1-A-5-양-5
논문일반(KCI)

LoRA와 ControlNet을 활용한 Stable Diffusion 기반 건축 입면 생성 방법과 평가에 관한 연구

2025. 10.

과 제 명	인공지능 기반의 건축설계 자동화 기술개발		
주 관 기 관	경북대학교 산학협력단		
총 연 구 기 간	2021. 04 . 01 - 2025. 12 . 31(4년 9개월)		
해당연도(3차년)	2025. 01 . 01 - 2025. 12 . 31(1년)		
구 성 기 술 명	구성기술 1	정형 건축물의 계획설계 지원자동화 기술개발	
세 부 과 제 명	1-A	지능형 공간계획 및 계획설계 제안 기술개발	
공 동 연 구 기 관	경북대학교 산학협력단, ㈜코스펙이노랩		
연 구 기 관	경북대학교 산학협력단	**연구책임자**	추승연

ISSN 2733-6239(Print) ISSN 2733-6247(Online)

https://doi.org/10.5659/JAIK.2025.41.8.85

LoRA와 ControlNet을 활용한 Stable Diffusion 기반 건축 입면 생성 방법과 평가에 관한 연구

A Generation Method and Evaluation of Architectural Facade Design Using Stable Diffusion with LoRA and ControlNet

박 정 민* 홍 순 민** 추 승 연***
Park, Jungmin Hong, Soonmin Choo, Seungyeon

* 경북대 건축학과 석사과정, Master's Course Student, School of Architecture, Kyungpook National University, Republic of Korea
** 경북대 건축학과 박사수료, Ph.D. Candidate, School of Architecture, Kyungpook National University, Republic of Korea
*** 경북대 건축학부 교수, Professor, School of Architecture, Kyungpook National University, Republic of Korea
(Corresponding author : choo@knu.ac.kr)

Abstract

This study proposes a novel approach for generating architectural facade images by combining the Stable Diffusion model with Low-Rank Adaptation (LoRA) and ControlNet. The standard Stable Diffusion model faces limitations in accurately reflecting architectural elements and material characteristics, which are critical in the design process. To address these challenges, this research integrates domain-specific fine-tuning using LoRA and precise shape control through ControlNet. LoRA allows the model to effectively learn architectural styles and details, ensuring better representation of essential design elements such as windows, balconies, and facade materials. Meanwhile, ControlNet utilizes Canny Edge and Depth Map information to enhance shape accuracy and spatial consistency, enabling more reliable image generation. The generated images were evaluated through Contrastive Language-Image Pretraining (CLIP) scores for quantitative analysis and GPT-4V-based qualitative evaluation, providing a more comprehensive understanding of architectural coherence and visual fidelity. The GPT-4V assessment offered insights into spatial relationships, contextual relevance, and material expression that are not easily captured through traditional metrics. This combined approach reduces the repetitive manual adjustments commonly required in text-prompt-based image generation and facilitates a more intuitive and efficient design process during the early stages of architectural planning. By improving control over detailed architectural features, the proposed method contributes to the automation of facade design, offering significant potential for real-world applications in architectural design and visualization. Future research will focus on expanding the dataset to include diverse architectural styles and validating its practical application in design and construction.

키워드 : 스테이블 디퓨전, 건축 매스, 입면 디자인, 생성형 AI, 로라, 컨트롤넷
Keywords : Stable Diffusion, Architecture massing, Facade Design, Generative AI, LoRA, ControlNet

1. 서 론

1.1 연구 배경 및 목적

건축 입면 디자인은 건축물의 미적 가치와 기능성을 결정하는 핵심 요소로 인식된다(Ching, 2023). 건축물의 외관은 전체적인 인상뿐만 아니라 사용자 경험에도 큰 영향을 미치며, 이를 통해 건물의 상징성과 활용도를 동시에 향상시킬 수 있어, 다양한 디자인 대안들이 필요하다. 그러나, 전통적인 CAD 및 3D 모델링 기반의 디자인 방식은 고해상도 렌더링과 반복적인 수정 작업으로 인해 시간과 인력 소모가 크다는 한계가 존재하며, 특히 대규모 프로젝트나 복합 용도 건축물의 경우 다양한 디자인 대안을 신속하게 도출하는 데 어려움이 있다.

이에 따라 최근 인공지능(AI)의 발전과 함께 생성형 AI를 활용한 이미지 생성 기법이 건축 디자인 분야에 도입되기 시작하였다. 그 중, Generative Adversarial Networks(GAN)과 Diffusion 모델은 대량의 데이터를 학습하여 창의적이고 현실적인 이미지 생성이 가능하게 하며, 특히 Diffusion 모델은 노이즈를 단계적으로 제거하는 방식으로 안정적인 학습 및 고품질 이미지 생성을 실현한다(Ho, Jain, & Abbeel, 2020). 이 중 Stable Diffusion(이후 SD로 표기) 모델은 텍스트 기반의 이미지 생성 기능을 제공하여, 사용자가 입력한 텍스트 프롬프트에 따라 다양한 디자인 대안을 신속하게 시각화할 수 있는 장점을 가지고 있다. 일반적으로

본 연구는 국토교통부/국토교통과학기술진흥원의 2025년도 지원으로 수행되었음. 과제번호 RS-2021-KA163269

학습된 모델은 창, 문, 기둥, 슬래브 등의 배치가 일관적으로 유지되어야 하는 구조적 일관성을 충분히 반영하지 못하는 한계가 있다. 게다가 건축물의 크기에 대한 비례와 스케일, 실제 사용 가능한 방식으로 건축물의 요소들이 배치되는 기능적 요소 등, 건축 도메인 특유의 세부적 요구사항 역시 충분히 고려하지 못한다. 따라서 본 연구는 SD 기반 이미지 생성 모델을 건축 입면 디자인에 최적화하기 위한 구체적인 방법을 제안하고자 한다. 미세조정 기법 중 하나인 Low-Rank Adaptation(LoRA)와 SD의 기능 중 하나인 img2img를 활용하여 건축 도메인의 전문적 특성을 효과적으로 반영하고자 한다. 또한, ControlNet 기능을 통해 Canny Edge 및 Depth Map 정보를 통합함으로써 세부 형상 제어를 구현하고자 한다. 이를 통해 설계 초기 단계에서 다양한 디자인 대안을 신속하고 체계적으로 도출할 수 있으며, 기존 방식에서 소요되던 인적 자원과 시간을 절감하는데 기여하고자 한다.

1.2 연구의 범위 및 방법

본 연구는 현대 건축물의 입면 이미지를 대상으로 하며, 특히 유리 파사드, 금속, 철근 콘크리트, 복합 패널 등 다양한 외장 재료가 혼합된 중·고층 규모의 주거, 오피스 및 복합 용도 건축물 외관 이미지를 연구 데이터로 활용한다. 또한 SD SDXL모델(SDXL모델), LoRA기법, img2img 기능, ControlNet 기능을 통해 건축 입면의 세부 구성 요소를 정밀하게 제어하고자 한다. 평가모델로는 건축 도메인을 평가할 수 있도록 미세조정된 CLIP모델과 이미지와 텍스트를 동시에 이해하고 해석할 수 있는 GPT-4V와 전문가 평가를 이용해 생성된 건축이미지를 평가한다.

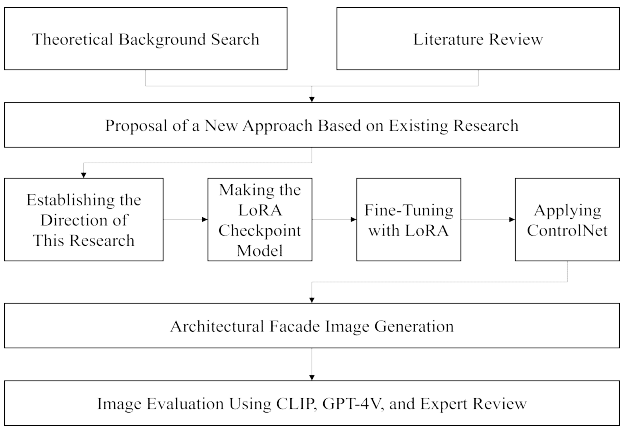


Figure 1. Research flow

연구의 방법은 Figure 1과 같다.

첫 번째로 Diffusion 모델, LoRA, ControlNet과 같은 생성형 AI 기술에 대한 이론적 배경과 선행연구를 살펴본 뒤, 이를 토대로 본 연구의 방향을 수립한다.

두 번째로 건축 입면 데이터셋을 바탕으로 구축한 LoRA 체크포인트 모델을 구축한 뒤, SD에 적용하여 모델을 미세조정한다. 그리고 Canny,Depth 등의 정보를 활용한 ControlNet을 통해 건축적 요소를 정밀 제어하도록 알고리즘을 구성한다.

