

인공지능 기반 이미지 생성 알고리즘과 사진 AI-based Image Generation Algorithm and Photography

박평종 / 중앙대학교 HK연구교수

Pyung-Jong Park / HK Research Professor, Chung Ang University

- I. I. 서론
- II. GAN의 원리와 유형
- III. 기술 이미지: 숫자의 세계
- IV. 결론

* 이 논문은 2017년 대한민국 교육부와 한국연구재단의 지원을 받아 수행된 연구임 (NRF-2017S1A6A3A01078538).

* This work was supported by the Ministry of Education of the Republic of Korea and the National Research Foundation of Korea (NRF-2017S1A6A3A01078538).

국문 초록

이 글은 인공지능 기반의 알고리즘을 활용하여 생산한 이미지가 기존의 사진술과 어떤 차별성을 갖는지를 다룬다. ‘생성적 적대 신경망(Generative Adversarial Networks)’은 생성자와 판별자의 균형 있는 학습을 통해 가짜 데이터를 산출하는 알고리즘으로, 원본 사진을 학습 데이터로 활용할 경우 시각적으로 사진과 거의 구분되지 않는 이미지를 만들어 낸다. 이 이미지가 제기하는 논점은 두 가지다. 첫째, 알고리즘 이미지는 지표이론의 관점에서 볼 때 사진이라 할 수 없으나 원본 사진의 픽셀 값을 바꾸어 새로운 이미지를 만들어 낸다는 점에서 기존의 디지털 사진과 다를 바 없다. 둘째, 알고리즘 이미지는 플루서가 기술 이미지의 핵심으로 규정한 프로그램의 자동성과 인간배제가 고도화된 형태다. 인간은 GAN 알고리즘 이미지의 생산과정으로부터 배제돼 있다. 그런 점에서 GAN의 생성자는 블랙박스다. 프로그램의 자동성이 강화될수록 인간은 이미지 생산을 통제하지 못하고 단순 소비자로 전락하고 만다. 따라서 이제 인간은 프로그램이라는 블랙박스를 어떻게 ‘투명한 상자’로 만들 수 있을지 고민할 시점이 됐다.

핵심어 | 디지털 사진, 빌렘 플루서, 생성적 적대신경망, 알고리즘, 포스트 포토그래피, 프로그램

ABSTRACT

This article deals with how images produced using artificial intelligence-based algorithms differ from existing photography. A generative adversarial network (GAN) is an algorithm that calculates fake data through balanced learning of generators and discriminators. When using original photos as learning data, it creates images that are visually indistinguishable from photos. There are two issues raised by this. First, the algorithmic image cannot be called a photograph from the point of view of index theory, but it is no different from the existing digital photograph in that it creates a new image by changing the pixel value of the original photograph. Second, the algorithmic image is an advanced form of program automatism and human exclusion, which Vilém Flusser defined as the core of the technical image. Humans are excluded from the production process of the GAN algorithm image. In that sense, the generator of the GAN is a black box. As the automaticity of the program increases, humans do not control image production and become simple consumers. Therefore, it is time for humans to think about how to turn a black box called a program into a “transparent box.”

Key Words | Algorithm, Digital photo, GAN, Post photography, Program, Vilém Flusser

I. 서론

이 글은 인공지능 기반의 알고리즘을 활용한 이미지 생성 기술의 원리를 살펴보고 그것이 포토몽타주나 포토샵 등을 통해 시뮬라크루를 생산해 온 기존의 사진술과 어떤 차별성을 갖는지 검토해 보는 것을 목표로 한다. 알고리즘 이미지의 종류와 유형은 다양하며 빠르게 진화하는 추세다. 특히 딥러닝을 통해 이미지의 식별과 분류 기능이 강화되면서 이와 연계된 이미지의 '생성' 능력 또한 눈부시게 발전하고 있다. 불과 수년 전까지만 하더라도 알고리즘의 이미지 생성력은 보잘 것 없었다. 그런데 새로운 딥러닝 기술을 적용한 알고리즘이 속속 개발되면서 이미지 생산의 품질과 속도는 과거와 비교할 수 없을 정도로 발전했다. 딥러닝에 기초한 이미지 생성 기술 중 최근 가장 주목을 끄는 알고리즘은 '생성적 적대 신경망(Generative Adversarial Networks, 이하 GAN)'이다. 2014년 이안 굿펠로우(Ian Goodfellow)가 발표한 이 알고리즘은 기존의 딥러닝 기술을 수정하여 '비지도학습(Unsupervised learning)' 원리에 따라 컴퓨터가 스스로 학습을 통해 이미지를 생성하도록 고안됐다.¹ GAN의 원리는 이후 개량을 거듭하여 DCGAN, PGGAN, CycleGAN, StyleGAN 등 다양한 이미지 생성, 수정 알고리즘으로 진화했다.

이 알고리즘을 적용하여 실제 이미지 생산에 활용하는 경우도 늘고 있다. 대표적인 예는 인공지능 컴퓨팅 분야의 주요 기업 중 하나인 엔비디아(Nvidia)가 2019년부터 StyleGAN을 활용하여 생산한 이미지를 서비스하는 플랫폼이다. "Generated Photos"(이하 GP)라는 명칭의 이 플랫폼은 "전적으로 인공지능이 생산한 사진"² 이백만 장 이상을 보유하고 있는데, 주로 사람 얼굴 이미지를 공급하고 있다. 수요자는 필요에 따라 나이, 성별, 인종, 머리카락 길이와 색깔, 눈동자 색깔 등을 선택할 수 있으며, 플랫폼은 그 선택에 부합하는 초상을 무작위로 제공한다.

이 과정에서 생겨난 핵심 논점은 크게 두 가지다. 첫째, 생산 주체의 문제다. 알고리즘 이미지는 문자 그대로 컴퓨터 프로그램에 따라 생산된다. 즉 이미지의 생산자는 인간이 아닌 기계며, 생산 과정에서 인간은 배제돼 있다. 이 문제는 자동성과 직결된다. 인간의 개입이 차단된

1. 알고리즘은 새로운 데이터를 만들어내는 것이 목표다. 이 목표를 달성하려면 주어진 데이터를 학습해야 한다. 지도학습(Supervised learning)은 데이터에 레이블을 붙여 주요 특성을 학습시키는 방법이다. 예컨대 강아지 사진은 강아지로, 고양이 사진은 고양이로 분류하도록 학습시키는 과정에서 기계는 식별과 분류 능력을 배운다. 반대로 비지도학습에서는 레이블이 없는 데이터를 기계 스스로 학습한다. 방법은 데이터에서 유사한 패턴들을 군집화 하는 것이다. 예를 들면 수많은 고양이 사진을 학습하는 과정에서 기계는 고양이의 주요 특성(뽀족한 귀, 동그란 눈, 긴 꼬리 등)을 모아 분류하고 그와 유사한 패턴을 보이는 사진은 고양이로 식별한다.

2. 이 표현은 플랫폼 초기 화면에 등장하는 소개 문구다. <https://generated.photos/> (2020년 12월 2일 최종 접속).

상태에서 이미지가 ‘자동으로’ 형성되기 때문이다. 이미지 생산의 자동성을 언급한 이는 앙드레 바쟁이다. 그에 따르면 사진은 최초의 ‘자동생성’ 이미지다. 즉 “처음으로 외부 세계의 이미지가 엄격한 결정론에 따라 인간의 창조적 개입 없이 자동적으로 형성”³⁾된다는 것이다. 플루서는 장치 개념을 통해 이 문제를 좀 더 멀리 밀고 나간다. 카메라는 최초의 장치며 그 본질은 자동성을 지향하는 프로그램에 있다. 말하자면 사진은 프로그램에 따라 인간이 배제된 상태에서 자동으로 생성된 이미지다. 알고리즘 기반의 이미지는 이런 특징, 요컨대 인간배제와 자동성이 고도화된 형태다.

