Knight, T.W. Transformations in Design: A Formal Approach to Stylistic Change and Innovation in the Visual Arts; Cambridge University Press: Cambridge, UK, 1994.
49. Biomimicry Institute. AskNature database. Available online: https://asknature.org (accessed on 19 February 2026).

50. Benyus, J.M. Biomimicry: Innovation Inspired by Nature; William Morrow: New York, NY, USA, 1997.
51. Alexander, C.; Ishikawa, S.; Silverstein, M. A Pattern Language: Towns, Buildings, Construction; Oxford University Press: New York, NY, USA, 1977.
52. Norberg-Schulz, C. Genius Loci: Towards a Phenomenology of Architecture; Rizzoli: New York, NY, USA, 1980.
53. Ching, F.D.K. Architecture: Form, Space, & Order, 4th ed.; John Wiley & Sons: Hoboken, NJ, USA, 2015.
54. Oxman, R. Prior knowledge in design: A dynamic knowledge-based model of design and creativity. Design Studies 1990, 11, 17–27.
55. Zhang, Y.; Al-Hussein, M.; Li, X. UCRD–Onto: A knowledge framework for user-centred design in residential built environments. Engineering, Construction and Architectural Management 2026, in press. DOI: 10.1108/ECAM-01-2025-0015.
56. Pallasmaa, J. The Eyes of the Skin: Architecture and the Senses, 3rd ed.; Wiley: Chichester, UK, 2012.
57. Royal Institute of British Architects (RIBA). RIBA Plan of Work 2020: Overview; RIBA: London, UK, 2020.
58. Lawson, B. How Designers Think: The Design Process Demystified, 4th ed.; Elsevier/Architectural Press: Amsterdam, The Netherlands; Boston, MA, USA, 2005.
59. Purcell, A.T.; Gero, J.S. Drawings and the design process: A review of protocol studies in design and other disciplines. Design Studies 1998, 19(4), 389–430. https://doi.org/10.1016/S0142-694X (98) 00015-5.
60. Goldschmidt, G. The dialectics of sketching. Creativity Research Journal 1991, 4(2), 123–143.
61. Ren, Q.; Sun, M. Using AHP-Entropy method to explore the influencing factors of spatial demand of EVs public charging stations: A case study of Jinan, China. Journal of Cleaner Production 2025, 491, 144779. https://doi.org/10.1016/j.jclepro.2025.144779.