

Article

CorbuAI: A Multimodal Artificial Intelligence-Based Architectural Design (AIAD) Framework for Computer-Generated Residential Building Design

Yafei Zhao ^{1,2} , Ziyi Ying ^{3,4,*}, Wanqing Zhao ⁵, Pengpeng Zhang ^{1,2}, Rong Xia ⁶, Xuepeng Shi ⁷ , Yanfei Ning ¹, Mengdan Zhang ¹, Xiaoju Li ⁸ and Yanjun Su ⁸

¹ School of Smart City Engineering, Qingdao Huanghai University, Qingdao 266427, China; zhaoyf@qdhhc.edu.cn (Y.Z.); zhangpp@qdhhc.edu.cn (P.Z.); 202502201073@qdhhc.edu.cn (Y.N.); 202502201078@qdhhc.edu.cn (M.Z.)

² Research Institute of Spatial Intelligence (RISI), Qingdao Huanghai University, Qingdao 266427, China

³ School of Architecture and Energy Engineering, Wenzhou University of Technology, Wenzhou 325035, China

⁴ Institute of Urban and Sustainable Development, City University of Macau, Avenida Padre Tomás Pereira, Taipa, Macau 999078, China

⁵ College of Pharmacy, Qilu Medical University, Zibo 255300, China; zhaowanqing12@qlmu.edu.cn

⁶ College of Civil Engineering and Architecture, Zhejiang University, Hangzhou 310058, China; sharon.xia@zju.edu.cn

⁷ School of Architecture and Urban Planning, Qingdao University of Technology, Qingdao 266033, China; shixuepeng@qut.edu.cn

⁸ School of Design and Architecture, Zhejiang University of Technology, Hangzhou 310023, China; 221124150105@zjut.edu.cn (X.L.); 202005540910@zjut.edu.cn (Y.S.)

* Correspondence: 20210421@wzut.edu.cn

Abstract

Integrating artificial intelligence (AI) into residential architectural design faces challenges due to fragmented workflows and the lack of localized datasets. This study proposes the CorbuAI framework, hypothesizing that a multimodal AI system integrating Pix2pix-GAN and Stable Diffusion (SD) can streamline the transition from floor plan generation to elevation and interior design within a specific regional context. We developed a custom dataset featuring 2335 manually refined Chinese residential floor plans and 1570 elevation images. The methodology employs a specialized U-Net V2.0 generator for functional layout synthesis and an SD-based model for stylistic transfer and elevation rendering. Evaluation was conducted through both subjective professional scoring and objective metrics, including the Perceptual Hash Algorithm (pHash). Results demonstrate that CorbuAI achieves high accuracy in spatial allocation (scoring 0.88/1.0) and high structural consistency in elevation generation (mean pHash similarity of 0.82). The framework significantly reduces design iteration time while maintaining professional aesthetic standards. This research provides a scalable AI-driven methodology for automated residential design, bridging the gap between schematic layouts and visual representation in the Chinese architectural context.



Academic Editor: John C. Brigham

Received: 6 January 2026

Revised: 27 January 2026

Accepted: 1 February 2026

Published: 5 February 2026

Copyright: © 2026 by the authors.

Licensee MDPI, Basel, Switzerland.

This article is an open access article distributed under the terms and conditions of the [Creative Commons Attribution \(CC BY\)](https://creativecommons.org/licenses/by/4.0/) license.

Keywords: artificial intelligence-based architectural design; multimodal generative design; residential building design; conceptual architectural design; generative adversarial networks (GANs); diffusion models

1. Introduction and Background

As smart technologies gain traction in the construction industry, the implementation strategies and methods for intelligent design within architectural firms have garnered

increasing attention [1]. Building informatics has evolved into an applied science, utilized to explore the application of computer science methods and information technology in the construction sector. Since September 2022, the technology of AI-generated content (AIGC) has experienced rapid development in the design, creation, and content industries, progressing at a swift pace.

In traditional architectural design, a complex process entails consideration of numerous parameters and various evaluation criteria. Typically, architects rely on their past experiences and creativity to generate new designs [2]. Architectural design is an iterative process [3], involving repeated attempts during conception and creation based on specific requirements. Therefore, the introduction of AI-generated content (AIGC) represents an exploration of needs and potential solutions within an uncertain space [2]. Research conducted by Harapan et al. indicates that over 50% of AI software is utilized to assist architects in conceptual design [4]. Before the emergence of AI-generated content (AIGC), most AI technologies were primarily developed for predefined objectives, utilizing functions tailored to address the initially defined scope [5].

A typical lifecycle of a building project can be observed in the stages depicted in Figure 1. Imdat As et al. identified the lack of viable tools in the early design stages as a problem in the project cycle that could be effectively addressed by AI-based products [6]. The MacLeamy Curve, proposed by Patrick MacLeamy of Hellmuth Obata Kassabaum, USA, is utilized to represent cost comparison at different stages of the building lifecycle. MacLeamy argued that changes are less likely to occur in the later stages of a building project [7,8]. Additionally, Mazurek et al. argued that the likelihood of design changes usually increases early in the design process and decreases in the later stages [9], thereby highlighting the importance of the pre-planning and conceptual scheme stages.

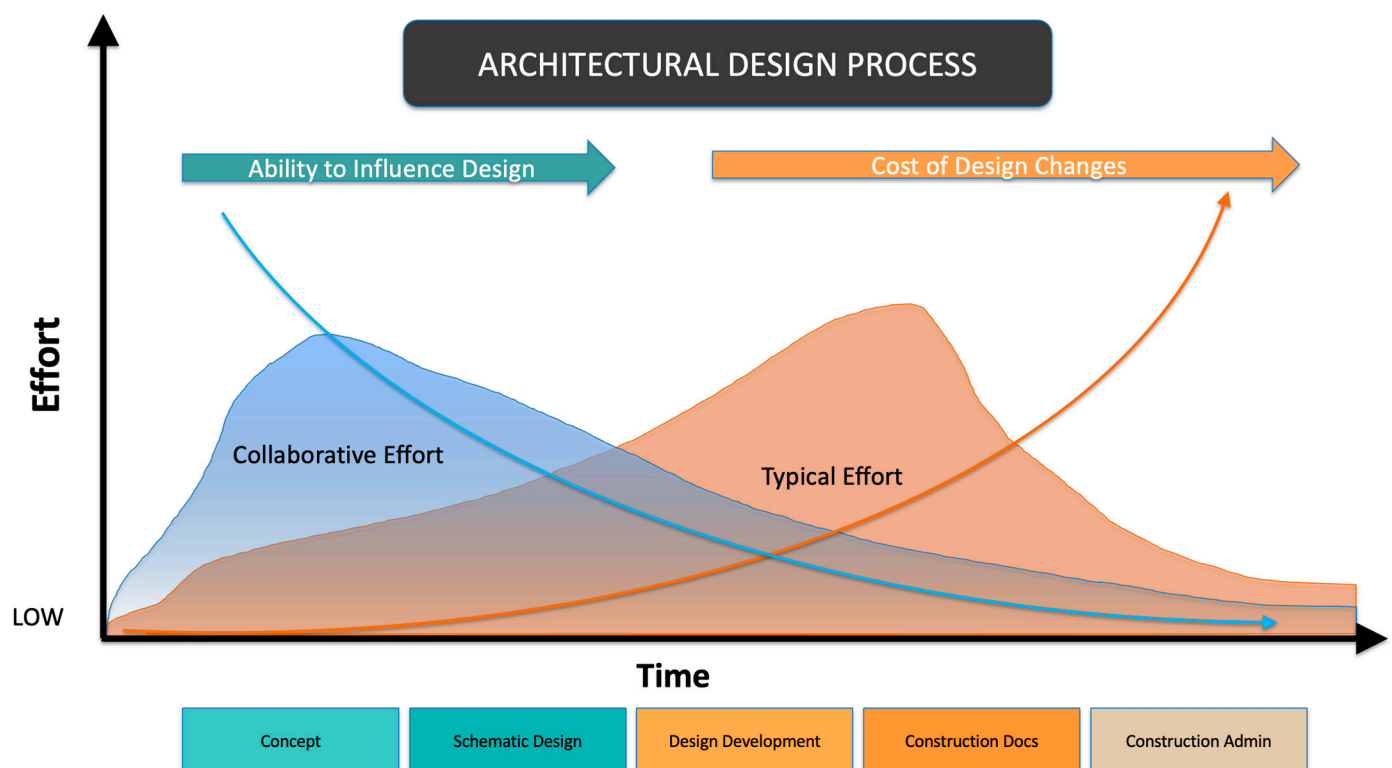


Figure 1. MacLeamy Curve for Architectural Design Process (Original concept by Davis, D. redrawn by author) [7,8].

In the conceptual and scheme stages, architectural design processes encompass various subsections such as floor plan, facade design, and space design. These processes exhibit

a relatively high standardization of geometric compositions characterized by repetitive elements and regular arrangements. Consequently, adopting computer programming languages to analyze, learn, and imitate a large number of design cases can efficiently generate design results. Drawing review and design specifications for building floor plans, facades, and interior spaces are not as stringent as those for building structures [10]. In practice, designs generated by computer programs can be utilized as intention drawings and reference drawings, facilitating clear and rapid evaluation of plane, facade, and spatial designs. Furthermore, floor plans, facades, and spatial forms all serve as expressions of a building's style and contribute to visualizing its architectural aesthetic characteristics. Additionally, the design generation function in the design phase is identified as one of the desired functions based on research and analysis of interviews with potential users.

Building upon this foundation, according to the national economic classification of industries, the "housing construction engineering industry" is a sub-sector of the "construction industry" [11]. "Residential building construction" consistently ranks first in terms of completed area and stock within the housing construction engineering industry across all sectors each year. This type encompasses almost the entirety of residential building construction, making it representative of the general housing construction engineering industry. Consequently, the previous study focuses on the specific type of residential building construction for in-depth examination, representing the broader housing construction engineering industry, and the previous results are shown in Figure 2. Based on the requirement analysis of the initial three parts, the model framework and development of the AI building design platform will be conducted during the early stages of residential buildings, specifically in the conceptual design stage.

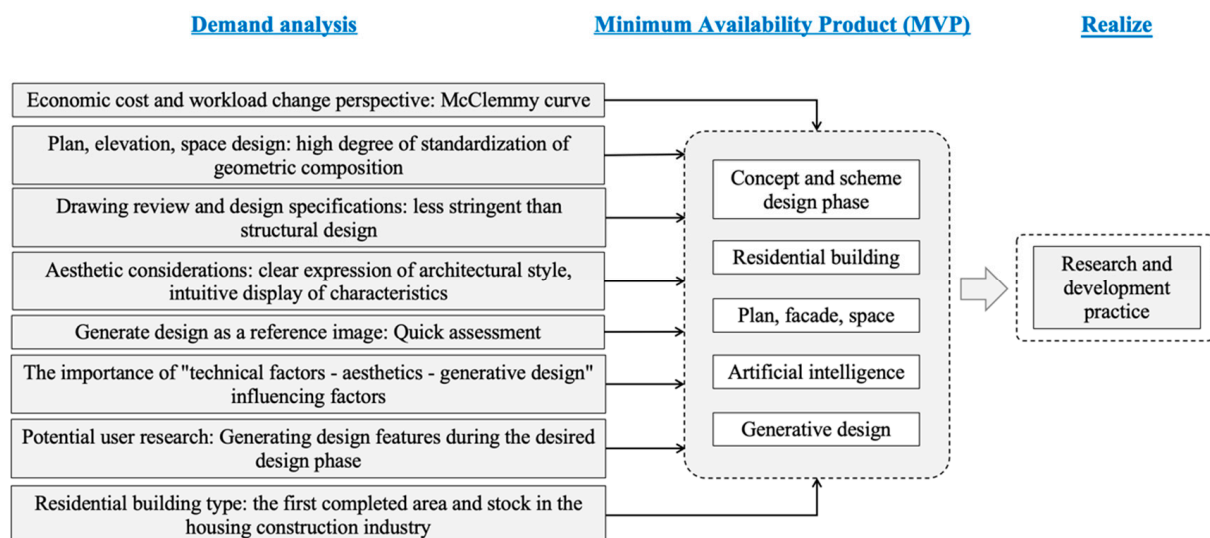


Figure 2. Requirement boundary definition of minimum viable product.

The study's background highlights the growing application of artificial intelligence (AI) technology in the construction industry, emphasizing the importance of developing an intelligent system to enhance the efficiency of architectural design. However, traditional architectural design processes face limitations in terms of efficiency and technology, creating the need to define the functional requirements of an AI-based architectural design platform. The objective of the research is to determine the platform's requirements by identifying stakeholders and collecting and filtering these requirements.

In this research, a layered architecture of the platform is developed, and discussions on the technology stack and key technologies are carried out. The study primarily addresses a significant gap in existing literature, which is the lack of practical applications tailored to

the Chinese context, especially regarding the availability of datasets and algorithm training. On this basis, the requirements of the minimum viable product (MVP) in this study solely pertain to the “conceptual and scheme design stage,” “residential building,” “plan, facade, space,” “AI,” and “generative design.” Within the scope of this study, the “multimodal AI-based architectural design platform” is conceptualized and deployed as the optimal implementation practice based on existing research.

To guide this investigation, we formulated the following three research questions (RQs) and corresponding hypotheses (Hs):

RQ1. *Can a localized, expert-refined dataset significantly improve the functional logic of AI-generated residential layouts compared to generic models?*

H1. *A domain-specific dataset tailored to regional architectural norms will result in higher spatial allocation accuracy and structural viability.*

RQ2. *How does a multimodal integration of GANs and diffusion models affect the consistency between 2D schematic design and 3D aesthetic rendering?*

H2. *A synchronized multimodal pipeline will maintain a consistent structural “DNA” across different design stages, reducing topological errors in the transition from 2D to 3D.*

RQ3. *To what extent can the CorbuAI framework enhance design efficiency in professional architectural practice?*

H3. *The automated generative workflow will reduce the early-stage schematic iteration cycle by over 90% compared to traditional manual methods.*

As a solution, the study uses original and novel datasets specific to China’s residential buildings and optimizes GAN/AI algorithms to better fit China’s practical application environment. In terms of methodology, the study analyzes platform requirements, architecture, and key technologies, leading to the proposal of CorbuAI (computer-generated residential building, CORBU), an AI-based architectural design platform leveraging multimodal AI technologies. This platform consists of three core features: building floor plan generation, building facade generation, and interior space style transfer. The research process involves data collection and preprocessing, fine-tuning pre-trained models, and evaluating model training results. Finally, the study successfully developed an integrated platform and verified its feasibility in real-world architectural design scenarios. The platform provides both theoretical and practical support for applying AI technology within the architectural design industry.

2. Literature Review

2.1. Historical Evolution of AI in Architectural Design

With the development of automated architectural design, computer-aided design and manufacturing (CAD-CAM) technology has been introduced into the traditional architectural field [12]. Warszawski and Sangrey argue that the combination of CAD-CAM provides the basis for future architectural systems [13]. Although rule-based design models dominate in academia and have achieved positive results in some studies [14], many architectural design problems remain non-deterministic. For example, Sacks points out that automation in architecture often faces challenges [12], and there may be multiple solutions for the same input. Therefore, the use of appropriate optimization methods is crucial in complex design scenarios [15]. Evolutionary algorithms, swarm intelligence, artificial neural networks, and other biologically inspired techniques are employed to assist architects in finding high-performance design solutions [16].

The visualization of designers' ideas is essential in the architectural design process. However, processes such as conceptualizing and creating design drawings are expensive and time consuming [17]. AI-based architectural design is widely applied in the field of "generative design", primarily to assist in floor plan creation, facade design, and rendering of effect drawings. For many years, researchers have been employing these algorithms to enhance architectural design work [18].

Oxman formalized the concept of "design interpretation" for defining, classifying, and designing expert system performance, facilitating mappings between design requirements and solutions, as well as between solutions and performance evaluations [19]. Zandi-Nia explored generative design at the topological level, emphasizing the need to clarify interactions between building locations based on design data and site conditions, utilizing clustering and hierarchical path selection to develop a hybrid integrated design system with neural networks [20]. Ali and Lee studied integrated bottom-up digital design, proposing a framework for early design exploration using deep convolutional generative adversarial networks to create intricate facade designs for urban infill [3]. Han and Shin developed a sketch-based effect drawing model utilizing the Pix2pix model [17].

Automated decision making in architectural design is primarily reflected in assisting design evaluation, material selection, and structural calculations. Through technologies such as natural language processing and visual monitoring, input and output data can be assessed, reducing human errors and allowing for the early evaluation of design results. AI algorithms help architects complete designs more quickly and accurately, improving design quality. Architects can utilize computer-aided design software to automatically generate design solutions that meet provided requirements. These solutions can then be analyzed and optimized using machine learning techniques to provide the best possible outcome. The multidisciplinary design optimization framework proposed by Lin et al. aims to obtain feedback on design decisions in the early stages [21,22]. In space planning, Currie and Tate studied the challenges in generating planning solutions and proposed an optimization scheme along with a demonstration system [23]. Wang et al. employed the Stable Stage Island Evolutionary Algorithm to generate diverse design outcomes, assisting architects in exploring different design directions [24].

In recent decades, a number of innovative firms have emerged, focusing on applying technologies developed in other industries to the construction sector, thus forming an ecosystem known as "ConTech." In the United States, investment in architectural technology has increased from approximately USD 250 million in 2013 to over USD 1 billion in 2018 [25].

Automation in the construction industry has become a common development goal for both academia and industry. The transition from traditional methods to intelligent ones has passed through several stages: expert systems in the 1970s and 1980s, computer-aided design in the 1990s, parametric design around 2000, and algorithm-driven models in the 2010s. This transformation can be attributed to improvements in computer hardware and enhanced computing power (Table 1).

Mehul Bhatt suggests that the next generation of products will consist of human-centered computational cognitive systems and aided design technologies. The method of applying "predictive analysis of designs that have not yet been created" in architectural design will address spatial problems from the perspective of formal modeling and computation, which is related to ontology and formal semantics [26]. During this transformation, a key issue that needs to be addressed is whether there are adequate infrastructure tools and platforms available to make architectural design more intelligent.

Emerging technology, centered on algorithms, requires development and deployment to realize its technical value, which is adopted due to its wide dissemination. Traditional

architectural design processes face efficiency and technological limitations, often seen as an “applied lag” industry, taking years to adopt mature technologies from other fields. To improve design efficiency and reduce costs for architectural design firms, it is necessary to develop functional requirements for an AI-based architectural design platform based on the analysis of impact mechanisms and adoption implementation, followed by the development of a prototype platform and practical testing with selected cases.

Table 1. Summarized historical evolution of AI in architectural design from literature.

Time Period	Event
1970s–1980s	Emergence of representative expert systems.
1980s	Warszawski and Sangrey proposed the combination of CAD-CAM technology with traditional architectural fields, providing a basis for future architectural systems [13].
1990s	Automated computational design methods became a focus for research and commercialization.
2000s	Rise of parametric design.
2010s	Algorithm-driven models became a mainstream focus for research and commercialization.
2013	Venture capital investment in US architectural technology was approximately USD 250 million annually.
2014	Lin et al. proposed a multidisciplinary design optimization (MDO) framework [21].
2016	Mehul Bhatt proposed that the next generation of products will be human-centered computational cognitive systems and aided design technologies [26].
2018	Venture capital investment in US architectural technology grew to over USD 1 billion annually [25].
2020	Challenges of automated architectural design [12]; Introduction of CAD-CAM technology into traditional architectural field [13]; Utilization of Stable Stage Island Evolutionary Algorithm (SSIEA) for diverse design optimization in architectural design [27].
2021	AI-based generative design widely used to assist in floor plan creation, facade design, and rendering of effect drawings [18]; development of a sketch-based effect drawing model using the Pix2pix model [17].

2.2. Identified Issues of AI in Architectural Design from Literature

Architectural design is a problem-solving process [28–30]. From the perspective of the architectural design industry, architecture, as a “delayed application” discipline, requires the support and empowerment of new technologies, especially emerging computer science and artificial intelligence algorithms. This demand also presents challenges to the speed of algorithm development and debugging efficiency, particularly for architectural design and its theoretical directions, in order to solve related problems encountered in practice. This demonstrates the rationale for applying artificial intelligence technology to architectural design.

Machine learning, belonging to the field of artificial intelligence, essentially seeks patterns in data and finds the optimal solution to a problem. It can be considered as exploring solutions within a relatively fixed range. When the optimization algorithm converges to a local minimum or a static point, it means that the optimal solution has been obtained.

The facade design of an architectural building collectively showcases its overall aesthetic style. However, traditional architectural design is a complex and time-consuming process that cannot quickly produce aesthetically expressive and creative works. In recent years, many scholars and technology developers have used machine learning techniques to assist various tasks in the architectural design process.

Building upon existing AI algorithms, this research makes three improvements: at the generator algorithm level, U-Net V2.0 is used, adding a processing layer between the encoder and decoder to better segment noisy images; during model training, a dataset of

newly built residential buildings in China is primarily used, the feature elements of labeled buildings are improved, and the iteration cycle and number of iterations are increased to enhance model accuracy; at the model application level, a multimodal approach is adopted, encapsulating the dataset and pre-trained model into a web application and integrating a front-end interface, allowing users to operate without any programming skills. The final research outcome is an AI-powered architectural design platform with a visual user interface, which can reduce the time and effort required in the architectural design process.

