Discussion on the Ontology Inconsistency & 3D Model Generation in Architectural Design Assisted by AIGC

Yunfan Ye[1], Fan Wang[2], Haojia Xu[3], Hui Wang[4]

Abstract: AIGC technology is playing an increasingly crucial role in generative design. At present, many AIGC tools are used to generate images in the architectural design, such as Midjourney, Stable Diffusion. Although the single image generated by AIGC tools might have high quality, it is difficult to maintain the consistency of the ontology when representing the same building from different angles due to the inevitable randomness of the generation process, which means different pictures contradict each other in details. This article discusses the possible solutions of ontology inconsistency problem with the guidance of different AIGC tools, and points out that generating 3D model is the most effective way to solve the problem, but the current generation tools still have various defects. With the contrast of different approaches, the technical path of Sketch-Multiple Pictures-Contradiction Elimination-3D Model is considered to be more appropriate for architectural design process. This path also puts forward technical requirements for the future specialization of current AIGC tools.

Keywords: Ontology Inconsistency, AIGC, Architectural Design, 3D Model Generation

1 Introduction

The rapid development of Artificial Intelligence Generated Content (AIGC) has brought new patterns of design work and also driven a significant reform in the field of architectural design. AIGC refers to the use of generative AI algorithms to assist or replace humans in creating rich personalized and high-quality content, such as text, painting, music, video, and even interactive 3D content, at a faster pace and lower cost, according to user input and requirements.[4] Through a large number of model training, AI understands the patterns and rules, and generates some new content with similar characteristics to the training data. Presently, AI is able to process and generate text, images, videos and other modes of information, and obtain a large number of outputs in a short period of time, which greatly improves design efficiency.

The research on AIGC in architectural design field is in the ascendant. Yanfeng Huang et al. summarized the current three generation methods of AIGC, and proposed an optimization method for the full generation of architectural design: generating 3D models from 2D images and prompts, and then slicing 3D models into 2D slices to generate specific planes.[2] Guo Li et al. generalized the main application scenarios of generative AI in architectural design fields, developed by various design institutes and universities.[3] Taking *LookX AI Cloud* as an example, Wanyu He et al. discussed how to integrate generative AI technology and proposed an intelligent workflow that integrates language model to realize conversational interaction design.[1] Chao Yuan explored the

---

[1] Yunfan Ye, College of Civil Engineering and Architecture, Zhejiang University

[2] Fan Wang, College of Civil Engineering and Architecture, Zhejiang University

[3] Haojia Xu, College of Civil Engineering and Architecture, Zhejiang University

[4] Hui Wang (✉), College of Civil Engineering and Architecture, Zhejiang University, e-mail: wang_hui@zju.edu.cn

multi-dimensional generation mode of AI and put forward a new architectural design mode under the influence of generative AI.[5]

This article focuses on the problem of ontology inconsistency in term of the deep use of AIGC in architectural design, which aims at a throughout generation of complete design scheme rather than generating a single perspective image. The problem will be demonstrated firstly, and then the possible solutions within the image generation tools will be explored. Aiming at solving the problem completely in the future, different tools of generating 3D models and the feasible working path in architectural design is also to be discussed.

