

확산모델의 미세조정을 통한 웹툰 생성연구

(A Study on the Generation of Webtoons through Fine-Tuning of Diffusion Models)

유경호*, 김형주**, 김정인***, 전찬준*, 김판구*

(Kyungho Yu, Hyungju Kim, Jeongin Kim, Chanjun Chun, Pankoo Kim)

요약

본 연구에서는 웹툰 작가의 웹툰 제작 과정을 보조하기 위해 사전학습된 Text-to-Image 모델을 미세조정하여 텍스트에서 웹툰을 생성하는 방법을 제안한다. 제안하는 방법은 웹툰 화풍으로 변환된 웹툰 데이터셋을 사용하여 사전학습된 Stable Diffusion 모델에 LoRA 기법을 활용하여 미세조정한다. 실험 결과 3만 스텝의 학습으로 약 4시간 반이 소요되어 빠르게 학습하는 것을 확인하였으며, 생성된 이미지에서는 입력한 텍스트에 표현된 형상이나 배경이 반영되어 웹툰 이미지가 생성되는 것을 확인하였다. 또한, Inception score를 통해 정량적인 평가를 수행하였을 때, DCGAN 기반의 Text-to-Image 모델보다 더 높은 성능을 나타냈다. 본 연구에서 제안된 웹툰 생성을 위한 Text-to-Image 모델을 웹툰 작가가 사용한다면, 웹툰 제작에 시간을 단축시킬 수 있을 것으로 기대된다.

■ 중심어 : Text-to-Image ; 확산모델 ; 웹툰 생성 ; 딥러닝

Abstract

This study proposes a method to assist webtoon artists in the process of webtoon creation by utilizing a pretrained Text-to-Image model to generate webtoon images from text. The proposed approach involves fine-tuning a pretrained Stable Diffusion model using a webtoon dataset transformed into the desired webtoon style. The fine-tuning process, using LoRA technique, completes in a quick training time of approximately 4.5 hours with 30,000 steps. The generated images exhibit the representation of shapes and backgrounds based on the input text, resulting in the creation of webtoon-like images. Furthermore, the quantitative evaluation using the Inception score shows that the proposed method outperforms DCGAN-based Text-to-Image models. If webtoon artists adopt the proposed Text-to-Image model for webtoon creation, it is expected to significantly reduce the time required for the creative process.

■ keywords : Text-to-Image ; Diffusion model ; Webtoon Generation ; Deep Learning

I. 서론

웹툰은 웹과 카툰의 합성어로 온라인에서 연재되는 만화를 일컫는다. 웹툰과 기존만화의 차이점은 세로 스크롤 방식으로 컷을 배치하여 모바일 디바이스에 읽기 쉬운 것이 특징이다. 웹툰

산업은 한국 포털 기업의 주도로 발전하고 있으며 국내뿐만 아니라 해외에 진출하여 높은 수익을 거두고 있는 부가가치가 높은 산업 중 하나이다. 웹툰의 제작은 스토리 구상 및 트리트먼트 제작, 콘티 제작, 스케치, 펜터치, 채색, 배경 그리기의 단계로 나뉜다. 이러한 여러 단계로 이루어진 웹툰의 제작은 각 단계마다 작가가 직접 그

* 정회원, 조선대학교 컴퓨터공학과

** 정회원, 조선대학교 AI융합연구원

*** 정회원, 조선대학교 치의학전문대학원 미생물학 및 면역학 교실

이 논문은 2021학년도 조선대학교 학술연구비의 지원과 연구개발특구진흥재단의 '기술사업화 협업 플랫폼' 사업으로 수행되었습니다.(과제명: 인공지능 산업 육성 및 기술사업화를 위한 지능형 디지털 콘텐츠 제작 기술 개발 및 플랫폼 구축 사업, 과제고유번호: 1711177250)

접수일자 : 2023년 08월 09일

게재확정일 : 2023년 08월 29일

교신저자 : 김판구 e-mail : pkkim@chosun.ac.kr

려야하기 때문에 많은 노동력이 필요하다[1].

