

생성 AI 모델을 위한 텍스트 프롬프트 추정 기반 한국화 재생성에 대한 연구

문성원¹, 이지원¹, 남도원¹

¹한국전자통신연구원 콘텐츠연구본부

moonstarry@etri.re.kr, ez1005@etri.re.kr, dwnam@etri.re.kr

A Study on the Reproduction of Korean Painting through Text Prompt Estimation for Generative AI Models

Sung-Won Moon¹, Jiwon Lee¹, Dowon Nam¹

¹Content Research Division,

Electronics and Telecommunications Research Institute(ETRI)

요 약

스테이블 디퓨전(Stable diffusion)과 같은 텍스트 프롬프트 입력 기반 이미지 생성 AI 기술의 발전으로 원하는 형태의 고품질 이미지를 누구나 손쉽게 생성가능할 것으로 기대하였으나 대부분의 경우 원하는 이미지를 얻기 위해서는 텍스트 프롬프트를 정교하게 조정해야 하며 많은 실패를 겪어야만 한다. 이러한 한계를 극복하기 위해 클립 인터로게이터(CLIP Interrogator)와 같은 유사 이미지를 재생성하기 위한 텍스트 프롬프트 추정 기술이 개발되었으며 몇몇 이미지에 대해 뛰어난 결과를 보였다. 본 논문에서는 이러한 텍스트 프롬프트 추정 기술의 활용이 한국화를 재생성할 수 있는지 실험을 통해 가능성을 확인하고 향후 한국화 재생성을 위한 연구 방향을 제안하고자 한다.

1. 서론

최근 미드저니(Midjourney), 달리 2(DALL·E 2), 스테이블 디퓨전(Stable diffusion)[1]과 같은 텍스트 기반 이미지 생성 AI 모델의 급격한 발전으로 인해 사용자들이 고품질의 이미지를 AI를 통해 자유롭게 생성하고 공유하고 있다. 하지만 초거대 생성 모델들이 출력 가능한 다양한 이미지를 원하는 스타일과 형태로 생성하기 위해서는 사용자 스스로 다양한 텍스트 프롬프트를 입력으로 적용하고 실패를 거듭해 가며 경험적으로 원하는 형태를 잘 생성할 수 있는 프롬프트를 찾아내어야 한다. 이러한 과정의 어려움으로 인해 초거대 AI가 원하는 출력을 잘 생성할 수 있도록 프롬프트를 구성하는 작업을 프롬프트 엔지니어링으로 명명할 정도로 적절한 프롬프트를 구성하는 것은 생성 AI 활용에 매우 중요한 영역을 차지하고 있다.

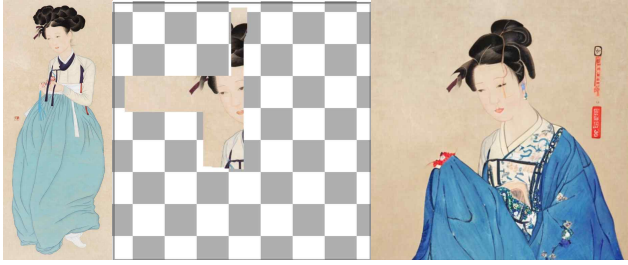
개인의 경험과 역량에 의존적인 프롬프트 구성을 대체하고 누구나 생성 AI 모델을 통해 원하는 형태의 이미지를 생성할 수 있도록 지원하기 위해 입력 이미지를 생성하기 위한 최적의 프롬프트를 자동으

로 추정하는 클립 인터로게이터(CLIP Interrogator)[2]와 같은 기술이 개발되었다. 클립 인터로게이터는 대표적인 생성 모델 활용 인터페이스 중 하나인 스테이블 디퓨전 웹 UI에도 적용되어 많은 사람이 활용성을 검증하였으며 실제로 다양한 이미지 내의 객체, 스타일, 작가 등을 추정하는 것에 어느 정도 성과를 보였다. 하지만 이러한 검증은 미국 카툰, 빈센트 반 고흐의 그림과 같이 매우 유명하고 특징적인 형태의 이미지에 한정되어, 상대적으로 초거대 AI 학습에 적게 활용되었을 것으로 예상되는 한국화 재생성을 위한 텍스트 프롬프트 추정에 기존 기술을 활용할 수 있을지는 미지수이다.

본 논문에서는 실험을 통해 스테이블 디퓨전이 한국화와 유사한 이미지를 생성 가능한지 확인하고, 기존 텍스트 프롬프트 추정 기술을 통해 추정한 프롬프트 입력이 한국화 재생성에 활용되었을 때 어떠한 결과물이 출력되는지 확인하고자 한다.