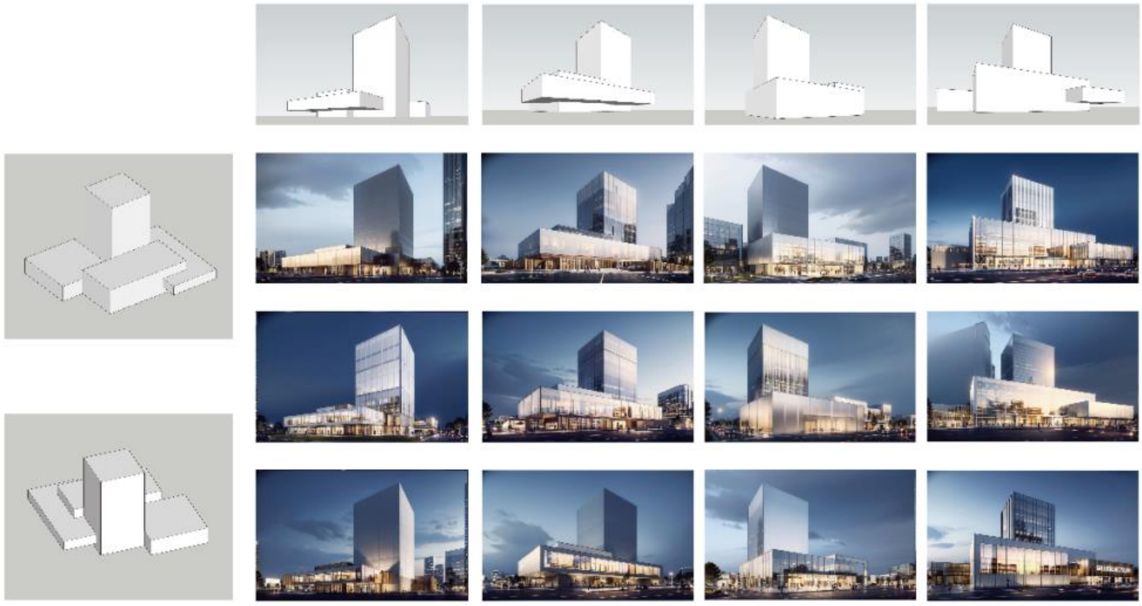

Fig. 1 Inconsistencies in generating multi-angle perspectives based on wireframes (Drawn by the authors)

2 Ontology Inconsistency

At present, the architectural AIGC mostly focuses on generating a certain style of renderings through sketches (including hand-drawn drawings or wireframes of simple models), which can be compared with a variety of styles in the early research stage and inspire design ideas. Tools such as Midjourney, Stable Diffusion and Dall-E are commonly used in this approach. However, the further use of AIGC tools will inevitably encounter the problem of ontology inconsistency. It refers to the generated results cannot maintain the consistency of forms and details, when generating different perspectives of the same building. Adding more manual details to control the model will reduce the proportion of AI-generated design, and is equivalent to reducing the AIGC tool to a renderer. This problem has a great impact on the deep use of AI in architectural design.

For example, in Stable Diffusion, the checkpoint and LoRA of intended design style should be firstly determined before input the corresponding positive and negative prompts. Then the outline of one perspective of a building should be imported into *ControlNet* under the control of *Lineart*. Next, different parameters, like the iterative steps and control weights should be adjusted to generate multiple renderings of that perspective. After that, the outlines of other perspectives could be imported into *ControlNet* for wireframe control, to generate other renderings, with the mostly consistent parameters and prompts. It is the results of four different perspectives of a building model (Fig.1). As the checkpoint, LoRA and the basic parameters remain unchanged, the generated renderings have the similar style and surroundings. However, there are still obvious problems such as inconsistencies and contradictions within the volume, facades and material details, which means the ontology is not the same one.

The fundamental cause of the problem lies in the deep generative mechanism of AIGC tools. In the field of image generation, there are four major generation models in recent years: GAN, VAE, Flow based Model, Diffusion Model, etc.(Fig.2), which are mainly based on deep learning as the training model. Starting in 2022, the main popular image generation model is Diffusion Model. For example, Stable Diffusion uses latent diffusion model (Fig.3). Historically, probabilistic models suffer from a trade-off between two conflicting objectives: tractability and flexibility. Diffusion models are inspired by non-equilibrium thermodynamics. It defines a Markov chain of diffusion steps to slowly add random noise to data and then learn to reverse the diffusion process to construct desired data samples from the noise. However, when random noise increases the diversity of the generated result, it also causes the inevitable randomness.

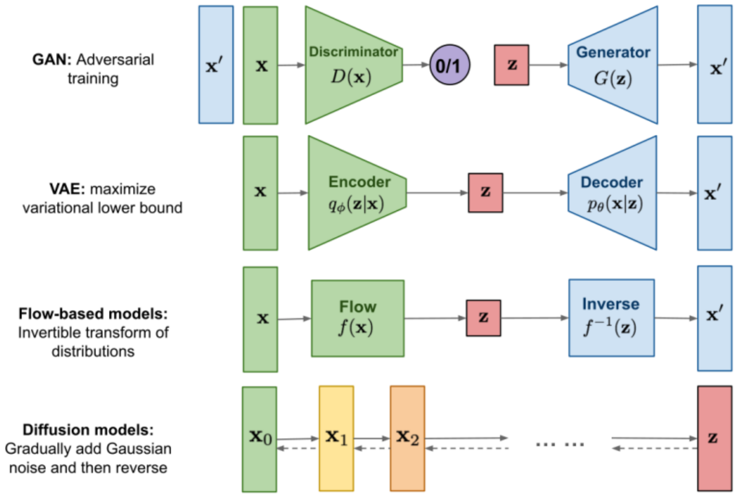

Fig. 2 Overview of different types of generative models (Image source: Rombach & Blattmann, et al. 2022)

