

## 생성모델의 시각적 최적화를 위한 학습데이터 제작기법

\*조형래      \*\*박구만

\*서울과학기술대학교 일반대학원 미디어IT공학과

\*\*서울과학기술대학교 전자미디어IT공학과

\*artjow@naver.com

Learning data production technique for visual optimization of generative models

\*Hyeongrae Cho      \*\*Gooman Park

\*Dept. of Media IT Engineering

\*\*Dept. of Electronics and IT Media Engineering

Seoul National University of Science and Technology

### 요약

본 논문은 생성모델의 학습데이터 제작기법에 대한 실험 및 결과와 향후 관련 연구의 방향을 기술한다. GAN으로 대표되는 생성모델이 아티스트에게 얼마만큼의 만족도와 영감을 주는지를 비교 실험 및 평가하기 위해서는 정제된 학습데이터가 필요하다. 하지만 현실적으로 아티스트의 작품은 데이터 세트를 만들기에는 그 수가 적고 인공지능이 학습하기에도 정제되어있지 않다. 2차 가공작업을 통하여 아티스트의 원본 작업과 유사한 데이터 세트의 구축은 생성모델의 성능향상을 위해 매우 중요하다. 연구의 결과 생성모델이 표현하기 어려운 스타일의 작가 작품을 선정한 뒤 최적의 학습데이터를 만들기 위한 다양한 실험과 기법을 통해 구축한 데이터 세트를 생성모델 알고리즘에 적용하고 실험을 통해 창작자의 작품제작 의도인 작가 진술에 최대한 유사한 이미지의 생성과 더 나아가 작가가 생각하지 못했던 창조적 모방의 결과물을 도출하였고 작가평가를 통해 높은 만족도를 얻었다.

키워드 : 학습 데이터셋, 생성모델, GAN, 데이터 가공 기법

### I. 서론

본 논문은 시각적 생성을 공통분모로 5명 작가(장르별 회화, 미디어아트, 설치미술)의 자유로운 창작품과 그 창작기술이 담긴 이미지와 작가 진술을 학습한 인공지능이 제작한 작품을 단계별로 비교 검토하고 생성모델의 새로운 창조적 가능성을 시현하기 위한 융합형 연구이다. 본 논문에서 실험을 위해 사용된 알고리즘은 StyleGAN2-ada[1]와 Pix2pix HD[2]이며 유의미한 결과를 얻은 학습데이터 세트 제작기법과 조형학습을 통해 생성모델이 표현하기 어려운 기하학적인 도형인 원을 비교적 정확하면서 다양한 형태로 생성하는 결과를 얻었다.

### II. 선행연구

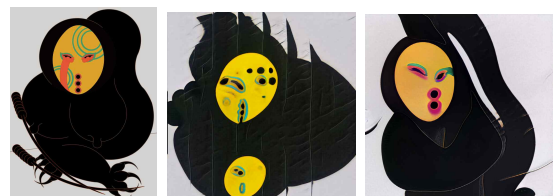
최근 이미지 생성과 관련된 대표적인 연구 중에는 텍스트를 이미지로 변환 및 생성하는 AttnGAN, CLIP[1], DALL-E[2], Stack GAN 등이 있다. 딥러닝이 좋은 성능을 내기 위해서는 네트워크망이나 알고리즘 기법과 함께 데이터가 매우 중요한다. 선행연구 중에서 DALL-E는 GPT-3로부터 파생된 것으로 설명과 단서만으로 추가 교육 없이 제로 샷 추론 기능을 활용하여 실제 존재하지 않는 이미지를 생성해 낸다. 이를 위해 120억개 매개변수가 사용되었다. CLIP은 이미지와 텍스트 데이터를 사용하여 라벨링 없이 웹 크롤링을 통하여 자동으로 이미지와 연관된 자연어 텍스트를 추출하여 이미지와 텍스트 쌍을 가진 4억 개의 거대한 데이터 세트를 구축하였다. 하지만 DALL-E와 CLIP은 데이터 세트의 크기

가 너무 커서 학습에 많은 시간과 비용이 소요되는 단점을 가지고 있다. 이에 본 연구에서는 적은 데이터 세트로도 아티스트에게 만족감과 영감을 줄 수 있는 이미지의 생성을 목표로 한다.

### III. 학습데이터 세트 구축

#### 1. 학습데이터 제작

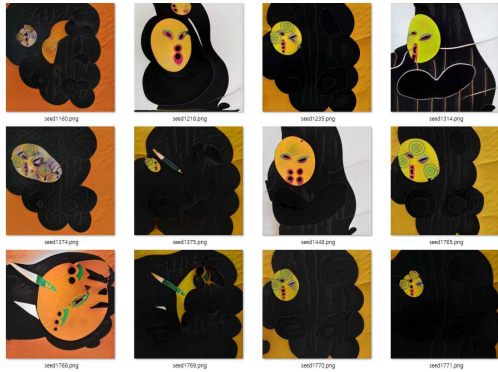
본 연구를 위해 자체적으로 학습용 데이터 세트 203,500개 이상을 구축하였으며 학습데이터를 가공하고 잘 정제하여 원, 면과 같은 조형 이미지를 만들어 학습시키는 방법을 통해 생성물에 대한 작가의 만족도를 향상 시키는 결과를 가져왔다. 아래 [그림 1]은 양대원 작가의 동글인 캐릭터의 1차 실험 결과에서는 얼굴 원 형태를 제대로 표현하지 못했으나, 2차 실험에서 다양한 원을 학습시켜 개선되었다. 그러나 원 내부의 세밀함과 작은 원까지 정확히 표현하지는 못했다.



