JAIST Repository
https://dspace.jaist.ac.jp/

Title	Building Image Generation Using Retrieval Augmented Diffusion Models
Author(s)	王, 征洋
Citation	
Issue Date	2025-03
Type	Thesis or Dissertation
Text version	author
URL	http://hdl.handle.net/10119/19785
Rights	
Description	Supervisor: 謝 浩然, 先端科学技術研究科, 修士 (情報科学)

JAIST
JAPAN
ADVANCED INSTITUTE OF
SCIENCE AND TECHNOLOGY

Japan Advanced Institute of Science and Technology

Building Image Generation Using Retrieval Augmented Diffusion Models

2310021 Zhengyang Wang

Despite recent advances in text-to-image generation greatly improving image quality, there is a growing demand for specifying image details through text prompts. Especially in the field of architectural image generation, current models struggle to achieve precise control of generated image details, such as the number of floors and other architectural components. For architectural design, it is essential to precisely control such details including the number of floors and architectural components, such as the shape and proportion of windows, as these factors directly influence the accuracy of the generated images in reflecting the architect's design intentions. Especially in the early stages of architectural design, the lack of detailed expression in generative models may directly impact the accurate communication of design intent and the final quality of the design proposal.

To address this challenge, a multi-stage retrieval-augmented diffusion model is proposed for precise control over image details such as the number of floors and details in architectural image generation. First, an initial image with architectural style is generated, and a sketch is extracted from it. Then, based on the number of floors specified in the text prompt, images are retrieved from an external architectural image database that are nearly consistent with the specified structure. The retrieved building structures are used as conditional inputs for a diffusion model, combined with features from the initial style image, to generate a building image with the specified number of floors and styles. Next, the generated building image sketch is processed using segmentation algorithms to extract key architectural components, such as doors and windows. These segmented architectural components serve as query conditions to match and retrieve relevant components from an external component database. The retrieved architectural components are fused with the local features in the sketch input, generating a detailed sketch. Finally, the detailed sketch along with the text prompt is input into the latent diffusion model to generate a high-quality architectural facade rendering.

The proposed framework effectively addresses the limitations of existing models in controlling details during the generation phase, with significant advantages in controlling the number of floors and components in architectural images. The proposed framework offers several contributions. First, an innovative multi-stage retrieval-augmented approach is introduced that effectively leverages external datasets to enhance the precision and realism of architectural image generation. Second, the proposed framework achieves precise control in generating building images with specified floor counts and

enables accurate depiction of architectural components and structural details. Finally, three specialized datasets are established: one for retrieving architectural building structures, one for augmented architectural components, and one for pairing sketches with their corresponding high-quality renderings, which collectively provide robust support for generating precise and realistic architectural designs.

Experimental results demonstrate that the proposed framework exhibits strong robustness and scalability in generating fine-grained architectural design images, providing an accurate architectural representation and detailed presentation in the early stages of architectural design, offering an efficient and flexible solution for architectural design tasks.