둘째, 알고리즘 이미지의 성격이다. GAN의 원리에 따라 생성된 이미지는 실재와 구분이 거의 불가능하다. 즉 사진처럼 보인다. 예컨대 GP가 제공하는 이미지는 실제 존재하는 인물이라고 착각할 정도로 놀라운 사실성을 지니고 있다. 그러나 그 인물들은 어디에도 없다. 가상의 인물이기 때문이다. GP가 활용한 StyleGAN 알고리즘은 학습용 데이터(사진)로부터 인물을 구성하는 특질들을 추출하여 이를 새롭게 조합해낸다. 그렇다면 GP 이미지는 사진일까? 플랫폼은 ‘생성된 사진(generated photos)’이라 규정하고 있다. 그리고 실제 그 이미지들은 사진과 구분이 어렵다. 그러나 그 ‘사진’의 생성 원리는 일반적인 사진의 그것과는 다소 차이가 있다. 사진의 ‘특수성’을 규정하고자 했던 ‘전통적인’ 사진이론에 따르면 사진은 빛을 통해 전달된 이미지가 감광물질의 작용으로 고정된 형태다. 따라서 사진 찍힌 대상과 이미지는 ‘물리적’ 접촉이라는 특수한 관계를 맺고 있다. 이런 생각은 1980년대 이후 사진이론을 지배했던 소위 ‘지표이론’으로 수렴한다. 그런데 알고리즘 이미지는 ‘시각적으로’ 사진과 다를 바 없으나 생성원리는 다르다. 사진은 실제 존재하는 대상이 렌즈 앞에 ‘출현’해야 하나 알고리즘 이미지는 단지 학습용 사진의 픽셀 값을 계산하여 패턴을 추출하여 그와 유사한 형태를 만들어낼 뿐이다.

이 점을 고려할 때 알고리즘 이미지는 다음과 같은 논의의 필요성을 제기한다. 만약 GAN 이미지를 사진으로 규정할 수 있다면 ‘전통적인’ 사진이론은 수정되어야 한다. 이 관점에 따르면 사진은 대상과 이미지 사이의 인과관계, 물리적 접촉 등을 요구하는 ‘지표이론’과 결별한다. 반대로 GAN 이미지가 지표이론의 관점에서 사진이 아니라면, 그에 대한 새로운 개념 규정이 요구된다. 이 때 알고리즘 이미지는 사람의 손을 떠나 장치 프로그램에 따라 생산되는 ‘기술 이미지(플루서)’의 범주에 속한다. 이런 문제의식을 바탕으로 이 글은 우선 GAN 알고리즘의 원리를 살펴보고 그 원리를 적용하여 실제 이미지 생성과 수정에 활용되고 있는 다양한 GAN 알고리즘의 유형들을 정리, 소개하고자 한다. 나아가 그 이미지가 기존의 디지털 사진 및 합성과 수정을 통해 변형된 가상 이미지와 어떤 공통점과 차별성을 갖는지, ‘전통적인’ 사진이론의 관점

3. André Bazin, “Ontologie de l'image photographique”, *Qu'est-ce que le cinéma?* (Paris: Cerf 1945), p. 13.

에서 어떤 의미를 지니는지 살펴볼 것이다. 이를 통해 알고리즘 이미지가 진화한 형태의 ‘기술 이미지’임을 주장하고자 한다.

II. GAN의 원리와 유형

1. GAN의 기본 구조

GAN의 핵심 개념은 2014년 이안 굿펠로우가 발표한 논문 「생성적 적대 신경망 Generative Adversarial Nets」⁴에서 처음 공식화되었다. 이 알고리즘은 딥러닝에 비지도학습(Unsupervised learning) 방식을 적용하여 정보를 산출하도록 고안된 프로그램이다. 이 학습방법에 적용된 모델은 두 가지다. 첫째는 생성자(Generator), 둘째는 판별자(Discriminator)다. 생성자는 학습 데이터를 토대로 그와 유사한 새로운 정보를 산출하는 것이 목표다. 판별자는 생성자가 산출한 데이터가 원래 데이터와 같은지 다른지를 식별하는 것이 목표다. 말하자면 두 모델은 서로 적대적 경쟁관계에 있다. 생성자는 진짜에 가까운 가짜를 생산하려 하고, 판별자는 판별을 통해 가짜의 생산을 억제하려 한다. 이 구조를 논문은 위조지폐범과 경찰의 관계로 설명한다. 위폐범은 경찰의 눈을 속이기 위해 진짜와 구분이 어려울 만큼 그럴듯한 가짜를 생산하려 한다. 생성자의 목표 또한 학습한 데이터를 토대로 원본과 흡사한 가짜 데이터를 산출하는 데 있다. 이때 두 모델 중 어느 한편의 능력이 과도하게 뛰어나면 상대 모델의 능력이 발전할 수 없다. 만약 학습의 초기 단계에서 생성자가 산출하는 정보가 원본과 크게 다름에도 판별자가 진짜라고 판별한다면 생성자는 더 이상 발전하지 못한다. ‘대충’ 생산해도 판별자의 눈을 속일 수 있기 때문이다. 반대로 판별자의 식별력이 한 단계 발전하면 생성자는 판별자를 속이기 위해 훨씬 그럴듯한 가짜를 만들어 내고자 할 것이다. 감식안이 뛰어난 경찰을 속이기 위해 위폐범들의 기술이 고도화되는 것과 같은 이치다. 따라서 판별자의 능력이 고도화됨에 따라 생성자의 능력도 함께 발전하며, 그 역도 마찬가지다. 즉 두 모델의 능력이 비슷해야 학습이 이루어질 수 있다.

여기서 새로운 데이터를 산출하는 핵심 원리는 손실함수(Loss function)다. 입력 값과 출력 값이 달라야 ‘새로운’ 가짜 데이터가 나오는데 이는 입력데이터에 변수를 추가함으로써 가능하다. 변수가 주어지면 입력 값과 출력 값 사이에 손실이 생기고, 이 편차는 다양한 확률분포를 보이면서 커지거나 작아진다. GAN의 목표는 원본과 흡사한 가짜 데이터를 산출하는 데 있

4. Ian J. Goodfellow et al., “Generative Adversarial Nets”, arXiv:1406.266, 1v1 (Jun, 10, 2014).

기 때문에 가장 가까운 근사치를 만들어내야 한다. 즉 손실(loss)을 줄여야 한다. 따라서 딥러닝에서 손실함수는 입력 데이터와 산출 데이터 사이의 차이를 줄이는 방향으로 연산을 진행한다. 손실함수는 입력 값과 산출 값이 같을 때 0으로 처리하므로 GAN에서 이 값은 무한히 0으로 수렴한다. 물론 GAN에게 정답은 없으며, 랜덤하게 정답에 가까운 오답을 무한히 만들어낼 따름이다. 생성자가 산출한 데이터에 손실이 크면 원본과 달라 판별자는 가짜로 판별하고 이 과정이 되풀이됨에 따라 손실 값은 줄어들어 원본과 거의 흡사해지는 구조다.

GAN은 본래 이미지 데이터뿐 아니라 음성을 포함한 오디오 파형, 자연어 등 다양한 종류의 데이터를 학습할 수 있도록 고안됐다.⁵ 그러나 적대적 경쟁관계가 균형 있게 유지되어야만 학습이 제대로 이루어지는 특수성 때문에 만족스런 결과물을 내놓지 못했다. 특히 불연속적 기호로 구성된 데이터셋, 예컨대 자연어의 경우 학습이 쉽지 않았다. 반면 이미지처럼 연속적 기호로 구성된 데이터의 경우 상대적으로 학습 효과가 뛰어나지만 실제로는 매우 불안정했다.⁶ 생성자와 판별자가 균등하게 발전해야 하는데 실제 학습과정에서는 판별자가 생성자를 압도하는 구조로 진행되는 경우가 많았기 때문이다. 그래서 GAN을 활용하여 생산한 이미지는 원본 데이터와 달리 터무니없게 변형되는 경우가 발생했다.⁷ 이런 문제를 해결하기 위해 고안된 알고리즘 중의 하나가 DCGAN이다.