2.3. Synthesis of Literature and Identification of Research Gaps

The current body of research in AI-based architectural design (AIAD) demonstrates a clear trajectory from simplified generative adversarial networks to high-fidelity multimodal diffusion models. However, through a systematic synthesis of existing studies, several critical limitations emerge.

First, existing research often exhibits a “functional-aesthetic decoupling.” Studies focusing on spatial layout generation, such as those employing Pix2pix or House-GAN [31], excel in functional topology but frequently lack the capacity for high-quality visual representation. Conversely, recent advancements in latent diffusion models provide sophisticated aesthetic rendering but often struggle to maintain the rigorous structural logic required in professional architectural practice, leading to “architectural hallucinations” where generated forms lack functional viability. Second, there is a prominent “geographic and contextual void” in standardized datasets. The majority of open-source datasets utilized in previous literature are based on Western residential typologies, which differ significantly from the high-density, standardized residential modules prevalent in the Chinese urban context. The lack of a manually refined, localized dataset limits the practical applicability of AIAD models for Chinese architectural practitioners. Third, the integration of multimodal workflows remains fragmented. Most frameworks operate as isolated tools for specific design stages—either floor plan synthesis or facade rendering—without a unified pipeline that ensures logical consistency across the entire design evolution from 2D to 3D.

In contrast to previous studies, CorbuAI addresses these gaps through three primary innovations:

- (1) **Multimodal integration:** Unlike single-model approaches, our framework utilizes an optimized U-Net V2.0 and Stable Diffusion v2 pipeline to bridge the gap between functional layout logic and aesthetic expression.
- (2) **Context-specific dataset:** We contribute a custom-curated dataset of 2335 floor plans and 1570 facades specifically tailored to the Chinese residential context, ensuring the model’s outputs align with regional design standards.
- (3) **End-to-end workflow:** Our study introduces a seamless transition mechanism that converts schematic functional diagrams into professional-grade elevations and interior visualizations within a single, coherent framework.

3. Materials and Methods

3.1. Modeling for the AI Building Design Platform Based on Domain-Driven Design

In scenario analysis, a comprehensive user operation process is delineated, depicting a specific user’s actions at a particular time and place to fulfill a particular functional requirement. It serves as a key indicator often utilized to reflect and evaluate specific functionalities. During this step, it is imperative to consider typical scenarios within the domain model of the AI architectural design platform (Table 2 and Figure 3). Matters such as target population, permission assignment, user events, business flow, and usage commands require exploration, thereby establishing the groundwork for the subsequent domain modeling phase. The following figure illustrates the state diagram of scenario

analysis, portraying the activities undertaken by the user during software usage, and delineating the initiation and conclusion of the flow from the user's initial interaction with the first page to the final module.

Table 2. Development, initial logical process flow diagram.

Development Phase	Main Tasks/Objectives
Modeling Based on Domain-Driven Design	Identifying requirements and core concepts
Technical Architecture and Key Technology Analysis	Selecting the tech stack and building the architecture
Multimodal Model Training and Fine-Tuning	Data processing and model training
Floor Plan Generation	Using Pix2pix-GAN to generate building floor plans
Elevation Drawing Generation	Extending Pix2pix-GAN to generate building elevation drawings
Interior Style Transfer	Applying diffusion models for style transfer
UI Integration and Interaction	Providing user interface support and visualizing generated results

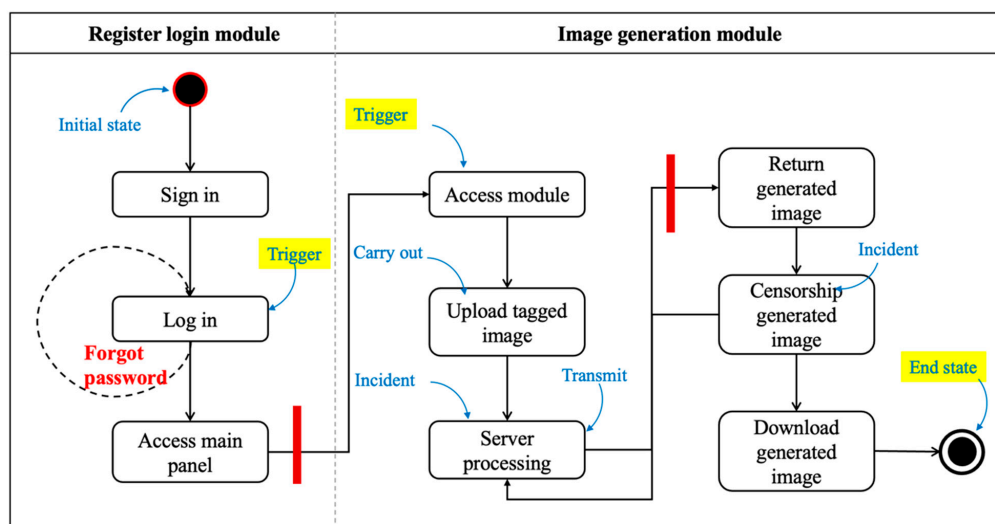


Figure 3. State diagram of scenario analysis.

The fundamental business service unit and the analysis of users' activities within each execution segment can be derived through the analysis of business scenarios and the screening of users' requirements as the criteria for evaluation. The basic business service unit serves as the foundational element for mapping relationships. The aforementioned core functions entail distinct initiation, operation, and conclusion phases, along with feedback on related outcomes. Therefore, it is imperative to construct an interactive interface for the platform. This interface aids users in comprehending the content and embedded business rules, enabling interactive operations throughout the design process to accomplish the functions outlined in the explanatory model. The primary functions of the platform are described in Table 3 as below.

Table 3. The main functions of the minimum viable platform.

Name	Introduction
User management	Login and registration of user
Upload files	Upload an image file in the input box, with the picture formats jpg or png
Generate results	Load the generated results for the user to view
Save and download	Click the button to download the generated images
Log out	The user will be able to log out from the system

Based on the explanatory model, it is imperative to separate domain objects from other functions within the system. The layered architecture pattern is commonly employed for this purpose, wherein each layer represents a distinct function associated with specific roles (Figure 4). Information is passed between layers in a predefined manner. The layered architecture diagram for this study is depicted in the following figure.

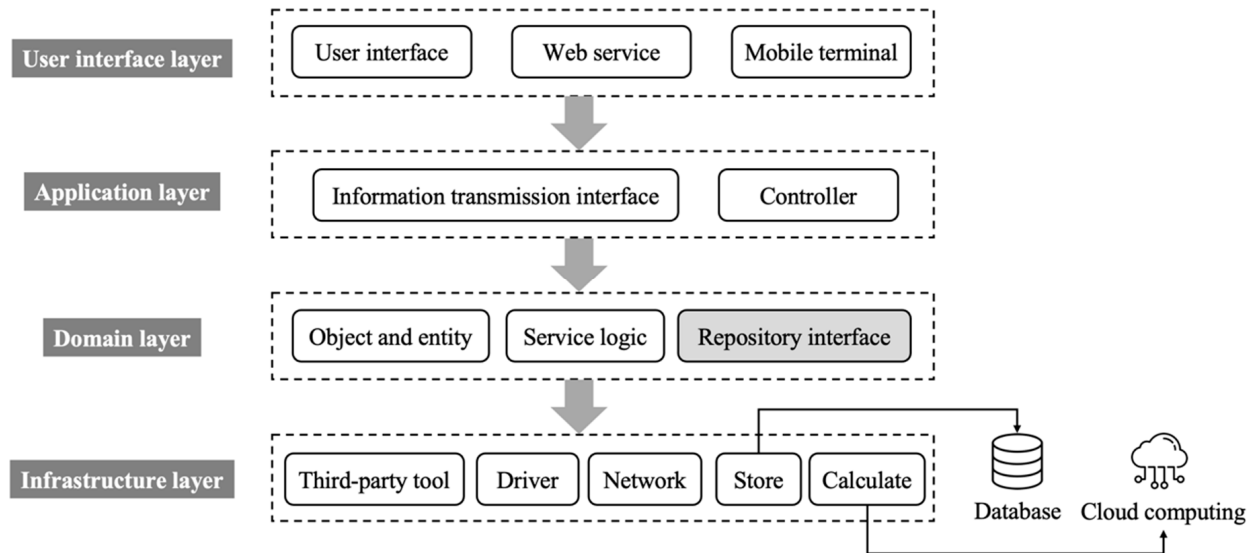


Figure 4. Layered architecture design.

The AI architectural design platform is segmented into four layers following the layered architecture model: the user interface layer, application layer, domain layer, and infrastructure layer. The user interface layer, directly interacting with users, facilitates the acquisition of user commands and visually displays the execution results of procedures. The application layer facilitates application interfaces and message transmission, enabling information exchange between modules and coordinating the synergy of multiple domain layers. Positioned as an intermediate tier, the application layer hosts the system's main functions, such as the controller and specific execution code. The domain layer, at the core of the domain model, implements core business logic without impacting other business models during the implementation of specific technologies. It encompasses resource libraries, value objects, and entities that isolate core service logic from external applications and basic designs. The infrastructure layer acts as the support layer for the entire system, encompassing hardware facilities for storage, computing, and networking, as well as third-party tools and drivers. This technology-intensive segment provides technical support for all its sublayers, all of which rely on the infrastructure layer for normal operation.

Aligned with the logical flow of software usage, the user interface layers correspond to various user actions, including user registration and login, and uploading and generation of graphic files. Due to the utilization of third-party libraries and open-source frameworks, this study and subsequent chapters focus solely on the analysis of core functions and algorithms, excluding processes involving auxiliary or secondary functions.

3.2. Analysis of the Technology Stack and Key Technologies of the AI Architectural Design Platform

The development of an AI architectural design platform is inherently complex, primarily due to the intricacies of the architectural design domain to which the software is applied. Therefore, an efficient domain model, capable of rational analysis, control, and optimization of these complex problems, is essential. According to Eric Evans, "a truly robust domain model evolves over time" [32], illustrating the importance of the iterative process. Itera-

tion enhances the efficiency of the design and development team by addressing product requirements in phases, from software product planning to release.

Technology stack refers to the combination of multiple technologies as an organic whole that helps achieve a specific purpose or function. The following Table 4 is the technology stack used in the core module of the minimum viable product of the architectural design platform based on AI technologies.

Table 4. Technology stack of the core module for the minimum viable product.

Programming language	Python 3.9: Python is a programming language commonly used for machine learning and AI applications. Python has a diverse ecosystem of libraries and frameworks that are suitable for the construction of complex software systems.
	PHP 7.1: PHP is a server-end scripting language used to develop static websites or dynamic websites or web applications.
Frameworks and libraries	TensorFlow2.11.0: TensorFlow is an open-source machine learning framework developed by Google. It is widely used to build and train neural networks and can be used to implement AI algorithms in AI software.
	Keras. Keras, as an advanced neural network API, is operated based on TensorFlow. It provides a simple and user-friendly interface for building and training neural networks, making it ideal for rapid prototyping.
	U-Net: U-Net is a convolutional network for image segmentation proposed by Ronneberger in 2015 [33].
	Pix2pix: Pix2pix is a generative adversarial network model for image-to-image transformation, developed by Isola [34].
	Stable Diffusion v1.5: Stable Diffusion is a variant of the diffusion model, and image transformation algorithm model developed by the Machine Vision Learning Group (CompVis) at the University of Munich in collaboration with the startups StabilityAI and Runway.
Database	Relational Database Management System (RDBMS): Cloud SQL
Other tools	Git: Git is a version control system where users manage and use different versions of code.
Operating environments	CentOS 6.0: A system environment for deployment of execution code of minimum viable products

The back end, or server end, represents the data and business logic part of the management system and is responsible for storing and retrieving data, applying business rules, and providing data to the front end. The back end is usually implemented using server-end technologies such as PHP, Java, Python, etc. These technologies typically operate on a server that is accessible over the network by the front end. The server-end code or back-end code usually assists in the interaction and management of data [35].

As mentioned earlier, the core idea of a distributed database is to provide database services via multiple distributed multi-copy resources as a whole, thus improving the performance and disaster tolerance of data servers. Mainstream cloud computing companies provide cloud-native databases as data infrastructure. In this study, GCP's Cloud SQL product is selected to enhance the elastic scaling capability of distributed databases. To achieve global data management, a central data repository needs to be deployed for the distributed database, and data information can be shared with servers of replica resources. When information storage, retrieval, data reading, and writing are performed, the central data repository and replica resources have a synchronization mechanism to maintain data consistency.

4. Multimodal Model Training and Fine-Tuning

The CorbuAI platform serves as an auxiliary architectural design tool for architects and designers. Users can input specific images or design parameters, resulting in the generation of multiple design images that meet the requirements. These generated images can be reviewed and saved. The platform primarily aims to streamline the architectural design process, improve efficiency and effectiveness, generate architectural design images more efficiently, and save design time and effort, ultimately aligning with corporate goals and business strategies. Auxiliary design decision making is an interactive process wherein AI technologies are integrated to enhance creation efficiency, quickly obtain a large number of generated images meeting requirements, and assist architects and designers in maintaining critical control over creativity and design quality.

The section on Multimodal Model Training and Fine-tuning will be divided into three parts. Part 1 involves multimodal AI model training and debugging, which includes Pix2pix-GAN-based building floor plan generation, Pix2pix-GAN-based building facade representation generation, and SDv2-based building interior space style migration. This part encompasses data collection and pre-processing, fine-tuning of pre-trained model parameters, evaluation of model training results, and assessment of generated images. Part 2 covers user-end interface design, as well as the interaction architecture and process involving cloud server, platform server, and user end. Part 3 focuses on the platform's function during the actual project application stage, including project background introduction and evaluation of the generated images.

Figure 5 illustrates the architecture of the platform's design-assisted process, wherein the building floor plan is a key outcome influencing facade representation and interior space. Therefore, architects and designers need to establish the interior footprint area before generating the floor plan. Subsequently, they manually deduce the facade and build a low-precision model of the interior space, enabling the algorithm to better identify input images of facades and interior spaces.

After analyzing the platform requirements and architecture and proposing a reliable technology stack and architecture, it is necessary to elaborate on the R&D process of the core functions and select a real case to verify its usability. In this part, an "architectural design platform based on multimodal AI technology" will be established, comprising three parts: building floor plan generation, building facade representation generation, and building interior space style migration. The research process for each part follows three steps: data collection and pre-processing, fine-tuning of the pre-trained model, and evaluation of the model training results. Finally, evaluation of the generated images is carried out from three aspects: subjective evaluation, objective evaluation, and actual case evaluation.

Subjective evaluation involves inviting architects and designers for visual inspection. The model's predictions are visualized and compared with the actual building facade, comprehensively evaluating the generated images in terms of image quality, facade details, and other factors. Objective assessment can quantify the quality of the generated images to some extent, complementing the subjective assessment, and assisting in design decisions. In the stage of actual case application and evaluation, real projects are selected to generate building floor plans, building facade representations, and building interior space styles. Additionally, evaluation and feedback of the design results are collected from the commissioner under specific conditions.

4.1. Pix2pix-GAN-Based Building Floor Plan Generation

4.1.1. Data Collection and Pre-Processing

As the input content of the model, the dataset is a crucial component of AI (machine learning). Its importance lies in the success of machine learning. A dataset represents a

collection of labeled data used for model training. These training images can be utilized for various tasks such as research on computer vision and machine learning, architectural and urban planning, and application of virtual and augmented reality technologies. The data in a dataset usually share the same or similar properties. In a sense, it can be called a “database with certain characteristics”. In this study, the database includes various building types, styles, and locations, providing detailed annotated information such as floor plans, materials, and dimensions. The steps are shown in Table 5 and Figure 6.

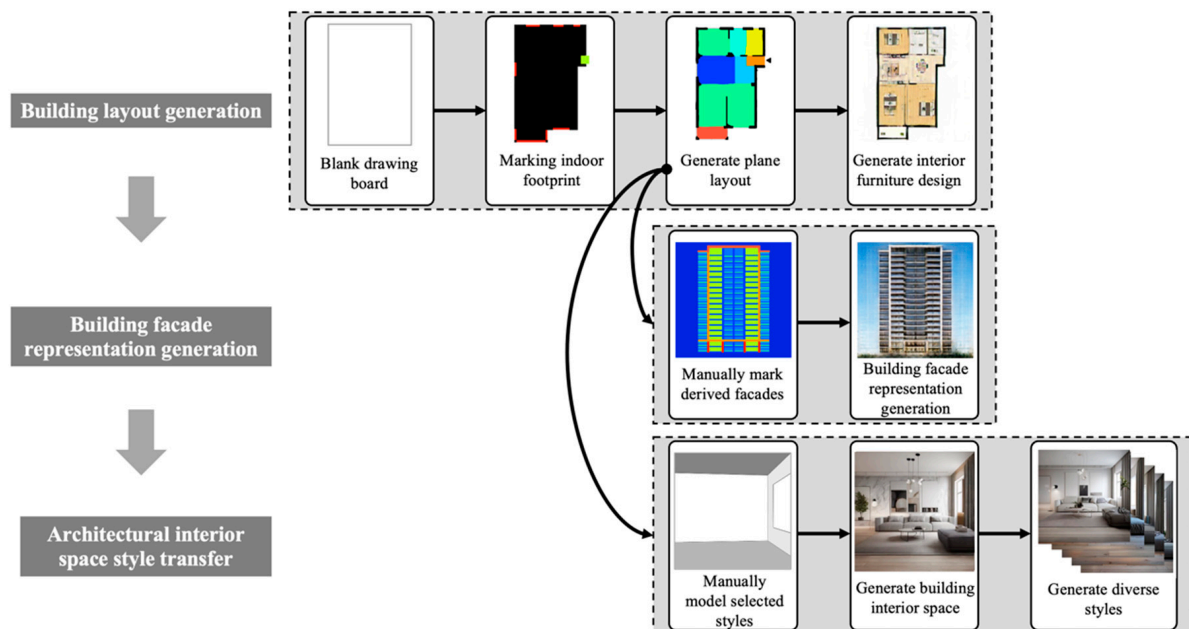


Figure 5. Flow of design-assisted architectural design platform based on multimodal AI technology.

In previous studies, some datasets related to floor plans of residential buildings were obtained from the Rakuten Institute of Technology (RIT) in Japan [36] and the Computer Vision Center (CVC) in Spain [37]. However, most of these datasets consist of floor plans of typical buildings in foreign countries, which are not applicable to this study. Therefore, it is necessary to create a floor plan dataset mainly consisting of typical residential buildings in China (Figure 7).

Table 5. Research process flowchart.

Phase	Description
Data Collection and Preprocessing	Collecting floor plan data of residential buildings and preprocessing it.
Image Labeling and Similarity Calculation	Upload an image file in the input box, with the picture formats jpg or png.
Style Transfer and Generative Models	Style transfer in building interior spaces; training results for different styles (Modern, New Chinese, Industrial).
Practical Project Application Phase	Detailed explanation of the application process in practical projects.
Conclusion and Evaluation	Objective evaluation of the generated results and measurement of the rationality.

In this study, partnerships have been established with several real estate websites and housing agency websites. With the authorization of third parties, the right to use over 10,000 residential building floor plans in China for academic research has been obtained. Based on this, a dataset containing 2335 manually corrected images of typical floor plans of residential buildings in China has been produced after screening high-resolution images.

These images have been classified into Training Set, Validation Set, and Test Set in the ratio of 60% (1401), 20% (467), and 20% (467), respectively, for the development, training, and evaluation of the Pix2pix-GAN model (Figure 8).

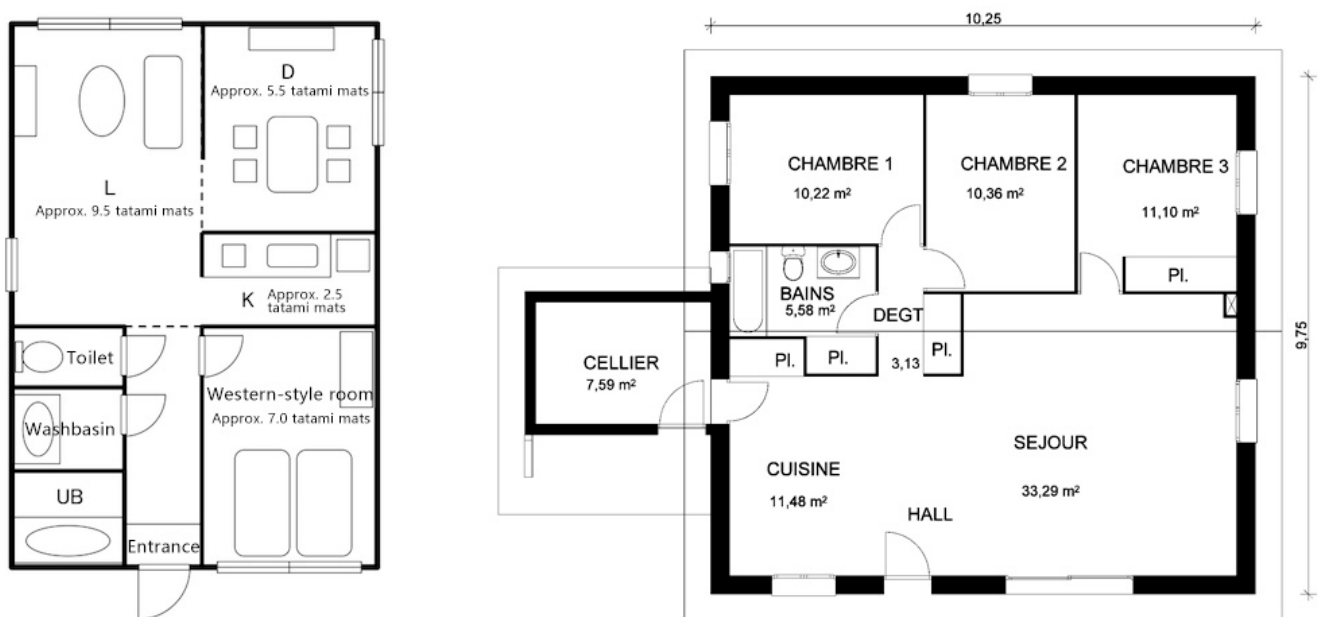


Figure 6. Floor plan dataset from Rakuten Institute of Technology [36] and Computer Vision Center [37].