2. 본론

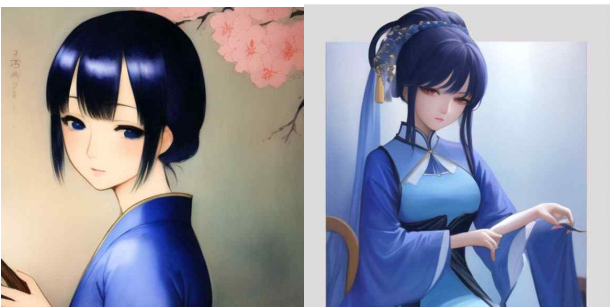
한국화 재생성을 위한 텍스트 프롬프트 추정의 효과를 확인하기 위해서는 먼저 생성 AI 모델의 한



(그림 1) 좌로부터 이미지 원본[3], 입력을 위해 잘라낸 이미지 일부분, 스테이블 디퓨전을 활용한 출력 결과

국화 출력 가능 여부를 확인하여야 한다. 이를 위하여 대표적인 한국화 중 하나인 혜원 신윤복의 미인도 모사작[3]을 활용하였다. 해당 이미지의 일부분을 잘라낸 후 이를 입력으로 활용하여 스테이블 디퓨전 인피티니[4]로 출력한 결과는 그림 1과 같다. 그림 1을 통해 원본과 동일한 모습은 아니나 생성 모델이 한국화 혹은 동양화의 화풍이나 채색을 재현하는 것이 가능함을 알 수 있다.

클립 인터로게이터에 미인도 모사작을 입력하여 추정된 텍스트 프롬프트는 [A painting of a woman in a blue dress by Takehisa Yumeji, cgsociety, shin hanga, portrait of female korean idol, korean hanbok, korean art nouveau anime] 이다. 그림 내의 객체가 푸른 옷을 입은 여성임을 추정한 것과 해당 그림에 대한 추정 중 한복, 한국화와 같이 올바른 추정도 있으나 해당 그림이 혜원 신윤복의 스타일임을 추정하는 것에는 실패 하였다. 위의 텍스트 프롬프트를 입력한 결과는 그림 2와 같다. 그림 1의 결과와는 상이하며 전통적인 한국화보다는 만화나 게임의 캐릭터와 유사한 형상의 결과물이 출력 되었다. 해당 프롬프트의 작가에 해당하는 부분을 클립 인터로게이터의 작가 목록 중 신윤복에 해당하는 Shin Yun-bok으로 교체한 결과는 그림 3과 같다. 그림 2와 화풍은 다소 다르나 여전히 신윤복의 스타일을 모사했다고 보기에는 어려운 결과이다.



(그림 2) 클립 인터로게이터를 통해 추정한 텍스트 프롬프트 입력에 대한 출력 결과 예시



(그림 3) 텍스트 프롬프트의 작가에 해당하는 입력을 신윤복으로 변경한 프롬프트 입력에 대한 출력 결과 예시

위 실험 결과를 통해 얻을 수 있는 결론은 다음과 같다. 기존 텍스트 프롬프트 추정 기술은 학습 데이터 구성의 한계로 한국화의 작가 추정 성능이 떨어지며, 이를 보정 하여 입력하는 경우에도 한국화 스타일을 출력하는 것은 어렵다. 따라서 별도의 생성 모델 학습 없이 한국화 형태의 데이터를 출력하기 위해서는 이를 가능케 하는 추가적인 입력 혹은 새로운 텍스트 프롬프트 추정 기법이 필요하다.

3. 결론

본 논문에서는 기존 기술의 한국화 재생성을 위한 텍스트 프롬프트 추정 활용 가능성을 확인하였다. 상대적으로 전체 데이터 대비 수가 적은 한국화의 특성상 한국화를 기존 기술의 조합으로 재현하는 것은 어려웠으며, 향후 이를 가능하게 하는 텍스트 프롬프트 추정 및 구성 방법을 연구하고자 한다.

사사

본 연구는 문화체육관광부 및 한국콘텐츠진흥원의 2023년도 문화체육관광연구개발사업으로 수행되었음(과제명: 사용자들의 다양한 관심사 맞춤형 문화유산 스토리텔링 AI 기술 개발, 과제번호: RS-2023-00220195, 기여율: 100%)

참고문헌

- [1] R. Rombach, A. Blattmann, D. Lorenz, P. Esser, and B. Ommer, "High-Resolution Image Synthesis with Latent Diffusion Models." *arXiv*, 2021.
- [2] pharmapsychotic, "pharmapsychotic/clip-interrogator," GitHub, 2023.
- [4] 혜원 신윤복 '미인도' 모사작, 연리제 (https://m.blog.naver.com/seo_000/), 2019
- [5] lkqw007, "LKWQ007/stablediffusion-infinity: Outpainting with stable diffusion on an infinite canvas," GitHub, 2023.