

ComicGen: LLM 및 Stable Diffusion 기반 자동 만화 생성 시스템*

김도영⁰¹, 김경중²
광주과학기술원 AI 대학원¹
광주과학기술원 융합기술학제학부²
doyoungk@gm.gist.ac.kr, kjkim@gist.ac.kr

ComicGen: LLM, Stable Diffusion Based Automated Comic Generation System

Doyoung Kim⁰¹, Kyung-Joong Kim²
¹Artificial Intelligence Graduate School, Gwangju Institute of Science and Technology
²Institute of Integrated Technology, Gwangju Institute of Science and Technology

요약

본 연구는 거대 언어 모델(LLM)과 Stable Diffusion을 활용하여 소설 텍스트를 자동으로 만화 형태로 변환하는 시스템을 구현하였다. 이 시스템은 장면 생성(Scene Generation), 프롬프트 생성(Prompt Generation), 이미지 생성(Image Generation)의 세 단계로 구성되며, 거대 언어 모델을 통한 장면 분리와 프롬프트 생성, Stable Diffusion 기반 이미지 생성을 통해 소설을 친숙한 만화 형식으로 시각화한다. 제한된 데이터셋을 사용해 특정 화풍의 이미지를 안정적으로 생성할 수 있음을 확인했으며, 추후 연구에서는 대사와 말풍선 자동 삽입 기능을 추가하여 만화화 시스템의 완성도를 높일 계획이다.

1. 서론

이 연구는 거대 언어 모델(LLM)과 Stable Diffusion을 활용해 텍스트 형태의 소설을 만화로 자동 변환하는 시스템의 구현에 대해서 다룬다.

스마트폰과 태블릿이 보급되며 전자책(eBook)은 그 어느 때보다 쉽게 접근할 수 있게 되었다. 소비자들은 한 기기에 여러 권의 책을 담아 다닐 수 있고, 방대한 도서들에 쉽게 접근할 수 있다. 이처럼 많은 책이 텍스트의 형태로 제공되고 있다. 따라서, 텍스트 형태의 도서를 다양하게 활용할 수 있는 방법을 고안할 필요가 있다. 이러한 전자책 독자들은 책을 직접 읽는 대신, TTS(Text-to-Speech)를 적용해 오디오북으로 듣는 경우가 많다. 그렇다면, 단순한 음성에서 나아가 소설 형태의 텍스트를 더욱 접근성이 높은 만화의 형태로 제공해줄 수 있겠다는 아이디어에서 착안해 소설 텍스트 기반의 자동화된 만화 생성 시스템을 구현했다. 우선 소설 형태의 텍스트를 만화에 적합하

게 변환하기 위해 대형 언어 모델(Large Language Model)로 전체적인 구조를 변경했다. 이후 변경된 텍스트로 적절한 프롬프트를 생성하여 최종적으로 Stable Diffusion 모델을 사용해 적절한 만화 이미지를 생성하였다.

2. 본론

자동화된 만화 생성 과정은 그림 1과 같이, 장면 생성(Scene Generation), 프롬프트 생성(Prompt Generation)과 이미지 생성(Image Generation)의 총 3가지 단계로 이루어진다. 우선, Scene Generation 단계에서는 연속된 장면으로 구성된 소설을 개별 장면으로 구분된 대본의 형태로 변환했다. 이후, Prompt Generation 단계에서는 이전 단계의 대본을 바탕으로 Stable Diffusion 모델이 적합한 이미지를 생성할 수 있도록 프롬프트(Prompt)를 생성했다. 마지막 Image Generation 단계에서는 최종적인 프롬프트를 바탕으로 Stable Diffusion을 사용해 적절한 만화 이미지를 생성한다. 모든 과정은 Streamlit을 사용한 웹 애플리케이션 상에서 자동화했다. 해당 웹 애플리케이션은 OpenAI 및 Stable Diffusion 이미지 생성을 위한 서버와 API 통신하여 결과를 저장한다.