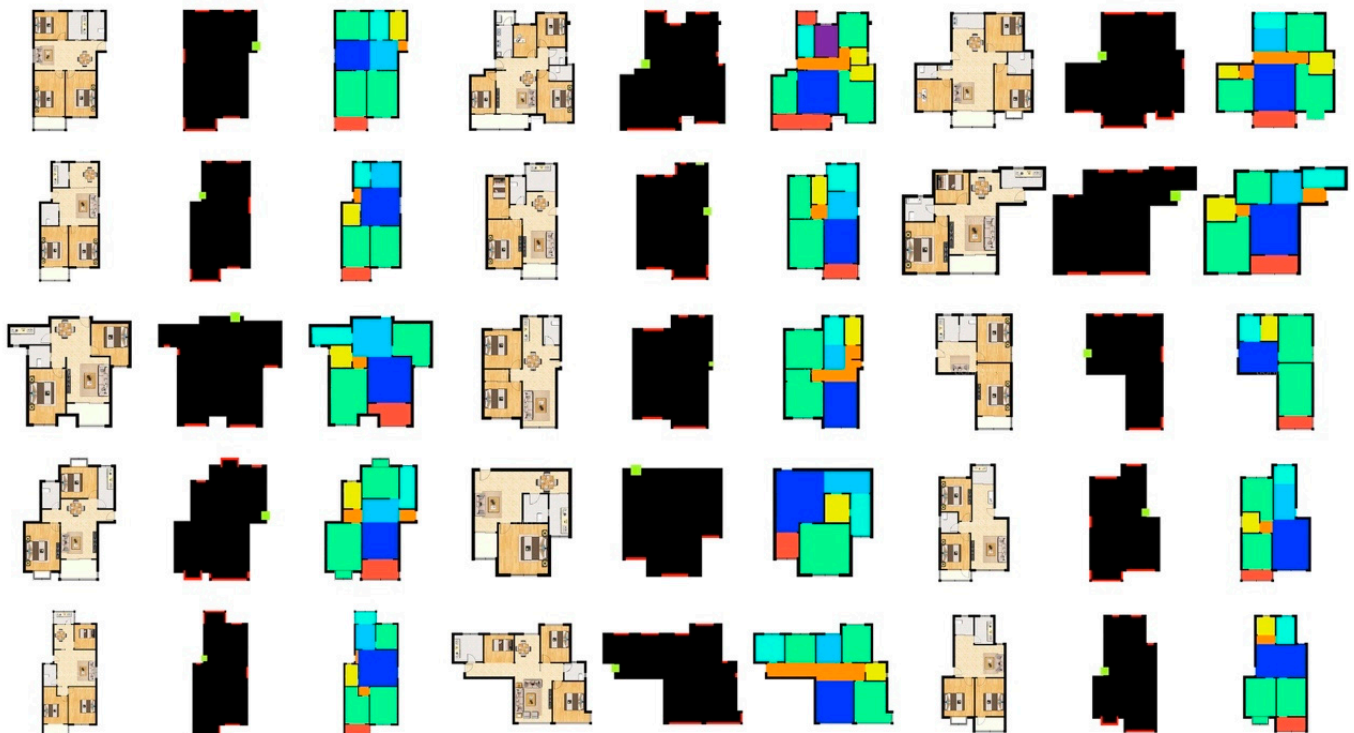


Figure 7. Floor plan dataset of residential buildings in major cities in China.

The dataset includes a variety of residential building floor plans, ranging from one-bedroom houses to three-bedroom houses. To facilitate content recognition and learning algorithms, as well as to ensure the generality of the training model results, a filtering rule for image content has been established in the study: plans with clear, regular geometric lines and wall contour lines are retained, while those with irregular or non-standard geo-

metric shapes are excluded. Each image case in the dataset is marked with different color blocks to represent various component elements, thereby reflecting the location information and image attributes of the components in the floor plan. This method builds upon the work of Zheng, Wan, Zhao, Rodrigues, and Chaillou [38–42] to better suit the characteristics of floor plans of typical residential buildings in China. The rules and indexing diagrams of the labeled images are in Table 6 as follows.

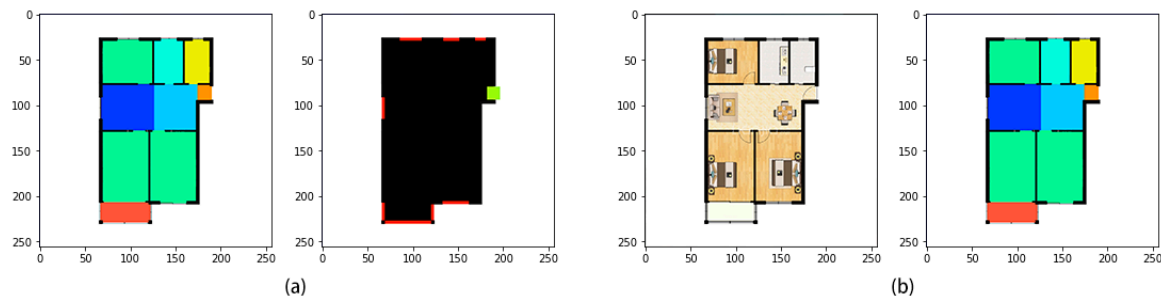


Figure 8. Data annotation of training set. (a) Training stage. (b) Validation stage.

Table 6. Rules for marking (annotation) images of dataset for residential building floor plan.

Example of Color Block	Number	Floor Plan Element	Color Name	RGB Value	HEX16 Color Code
	1	Indoor footprint	Black	(0, 0, 0)	000000
	2	Living room	Celestine blue	(28, 60, 241)	1C3CF1
	3	Public area *	Dark gray	(156, 156, 156)	9C9C9C
	4	Bedroom	Gray green	(143, 245, 160)	8FF4A0
	5	Kitchen	Dark blue	(118, 249, 222)	78F7DE
	6	Bathroom	Mimosa yellow	(238, 238, 87)	EDEE56
	7	Entrance door	Tender green	(198, 248, 96)	C5F85E
	8	Corridor	Sandy brown	(240, 151, 56)	F19838
	9	Balcony	Soda orange	(235, 89, 61)	EB593D
	10	Window	Madder red	(227, 62, 36)	E33E24
	11	Dining room	Bright sky blue	(92, 198, 250)	5BC5FB
	12	Study room	Purple	(99, 38, 151)	632697
	13	Outdoor equipment **	Dark rock blue	(57, 83, 135)	395387

Notes: * Public area refers to the parts including elevator rooms, staircases, equipment facilities and equipment tube wells, and public corridors, which are located outside of the rooms on the floor plan. ** Outdoor equipment refers to the elements that are located outside the rooms, such as chillers, new ventilation units, and outdoor units of air conditioners, in the floor plan.

After the original image is labeled, it is necessary to apply random jitter and mirroring for image pre-processing operations to avoid the phenomenon of overfitting. Firstly, the image is cropped into a square image of size 256×256 and resized to 286×286 . Then

random cropping is applied to each image, which is then resized back to 256×256 . The image is randomly flipped horizontally to generate the mirrored image, and finally, the image is normalized into the range of $[-1, 1]$, as shown in Figure 9.

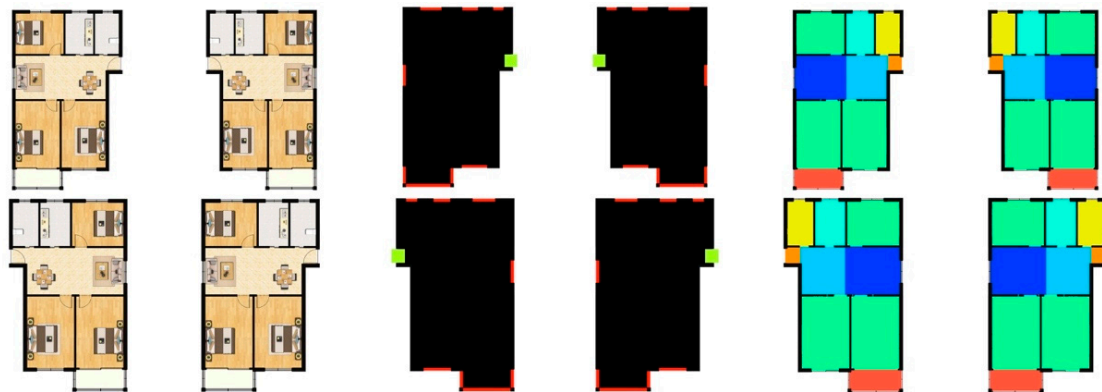


Figure 9. Random jitter and mirroring process.

The original image is combined with the labeled image, and each combined image has a size of 256×512 , containing two square images of size 256×256 each. The combination results are shown in Figure 10 as follows.

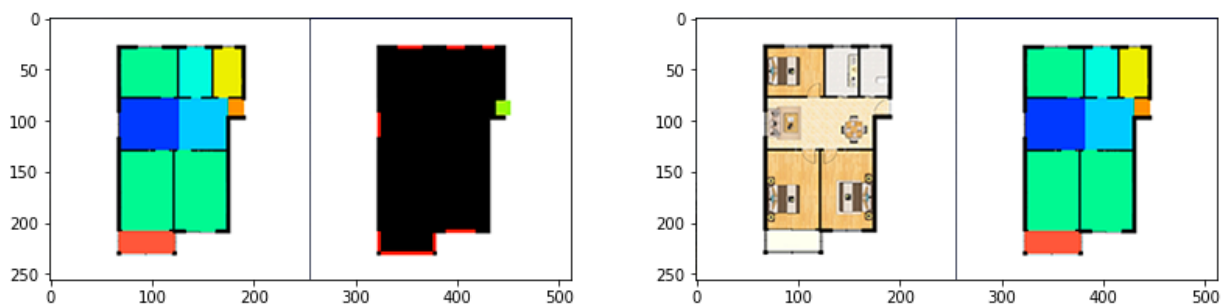


Figure 10. Completion of data pre-processing image.

4.1.2. Fine-Tuning of Pre-Training Model Parameters

In the Pix2pix conditional generative adversarial network (cGAN), the output image is generated based on the input image, which serves as a precondition for the generation process [34]. Considering the convenience of cloud computing, TensorFlow is selected as the training platform for the model in this study. TensorFlow is used to achieve Pix2pix. The generator and discriminator networks are defined and trained using TensorFlow layers and optimizers. The generator creates new images, while the discriminator verifies the authenticity of the generated images. The parameters of the generator are tuned during training to enhance the authenticity of the generated images.

Pix2pix-GAN is a conditional generative adversarial network that generates corresponding output images based on input images. This architecture was first introduced by Mirza and Osindero in their 2014 paper “Conditional Generative Adversarial Networks” [43]. The network architecture includes two main components: a generator and a discriminator. The discriminator model is trained directly on real and generated images, while the generator model is trained by the discriminator model [44]. In this study, the U-Net V2.0 architecture is adopted as the generator. U-Net preserves details and spatial resolution when generating output images, and its V2.0 version improves image segmentation. The method involves programming 64 channels of the input image by convolution twice. Through multiple steps, each of which halves the standardized image size, the images are combined stepwise. Both U-Net generators, Functional Segmentation

Layout (FSL), and Building Floor Plan (BFP) contain a set of paired floor plan labeled images [38]. The discriminator is represented by the convolutional PatchGAN classifier, a classifier proposed in the Pix2pix paper [34], used to distinguish real images from generated images. The following figure shows the flowchart of the whole model training and image generation process. The dashed boxes indicate the gaming process of the generator and the discriminator in Figure 11.

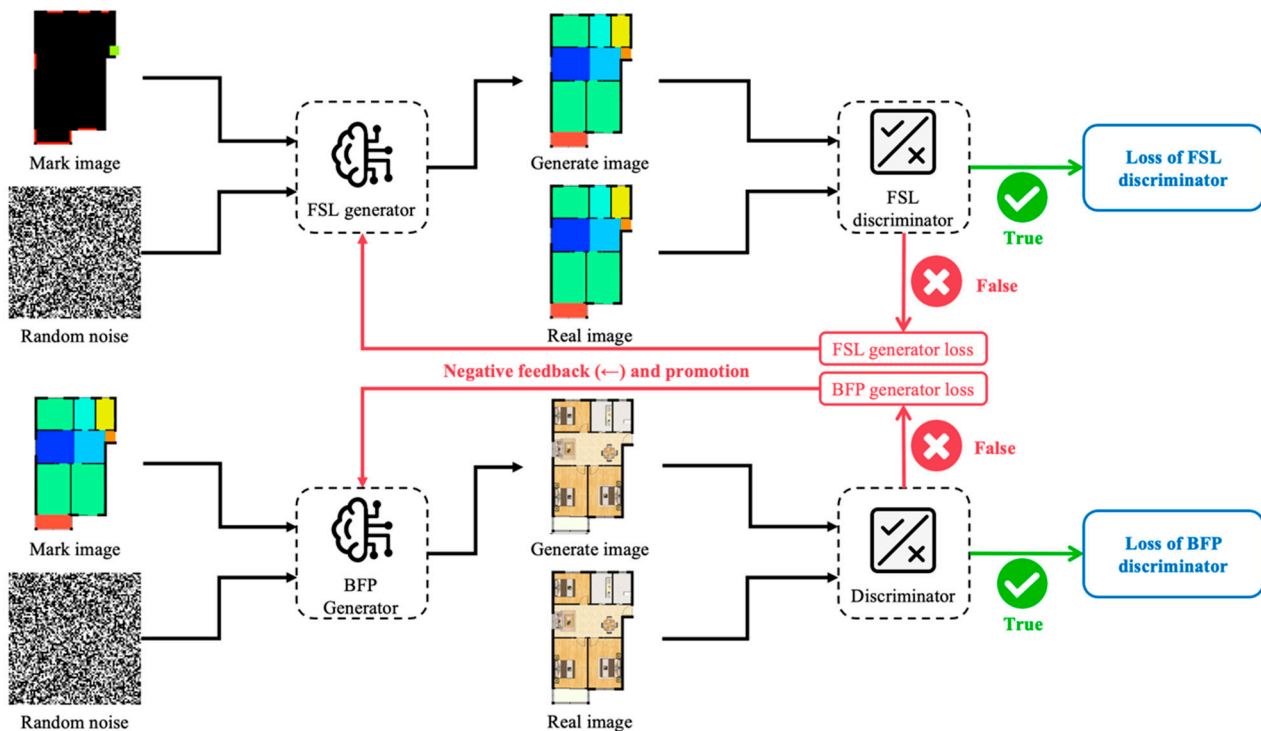


Figure 11. Gaming process of generator and discriminator.

The figure below depicts the generator based on U-Net V2.0 [33] and the discriminator represented by the convolutional PatchGAN classifier [34] in Figures 12 and 13.

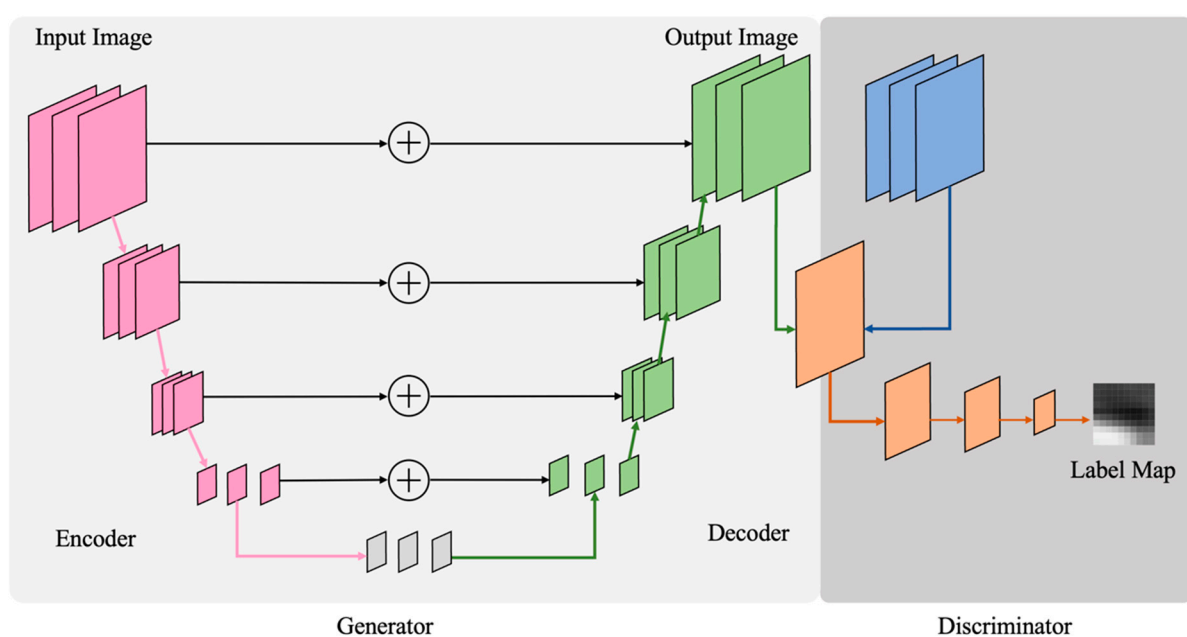


Figure 12. Generator and discriminator architectures.

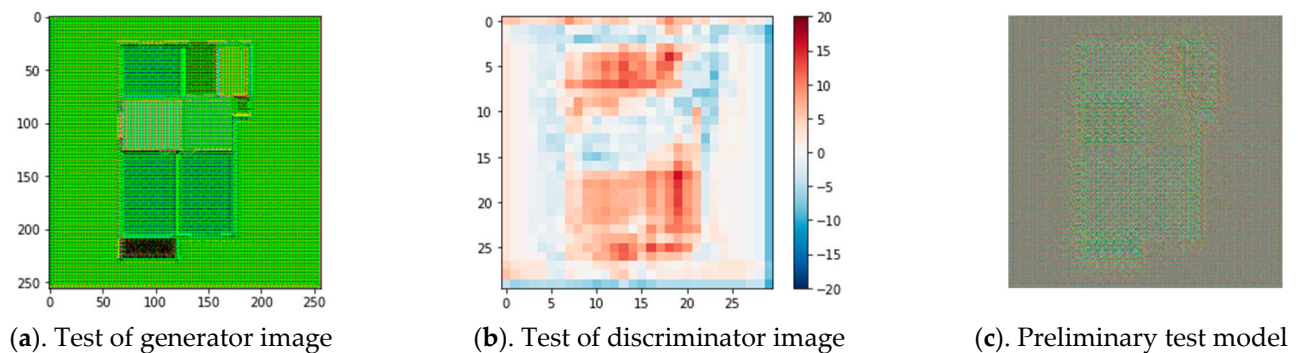


Figure 13. Test of the generator and discriminator.

When training the model, it is essential to calculate the losses of both the generator and discriminator in TensorBoard, presenting them graphically. Following this, the model debugging phase commences, and then, debugging the AI model involves training it on a new dataset while utilizing the previously trained weights as a starting point. This process is intricate, requiring continual adjustment of generator and discriminator parameters to generate image results with minimal loss values. When the new dataset resembles but differs slightly from the original dataset used for training the model, real-time insights into these discrepancies can be gained during debugging. In this phase, the model's learning progress (i.e., the checkpoints saved every 5000 steps in the previous section) can be adjusted based on the new datasets. Consequently, the model does not need to be trained from scratch, leading to improved performance and shorter training times. For smaller new datasets, fine-tuning can mitigate the overfitting problem. Pre-trained models have already acquired general features from large datasets, and fine-tuning on smaller datasets can further enhance performance.

4.1.3. Evaluation of the Generated Images

After the Pix2pix-GAN is trained, the quality of the generated images needs to be judged based on the corresponding evaluation criteria, which include the realism, diversity, confrontation and robustness, and user's evaluation. Realism is the degree to which the generated images match the target images under this method. There are different variations of the target images generated by the evaluation model under this method. It can be measured by generating multiple images from the same input and calculating the similarity between them, which causes the diversity. The "confrontation and robustness" indicator is used to assess how well the model generates images that resist various types of image processing operations. It is measured by the training of networks that confront each other to distinguish the generated images from the target ones, shown in Figure 14. The user's evaluation is used to assess the subjective quality of the generated images and it is measured by the conduction of a user study. In a user study, the evaluator evaluates the generated images based on factors such as realism, aesthetic appeal, and semantic consistency.

In this section, a subjective quantitative scoring method and cross-validation method are used to evaluate the effectiveness of the generated images. According to the method proposed by Wan et al. [38], the generated results of the Functional Segmentation Layout (FSL) model are scored according to spatial allocation clarity, functional distribution rationality, and color block boundary clarity. The generated results of the building floor plan (BFP) model are scored based on wall generation accuracy and furniture generation accuracy. A score from 0 to 5 indicates the acceptance level from low to high, respectively. In addition, the floor plan datasets from the Rakuten Institute of Technology and the Computer Vision Center (Figure 8) are used as the images for cross validation. The data annotation needs to be created according to image tagging rules before the model is imported, shown in Table 7.