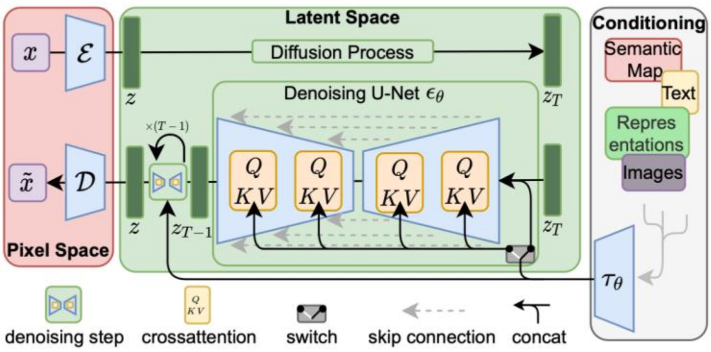

Fig. 3 The architecture of the latent diffusion model (LDM) (Image source: Rombach & Blattmann, et al. 2022)

The issue of ontology inconsistency is currently a widespread problem in the field of AIGC, especially in the case of generating multiple images or videos of the same subject. How to maintain the consistency of the ontology with less manual control needs more exhaustive exploration. Developers of AIGC tools have taken note of this issue and have attempted to address it in certain ways, such as by adding reference images to enhance similarity. Recently Midjourney has released the *character reference* tool in version 6.0 to maintain consistent characters across multiple images. It can optimize the character traits such as face, hair and dress, to obtain a roughly consistent result. But it cannot achieve accurate sameness and is not efficient in architectural design.

3 Optimization within Image Generation Tools

The possible solutions within 2D image generation tools can be explored. In Stable Diffusion, there is also a reference function similar to the character reference function in Midjournal. It can guide diffusion using single image as a reference without any control model. However, currently this function only has good diffusion results for characters, animals, objects and so on. For architecture, even with the control of line draft and depth of field, it can only generate images with similar styles, and there is a significant randomness in the details. The Seed parameter in Stable Diffusion determines the content of the generated image and ensures the reproducibility of the generated results. As long as the same number of seeds and other identical parameters are used, the same image result can be obtained. In the process of generating architectural renderings, modifying other parameters appropriately and using the same number of seeds can obtain other renderings with the same style features and details.

After research, it was found that the number of seeds can control the similarity of generated images, but to maximize the consistency of the generated result model ontology, more controls need to be added. For example, in Stable Diffusion, it is viable to optimize the generated contents by adjusting parameters and importing images circularly into *ControlNet*. The seed of the first image from the above generation process could be used to control the generation of the other perspectives. The generated images could be imported into *ControlNet* to add control by *Reference*, which is beneficial to generate similar images. Setting parameters and sending images to *inpaint* to repaint the details also contribute to the consistency of the pictures. If there are differences among the first one and other perspectives, the images from other perspectives could be used to add control by *Reference* with the same seed. The regenerated picture could develop in the way as above to improve the details.

Under multiple iterations by this method, the 2D generated images could improve the result of Fig.1to a certain extent, with the improved consistency of different perspectives (Fig.4). However, there are still discrepancies in details. To achieve further consistency of 2D images, other tools need to be explored.