2. 학습의 안정성: DCGAN과 BEGAN

DCGAN의 원리를 간략히 정리하면 기존의 GAN과 합성곱신경망(Convolutional Neural Networks, 이하 CNN)을 합한 알고리즘이라 할 수 있다. 본래 CNN은 이미지 식별에 사용되는 인공신경망이다. CNN이 이미지를 식별할 때 사용되는 기계학습은 지도학습의 원리를 따르는데, 이를 GAN의 비지도학습에 적용한 셈이다. CNN의 핵심 원리는 인공신경망 구축에 합성곱(Convolution)이라 불리는 레이어를 여러 단계로 구축하는 데 있다. CNN은 이미지 식별을 위해 레이블이 붙은 데이터를 학습해야 한다. 합성곱은 학습 과정에서 사람의 뇌를 모방한 인공신경망을 겹겹이 여러 층으로 구축함으로써 식별력을 높여준다. 이 기술을 GAN과 접목시킨

5. Goodfellow et al., "Generative Adversarial Nets", p. 1.

6. 불연속적인 기호의 학습이 쉽지 않은 이유는 예컨대 문자가 '차이'에 기초한 기호이기 때문이다. 입력 값과 출력 값 사이의 편차를 아주 조금씩 줄여나가야 학습이 진행되는데 이 경우 '미세한' 손실을 만들어내기 어렵다. 반면 연속적인 기호는 '유사'의 원리를 따르기 때문에 미세한 편차를 만들어내기 쉽다. GAN이 이미지 학습에서 상대적으로 만족스런 결과를 내는 이유는 그 때문이다.

7. Alec Radford & Luke Metz, "Unsupervised Representation Learning with Deep Convolutional Generative Adversarial Networks", arXiv:1511.06434v2 (Jan. 7, 2016), p. 1.

DCGAN은 GAN의 학습능력을 향상시켜 이미지 생성에 적합한 알고리즘으로 진화했다. 나아가 이후 GAN의 원리를 적용하여 개량된 대부분의 알고리즘은 DCGAN의 기본 구조, 즉 합성곱이라는 다수의 레이어를 활용하여 학습능력을 향상시켰다.

GAN의 불안정성을 줄여 이미지의 품질을 비약적으로 높여준 또 다른 알고리즘이 BEGAN이다. 기본 원리는 GAN이지만 학습능력 향상을 위해 생성자와 판별자의 균형 있는 발전을 위해 몇 가지 다른 알고리즘을 추가한 모델이다. 2017년 발표된 논문에서 저자들은 이 알고리즘의 목적을 다음과 같이 밝힌다. 빠르고 안정되게 수렴하는 표준학습 방법을 탑재한 GAN의 구축, 생성자에 작용하는 판별자의 힘을 조절해주는 균형 개념의 도입, 이미지의 다양성과 품질 사이의 균형을 잡아주는 새로운 방법의 도입 등이다.⁸ 실험결과로 논문이 제시한 이미지들은 기존 GAN을 통해 생성된 ‘비현실적인’ 이미지와 달리 사진과 매우 유사할 뿐 아니라 고해상도의 품질을 얻어낼 수 있었다.⁹

3. 해상도 향상: SRGAN과 PGGAN

이미지 품질, 즉 해상도를 높이는 문제는 모든 유형의 GAN에게 중요한 사안이다. 해상도가 낮을 경우 확대했을 때 픽셀의 형태가 깨져 사실적인 이미지를 얻어낼 수 없기 때문이다. 따라서 원본 사진과 유사한 이미지를 얻기 위해서는 해상도가 높은 사진을 학습해야 한다. 그런데 초기의 GAN에서 해상도가 낮은 사진의 학습은 상대적으로 원활했지만 고해상도의 사진에서는 학습이 제대로 이루어지지 않았다. 따라서 생성자는 고해상도의 ‘사실적’ 이미지를 산출하기 어려웠다. 이 문제를 해결하는 데 기여한 알고리즘이 SRGAN이다.¹⁰ 고해상도 이미지를 얻기 위한 기존의 방법들이 봉착했던 문제는 섬세한 텍스처를 복원하기 어렵다는 것이었다. 예컨대 손실함수는 딥러닝에서 정답에 대한 오류를 숫자로 표현한 것으로 픽셀로 구성된 이미지 학습에서는 주로 평균제곱오차(Mean Squared Error)가 사용된다. 그런데 MSE는 픽셀들 간의 차이에 대한 평균값을 계산하는 함수다. 따라서 복원된 이미지는 모든 픽셀들을 일률적인 평균값으로 처리한 탓에 지나치게 부드럽고 흐릿한 상태로 처리되곤 했다. 이 문제를 해결하기 위해 SRGAN은 복구할 이미지와 산출될 이미지 사이의 손실을 최소화시키고자 한다. 방법은 GAN을 활용한 학습에 있다. 우선 고해상도 사진(High resolution)에 노이즈 등을 주입하여 학습용

8. David Berthelot, Thomas Schumm & Luke Metz, “BEGAN: Boundary Equilibrium Generative Adversarial Networks”, arXiv:1703.10717v4 (May 31, 2017), p. 1.

9. 해상도 32x32와 64x64의 학습용 데이터는 실험결과 128x128로 향상되었다. Berthelot, “BEGAN: Boundary Equilibrium Generative Adversarial Networks”, p. 7.

10. Christian Ledig et al., “Photo-Realistic single image Super-Resolution using a Generative Adversarial Network”, arXiv:1609.04802v5 (May 25, 2017).

저해상도 사진(Low resolution)으로 변환한다. 이 사진을 다시 알고리즘의 궁극적 목표인 초해상도(Super resolution)로 바꾸도록 훈련시킨다. 이 과정에서 판별자는 MSE처럼 픽셀의 평균 값을 비교하는 것이 아니라 세부적인 디테일간의 손실까지 비교하여 복원을 진행한다. 결국 학습이 충분히 진행되면 판별자는 HR과 SR 이미지를 구분하기 힘들 정도가 되며 결과물은 “포토-리얼리스트틱”한 이미지로 산출된다.¹¹

이미지의 해상도를 높이는 데 크게 기여한 또 다른 알고리즘이 PGGAN이다.¹² 이 알고리즘은 해상도 1024x1024 픽셀의 이미지를 생성하는 놀라운 결과를 보여주었다.¹³ 이 결과물을 얻기 위해 PGGAN이 적용한 방법은 한 마디로 생성자와 판별자 모두를 점진적으로 성장시키는 것이다. GAN의 고질적 문제였던 양자의 불균형을 해결하기 위해 더디지만 천천히 훈련시키는 방법을 선택한 셈이다. 학습은 4x4 픽셀의 저해상도 이미지에서 출발한다. 저해상도 사진의 학습은 쉽기 때문에 이 단계가 완료되면 좀 더 어려운 학습으로 넘어가 8x8 픽셀의 사진을 레이어로 추가하여 새로운 값을 예측(산출)하도록 학습시킨다. 여기서 생성자가 산출한 예측 값을 판별자가 진짜로 판별하면 다시 16x16의 새 레이어를 추가하여 동일한 방식으로 학습을 진행한다. 이처럼 새로운 레이어를 추가하는 과정은 최종적인 이미지, 즉 1024x1024 픽셀의 고해상도 이미지를 산출할 수 있을 때까지 반복된다 [도 1]. 말하자면 레이어는 생성과 판별이 진행되는 동안 계속 추가되면서 새로운 학습 자료로 제공되는 셈이다. 이 때 인공지능망에 새로운 레이어가 추가되면 전 단계의 레이어는 삭제한다. 이는 이미 충분히 학습된 저해상도 레이어로부터 발생할 수 있는 변수를 제거하기 위함이다.¹⁴ CelebA 데이터셋을 활용하여 산출한 이미지는 대단히 정교하고 섬세하며 픽셀 값은 모두 1024x1024로 동일하다. 실상 CelebA의 원본 데이터는 매우 다양해서 해상도는 43x55에서 6732x8984로 제각각이었으나¹⁵ 결과물은 하나로 통일되었다.

11. Ledig et al., “Photo-Realistic single image Super-Resolution using a Generative Adversarial Network”, p. 2.

12. Tero Karras et al., “Progressive Growing of GANs for improved quality, stability, and variation”, arXiv:1710.10196v3 (Feb. 26, 2018).

13. PGGAN이 활용한 데이터셋은 CelebA, Lsun, Cifar10 세 가지다. CelebA는 유명 인사들의 얼굴사진 200,000장으로 구성되어 있으며, Lsun과 Cifar10은 일상에서 볼 수 있는 다양한 물건이나 동물, 풍경 등을 촬영한 사진 데이터다. Cifar10은 32x32 비트 사진 60,000점을 보유한 라이브러리다.