Figure 14. Comparison of generated floor plan images at different progressions.

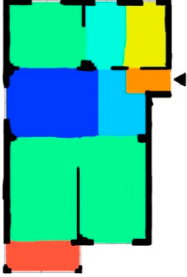
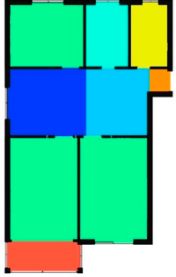
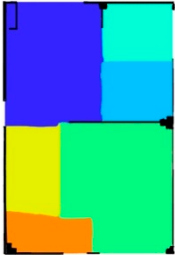
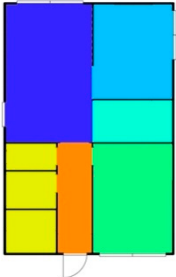
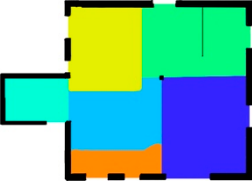
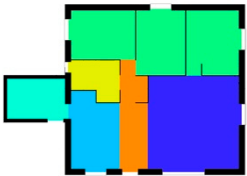
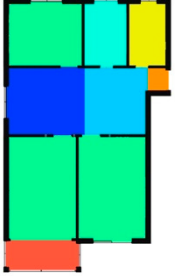
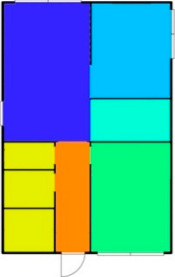
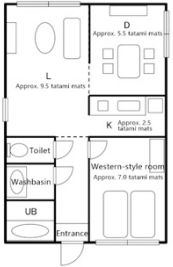
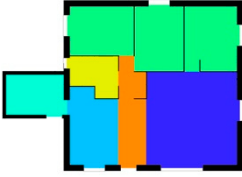
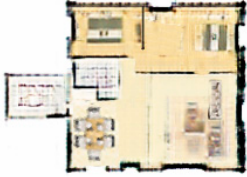
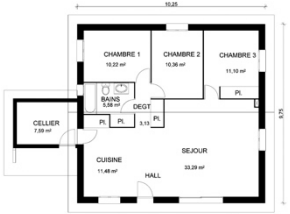
It can be seen from the evaluation results of the cross-validation generated images that the functional segmentation layout (FSL) model and the building floor plan (BFP) model can generate excellent floor plan images, and the difference between the output image and the real image is within the acceptable range. However, the image clarity is related to the spatial distribution of the input image and the location of the building elements, and the color block boundary and furniture accuracy of some results are not clear enough. When the actual project inspection is conducted, the arithmetic power and the accuracy of the input images should be improved so as to obtain image results that can meet the requirements (Table 8). Google cloud platform (GCP) allows the expansion and management of the deployed models, as well as provides, stores, and monitors other integration services. In this platform, models can be deployed via the use of services such as Cloud Machine Learning Engine, Cloud Functions, Kubernetes Engine, or App Engine. Also, considering the cost and elasticity of the chosen deployment method, the Google Cloud ML Engine of GCP is selected in this study to deploy the pre-trained Pix2pix-GAN model. Meanwhile, TensorFlow is used for model training and TensorFlow-Hub is used to manage the model.

4.2. Generation of Pix2pix-GAN-Based Building Facade Representation

4.2.1. Data Collection and Pre-Processing

Based on a dataset of 606 manually annotated and corrected facade images produced by a team from the Center for Machine Perception (CMP) at the Technical University of Prague, Czech Republic [45], Isola's team at UC Berkeley proposed a Pix2pix-GAN for image-to-image transformation via the use of conditional adversarial networks (GAN) [34]. However, these datasets only involve building facade images of major European cities, and are not in line with the research target and research purpose of this chapter, so it is necessary to produce a dataset of facade images mainly for residential buildings in major Chinese cities. This study has been conducted under the cooperation with several non-commercial architectural database websites and the rights to use over 5000 Chinese residential building facade images have been obtained for academic research with the authorization of the third party (Figure 15).

Table 7. Cross-validation generated image results.

Model	Input Image	Output Image	Real Image
Functional segmentation layout (FSL) model			
			
			
Building floor plan (BFP) model			
			
			

These images include not only actual photos of the completed projects, but also residential building projects that are not actually built, like conceptual schemes, intentional drawings, renderings, and bid proposals. On this basis, datasets containing 1570 manually corrected facade images of new residential buildings in China were finally produced after the high-resolution images were screened. They were allocated into training datasets (60%), validation datasets (20%), and test datasets (20%), which respectively include 942, 314, and 314 facade images, for the development and testing of Pix2pix-GAN model training and evaluation (Figure 16).

The datasets involve a diverse collection of architectural styles of new residential building facades in different Chinese cities, such as modern European style, new Asian style, and new Chinese style. Also, there are various types of residential buildings, including multi-story building, small high-rise building, and foreign-style house. Considering the generality of the training model results, to facilitate the algorithm for content recognition and learning, a filtering rule for the image content is set up in the research: the facade design images with clear and regular geometric compositions are retained, and some of the facade design images with alien or non-standard geometric shapes are excluded. The image cases in the dataset are marked with different color blocks representing different component elements to reflect the location information and image attributes of the components in the interior layout. The adopted method was improved on the basis of the study by Isola et al. [34] to suit the characteristics of typical new residential buildings in China, and the index plan of the labeled images is shown in Table 9.

Based on the above method, the input images are randomly jitterred and mirrored, and the paired data containing two images of 256×256 size are merged (Figures 17 and 18).

Table 8. Evaluation of generated images of the model.

Model	Scoring Items	Score			Average Score
Functional segmentation layout (FSL) model	Space allocation clarity	4	3	3	3.3
	Functional distribution rationality	3	4	4	3.7
	Color block boundary clarity	4	2	2	2.7
Building floor plan (BFP) model	Wall generation accuracy	3	3	2	2.7
	Furniture generation accuracy	3	2	2	2.3



Figure 15. Dataset for facades of new residential buildings in major Chinese cities.

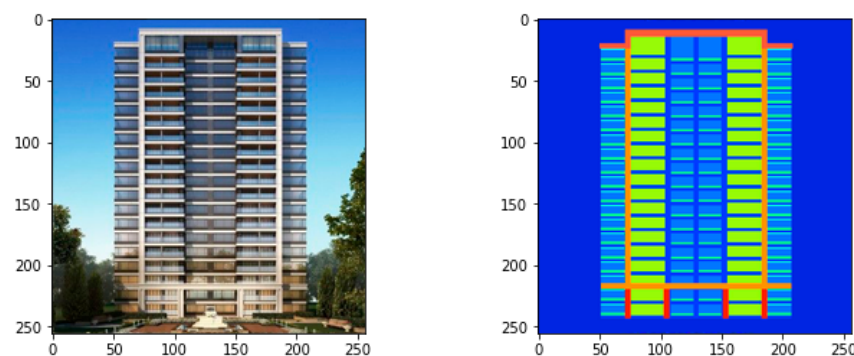


Figure 16. Training set data annotation.

Table 9. Image marking rules for the facade dataset of new residential buildings.

Example of Color Block	Number	Component	Color Name	RGB Value	HEX16 Color Code
	1	Background	Medium Blue	(15, 41, 209)	0F29D0
	2	Wall	Celestine blue	(28, 60, 241)	1C3CF1
	3	Window	Dodger blue	(50, 115, 246)	3273F6
	4	Window sill	Gray green	(143, 245, 160)	8FF4A0
	5	Sunblind	Dark blue	(118, 249, 222)	78F7DE
	6	Window blind	Mimosa yellow	(238, 238, 87)	EDEE56
	7	Balcony	Tender green	(198, 248, 96)	C5F85E
	8	Decoration	Sand brown	(240, 151, 56)	F19838
	9	Roof	Soda orange	(235, 89, 61)	EB593D
	10	Pillar	Madder red	(227, 62, 36)	E33E24
	11	Entrance	Bright sky blue	(92, 198, 250)	5BC5FB

4.2.2. Fine-Tuning of Pre-Training Model Parameters

The model architecture is proposed based on the above method, a typical Pix2pix, which contains a generator and a discriminator. The following figure shows the flowchart of the model training and image generation process in the facade generation phase. Here, the dashed boxes indicate the gaming process of the generator and discriminator (Figures 19 and 20).

The result shown in step 1 for the preliminary operation demonstrates that the images can be successfully passed from the test set to the generator and that the input image can be successfully converted to the output image via the generator (Figure 21).

After that, the number of training iterations needs to be set. A complete iteration process is called an Epoch. There are 942 images for the training datasets in the research, signifying that one iteration needs 942 training steps. In this study, the number of training steps is set to be 80,000, and a checkpoint is saved for every 5000 steps for the convenience of subsequent fine-tuning operations. After the test training loop code is operated, it can be

seen that the average time required per 1000 steps if the GPU and TPU are not used for computation is 2158.76 s (Figure 22).

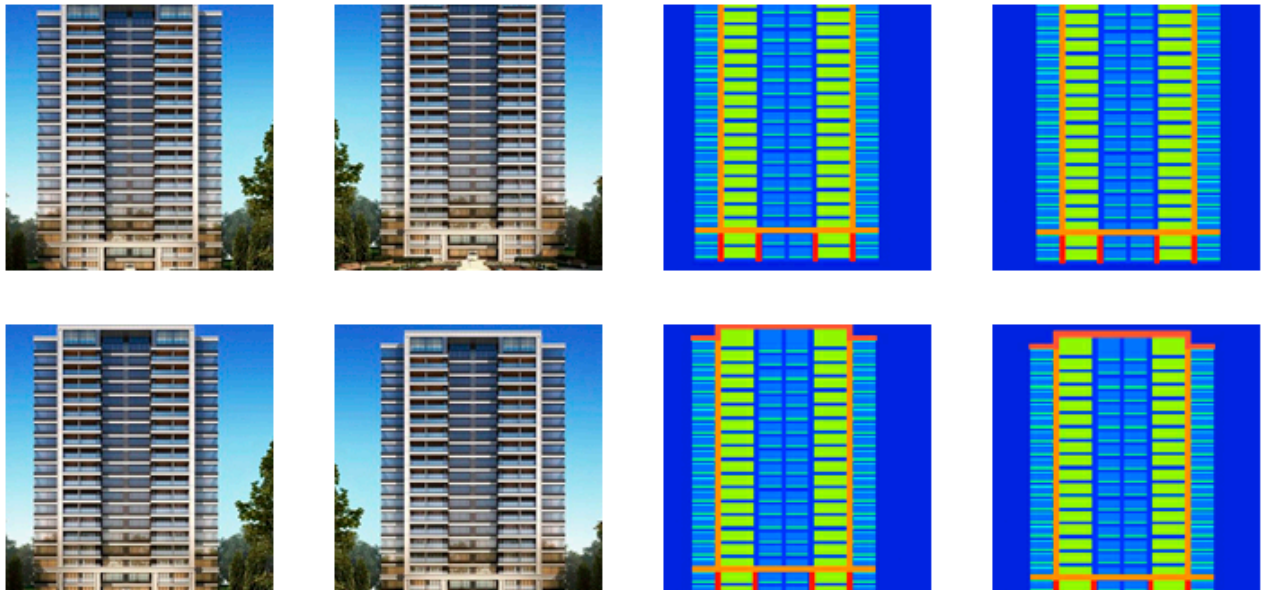


Figure 17. Random jittering and mirroring process.

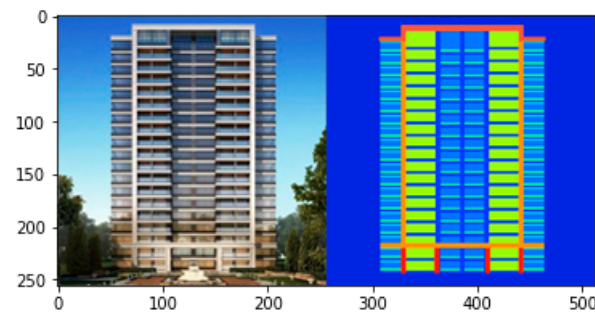


Figure 18. Completion of image of data pre-processing.

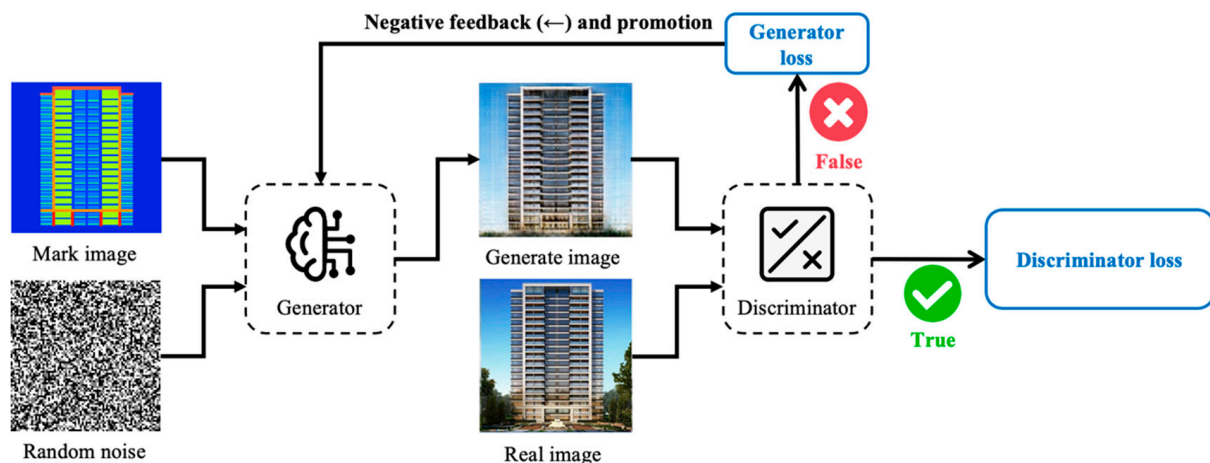


Figure 19. Gaming process of the generator and discriminator.

4.2.3. Model Training Results and Evaluation of Generated Images

Since the training logs of the model visually reflect the quality of generated images, and the model prediction accuracy can be visually measured by quantified numbers and computational metrics. Multiple loss log plots need to be considered to comprehensively evaluate the training effect of the model, shown in Figure 23.

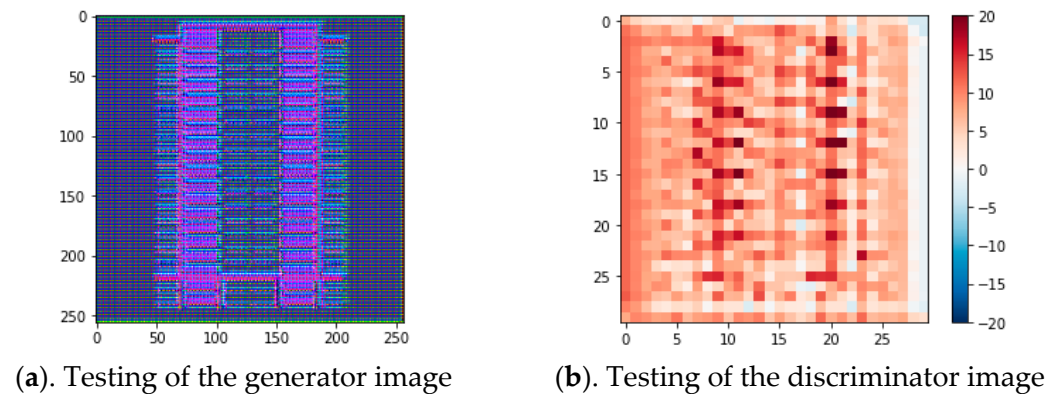


Figure 20. Testing of the generator and discriminator.

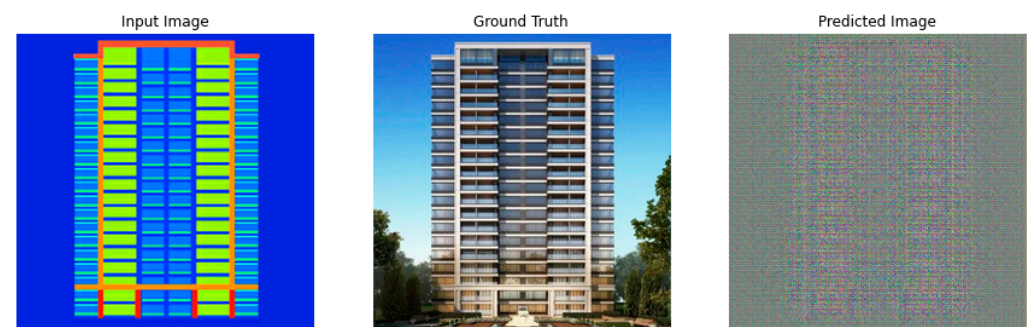


Figure 21. Testing of the function of the generated images.

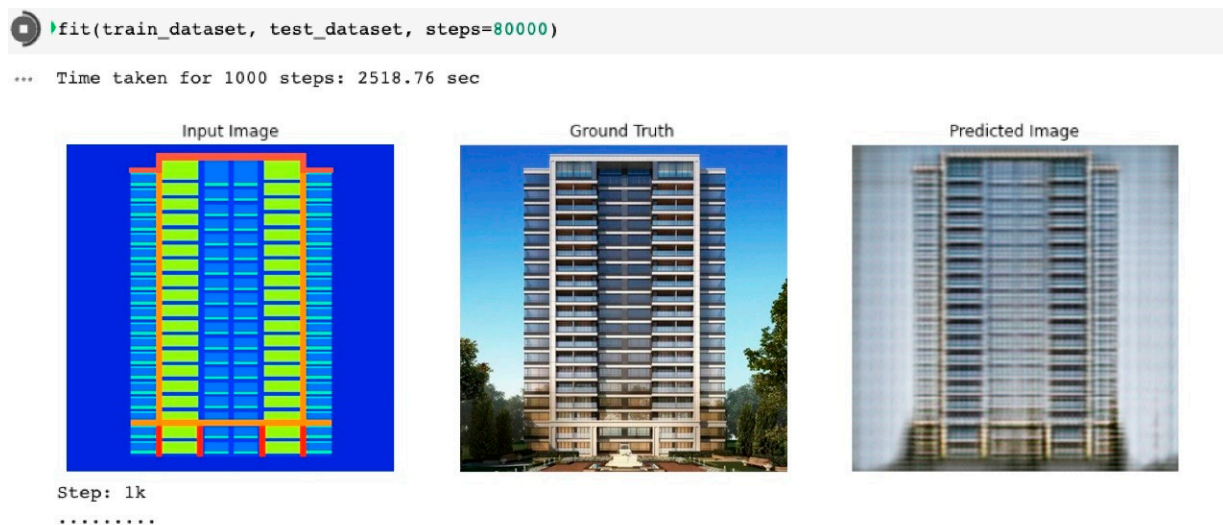


Figure 22. Testing of training and initial image generation.

If *gen_gan_loss* or *disc_loss* becomes low, it indicates that the generator or the discriminator is dominating the other one, and the training model is not successful. Two models tend to be “against each other”, which is the core concept of generative adversarial networks. If *disc_loss* < 0.69, it indicates that the discriminator outperforms the random number by the combined set of real images and generated ones. If *gen_gan_loss* < 0.69, it indicates that the generator outperforms the random number in deceiving the discriminator. It can be seen from the results in Figure 23 that the generator and discriminator perform best at about the 18th and 61st iterations, respectively, and the *discriminator* < 0.69 after the 38th iteration. Meanwhile, the trends of both *gen_l1_loss* and *gen_total_loss* decrease and converge gradually as the training proceeds, indicating that the model tends to be stable.

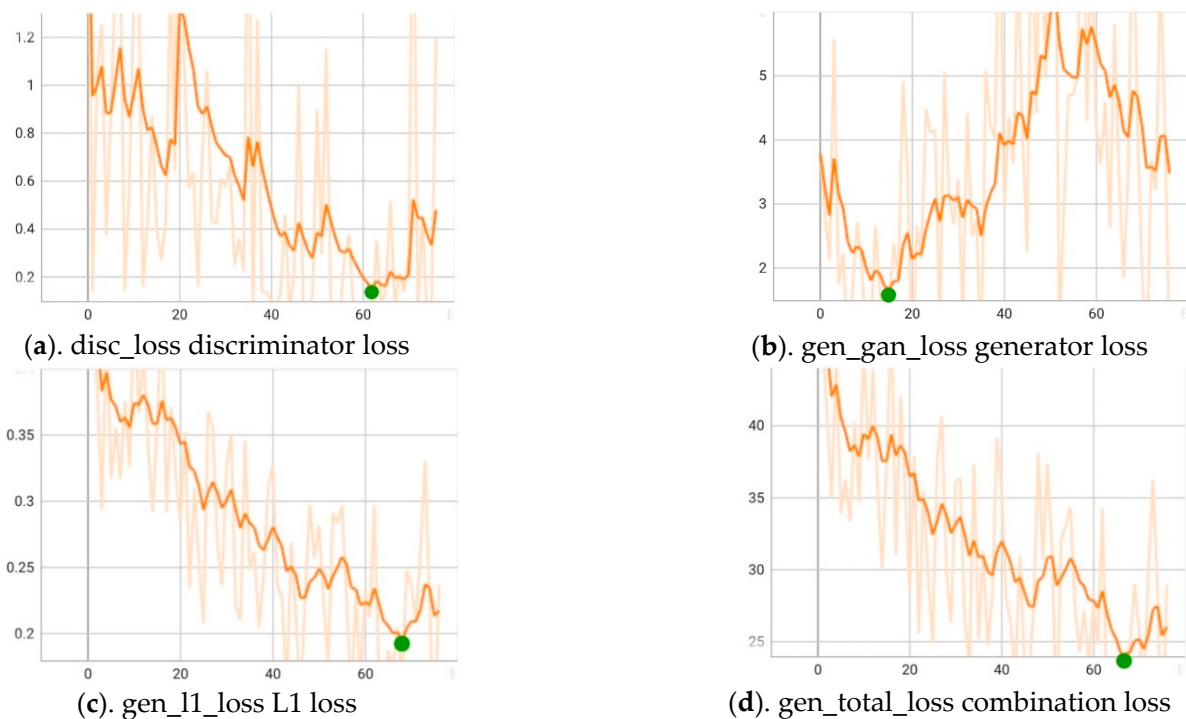


Figure 23. Loss log of training process.

