

刘函宁¹ 吴昊¹ 谢星杰¹ 袁梦豪¹ 袁烽^{1*}

1. 同济大学建筑与城市规划学院; philipyuan007@tongji.edu.cn

Liu Hanning¹ Wu Hao¹ Xie Xingjie¹ Yuan Menghao¹ Yuan Feng^{1*}

1. College of Architecture and Urban Planning, Tongji University; philipyuan007@tongji.edu.cn

国家重点研发计划“政府间国际科技创新合作”项目(2022YFE0141400);上海市级科技重大专项(2021SHZDZX0100);教育部第二批产学合作协同育人项目(202102560007)

生成式人工智能工具辅助建筑设计中的提示词撰写方法研究

——以城市露营地设计为例

Prompt Writing Approach in GAI Tools Aided Architectural Design: Taking Urban Camp Center Design as an Example

摘要:近期,生成式人工智能(Generative Artificial Intelligence, GAI)工具将扩散模型与提示词工程相结合,能够通过输入描述性文字或参考图片的方式,快速生成建筑方案图像,从而提高建筑师的工作效率。但是,许多建筑师在使用这类工具时通常无法得到与他们的设计意图相符的结果。所以,本研究的目的是探索这类工具如何融入建筑师的设计流程中,并提出一种系统化的提示词撰写方法,帮助使用者获得更符合他们设计意图的结果。首先,本文介绍了GAI技术的发展历程、人工智能生成内容(Artificial Intelligence Generated Content, AIGC)的概念定义以及4种图像类GAI工具。其次,本文采用了对照实验的方法,设立4个对照组,分别将相同的文字、图片作为提示词输入给四种工具,用FID的一致性评价方法分析各个工具生成的图像结果。最后,本文选定Stable Diffusion这一工具,并以城市露营地设计为例,探索它在建筑设计流程中的作用,展示了一套提示词撰写流程,包括“收集设计意向图”“结合人工智能与人工构思处理大量图像信息”“参照建筑专业术语类目表完成文本提示词的撰写”3个阶段。总的来说,本文提供了一套系统化的提示词撰写方法,使得GAI工具能够输出符合建筑师设计意图的结果,极大地帮助建筑师提高工作效率。

关键词:生成式人工智能工具;扩散模型;提示词工程;人工智能辅助建筑设计;城市露营地设计

Abstract: Recently, Generative Artificial Intelligence (GAI) tools combine diffusion models with prompt engineering, which can quickly generate architectural images by inputting text or images, thereby improving the work efficiency of architects. However, many architects often do not get results that match their design intent when using such tools. Therefore, the purpose of this study is to propose a systematic method of prompt writing to help users obtain results that are more in line with their design intentions. First, this paper introduces the development history of GAI technology, the concept definition of Artificial Intelligence Generated Content (AIGC), and four image-based GAI tools. Secondly, this paper adopts the method of controlled experiment, setting up 4 control groups, inputting the same text and images as prompt into the four tools respectively, and analyzing the image results generated by each tool with the consistency evaluation method of FID. Finally, this paper selects Stable Diffusion as a tool, and takes the design of urban camp center as an example to explore its role in the architectural design process, showing a set of prompt writing process in three stages: “collecting design references”, “combining artificial intelligence and human conception to process a large amount of image information”, and “completing the text prompt writing with reference to the architectural terminology table”. In general, this paper provides a systematic prompt writing methods, so that the GAI tool can output the results that meet the architect’s design intention, which greatly helps the architect to improve work efficiency.

Keywords: GAI Tools; Diffusion Model; Prompt Engineering; AI-Aided Architectural Design; Urban Camp Center Design

1 引言

人工智能(AI)在过去几年取得了突飞猛进的发展,从2016年AlphaGo战胜世界围棋冠军李世石,到2022年ChatGPT与Stable Diffusion等生成式人工智能(Generative Artificial Intelligence, GAI)工具的出现,再到2023年多模态模型GPT-4的发布,AI不断向前推进着技术的边界,在各个领域展现出惊人的应用潜力。

GAI工具让AI技术真正走入大众的视野,GAI工具生成的内容被称为人工智能生成内容(Artificial Intelligence Generated Content, AIGC)。AIGC较高的平均水准以及GAI工具的易用性,使得每一个渴望探索并亲手使用AI技术的人,都能够创造出效果惊人的内容。