세 번째로 앞서 구축한 기술들을 이용해 다양한 형태의 건축 입면 이미지를 생성한다. 네 번째로 생성된 이미지는 이미지 평가에 사용되는 Contrastive Language-Image Pretraining(CLIP) 모델[1]과 GPT-4V[2], 전문가 평가를 활용해 평가한다.

2. 이론적 배경 및 선행연구 고찰

2.1 생성형 AI 모델의 이론적 배경

생성형 AI는 대규모 데이터를 학습하여 새로운 이미지를 생성하는 기술로, 대표적으로 GAN과 Diffusion 모델이 있다. GAN은 생성자와 판별자의 경쟁적 학습을 통해 현실적인 이미지를 생성하는 모델이다(Goodfellow et al., 2014). 그러나 학습 과정의 불안정성과 다양한 이미지를 생성하지 못하는 모드 붕괴(mode collapse)가 자주 발생해 해상도나 세부 표현이 필요한 작업에 한계를 보였다(Zhao & Li, 2023).

Zhang et al.(2018)은 이러한 문제를 해결하기 위해 텍스트 정보를 바탕으로 저해상도 이미지를 생성한 후, 점진적으로 고해상도 이미지로 변환하는 다단계 GAN 구조를 도입하였다. 또한 Xu et al.(2018)은 단어 단위의 텍스트 정보를 활용하여 이미지의 특정 영역을 강조하는 어텐션 메커니즘을 적용하였다. 이러한 연구들을 통해 텍스트에서 이미지로 변환하는 모델의 성능이 개선되었으나, 텍스트와 이미지 간의 정밀성 문제는 해결되지 않았다.

Radford et al.(2021)은 CLIP모델을 발표하면서 텍스트와 이미지 간의 관계를 학습하는 새로운 방식을 제안하였다. CLIP을 기반으로 한 DALL·E 모델은 GAN을 사용하지 않고도 텍스트 기반 이미지 생성을 가능하게 하였다. 그러나 DALL·E 및 기존 GAN 기반 모델들은 여전히 고해상도 이미지 생성의 어려움과 세부 요소 표현의 한계를 가지고 있었다.

Diffusion 모델은 이러한 한계를 개선한 모델로서 GAN과 달리 노이즈를 점진적으로 제거하는 방식으로 이미지를 복원하며, 안정적인 학습과 고해상도 표현에 강점이 있다(Ho, Jain, & Abbeel, 2020). SD, DALL·E 등은 텍스트 조건부 이미지 생성에 Diffusion 방식을 적용해, GAN의 단점을 보완하고 정밀한 제어와 높은 해상도를 가능케 했다. 특히 건축 디자인에서는 건물 외관의 복합 요소를 고려해야 하므로, Diffusion의 안정성과 풍부한 표현력이 유리하다. 따라서 본 연구에서는 Diffusion 모델의 안정성과 고해상도 표현 능력을 활용하여 건축 디자인에 적합한 이미지 생성 방법을 모색하고자 한다.

1) OpenAI에서 2021년 개발한 인공지능 모델로, 이미지와 텍스트를 같은 벡터 공간에 대응시켜 텍스트 설명과 이미지 간의 유사성을 정량적으로 측정할 수 있으며, 이미지 분류, 검색, 평가 등 다양한 분야에 활용되고 있다.

2) OpenAI가 개발한 멀티모달(Multimodal) 인공지능 모델로, 텍스트뿐 아니라 이미지까지 이해하고 분석할 수 있다. 텍스트와 이미지 간 맥락을 파악하여 이미지 묘사, 시각 정보 분석 및 평가 등의 다양한 시각적 작업 수행이 가능하다.

2.2 건축 이미지 생성에 관한 선행연구

최근 건축 분야에서도 AI 기반 이미지 생성에 관한 연구가 많이 진행되고 있다. 크게 Diffusion 모델, img2img, LoRA, ControlNet 4가지 유형으로 분류 할 수 있다.

Diffusion 관련 연구로는 Denoising Diffusion Probabilistic Models(DDPMs)을 통해 고품질·고해상도 이미지 생성이 가능함을 보였다(Ho, Jain, & Abbeel, 2020). Rombach et al.(2022)은 Latent Diffusion Models를 통해 대규모 이미지 생성 시 연산 효율성을 확보하면서도 재질과 형태 같은 세밀한 요소까지 표현할 수 있는 방법론을 선보였다. 또한 Saharia et al.(2022)의 Photorealistic Text-to-Image Diffusion Models은 텍스트만으로 사실적인 이미지 구현이 가능함을 증명하였다.

Img2img관련 연구로는 Meng et al.(2021)이 SDEdit을 활용해 노이즈 추가-역확산 과정을 통해 구조적 특징을 자연스럽게 보존하는 편집 기법을 제안하였다. 이를 바탕으로 Wang & Zhang(2024)은 SD 기반 img2img 기법을 이용해 텍스트 조건과 사용자 지정 마스크를 결합하여 건축 입면의 특정 영역을 효과적으로 편집하는 방법을 제시했으며, Lee & Ko(2023)는 SD 기반 AI 이미지 생성기를 활용한 실험을 수행하였다.

Table 1. Analysis of related research

Research	Author	Contents
Diffusion Models	Ho, Jain, & Abbeel, (2020)	Introduced DDPMs for stable high-quality image generation through progressive restoration.
	Rombach et al. (2022)	Proposed Latent Diffusion Models for efficient large-scale image generation with detailed expression.
	Saharia et al. (2022)	Demonstrated photorealistic image generation from text-based inputs using Diffusion Models.
Img2img	Meng et al. (2021)	Presented structure-preserving image editing with stochastic differential equations (SDEdit).
	Lee & Ko (2023)	Applied Img2img for architectural refinement, converting conceptual sketches into detailed designs.
	Wang & Zhang (2024)	Proposed blended diffusion with user-defined masks for selective facade editing.
LoRA	Hu et al. (2021)	Introduced LoRA for domain-specific adaptation while reducing model size and training cost.
	Jo & Lee (2023)	Applied LoRA for creating facade designs reflecting regional identity in architecture.
	Yoo & Lee (2023)	Used LoRA to learn architect styles and visualize single-family house exteriors.
ControlNet	Zhang, Rao, & Agrawala (2023)	Developed ControlNet for fine-grained shape control using edge, depth, and sketch inputs.
	Ma & Zheng (2023)	Combined CMP Facades with LoRA and ControlNet for precise architectural facade generation.

LoRA관련 연구로는 Hu et al.(2021)이 LoRA 기법을 제안하여 대규모 모델 재학습 없이 도메인 특화 정보를 효과적으로 반영할 수 있도록 하였다. Jo & Lee(2023)는 이를 활용해 지역적 특성이 반영된 건물 입면을 생성하였고, Yoo & Lee(2023)는 건축가 스타일을 LoRA로 학습하여 단독 주택 외관을 효과적으로 시각화하였다.

ControlNet 관련 연구로는 Zhang, Rao, & Agrawala(2023)이 ControlNet 기법을 도입하여 텍스트-이미지 확산 모델에 공간적 조건 제어 기능을 추가함으로써 이미지 생성의 정밀도를 높였다. 이를 확장하여 Ma & Zheng(2024)는 CMP Facades 데이터셋을 활용해 LoRA와 ControlNet을 결합, 창호·발코니·입구 등의 세부요소를 정밀하게 구현하는 방법을 제안했으며, 이를 통해 건축 입면 디자인의 품질을 향상시킬 수 있음을 입증하였다.

그러나, 기존 선행연구에서 Diffusion 모델로 고해상도 건축 이미지를 구현했음에도, 세부 요소를 정교하게 제어하는 데에는 한계가 있었다. Img2img 기법은 기본 구조를 유지하면서 이미지를 변형할 수 있지만, 특정 디자인 요소를 원하는 위치에 반영하기가 쉽지 않았다. 또한, LoRA와 ControlNet을 통해 건축 이미지를 보다 정밀하게 표현하려는 시도도 있었으나, 복잡한 형태와 다층적 외장 요소를 완벽히 제어하기에는 아직 부족하였다. Ma & Zheng(2023)의 연구는 CMP Facades 데이터셋을 기반으로 Stable Diffusion에 LoRA를 적용하고, 다양한 ControlNet 조건(Canny, Depth, Segmentation 등)을 실험하여 텍스트 기반의 건축 입면 이미지 생성 가능성을 제시하였다. 그러나 해당 연구는 주로 역사적 정면 중심의 CMP Facades 데이터셋에 기반하여 다양한 재료와 입체적 구조가 반영된 현대 건축물에 대한 표현력은 제한적이었고, 이미지 생성 결과에 대한 체계적인 정량 평가 또한 수행되지 않았다.