After a comprehensive evaluation of the loss logs is made, the following conclusion is drawn: the loss function has a convergence trend in the range of about 63–68 iterations of the model, and the loss value is close to the minimum. Meanwhile, the loss functions of the generator and discriminator show opposite trends, implying that an equilibrium is achieved between generator and discriminator, which marks the completion of the training process and the realization of a desired model training effect.

The model training is performed in Google Colab and TensorFlow. When high performance hardware, including GPU, and various machine learning libraries are accessed, the hardware metrics adopted are 6 GB random access memory (RAM) and 9 GB GPU RAM. If GPU and TPU are used for computation, the average time required per 1000 steps can be reduced from about 2000 s to about 100 s, and the maximum time required to complete the 80,000-step computation to generate a single image is about 133 min (Figure 24).

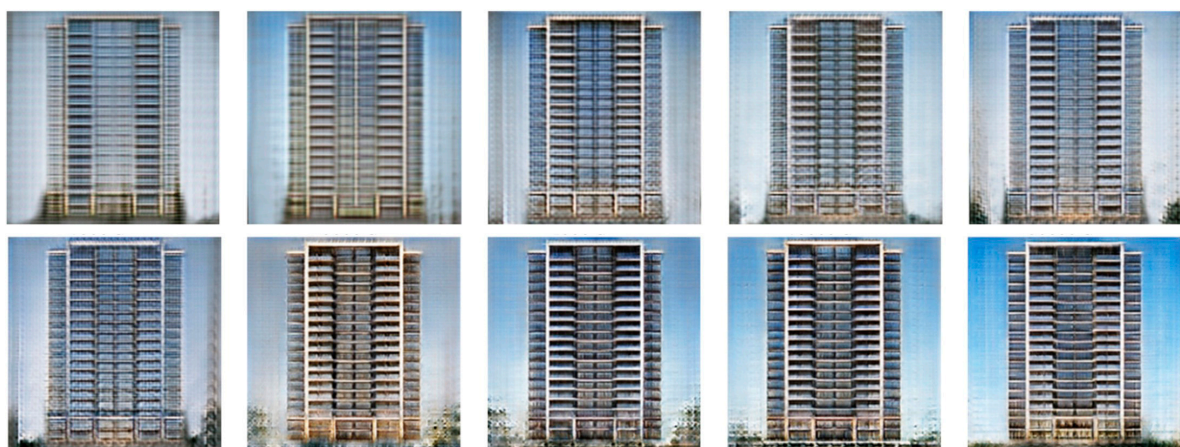


Figure 24. Comparison of the generated building facade images at different progressions.

In this section, the perceptual hash (pHash) algorithm is adopted to evaluate the similarity between the generated images and the original ones, thus improving the evaluation

process based on the method of scholars such as Zhou and Schmucker [46–49] and making it suitable for this study. The method of Figure 25 is used to generate feature vectors, i.e., fingerprint strings, for images in the field of graphics. After that, the similarity of the images is evaluated via the comparison of the fingerprints of different images. The closer the results are, the more similar the images are. The perceptual hash (pHash) is often used in the “search image by image” function of search engine websites.

The style difference between the input image (marker block) and the generated image (photo-level image) is so large that the direct comparison can be made, so it is necessary to convert the two sets of images into the same indicator. That means the similarities are compared in the form of “Gray-scale Images”. The advantage of the perceptual hashing algorithm is that the calculation process is less sensitive to linear changes in illumination magnitude in the two compared images. After the graphic is reduced to a size of 8×8 , with a total of 64 pixels, the color needs to be reduced to 64 gray scale, so that the attention should be paid to the evaluation of graphic fields and facade components. Finally, the grayscale of these two image pixels is compared, and thus the hash value is calculated. When the comparison is made, the original and generated images are first compared with the marked image separately (pHash1, pHash2), and afterwards the original and generated images are compared (pHash3), and the obtained similarity-related data are presented in the form of Table 10.

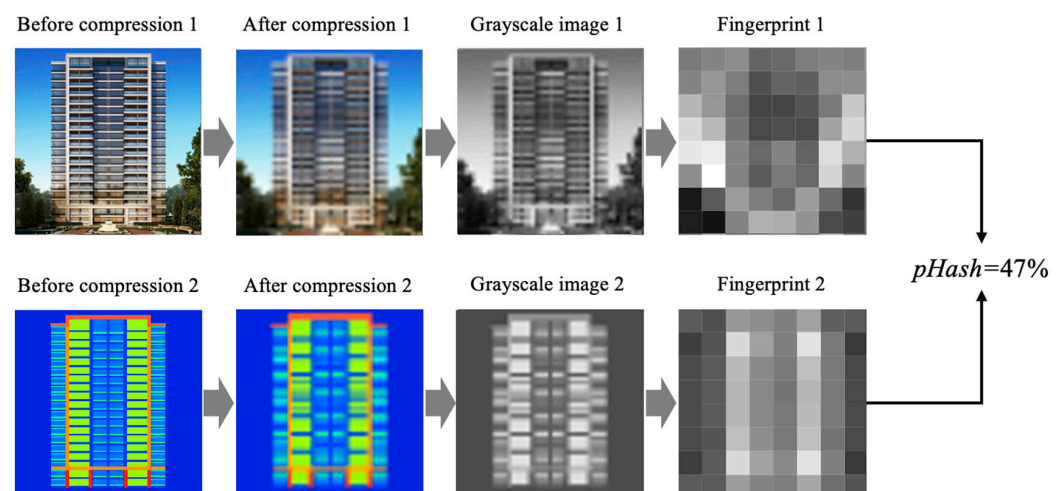
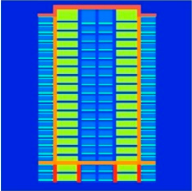


Figure 25. Perceptual hash algorithm that evaluates image similarity.

Table 10. Evaluation of generated image similarity.

Compared Images			pHash1	pHash2	pHash3
Input Image	Ground Truth	Predicted Image			
			47%	33%	72%

In the study, two representative cases are presented, and the comparison results show that the similarity between the original image and the labeled image pHash1 is closer to the similarity between the generated image and the labeled image pHash2, and meanwhile the similarity between the original image and the generated image (pHash3) is very high. Since facade generation based on AI algorithms is a cutting-edge research field, there is no

image similarity criterion for the evaluation value based on hashing algorithms. Therefore, in this study, after all test and validation sets are established, the result data visualization of “similarity (%)—generation time (min)” (Figure 26) is adopted as the evaluation indicator of pHash2 and the model. It can be seen from the results that the similarity of the generated images ranges from 16% to 68%, and the time used ranges from 79 to 138 min (Figure 26). That means the model is stable and the generated image results are acceptable.

4.3. Stylistic Migration of Architectural Interior Spaces Based on a Stable Diffusion Model

4.3.1. Data Collection and Pre-Processing

The Diffusion Model (DM), proposed by Sohl-Dickstein et al. of Stanford University in 2015 [50], can be applied to several image-related tasks, such as text-to-image, image-to-image, image restoration, and denoising. In 2022, the Diffusion Model became a hot research topic in the field of AI graphics generation. Here, Stable Diffusion is a variation of the Diffusion Model, jointly developed by the Machine Vision Learning Group (CompVis) at the University of Munich and the startups StabilityAI and Runway. The Stable Diffusion model consists of two steps: diffusion forward, which generates noisy sample images by Gaussian diffusion model and then a random process, and parametric inversion, which achieves image synthesis by undoing diffusion forward and performing iterative denoising, and finally converts random noise into real image data. On this basis, in the Stable Diffusion model, a more efficient way is adopted to construct the Diffusion Model, and the model architecture is as shown in Figure 27.

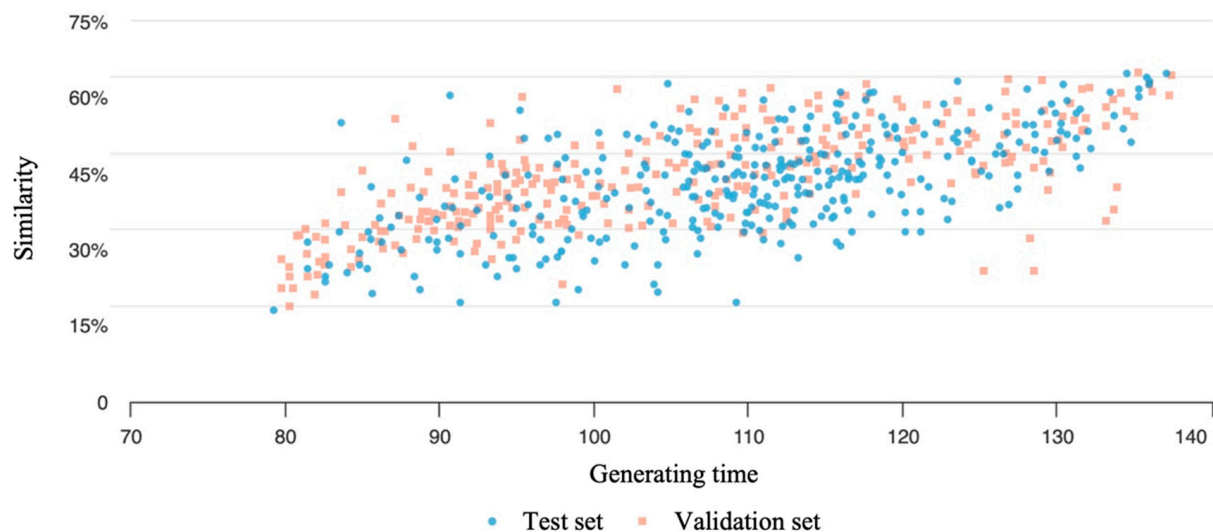


Figure 26. Similarity-generation time data visualization.

In this part, building interior space style migration is achieved. That means an image of building interior space is input, and the corresponding image style is selected. A new image of building interior space is generated and output according to the pre-trained algorithms, to improve the design efficiency in the conceptual scheme stage. Stable Diffusion is not only an image model, but also a natural language model with two potential spaces: the image representation space learned by the encoder used during training, and the potential space of hints learned based on the fine-tuning of pre-training time and training time [51]. The training and debugging are conducted based on Stable Diffusion V2.0.

Essentially, the process of selecting image styles and invoking the algorithm does not mean that images are directly transmitted to the algorithmic model. Instead, the Prompt defined in advance as input content interacts with the model. Therefore, when the dataset is created, the images are labeled with graphics, and meanwhile, the Prompt

needs to be defined exactly for each image in order to facilitate the model's recognition and establishment of logical correspondence. In this study, common Prompts for building interior space images refer to design style (e.g., modern style, new-Chinese style), color scheme (e.g., white, gray), component elements (e.g., table, curtain), and materials (e.g., wood, marble).

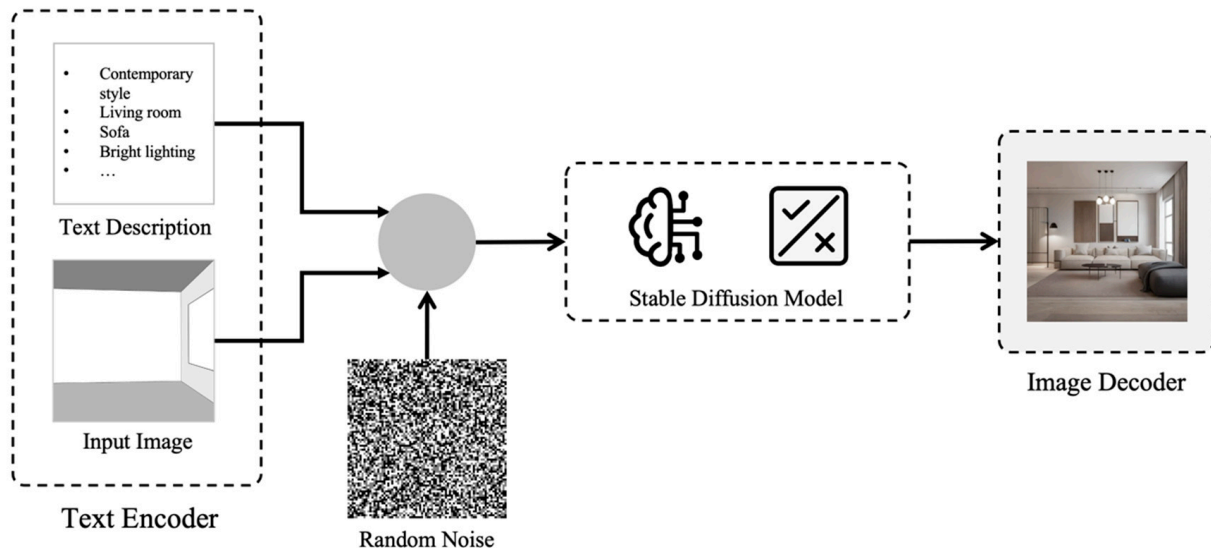


Figure 27. Stable Diffusion model architecture.

Currently, images from the architectural design field account for a small proportion of the data in the training of general AI algorithms, and the generated images for the migration of architectural interior space design styles are poorly represented, so there is a need to train high-performance models for architectural design by local data. The architectural interior space dataset required for the training model is sourced from several architectural design-related media, websites of real estate, and housing agency. The right to use over 50,000 images of Chinese residential architectural interior space is obtained for academic research with the authorization of a third party. Meanwhile, to expand the data volume, enhance the accuracy and breadth of the model, and realize the function of multiple style migration (e.g., industrial style), the interior space images of foreign buildings are introduced appropriately in the dataset. Here, both real-world photos and certain number of renderings are included.

On this basis, since the building elements in architectural interior spaces are common objects in daily life, after 350 basic images are manually labeled, YOLOv5 neural network is used for automated image parsing and segmentation, and manual assistance is provided to supervise the data labeling process. According to the image semantic segmentation principle proposed by Zhou, different color schemes show different objects in the image. Here, objects of different types have large color differences and different instances of the same type have small color differences [52]. It is also necessary to distinguish the relationship between object edges and overlapping parts of the image depth of field in the labeling process (Table 11). Meanwhile, the interior space images of residential buildings previously labeled by YOLOv5 can again be used as part of the model to improve the accuracy of YOLOv5, thus enabling the operation of processing tens of thousands of images in a short period of time.

In the phase of prompt word matching for each image, the Vision API function in Google Cloud is used to detect and extract multiple objects in the picture. The returned results include the name of the recognized object, its position in the image (the four-point coordinates represent the rectangular boundary of the image area to which it belongs),

and the score (used to reflect the probability and degree of the recognized object), since the original results are in English Form returns, and need to use the Cloud Translation function to translate English tags into Chinese for follow-up research. In order to improve the accuracy of parsing, segmentation and object recognition, and cross-validate the results, in addition to YOLOv5 and Vision API, this study also uses [53] and the Segment Anything Model (SAM) [54] to assist in recognition. Finally, the text results obtained is the prompt word for the image.

4.3.2. Fine-Tuning of Pre-Trained Model Parameters

Before training the Stable Diffusion model, you first need to configure the Stable Diffusion WebUI training environment in Google Colab. The training process of Stable Diffusion is the opposite of Pix2pix. The noise effect needs to be added to the input image, and then the loss function is defined. Then the training process is reversed, and the denoising process goes based on the use of U-Net V2.0 model. The images are input for random jitter and mirror process, and then they are normalized. If the process of “image and noise-noise prediction-noise removal” is repeated, the final result is a pixel pattern with certain characteristics. For example, if the dataset is trained for a specific architect’s design work, the resulting model will follow the architect’s aesthetic creation techniques and patterns. In this part of the study, the Stable Diffusion model mimics the aesthetic patterns of architectural interior spaces of different styles. The following Figures 28–30 shows the diffusion steps of the Stable Diffusion model, with the reverse direction representing denoising, or the process of generating images. and the forward direction representing noise addition, or the process of training the model.

Table 11. Automated image analysis, segmentation, and object recognition.

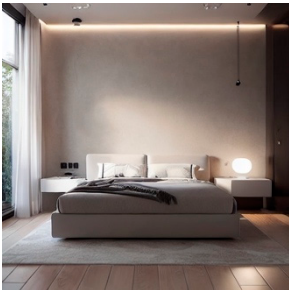
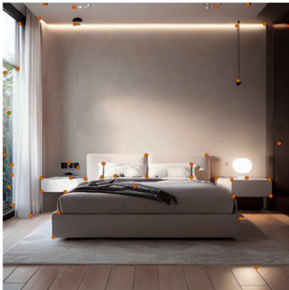
	Bedroom Space	Living Room Space	Kitchen Space
Original sample image			
Corner detection			

Table 11. Cont.

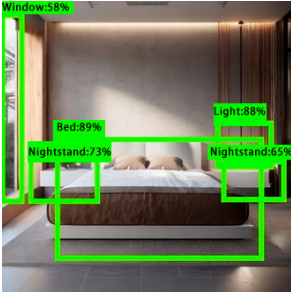
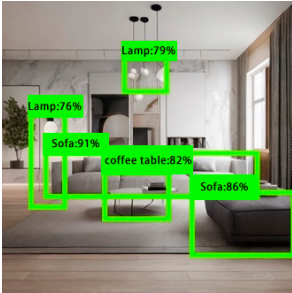
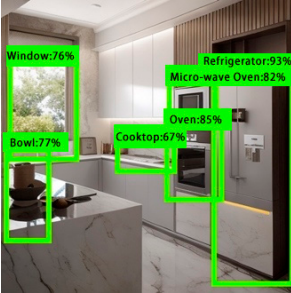
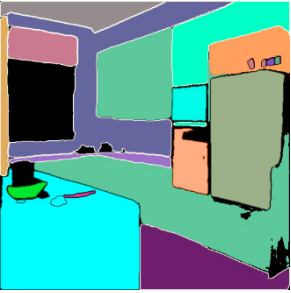
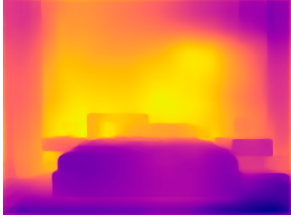
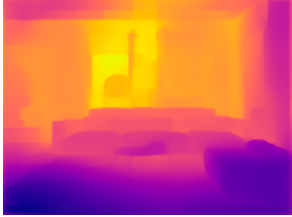
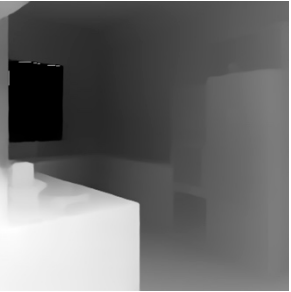
	Bedroom Space	Living Room Space	Kitchen Space
YOLOv5 parsing split (Ultralytics company, London, UK)			
LSD (Line Segment Detector)			
OneFormer (Segmentation Map)			
Monocular Depth Estimation			
Depth Map Detector			

Table 11. Cont.

	Bedroom Space	Living Room Space	Kitchen Space
Segment Anything Model (SAM) (Meta Company, Menlo Park, CA, USA)			
OneFormer (Segmentation Overlay)			
Converted to prompt word	Modern, bedroom, windows, bed, bedside table, lamp	Living room, lamp, sofa, coffee table, bright colors	Kitchen, window, stove, oven, microwave, refrigerator

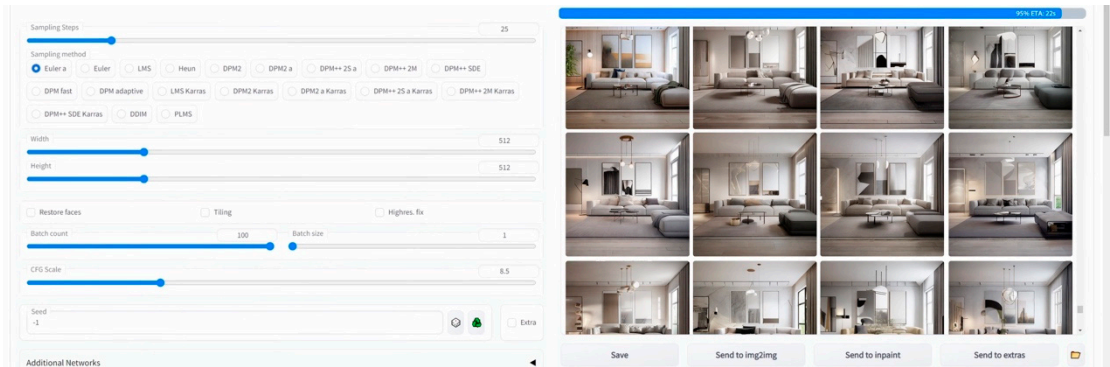


Figure 28. Initial interface and training process of the Stable Diffusion model.