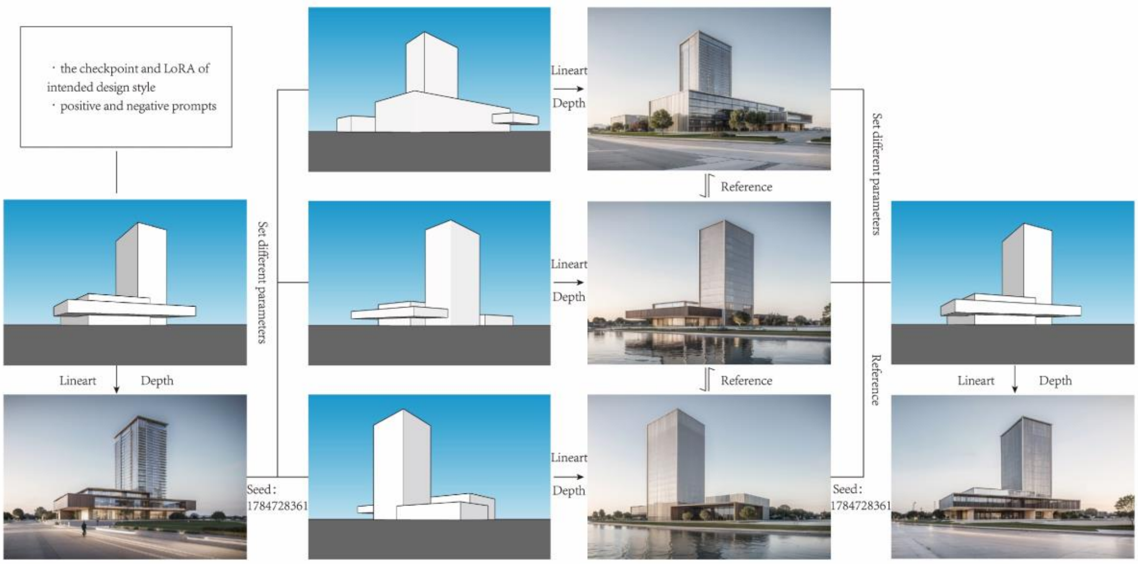

Fig. 4 The process of generating multi-perspective images with higher consistency in Stable Diffusion (Drawn by the authors)

4 Present Approaches of 3D Model Generation

Due to the inevitable randomness in the image generation process, it can hardly solve the problem of ontology inconsistency solely by relying on graphical AIGC tools. It can be inferred that generating a unified 3D model is the most effective and complete solution, which is also beneficial for interfacing with the subsequent architectural design process. At present, there are two common approaches of 3D generation: text-to-3D and image-to-3D, which are mostly used in the fields of game and animation generation.

4.1 Text-to-3D

Some AI tools for generating 3D models have emerged, such as Meshy, Masterpiece X and Genie (Table 1). These tools allow users to input texts (prompts) to generate 3D models, with output formats typically including *. obj, *. fbx, etc. These tools can generate models with textures in short time, and some can generate multiple models at once. In the early stages of architectural design, it may be useful for rough morphological research. However, as it main focuses on the field of game design, there are problems such as the models lack a sense of scale and rounded edges.

Table 1 Overview of different types of text-to-3D tools

Tool	Input	Output	Advantage	Disadvantage
Meshy	prompts	3D model with maps (*.fbx/*.obj/*.glb/*.usdz/ *.stl/*.blend)	quick generation process; high quality and modifiable model	unsharp delineation; vague model components; non-editable model
Master piece X	prompts	3D model with maps (*.obj)	model with sharp corners and a strong sense of Volume; ability of importing into a virtual reality environment; clear components;	combination of building clusters; glued-together blocks; poor quality model
Genie	prompts	3D model with maps (*.fbx/*.obj/*.glb/*.usdz/ *.stl/*.blend)	multiple generation models; iterative operation	sticky components; poor editing capabilities

4.2 Image-to-3D

Generating 3D models from images is another possible solution. Such tools can be used to generate a model of the whole building by inputting images of the building facade. At present, several AI tools for generating 3D models have emerged (Table 2), and some of them also support the import of video.

Compared with Text-to-3D, the approach of image-to-3D is closer to the working flow of architectural design. The architect could generate multiple images using prompts or sketches, and then use image-to-3D tools to generate building models. Videos generated by SORA could also be imported into Luma AI to generate the 3D model (Fig.5). However, due to the expression of complex bodies, like spot light, in Sora video is still not perfect, and the model is somewhat blurred.

Overall, the current AI tools for generating models based on text or image are not mature, but tools like Luma AI which can provide accurate models based on multiple images are expected to be applied in architectural design in the future.