14. Karra et al., “Progressive Growing of GANs for improved quality, stability, and variation”, p. 2.

15. PGGAN에게 학습용으로 주어진 CelebA 원본 데이터셋은 총 30,000장의 사진으로 해상도뿐만 아니라 형태도 제각각이었다. 인물의 얼굴은 한 명에서 군중까지 천차만별이었으며, 촬영 거리와 각도도 달라 얼굴 형태는 모두 정면을 보여주지는 않았다. 새로운 데이터셋을 만들기 위해 연구팀은 이 사진들을 가공하여 표준화시킨 후 학습데이터로 활용했다. Karra et al., “Progressive Growing of GANs for improved quality, stability, and variation”, p. 14.



[도 1] PGGAN의 학습 구조와 생성된 고해상도 이미지

4. 이미지 치환: Pix2Pix와 CycleGAN

GAN의 원리는 이미지 치환에도 유용하게 활용된다. Pix2Pix나 CycleGAN은 입력 이미지를 그와 짝을 이루는 다른 이미지로 변환시키는 알고리즘으로 개발되어 큰 반향을 불러일으켰다. Pix2Pix는 흑백사진을 컬러사진으로 바꾸거나 윤곽선으로 구성된 그래픽을 사진으로 바꾸는 이미지 변환 기술이다 [도 2]. 이 알고리즘의 개발자들은 예컨대 선형그래픽 이미지(엣지맵)를 사진(비트맵)으로 바꾸는 과업이 영어를 프랑스어로 번역하는 것과 유사하다고 생각하여 이를 이미지 번역(Image-to-image translation)이라 명명했다.¹⁶ 충분한 훈련 데이터가 주어졌을 때 자동번역기의 번역능력이 향상된다는 점에서 이미지 치환 알고리즘은 상대적으로 쉽게 과업을 수행한다. 원리는 짝을 이루는 두 이미지를 모두 학습시켜 상호 ‘번역’이 가능하도록 훈련시키는 것이다. 예컨대 엣지맵으로 구성된 구두 이미지를 구두사진으로 변환할 때 Pix2Pix는 두 유형의 이미지를 모두 학습한다. 영어와 프랑스어의 상호번역을 위해서는 두 언어의 문법을 모두 학습해야 하는 것과 같은 이치다.

한편 구조가 다르거나 불규칙한 이미지의 변환은 쉽지 않다. 짝을 이루는 두 이미지의 상호 변환은 상대적으로 간단하지만 짝이 없는 이미지의 경우 그에 상응하는 이미지를 만들기는 어렵다는 뜻이다. 즉 사과 사진을 오렌지 사진으로 바꾸거나 말 사진을 얼룩말 사진으로 변환하는 과업, 나아가 그림을 사진으로, 반대로 사진을 그림으로 변환하는 것은 다른 차원의 문제다. 이 과업을 해결한 알고리즘이 CycleGAN이다.¹⁷ 개발자들은 논문에서 이 문제에 대한 아이디어

16. Phillip Isola et al., “Image-to-Image Translation with Conditional Adversarial Networks”, arXiv:1611.07004v3 (Nov. 26, 2018), p. 1.

17. Jun-Yan Zhu et al., “Unpaired Image-to-Image Translation using Cycle-Consistent Adversarial Networks”,



[도 2] Pix2Pix가 수행한 과업의 예

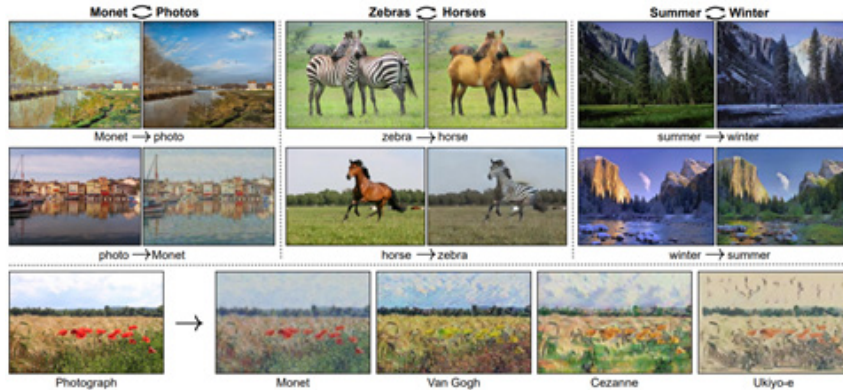
를 다음과 같이 설명한다. 만약 1873년 화창한 어느 봄날 모네가 아르장퇴유(Argenteuille) 부근의 센 강변에 이젤을 놓고 앉아있었다면 풍경은 그에게 마치 한 장의 컬러사진처럼 펼쳐졌을 테고 그는 그 ‘사진’을 자신의 화풍에 따라 그려나갔을 거라고 가정할 수 있다는 것이다.¹⁸ 우리가 사진과 그림이 지닌 스타일의 차이를 알고 있는 한 사진이 그림으로 ‘번역’되더라도 양자가 왜 같은 장면을 재현하고 있는지 알 수 있다는 것이 개발자들의 논리다.

이런 가설에 따라 어떤 사진을 모네 풍의 그림으로 변환하고자 할 때 봉착하는 난관은 그에 대응하는 모네의 그림이 없다는 것이다. 말 사진을 얼룩말 사진으로 변환하고자 할 때도 정확히 같은 자세를 취한 얼룩말 사진이 없다면 변환은 쉽지 않다. 이 문제를 해결하기 위해 CycleGAN은 생성자와 판별자를 순환 구조로 바꾼다. 일반적인 GAN과 달리 CycleGAN은 두 개의 생성자/판별자를 사용하는 것이다. 입력 값에 대응하는 출력 값을 결과물로 산출할 때 출력 값은 원본과 짝을 이루지 않을 수 있다. 즉 위에서 예로 든 모네의 경우 입력한 사진과 같은 장면을 보여주는 모네 그림이 나오지 않을 확률이 높다. 결국 CycleGAN은 역방향으로 생성과 판별을 진행하는 또 다른 GAN구조를 추가했다. 요컨대 하나의 생성자/판별자는 그림을 사진으로, 다른 생성자/판별자는 사진을 그림으로 변환하는 학습을 무수히 반복하는 구조다. 즉 순환구조를 통해 두 이미지가 짝을 이룰 수 있도록 학습을 진행시켜 근사치를 얻어내는 것이다. 학습이 성공적으로 진행되어 나온 결과물은 매우 다양하다. CycleGAN은 말을 얼룩말로, 겨울 풍경을 여름풍경으로, 평범한 풍경사진을 모네풍, 고흐풍, 세잔풍 그림으로 변환시키는 것은

arXiv:1703.10593v6 (Nov. 15, 2018).

18. Zhu et al., “Unpaired Image-to-Image Translation using Cycle-Consistent Adversarial Networks”, p. 1.

물론이고, 사과를 오렌지로, 강아지를 고양이로 바꾸는 과업까지도 성공적으로 수행했다 [도 3].



[도 3] CycleGAN이 수행한 이미지 생성의 예

5. 이미지 속성 변환: StarGAN

이미지 변환 기술에 또 하나의 변화를 가져온 알고리즘이 StarGAN이다.¹⁹⁾

이 알고리즘은 기존의 Pix2Pix나 CycleGAN과 달리 이미지를 구성하는 구체적인 특성들을 변환하는 것이 목표다. 차별성을 명확히 밝히기 위해 논문의 저자들은 ‘도메인(domain)’과 ‘속성(attribute)’을 구분한다. 속성은 “하나의 이미지에 내재된 의미 있는 특성, 예컨대 검은 머리/금발 머리/갈색 머리”등이며, 도메인은 “동일한 속성을 공유하는 이미지들의 집합”으로, 예를 들면 “여성들의 이미지는 하나의 도메인을, 남성들의 이미지는 또 다른 도메인을 표상”한다.²⁰⁾ 검은 머리는 한 인물의 속성에 해당하나 머리카락 색깔은 도메인을 표상한다. 같은 의미로 여성은 성별이라는 도메인에 해당하는 속성이다. 따라서 ‘구체적’ 속성들의 집합은 도메인이 될 수 있다. 개별과 보편의 차이라 할 수 있겠다. Pix2Pix나 CycleGAN은 도메인 대 도메인 변환이며, 입력된 이미지의 속성을 바꾸지는 못한다. 한편 StarGAN은 하나의 도메인 안에서 속성을 변환시키는 알고리즘으로 예컨대 인물의 머리색깔이나 피부색, 나이가 나이와 성별, 표정까지 바꾼다. 즉 이 알고리즘은 하나의 도메인에 해당하는 속성을 바꾸는데, 그 도메인의 종류는

19. Yunjey Choi et al., “StarGAN: Unified Generative Adversarial Networks for Multi-Domain Image-to-Image Translation”, arXiv:1711.09020v3 (Sep. 21, 2018).