대표적인 유사한 연구로, SEED-Story는 대형 언어 모델

* 본 연구는 문화체육관광부 및 한국콘텐츠진흥원의 2024년도 문화기술 연구개발사업으로 수행되었음
(과제명 : 청각장애인과 청인의 의사소통을 위한 인공지능 기반 수어 통·번역 서비스 기술 개발, 과제번호 : RS-2023-00303791, 기여율: 33%)

Alice was beginning to get very tired of sitting by her sister on the bank, and of having nothing to do: once or twice she had peeped into the book her sister was reading, but it had no pictures or conversations in it, "and what is the use of a book," thought Alice "without pictures or conversations?"

```

0 : {
  "scene_number": 1
  "background": "
  "Alice and her sister are sitting on a bank, with Alice growing increasingly tired and bored."
  "characters": [
    0 : "Alice"
    1 : "Alice's sister"
  ]
  "dialogue": [
    0 : "
    "Alice: I'm starting to get very tired of just sitting here and having nothing to do."
    1 : "
    "Alice's sister: (silent, reading a book)"
    2 : "
    "Alice: (peeping into the book her sister is reading) This book has no pictures or conversations. What is the use of a book without pictures or conversations?"
  ]
}
    
```

(a) 소설 텍스트 (b) 생성된 대본 텍스트
 그림 2 소설 텍스트 입력 및 생성된 대본 텍스트

을 활용한 멀티모달 기반 장편 이야기 생성을 목표로 하며, 텍스트, 이미지, 오디오를 통합해 스토리텔링의 표현력을 극대화하는 기술적 성과를 보였다[1]. 반면, 본 연구는 텍스트를 기반으로 한 소설을 만화 형태로의 변환에 초점을 두고 있다. SEED-Story와 비교하여 본 연구는 장면 분할, 텍스트-이미지 프롬프트 생성, 그리고 Stable Diffusion 모델의 파인튜닝을 통해 만화라는 특정 매체에 최적화된 결과물을 생성하는 데 주력한다는 점에서 차별성을 두었다.

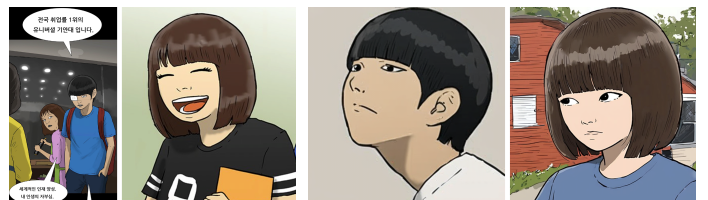
number), 배경(background), 캐릭터(character), 대사(dialogue)의 네 가지 요소로 이루어진 JSON(JavaScript Object Notation) 형식의 텍스트가 된다. 그림 2는 고전 소설 <이상한 나라의 엘리스(Alice's Adventures in Wonderland)>의 도입부 텍스트를 변환한 예시이다.

2.2 프롬프트 생성(Prompt Generation)

Stable Diffusion 모델과 같은 텍스트-이미지 생성에서 이미지의 적합성과 품질을 개선하기 위해 프롬프트를 최적화하는 여러 연구가 이루어지고 있다[2]. 또한, 대형 언어 모델은 프롬프트를 자동으로 생성하고 최적화할 수 있어 출력물의 품질과 적합성을 크게 향상할 수 있다[3].

이 단계에서는 GPT-4 모델을 사용해서 적절한 프롬프트가 출력되도록 프롬프트 엔지니어링을 거쳤다. 앞선 장면 생성 단계로부터 JSON 형태의 장면을 입력받아 개별 단어의 리스트로 이루어진 Stable Diffusion 프롬프트를 생성하도록 하였다. 해당 프롬프트는 장면의 배경, 등장인물, 행동에 대한 자세한 묘사를 개별 영문 단어로써 묘사하게 된다.