따라서, 본 연구에서는 img2img, LoRA, ControlNet 기법을 함께 활용하여 건축 도메인에 적합한 이미지 생성 방법을 개발하고자 한다. 이를 위해 현대 건축물의 외장 재료와 구조적 특성을 반영한 데이터셋을 구축하고, LoRA를 적용하여 SD 모델을 세부적으로 조정한다. 이후, SD의 img2img 기능을 LoRA 및 ControlNet과 결합하여 창호, 발코니, 출입구 등의 세부 요소를 정밀하게 제어함으로써, 기존 모델의 한계를 보완하고 건축적 맥락과 스타일을 구현할 수 있는 이미지 생성 방안을 제안한다.

생성 결과에 대해 CLIP Score와 GPT-4V 기반 평가 항목을 활용하여 정량적 성능을 분석하고, 전문가 평가를 병행함으로써 정성적 타당성 역시 확보하며, 생성 이미지의 건축적 일관성과 표현 타당성을 다층적으로 검증하였다는 점에서, 단순 스타일 재현을 넘어서 실질적 설계 도구로서의 활용 가능성을 평가하였다는 데에 차별성이 있다.

3. LoRA-ControlNet 기반 건축 입면 생성 방법

3.1 건축 입면 이미지의 생성 전략

건축 입면 이미지 생성 프로세스는 Figure 2와 같다.

첫째, LoRA 생성 단계에서는 현대 건축물 입면 이미지를 수집하고 전처리한 후, 건축적 특징을 반영한 태그를 부여해 데이터베이스를 구축한 뒤 LoRA 체크포인트 모델을 생성한다.

둘째, 건축 도메인에 알맞은 프롬프트들을 구축하고, 이를 원하는 입면의 특성에 맞추어 재구성한다.

셋째, img2img 기법을 이용해 이미지를 입력값으로 받고, LoRA 기법을 이용하여 이미지를 생성한다.

넷째, LoRA를 통해 생성한 이미지에 ControlNet을 적용시켜 이미지에 대한 상세한 조정과 특징을 구체화한다.

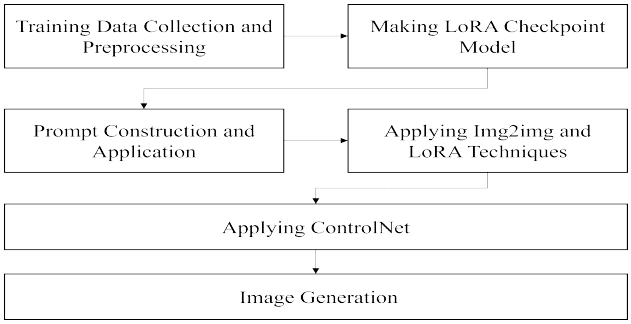


Figure 2. Flow of facade image generation

3.2 LoRA 학습을 통한 체크포인트 파일 생성

(1) 데이터 전처리 및 학습 데이터 베이스 구축

Table 2. Category for tagging

Main Category	Sub Category	Examples	Number of Images Tagged (Multiple Inclusion)
Architectural Style & Typology	Building Typology	Contemporary Office Building, Mixed-Use High-Rise, Single-Family Residence, Luxury Villa, Urban Apartment Complex	43
	Architectural Style	Sustainable Office Building, Modernist, Parametric Design, Minimalist Modern, Deconstructivist	52
	Form Characteristics	Biophilic Design, Modular Facade, Cantilevered Terraces, Geometric Facade Articulation, Glass Pavilion Concept	44
Material & Structural Elements	Exterior Materials & Structural Systems	Glass Curtain Wall, Prefabricated Concrete Panels, Aluminum Window Frames, Textured Stone Cladding, Exposed Concrete Panels	63
	Structural Framework	Reinforced Concrete Framework, Structural Steel Frame, Steel-Frame Core, Reinforced Concrete Shell, Reinforced Concrete Core	16
	Additional Structural Elements	Green Roof & Vertical Garden, Cantilevered Overhang, Glass Balustrades, Integrated Planters, Decorative Metal Screen	20

데이터베이스를 구축하기 위해 금속, 유리, 철근 콘크리트, 복합 패널 등 다양한 외장 재료를 사용하는 현대 건축물의 고해상도 이미지를 총 56장 수집하였다. 이미지는 Unsplash, Pexels 등 저작권이 명시적으로 해제된 공개 이미지 플랫폼(CC0 라이선스 기반)에서 확보하였으며, Google 검색은 참고 용도로만 활용하되, 사용된 이미지의 경우에는 별도의 라이선스 명시가 확인된 출처에 한하여 포함하였다. 각 플랫폼은 상업적·비상업적 목적의 이미지 사용을 명시적으로 허용하고 있으며, 해당 조건을 준수하여 연구 목적으로 활용하였다. 또한, 수집 과정에서 품질이 낮거나 왜곡이 심한 이미지는 데이터베이스에서 제외하였다.이후, 수집한 이미지는 1024×1024 픽셀로 크기를 조정하고, 노이즈 및 색상 편향을 제거 및 보정하는 전처리 과정을 진행한다. 전처리된 이미지는 Table 2와 같이 건축 양식 및 유형(Architectural Style & Typology)와 재료 및 구조 요소(Material & Structural Elements)를 기준으로 구축된 건축 입면 이미지에 알맞은 태그를 부여하는 태깅(tagging)[3] 작업을 수행하였다. 각 카테고리는 최소 3개 이상의 구체적이고 명확한 세부 요소로 구성되었으며, 의미가 중복되거나 모호한 태그는 배제하였다. 특히, 하나의 이미지에는 둘 이상의 카테고리 태그가 중첩될 수 있기 때문에, Table 2에 제시된 태그 분포는 포함된 이미지 수가 아닌, 태그의 포함 횟수(중복 포함 기준)를 기준으로 정리하였다. 전체 태깅 구조와 예시는 Figure 3과 같다.

Figure 3. Example of an architectural facade tag

(2) LoRA 체크포인트 파일 생성

구축한 현대 건축 입면 데이터셋을 활용하여, LoRA 파일 학습을 진행하였다. 학습에는 Kohya_ss 라이브러리를 사용하였으며, Batch Size, Epoch, Learning Rate, Optimizer 등 하이퍼파라미터를 다양하게 실험하면서 모델을 점진적으로 개선하였다. 학습 과정 중간마다 LoRA 체크포인트 파일을 생성하고, 손실값 변화를 Epoch 단위로 모니터링하며 모델의 수렴 여부와 과적합 발생 가능성을 점검하였다.

Figure 4. Loss graph

Figure 4은 LoRA 체크포인트 모델 학습 과정에서 평균 손실값 변화와 실시간 손실값 변화, Epoch별 손실값 변화를 나타내고, 이들은 손실값의 안정적 감소와 모델 수렴 과정을 보여준다. loss/average 그래프는 손실값이 점진적으로 줄어드는 추세를 나타내며, loss/current는 실시간 손실값이

3) 데이터나 객체에 특정한 레이블(label)이나 태그(tag)를 부여하여 이를 식별하고 분류하는 과정을 뜻한다. 태깅을 함으로써 SD이 데이터를 보다 정확하게 이해하고 예측할 수 있다.

변동 속에서 점차 안정화되는 모습을 보여준다. loss/epoch 그래프에서는 4~5 Epoch 이후 손실값이 꾸준히 감소하며, 모델이 효과적으로 학습되었음을 확인되어, 모델은 점진적으로 최적화되는 것을 확인할 수 있다.

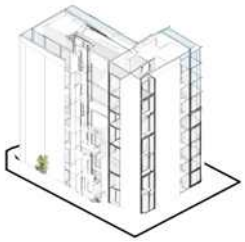


Figure 5. Midpoint output during training

이와 같이, LoRA 체크포인트 모델이 y축에서 0으로 수렴하면, 중간 생성 이미지가 Figure 5와 같이 시각적으로 현대 건축 이미지에 준하는 수준으로 나타난다. 이때 생성 이미지는 모듈형 파사드, 루버 등의 건축 요소 표현, 재료의 현실성, 형태 구성의 균형을 기준으로 검토되었으며, 문제가 발생한 경우 태그나 하이퍼파라미터를 조정하는 반복 학습을 수행하였다. 예를 들어, 루버 표현이 흐릿하게 나타나거나 유리와 금속 재료 간 경계가 불명확한 경우, 해당 프롬프트의 재료 관련 태그나 하이퍼파라미터를 조정하여 학습을 다시 진행하였다. 최종적으로 결정된 하이퍼파라미터는 Table 3과 같다.