20. Choi et al., “StarGAN: Unified Generative Adversarial Networks for Multi-Domain Image-to-Image Translation”, pp. 1-2.

여럿이다. 말하자면 다양한 데이터셋으로부터 복수 도메인(Multiple domains)을 학습하여 도메인의 교차 변환이 가능하도록 설계된 모델인 셈이다. 논문이 제시하는 예에서 입력된 이미지(얼굴사진)는 4개의 도메인, 즉 머리 색깔, 성별, 나이, 피부색을 성공적으로 변환시켰다 [도 4].

한 장의 사진에서 복수의 도메인을 바꾸기 위해 StarGAN은 생성자에게 원본 데이터의 속성

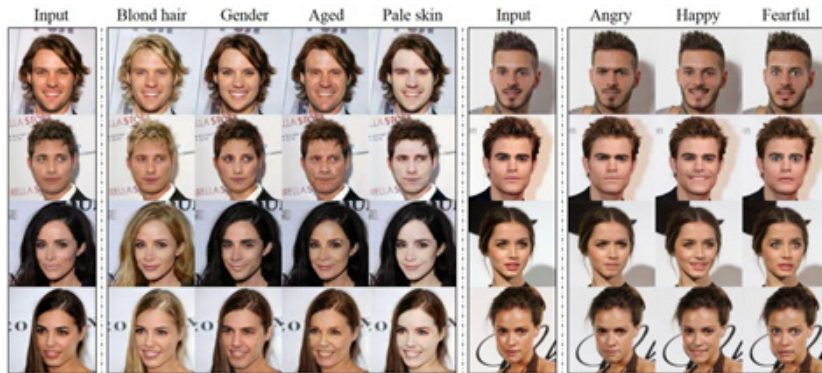


Figure 1. Multi-domain image-to-image translation results on the CelebA dataset via transferring knowledge learned from the RaFD dataset.

[도 4] StarGAN이 수행한 생성 이미지의 예

뿐만 아니라 도메인 정보도 함께 학습하도록 했다. 이 때 생성자가 학습한 도메인 정보는 하나가 아니라 여럿이다. 이는 컴퓨터가 도메인 간의 차이를 스스로 구분할 수 없기 때문이다. 요컨대 어떤 데이터가 머리 색깔에 해당하는지, 성별에 해당하는지 모른다는 것이다. 따라서 본래 GAN은 비지도학습의 원리를 따르지만 StarGAN이 도메인 정보를 학습할 때는 지도학습 방식을 취한다. 즉 도메인의 레이블 정보를 함께 입력한다는 뜻이다. StarGAN이 활용한 학습데이터는 CelebA와 RaFD로 각 데이터셋에는 도메인 레이블이 있다. 예를 들면 한 인물사진에 대한 레이블은 검은 머리, 남성, 흰 피부 등이거나 감정 상태를 규정한 ‘행복한’, ‘화난’, ‘두려운’ 등이다. 생성자는 결국 입력 데이터로부터 이미지 속성과 도메인을 함께 학습하여 손실함수를 적용하여 가짜 데이터를 계속해서 산출한다. 이 때 발생할 수 있는 문제는 정확히 목표로 하는 도메인의 변화가 나오지 않을 수 있다는 점이다. 왜냐하면 각 데이터셋마다 붙어있는 도메인 레이블에는 차이가 있고 어떤 도메인 정보는 아예 누락되어 있기 때문이다. 예컨대 CelebA에는 성별, 머리색, 나이 등의 도메인 정보가 표기돼 있지만 RaFD에는 감정 도메인 정보만 기록돼 있다. 결국 개발자들은 데이터 입력 시 알 수 없는 레이블은 무시하고 명확한 데이터셋의 레이블 정보만을 학습시키는 방법을 택했다. 이처럼 복수의 도메인 정보를 교차 학습하여 StarGAN이 내놓은 결과는 한 인물의 얼굴 사진으로부터 그 인물의 머리색, 성별, 나이, 피부색을 변환

시킨 이미지들이거나(CelebA 데이터셋), 동일 인물의 화난 표정, 웃는 표정, 두려워하는 표정으로 변환된 이미지다(RaFD 데이터셋).

6. 이미지 스타일 변환: StyleGAN

이처럼 다양한 유형의 GAN 알고리즘이 개발되면서 이미지의 품질과 해상도는 이전에 비해 비약적으로 향상됐다. 그러나 실제 이미지 생성이 어떻게 진행되는지는 알고리즘 설계자인 프로그래머조차도 알지 못한다. 즉 “생성자는 블랙박스처럼 작동”하며, “이미지 합성 과정의 다양한 측면, 예를 들면 이미지의 확률적 특성의 기원에 대해서 알 수 없”다는 것이다.²¹ StyleGAN의 설계자들은 이미지 합성을 통제할 수 있는 새로운 방법을 찾기 위해 이 알고리즘을 고안했다. 방법은 입력 값을 학습한 생성자가 학습이 진행됨에 따라 각각의 레이어에서 해당 이미지의 스타일을 조정하는 것이다. 이미지는 여러 스타일의 조합이라는 생각이 바탕에 깔려있는 셈이다. 이 알고리즘의 명칭이 스타일 기반(style-based) 생성자인 까닭도 그 때문이다.²²

StyleGAN이 각 레이어에서 스타일을 조절하는 과정은 크게 세 단계로 구분된다. 첫째는 거친 해상도(coarse resolution, 4^2 - 8^2) 단계로, 이 레이어에서 조절하는 스타일은 성별, 포즈, 일반적인 헤어스타일, 얼굴 형태, 안경 착용여부 등이다. 둘째는 중간 해상도(middle resolution, 16^2 - 32^2)단계다. 여기서는 얼굴의 부분적 특성, 디테일한 헤어스타일, 입의 개폐여부 등을 조절한다. 셋째는 미세 해상도(fine resolution, 64^2 - 1024^2)단계의 레이어로 여기서는 눈동자 색깔, 머리카락 색깔, 조명에 따른 명암이나 얼굴색의 변화, 그밖에 얼굴의 미시적 특성을 조절한다.²³ 이처럼 최초 입력 값을 합성곱 레이어를 지날 때마다 새로운 스타일을 부여받게 되며 이 과정이 반복되면 매우 자연스러운 이미지가 산출된다. 이미지의 부분적인 특성을 조정하는 것이 아니라 전체적인 스타일을 조정하는 까닭에 결과물은 실제 인물처럼 보인다 [도 5]. 결국 StyleGAN은 가상의 이미지를 생성하는 알고리즘이지만 그 원리를 적용하여 우수한 이미지 합성 기술로도 활용할 수 있다. 여러 장의 사진으로부터 각 레이어마다 원하는 스타일을 선택하여 섞을 수 있기 때문이다. 실제 StyleGAN을 처음 개발했던 앤비디아는 연구를 진척시켜 포토샵을 능가하는 사진합성 기술로 이 알고리즘을 개량한 바 있다.

21. Tero Karra, Samuli Laine & Timo Aila, “A Style-Based Generator Architecture for Generative Adversarial Networks”, arXiv:1812.04948v3 (Mar. 29, 2019), p. 1.