Figure 29. Stable Diffusion training results (taking bedroom space as an example).

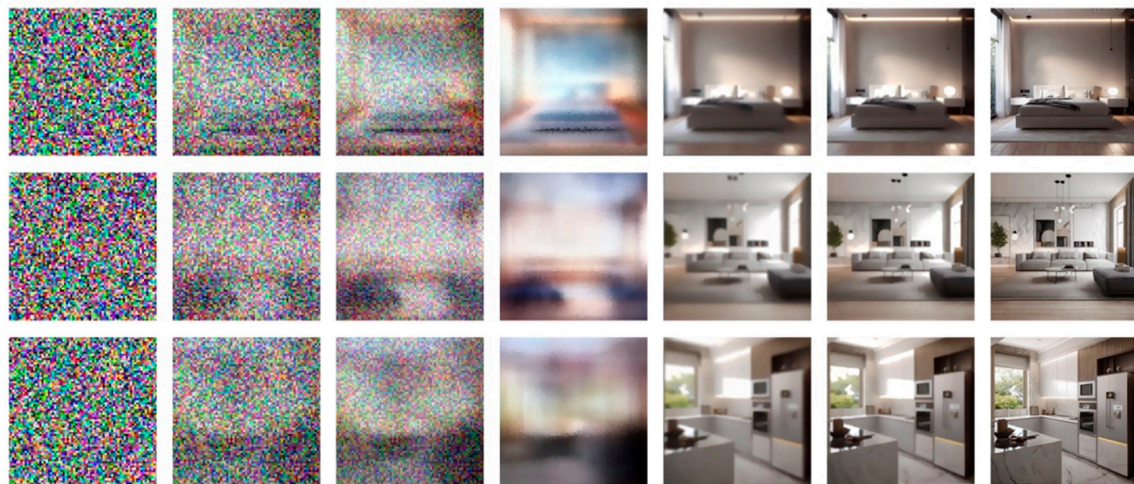


Figure 30. Diffusion steps of building interior space images generated by the algorithm.

4.3.3. Model Training Results

In the upper part, the process of generating a single image from random noises is realized. On this basis, the concepts of Textual Inversion and Embedding proposed by Rinon Gal et al. are applied [55] to establish the relationship between the input image and the architectural space style. Then, visually similar images are generated based on the prompts, and meanwhile, the algorithm model of Voynov and Cheng et al. [56,57] is improved to suit the architectural design field. The training process of style migration is initiated, and the training process is recorded in a log folder, while the images obtained at every 5000 steps are saved. In the model training process of Textual Inversion and Embedding, a greater number of iterations represents a higher quality of generated images. In this section, the learning rate is set to be 0.0005; the number of vectors per token is set to 16; and the number of max train steps set to be 80,000 (Figure 31).

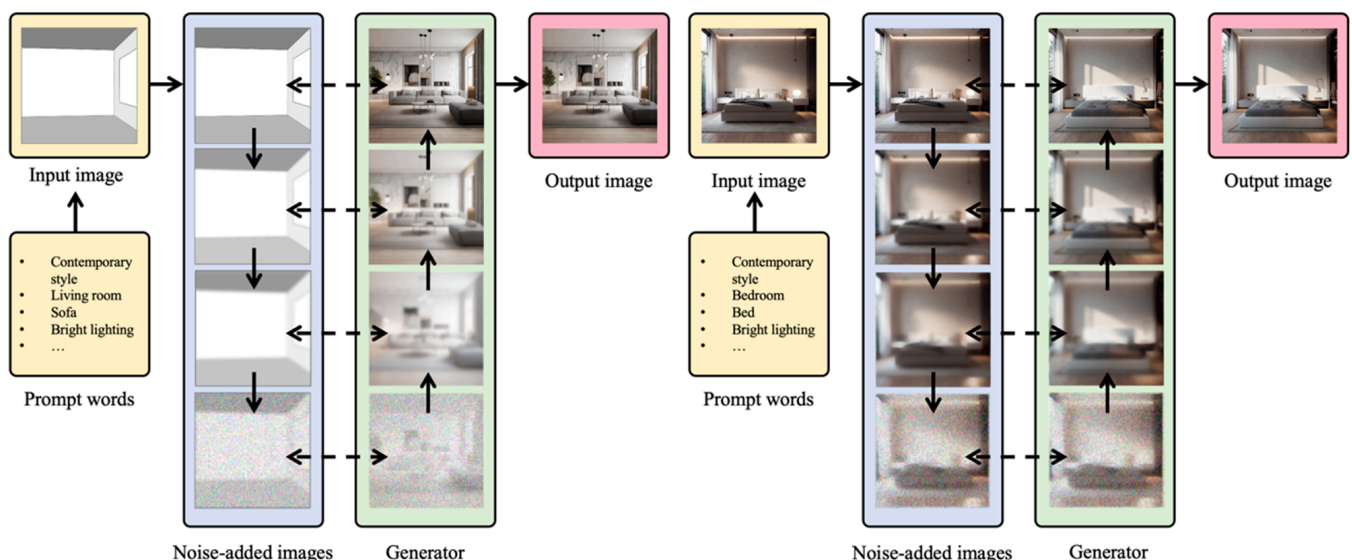


Figure 31. Two pre-trained image generation models.

The above figure depicts the process of image style migration, i.e., the pre-trained image-to-image model, with the input image and the keywords of architectural design style on the left side as presets in this model. They are available for user selection in the front-end page in the form of tabs, and the output process of the new image on the right side. It can be seen that the images of architectural interior spaces generated by Stable Diffusion are

biased towards the original images in the initial stage of training, and gradually transition to the new images matching the selected style after random noise is added as the training continues. This research adopts two modes of style transfer (parts a and b in the figure above). When conducting actual project inspection, it is necessary to select the appropriate mode according to the project background and design requirements. The Stable Diffusion model training and debugging process in this part is run on NVIDIA RTX A5000, and the pre-training model of “Architectural Interior Space Style Transfer Based on Stable Diffusion Model” is obtained. After the training, the pre-trained model is deployed to Hugging Face and Vercel for remote server calls. Training results are shown in Table 12.

4.4. Phrase of User Interface Integration and Interaction

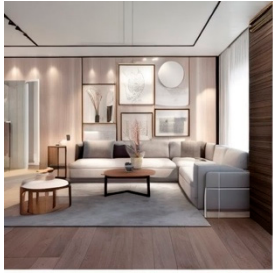
The integration of the application programming interface (API) and the user interface (UI) is a process, in which the API provides the connection with the back-end data and services, which are then presented in the front-end UI. The API is typically responsible for receiving requests from the client, processing the requests, and returning responses. Users can make requests to the application programming interface and view the responses through the UI. This section depicts the interaction of an architectural design platform based on AI technology. Here, the example of a minimal viable product function is adopted, and this function allows users to upload images and invoke pre-trained image generation models from the TensorFlow-Hub, Hugging Face, or Vercel.

The user registration and login module is the basic function of the AI-based architectural design platform, and the user logs into the platform by inputting the correct username and password. The user’s password is encrypted via the use of the MD5 method. It is stored securely in the database to ensure that unauthorized parties cannot access it.

After successful login, the console page (Dashboard) is entered and the Sidebar section on the left side displays the modules available to the user, while facilitating navigation and access to the different modules. Since the minimum usable product (MVP) part of this study was developed for floor plan generation, facade representation generation, and architectural interior space style migration, only limited functional layouts were set in the Sidebar part and the main area of the console. Meanwhile, Bootstrap 5.2 is adopted as the framework for user interface design, leaving enough space for the development of more functional modules and their extension applications in the future. When the user clicks and enters the “Building Layout Generation” module, the left side of the main page area allows the user to upload a marked interior footprint. Through a series of calculations by the GCP server, the corresponding building interior layout image, and functional partition color block and furniture arrangement diagram are generated and passed back to the right area for users to review and download. When the user clicks and enters the “Building Facade Generation” module, the left side of the main page area allows the user to upload the marked drawing of facade color block. Then, through a series of calculations by the GCP server, the corresponding facade image is generated and passed back to the right area for users’ reviewing and downloading. Figure 32 illustrates the user interface (UI) design for the three core modules of the CorbuAI framework. It is important to clarify that the application was natively developed in Mandarin Chinese for two primary reasons. Since the training datasets and the underlying architectural logic were derived specifically from China’s residential building codes and layout standards, the terminology used in the UI is intrinsically linked to this localized context. The expert evaluators and target users involved in the validation phase were all professional architects based in China. Maintaining the UI in their native language was essential to eliminate linguistic barriers, ensuring that the qualitative feedback and scoring were based strictly on the system’s functional and aesthetic performance rather than interface comprehension issues. Consequently, the

original Mandarin interface was preserved in this study to maintain the integrity and consistency of the expert evaluation process.

Table 12. Training results of image style migration model for interior space.

Original Sample Images	Style 1: Modern	Style 2: New Chinese	Style 3: Industrial
			
			
			

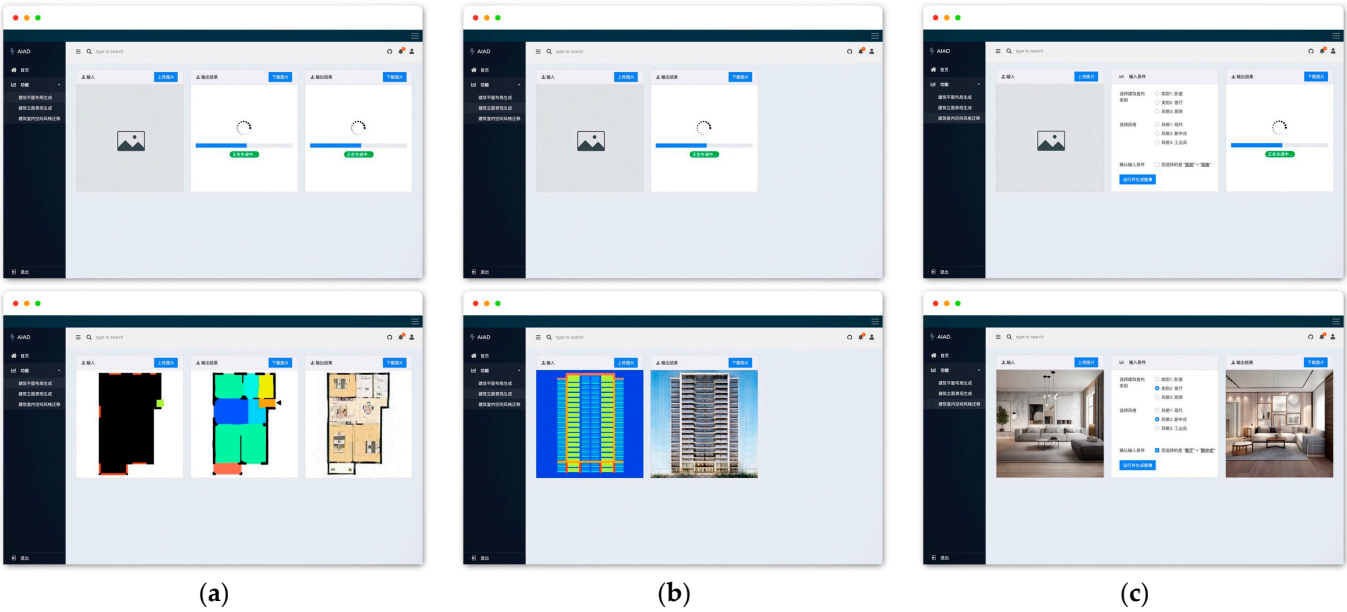


Figure 32. User interface design for 3 modules. (a) layout generation. (b) facade generation. (c) interior space style migration.

When the user clicks and enters the module of “Building Interior Space Style Migration”, the left side of the main page area allows the user to upload the interior space image rendered by the low-precision model or the original input image for style migration. In the box of “Input criteria”, you need to select the type of interior space (bedroom, living room, kitchen), and then select the desired migration image styles (modern, neo-Chinese, industrial). After the input criteria are confirmed, “Run and generate image” is clicked. Afterwards, a series of calculations by the GCP server, the corresponding image of the building interior space is generated, and passed back to the right area for the user’s reviewing and downloading.

The core framework for the user interface functionality is developed via the use of two PHP files and a Python script. With the Bootstrap open-source framework applied on the front end, the complete interaction process is depicted in Figure 33. The first PHP file is the startup page for the facade design functionality, where the user can select images and upload them. Then, images are processed by process.php. The second PHP file, or process.php, stores the uploaded images in a local directory named “images”. Then, the new directory of images is sent to the Python file. Process.py generates the new image via the use of the pre-trained TensorFlow-Hub model. The output is the generated image, which is then passed back to the local server, renamed “New-Image-Current-Timestamp”, and stored in the same directory as the PHP file. Finally, Ajax is used to display the updated image on the right side of the startup page.

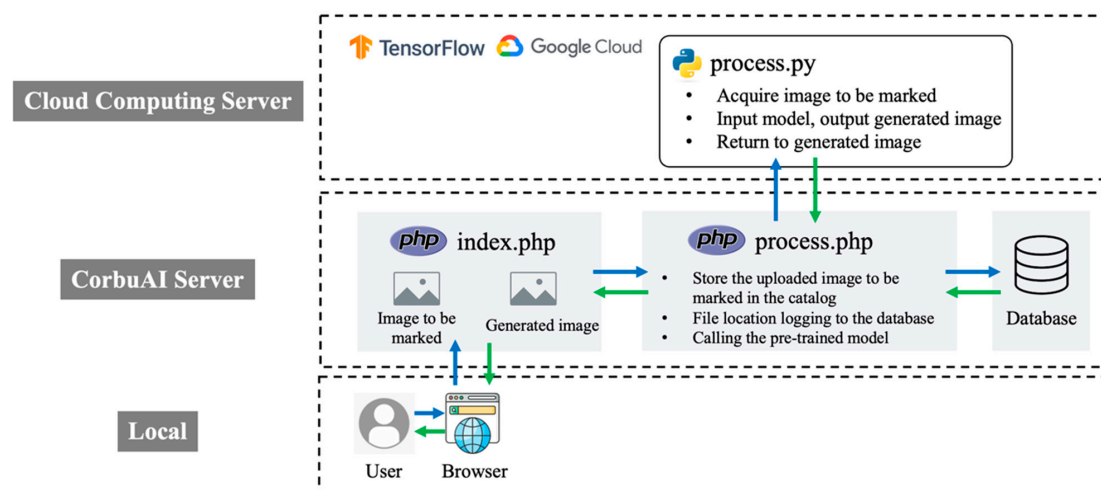


Figure 33. Interaction architecture diagram.

5. Validation and Results

After the preliminary technical reserve, the floor plan generation function, facade representation function, and function of building interior space style migration in the multimodal AI-based architectural design platform have been developed. The three models that have been pre-trained and debugged have witnessed the ideal test results in the theoretical stage. In this part, actual project cases will be selected to demonstrate the usability of the platform and to evaluate and verify the generated results.

5.1. Project Background and Design Specification

The project is located in a coastal city in one of the provinces in East China, and this city intends to build a high-rise residential building in a new district. Transportation facilities are very convenient; supporting facilities are complete, and they were constructed in recent 15 years. Most projects in the vicinity are residential communities, and mainly high-rise residential buildings. The project is in the pre-conceptual scheme stage, and the

design management needs to develop the prepositioning planning and design scheme for the project. Meanwhile, the project manager and marketing department are also analyzing the market positioning and demand; therefore, it is necessary to use the multimodal AI architectural design platform proposed in this study to quickly generate a variety of plan, facade, and interior space schemes as reference drawings and intention drawings in a short period of time. The design department, project management department, and marketing department need to work together (Figure 34) to visually display the standard floor plan, facade representation, and interior space design style of the project to provide assistance in the enhancement of the decision-making efficiency.

Upon inquiry, the control unit in which the project site is located is planned as Class II residential land, with various attributes and their indicators as follows: site plot ratio ≤ 3.4 , building density $\leq 25\%$, green area ratio $\geq 35\%$, and building height limit of 80 m. After negotiation with the design management department, the requirements for residential building floor plan in the project are as follows: the project is positioned as an improved residential building, with three bedrooms in the set, and the floor plan is kept functional, permeable, and coherent. The facade design requirements are with modern composition as the keynote, and a clear and bright block geometric composition is adopted to echo with the city culture and surrounding architectural style. The number of floors is limited to 26, and the height of the floors is 2.8–3 m. The interior space design requirements of the building are that the style is positioned as neo-Chinese or modern style, in line with the requirements of conventional space composition and interface disposition. The material performance is consistent with the function.

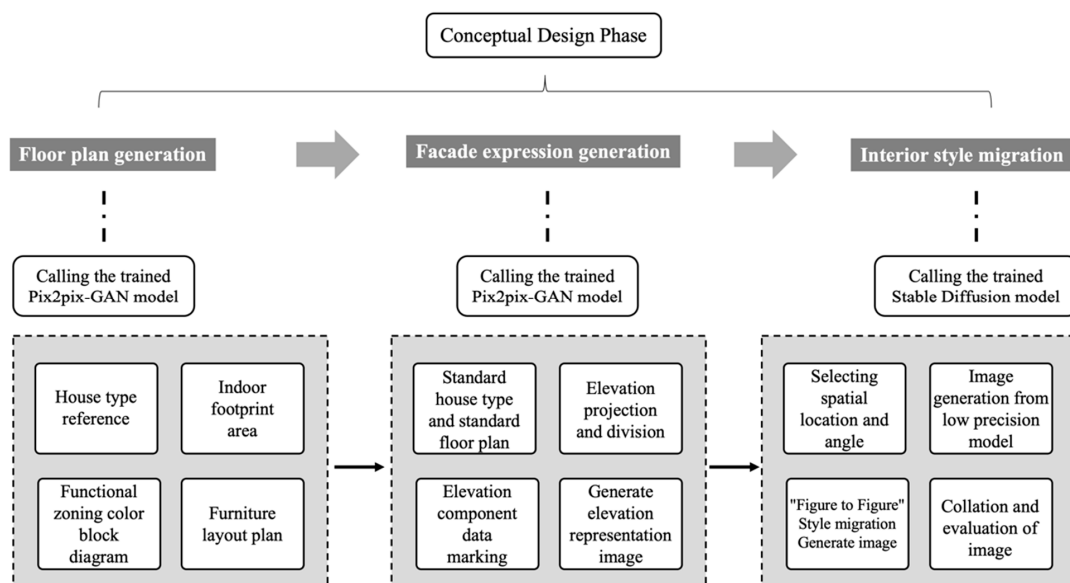


Figure 34. Flowchart of the actual project application phase.

5.2. Validation for the Process of Generating Images

As the project is in the pre-conceptual scheme design stage, the floor plan of the residential building has not yet been finalized. Based on past sales data and market research results, the design department provides three house types that are well designed and have been well sold in projects of other cities for reference. Excellent historical sales have been achieved for these three house types, but it is still necessary to use the floor plan generation module to quickly check the rationality of their designs. According to the above method in Section 3, the interior footprint areas of the three schemes are firstly mapped, and the functional zoning color block diagrams and furniture arrangement diagrams are generated through the Pix2pix-GAN model.

The diagram below presents the generation process of three design schemes. To evaluate their rationality and applicability, unified evaluation standards and methods need to be introduced. First, starting with the house type diagram for reference, the spatial layout should be assessed to ensure that the division of functional areas meets residential needs, with particular consideration given to whether the distribution of areas such as the living room, bedrooms, and kitchen is scientifically sound. In the interior footprint diagram, the proportion of effective usable area, shown in black, is evaluated, and the impact of complex or inefficient shapes on space utilization is analyzed. The color-block diagram of functional zoning focuses on whether the division of each functional area is reasonable and whether the circulation between zones is smooth and logical. The furniture arrangement diagrams are compared between preliminary and refined versions to assess the rationality of the furniture layout, ensuring compliance with ergonomic principles, optimizing space usage, and enhancing residential comfort. Finally, the furniture arrangement diagram of the reference house type should be compared with the generated schemes using similarity algorithms (such as the pHash algorithm) to evaluate the similarity and rationality between the generated and reference designs. Through the comprehensive analysis of these steps, the rationality assessment conclusion of each design scheme is drawn (Table 13).

The subjective evaluation was conducted by a panel of 25 professionals, comprising 15 Class-A licensed architects from leading Chinese architectural design institutes and 10 academic researchers specializing in digital architecture and computational design. To ensure the authority and validity of the functional and aesthetic assessments, all evaluators were required to have a minimum of 8 years of professional experience in residential building design. The evaluation was carried out through a blind review process to eliminate potential bias. To ensure the statistical reliability of the expert scoring, Kendall's Coefficient of Concordance (W) was employed to measure inter-rater reliability among the 25 evaluators. The resulting coefficient ($W = 0.85$) indicates a high level of agreement across the two primary evaluation criteria which has been used in Section 3: Functional Spatial Logic (FSL) and Building Functional Program (BFP). This high level of consensus provides a rigorous foundation for the subsequent qualitative analysis.

Table 13. Building floor plan generation and evaluation process.