Table 3. Model training parameters

Parameter	Value
Batch size	4
Epoch	8
Learning Rate	0.0001
Learning Rate Scheduler	Cosine
Optimizer	AdamW

3.3 건축 이미지 생성의 LoRA 기법 적용

(1) 건축 입면 프롬프트 구성

LoRA 모델을 활용한 건축 이미지 생성에서 텍스트 프롬프트의 구성은 최종 출력 이미지의 품질, 세부 표현, 스타일적 일관성에 큰 영향을 미친다. 프롬프트의 세부적인 구성 요소가 이미지 생성 과정의 방향성을 결정하며, 이를 정교하게 설계할수록 사실적이고 명확한 건축적 표현이 가능하다.

Table 4는 건축 이미지 생성을 위한 효과적인 프롬프트 구성을 위해 건물 유형(Building Type), 건축 형태 및 구조(Form & Structure), 외장 마감재(Exterior Materials), 창문 및 개방 요소(Windows & Openings), 조명 및 디테일(Lighting & Details), 건축 스타일(Architectural Style), 건축 분위기 및 환경(Atmosphere & Context)의 총 7가지 주요 항목을 제시하고 있다. 각 항목은 앞서 3.2절에서 수행한 태깅 분류를 바탕으로 구성하였다.

Table 4. Prompt classification table

Building Type	High-Rise	contemporary high-rise, modern high-rise
	Mid-Rise	mid-rise residential
	Mixed-Use & Residential	mixed-use tower, urban apartment complex, open-plan apartment complex,open-plan apartment complex
	institution	contemporary institutional building
Form & Structure	Form & Massing	rectangular high-rise form, geometric massing,clean geometric forms
	Facade & Composition	modular geometric facade, detailed parametric exterior, precise geometric composition
	Primary Structure	reinforced concrete core, structural steel frame
	Structural Emphasis	bold structural framing, cantilevered frame
Exterior Materials	Metal & Glass	dark metal cladding, glass curtain wall, alternating glass and metal panel facade, decorative metal screen integration, exposed steel frame
	Wood & Exposed Elements	wooden cladding, exposed steel frame, reinforced concrete frame, textured stucco finish
Windows & Openings	Windows	large glass windows, floor-to-ceiling glass windows, black window frames, vertical metal louvers,floor-to-ceiling glass windows
	Balconies & Protrusions	open balconies, recessed balconies, cantilevered balconies, cantilevered glazed volume, glass railings,recessed balconies
Lighting & Details	Lighting Elements	integrated LED lighting along facade, high-end architectural lighting, integrated LED linear lighting on cantilever edges, dynamic lighting, realistic glass reflections
	Roofline & Reflections	overhanging roofline, floating roofline with overhanging eaves, realistic glass reflections
Architectural Style	Contemporary Style	sleek minimalist aesthetic, elegant structural composition, sophisticated contrast of materials, sharp angular elements, minimalist yet expressive facade
	Minimalist Structure	modern minimalist, precise architectural composition, clean geometric forms, refined proportions, contemporary minimalist
Atmosphere & Context	Urban Context	modern urban development, contemporary urban aesthetic, high-tech cityscape, professional commercial environment,modern urban aesthetic
	Spatial Quality & Composition	ultra-realistic, structural clarity, seamless indoor-outdoor transition, well-balanced proportions

또한, 본 연구에서는 건축적 특징을 명확히 반영하기 위해 Positive Prompt와 Negative Prompt를 함께 사용하는 방식을 채택하였다. Positive Prompt에는 원하는 외장 재료나 양식 창문 배치 등 구체적인 요소를 서술함으로써 모델이 이러한 정보를 우선적으로 반영하도록 하며 Negative Prompt에는 이미지 생성 시 회피하고자 하는 왜곡이나 품질 저하 요소, 저해상도 표현 등을 기재하여 모델이 해당 표현을 최소화하도록 유도한다.

이를 활용해서 Table 5와 같이 건물 유형이나 형태 구조에 관한 주요 정보는 문장의 앞쪽에 배치하여 우선순위를 높이고, 나머지 재료나 조명 등의 보조 정보는 뒤쪽에 배치하여 우선 순위를 낮추어 프롬프트로 작성한다.

Table 5. Prompt contexts

Positive Prompt	architecture, facade, contemporary geometric design (Facade & Composition), luxury multi-family residence (Building Type), urban apartment complex (Building Type), reinforced concrete frame (Primary Structure), textured stucco finish (Wood & Exposed Elements), cantilevered balconies (Balconies & Protrusions), glass railings (Balconies & Protrusions), decorative metal screen integration (Metal & Glass), sharp angular elements (Architectural Style), bold structural framing (Structural Emphasis), minimalist yet expressive facade (Architectural Style), high-end residential building (Urban Context), modern urban aesthetic (Urban Context), well-balanced proportions (Spatial Quality & Composition), sophisticated contrast of materials (Architectural Style), ultra-realistic (Spatial Quality & Composition)
Negative Prompt	low quality, blurry, distorted, overexposed, watermark, low-resolution, cartoonish, excessive reflections, asymmetrical structure

(2) img2img와 LoRA 기법의 입면 이미지 적용

Img2img는 입력된 기본 이미지의 구조를 유지하면서도, 학습된 건축 양식과 재료 특성을 반영하여 더욱 구체적이고 세부적인 형태로 변환하는 역할을 수행한다.

예를 들어, Figure 6와 같은 단순한 매스 스터디의 볼륨을 입력하면, SD 모델이 학습된 건축적 특징을 바탕으로 이를 현대적인 건축 입면으로 구체화한다.

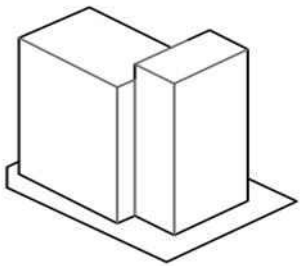


Figure 6. Simple Mass Image

Figure 7은 Figure 6를 활용하여 생성된 img2img 변환 결과를 시각화한 것으로, Table 4에 제시된 Positive Prompt와 Negative Prompt를 기반으로 생성된 이미지를 보여준다. 이를 통해 단순한 매스가 현대 건축물의 입면으로 구체화되는 과정을 확인할 수 있다. 하지만, Figure7에서 보이듯 건축적 맥락을 반영하지 못하고, 재료 또한 단순하게 표현된 것을 확인할 수 있다.

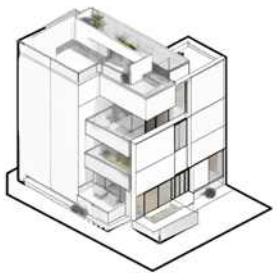


Figure 7. Mass Image using img2img

이때 앞서 학습된 LoRA 모델을 SD과 결합하여 사용하면, 같은 프롬프트를 입력하더라도 앞서 학습한 데이터베이스에서의 예시들에 담긴 건축적 맥락과 재료 및 형태의 구체적 디테일이 강화된 결과를 얻을 수 있으며, 설계자의 의도가 반영된 이미지를 생성할 수 있다.

따라서, Figure 7과 같은 프롬프트이지만 LoRA를 적용함으로써 원하는 설계 의도를 반영한 Figure 8와 같은 건축 입면 이미지를 얻을 수 있다. 하지만, 건물의 깊이나 입면의 입체감, 재료의 물성들에 대한 세부 형상이 실질적인 형태에 미치지 못한 것을 확인할 수 있다.

Figure 8. Mass Image using LoRA

3.4 ControlNet 기법을 활용한 세부형상 조정

Figure 9은 Figure 8과 같은 프롬프트와 조건에서 ControlNet의 Canny&Depth 모듈을 적용하여 생성된 이미지를 제시한다. Canny 모듈은 건물 실루엣과 창문·출입구 등의 윤곽선을 추출하여 프롬프트에 기반한 요소들을 선명하게 표현한다. 또한, Depth 모듈은 전후 관계와 깊이감을 사실적으로 반영하며 층간 간격, 테라스 돌출부, 건물의 볼륨감을 입체적으로 표현한다. 이를 통해, Canny&Depth 모듈과 LoRA를 함께 적용하면 건물의 깊이나 입면의 입체감, 재료의 물성들에 대한 세부 형상 그리고 건축적 맥락을 포함하는 원하는 세부 형상을 생성할 수 있다.

