跨平台图像喂入对 AIGC 文生图生成稳定性的影响——基于统一提示条件的实证研究

谢尚东，王凯宏，王旷，彭灏琳
广东外语外贸大学，广东 广州 510006
DOI: 10.61369/SSSD.2026040017

摘　　要： 随着文生图工具进入创作流程，重复生成的不稳定性影响人物一致性与产出效率。本文在统一结构化 Prompt（P2）条件下，以四个平台开展真实生成实验（K=15），对比 Feed=0 与 Feed=5 两种喂图条件下的人物主体一致性变化。结果显示，多数平台在喂入参考图后稳定性提升，但提升幅度存在平台差异。研究为跨平台创作中图像喂入策略的使用提供可审查的经验依据。

关 键 词： AIGC；文生图；图像喂入；生成稳定性；跨平台比较

The Impact of Cross-Platform Image Feeding on the Generation Stability of AIGC Text-to-Image: An Empirical Study Based on Unified Prompt Conditions

Xie Shangdong, Wang Kaihong, Wang Kuang, Peng Haolin
Guangdong University of Foreign Studies, Guangzhou, Guangdong 510006

Abstract: As text-to-image tools become integrated into creative workflows, the instability of repeated generation undermines character consistency and production efficiency. Under unified structured Prompt (P2) conditions, this paper conducts real-world generation experiments (K=15) across four platforms, comparing changes in character subject consistency between two image feeding conditions: Feed=0 and Feed=5. Results indicate that most platforms show improved stability after feeding reference images, though the magnitude of improvement varies across platforms. This study provides a verifiable empirical basis for the application of image feeding strategies in cross-platform creative processes.

Keywords: AIGC; text-to-image; image feeding; generation stability; cross-platform comparison

引言

近年来，随着生成式人工智能技术的快速发展。其中，以扩散模型为代表的生成方法已成为当前文生图系统的主流技术路线。能够根据文本提示生成具有较高完成度和视觉一致性的图像结果[1-3]。

文本提示保持一致，多次生成仍可能出现人物主体外观差异，从而增加筛选成本并降低连续创作任务中的可控性。相关研究从模型结构与采样机制角度对这一问题进行了分析，指出扩散过程中的随机性是导致结果波动的重要原因之一[1-3,7-8]。

为提升生成结果的可控性，各类文生图平台在模型应用层面引入了多种条件控制策略。图像喂入（image conditioning）通过提供参考图像对生成施加视觉约束，常用于提升人物主体一致性[4-6]。

文本提示（Prompt）的设计方式也被认为是影响文生图生成结果的重要因素。已有研究指出，相较于自然语言描述，结构化 Prompt 通过明确限定生成主体、视角及场景要素，有助于减少文本语义歧义，从而提升生成结果的一致性与可预期性[11,12]。在创作实践中 Prompt 与图像喂入往往同时被使用，但二者在不同平台中的协同效果仍缺乏系统性的实证分析。

多数的研究基于受控环境，跨平台真实实践不足；平台实现差异导致策略效果不必然一致，所以需要统一输入对比。

图 1-1 展示了本文的整体研究思路与实验流程框架，从研究问题提出、实验条件设置到结果分析的主要步骤均在图中予以概括，为后文方法与实验结果的展开提供整体参考。

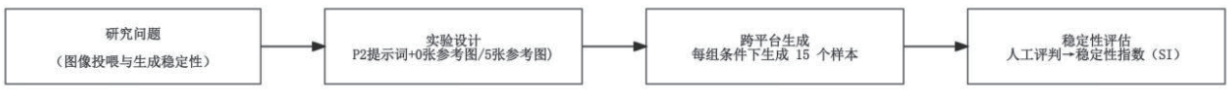

图 1-1实验流程框架

一、文献综述与研究定位

（一）文生图生成稳定性的相关研究

关于文生图相关工作从不同角度探讨了文本提示、参考图像等输入条件对生成结果的影响，相关研究从模型结构与方法层面对文生图生成过程进行了系统探讨，为理解其运行机制提供了基础，但真实平台环境中的跨平台实践研究仍相对有限[1-3]。

生成稳定性是评价文生图系统表现的重要维度之一。部分研究从模型结构角度出发，通过改进扩散过程或引入约束机制，以降低生成结果在多次采样中的随机波动。这类研究通常以生成一致性、语义对齐程度或视觉相似性作为分析的对象，探讨模型在不同条件下的稳定表现[1-3,7]。也有研究尝试从评价方法角度对生成一致性进行量化分析，为稳定性比较提供技术支撑[8,9]。