텍스트에서 이미지 생성기술은 크게 적대적 생성 신경망(Generative Adversarial Networks)과 확산모델(Diffusion model)을 기반으로 하며 최근 확산모델은 적대적 생성모델의 성능을 크게 웃돌고 있다[2,3]. 대표적인 확산모델 기반의 이미지 생성 모델은 OpenAI의 Dalle-2(혹은 unCLIP)[6], Stable AI의 stable diffusion이 있으며[7], 이와같은 생성모델은 실제 사람이 그린것과 유사한 이미지를 생성할 수 있다. 또한, 이미지 생성 뿐만아니라 생성된 이미지의 일부를 수정하는 'inpaint' 와 텍스트를 가이드로하여 이미지를 변환시키는 'image-to-image'까지 가능하다. 학습한 확산모델을 새로운 이미지를 사용하여 미세조정 할 때 적은수의 이미지를 사용하여 학습한 이미지의 화풍에 유사한 이미지를 생성할 수 있는 Dreambooth[8]나 LORA(Low-Rank adaptation of Large Language Models)[9]와 같은 기술이 연구되었다.

인공지능 기술의 발전으로 다양한 분야에서 생성모델을 사용하고 있다[4,5]. 최근에는 생성모델을 사용하여 웹툰의 저작단계에 도움을 줄 수 있는 기술이 보급되고 있다. 웹툰을 서비스하고 있는 네이버에서는 인공지능기술을 웹툰 저작에 적용하기 시작하여 "웹툰 AI페인터"[10]를 서비스하고 있으며 스케치나 펜터치 작업 단계에서 작가가 기본적인 형상을 그리면 자동으로 그려주는 연구가 수행되었다[11,12]. 또한, 웹툰에 대한 설명을 나타내는 트리트먼트와 웹툰을 Text-to-Image 모델에 학습하여 트리트먼트를 입력하였을 때 웹툰을 생성하는 연구가 수행되었다[13]. 그리고 인공지능 기술을 웹툰에 적용하였을 때 상대적 이점, 적합성, 복잡성, 시도가능성, 관찰가능성과 같은 요인을 기반으로 기술에 대한 혁신 채택과정을 분석하는 혁신 확산 이론으로 인공지능기술의 웹툰 적용에 대한 분석을 하였을 때 인공지능 웹툰의 긍정적 영향에 대한 전망을 보였다[1].

본 연구에서는 사전학습된 Text-to-Image 모델을 미세조정하여 텍스트에서 웹툰을 생성할 수 있는 방법을 제안한다. 제안하는 방법은 공개 데이터 셋을 웹툰 화풍으로 변환한 웹툰 데이터 셋을 사전학습된 Text-to-Image 모델에 미세조정하여 텍스트를 입력하였을 때 웹툰을 생성할 수 있다. 미세조정에 사용한 기법은 빠른시간안에 최소한의 데이터 셋을 사용하여 학습할 수 있는 LoRA를 사용한다. 본 연구에서 제안하는 방법을 통해 학습된 Text-to-Image 모델을 웹툰 저작에 사용한다면 사용자가 그리고자하는 웹툰에 대한 텍스트에서 웹툰을 생성할 수 있기 때문에 웹툰 저작에 도움을 줄 수 있을것으로 예상된다.

본 연구의 구성은 2절에서는 본 연구와 관련높은 확산모델과 딥러닝 기반 Text-to-Image에 대해 설명하고 3장에서는 본 연구에서 제안하는 사전학습된 Text-to-Image 모델에 LoRA 기법을 사용하여 미세조정하는 방법을 설명한다. 4장에서는 본연구에서 제안하는 웹툰 생성을 위한 Text-to-Image 모델을 사용한 실험결과이며, 5장은 본 연구의 결론과 향후 연구에 대해 논한다.