Figure 9. Mass Image using ControlNet

4. 건축 입면 생성 이미지의 분석 및 평가

4.1 생성 이미지의 평가 전략

SD의 img2img, LoRA, ControlNet을 활용하여 건축 입면 이미지를 생성하고, 이미지 평가 분석에 주로 사용하는 CLIP과 GPT-4V 기반의 평가 시스템를 통해 결과를 분석하였

다(Liang et al., 2023; Wang, Chan, & Loy, 2022). 그러나 CLIP의 최대 토큰 길이가 77개로 제한되면서, Table 5에 제시된 프롬프트의 일부 정보가 반영되지 않을 가능성이 있다. 이를 보완하기 위해, 건축적 요소를 중심으로 프롬프트를 체계화하고, 핵심 키워드를 중심으로 문장을 구성하였다. 또한, 같은 방식을 GPT-4V 평가에도 적용하여, 질문지를 체계적으로 구성한 후 개념별로 분리·재구성하여 분석을 수행하였다.

CLIP 점수는 텍스트 프롬프트와 이미지 간 의미적 일치도를 정량적으로 평가하는 지표로, Figure 10에서 제시된 파이썬 라이브러리를 활용하여 각 프롬프트별 점수를 산출한 뒤 평균값을 지표로 활용한다.

```
# Path to the image for evaluation
image_path = r"D:\12345\00236-65756554.png"
image = preprocess(Image.open(image_path)).unsqueeze(0).to(device)

# Optimized prompts for evaluation
text_prompts = [
    "A modern high-rise mixed-use tower with a modular geometric facade, reinforced concrete core, and glass curtain wall.",
    "A contemporary skyscraper featuring metal louvers, a cantilevered frame, and recessed balconies.",
    "A parametric architectural facade with precise grid structure and seamless glass and steel integration.",
    "An ultra-modern cityscape with high-tech design and structural clarity, emphasizing refined proportions."
]

# Tokenize the text prompts for CLIP
text_inputs = clip.tokenize(text_prompts).to(device)
```

Figure 10. Example of CLIP model code

GPT-4V는 특정 요소의 존재 여부뿐만 아니라 맥락적 타당성과 미적 조화를 평가하여 설계 의도와의 일치도를 분석한다. 이를 위해 Figure 11의 파이썬 라이브러리를 활용해 개별 점수를 산출하고, 평균값을 지표로 사용한다.

```
# ✅ Evaluation prompt for GPT-4V (in English-style prompt below)
prompt = """
You are an expert in architectural design evaluation.
Please assess the following architectural image based on the criteria below, and assign a score from 0 to 10 for each item:

Q1. Does this building have a high-rise tower form with a modular geometric facade?
👉 (0 = Not at all / 10 = Perfectly matches) ___ points

Q2. Is the building's structure well-reflected through a reinforced concrete core and glass curtain wall system?
👉 (0 = Not at all / 10 = Perfectly matches) ___ points

Q3. Does the building facade include cantilevered structures, emphasizing metal louvers or frame elements visually?
👉 (0 = Not at all / 10 = Perfectly matches) ___ points

Q4. Are the overall proportions and forms of the building refined and well-balanced?
👉 (0 = Not at all / 10 = Perfectly matches) ___ points

Please provide numeric scores only for each item (0–10), followed by a brief explanation.
Then, calculate and report the final average score based on the four categories.
"""
```

Figure 11. Example of GPT-4V model code

그러나, 이러한 자동화된 평가 시스템만으로는 건축적 설계의 실무적 판단 기준까지 충분히 반영하기 어렵다는 한계점이 존재한다. 따라서 본 연구에서는 CLIP 및 GPT-4V 기반의 자동 평가 결과를 보완하고, 보다 신뢰성 있는 분석을 도출하기 위해 건축 분야의 실무 경험을 갖춘 전문가들을 대상으로 한 평가를 Figure 12와 같이 병행하였다. 전문가 평가는 앞선 자동 평가와 동일한 평가 항목을 바탕으로 한 문항으로 조정하여, 건축적 관점이 함께 반영된 분석이 가능하도록 하였다.

The following images correspond to three different generation methods: (A) img2img only, (B) img2img with LoRA, and (C) img2img with LoRA and ControlNet.

Please assess each image individually according to the provided criteria.

Mass1	Practical Evaluation	Q1. Does the façade composition effectively convey the concept of modularity within a high-rise typology? (0~10) (A) ___ (B) ___ (C) ___
		Q2. To what extent does the structural system visibly support the architectural expression? (0~10) (A) ___ (B) ___ (C) ___
		Q3. Are key design features such as louvers and cantilevered elements appropriately emphasized in the elevation? (0~10) (A) ___ (B) ___ (C) ___
		Q4. How well does the overall form maintain proportional clarity and compositional coherence? (0~10) (A) ___ (B) ___ (C) ___
Mass2	Practical Evaluation	Q1. Does the building reflect the characteristics of a commercial tower with clarity and functional articulation? (0~10) (A) ___ (B) ___ (C) ___
		Q2. Is the articulation of façade elements effectively contributing to spatial rhythm and hierarchy? (0~10) (A) ___ (B) ___ (C) ___
		Q3. How convincingly are lighting effects and reflections integrated into the urban setting? (0~10) (A) ___ (B) ___ (C) ___
		Q4. Does the configuration of the building mass offer spatial depth or architectural dynamism? (0~10) (A) ___ (B) ___ (C) ___

Figure 12. Example of expert evaluation questionnaire

4.2 건축 입면 이미지 생성

이미지 생성을 위하여 3장에서 수행한 실험 절차를 동일하게 적용하되, Negative Prompt는 Table 4의 내용과 같이 일관되게 유지하였고, 다양한 결과를 확보하기 위하여 Table 6에 제시된 조건에 따라 Positive Prompt를 세분화하여 설정하였다.

각 Mass의 특징을 간략히 살펴보면, Mass 1은 강화 콘크리트 구조, 유리 커튼월 및 금속 루버를 적용한 기하학적 캔틸레버 구조의 현대적 고층 복합용도 건물이고, Mass 2는 캔틸레버 발코니와 금속 루버를 활용한 모듈식 외관으로 구성된 도시형 상업 건물이다.

또한, Mass 3은 철골 프레임과 대형 창호를 활용해 투명성과 불투명성을 조화롭게 표현한 현대적 오피스 건물이며, Mass 4는 LED 조명, 노출 철골 구조, 독특한 지붕 형태 및 기하학적 조경 요소가 어우러진 세련된 기업용 건축물이다. 이렇게 설정된 조건을 바탕으로 img2img, LoRA, ControlNet을 각각 적용해 이미지를 생성하였다.

4.3 생성 이미지의 평가 방법

(1) CLIP점수 평가 방법

CLIP 점수 평가는 앞서 제시한 4.1장의 기준에 따라 Table 6의 프롬프트를 분석해 Table 7과 같이 구성하였다. 이때 건축적 요소를 체계적으로 평가하기 위해 A는 건물의 전체 개념·용도 정의, B는 입면·외관 디자인 요소, C는 구조적·공간 구성 특성, D는 도시적 맥락·환경 관계라는 4가지 기준으로 프롬프트를 나누었다.

이후, CLIP 모델을 활용하여 생성된 건축 이미지와 해당 텍스트 프롬프트 간의 의미적 일치도를 Logits 기반 점수로 산출하였는데, 이는 프롬프트가 이미지에 어느 정도 반영되었는지를 정량적으로 평가할 수 있는 역할을 한다.

Table 6. Comparison of generated images

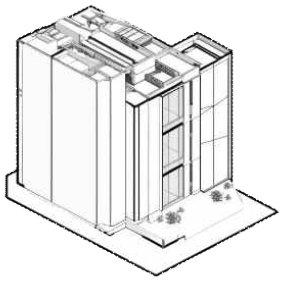
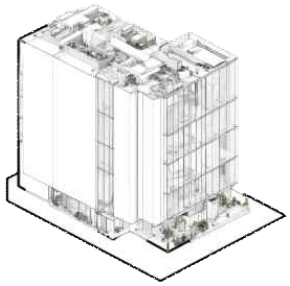
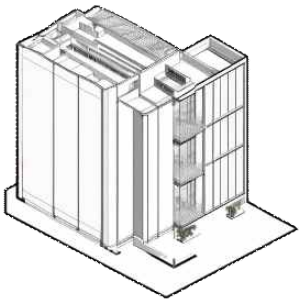
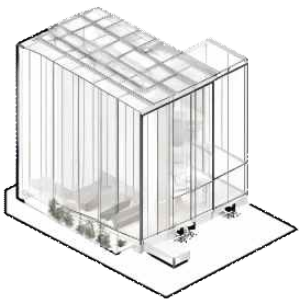
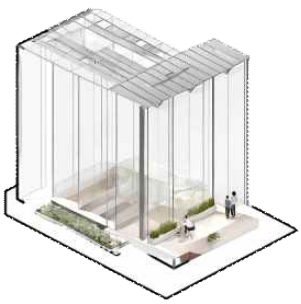
	img2img	LoRA	ControlNet
Mass1			
Prompt	Architecture, facade, contemporary high-rise, mixed-use tower, modular geometric facade, reinforced concrete core, glass curtain wall with metal louvers, cantilevered frame with recessed balconies, ultra-modern skyscraper, precise grid structure, high-tech cityscape, structural clarity, sharp lines, seamless integration of glass and steel, refined proportions, detailed parametric exterior, ultra-realistic		
Mass2			
Prompt	Architecture, facade, contemporary high-rise, mixed-use tower, modular geometric facade, reinforced concrete core, glass curtain wall with metal louvers, cantilevered frame with recessed balconies, urban skyscraper, precise architectural composition, professional commercial environment, realistic glass reflections, dynamic lighting		
Mass3			
Prompt	Architecture, facade, A minimalist modern office building with a rectangular high-rise form, alternating glass and metal panel facade, structural steel frame with reinforced concrete core, floor-to-ceiling windows with opaque and transparent sections, ground-level open lobby with dark-tinted glass, flat roof with minimal overhang, maintaining the original massing structure		
Mass4			
Prompt	Architecture, facade, A contemporary institutional building with a transparent corporate headquarters design, featuring a cantilevered glazed volume, full-height curtain wall glazing with minimal vertical mullions, integrated LED linear lighting on cantilever edges, exposed steel structural system with horizontal emphasis, floating roof line with overhanging eaves, and a landscaped plaza with geometric hardscape elements, maintaining the original massing structure		