1.1 GAI技术的发展历程

在2014年,生成对抗网络(Generative Adversarial Networks, GANs)^[1]的概念一经提出,就被视为图像生成领域的里程碑。GANs由生成器(Generator)和判别器(Discriminator)两部分组成,其中生成器扮演艺术家的角色,不断生成新的作品,而判别器扮演艺术品鉴别员的角色,鉴别作品的优劣并给出反馈来帮助生成器生成更逼真的图像。

在2017年,Ashish等发布的*Attention Is All You Need*^[2]一文中首次提到了Transformer模型的概念,该模型由多个编码器和解码器组成,强调了自主注意力机制在人工智能模型中的重要性。

在2021年出现的CLIP(Contrastive Language-Image Pre-Training)^[3]能够分别学习文本和图像的特

征,将图片分类任务转换为图文匹配任务,从而实现高效的多模态识别、融合与转换,标志着人工智能在多模态领域迈出了重要的一步。

在2022年,扩散模型(Diffusion Model)^[4]开始流行。扩散模型的运行原理与GANs类似,但其训练过程却更加稳定高效,通过多轮的前向过程(给图片加入噪声的过程)和逆向过程(给图片去除噪声的推断过程)快速生成图像或文字结果,再次推动了GAI技术的变革与创新。

1.2 AIGC的概念定义

从内容生产者的角度来讲,AIGC是继专业生产内容(professional-generated content, PGC)与用户生产内容(user-generated content, UGC)之后的一种新的内容生产者创作的内容,强调创作者的角色从人类转换为人工智能。从内容形式的角度来讲,AIGC可以是文字、图像、语音、音乐、视频、代码等,强调AI生成的内容形式各种各样。

总的来说,AIGC既是从内容生产者视角进行分类的一类内容,又是一种内容生产方式,还是用于内容自动化生成的一类技术集合^[5]。

1.3 图像类GAI工具的介绍

由于建筑设计师需要借助大量图像来表达自己的设计意图和想法,所以在众多类型的GAI工具中,与建筑学专业相关度最高的是图像类生成工具。

以2022年之后发布的四种图像类GAI工具为例,它们都以文字或图片作为提示词(Prompt),都具备文生图、图生图的功能,各自还具备不同的专项功能与擅长的风格。关于这四种图像类GAI工具的简介,见下文中的表1。

表1 常见的几种图像类GAI工具简介

工具名称	Midjourney	DALL·E 2	Stable Diffusion	FUGenerator
发布时间	2022年3月	2022年1月	2022年8月	2023年3月
发布者	David Holz	OpenAI	Stability·AI	Tongji CAUP
特点	擅长艺术风格与逼真图像 有针对卡通风格优化的模型	擅长写实风格 有局部调整与补全功能	可调参数多 成果水平参差不齐	提供包含建筑专业词汇的 预训练模型与参数

(来源:作者自绘)

2 研究方法概述

2.1 研究内容

GAI工具在生成图像时,需要使用者决定的内容包括两部分:大量的可调参数以及文本或图像形式的提示词。其中参数调整的部分对于图像自身的质量好坏影响较大,而提示词的部分对于图像是否符合建筑

设计师的设计意图和想法影响较大。

所以本研究的内容重点为:分析4种图像类GAI工具生成的图像特征,对比它们在建筑设计领域的适用性,挑选出最适合用于建筑设计工作的GAI工具。然后,使用这个工具来完成城市露营地的设计工作,并提出一个较为系统的提示词撰写方法。

2.2 研究方法

在 GAI 工具对比阶段,采用对照实验的方法,分别将相同的文字、图像作为提示词输入给四种工具,用 FID 的评价方法分析各个工具生成的图像结果与建筑师设计意图的一致程度。

在城市露营地设计阶段,采用归纳总结的方法,将提示词的撰写方法归纳为“收集意向图”“结合人工智能与人工构思处理大量图像信息”“参照建筑专业术语类目表完成文本提示词的撰写”这 3 个阶段。

3 实验内容

常规的建筑设计工作流程需要建筑设计师先了解设计任务的背景,在此基础上思考建筑的材料、结构、形态,并搜集能够激发创作灵感的图像。然后,将这些繁杂的信息进一步整理,绘制初步的方案草图,再对方案草图不断调整,直到获得最终满意的结果。

借助图像类 GAI 工具来进行的建筑设计工作流程(图 1)与常规的建筑设计工作流程类似,只是由人工智能代替人工完成了图像信息的解读、草图绘制以及草图深化这几个阶段的工作。