II. 관련 연구

1. 확산모델(Diffusion model)

확산 모델(Diffusion Model)은 이미지 생성에 사용되는 확률적인 모델링 방법이다. 확산 모델은 이미지의 확률 분포를 모델링하고, 이를 기반으로 새로운 이미지를 생성할 수 있다. 확산 모델은 이미지 생성 시 다양한 결과물을 얻을 수 있는 장점을 가지고 있으며, 이미지 생성 과정에서 학습된 정보를 재 사용하여 효율성을 높일 수 있다[14].

확산 모델의 학습은 두 가지 구성 요소로 이루어진 확산 과정(diffusion process)을 통해 이루어진다. 먼저, 정방향 확산 과정으로, 이미지의 픽셀 값

에 스텝마다 노이즈를 조금씩 추가하여 이미지를 확산시키는 과정이다. 두 번째는 역방향 확산 과정으로, 이미지를 역으로 확산시켜 노이즈에서 원본 이미지를 복원하는 과정이다. 이 두 과정을 반복하고, 이미지 생성은 역방향 확산 과정을 사용하여 노이즈에서 새로운 이미지를 생성할 수 있다.

확산 모델은 대규모 데이터셋을 사용하여 학습되며, 학습 과정에서는 역방향 확산 과정을 통해 학습된 이미지와 실제 이미지의 차이를 최소화하는 방향으로 모델이 업데이트된다. 학습된 모델은 이후 정방향 확산 과정을 통해 이미지 생성 과정에서 사용되며, 노이즈를 추가하여 다양한 이미지를 생성할 수 있게 된다[14,15].

확산 모델은 이미지 생성 분야에서 정량적, 정성적 평가에서 적대적 생성신경망보다 높은 성능을 나타내었으며 다양한 응용 분야에 활용되고 있다. 이미지 생성에서 확률적인 모델링 방법을 사용하기 때문에, 적대적 생성 신경망보다 다양한 이미지를 생성할 수 있으며 이미지 생성이 가능하며, LoRA와 같은 미세조정 기법을 통해 작은 데이터 셋에 나타나 있는 이미지에 나타나 있는 표현을 학습하여 이미지를 생성할 수 있다.

2. 딥러닝 기반 Text-to-Image

Text-to-Image 기술은 텍스트를 Text-to-Image 모델에 입력하였을 때 이미지를 생성하는 것을 말한다. 사용자가 원하는 이미지에 대한 설명에서 이미지를 생성할 수 있으므로 온라인 광고, 웹툰 등 엔터테인먼트 분야뿐만 아니라 다양한 분야에서 활용이 가능한 기술이다[13].

Text-to-Image 기술은 텍스트와 이미지 데이터 셋에서 특징을 추출하고 연관성을 분석하는 부분과 이미지를 생성하는 부분으로 나눌 수 있다. 텍스트와 이미지 사이의 특징과 연관성을 분석하는 부분은 BERT와 ViT(Vision Transformer)를 인코더로 사용하여 특징을 추출하고 대조 학습(contrastive learning)을 통해 텍스트와 이

미지 사이의 관계를 학습한다[16]. 그리고 이미지 생성 부분은 적대적 생성신경망과 확산 모델과 같은 이미지 생성모형을 기반으로 한다. 따라서 Text-to-Image 기술의 발전은 텍스트, 이미지의 특징을 추출하는 모델과 그 관계를 학습하는 모델, 이미지를 생성하는 생성모형의 발전에 따라 성능이 향상된다.

Text-to-Image 기술은 DCGAN 모델을 기반으로 하여 시작하였다. 그 후 어텐션 메커니즘이 도입된 AttnGAN[17], StackGAN[18], DM-GAN[19]이 연구되었으며 최근에는 확산모형을 생성 모델로 하는 Dalle-2, Stable diffusion 와 같은 모델이 주로 사용되고 있다. 이 두가지 모델은 유사한 구조를 나타내고 있으며 이 중 Stable diffusion은 구성요소는 CLIP 기반의 텍스트 인코더, UNet과 scheduler로 구성된 정보 생성자, 이미지를 생성하는 디코더로 구성된다. Stable diffusion은 픽셀 공간(pixel space)에서 이미지를 생성하는 것이 아니라 잠재 공간(latent space)에서 이미지를 생성하기 때문에 빠른 이미지를 생성할 수 있다.