Table 7. Mass CLIP prompt

Mass1	A	A modern high-rise mixed-use tower with a modular geometric facade, reinforced concrete core, and glass curtain wall
	B	A contemporary skyscraper featuring metal louvers, a cantilevered frame, and recessed balconies
	C	A parametric architectural facade with precise grid structure and seamless glass and steel integration
	D	An ultra-modern cityscape with high-tech design and structural clarity, emphasizing refined proportions
Mass2	A	A contemporary high-rise mixed-use tower featuring a modular geometric facade, reinforced concrete core, and a glass curtain wall with metal louvers
	B	An urban skyscraper with a cantilevered frame and recessed balconies, designed with precise architectural composition
	C	A professional commercial environment with realistic glass reflections and modern dynamic lighting
	D	A modern high-rise building with structured facade elements, creating a striking visual impact in an urban setting
Mass3	A	A minimalist modern office building with a rectangular high-rise form, featuring an alternating glass and metal panel facade
	B	An architectural structure with a reinforced concrete core and a structural steel frame, designed with precision
	C	A contemporary office building with floor-to-ceiling windows, combining opaque and transparent sections
	D	A well-balanced urban skyscraper with a dark-tinted glass lobby, a flat roof with minimal overhang, and preserved massing structure
Mass4	A	A contemporary institutional building designed as a transparent corporate headquarters, featuring a cantilevered glazed volume and full-height curtain wall glazing with minimal vertical mullions
	B	An architectural structure integrating LED linear lighting along cantilever edges and an exposed steel structural system with a horizontal emphasis
	C	A modern institutional building with a floating roofline extending with overhanging eaves, emphasizing contemporary aesthetics
	D	A well-balanced corporate plaza incorporating geometric hardscape elements, seamlessly maintaining the original massing structure

(2) GPT-4V 평가 방법

Table 8. Architectural evaluation criteria for GPT-4V

Mass1	a	Does this building have a high-rise tower form with a modular geometric facade?
	b	Is the building's structure well-reflected through a reinforced concrete core and glass curtain wall system?
	c	Does the building facade include cantilevered structures, emphasizing metal louvers or frame elements visually?
	d	Are the overall proportions and forms of the building refined and well-balanced?
Mass2	a	Does this building appear as a modern high-rise office tower?
	b	Does the architectural composition include dynamic spatial elements such as cantilevers and balconies?
	c	Are the reflections and lighting effects realistic and well-aligned with an urban atmosphere?
	d	Does the building facade utilize structural patterns (metal louvers and geometric panels) to create strong visual effects?
Mass3	a	Does this building resemble a minimalist and modern office building?
	b	Is the facade design composed of alternating glass and metal panels?
	c	Is the building's geometric form (rectangular shape, emphasized structural frame) clearly visible?
	d	Does the lobby design maintain symmetry with dark glass while providing an open and spacious environment?
Mass4	a	Is this building suitable as a modern corporate headquarters?
	b	Does the facade design feature cantilevered glass volumes and curtain walls?
	c	Does the architectural design incorporate LED lighting lines, exposed steel structures, and horizontal emphasis elements?
	d	Does the building's form harmonize with the surrounding topography while maintaining the originality of its mass?

GPT-4V 기반 평가도 앞서 제시한 4.1장의 기준에 따른 네 가지 기준을 바탕으로 Table 6의 동일한 프롬프트를 분석한 뒤, 앞서 CLIP점수 평가방법에서의 4가지 분류방식과 같은 방식으로 a,b,c,d를 나누었다.

이를 Table 8에서 제시된 평가 항목으로 재구성하여 0점부터 10점 사이의 점수를 매겼다. 이 과정에서는 이미지 속 요소가 단순히 존재하는지뿐만 아니라, 맥락적 타당성과 미적 완성도가 설계 의도에 부합하는지를 정성적으로 살핀다.

(3) 전문가 평가 방법

전문가 평가는 CLIP 및 GPT-4V의 평가 체계를 보완하기 위해 도입되었으며, 건축 실무 관점에서 생성 이미지의 타당성과 구현 가능성을 정성적으로 평가할 수 있도록 구성되었다. 각 평가는 설계 의도와의 부합성, 공간 구성의 타당성, 재료 및 구조 표현의 적절성 등을 반영한 항목으로 구성되었으며, 0점에서 10점 사이의 점수를 기준으로 이루어졌다. 평가 인원은 총 11명이며, 평균 실무경력 16.7년(경력 7년~23년)으로 이는 실무 경력과 평가 신뢰도를 고려한 적정 규모이다.

이는 Lu et al. (2024)의 사진 품질 평가 연구에서 활용된 10명 수준 및 핵의학 영상 품질 평가에 15명의 전문가를 활용한 예인 Chung et al.(2025) 등과 유사한 맥락에서 판단된 것이다. 평가 항목은 CLIP과 GPT-4V의 기준을 바탕으로 재구성된 a', b', c', d'로 제시되며, Table 9와 같다.

Table 9. Architectural evaluation criteria for expert

Mass1	a'	Does the façade composition effectively convey the concept of modularity within a high-rise typology?
	b'	To what extent does the structural system visibly support the architectural expression?
	c'	Are key design features such as louvers and cantilevered elements appropriately emphasized in the elevation?
	d'	How well does the overall form maintain proportional clarity and compositional coherence?
Mass2	a'	Does the building reflect the characteristics of a commercial tower with clarity and functional articulation?
	b'	Is the articulation of façade elements effectively contributing to spatial rhythm and hierarchy?
	c'	How convincingly are lighting effects and reflections integrated into the urban setting?
	d'	Does the configuration of the building mass offer spatial depth or architectural dynamism?
Mass3	a'	Is the design consistent with contemporary standards of minimalism in office architecture?
	b'	To what extent are material transitions articulated clearly in the facade?
	c'	Does the form successfully express geometric discipline and formal legibility?
	d'	How effectively does the entrance zone or lobby area support spatial openness and public engagement?
Mass4	a'	Is the design appropriate for a corporate headquarters in terms of symbolism and public image?
	b'	How well are the cantilevered glass volumes and curtain wall elements integrated into the massing?
	c'	Are technical components sufficiently detailed in the design language?
	d'	Does the mass harmonize with site conditions while maintaining formal identity and compositional integrity?

4.4 건축 입면 생성 이미지의 평가

(1) CLIP 점수

Table 10. CLIP score evaluation table for generated images

Mass	Prompt	img2img	img2img+ LoRA	img2img+ LoRA+ ControlNet
1	A	23.6	25.4	27.1
	B	23.8	22.1	28.3
	C	22.0	25.3	23.1
	D	22.9	22.8	25.3
	Avg	23.1	23.9	26.0
2	A	24.3	27.8	23.3
	B	23.2	27.1	22.6
	C	17.6	26.7	26.8
	D	16.4	24.1	21.7
	Avg	20.4	26.4	23.6
3	A	29.4	32.8	35.7
	B	29.7	28.1	28.3
	C	27.2	36.4	36.8
	D	26.9	26.2	26.7
	Avg	28.3	30.9	31.9
4	A	30.6	30.4	26.8
	B	26.0	22.9	26.1
	C	23.4	22.2	25.2
	D	19.8	19.2	27.6
	Avg	24.9	23.7	26.4

3개 항목의 평가에서 img2img를 SDXL 베이스라인으로 두고, LoRA·ControlNet 적용 방식과의 점수를 비교하였다.

