

프롬프트 기반 생성형 게임 에셋 합성 및 평가 파이프라인

이용태, 김준성, 이흥노
광주과학기술원

lyt98313@gm.gist.ac.kr, wnstjd123@gm.gist.ac.kr, heungno@gist.ac.kr

A Prompt-Based Pipeline for Generative Game Asset Synthesis and Evaluation

Yongtae Lee, Junsung Kim, Heung-No Lee
Gwangju Institute of Science and Technology

요 약

자연어 프롬프트 기반 이미지 생성은 게임 에셋 제작의 자동화를 가능하게 하지만, 생성 결과의 품질과 사용 가능성을 정량적으로 검증한 사례는 많지 않다. 본 논문에서는 사용자가 입력한 자연어 지시문을 기반으로 시각적 게임 에셋 이미지를 생성하고, 이를 절차적으로 생성된 2D 게임 플레이 환경에 통합하는 생성 파이프라인을 제안한다. 제안하는 시스템은 지시문으로부터 유도된 테마와 분위기를 반영한 타일 및 스프라이트 이미지를 생성하여, 동일한 규칙 하에서도 매 플레이 세션마다 시각적으로 상이한 환경을 구성할 수 있도록 한다. 또한 생성된 이미지 에셋에 대해 CLIP score 를 활용한 정량적 평가를 수행하고, 기존 게임에 사용된 에셋과의 비교를 통해 생성 이미지의 의미적 일관성과 시각적 완성도를 분석함으로써 제안한 파이프라인이 실제 게임 환경에서 활용 가능한 수준의 이미지를 안정적으로 생성할 수 있음을 보인다.

I. 서 론

게임 개발 과정에서 시각적 에셋 제작은 높은 비용과 시간이 요구되는 작업이다. 생성형 이미지 모델과 자연어 프롬프트 기반 생성 기법은 사용자의 지시문으로부터 특정 테마나 분위기를 반영한 이미지를 생성할 수 있음을 보여주며, 게임 에셋 생성에 적용될 가능성을 제시하고 있다. 그러나 이러한 생성 결과가 실제 게임에 적용 가능한 수준의 품질과 일관성을 갖추었는지에 대해서는 정량적 검증이 충분히 이루어지지 않았다.

게임 에셋은 단일 이미지의 시각적 완성도뿐만 아니라, 프롬프트 의미와의 일관성, 반복 사용 시의 안정성을 동시에 요구한다. 기존 연구들은 주로 생성 가능성이나 표현 다양성에 초점을 두었으며, 게임 에셋 적용을 전제로 생성 이미지의 품질을 정량적으로 분석한 연구는 제한적이다. 본 연구는 이러한 평가를 수행하기 위한 실험 환경으로서, 용도와 해상도가 명확히 정의되는 2D 게임 환경을 대상으로 설정하였다.

관련 연구로서, Zammit et al.의 CrawlLLM 프로젝트는 대규모 언어 모델을 활용한 게임 시각 테마 변수 가능성을 제시하였으나, 생성 이미지에 대한 정량적 품질 평가를 포함하지는 않았다[1].

본 연구에서는 자연어 프롬프트를 입력으로 받아 게임용 시각적 에셋 이미지를 생성하고, 이를 2D 게임 파이프라인에 적용하는 시스템을 제안한다. 생성된 타일 및 스프라이트에 대해 정량적 평가를 수행함으로써, 프롬프트 기반 이미지 생성이 실제 게임 에셋 제작에 활용 가능한 수준의 품질과 안정성을 확보할 수 있음을 실험적으로 분석한다.