22. StyleGAN이 지닌 또 다른 특징 중의 하나는 생성자의 구조에 집중하는 반면 기존 GAN의 판별자 구조를 바꾸지는 않았다는 점이다.

23. Karra et al., “A Style-Based Generator Architecture for Generative Adversarial Networks”, p. 4.

StyleGAN 이미지의 리얼리티는 스타일을 여러 단계에서 조정하는 데서만 나오는 것은 아니다. 이 알고리즘은 확률변수(Stochastic variation)를 활용하여 각 레이어마다 랜덤하게 노이즈를 추가함으로써 특정 개인들만 가질 수 있는 변칙적 요소를 끌어들인다. 예를 들면 “까칠한 수



[도 5] StyleGAN이 생성한 이미지의 예

염이나 주근깨, 피부모공처럼 확률적으로 여겨질 수 있는 특질들”을 이미지에 부여하는 것이다.²⁴ 논문의 저자들은 실험 결과 확률변수를 통해 입력된 노이즈는 이미지의 전체적인 지각에 영향을 미치지 않고 정확히 배치되면, 즉 엉뚱하게 추가되지 않는 한 리얼리티를 증폭시켜줄 수 있다고 주장한다. 예컨대 노이즈를 일부러 추가하지 않은 경우 이미지는 별 특징 없이 ‘그림 처럼’ 보이며, 초기 레이어(거친 해상도 단계)에 투입된 레이어는 헤어스타일을 과장되게 만들고, 후반 레이어(미세 해상도 단계)에 노이즈를 추가하면 머리카락의 디테일 변화를 유발한다는 것이다. 반대로 매 레이어마다 노이즈를 입력했을 때 이미지의 리얼리티가 가장 높았다.²⁵

Ⅲ. 기술 이미지: 숫자의 세계

1. 프로그램과 무작위성

앞에서 살펴보았듯이 GAN의 유형은 다양하고 각 알고리즘이 구체적으로 지향하는 목표는 상이하다. 그러나 근본 구조와 원리는 같다. 거시적 관점에서 보면 GAN 이미지는 컴퓨터 프로그램의 산물이다. 왜 프로그램을 활용하여 이미지를 생산할까? 프로그램은 제기된 문제에 대

24. Karra et al., “A Style-Based Generator Architecture for Generative Adversarial Networks”, p. 5.

25. Karra et al., “A Style-Based Generator Architecture for Generative Adversarial Networks”, p. 5.

한 해결책을 찾기 위해 기계의 연산에 의존하는 방법이다. 왜 기계의 힘을 빌려야 할까? 사람의 손으로 ‘제작한’ 이미지도 있고, 카메라를 활용하여 ‘자동으로’ 찍어낸 사진도 있는데 굳이 컴퓨터 프로그램에 의존하는 이유는 무엇일까? 연산기계의 프로그램 문제에 대해 로체스터(N. Rochester)는 근본적인 질문을 던진다. 이 제안은 인공지능에 대한 핵심 문제들을 제기하고 논의의 장을 처음 마련했던 <다트머스 컨퍼런스 Dartmouth Conference>에서 나온 것으로 주요 논점을 요약하면 다음과 같다.²⁶

인간은 특정한 문화 환경 속에서 끊임없이 문제에 부딪히면서 해답을 찾고자 한다. 그 과정에서 과거에 이미 해결됐던 문제들에 대한 해답은 데이터로 축적되어 있다. 만약 인간에게 새로운 문제가 주어졌을 때 이미 데이터 속에 해답이 있다면 우리는 쉽게 문제에 대한 해결책을 예측할 수 있다. 프로그램은 이처럼 어떤 사태에 대한 해답을 포함한 규칙들을 담고 있다. 즉 프로그램은 사전(pro)에 문자(gram)로 정리된 규칙들의 집합으로 유사한 문제가 주어졌을 때 이를 적용할 수 있다. 그런데 복잡한 사태와 만나면 해결 규칙도 복잡해진다. 이 경우 어떤 것이 올바른 해답인지 판단하기 위해 여러 가능한 경우를 테스트해야 한다. 사태가 더욱 복잡해지면 불확실성도 커지고 그에 따라 적용 규칙도 복잡해진다. 예컨대 틱택토 게임(Tic tac toe)에서 참가자는 자신의 결정뿐만 아니라 주변 환경도 함께 고려해야 한다. 즉 상대방이 내릴 수 있는 가능한 다양한 결정들도 예상해야 한다.²⁷ 나아가 부단한 노력에도 불구하고 어떠한 개인도 해답을 찾지 못한 문제가 있을 수 있다. 요컨대 바로 그 문제의 해결을 위한 규칙들을 담고 있는 프로그램(사전에 축적된)들이 있으나 규칙이 잘못된 경우다. 이 때 개인은 그 문제를 해결하기 위해 축적된 지식과 경험에 의거해서 다른 방법, 말하자면 규칙에 없는 방법을 적용해야 한다. 사전에 규정된 규칙과 동일한 방법으로는 해결할 수 없기에 그는 무작위로 예측 범위 밖에서 방법을 찾아야 한다. 따라서 무작위로 적용된 새로운 방법은 ‘전통적인’ 관점에서는 기이한 일탈이다. 만약 그에 대한 해답을 찾았다면 그 방법은 완전히 새로운 수밖에 없다. 기존 규칙의 목록에 없었던 것이기 때문이다. 이런 방법을 로체스터는 ‘몬테 카를로 방법(Monte Carlo

26. 인공지능이라는 용어가 처음 사용된 이 컨퍼런스의 공식명칭은 “인공지능에 대한 다트머스 여름 연구프로젝트(Dartmouth Summer Research Project on Artificial Intelligence)”로 뉴햄프셔의 다트머스 대학에서 개최됐다. 주요 발표자는 존 매카시(J. McCarthy), 클로드 섀넌(C. E. Shannon), 마빈 민스키(M. I. Minsky), 니대니얼 로체스터(N. Rochester)며, 컴퓨터 공학 관련 분야 전문가들이 참여했다. John McCarthy et al., “A Proposal for the Dartmouth Summer Research Project on Artificial Intelligence”, in: *AI Magazine*, vol. 27, no. 4 (August 31, 1955), pp. 12-14 (DOI: 10.1609/aimag.v27i4.1904).

27. McCarthy et al., “A Proposal for the Dartmouth Summer Research Project on Artificial Intelligence”, p. 7. 틱택토는 3x3 사각형 도형의 9개 칸 속에 참가자가 O나 X를 번갈아가며 채워 넣는 게임으로 연달아 3개의 O나 X를 채우면 이긴다.

Methode)’을 확대 적용한 것이라고 밝힌다.²⁸

왜 문제의 해답을 찾기 위해 무작위성에 의존할까? 첫째 이유는 앞에서 언급했듯이 사전에 축적되어 있지 않은 새로운 규칙을 찾아낼 필요성 때문이다. 무엇이 정답인지 알 수 없는 탓에 무수한 가능성들을 하나씩 대입시켜 볼 뿐이다. 모래밭에서 바늘 찾는 격이다. 둘째, 기계는 인간, 예컨대 프로그래머의 편견이나 선입견을 극복하고 무작위로 가능한 경우의 수를 차례로 대입시켜 나감으로써 근사치에 해당하는 답을 찾아낼 수 있다.²⁹ 기계의 연산 능력이 고도화되면 해답을 찾아내기 위한 데이터 처리 속도가 빨라져 임무를 수행할 수 있다는 것이다.

GAN은 기존의 알고리즘이 봉착했던 한계, 말하자면 딥러닝을 통해 무작위로 내놓는 결과물이 목표로 하는 근사치와 너무 멀다는 문제를 극복하기 위해 고안되었다. 이미지의 경우 다른 생성 알고리즘들의 산출물은 사진처럼 보이지 않았다. 이를 해결하기 위해 끌어들이는 규칙은 비지도학습과 생성자/판별자의 경쟁구도다. GAN의 원리는 그럴 듯 했으나 학습이 불안정해서 실제 ‘사진 같은’ 이미지의 생성은 쉽지 않아 이 문제를 해결하기 위해 개발된 알고리즘이 DCGAN이나 BEGAN같은 프로그램이다. 앞에서 살펴보았듯이 해결책은 수많은 레이어를 통해 ‘랜덤하게’ 학습을 여러 차례 진행하는 방법이다. 이미지의 해상도를 비약적으로 높여주는 과업을 달성한 모델은 SRGAN과 PGGAN으로, 여기서도 문제는 근사치를 통해 해결되었다. Pix2Pix나 CycleGAN은 유사 이미지를 치환하는 난제를 동일한 방법으로 해결했다. 나아가 StarGAN과 StyleGAN은 기존의 GAN으로는 풀 수 없었던 이미지 변형과 합성을 탁월하게 수행했다. 각각의 GAN이 해결한 과업들은 실상 과거에 제기되었던 질문들이다. 다만 과거에 축적되어 있던 규칙들만으로는 그 문제를 해결할 수 없었을 뿐이다. 해결책은 복합적이다. 기존 프로그램의 규칙들로부터 새로운 규칙들을 고안하고 그것을 랜덤하게 적용해 나감으로써 문제를 해결할 수 있었던 것이다.