Mass 1에서는 img2img,LoRA,ControlNet을 사용한 부문이 B 항목 28.3점, C 항목 23.1점을 포함해 평균 26.0점을 기록하였다. 금속 루버와 캔틸레버 프레임 같은 외관 디테일을 명확히 구현한 점이 전반적 점수 향상에 기여한 것으로 해석된다.

Mass 2의 CLIP 평가는 LoRA과 img2img를 같이 사용한 방식이 평균 26.4점을 기록하였다. 이는 유리 커튼월의 반사, 동적 조명이 정교하게 재현한 결과로 해석된다. 반면 구조·공간 구성인 C 항목에서는 ControlNet이 26.8점으로 가장 높은 수치를 기록하여, 세부 형상 제어와 매스 윤곽의 정확도 면에서는 우위를 유지하고 있음을 시사한다.

Mass 3에서는 img2img,LoRA,ControlNet을 사용한 부문 C 항목 36.8점, A 항목 35.7점을 바탕으로 평균 31.9점을 기록하였다. 금속, 유리 패널의 교차 배열와 같은 구조적 대비를 반영해 높은 평가를 얻은 것으로 분석된다.

Mass 4에서도 img2img,LoRA,ControlNet을 사용한 부문 D 항목 27.6점, C 항목 25.2점을 포함해 평균 26.4점으로 최고 점수를 나타냈다. 캔틸레버 볼륨과 하드스케이프 등 도시 맥락 요소를 명확히 제시한 점이 우수한 평가로 이어진 것으로 해석된다.

CLIP 평가를 평균값 기준으로 종합하면, img2img,LoRA, ControlNet이 Mass 2를 제외한 모든 사례에서 최상위 평균을 기록하며 가장 일관된 우수성을 보였다. Mass 2에서는 img2img,LoRA가 최고점을 달성했으나, 구조·공간 세부 형상 제어부분에서는 ControlNet이 우위임을 확인할 수 있었다. 반면 img2img는 기본 형태 유지에는 유효하지만 복합적인 건축 디테일과 맥락 표현에는 한계를 드러냈다.

(2) GPT-4V 점수

Table 11. GPT-4V score evaluation table for generated images

Mass	Criteria	img2img	img2img+ LoRA	img2img+ LoRA+ ControlNet
1	a	3.0	6.0	8.0
	b	7.0	8.0	7.0
	c	5.0	3.0	4.0
	d	8.0	7.0	9.0
	Avg	5.8	6.0	7.0
2	a	7.0	8.0	8.0
	b	5.0	6.0	7.0
	c	4.0	5.0	5.0
	d	6.0	4.0	6.0
	Avg	5.5	5.8	6.5
3	a	10.0	9.0	9.0
	b	5.0	2.0	8.0
	c	10.0	10.0	10.0
	d	4.0	8.0	7.0
	Avg	7.2	7.2	8.5
4	a	8.0	8.0	9.0
	b	8.0	7.0	7.0
	c	4.0	5.0	4.0
	d	6.0	7.0	8.0
	Avg	6.5	6.8	7.0

GPT-4V 또한 3개 항목의 평가에서 CLIP과 같은 방식으로 점수를 비교하였다.

Mass1은 img2img,LoRA,ControlNet을 사용한 부문이 평균 7.0을 기록하며 가장 높은 결과를 보였다. d항목에서 9점, b항목에서 7점을 받았으나, c항목에서 4점을 기록하여 평균 7점으로 평가되었다. 이는 전반적으로 비례와 형태가 현대적이고, 구조적 요소가 잘 반영되었으나 특정 디자인 요소 강조가 부족했다고 평가되었다.

Mass2는 img2img,LoRA,ControlNet을 사용한 부문이 평균 6.5점으로 가장 높은 결과를 보였다. a항목에서 8점, d항목에서 7점을 받았으나, c항목에서 5점을 기록하며 평균 6.5점으로 나타났다. 건물의 기본 특성,금속 루버와 구조적 패턴은 충족했으나, 시각적 효과 개선이 필요했다고 평가되었다.

Mass 3는 img2img, LoRA, ControlNet을 사용한 부문이 평균 8.5점으로 가장 높은 결과를 보였다. c항목에서 10점, b항목에서 8점을 받았으나, d항목에서 7점을 기록하며 평균 8.5점으로 나타났다. 명확한 기하학적 구조와 미니멀리즘 디자인은 성공적으로 반영되었으나, 로비 디자인과 유리 표현에서는 일부 한계가 있는 사례로 평가되었다.

Mass4는 img2img,LoRA,ControlNet을 사용한 부문이 평균 7.0점으로 가장 높은 결과를 보였다. a항목에서 9점, b항목에서 7점을 받았으나, c항목이 4점을 기록하며 평균 7점을 기록했다. 균형 잡힌 디자인와 유리의 시각적 연속성이 강점이지만, 세부 구조 표현이 약했다고 평가되었다.

종합적으로는 GPT-4V 평가에서는 img2img,LoRA,ControlNet을 결합한 방식이 모든 사례에서 최상위 평균 점수를 기록하며, 기하학적 형태와 디자인 일관성이 높을수록 높은 점수를 기록하는 경향을 보였다.

(3) 전문가 평가 점수

Table 12. Expert score evaluation table for generated images

Mass	Criteria	img2img	img2img+ LoRA	img2img+ LoRA+ ControlNet
1	a’	4.4	5.2	6.0
	b’	3.5	5.3	6.5
	c’	3.5	4.7	6.1
	d’	3.7	4.5	5.8
	Avg	3.8	4.9	5.9
2	a’	2.7	3.8	5.0
	b’	2.8	3.5	4.6
	c’	2.8	3.7	4.7
	d’	3.1	3.7	4.8
	Avg	2.9	3.7	4.6
3	a’	5.7	7.5	8.6
	b’	5.7	7.3	8.7
	c’	6.2	7.0	8.3
	d’	5.0	7.1	8.5
	Avg	5.7	7.2	8.5
4	a’	5.0	6.3	7.4
	b’	4.6	6.3	7.2
	c’	4.5	5.9	7.3
	d’	4.5	5.7	7.9
	Avg	4.7	6.1	7.5

전문가 평가도 앞선 CLIP, GPT-4V와 같은 방식으로 3개 항목의 평가에서 점수를 비교하였다.

Mass 1은 img2img, LoRA, ControlNet을 사용한 부문이 평균 5.9점으로 가장 높은 결과를 보였다. b'항목에서 6.5점, c' 항목에서 6.1점을 받았으나, d'항목에서 5.8점을 기록하였다. 구조 시스템의 시각적 구현과 루버·프레임 등 디자인 요소의 통합은 긍정적으로 평가되었으며, 비례감과 조형적 일관성, 공간 구성 측면에서는 설득력이 다소 부족하다고 평가되었다.

Mass 2는 img2img, LoRA, ControlNet을 사용한 부문이평균 4.6점으로 가장 높은 결과를 보였다. a'항목에서 5.0점, d' 항목에서 4.8점을 받았으나, b'항목에서 4.6점을 기록하였다. 상업용 고층 건물의 기능적 구획과 매스의 공간적 역동성은 일정 부분 긍정적으로 평가되었지만, 구조 표현의 명확성과 기능–형태 통합 면에서는 미흡하다고 평가되었다.

Mass 3은 img2img, LoRA, ControlNet을 사용한 부문이평균 8.5점으로 가장 높은 결과를 보였다. b'항목에서 8.7점, a' 항목에서 8.6점을 받았으나, c'항목에서 8.3점을 기록하였다. 재료 표현과 미니멀리즘 구현, 설계 언어의 일관성도 긍정적으로 평가되었으나, 기하학적 질서와 시각적 효과는 상대적으로 부족하다고 평가되었다.

Mass 4는 img2img, LoRA, ControlNet을 사용한 부문이 평균 7.5점으로 가장 높은 결과를 보였다. d'항목에서 7.9점, a' 항목에서 7.4점을 받았으나, b' 항목에서 7.2점을 기록하였다. 상징성 표현과 맥락 대응력과 도심 환경에 적합한 조형 정체성 구현은 우수하게 평가되었으나, 구조 시스템과 재료 표현의 명확성은 다소 부족하다고 평가되었다.

종합적으로 전문가 평가를 평가하면, img2img, LoRA, ControlNet을 결합한 방식이 모든 사례에서 최상위 평균 점수를 기록하며 가장 일관된 우수성을 보였다. 이는 구조 시스템의 시각화, 재료 표현, 조형적 정합성, 맥락 대응력 등 다면적인 건축적 평가 항목에서 균형 있는 성과를 낸 것으로 해석된다.