III. 확산모델 기반 텍스트에서 웹툰 생성

본 절에서는 텍스트에서 웹툰을 생성하기 위해 사전학습된 Text-to-Image 모델인 Stable diffusion 1.5 버전 모델을 자체 구축한 웹툰 데이터 셋을 사용하여 LoRA 기법을 사용하여 미세 조정하는 방법에 대해 설명한다.

1. 확산모델 기반의 Text-to-Image 모델

Stable Diffusion 모델은 확산 모델을 기반으로 텍스트에서 이미지를 생성하는 모델이다. Stable Diffusion은 의 구성요소는 CLIP의 텍스트 인코더, UNet, VAE(Variational Autoencoder)로 구

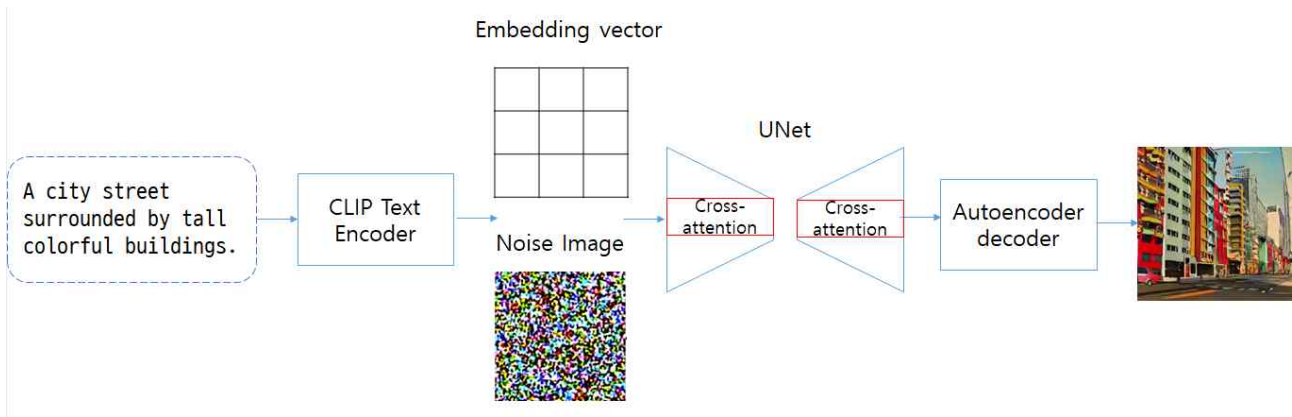


그림 1. Stable Diffusion을 이용한 텍스트에서 이미지 생성 과정

성되어 있다. CLIP(Contrastive Language-Image Pretraining)는 텍스트와 이미지의 특징을 추출하고 대조학습(Contrastive Learning)을 통해서 다른 두 데이터의 관계를 학습한 모델이다. Stable Diffusion에서는 CLIP의 텍스트 인코더를 사용하여 텍스트의 특징을 추출하는데, 생성하고자 하는 이미지에 대한 설명을 입력하면 CLIP의 텍스트 인코더를 통해 토큰(token)형태로 변환하여 이 토큰을 Unet에 전달한다. Unet에서는 토큰을 기반으로 하여 생성된 노이즈를 디노이징 과정(Denoising Process)을 한다. 디노이징은 이미지를 생성하기 위해 이미지에서 노이즈를 제거하며 이미지를 생성한다. 마지막으로 이미지 디코더는 VAE를 사용하여 이미지의 잠재 변수에서 이미지를 픽셀로 변환하여 최종적으로 이미지를 생성한다. Stable Diffusion은 대용량의 텍스트-이미지로 구성된 LAION 데이터 셋을 사전에 학습한 모델로, 입력된 텍스트에 대해 다양한 형태의 이미지를 생성할 수 있다. 본 연구에서는 대용량의 데이터 셋을 학습하여 다양한 이미지를 생성하기 위해 사전학습된 Stable Diffusion을 사용한다.