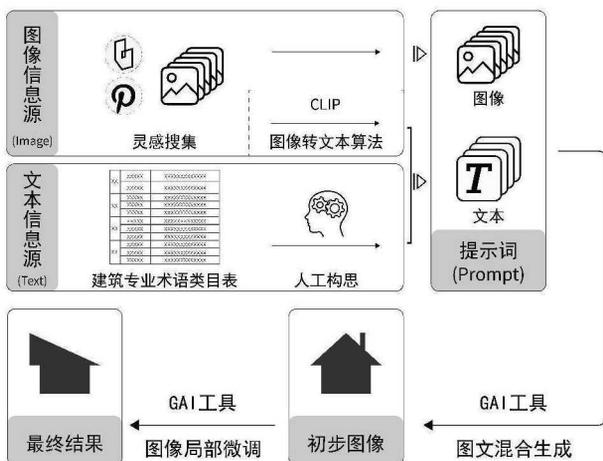


图 1 GAI 工具辅助建筑设计的典型工作流程

(图片来源:作者自绘)

3.1 不同 GAI 工具的效果对比

为了让 GAI 工具生成图像,我们需要输入提示词。而提示词的来源分为图像和文本两种类型,其中图像可以直接作为提示词输入,而文本需要通过 CLIP 算法与人脑的创意构思相结合,归纳总结得到提示词。将相同的提示词输入到 4 种图像类 GAI 工具中,保持 GAI 工具内的其他参数为默认状态,观察生成的结果(图 2)。

我们可以发现 4 种 GAI 工具都能生成质量较高的图像,但是图像之间在风格上有较大差异。为了客观地评价各 GAI 工具生成的图像结果孰优孰劣,我们生成更多图像,借助 Fréchet Inception Distance (FID) 算

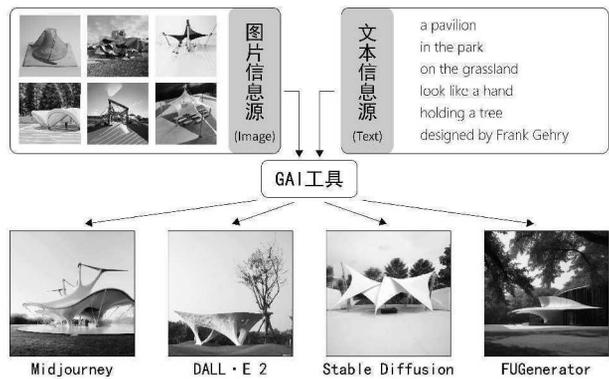


图 2 GAI 工具依据相同的提示词生成的结果

(图片来源:作者自绘)

法^[6]来计算生成图像和真实图像的距离,计算公式见式(1),计算结果数据越小越好(图 3)。

$$FID = \|\mu_r - \mu_g\|^2 + Tr[\Sigma_r + \Sigma_g - 2(\Sigma_r \Sigma_g)^{1/2}] \quad (1)$$
 式中, μ_r 为真实图像的特征均值; μ_g 为生成图像的特征均值; Σ_r 为真实图像的协方差矩阵; Σ_g 为生成图像的协方差矩阵; Tr 为矩阵的迹。

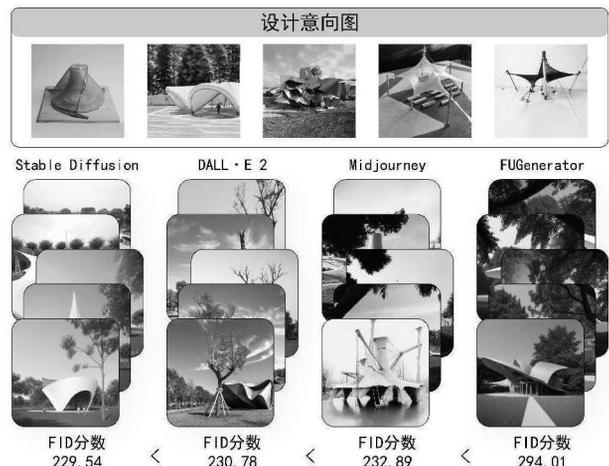


图 3 不同工具生成的图像结果与 FID 评分对比

(图片来源:作者自绘)

从 FID 评分的对比(图 3)中,我们能够看出 FUGenerator 的 FID 分数数值最大,由于它的预训练数据库中缺乏装置艺术类图像,暂不适用于露营地的设计;而 Stable Diffusion 的 FID 分数数值最小,生成的图像最符合建筑师的设计意图。因此,本研究选择该工具进行下一步研究提示词的系统化撰写方法。