II. 본 론

i. 시스템 구현 개요

본 연구에서 제안하는 시스템은 자연어 프롬프트를 입력으로 받아 시각적 게임 에셋 이미지를 생성하고, 이를 자동 생성된 2D 게임 환경에 적용하는 파이프라인으로 구성된다. 사용자가 입력한 자연어 지시문은 생성형 이미지 모델에 직접 전달되지 않으며, 해당 지시문이 요구하는 테마와 분위기를 반영한 생성용 프롬프트로 변환된 후 이미지 생성에 사용된다. 생성된 타일 및 스프라이트 이미지는 동일한 게임 규칙과 구조를 유지한 상태에서 플레이 환경에 매핑되며, 시각적 에셋 생성 과정은 게임 로직과 분리되어 설계되었다. 그림 1 은 전체 파이프라인의 구성과 데이터 흐름을 나타낸다. 본 연구는 시각적 게임 에셋의 생성 및 정량적 평가에 초점을 두므로, 절차적 맵 생성이나 게임 구조 자체에 대한 상세한 논의는 생략한다.

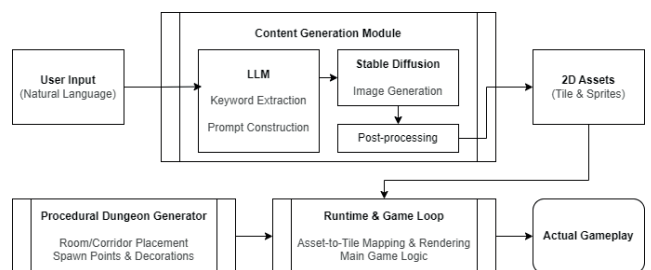


그림 1. 전체 시스템의 구성 및 데이터 흐름을 나타낸 도식

ii. 자연어 프롬프트 기반 시각적 생성

본 시스템은 사용자의 지시문을 에셋 생성 조건으로 사용하며, 대규모 언어 모델을 통해 이미지 생성에 적합한 텍스트 프롬프트를 구성한다. 이 과정에서 언어 모델은 지시문의 의미를 바탕으로 시각적 속성과 스타일을 명시적으로 포함하는 프롬프트를 생성한다. 본 연구에서는 별도의 미세 조정 없이 GPT-4o mini 를 사용하였다.

구성된 프롬프트는 Stable Diffusion 기반 이미지 생성 모델에 입력되며, 생성 결과는 맵을 구성하는 타일 이미지와 독립 오브젝트를 묘사한 스프라이트 이미지로 구분되어 게임 환경에 적용된다. 사용된 생성형 모델들은 추후 개선된 결과 생성을 원할 시 교체가 가능하다.

iii. 후처리 및 정규화

생성된 이미지들은 시각적으로 완성도가 높더라도 해상도 및 표현 방식의 불일치로 인해 게임 에셋으로 즉시 활용하기에는 적합하지 않은 경우가 많다. 이에 본 연구에서는 동일한 렌더링 환경에서 일관되게 사용될 수 있도록 후처리 및 정규화 과정을 적용하였다. 구체적으로 배경 제거, 팔레트 양자화 및 픽셀아트화, 규격 통일을 위한 리사이징을 수행하여 에셋이 게임 환경에 안정적으로 통합될 수 있도록 한다.

iv. 이미지 평가

생성된 게임 에셋 이미지의 의미적 일관성과 반복 생성 시 안정성의 정량적 평가를 수행하기 위해 CLIP score[2]를 사용하여 두 가지 실험을 수행하였다.

첫 번째 실험은 프롬프트-이미지 간 의미 일관성 평가로, 생성 이미지가 입력 프롬프트의 의미를 적절히 반영하는지를 확인한다. 총 15 개의 서로 다른 캐릭터 묘사 프롬프트를 사용하여 이미지를 생성하고, 각 프롬프트-이미지 쌍에 대한 점수를 계산 하였다. 반복적 타일 패턴 이미지는 의미 유사도 평가에 적합하지 않아 제외하였다.

두 번째 실험은 반복 생성 안정성 평가로, 하나의 프롬프트를 선택하여 서로 다른 무작위 초기값을 적용해 이미지를 20 회 반복 생성하고, 점수 분포를 통해 생성 안정성을 평가하였다.