2. Text-to-Image 모델의 미세 조정

확산모델을 사용하여 이미지를 생성할 때 사용자가 원하는 방향의 이미지를 생성하기 위해 미

세조정을 하는 것은 시간이 많이 소요된다. 그렇기 때문에 시간을 단축시키는 방법으로는 DreamBooth이나 Textual Inversion 그리고 LoRA를 사용하는 방법이 있다. 따라서, 본 연구에서는 사전학습된 확산모델인 Stable diffusion 모델을 LoRA방식으로 웹툰 데이터 셋을 학습하여 텍스트에서 웹툰을 생성하도록 한다. LoRA(Low-Rank Adaptation of Large Language Models)는 대규모의 언어모델의 저차원 적응 기법이다. 이 기법은 많은 수의 파라미터를 가지고 있는 대규모의 언어모델을 빠르고 효과적으로 미세조정하기 위해 고안된 기법으로써, 전체계층을 모두 재학습하지 않고 미세조정하기 위한 데이터를 사용하여 일부 매개변수만을 업데이트한다. 다시말하자면, 고정된 가중치를 갖는 사전학습 모델에 학습이 가능한 랭크 분해(rank decomposition) 행렬을 삽입하여 계층에 학습이 가능한 파라미터를 삽입하여 미세조정한다. 이렇게 LoRA를 통해 미세조정된 모델은 LoRA는 타겟 레이블이나 타겟 데이터에 맞게 모델을 미세조정 할 수 있어 사용자가 원하는 값을 출력할 수 있다. 이러한 장점으로 LoRA는 언어모델 뿐만아니라 이미지 분류, Text-to-Image와 같은 모델에 사용된다[20]. Stable Diffusion에서 LoRA를 적용하는 방법은 이미지의 표현과 사용자가 입력한 텍스트에 대한 크로스 어텐션 레이어(cross-attention layers)에 적용할 수 있다(그림 1).

본 연구에서는 공개데이터 셋은 MSCOCO 데이터 셋[21]을 CartoonGAN[22]을 사용하여 웹툰 화풍을 변환하여 구축한 웹툰 데이터 셋을 LoRA기법을 사용하여 미세조정한다. 웹툰 데이터 셋을 타겟 데이터 셋으로 사용하여 LoRA기법으로 미세조정된 Stable diffusion은 텍스트를 입력하였을 때 사실적인 이미지를 생성하는 것이 아닌, 웹툰 화풍의 이미지를 생성할 수 있다. 웹툰 화풍으로 미세조정된 Stable diffusion 모델을 사용한다면 기존의 생성된 이미지를 별도의 스타일 변환을 하지 않아도 웹툰 화풍의 이미지를 생성할 수 있다.

IV. 실험결과

본 절에서는 웹툰 데이터 셋을 타겟으로 하여 Stable Diffusion을 미세조정된 실험결과를 설명한다. 웹툰 데이터 셋은 공개 데이터 셋인 MSCOCO 데이터 셋을 CartoonGAN을 사용하여 웹툰 화풍으로 변환시킨 데이터 셋으로 웹툰 화풍으로 변환된 이미지와 이 이미지에 대한 영어문장 5가지로 구성되어있는 데이터 셋이다. 구축한 웹툰 데이터 셋은 82,783장의 이미지로 구성된 학습 데이터, 40,504장의 이미지로 구성된 평가 데이터이며 1장당 5개의 설명문으로 구성되어있다.