3.2 提示词的系统化撰写方法

3.2.1 处理大量图像信息

在建筑设计的前期阶段,建筑师通常会收集大量的设计意向图。然而,由于 GAI 工具只能接受一张图像作为图像提示词,因此我们需要结合 CLIP 算法与人工构思来将其余的图像信息转化为文本信息(图 4),并根据建筑专业术语类目表进行分类,作为文本提示词。

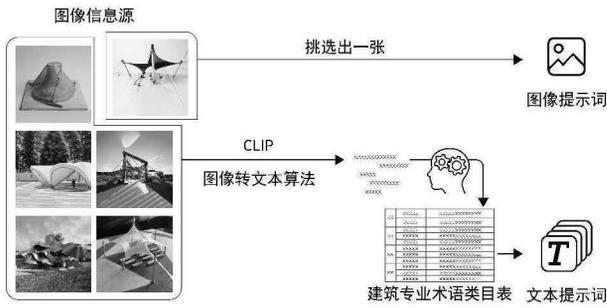


图 4 大量图像信息的处理方法
(图片来源:作者自绘)

3.2.2 参照建筑专业术语类列表

为了获得符合预期的图像结果,使用者应该尽可能清晰、详细地描述自己的意图,以帮助 GAI 工具更好地理解并生成符合要求的图像。随时查看建筑要素类列表(表 2)能够帮助我们查漏补缺,撰写出较完整的提示词。并且,类列表的前两级分类可以在我们进行其他类型的建筑设计任务时重复使用。

表 2 建筑专业术语类列表(以城市露营地设计为例)

一级分类	二级分类	具体提示词
图像的类型	图像内容的所属领域	A pavilion, A service center
主体的形态	形式	Look like a tent, Smooth shape
	面积、高度、层数	250 square meters, Height is 8m, Two floors
	特色的建筑构件	A slope towards roof
	材料与结构	Wooden structure, Membrane structure
	设计者	Designed by ZAHA
周围的环境	自然环境	Around a tree
	使用者的数量	Many people standing and eating around it
	使用者的行为	Attending a weeding party or an film festival
	所处的具体位置	In the city park, On the lawn
杂项	图像的视角	Bird view, exterior
	图像的制作方式	V-Ray render image, Photography
	发布的媒体	Archdaily, Dezeen
	分辨率/清晰度	4k, uhd

(来源:作者自绘)

4 总结

城市露营地作为一种小众的建筑类型,更偏向于装置艺术领域,各 GAI 工具的大模型数据库中的相关信息较为欠缺。但是只要我们依靠系统化的提示词撰写方法,借助 CLIP 算法与建筑专业术语类列表,就能够克服这些困难,获得与设计意图相符的成果(图 5)。

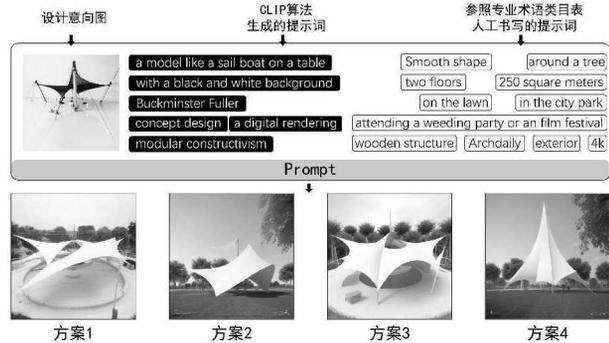


图 5 最终成果
(图片来源:作者自绘)

参考文献

- [1] GOODFELLOW I, POUGET-ABADIE J, MIRZA M, et al. Generative adversarial networks[J]. Communications of the ACM, 2020, 63(11): 139-144.
- [2] VASWANI A, SHAZEER N, PARMAR N, et al. Attention is all you need[J]. arXiv, 2017, 30:1706.
- [3] RADFORD A, KIM J W, HALLACY C, et al. Learning transferable visual models from natural language supervision[C]//International Conference on Machine Learning. PMLR, 2021: 8748-8763.
- [4] ROMBACH R, BLATTMANN A, LORENZ D, et al. High-resolution image synthesis with latent diffusion models[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2022: 10684-10695.
- [5] 李白杨,白云,詹希旒,等.人工智能生成内容(AIGC)的技术特征与形态演进[J].图书情报知识, 2023, 40(1): 66-74. DOI: 10.13366/j. dik. 2023. 01. 066.
- [6] HEUSEL M, RAMSAUER H, UNTERTHINER T, et al. Gans trained by a two time-scale update rule converge to a local nash equilibrium[J]. Advances in Neural Information Processing Systems, 2017, 12: 6629-6640.