Table 2 Overview of different types of image-to-3D tools

Tool	Input	Output	Advantage	Disadvantage
Wonder 3D	single image	3D model with maps (*.ply)	strong ability to recognize shapes	unclear relationship between shapes; model with undulating plane and over-smooth edges; non-ideal generating effect
IMAGE to STL	single image without background	3D model with maps (*.glb/ *.obj/ *.ply)	fast model generation; rich output format; suitable mode for LOGO or terrain modelling	inaccurate image recognition; non-ideal quality, detail and texture of model; model doped with blocks
Tripo AI	text / single image	3D model with maps (*.glb/)	strong ability to recognize planes and basic shapes; better quality of generation	no adjustable parameters; model of a single trilateral closed grid.
Luma AI	a video with a stable scene; multiple pictures from different perspectives for a subject	3D model (Luma field format)	outstanding ability to draw model and identify the form, especially the transparent components such as glass.	non-editable model; incompatible with architectural software

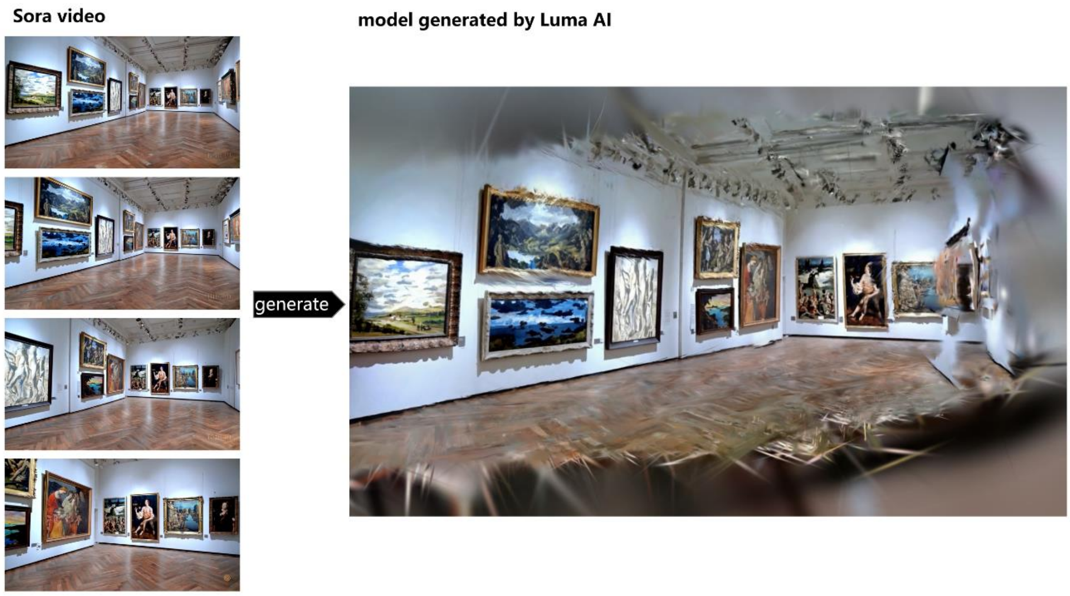

Fig. 5 The model generated by Luma AI based on a video generated by Sora (Drawn by the authors)

5 Exploration of the Feasible AIGC Path for Generating Architectural Models

To generate 3D model is an ultimate object of AIGC-assisted architectural design. It requires at least the following four aspects: (1) generating results in line with the spatial composition logic of architectural design; (2) reducing manual intervention as much as possible; (3) having high generating efficiency; (4) the output files are suitable for the needs of the architectural design work in the subsequent stages. It is difficult to achieve the above objectives at the same time with the current technology level. But based on the work logic of architectural design, the future technical path of AI-assisted architectural design can be discussed.

5.1 Path 1：Text-to-3D Models

Text-to-3D path has an advantage in dealing with the consistency of multiple perspectives of single model. This path enables AI to modify model details in a holistic and continuous manner, effectively circumventing the limitations that 2D-to-3D tools only define one or a few views of the model. However, according to the currently available Text-to-3D tools, there are three insurmountable difficulties. Firstly, the poor quality of model generation for uncertainty cannot be applied with simple types. Secondly, it is difficult to accurately describe the composition of the building in a textual way. Thirdly, the results of generation that adhere to the functional logic and the myriad of design-specific rules is a formidable obstacle. Consequently, the feasibility of utilizing this path for architectural design in the immediate future appears tenuous.