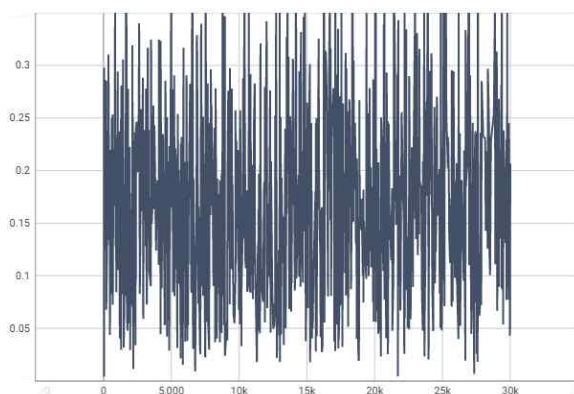


그림 2. LoRA기법을 사용한 Stable Diffusion의 미세조정 시 train loss

학습에 사용한 주요 하이퍼파라미터는 학습 step은 30000번의 step을 학습하였으며 배치 크기는 4로 설정하였다. 배치 크기가 작은 이유는 학습 시 이미지의 해상도가 512이기 때문이다. 학습에 사용한 GPU는 NVIDIA의 A5000모델 4장을 사용하여 학습하였으며 학습 82,783장의 이미지를 학습하는데 약 4시간 40분이 소요되었다. 그림 2는 LoRA기법을 사용하여 Stable Diffusion을 미세조정할 때 학습 시 loss를 나타낸 그림이다. 그림에서 볼 수 있듯이 낮은 loss 값을 가지며 균일하게 학습이 된 것을 확인할 수 있다.

웹툰 데이터 셋 중 테스트 데이터를 사용하여 미세 조정된 Stable Diffusion을 이용하여 웹툰을 생성한 결과는 그림 3과 같다. 그림 3에서 (a)는 Stable Diffusion에 입력하는 텍스트이며 (b)는 텍스트에 대한 실제 이미지이며 (c)와 (d)는 미세 조정된 Stable Diffusion을 사용하여 생성한 이미지이다. 생성한 이미지를 보았을 때 입력한 텍스트에 나타나있는 표현이 잘 반영되어 생성된 것을 확인할 수 있었으며 화풍 또한 실제 이미지에 표현된 웹툰 화풍이 반영되어 생성된 것을 확인할 수 있었다.

	DCGAN	Stable Diffusion
Inception score	4.992[10]	13.86

표 1. DCGAN, Stable Diffusion 기반의 Text-to-Image 모델의 Inception score

정량적 평가는 Inception score를 사용하여 측정하였다. 평가에 비교한 Text-to-Image모델은 DCGAN기반의 Text-to-Image 모델과 본 연구에서 제안하는 Stable Diffusion기반의 Text-to-Image 모델을 비교하였다. 본 연구에서 제안하는 LoRA기법을 사용하여 Stable Diffusion의 inception은 13.86이며 DCGAN기반의 Inception score는 4.92이며 비교결과 Stable Diffusion이 높은 것을 확인할 수 있었다.

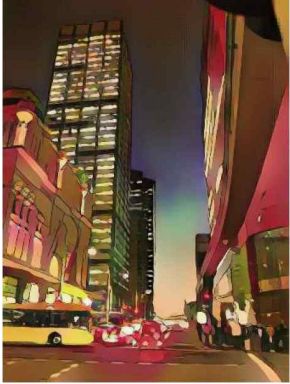


텍스트 (a)	실제 이미지 (b)	생성된 이미지-1 (c)	생성된 이미지-2 (d)
<p>A city street surrounded by tall colorful buildings.</p>			
<p>A man with a red helmet on a small moped on a dirt road.</p>			
<p>A smiling person holding a snowboard standing on a snow-covered hill.</p>			