2.3 이미지 생성(Image Generation)



(a) 원본 웹툰 이미지 (b) 모델 생성 이미지
 그림 3 원본 이미지와 파인튜닝을 거친 모델 생성 이미지

프롬프트를 사용한 이미지 생성 이전에, 기반 소설에 적

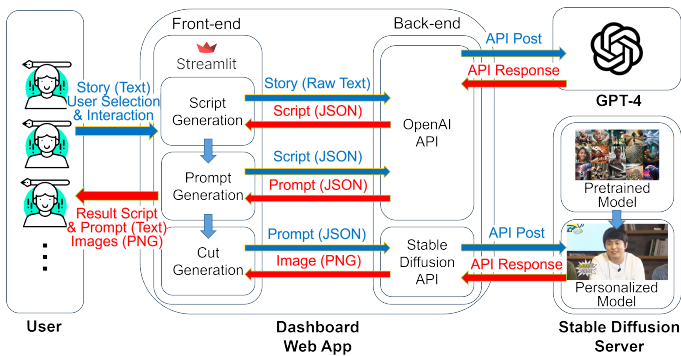


그림 1 ComicGen의 전체 개요도

2.1 장면 생성(Scene Generation)

이 단계에서는 기존의 소설 텍스트를 개별 장면으로 나누어진 JSON 텍스트로 변환한다. 기존의 소설 텍스트를 개별적인 장면으로 분리하기 가장 좋은 포맷이 연극 또는 영화의 장면(Scene)으로 나누어진 대본과 같은 형태라고 판단하였다. 소설의 내용을 누락하지 않고, 충실히 구현하도록 OpenAI의 GPT-4 모델을 사용해서 적절한 결과가 나올 수 있게 프롬프트 엔지니어링을 거쳤다. 이 단계에서 소설 텍스트를 변환 시 각 장면이 장면 번호(scene

합한, 특정한 화풍을 적용하기 위해 LoRA 기법을 사용해 Stable Diffusion 모델을 파인튜닝 하였다. LoRA 기법은 저랭크 적응(low-rank adaptation)에 중점을 두고 있어, 소량의 이미지 데이터셋을 활용하는 이미지 생성 상황에서 Stable Diffusion 모델에 적용될 수 있다. 사전 훈련된 모델의 가중치를 고정하고 저랭크 구성 요소만을 적응시키는 방식으로, 제한된 데이터로도 거대 모델을 효율적으로 미세 조정 가능하다. 이를 통해 높은 성능을 유지하면서도 계산 비용을 절감하는 데 도움을 줄 수 있다[4]. 기존의 화풍보다 독특한 화풍 또한 학습될 수 있는지 확인하기 위해, 개성 있는 화풍의 만화 이미지를 학습에 사용하였다. 기반 Stable Diffusion 모델에 특정 만화 작가의 108장의 이미지를 데이터 증폭해 RTX 4090 그래픽 카드 서버에서 2시간 동안 추가로 학습을 진행했다. 그 결과 개성 있는 화풍과 흡사한 이미지가 생성되는 것을 확인하였다(그림 3). 최종적으로, 간단한 임의의 텍스트를 전체 과정을 거쳐 생성한 결과의 예시는 그림4와 같다.

3. 결론

이 연구에서 소설 형태의 텍스트를 만화로 자동 변환하는 시스템을 제시했다. 시스템은 Scene Generation, Prompt Generation and Image Generation의 총 3가지 단계로 이루어지며, 프롬프트 엔지니어링을 거친 LLM과 Stable Diffusion 기반의 이미지 생성을 거쳤다.

현재 이 시스템은 단순히 소설의 장면에 대한 이미지를 생성하는 것에 그치는 한계점이 있다. 각 장면에 대한 대사는 제공되지만, 생성된 이미지에서는 사용자가 직접 말풍선과 대사를 수동으로 추가해야 한다. 추후 이미지 프로세싱을 거쳐 각 등장인물의 위치를 인식하고, 적절한 위치에 말풍선과 대사를 추가할 수 있도록 개선할 계획이다. 이 연구를 통해 더 많은 소설과 전자책들이 높은 접근성을 가지도록 도울 수 있기를 바란다.

참고 문헌

- [1] YANG, Shuai, et al. Seed-story: Multimodal long story generation with large language model. arXiv preprint arXiv:2407.08683, 2024.
- [2] HAO, Yaru, et al. Optimizing prompts for text-to-image generation. Advances in Neural Information Processing Systems, 2024.
- [3] ZHOU, Yongchao, et al. Large language models are human-level prompt engineers. arXiv preprint arXiv:2211.01910, 2022.
- [4] HU, Edward J., et al. Lora: Low-rank adaptation of large language models. arXiv preprint arXiv:2106.09685, 2021.



그림 4 웹툰 형식의 자동화된 만화 생성 예시