需指出上述研究多基于受控模型环境或特定算法框架展开，结论在一定程度上依赖于具体实现条件。实际应用场景中，文生图系统往往以平台化形式呈现，用户难以直接干预模型结构或采样机制。所以模型层面的稳定性研究与创作实践之间仍存在一定距离。

（二）图像喂入与条件控制机制研究

图像喂入作为一种常见的条件输入方式被广泛用于在生成过程中引入额外的视觉约束。相关研究表明，通过向模型提供参考图像，可以在一定程度上影响生成结果的主体结构、外观特征或整体风格，从而提升角色一致性或主体保持能力[4-6]。在人物生成与风格延续等任务中，基于参考图像的条件控制方法已被证明具有一定效果。

目前现有研究多集中于单一模型或特定技术框架，不同平台在图像喂入实现方式上的差异关注较少。所以有必要在跨平台情境下对图像喂入策略进行实证比较，以更贴近真实创作环境。

（三）Prompt 设计与生成控制研究

Prompt 的设计方式被认为对文生图生成结果具有重要影响。已有研究从 Prompt 工程的角度探讨了文本描述结构对生成结果的影响，指出通过结构化或分层描述，可以在一定程度上减少语义歧义并提升生成结果的一致性[11,12]。相关工作主要关注 Prompt 的表达方式、信息密度及其对生成质量的影响。

Prompt 设计的重要性已得到广泛讨论，现有研究往往将 Prompt 作为主要变量展开比较，较少将它作为受控条件，用以考察输入策略的效果。在实际创作中 Prompt 与图像喂入通常同时被使用，其协同作用关系仍缺乏系统性的实证分析。

（四）研究定位

现有研究主要从模型机制与控制策略两个方向探讨文生图生成稳定性问题，但多数研究基于受控实验环境，较少涉及真实平台条件下的跨平台比较。基于此，本文在统一结构化 Prompt（P2）条件下，通过真实平台环境中的生成实验，比较不同喂图条件下生成稳定性的变化，并考察该策略在不同平台中的表现差异。

二、研究设计与方法

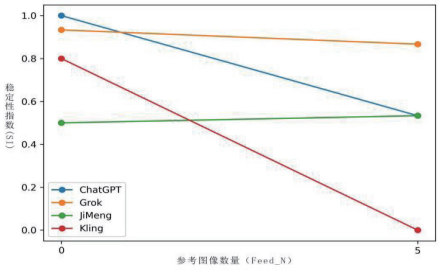

图 3-1 图像投喂对不同平台生成稳定性的影响 ,K=15

本研究在统一文本提示条件下考察图像喂入策略对文生图生成稳定性的影响，并比较其在不同平台中的表现。为贴近实际创作环境，实验在真实平台条件下完成，并注重流程的可审查性与可复现性。

（一）实验平台选择

本文选取四个当前创作实践中使用频率较高的文生图平台作为研究对象，分别为 JiMeng、Kling、Chat GPT 和 Grok。采用多平台实验有助于避免单一系统结论的局限。

说明本文不对平台性能进行评价，而仅比较相同输入条件下图像喂入策略的表现差异。

（二）Prompt 设置与受控条件

为减少文本提示对生成结果的干扰，实验统一采用结构化 Prompt（P2）。该 Prompt 明确限定人物主体、视角、表情及光照条件，以降低语义歧义并提高生成结果的可比性。

各平台均使用完全一致的 Prompt 文本，未进行平台定制化调整。

（三）图像喂入条件设计

图像喂入条件设置上，本文对比两种情形：不引入参考图像（Feed=0）与引入多参考图像（Feed=5）。参考图像选取同一人物的多张真实照片，用于为生成过程提供稳定的视觉约束。

（四）实验流程与重复生成设置

实验流程如图3-1所示。对于每一平台，两种图像喂入条件下分别进行重复生成实验。每一条件下生成次数设为 K=15，以减弱单次生成结果的偶然性影响。实验过程中，所有生成操作均由人工完成，并记录生成时间、运行编号及结果文件名。

（五）数据记录与分析方法

为保证结果可审查性，对每次生成结果进行记录，并基于人物主体一致性进行人工稳定性判断。判断标准主要关注人物是

否保持一致、是否出现明显性别偏移或面部结构变化，并据此将结果划分为“低”“中”“高”三个等级。实验结束后整理运行日志，以描述性方式比较不同平台在两种喂图条件下的稳定性变化。