[그림 1] (좌)원본이미지, (중앙)1차 실험 결과, (우)2차 실험 결과

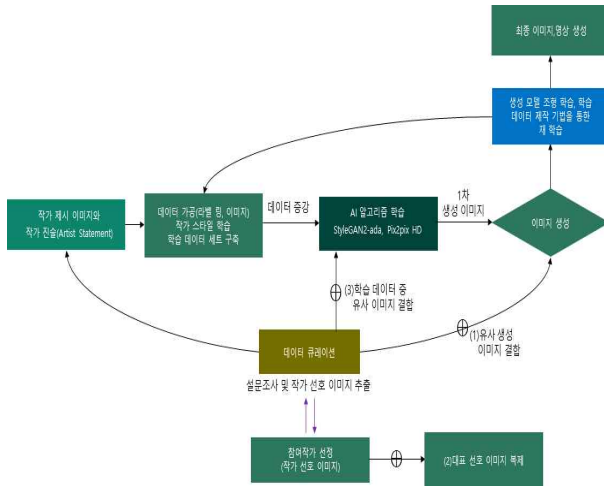
### 2. 학습데이터 제작기법

이미지 생성을 잘 하기 위해서는 학습데이터의 형태, 개수, 이미지 형태의 차이나 비율 등이 중요하다. 다음은 본 연구를 통해 유의미한 결과를 얻은 학습데이터 제작기법을 소개하고자 한다. 먼저 생성결과로부터 작가의 만족도 높은 이미지가 나오면 이와 유사한 생성 이미지 60%와 사람이 만든 학습데이터 중에서 유사한 이미지 30%를 선별한다. 이후 가장 만족도 높은 이미지 10%를 복제하여 데이터 세트를 만든다. 학습한 결과는 [그림 2]와 같이 더욱 정확한 표현과 다양한 모습으로 생성되는 것을 볼 수 있다.



[그림 2] 실험 결과 더욱 형태가 정확하고 다양한 이미지가 생성된다.

### 3. 학습데이터 제작 프로세스



[그림 3] (1)유사한 생성이미지 + (2)만족도 높은 대표이미지 복제 + (3)학습 데이터 중 만족도 높은 이미지와 유사한 이미지 모두를 합하여 데이터 세트를 만든 후 재학습을 통해 이 전에 비해 더욱 정확하고 다양한 이미지를 생성하였다.

### IV. 실험 및 결과

GAN의 경우 전체적으로 글로벌 한 형태와 질감은 잘 표현하나 작고 국부적인 특징과 관련된 '규칙'을 배워 표현하기에는 어려운 점이 있다. 따라서 학습 과정에서 조형적 요소에 대한 학습이 필요하다. 조형학습을 위해서는 학습데이터와 함께 알고리즘 기법, 네트워크 망도 중요한데 본 논문에서는 학습데이터 제작을 위한 정교한 데이터 구축을 통해 이미지 생성능력을 높이고자 한다. 아래 [표 1]과 같이 학습데이터를 그대로 학습 후 생성하는 것보다 조형학습을 시킨 것이 성능이 좋고, 학습기법을 더한 것이 더 좋은 결과를 보였다.

작가	학습데이터	1차 실험 결과	조형학습	조형학습 + 학습기법
홍경택				
유한이				
김창겸				
이돈순				
양대원				

[표 1] 작가별 학습기법 유무에 따른 생성물의 차이

### V. 결론

본 논문에서는 생성모델의 결과물이 아티스트에게 만족감과 영감을 주기 위해서 먼저 생성모델의 문제점을 파악하고 작가별 목표를 정하고 이를 실험을 통해 달성하고자 하였다. DALL-E나 CLIP처럼 커다란 학습데이터 없이 정제된 데이터와 조형학습 그리고 학습데이터 제작 기법을 이용하여 비교적 만족도 높은 결과에 도달하였다. 하지만 일부 작가의 기하학적인 도형, 패턴, 세밀한 모양, 고유성을 가지는 디자인 등에서는 학습이 쉽지 않았는데 이의 해결을 위해 향후 작은 디테일과 전체 이미지 구조에서 정확하게 기하학적인 모양을 그리는 능력을 위해 Self attention, Transformers를 활용한 실험, 조형학습을 한 네트워크 망에 대한 연구, 2D to 3D 변환을 통한 GAN의 문제점 보완과 관련된 연구를 진행하고자 한다.

### 감사의 글

이 논문은 2021년도 정부(과학기술정보통신부)의 재원으로 정보통신기획평가원의 지원을 받아 수행된 연구임 (No.2021-0-00751, 0.5mm 급 이하 초정밀 가시·비가시 정보 표출을 위한 다차원 시각화 디지털 트윈 프레임워크 기술 개발)

### 참고문헌

[1] Alec Radford, Jong Wook Kim, Chris Hallacy, Aditya Ramesh, Gabriel Goh, Sandhini Agarwal, Girish Sastry, Amanda Askell, Pamela Mishkin, Jack Clark, Gretchen Krueger, Ilya Sutskever, "Learning Transferable Visual Models From Natural Language Supervision," <https://arxiv.org/pdf/2103.00020.pdf>, arXiv:2103.00020v1 [cs.CV] 26 Feb 2021

[2] Aditya Ramesh, Mikhail Pavlov, Gabriel Goh, Scott Gray, Chelsea Voss, Alec Radford, Mark Chen, Ilya Sutskever, "Zero-Shot Text-to-Image Generation," <https://arxiv.org/pdf/2102.12092.pdf>, arXiv:2102.12092v2 [cs.CV] 26 Feb 2021