28. 몬테 카를로 알고리즘은 난수(무작위로 추출된 수)를 이용해서 함수의 값을 확률적으로 계산하는 방법인데, 복잡하고 난이도가 높은 문제를 해결할 때 근사치를 찾아냄으로써 문제를 해결한다. 결정론과 달리 확률론의 입장에서 무작위로 샘플링을 통해 답을 얻어내는 방법이다. McCarthy et al., “A Proposal for the Dartmouth Summer Research Project on Artificial Intelligence”, p. 8.

29. 로체스터는 기계가 인간에 대해 갖는 비교우위를 다음과 같이 기술한다. 새로운 문제에 봉착했을 때 한 과학자가 평생 노력해도 해답을 찾기 어렵다면 많은 다른 과학자들의 노력이 함께 요구된다. 이 경우 그들은 중복 없이 랜덤하게 하나씩 새로운 방법을 적용할 수 있지만 대단히 미세한 소통이 필요하다. 또한 경쟁 관계에 있는 개인들은 경쟁에서 이기기 위해 복잡한 전략을 취하는 것 보다 무작위적인 방법을 선호할 수 있다. 그러나 기계에게 그런 복잡한 고려는 없으며, 근시안적 사고나 선입견도 없다. McCarthy et al., “A Proposal for the Dartmouth Summer Research Project on Artificial Intelligence”, p. 8.

2. 픽셀 지도

GAN의 원리를 적용하여 생성된 이미지의 지위와 성격을 어떻게 규정해야 할까? 여러 차례 언급했듯 GAN 이미지는 사진처럼 보인다. GAN의 본래 목표가 원본 데이터와 ‘무한히’ 유사한 가짜 데이터를 만들어내는 데 있기 때문이다. 게다가 학습용 원본 데이터는 무작위로 제공된 사진, 그것도 수만 점 이상의 ‘빅데이터’다. 그런데 실제 사진으로부터 생성자가 산출한 이미지는 픽셀 값을 계산하여 수정된 ‘데이터’다. 이 데이터는 GAN의 핵심 원리인 ‘손실’에 바탕을 둔다. 즉 원본 사진의 픽셀 값과 편차가 있다. 한 마디로 원본 사진과 다르다는 것이다. 무엇이, 어떻게 다를까?

사진을 포함한 모든 컴퓨터 이미지는 파일로 저장된다. 파일을 구성하는 것은 픽셀로 기본 단위는 비트맵(Bitmap)이다. 즉 한 장의 이미지는 픽셀들로 구성된 일종의 ‘지도(map)’다. GAN의 학습에 주로 사용됐던 사진들은 비트맵의 확장 파일인 JPG나 GIF, PNG 등으로 모두 픽셀 데이터로 이루어져 있다. 픽셀은 그저 하나의 격자다. 모눈종이 위의 수많은 빈칸들인 셈이다. 이 격자는 홀로 존재할 수 없으며 아무런 정보도 갖고 있지 않은 ‘빈 방’과도 같다. 픽셀은 촬영 시 디지털 카메라의 광센서(CCD, CMOS)가 받아들인 빛을 숫자 정보로 저장하고 서브픽셀이라 부르는 RGB값을 갖게 되면서 이미지 정보를 부여받는다. 흑백을 포함한 모든 색상 정보는 이 숫자로 환산된 색상의 조합을 통해 결정된다. 그리고 이렇게 ‘의미 있는’ 격자로 변환된 픽셀들이 모여 형성된 ‘지도’가 이미지 파일이다. 이 때 이미지 전체의 데이터는 모든 픽셀 정보의 총합으로 구성된다. 그리고 비트맵을 기본 단위로 한 사진 이미지의 수정, 변형은 픽셀 값의 변화로 가능하다.³⁰⁾ 이렇게 픽셀로 구성된 디지털 이미지를 임의로 ‘픽셀 지도’라 부른다. 이 ‘지도’를 구성하는 각각의 픽셀들은 자신의 고유한 좌표 값을 갖고 있으며, 이후 새로운 색상 값을 부여받으며 형태와 색의 변화를 통해 이미지 전체를 변형시킨다.

그렇다면 GAN이 만든 가상 이미지와 디지털 합성이나 수정으로 ‘변형된’ 사진, 요컨대 ‘포스트 포토그래피’의 범주에 속하는 가상 이미지 사이에는 어떤 차이가 있을까? 디지털 합성 또한 결과물은 허구와 가상이나 그 재료는 사진이다. 예컨대 포토샵을 활용하여 두 장 이상의 사진을 합성할 경우 만약 이 레이어가 사진으로 구성된다면 합성된 사진에는 각각의 원본이 지닌 픽셀 지도가 그대로 이동한다. 이미지를 구성하는 기본 단위인 픽셀의 구조는 변하지 않는다는 뜻이다. 사진의 형태를 변형시키는 경우는 어떨까? 포토샵으로 사람의 얼굴 형태를 길게 늘이면 형태가 변한다. 이 때 바뀌는 것은 형태 변화를 야기한 선에 해당하는 픽셀 값이다. 선을 구성하는 각각의 픽셀은 새로운 색상 값을 부여받을 뿐이다. 예를 들면 배경의 면을 구성하고 있

30. James R. Parker, *Generative art: algorithms as artistic tool* (Calgary: Alberta 2020), pp. 120–121.

던 픽셀이 얼굴 피부의 픽셀 값으로 바뀌는 것이다. 이 때 픽셀의 위치는 바뀌지 않는다. 즉 픽셀은 ‘이동’하지 않는다. 각각의 픽셀은 픽셀 지도에서 고유한 위치를 점유하고 있으며 형태 변화 시 색상 값은 바뀌나 좌표 값은 그대로다. 이 점이 가리키는 바는 명확하다. 즉 디지털 카메라가 촬영 시 ‘결정했던’ 픽셀 지도는 사진의 수정 과정에서도 그대로 유지된다는 것이다. 결국 바뀌는 것은 각 픽셀의 색상 값뿐이며, 한 장의 이미지를 구성하는 픽셀 지도는 동일하다.

이런 관점에서 보면 디지털 합성 및 수정을 통해 얻어낸 이미지는 여전히 ‘전통적인’ 사진의 생성원리에서 크게 벗어나지 않는다고 할 수 있다. 촬영 시 광센서에 저장된 픽셀 지도는 변하지 않기 때문이다. 즉 이미지를 구성하는 모든 픽셀의 위치와 배열은 고정이며, 수정 시 바뀌는 것은 색상 값뿐이다. 따라서 디지털 변형은 ‘시각적으로는’ 형태 변화처럼 보이지만 실제로는 각각의 픽셀에 색을 입힌 결과다. 이런 관점은 디지털 카메라가 이미지를 숫자정보로 저장한다 할지라도 생성원리는 근본적으로 지표주의자들의 주장처럼 빛과 감광판의 물리적 접촉이라는 특수한 관계에서 비롯된다는 생각과 맞닿아 있다. 즉 디지털 사진도 ‘지표성’을 지니며 바로 그 때문에 증명사진을 비롯한 각종 기록과 입증의 수단으로 쓰일 수 있다는 것이다.³¹ 물론 형태 변형이 심할 경우 이 규범은 적용될 수 없다. 그러나 이미지의 생성원리와 구조는 다르지 않다.