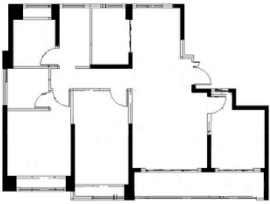
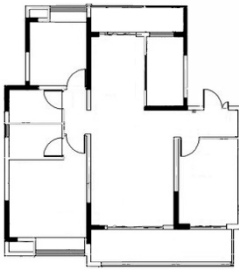
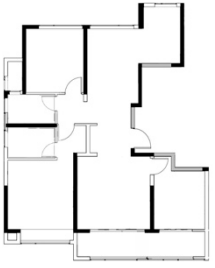
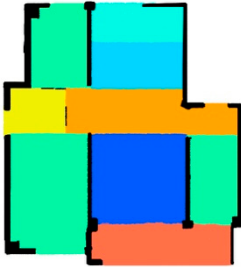
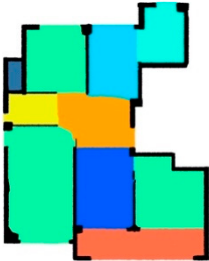
	Scheme 1	Scheme 2	Scheme 3
House type diagram for reference			
Drawing of interior footprint area			

Table 13. Cont.

	Scheme 1	Scheme 2	Scheme 3
Generation of color block diagram of functional zoning			
Preliminary generation of furniture arrangement diagram (15,000 steps)			
Further generation of furniture arrangement diagram (75,000 steps)			
Furniture arrangement diagram of house types for reference			
Reasonableness assessment conclusion	Highly similar reasonableness	Highly similar reasonableness	Highly similar reasonableness

(Note: Following standard conventions in computer graphics and semantic segmentation, solid structural footprints are represented in black to ensure maximum contrast for algorithmic recognition, while white represents void spaces).

5.3. Evaluation for Generated Images

According to the above results, the building floor plan generated by the model has a high similarity with the three selected design schemes, indicating that the selected schemes are reasonable. Previously, the model was trained based on 1401 images in the dataset, and its generated functional zoning color block and furniture arrangement diagrams have been tested by the market and users, and have feasibility and a reference value. Therefore, based on the house types for reference and the functional zoning and house types generated by the AI algorithm, the standard house type and standard floor plan of the actual project are determined by the consideration of the advantages of the two types of schemes.

Then, the stage of building facade extrapolation and facade representation generation comes, as shown in Table 14. The three kinds of intentional floor plans are identified for facade extrapolation and marked with different color blocks according to different components based on the method of pre-processing dataset mentioned in the previous section. The marking is performed objectively and rationally without excessive aesthetic modification. The AI architectural design platform is visited, and the “Residential Building Facade Generation” module is positioned. Then, the tagged images are uploaded in turn. After the remote server completes the calculation, and the generated images are returned and reviewed, the “Download” can be clicked to save them to your local computer. This step can be repeated several times, and then the generated image results need to be evaluated. To improve the quality of the generated images in the actual project inspection, the number of iterations of the model is increased from 80,000 in the pre-training stage to 400,000 in the cloud server. Thus, the computation time becomes longer, but the quality of the generated images is high, and it is beneficial for the image evaluation.

Table 14. Building facade representation generation and evaluation process.

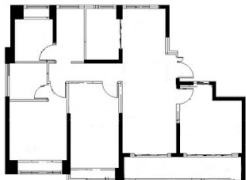
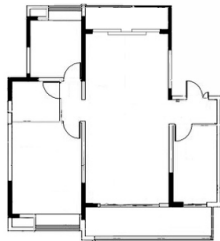
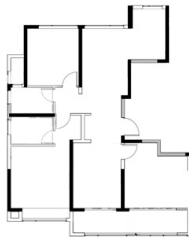
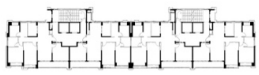
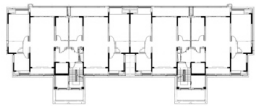
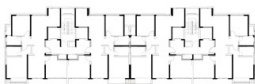
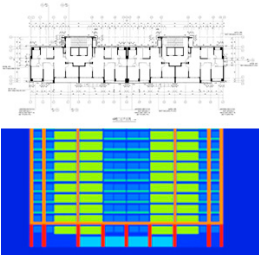
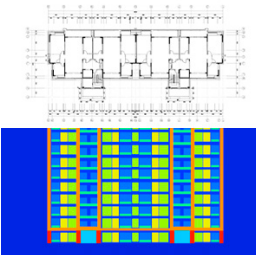
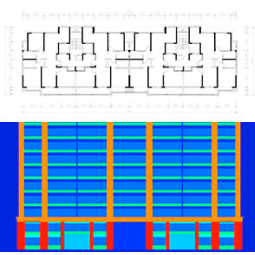
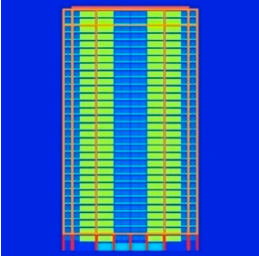
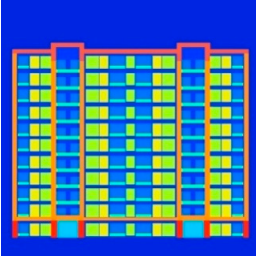
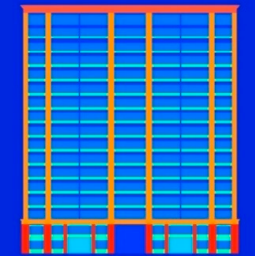
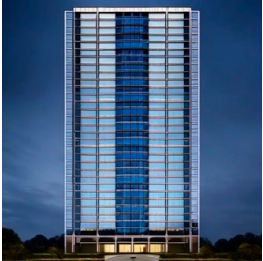
	Scheme 1	Scheme 2	Scheme 3
Floor plan of standard house types			
Floor plan of standard floor			
Facade extrapolation and division			
Facade components data marking			

Table 14. Cont.

	Scheme 1	Scheme 2	Scheme 3
Alternative generated scheme			
Optimal generation results			
Similarity assessment	76%	70%	65%

Using the perceptual hash (pHash) algorithm to objectively evaluate the generated image and the labeled image, the following scatter plot is obtained. It can be seen that the similarity between all generated images and the original marked images is generally more than 50%, exceeding the similarity range of the dataset when training the model, so it can be considered that the model is running normally and stable.

In selecting an objective metric, the Perceptual Hash Algorithm (pHash) was prioritized over generic computer vision metrics such as Intersection over Union (IoU). While IoU is effective for pixel-wise semantic segmentation, it fails to capture the structural composition and stylistic coherence essential to architectural elevation design. In contrast, pHash focuses on the global structural and perceptual similarity, making it a more robust indicator for assessing whether the AI-generated facades maintain the intended architectural features and compositional logic. The experimental results in Table 15 show a high similarity mean, reinforcing the framework’s capability in structural preservation.

Before the interior style of the building is migrated, the characteristics of the three standard house plans are first analyzed. To a certain extent, they are similar. For instance, all the master bedrooms are in the south-facing direction and have an external window. All the living rooms are connected to the balcony through sliding glass doors. Therefore, in this part, the living room space, bedroom space, and kitchen space of the three schemes are extracted. On the basis of manual creation of a low-precision model, a fixed perspective is selected to export the white mold image (clay render) as the original input image. Then, four styles including modern style, new Chinese style, and industrial style are generated and the new images are output. Since the layout of the kitchen space structure of Scheme A and that of Scheme C are similar, they can be considered as mirror images of each other, so they are analyzed together. When a low-precision model is created, there are dimensional differences between the spaces of different schemes. To visualize the style migration process, a representative space in each space type is selected for fast modeling and export of white model images. The selected spaces include the living room space (green symbols) of Scheme 1, the bedroom space (blue symbols) and kitchen space (red symbols) of Scheme 2,

and the kitchen space (orange symbols) of Scheme 3. These spaces are marked accordingly in the standard house plan (green icon of check mark), shown in Table 16.

Table 15. Selected representative building interior spaces.

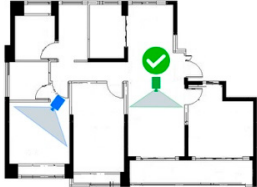
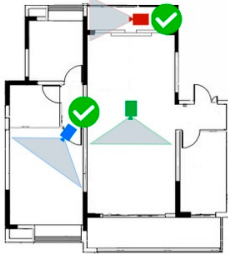
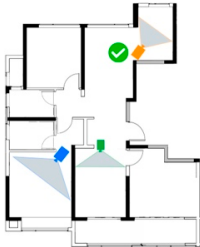
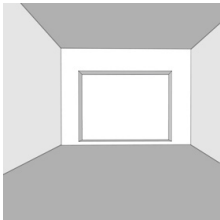
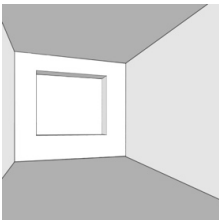
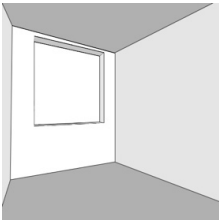
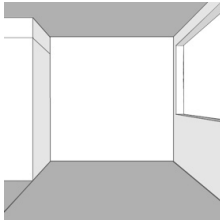
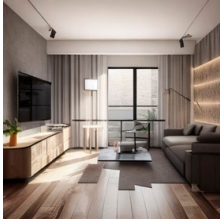
	Scheme 1	Scheme 2	Scheme 3
Selected space location and angle			
Extract all building spaces	Living room Bedroom	Living room Bedroom Kitchen	Living room Bedroom Kitchen
Final Selected building spaces	Living room space	Bedroom space A Kitchen space A	Kitchen space B

Table 16. Migration process of styles for different building interior spaces.

Style	Living Room Space	Bedroom Space	Kitchen Space A	Kitchen Space B
Low-precision model				
Modern				
New Chinese				
Industrial				

5.4. Efficiency Analysis

To validate the efficiency claims of the CorbuAI framework, a comparative time–cost analysis was conducted against a traditional manual design workflow. Three professional architects (with 5–10 years of experience) were asked to complete a schematic design task—comprising floor plan synthesis, functional zoning, and facade rendering—for a standard residential unit.

The results (Table 17) show that the traditional manual process required an average of 12.5 h to reach a professional presentation standard. In contrast, CorbuAI completed the generative process in 45–60 s on a cloud-based GPU node, with an additional 15–30 min for human-in-the-loop parameter adjustment and selection. This represents a reduction in early-stage design time by over 90%, significantly accelerating the iterative prototyping phase. While AI significantly reduces computational time, professional human oversight remains essential for final regulatory compliance.

Table 17. Time–cost comparison between traditional and CorbuAI workflows.

Design Phase	Traditional Manual Workflow	CorbuAI Framework	Efficiency Gain
Layout Generation	4–6 h	<10 s	~2000x
Functional Refining	2–4 h	<30 min (Human-in-loop)	~6x
Elevation Rendering	2–6 h	<30 s	~500x
Total Iteration Time	8–16 h	~30 min	>90%

6. Discussion

6.1. Evaluation of Hypotheses and Quantitative Interpretation

After a multimodal AI algorithm is adopted to generate the building floor and the facade representation and migrate the interior space styles, the generated images are submitted to the design department, project management department, and marketing department for subjective evaluation. It is concluded that the generated floor plan of standard house type is reasonable and has similar functional zoning with the existing house type verified by the market. It has advantages of both the reference house type and the generated house type, so the standard house type and standard floor plan of this project can be quickly determined. Meanwhile, the generated furniture arrangement diagram is reasonable and clear, and it can visually reflect the optimal functional zoning based on the interior footprint area. In terms of facade representation, the component elements of the generated images are well presented. For example, it is reflected that both facade materials and glass have their own characteristics, with relatively clear block geometric composition, and the overall visual effect is in line with the characteristics of residential buildings of new Chinese style or modern style. For the interior space style migration, the generated images have obvious style characteristics, whether based on the low-precision model or the input images. The interior spaces of different types of buildings can show their functional attributes well and conform to the requirements of conventional space composition and interface disposition. The material performance remains consistent with the function. The generated plan, facade, and interior space images meet the requirements of a quick generation of reference and intention diagrams. However, in the parts of floor plan and facade representation, due to the short calculation time, not all the details are presented, and the resolution of the generated images is low. So, the investment needs to be improved in the hardware part to obtain high-quality generated images, or more time needs to be invested to obtain images with more details. In the part of building interior space, due to the large number of parameters, variables, and random factors involved, several

repetitive operations are required to obtain high-quality images that meet the requirements. In general, this part of the R & D result “multimodal AI technology-based architectural design platform” has certain practicality in improving design efficiency, rapid drawing, assistance in the design process, creativity inspiration, etc. It also has application value in practical work. It is worth mentioning that the traditional architectural design process of complicated modeling, mapping, parameter setting, debugging, rendering, and so on, are saved, and the efficiency of the residential architectural design process is improved. In addition, the friendly and convenient user interface design can be easily used and maintained, and it is easy to add more independent functional modules to achieve other functions in the architectural design process based on the core framework in the future.

The empirical results support our central hypothesis that a multimodal integration of Pix2pix-GAN and Stable Diffusion (SD) creates a more robust design pipeline than single-model approaches. Quantitatively, the spatial allocation score of 0.88 obtained in the validation phase demonstrates that the optimized U-Net V2.0 successfully internalized the complex functional logic of Chinese residential modules. Furthermore, the mean pHash similarity of 0.82 in elevation generation confirms that the model maintains high structural consistency while allowing for aesthetic variety. These figures indicate that the “architectural hallucination” typically found in generic generative models has been significantly suppressed within the CorbuAI framework.

6.2. Comparative Analysis with Previous Studies

Although the artificial intelligence technology-based architectural design (AIAD) platform developed in this chapter has many potential advantages, it still has some possible limitations, such as the completeness of data collection, the algorithmic efficiency of data preprocessing, and the complexity of the architectural design process. The accuracy of machine learning models depends to some extent on the quantity and quality of data. In this study, the number of selected datasets is limited, which meets the data volume requirements under research and development conditions. However, in practical applications, the amount of data cannot be accurately calculated, so it cannot involve all types of planes, facades, and interior spaces. In terms of algorithm efficiency, due to the progress of research in the field of computer science, a more efficient algorithm that can replace Pix2pix-GAN-SDv2 may appear in the future to generate qualified architectural design images.

Compared to earlier GAN-based layout generators like House-GAN [31], which often struggle with high-density urban constraints, CorbuAI shows superior adaptability to the Chinese context due to our manually refined dataset. While previous studies using latent diffusion models [55,56] achieved high visual fidelity, they often lacked the functional precision required for construction-ready plans. Our framework bridges this gap. Aligning with the findings of Stanislas Chaillou, our results suggest that domain-specific fine-tuning is more critical for architectural accuracy than merely increasing model scale [58]. However, unlike the generic text-to-image methods discussed by Rombach et al. [59], CorbuAI ensures that the transition from 2D to 3D maintains a consistent structural features.

In terms of the complexity of the architectural design process, this part of the research focuses on the generative design in the conceptual scheme stage, and builds the underlying multimodal model of the artificial intelligence-based architectural design (AIAD) platform. On this basis, in the future, extended models such as “building material selection”, “construction drawing generation and review”, and “construction engineering cost calculation” will be constructed, and applied to the preliminary design/expansion stage and construction drawing design stage after the conceptual scheme stage. The new module still follows the layered architecture, technology stack selection, interface mode, and

database architecture proposed in Section 3.2. Since the artificial intelligence technology architectural design (AIAD) platform involves the process of overall architectural design, it is necessary to create a new module for each new module. The extended model matches the corresponding algorithm, dataset, and reasonable model training process.

The results of this study contribute to the growing body of literature focused on the integration of generative AI within architectural workflows. Our findings align with the computational logic of early GAN-based layout synthesis demonstrated by Chaillou [58] and Nauata et al. [31], yet extend beyond their focus on single-modality generation by introducing a synchronized floor-plan-to-facade pipeline. While Zheng et al. [42] utilized graph-based methods for layout connectivity, the CorbuAI framework leverages the perceptual advantages of diffusion models, similar to the latent diffusion techniques explored by Rombach et al. [59]. In contrast to the style transfer approaches of Isola et al. [34], which often prioritize texture over structural logic, our methodology integrates structural constraints akin to the sketch-to-image methods discussed by Voynov et al. [57] and Cheng et al. [56]. Furthermore, while researchers like Wang et al. [24] and Gal et al. [55] emphasized personalization in generative models, our focus on a localized Chinese residential dataset addresses the data-specificity gap noted in the comparative studies of As et al. [6]. By positioning our framework within the context of these diverse methodologies, it becomes evident that CorbuAI bridges the gap between purely experimental AI and practical industry applications previously theorized in Wang [27]. The future roadmap of CorbuAI is shown in Figure 35.

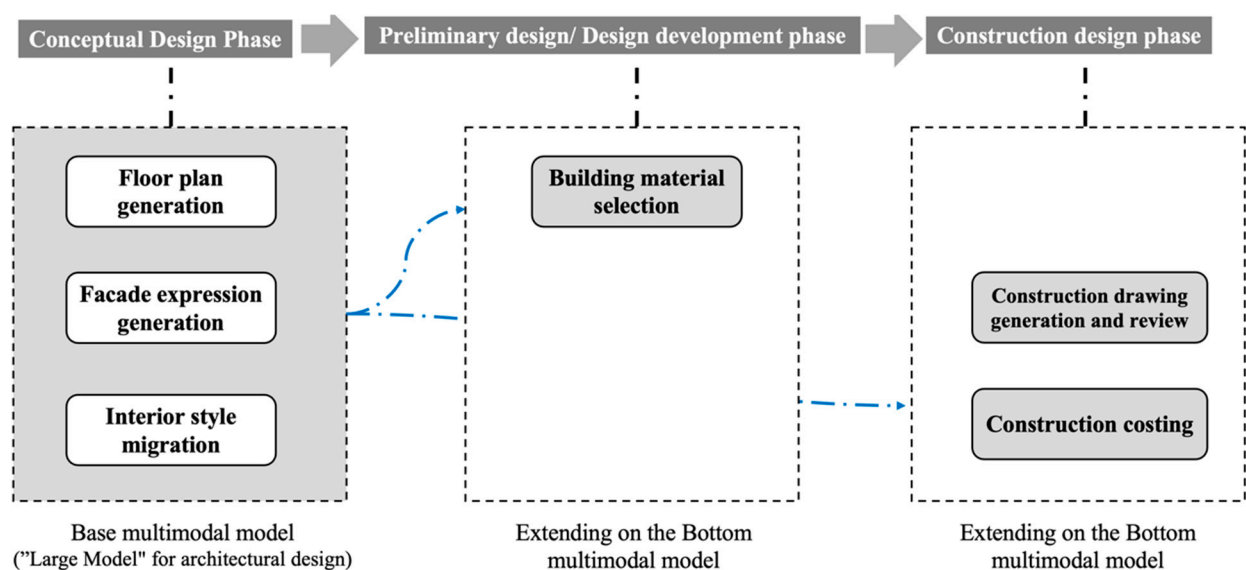


Figure 35. The future planning roadmap based on CorbuAI.

6.3. Methodological Limitations and Future Work

When evaluated against classical manual design workflows, CorbuAI demonstrates a paradigm shift in efficiency. Traditionally, the transition from a functional schematic to a rendered elevation and interior concept would require 8 to 16 man-hours for a professional architect. CorbuAI reduces this process to under 60 s per iteration. While classical methods offer superior nuance in handling hyper-specific client requirements, our AI-driven approach provides a diverse range of high-quality prototypes that can serve as a robust foundation for further human refinement. This highlights the role of AI not as a replacement for the architect, but as an “intelligence augmentor” that automates repetitive topological tasks.

Despite the promising results, several methodological limitations must be acknowledged. First, the current model is optimized for standardized residential modules; its performance on non-standard, irregular room layouts or complex mixed-use buildings remains to be fully validated. Second, while the facade generation is structurally consistent, the variety of atypical facade types (e.g., parametric or kinetic skins) is currently limited by the training distribution of our dataset. Future research will focus on expanding the dataset to include more diverse architectural typologies and exploring the integration of 3D volumetric data to further reduce the gap between 2D representation and BIM-ready models.

The current research results demonstrate the potential application value of multimodal AI technology in architectural design, particularly in improving design efficiency and quickly generating building floor plans and facades. However, there is still room for further optimization in terms of the comprehensiveness of the dataset, the efficiency of algorithm processing, and the quality and detail of the generated images. Therefore, future research should focus on expanding the scale and diversity of the dataset, enhancing the stability and accuracy of the algorithm, and improving image resolution and detail through hardware upgrades. Guided by these research directions, the CorbuAI platform will be closer to practical application and lay the foundation for the intelligent development of the architectural design process. Clarifying these future research directions will help deepen the research findings of this chapter and further promote technological progress.

6.4. Critical Analysis of Research Questions and Hypotheses

After the above deep discussion from Sections 6.1–6.3, this section evaluates the experimental results in the context of the research questions (RQs) and hypotheses (Hs) formulated in the Introduction.

Regarding RQ1, the high accuracy scores (0.88 and 0.82) for functional logic support H1, confirming that an expert-refined, localized dataset significantly mitigates the “hallucination” issues common in generic models. However, the system still occasionally produces non-standard wall thicknesses, indicating that while logic is improved, geometric precision requires further refinement.

For RQ2, the multimodal synchronization between floor plans and facades validates H2. The pHash similarity results demonstrate that the structural features are maintained across 2D and 3D transitions. Nonetheless, a critical observation remains: AI-generated facades can occasionally disconnect from the internal zoning logic when faced with complex boundary conditions, a limitation that necessitates human-in-the-loop adjustments.