그림 3. 미세 조정된 Stable Diffusion을 이용한 웹툰 생성

V. 결 론

본 연구에서는 작가의 웹툰 저작에 도움을 줄 수 있도록 사전학습된 Text-to-Image 모델을 미세조정하여 텍스트에서 웹툰을 생성할 수 있는 방법을 제안하였다. 제안하는 방법은 공개 데이터 셋인 MSCOCO 데이터 셋을 CartoonGAN을 사용하여 웹툰 화풍을 변환한 웹툰 데이터 셋을 사전학습한 Stable Diffusion 모델에 LoRA 기법을 사용하여 미세조정한다. 실험결과 3만 스텝의 학습횟수에서 약 4시간 반이 소요되어 학습이 빠르게 진행된 것을 확인할 수 있었다. 미세조정된 Stable Diffusion을 사용하여 이미지를 생성한 결과에서는 입력한 텍스트에 표현된 형상이나 배경이 반영되어 이미지가 생성되는 것을 확인할 수 있었으며 미세조정에 사용한 웹툰 데이터 셋의 화풍이 생성된 이미지에 반영된 것을 확인할 수 있었다. 정량적 평가는 Inception score를 사용하여 측정하였는데 13.86으로 DCGAN 기반의 Text-to-Image 모델보다 높은 것을 확인할 수 있었다. 본 연구에서 수행한 웹툰 생성을 위한 Text-to-Image 모델을 웹툰 작가가 사용한다면 웹툰 저작에 시간을 단축시킬 수 있을 것으로 예상된다. 본 연구의 제한점으로는 CLIP 인코더가 영어를 타겟으로 학습되어 다른 언어를 입력하였을 때에는 텍스트의 표현이 반영되지 않는 것을 확인할 수 있었다. 추후 연구에서는 다국어 입력하였을 때에도 품질 높은 웹툰을 생성할 수 있는 연구를 수행 하도록 하겠다.

REFERENCES

[1] 이승진, 왕덕원, "인공지능 웹툰 작가의 가능성과 성공에 대한 연구: 혁신 확산 이론을 중심으로," *엔터테인먼트연구*, pp 231-254, 2023년 6월

[2] Dhariwal, P. & Nichol, A., "Diffusion models beat gans on image synthesis," *Advances in neural information processing systems* 34, pp 8780-8794, 2021.

[3] 유경호, 노주현, 홍택은, 김형주, 김판구, "Self-Attention을 적용한 문장 임베딩으로부터 이미지 생성 연구," *스마트미디어저널*, 제10권, 제1

호, 63-69쪽, 2021년 03월

[4] 유경호, 양희덕, "Generative Adversarial Network를 이용한 카툰 원화의 라인 드로잉 추출," *스마트미디어저널*, 제10권, 제2호, 30-37쪽, 2021년 06월

[5] 김광진, 이철우, "Bi-LSTM 모델을 이용한 음악 생성 시계열 예측," *스마트미디어저널*, 제11권, 제10호, 65-75쪽, 2022년 11월

[6] Ramesh, A., Dhariwal, P., Nichol, A., Chu, C. & Chen, M., "Hierarchical text-conditional image generation with clip latents," arxiv.org/abs/2204.06125, 2022.

[7] Rombach, R., Blattmann, A., Lorenz, D., Esser, P. & Ommer, B., "High-resolution image synthesis with latent diffusion models," *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 10684-10695, 2022.

[8] Ruiz, N., Li, Y., Jampani, V., Pritch, Y., Rubinstein, M. & Aberman, K., "Dreambooth: Fine tuning text-to-image diffusion models for subject-driven generation," *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 22500-22510, 2023.

[9] Hu, E. J., Shen, Y., Wallis, P., Allen-Zhu, Z., Li, Y., Wang, S., ... & Chen, W., "Lora: Low-rank adaptation of large language models," [arXiv:2106.09685](https://arxiv.org/abs/2106.09685), 2021.

[10] 웹툰 AI 페인터. <https://ai.webtoons.com/ko/painter> (accessed Jul., 28, 2023).

[11] Kang, S., Choo, J. & Chang, J., "Consistent comic colorization with pixel-wise background classification," *In Proceedings of the NIPS*, Vol. 17, 2017.

[12] Xu, X., Xie, M., Miao, P., Qu, W., Xiao, W., Zhang, H., ... & Wong, T. T., "Perceptual-aware sketch simplification based on integrated VGG layers," *IEEE transactions on visualization and computer graphics*, Vol. 27, Issue 1, pp. 178-189, Jan. 2021.

[13] Yu, K., Kim, H., Kim, J., Chun, C., & Kim, P., "A Study on Generating Webtoons Using Multilingual Text-to-Image Models," *Applied Sciences*, 13(12), 7278, 2023.

[14] Ho, J., Jain, A. & Abbeel, P., "Denoising diffusion probabilistic models," *Advances in neural information processing systems* 33, pp 6840-6851, 2020.