항목	실험 1	실험 2
시행 횟수	15	20
평균	28.3468	29.3626
표준편차	2.3822	1.0863
최대	31.2680	31.2625
최소	23.0937	27.1606

표 1. 프롬프트 기반 생성 이미지의 CLIP score 요약 통계

v. 결과 및 분석

실험 결과를 표 1 에서 확인할 수 있다. 실험 1 은 서로 다른 프롬프트에 대한 의미 일관성 평가 결과이며, 실험 2 는 동일 프롬프트 반복 생성 시 안정성을 평가한 결과이다.

의미 일관성 평가 결과, 생성된 이미지들은 입력 프롬프트의 의미를 전반적으로 반영하는 경향을 보였다. 본 실험의 대상은 저해상도 도트 스타일의 이미지로, 고해상도 자연 이미지에 최적화된 CLIP score 의 특성상 절대적인 점수 수준은 제한적일 수 있다. CLIP 기반 점수가 스타일 일관성을 고려하지 않으며 기존 연구

에서도 유사한 조건에서 점수 상한이 낮게 형성됨이 보고된 바 있음을 고려할 때, 본 연구에서 관측된 점수 분포는 도트 기반 게임 에셋 생성 환경에서 의미적 일관성이 충분히 확보되었음을 시사한다[3].

반복 생성 안정성 평가에서는 생성한 이미지들의 점수 분산이 작게 나타났다. 이는 생성 결과가 단일 우연적 샘플에 의존하지 않고, 반복 실행 시에도 유사한 의미적 특성을 유지함을 의미한다. 이러한 결과는 제안된 생성 파이프라인이 반복 생성 환경에서도 안정적으로 작동함을 보여준다.

III. 결 론

본 연구는 자연어 프롬프트를 입력으로 받아 2D 게임 환경에 활용 가능한 타일 및 스프라이트 에셋을 생성하는 파이프라인을 제안하고, 이를 정량적으로 평가하였다. 제안된 시스템은 프롬프트 변환, 이미지 생성, 후처리, 게임 적용까지의 모든 과정을 하나의 실행 가능한 파이프라인으로 구현하였다.

실험 결과, 생성된 에셋은 프롬프트의 의미를 일관되게 반영하며, 동일 프롬프트 반복 생성 시에도 비교적 안정적인 결과를 보였다. 이는 저해상도 도트 스타일이라는 제약 하에서도 제안한 접근 방식이 실용적인 수준의 생성 품질과 재현성을 확보할 수 있음을 시사한다.

본 연구의 의의는 게임 에셋 생성 맥락에서 프롬프트 기반 이미지 생성 파이프라인을 구현하고, 이를 정량적 지표를 통해 검증한 점에 있다. 기존 연구가 주로 정성적 결과 제시에 집중한 것과 달리, 본 연구는 생성 결과의 일관성과 안정성을 실험적으로 입증함으로써 향후 자동화된 게임 콘텐츠 생성 연구의 기초를 제공한다.

ACKNOWLEDGMENT

본 연구는 과학기술정보통신부 및 정보통신기획평가원의 대학 ICT 연구센터지원사업의 연구결과로 수행되었음 (IITP-2026-RS-2021-II211835) 그리고 이 성과는 정부(과학기술정보통신부)의 재원으로 한국연구재단의 지원을 받아 수행된 연구임. (RS-2025-22932973) We appreciate the high-performance GPU computing support of HPC-AI Open Infrastructure via GIST S C E N T .

참 고 문 헌

- [1] Zammit, M., Liapis, A., and Yannakakis, G. N., "CrawLLM: Theming Games with Large Language Models", Proceedings of the IEEE Conference on Games (CoG), 2023.
- [2] Hessel, J., Holtzman, A., Forbes, M., Le Bras, R., and Choi, Y., "CLIPScore: A Reference-free Evaluation Metric for Image Captioning", Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP), 2021.
- [3] Li, H., Fan, Z., Wen, Z., Zhu, Z., and Li, Y., "AIComposer: Any Style and Content Image Composition via Feature Integration", arXiv preprint, 2025.