三、实验结果与分析

本章分析不同平台在两种图像喂入条件下的生成结果，考察参考图像对文生图生成稳定性的影响。

（一）生成稳定性的整体变化趋势

在不引入参考图像（Feed=0）的条件下，各平台生成结果的稳定性存在较为明显的波动。同 Prompt 多次生成中，人物主体外观差异较为常见，部分结果难以视为同一人物的连续呈现。

对比之下，在引入多参考图像（Feed=5）的条件下，多数平台的生成结果在人物主体一致性方面表现出更为稳定的趋势。人工稳定性判断中，“中”和“高”等级结果的比例整体有所上升，偏离预期的生成错误出现频率相应降低。该结果表明，参考图像能够在一定程度上缓解生成随机性带来的不稳定问题。

（二）不同平台在喂图条件下的表现差异

引入参考图像整体呈现正向效果，但不同平台的响应程度存在差异。部分平台在喂图条件下人物特征更为一致，而部分平台虽错误减少但仍存在一定差异。

由于本文实验环境以真实平台条件为前提，相关差异更多体现为创作实践层面的表现，而非模型结构层面的对比。

（三）典型生成问题的变化情况

在不喂图条件下，较为常见的问题包括人物性别偏移、面部特征不稳定以及整体风格难以保持一致等。这类问题在多次重复生成中具有一定的偶发性，但总体出现频率较高。

喂图条件下，上述问题的出现频率明显下降，尤其是人物性别偏移现象，其发生次数在多数平台中均有所减少。个别生成结果仍可能出现细节差异，但整体而言，生成结果更容易维持人物主体的连续性。

实验结果与第三章所述方法设计进行对照可以发现，统一结构化 Prompt（P2）为生成过程提供了稳定的文本语义基础，而图像喂入则在此基础上进一步强化了视觉约束。两者的结合，使生成结果在人物一致性方面表现得更为稳定。

需说明本文采用 K=15 的重复生成规模，为观察稳定性变化的趋势，而非严格的统计显著性检验。本文结果侧重于实践层面的经验分析，而非数值层面的精确比较。

（四）本章小结

实验结果可以看出在统一 Prompt 条件下，引入多参考图像在多数平台中能够有效提升文生图生成的稳定性，减少明显偏离语义预期的生成错误。不同平台在喂图响应程度上存在差异，但整体趋势支持图像喂入作为一种具有实践价值的生成控制策略。

本章结果为后续讨论图像喂入在创作实践中的意义及其局限性提供了实证基础。

四、讨论与结论

（一）图像喂入作为生成控制策略的意义

本次实验所设条件下，当保持文本提示不变并引入多参考图像，多数平台生成结果在人物主体一致性方面出现了较为明显的变化。靠文本提示相比，图像喂入为生成过程提供了额外的视觉约束，使人物主体在多次生成中更容易保持一致。该发现与已有关于生成一致性与可控性的研究结论在方向上保持一致，但它具体表现贴近真实平台环境中的创作实践[7]。

创作流程的角度来看，图像喂入不是用于替代文本提示，而是与 Prompt 形成互补关系。已有研究从 Prompt 工程与条件控制角度讨论了不同输入方式在生成过程中的作用机制，而本文的实验结果进一步表明，在统一 Prompt 条件下，参考图像能够在一定程度上缓解生成随机性带来的不稳定问题[11,12]。这一点对于需要反复生成并筛选结果的创作场景具有现实意义。

（二）跨平台差异与创作

不同平台生成结果进行对比后，各平台在引入参考图像后的变化幅度并不一样。整体趋势虽一致，但各平台在生成稳定性的提升幅度上并不完全相同。这种差异并不意味着平台技术水平的优劣，也许与平台在模型实现方式、参考图像融合策略及参数设置上的差别有关。相关研究指出，生成系统的可控性表现往往与其内部条件控制机制密切相关[4-6]。

这些差异提示创作者在跨平台使用文生图工具时，不能简单假设同一生成策略在不同平台中具有完全一致的效果。反之，结合具体平台特性，对图像喂入与 Prompt 设计进行适当调整，可能更有助于获得稳定的生成结果。这一判断也与人工智能艺术与创作研究中关于工具差异影响创作过程的讨论相呼应[13,14]。