5.2 Path 2：Text-Video-3D Models

Text-to-video tool is able to generate continuous video scenes, which has the potential to promote its application in architecture field. As an open-sourced software, Stable Video Diffusion 1.1 provides two input methods, prompt or image. Based on the prompt or image, high-quality video files can be obtained by adjusting camera parameters. The video or video-based multi-angle images can be input into softwares such as Luma AI for model generation.

However, generating a panoramic video of a building through text also faces the same problems in path 1: the building form is difficult to be described accurately, and the generated result is difficult to conform to the logical needs of the design. Furthermore, if the scene in the generated video is complex or contains contradictions, the existing tools cannot be used to generate 3D models. For example, it was found that a video of a seaside castle generated by Sora was invalid when imported into Luma AI.

5.3 Path 3: Sketch-Multiple Pictures-Contradiction Elimination-3D Model

Through the above analysis, it can be seen that the overall control of the architect is essential in the preliminary stage of design with AIGC tools. Generally a sketch drawn by architect is based on reasonable composition logic which determines the outline of the building according to the site and function requirements. It can give a direction of the AI generation while ensuring the rationality of the generated results.

According to the current technological level, the Image-to-Image mode cannot completely eliminate the contradictions between different pictures. A feasible solution is to optimize the generation process of the 3D model tools, such as specifying a prior picture or a prior part in one picture which will be adopted preferentially. The false content in other pictures will be ignored so that the consistency can be maintained.

It’s considered that the path of Sketch-Multiple Pictures-Contradiction Elimination-3D Model is a possible approach to architectural model generation. This path puts forward certain technical requirements for the specialization of 3D model tools at present, and points out a feasible direction for the development of AI-assisted architectural design.

6 Conclusion

As a widespread problem in the field of current AIGC, ontology inconsistency poses a significant barrier to the in-depth application of AIGC in the field of architectural design. This article describes the problem and attempts to optimize the solution within the image generation tool, resulting in a limited improvement in the generated outcomes. To completely resolve this issue and integrate with the architectural design workflow, generating 3D models is considered as an effective solution. The article further analyses the characteristics and advantages and disadvantages of different 3D model generation tools, such as *text to 3D* and *image to 3D*, and discusses their potential for application in the field of architectural design in the future. This article considers that the path of Sketch-Multiple Pictures-Contradiction Elimination-3D Model is more appropriate for architectural design logic. It is important that additional step should be added to eliminate contradictions and the model generation tools should support local re-generation. This path provides a reference for the further development of AI-assisted architectural design. However, the tools mentioned in the article are currently under rapid evolution, so the feasible paths presented here are only phased cognition. It is expected that with the iteration of AIGC, highly integrated tools will emerge in the coming future and solve the current ontology inconsistency problems.

Acknowledgements
This research was supported by the National Natural Science Foundation of China (52378043)

References

1. Wanyu He, Leong Shong Jackie Yong. (2023). Exploration of Generative Artificial Intelligence in Architectural Design A Case Study of LookX AI Cloud. Architectural Journal 10:36-41
2. Yanfeng Huang, Yirong Zhou. (2023). Exploration on the Generative Architecture Design Method with AIGC Technology: A Case of the Overall Design Process of Generating Architectural Image with Prompt as a Key Word[J]. Urbanism and Architecture 15: 202-206+213
3. Guo Li, Tiandu Zhang, Zhiwei Xing. (2023). On the Eve of the Technological Revolution: Architectural and Scene Design Innovation under the Wave of Generative AI Tools. Chinese & Overseas Architecture 9:24-28
4. Wang Y, Pan Y, Yan M, et al. (2023). A Survey on ChatGPT: AI–Generated Contents, Challenges, and Solutions. IEEE Open J Comput Soc 4: 280–302.
5. Chao Yuan, Hao Zheng. (2023). A New Architectural Design Methodology in the Age of Generative Artificial Intelligence. Architectural Journal 10:29-35