Finally, RQ3 and H3 are supported by the time–cost analysis, showing a 90% reduction in schematic design time. However, the efficiency gain is primarily computational. The analytical depth confirms that while CorbuAI excels at rapid prototyping, the professional responsibility for regulatory compliance and structural safety remains with the human architect, framing the AI as a powerful collaborative assistant rather than an autonomous designer.

7. Conclusions

In this study, the analysis of the rationality, necessity, urgency, and applicability of the influence mechanism and adoption model of architectural design-oriented AI technologies leads to the determination of stakeholders, collected requirements, and screening requirements to establish the requirement boundaries of the minimum viable product. Subsequently, the platform’s layered architecture is developed, and discussions regarding the technology stack and key technologies are undertaken. This encompasses the construction of the technology stack model selection and combination model, the design of the separated front and back-end interface model, and the design of the distributed architecture and

relational database. Building on this foundation, the problems in the R&D process are analyzed, and CorbuAI, an architectural design platform based on Pix2pix-GAN-SDv2 multimodal AI technologies, is proposed. This consists of three parts: building floor plan generation, building facade representation generation, and building interior space style migration, each following three steps: data collection and pre-processing, fine-tuning of the pre-trained model, and evaluation of the model training. After training and debugging the models, the user-end interface and interaction are determined, and the pre-trained models are deployed to the cloud server to realize the minimum viable product. Finally, in the stage of practical case application and evaluation, real projects are selected to generate building floor plans, building facade representations, and building interior space styles. The results demonstrate that the AI-based architectural design platform can undergo subjective evaluation by designers and meet the requirements for the rapid generation of reference and intention diagrams in the conceptual scheme stage.

From a theoretical perspective, this study enriches the theoretical system of interdisciplinary combination of architectural design theory and computer AI technology and improves empirical research and dynamic simulation of the willingness to adopt AI technologies in the field of architectural design. Moreover, it explores implementation methods based on the combination of theory and practice for the analysis and research work about the requirements for “AI-based architectural design platforms.” The research results indicate that architectural design companies and architects can utilize the tool to conduct work in the conceptual stage of actual architectural design, thus enhancing the innovation of research methods in this field. From a practical standpoint, the research results hold great significance for promoting the development of the architectural design industry and exploiting the potential value of AI technologies in architectural design projects, which might lead the collaboration between academia and design firms to foster real-world pilot studies. The use of the “AI-based architectural design platform” can enhance the efficiency of architects and designers, strengthen design capabilities, and serve as a “design assistant” for architects by enabling machines to assist in design through algorithms. Therefore, this study demonstrates product feasibility with practical significance.

Unlike generic AI generative models that often prioritize visual aesthetics over architectural logic, this study introduces CorbuAI, a framework uniquely tailored to the functional and regulatory constraints of the Chinese residential context. By bridging the gap between 2D topological layout and 3D-oriented elevation rendering, our work provides a novel solution to the fragmented workflows identified in previous literature.

The implementation results demonstrate a high degree of precision, with a spatial allocation accuracy of 0.88 and a mean pHash similarity of 0.82 in facade generation. These metrics confirm that the proposed multimodal framework significantly enhances the reliability of AI-generated architectural designs. While CorbuAI exhibits professional-grade performance in standardized residential design, a critical reflection reveals that the system remains dependent on the quality and diversity of its training data. The potential for “algorithmic bias” towards specific architectural styles and the challenge of handling highly irregular urban plots remain areas for further refinement.

Based on our findings, we offer the following practical conclusions, first, the multimodal AI frameworks can reduce the schematic design cycle by over 90% in standardized residential projects. To ensure functional viability, designers should utilize AI as an iterative prototyping tool rather than a final output generator, maintaining a “human-in-the-loop” approach to refine structural and social nuances. Future research will explore the integration of 3D geometric data to transition from image-based generation to volumetric BIM-compatible models. Additionally, we aim to expand the dataset to include diverse climate-responsive facade typologies to enhance the framework’s global applicability.

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

Funding: This research was supported by Shandong Provincial Association for Science of Arts & Culture (Grant No. L2025Y10190012). This research was supported by Shandong Philosophy and Social Science Innovation and Development Research Association (Grant No. Z-FZY25024). This research was funded by Qingdao Huanghai University Doctoral Research Foundation Project (Grant No. 2025boshi14). This research was funded by the Open Research Project of the Ministry of Education Base for the Inheritance of Chinese Excellent Traditional Culture (Han Embroidery), Wuhan Textile University, 2025 (Grant No. HXJD2025A19).

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

Acknowledgments: We thank all the support from the research team members, as well as the intangible help from their affiliations. The authors have reviewed and edited the output and take full responsibility for the content of this publication. The authors wish to express their thanks to the editors and reviewers for their constructive comments. This research was inspired by the transformative power of artificial intelligence and its profound impact on the field of architectural design.

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

Abbreviations

The following abbreviations are used in this manuscript:

Nomenclature

Abbreviation	Full name
AIAD	Artificial Intelligence-based Architectural Design
CORBU	COmputer-generated Residential BUilding
SD	Stable Diffusion
AIGC	AI-Generated Content
GAN	Generative Adversarial Network
CLIP	Contrastive Language-Image Pre-training
MVP	Minimum Viable Product
CAD-CAM	Computer-Aided Design And Manufacturing
MDO	Medium Density Overlay
SSIEA	Stable Stage Island Evolutionary Algorithm
MMML	Multimodal Machine Learning
IPD	Integrated Project Delivery
API	Application Programming Interface
GCP	Google Cloud Platform
GPU	Graphics Processing Unit
TPU	Tensor Processing Unit
CDN	Content Distribution Network
RIT	Rakuten Institute of Technology
CVC	Computer Vision Center
cGAN	Conditional Generative Adversarial Network
FSL	Functional Segmentation Layout
BFP	Building Floor Plan
CMP	Center for Machine Perception
RAM	Random Access Memory
DM	Diffusion Model
UI	User Interface

References

1. Fang, C.; Lai, Y.; Wen, Z. *Research on Key Technologies of Intelligent Integrated System for Architectural Design Enterprises*; IOP Publishing: Bristol, UK, 2020; Volume 455, p. 012209. [\[CrossRef\]](#)
2. Pena, M.L.C.; Carballal, A.; Rodríguez-Fernández, N.; Santos, I.; Romero, J. Artificial Intelligence Applied to Conceptual Design. A Review of Its Use in Architecture. *Autom. Constr.* **2021**, *124*, 103550. [\[CrossRef\]](#)
3. Ali, A.K.; Lee, O.J. Facade Style Mixing Using Artificial Intelligence for Urban Infill. *Architecture* **2023**, *3*, 258–269. [\[CrossRef\]](#)
4. Harapan, A.; Indriani, D.; Rizkiya, N.; Azbi, R. Artificial Intelligence in Architectural Design. *Int. J. Des. (INJUDES)* **2021**, *1*, 1–6. [\[CrossRef\]](#)
5. Narahara, T. Collective Construction Modeling and Machine Learning: Potential for Architectural Design. In Proceedings of the 35th eCAADe Conference, Rome, Italy, 20–22 September 2017.
6. As, I.; Pal, S.; Basu, P. Artificial Intelligence in Architecture: Generating Conceptual Design via Deep Learning. *Int. J. Archit. Comput.* **2018**, *16*, 306–327. [\[CrossRef\]](#)
7. Davis, D. The MacLeamy Curve. Available online: <https://www.danieldavis.com/macleamy/> (accessed on 12 December 2022).
8. MacLeamy, P. *Collaboration, Integrated Information and the Project Lifecycle in Building Design, Construction and Operation*; WP-1202; The Construction Users Roundtable: Cincinnati, OH, USA, 2004.
9. Płoszaj-Mazurek, M.; Ryńska, E.; Grochulska-Salak, M. Methods to Optimize Carbon Footprint of Buildings in Regenerative Architectural Design with the Use of Machine Learning, Convolutional Neural Network, and Parametric Design. *Energies* **2020**, *13*, 5289. [\[CrossRef\]](#)
10. Moghtadernejad, S.; Mirza, M.S.; Chouinard, L.E. Facade Design Stages: Issues and Considerations. *J. Archit. Eng.* **2019**, *25*, 04018033. [\[CrossRef\]](#)
11. Chai, C. Housing Construction Industry: What Are Its Sub-Sectors? What Is Their Size? Available online: <https://www.construction21.org/china/articles/h/%E6%88%BF%E5%B1%8B%E5%BB%BA%E7%AD%91%E4%B8%9A%E6%9C%89%E5%93%AA%E4%BA%9B%E7%BB%86%E5%88%86%E8%A1%8C%E4%B8%9A-%E8%A7%84%E6%A8%A1%E5%A4%9A%E5%A4%A7.html> (accessed on 22 November 2022).
12. Sacks, R.; Girolami, M.; Brilakis, I. Building Information Modelling, Artificial Intelligence and Construction Tech. *Dev. Built Environ.* **2020**, *4*, 100011. [\[CrossRef\]](#)
13. Warszawski, A.; Sangrey, D.A. Robotics in Building Construction. *J. Constr. Eng. Manag.* **1985**, *111*, 260–280. [\[CrossRef\]](#)
14. Rahbar, M. Application of Artificial Intelligence in Architectural Generative Design. In Proceedings of the 36th eCAADe Conference, Lodz, Poland, 19–21 September 2018.
15. Wortmann, T.; Nannicini, G. Black-Box Optimisation Methods for Architectural Design. In *Proceedings of the 21st International Conference on Computer-Aided Architectural Design Research in Asia (CAADRIA 2016)*; The Association for Computer-Aided Architectural Design Research in Asia (CAADRIA): Hong Kong, China, 2016. [\[CrossRef\]](#)
16. Cichocka, J.M.; Browne, W.N.; Rodriguez, E. Optimization in the Architectural Practice. In *Proceedings of the 22nd International Conference on Computer-Aided Architectural Design Research in Asia (CAADRIA 2017)*; The Association for Computer-Aided Architectural Design Research in Asia (CAADRIA): Hong Kong, China, 2017; pp. 387–396.
17. Han, S.-K.; Shin, D.-Y. A Study on Architectural Image Generation Using Artificial Intelligence Algorithm—A Fundamental Study on the Generation of Due Diligence Images Based on Architectural Sketch. *J. KIBIM* **2021**, *11*, 54–59. [\[CrossRef\]](#)
18. Zheng, H.; Yuan, P.F. A Generative Architectural and Urban Design Method through Artificial Neural Networks. *Build. Environ.* **2021**, *205*, 108178. [\[CrossRef\]](#)
19. Oxman, R.E. Expert System for Generation and Evaluation in Architectural Design. Ph.D. Thesis, Technion—Israel Institute of Technology, Haifa, Israel, 1988. Available online: <http://papers.cumincad.org/cgi-bin/works/paper/45b7> (accessed on 30 January 2026).
20. Zandi-Nia, A. TOPGENE: An Artificial Intelligence Approach to a Design Process. Ph.D. Thesis, Delft University of Technology, Delft, The Netherlands, 1992.
21. Lin, S.-H.; Gerber, D.J. Evolutionary Energy Performance Feedback for Design: Multidisciplinary Design Optimization and Performance Boundaries for Design Decision Support. *Energy Build.* **2014**, *84*, 426–441. [\[CrossRef\]](#)
22. Østergård, T.; Jensen, R.L.; Maagaard, S.E. Building Simulations Supporting Decision Making in Early Design—A Review. *Renew. Sustain. Energy Rev.* **2016**, *61*, 187–201. [\[CrossRef\]](#)
23. Currie, K.; Tate, A. O-Plan: The Open Planning Architecture. *Artif. Intell.* **1991**, *52*, 49–86. [\[CrossRef\]](#)
24. Wang, S. *Application Research of Artificial Intelligence Technology in Landscape Architectural Art Design*; Springer: Berlin/Heidelberg, Germany, 2020; pp. 341–348. [\[CrossRef\]](#)
25. Andersen, K.; Forr, T. *The State of Construction Technology*; Jones Lang LaSalle IP, Inc.: Chicago, IL, USA, 2018.
26. Bhatt, M.; Suchan, J.; Schultz, C.; Kondyli, V.; Goyal, S. Artificial Intelligence for Predictive and Evidence Based Architecture Design. *Proc. AAAI Conf. Artif. Intell.* **2016**, *30*, 4349–4350. [\[CrossRef\]](#)

27. Wang, L.; Janssen, P.; Ji, G. SSIEA: A Hybrid Evolutionary Algorithm for Supporting Conceptual Architectural Design. *AI EDAM* **2020**, *34*, 458–476. [\[CrossRef\]](#)
28. BAŞARIR, L. Modelling AI in Architectural Education. *Gazi Univ. J. Sci.* **2022**, *35*, 1260–1278. [\[CrossRef\]](#)
29. Frazer, J. *An Evolutionary Architecture*; Architectural Association: London, UK, 1995.
30. Klir, G.J. *Architecture of Systems Problem Solving*; Springer Science & Business Media: Berlin/Heidelberg, Germany, 2013. [\[CrossRef\]](#)
31. Nauata, N.; Chang, K.-H.; Cheng, C.-Y.; Mori, G.; Furukawa, Y. House-GAN: Relational Generative Adversarial Networks for Graph-Constrained House Layout Generation. In *Proceedings of the Computer Vision—ECCV 2020*; Vedaldi, A., Bischof, H., Brox, T., Frahm, J.-M., Eds.; Springer International Publishing: Cham, Switzerland, 2020; pp. 162–177. [\[CrossRef\]](#)
32. Evans, E.; Evans, E.J. *Domain-Driven Design: Tackling Complexity in the Heart of Software*; Addison-Wesley Professional: Boston, MA, USA, 2004.
33. Ronneberger, O.; Fischer, P.; Brox, T. U-Net: Convolutional Networks for Biomedical Image Segmentation. In *Proceedings of the Medical Image Computing and Computer-Assisted Intervention—MICCAI 2015: 18th International Conference, Munich, Germany, 5–9 October 2015*; Proceedings, Part III 18; Springer: Berlin/Heidelberg, Germany, 2015; pp. 234–241. [\[CrossRef\]](#)
34. Isola, P.; Zhu, J.-Y.; Zhou, T.; Efros, A.A. Image-to-Image Translation with Conditional Adversarial Networks. In *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*; IEEE: New York, NY, USA, 2017; pp. 1125–1134. [\[CrossRef\]](#)
35. Saffron Team. Why Should Your Website Have Separate Backend and Frontend? 2021. Available online: <https://www.saffrontech.net/blog/why-should-your-website-have-separate-backend-tech-and-frontend-tech> (accessed on 30 January 2026).
36. Rakuten Institute of Technology | Rakuten Technology Research Institute. Rakuten Data Release. Available online: https://rit.rakuten.com/data_release/ (accessed on 30 January 2026).
37. de las Heras, L.-P.; Terrades, O.R.; Robles, S.; Sánchez, G. Robles CVC-FP: Database for Structural Floor Plan Analysis. *Int. J. Doc. Anal. Recognit. (IJ DAR)* **2015**, *18*, 15–30. [\[CrossRef\]](#)
38. Wan, D.; Zhao, X.; Lu, W.; Li, P.; Shi, X.; Fukuda, H. A Deep Learning Approach toward Energy-Effective Residential Building Floor Plan Generation. *Sustainability* **2022**, *14*, 8074. [\[CrossRef\]](#)
39. Chaillou, S. Archigan: Artificial Intelligence x Architecture. In *Architectural Intelligence*; Springer: Berlin/Heidelberg, Germany, 2020; pp. 117–127. [\[CrossRef\]](#)
40. Rodrigues, R.C.; Imagawa, M.K.; Koga, R.R.; Duarte, R.B. Big Data vs Smart Data on the Generation of Floor Plans with Deep Learning. In *Proceedings of the XXV International Conference of the Iberoamerican Society of Digital Graphics*; Blucher Design Proceedings: São Paulo, Brazil, 2021.
41. Zhao, C.-W.; Yang, J.; Li, J. Generation of Hospital Emergency Department Layouts Based on Generative Adversarial Networks. *J. Build. Eng.* **2021**, *43*, 102539. [\[CrossRef\]](#)
42. Zheng, H.; An, K.; Wei, J.; Ren, Y. Apartment Floor Plans Generation via Generative Adversarial Networks. In *Proceedings of the 25th International Conference on Computer-Aided Architectural Design Research in Asia (CAADRIA)*; The Association for Computer-Aided Architectural Design Research in Asia (CAADRIA): Hong Kong, China, 2020; pp. 601–610.
43. Mirza, M.; Osindero, S. Conditional Generative Adversarial Nets. *arXiv* **2014**, arXiv:1411.1784. [\[CrossRef\]](#)
44. Brownlee, J. How to Develop a Pix2Pix GAN for Image-to-Image Translation. 2019. Available online: <https://machinelearningmastery.com/how-to-develop-a-pix2pix-gan-for-image-to-image-translation/> (accessed on 30 January 2026).
45. Tyleček, R.; Šára, R. Spatial Pattern Templates for Recognition of Objects with Regular Structure. In *Proceedings of the Pattern Recognition: 35th German Conference, GCPR 2013, Saarbrücken, Germany, 3–6 September 2013*; Proceedings 35; Springer: Berlin/Heidelberg, Germany, 2013; pp. 364–374. [\[CrossRef\]](#)
46. Ding, K.; Chen, S.; Meng, F. A Novel Perceptual Hash Algorithm for Multispectral Image Authentication. *Algorithms* **2018**, *11*, 6. [\[CrossRef\]](#)
47. Li, Z.; Xiang, J.; Wang, L.; Zhang, N.; Pan, R.; Gao, W. Yarn-Dyed Fabric Image Retrieval Using Colour Moments and the Perceptual Hash Algorithm. *Fibres Text. East. Eur.* **2019**, *5*, 39–46. [\[CrossRef\]](#)
48. Weng, L.; Preneel, B. A Secure Perceptual Hash Algorithm for Image Content Authentication. In *Proceedings of the Communications and Multimedia Security: 12th IFIP TC 6/TC 11 International Conference, CMS 2011, Ghent, Belgium, 19–21 October 2011*; Proceedings 12; Springer: Berlin/Heidelberg, Germany, 2011; pp. 108–121. [\[CrossRef\]](#)
49. Zhou, X.; Schmucker, M.; Brown, C.L. Perceptual Hashing of Video Content Based on Differential Block Similarity. In *Proceedings of the Computational Intelligence and Security: International Conference, CIS 2005, Xi'an, China, 15–19 December 2005*; Proceedings, Part II; Springer: Berlin/Heidelberg, Germany, 2005; pp. 80–85. [\[CrossRef\]](#)
50. Sohl-Dickstein, J.; Weiss, E.; Maheswaranathan, N.; Ganguli, S. Deep Unsupervised Learning Using Nonequilibrium Thermodynamics. In *Proceedings of the International Conference on Machine Learning*; MLResearchPress: Cambridge, MA, USA, 2015; pp. 2256–2265.
51. Stenbit, I.; Chollet, F.; Wood, L. Keras Documentation: A Walk Through Latent Space with Stable Diffusion. Available online: https://keras.io/examples/generative/random_walks_with_stable_diffusion/ (accessed on 15 March 2023).

52. Zhou, B.; Zhao, H.; Puig, X.; Xiao, T.; Fidler, S.; Barriuso, A.; Torralba, A. Semantic Understanding of Scenes through the Ade20k Dataset. *Int. J. Comput. Vis.* **2019**, *127*, 302–321. [[CrossRef](#)]
53. Jain, J.; Li, J.; Chiu, M.; Hassani, A.; Orlov, N.; Shi, H. OneFormer: One Transformer to Rule Universal Image Segmentation. *arXiv* **2022**, arXiv:2211.06220. [[CrossRef](#)]
54. Kirillov, A.; Mintun, E.; Ravi, N.; Mao, H.; Rolland, C.; Gustafson, L.; Xiao, T.; Whitehead, S.; Berg, A.C.; Lo, W.-Y.; et al. Segment Anything. *arXiv* **2023**, arXiv:2304.02643.
55. Gal, R.; Alaluf, Y.; Atzmon, Y.; Patashnik, O.; Bermano, A.H.; Chechik, G.; Cohen-Or, D. An Image Is Worth One Word: Personalizing Text-to-Image Generation Using Textual Inversion. *arXiv* **2022**, arXiv:2208.01618.
56. Cheng, S.-I.; Chen, Y.-J.; Chiu, W.-C.; Tseng, H.-Y.; Lee, H.-Y. Adaptively-Realistic Image Generation from Stroke and Sketch with Diffusion Model. In *IEEE/CVF Winter Conference on Applications of Computer Vision*; IEEE: New York, NY, USA, 2023; pp. 4054–4062. [[CrossRef](#)]
57. Voynov, A.; Aberman, K.; Cohen-Or, D. Sketch-Guided Text-to-Image Diffusion Models. *arXiv* **2022**, arXiv:2211.13752.
58. Chaillou, S. AI and Architecture: An Experimental Perspective. In *The Routledge Companion to Artificial Intelligence in Architecture*; Routledge: Abingdon-on-Thames, UK, 2021; pp. 420–441. [[CrossRef](#)]
59. Rombach, R.; Blattmann, A.; Lorenz, D.; Esser, P.; Ommer, B. High-Resolution Image Synthesis with Latent Diffusion Models. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition*; IEEE: New York, NY, USA, 2022; pp. 10684–10695. [[CrossRef](#)]

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