[15] Kim, G., Kwon, T. & Ye, J. C., "Diffusionclip: Text-guided diffusion models for robust image manipulation," *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 2426-2435, 2022.

[16] Radford, A., Kim, J. W., Hallacy, C., Ramesh,

A., Goh, G., Agarwal, S., ... & Sutskever, "I. Learning transferable visual models from natural language supervision," *International conference on machine learning*, pp. 8748-8763, 2021.

[17] Xu, T., Zhang, P., Huang, Q., Zhang, H., Gan, Z., Huang, X. & He, X., "AttnGAN: Fine-grained text to image generation with attentional generative adversarial networks," *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 1316-1324, Salt Lake City, USA, Jun. 2018.

[18] Zhang, H., Xu, T., Li, H., Zhang, S., Wang, X., Huang, X. & Metaxas, D. N., "StackGAN: Text to photo-realistic image synthesis with stacked generative adversarial networks," *2017 IEEE International Conference on Computer Vision (ICCV)*, pp. 5907-5915, Venice, Italy, Oct. 2017.

[19] Zhu, M., Pan, P., Chen, W. & Yang, Y., "Dm-gan: Dynamic memory generative adversarial networks for text-to-image synthesis," *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 5802-5810, 2019.

[20] PEFT. <https://github.com/huggingface/peft/tree/main> (accessed Jul., 28, 2023).

[21] MSCOCO. <https://cocodataset.org/#home>, (accessed Jul., 28, 2023).

[22] Chen, Y., Lai, Y. K. & Liu, Y. J., "Cartoongan: Generative adversarial networks for photo cartoonization," *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 9465-9474, Salt Lake City, USA, Jun. 2018.

저 자 소 개



유경호(정회원)
 2017년 조선대학교 컴퓨터공학과 학사 졸업.
 2019년 조선대학교 컴퓨터공학과 석사 졸업
 2023년 조선대학교 컴퓨터공학과 박사 졸업

<주관심분야 : 인공지능, 자연어처리, 멀티모달 AI>



김형주(정회원)
 1999년 조선대학교 전산통계학과 학사 졸업(이학사).
 2002년 원광대학교 컴퓨터공학과 석사 졸업(공학사).
 2018년 조선대학교 교육대학원 정보·컴퓨터교육전공 석사 졸업(교육석사).
 2022년 조선대학교 컴퓨터공학과 박사 졸업(공학박사).

<주관심분야 : 빅데이터 처리, 인공지능, 자연어처리, 딥러닝>



김정인(정회원)
 2019년 조선대학교 컴퓨터공학과 박사 졸업(공학박사).
 2020년 (3단계 BK21) 스마트인터넷 기반 융합콘텐츠기술 인력양성사업팀 연구원(Post-Doc).
 2023년 (4단계 BK21) 인공지능기반 초연결사회 IT융합인력양성팀 연구원(Post-Doc).
 2023년~현재 조선대학교 치과대학 구강미생물학 교실 연구원(Post-Doc).

<주관심분야 : 인공지능, 딥러닝, 빅 데이터 처리, 자연어 처리>



전찬준(정회원)
 2009년 한국기술교육대학교 정보기술공학부 학사 졸업
 2011년 광주과학기술원 전기전자컴퓨터공학 석사 졸업
 2017년 광주과학기술원 전기전자컴퓨터공학 박사 졸업
 2021년 한국건설기술연구원 수석연구원
 2021년~현재 조선대학교 컴퓨터공학과 조교수

<주관심분야 : 음성 및 오디오 신호처리, 딥러닝>



김관구(정회원)
 1988년 조선대학교 컴퓨터공학과 학사 졸업
 1990년 서울대학교 컴퓨터공학과 석사 졸업
 1994년 서울대학교 컴퓨터공학과 박사 졸업
 1994~현재 조선대학교 컴퓨터공학과 교수

<주관심분야 : 지능형정보처리, 시맨틱 웹, 온톨로지, 자연어처리, 데이터 마이닝 등>