（三）研究局限与未来研究方向

需说明本文的实验规模与条件仍存在一定局限性。第一，实验采用 K=15 的重复生成次数，目的在于观察生成稳定性变化的整体趋势，不是进行严格的统计显著性检验。本文聚焦于人物生成任务，统一使用结构化 Prompt（P2），未进一步比较不同 Prompt 类型或其他生成任务情境下的效果。第二，主要考虑了平台可访问性与时间成本因素，未对参考图像数量进行更细粒度的分组比较。相关参数的进一步细化，可能有助于揭示图像喂入效果的边界条件。

未来在扩大实验规模的基础上，引入更多生成任务类型，进

一步探讨 Prompt 设计与图像喂入策略之间的交互关系。结合定量评价指标与人工判断方法，可为生成稳定性的评估提供更为多维的视角。这类工作已在相关生成一致性与评价研究中被提出，仍有进一步发展的空间[8,9]。

（四）结论

最后，本文在真实平台环境中通过统一 Prompt 条件下的生成实验，验证了图像喂入策略在提升文生图生成稳定性方面的实际效果。不同平台在喂图响应程度上存在差异，整体结果支持图像喂入作为一种具有实践价值的生成控制手段。

平台的差异在同条件下呈现出的生成表现并一样，这种差异在实际创作过程中会直接影响生成结果的稳定程度。已有研究从人工智能美学与社会技术系统角度对这一问题进行了讨论，本文的实证结果为相关讨论提供了经验层面的补充[15,16]。

参考文献

[1]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. New Orleans: IEEE, 2022: 10684-10695.

[2]Ho J, Jain A, Abbeel P. Denoising diffusion probabilistic models[C]//Advances in Neural Information Processing Systems. Vancouver: NeurIPS, 2020: 6840-6851.

[3]Saharia C, Chan W, Saxena S, et al. Photorealistic text-to-image diffusion models with deep language understanding[C]//Advances in Neural Information Processing Systems. New Orleans: NeurIPS, 2022: 36479-36494.

[4]Zhang L, Rao A, Agrawala M. Adding conditional control to text-to-image diffusion models[C]//Proceedings of the IEEE/CVF International Conference on Computer Vision. Paris: IEEE, 2023.（Open Access 版）获取路径：https://openaccess.thecvf.com/content/ICCV2023/papers/Zhang_Adding_Conditional_Control_to_Text-to-Image_Diffusion_Models_ICCV_2023_paper.pdf

[5]Ruiz N, Li Y, Ouyang P, et al. DreamBooth: Fine-tuning text-to-image diffusion models for subject-driven generation[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Vancouver: IEEE, 2023: 22500-22510.

[6]Gal R, Alaluf Y, Atzmon Y, et al. An image is worth one word: Personalizing text-to-image generation using textual inversion[C]//Proceedings of the International Conference on Learning Representations. Vienna: ICLR, 2023.

[7]Liu B, Zhang Y, Hu H. Consistency and controllability in text-to-image generation: A survey[J]. ACM Computing Surveys, 2023, 56(6): 1-36.

[8]Liu F, Ren Y, Huang J. Evaluating visual consistency in text-to-image generation[C]//Proceedings of the ACM International Conference on Multimedia. Lisbon: ACM, 2022: 4152-4161.

[9]Hessel J, Holtzman A, Forbes M, et al. CLIPScore: A reference-free evaluation metric for image captioning[C]//Proceedings of the Conference on Empirical Methods in Natural Language Processing. Punta Cana: ACL, 2021: 7514-7528.

[10]Radford A, Kim J W, Hallacy C, et al. Learning transferable visual models from natural language supervision[C]//Proceedings of the International Conference on Machine Learning. Virtual: PMLR, 2021: 8748-8763.

[11]Oppenlaender J. A taxonomy of prompt modifiers for text-to-image generation[EB/OL]. (2022-04-20) [2025-12-22]. 获取路径：https://arxiv.org/abs/2204.13988

[12]Reynolds L, McDonell K. Prompt programming for large language models: Beyond the few-shot paradigm[C]//Proceedings of the CHI Conference on Human Factors in Computing Systems. 2021. DOI:10.1145/3411763.3451760.（条目信息来源：dblp）

[13]Manovich L. AI aesthetics[M]. Moscow: Strelka Press, 2018.

[14]Elkins J, Chun A. Can AI make art?[J]. Arts, 2020, 9(4): 1-14.

[15]Boden M A. Creativity and artificial intelligence[J]. Artificial Intelligence, 1998, 103(1-2): 347-356.

[16]Latour B. Reassembling the social: An introduction to actor-network-theory[M]. Oxford: Oxford University Press, 2005.