(4) 종합분석

본 연구에서는 CLIP 모델과 GPT-4V 모델 그리고 전문가 평가를 활용하여 생성된 건축 이미지를 정량적·정성적으로 평가하였다. CLIP 점수는 프롬프트와 이미지 간의 일치도를 평가하며, 특정 건축적 요소가 명확히 반영될수록 높은 점수를 기록하였다. 반면, GPT-4V 평가는 공간 구성, 형태의 균형, 재료 표현 등 건축적 맥락 속에서의 조화를 고려하여 이루어졌다. 여기에 더해진 전문가 평가는 CLIP이나 GPT-4V가 포착하지 못하는 구조 시스템의 표현력, 기능과 형태의 통합성, 맥락 대응력 등 실무적 타당성을 기준으로 보다 현실적인 판단을 제공하였다. 세 가지 평가 결과를 비교한 결과, 건축적 요소가 정돈되고 조형 언어의 일관성이 유지된 이미지일수록 공통적으로 높은 점수를 받았으며, 반대로 구조나 세부 표현이 부족한 경우에는 전 평가 방식에서 모두 낮은 점수를 기록하는 경향이 나타났다. 이는 CLIP, GPT-4V, 전문가 평가가 서로 다른 기준과 해석 관점을 갖고 있음에도 불구하고, 건축적 완성도를 판단하는 데 있어 일정한 일치성과 상호 보완적 관계를 형성하고 있음을 시사한다.

결론적으로, 세 가지 평가 방식은 각기 다른 관점을 통해 설계 결과물을 다각적으로 해석하고 보완하며, 이를 통합적으로 활용할 경우 신뢰도 높은 건축 디자인 검토 체계를 구축하는 데 기여할 수 있음을 확인하였다.

5. 결　　론

본 연구는 SD 모델이 건축 도메인의 세부 요구 사항을 충족하도록 LoRA와 ControlNet을 결합한 이미지 생성 방안을 제안하고 그 효과를 검증하였다. LoRA를 활용한 미세 조정으로 창호 배열, 발코니 배치, 외장 재료 등의 건축적 디테일이 정밀하게 표현됨을 확인했다. 또한 ControlNet의 Canny Edge 및 Depth Map 적용을 통해 형상과 깊이감을 정교하게 제어할 수 있음을 실험적으로 입증하였다. 또한, CLIP 점수와 GPT-4V 점수, 전문가 평가 점수를 동시에 수행한 결과, 설계 의도와 건축적 요소가 명확할수록 높은 평가를 받는 경향이 나타나, img2img,LoRA,ControlNet을 종합적으로 사용하는 것이 건축적 맥락을 반영하는 데 긍정적인 영향을 미친다는 사실을 확인하였다.

하지만, 본 연구의 데이터셋이 특정 건축 양식과 외장 재료에 편중되어 있어 다양한 스타일을 반영하는 데 한계가 있으며, ControlNet이 창호·발코니·출입구 등의 형상 제어에는 효과적이나, 다층적 패턴을 포함한 건축물에는 추가 조정이 필요할 수 있다. 그러나, 이러한 한계는 향후 연구에서 건축 양식과 재료 유형을 더욱 확장하고, 구조적 특징 및 환경적 요소까지 고려하는 데이터셋을 구축함으로써 개선될 수 있을 것이다.

결과적으로 본 연구는 LoRA와 ControlNet이라는 두 가지 핵심 기법의 결합이 건축 디자인 프로세스에서 실제 활

용될 만한 수준의 이미지 생성 품질을 확보할 수 있음을 입증하였다. 이는 설계 초기 단계에서 다양한 대안을 신속하게 생성·비교함으로써, 반복적인 시각화 작업에 따른 시간·인력 소모를 줄이는 동시에 창의적인 설계 아이디어를 탐색하는 도구가 될 수 있을 것으로 기대된다.

REFERENCES

1. Ching, F. D. K. (2023). *Architecture: Form, space, and order (5th ed.)*. John Wiley & Sons.
2. Chung, H. W., Kim, J. Y., Lee, J. S., Park, S. Y., Jang, J., Lee, J. W., & Lee, D. S. (2025). A fully automated, expert-perceptive image quality assessment system for whole-body [18F]FDG PET/CT. *EJNMMI Research*, 15(1), 1-12.
3. Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., & Bengio, Y. (2014). Generative adversarial nets. *Advances in Neural Information Processing Systems*, 27, 2672–2680.
4. Ho, J., Jain, A., & Abbeel, P. (2020). Denoising diffusion probabilistic models. *Advances in Neural Information Processing Systems*, 33, 6840-6851.
5. Hu, E. J., Shen, Y., Wallis, P., Allen-Zhu, Z., Li, Y., Wang, S., Wang, L., & Chen, W. (2021). LoRA: Low-rank adaptation of large language models. *arXiv*. https://doi.org/10.48550/arXiv.2106.09685
6. Jo, H., & Lee, J. (2023). Approaches to building facade design reflecting local identities using generative AI: Focusing on commercial streets in Seongsu-dong area. *Journal of the Korea Institute of the Spatial Design*, 18(7), 361-369.
7. Lee, D., & Ko, S. (2023). Experiment and evaluation of architectural image generation through artificial intelligence-based text image generation tool. *KIEAE Journal*, 23(5), 13-22.
8. Liang, P., Zou, Y., Zhang, C., & Ghosh, S. (2023). An early evaluation of GPT-4V(ision). *arXiv*. https://doi.org/10.48550/arXiv.2310.16534
9. Lu, X., Shahid, M., Li, W., He, Z., Hamidouche, W., & Ma, K. (2024). UHD-IQA benchmark database: Pushing the boundaries of blind photo quality assessment. *arXiv*. https://doi.org/10.48550/arXiv.2406.17472
10. Ma, H., & Zheng, H. (2023). Text Semantics to Image Generation: A method of building facades design base on Stable Diffusion model. *The International Conference on Computational Design and Robotic Fabrication* (pp. 24-34). Singapore: Springer Nature Singapore.
11. Meng, C., He, Y., Song, Y., Song, J., Wu, J., Zhu, J. Y., & Ermon, S. (2021). Sdedit: Guided image synthesis and editing with stochastic differential equations. *arXiv*. https://doi.org/10.48550/arXiv.2108.01073
12. Radford, A., Kim, J. W., Hallacy, C., Ramesh, A., Goh, G., Agarwal, S., & Clark, J. (2021). Learning transferable visual models from natural language supervision. *Proceedings of the 38th International Conference on Machine Learning (ICML)*, 139, 8748-8763.
13. Rombach, R., Blattmann, A., Lorenz, D., Esser, P., & Ommer, B. (2022). High-resolution image synthesis with latent diffusion models. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition,* 10684-10695.
14. Saharia, C., Chan, W., Saxena, S., Li, L., Whang, J., Denton, E., Ghasemipour, S. K. S., Ayan, B. K., Mahdavi, S. S., Lopes, R. G., Salimans, T., Ho, J., Fleet, D. J., & Norouzi, M. (2022). Photorealistic text-to-image diffusion models with deep language understanding. *Advances in Neural Information Processing Systems*, 35, 36479-36494.
15. Wang, J., Chan, K. C. K., & Loy, C. C. (2022). Exploring CLIP for assessing the look and feel of images. *arXiv*. https://doi.org/10.48550/arXiv.2207.12396
16. Wang, J., & Zhang, X. (2024). Exploring text-based realistic building facades editing application. *arXiv.* https://doi.org/10.48550/arXiv.2405.02967
17. Xu, T., Zhang, P., Huang, Q., Zhang, H., Gan, Z., Huang, X., & He, X. (2018). Attngan: Fine-grained text to image generation with attentional generative adversarial networks. *Proceedings of the IEEE conference on computer vision and pattern recognition,* 1316-1324.
18. Yoo, Y., & Lee, J. (2023). Generative AI-Based Construction of Architect's Style-trained Models and its Application for Visualization of Residential Houses. *Design convergence study*, 22(6), 103-116.
19. Zhang, H., Xu, T., Li, H., Zhang, S., Wang, X., Huang, X., & Metaxas, D. N. (2018). Stackgan++: Realistic image synthesis with stacked generative adversarial networks. *IEEE transactions on pattern analysis and machine intelligence*, 41(8), 1947-1962.
20. Zhang, L., Rao, A., & Agrawala, M. (2023). Adding conditional control to text-to-image diffusion models. *Proceedings of the IEEE/CVF International Conference on Computer Vision,* 3836-3847.
21. Zhao, Z., & Li, R. (2023). Modified generative adversarial networks for image classification. *Evolutionary Intelligence,* 16(6), 1899-1906.

(Received Mar. 13, 2025/ Revised Apr. 7, 2025/ Accepted Jun. 23, 2025)