GAN이 개발되기 전에도 알고리즘으로 픽셀 값을 변화시켜 이미지를 변형시키는 방법은 있었다. 사진에 회화적 효과를 부여하기 위해 의도적으로 흐린 이미지를 만드는 것이 그 예다. 방법은 하나의 픽셀을 둘러싼 주변 픽셀들이 형성하고 있는 구역을 설정하여 새로운 픽셀 값을 계산하는 모델이다.³² 원리는 간단하다. 즉 새로운 픽셀 값을 얻기 위해 인접한 픽셀들을 활용하고 원래 픽셀을 이 새로운 값으로 대체하는 것이다. 문제는 어떤 인접 픽셀을 사용하여 어떻게 계산할 것인지에 있다. 개별 픽셀은 하나의 이미지 안에서 고유한 좌표 값을 갖고 있다. 이미지의 일부를 활용할 때 하나의 픽셀은 좌표 속에서 8개(정방향)나 4개(상하좌우)의 인접 픽셀을 가진다. 새로운 픽셀 값을 찾아 이미지를 변형시키고자 할 때 8개나 4개로 구성된 인접 픽셀들의 평균치를 구한다. 그렇게 계산된 평균값이 각각의 픽셀에 부여되면 이미지의 변형은 완료된다. 이 때 부여된 새로운 픽셀 값은 색상 값(예컨대 RGB값)이며 픽셀의 크기는 바뀌지 않는다.³³ 이런 방식으로 변형된 이미지는 비록 픽셀 값은 다를지라도 포토샵을 통해 수정된 사진의 경우처럼 원본 사진과 동일한 픽셀 지도를 유지한다. ‘시각적으로는’ 그림처럼 보인다 할지라도 디지털 카메라가 ‘결정했던’ 원래의 픽셀 구조가 이 ‘그림 같은’ 사진의 바탕을 이루는 셈이다.

31. Tom Gunning, “La retouche numérique à l’index”, in: *Etudes photographiques*, no. 19 (2006), pp. 96–119, p. 2.

32. Parker, *Generative art*, p. 138.

33. Parker, *Generative art*, p. 140.

GAN의 원리에 따라 생성된 이미지의 경우 사정은 좀 더 복잡하다. 우선 이 경우도 이미지의 구성 요소가 일반적인 디지털 사진처럼 비트맵 구조, 즉 픽셀이라는 점에서 같다. 그런데 생성자와 판별자의 경쟁을 통해 형성된 새로운 이미지에는 원본의 형태가 거의 남아있지 않다. 이 판단은 인간 시지각의 차원에서 그렇다. 그렇다면 기계의 차원, 즉 픽셀의 관점에서는 어떨까?

GAN의 생성자가 새로운 이미지를 만들고자 할 때 손실함수는 원본과 비슷한 특성들을 학습용 사진의 픽셀들로부터 계산해 내야 한다. 어떤 픽셀들을 선택할 것인가? 샘플링이 필요하다. 요컨대 원본 사진의 특성을 지녔을 법한 픽셀들을 확률에 따라 무작위로 선택하여 하나씩 대입해 나가는 것이다. 이 때 하나의 픽셀은 아무 의미가 없다. 예컨대 64x64 사진의 픽셀 총수는 4,096인데 어떤 픽셀 하나가 이미지의 특성을 규정하는 지 알 수 없다. 게다가 하나의 픽셀은 다른 픽셀과 분리되어 별개로 존재할 수 없다. 따라서 샘플링은 이미지의 일부를 대상으로 픽셀 값을 계산하고 거기에 손실함수를 대입하여 편차를 만들어 낸다. 이때도 역시 바뀌는 것은 픽셀의 색상 값이다. 따라서 학습용 원본 사진의 픽셀 값은 생성자가 첫 손실을 만들어내는 순간부터 바뀐다. 그리고 그 과정이 ‘무수히’ 진행됨에 따라 원본의 픽셀 값은 ‘심각히’ 변형되어 거의 남아있지 않을 정도다. GAN 알고리즘 개발자들이 누차 지적하듯 생성자는 ‘블랙박스’처럼 작동하는 탓에 그 과정이 어떻게 이루어지는지는 누구도 알 수 없다. 다만 이 경우도 학습용 사진을 구성하는 각 픽셀의 기본 구조는 바뀌지 않는다는 점이 중요하다. 비트맵은 고정된 구조며, 픽셀의 RGB 값이 바뀌더라도 픽셀 지도는 그대로라는 뜻이다. 따라서 수만 번 새로운 픽셀 값이 부여됐더라도 최종 결과물로 산출된 이미지의 픽셀 지도는 같다. 이는 원본 사진을 찍었던 카메라의 ‘결정’이 변형된 이미지의 픽셀 구조, 말하자면 픽셀 지도를 여전히 지배하고 있음을 의미한다.

예를 들어 StyleGAN의 ‘거친 해상도’ 레이어에서 생성자는 인물의 전체 얼굴 형태와 머리카락의 특성을 지닌 픽셀들을 샘플링하여 그 값을 계산하고 근사치에 가까운 가짜 데이터를 만든다. 즉 새로운 RGB 픽셀 값을 부여한다. 이 값이 부여된 픽셀들의 좌표 값은 원본 사진의 픽셀 지도와 같다. 판별자가 이 픽셀 지도를 진짜라고 판단하면 다음 레이어, 즉 중간 해상도 레이어로 넘어가 같은 방식으로 새로운 RGB 값을 부여한다. 여기서도 전 단계 레이어의 픽셀 지도는 유지된다. 마지막 미세 해상도 단계에서도 과정과 결과는 같다.

이런 구조는 골턴(F. Galton)이 고안한 합성초상(Composite portrait)과의 비교를 통해 좀 더 명확해질 수 있다. 합성초상은 개별자로 구성된 유형(Type)의 시각적 평균을 추출하기 위해 고안된 방법으로 여러 장의 초상사진을 합성하여 한 장의 초상으로 만든 것이다. 예컨대 8명의 얼굴 사진을 합성하고자 할 때 8장 전체의 노출 값을 먼저 구한다. 그 값이 80초라면 각각의 사

진에 10초씩의 노출시간을 준다. 이 때 여러 번 반복해서 빛에 노출될수록 이미지는 또렷해지고 그렇지 않은 부분은 희미하게 남거나 시야에서 사라진다. 결국 개별 인물의 특성은 사라지고 8명의 인물이 공유하는 얼굴의 구성요소만이 최종 이미지로 형성되는 셈이다.³⁴ 이 결과는 StyleGAN에서 노이즈, 말하자면 개별인물에게 고유한 확률적 속성(주근깨나 점 등)을 의도적으로 추가하지 않았을 때 아무 특징 없는 평범한 이미지가 산출되는 것과도 유사하다.

이 방법을 통해 얻어낸 사진을 골틴은 ‘평균 이미지’라 규정했다. 합성초상이 보여주는 인물의 모습은 분명 ‘가상’의 이미지다. 그 사진이 가리키는 인물은 현실에 존재하지 않기 때문이다. 마치 GP가 생성한 가상의 인물처럼 말이다. 그런데 이 문제를 ‘입자’의 관점에서 보면 사정은 달라진다. 한 장의 얼굴 사진을 구성하고 있던 입자의 배열은 10초의 노출시간 동안 목적지인 다른 감광판으로 이동했다. 이 때 입자의 배치에는 변화가 없다. 시각적으로 명확히 드러나지 않더라도 말이다. 8장의 사진에 배열돼 있던 다른 입자들도 마찬가지다. 이는 당초 원본에 ‘지도’로 배치돼 있던 입자들의 구조가 바뀌지 않았다는 의미다. 만약 어떤 입자가 희미하게 보인다면, 즉 연한 회색이어서 흰 바탕과 구분이 어려울 정도라면 이는 비트맵 구조에서 픽셀의 RGB 값을 연회색으로 바꾼 것과 다를 바 없다. 결국 GAN 이미지에서 픽셀의 배열이 달라지지 않는 것처럼 골틴의 합성초상에서도 입자의 배열은 그대로 유지된다. 요컨대 입자의 구조와 픽셀 지도, 즉 비트맵의 구조는 같다.

위의 내용을 요약, 정리하면 다음과 같다. ‘전통적인’ 사진이론은 생성원리의 관점에서 지표이론을 적용하여 사진의 ‘특수성’을 규명하고자 했다. 그런데 이 ‘배타적’ 이론은 이후 디지털 합성과 수정을 통해 생산된 이른바 ‘포스트 포토그래피’를 포섭할 수 없었다. GAN 알고리즘이 원본 사진을 학습하여 생산한 이미지에 이르면 지표이론의 폐쇄성은 더욱 분명해진다. 그럼에도 아날로그 사진과 디지털 사진, 알고리즘 이미지의 연결고리는 분명하다. 즉 이미지의 생성을 결정했던 빛 입자가 출발이라는 것이다. 디지털 수정과 GAN의 계산은 당초 결정된 픽셀 지도를 바탕으로 진행된다. 비록 이 ‘격자’ 위에 새로운 형태를 ‘그리더라도’ 본래의 픽셀 구조는 같다. 이 점이 세 유형의 이미지, 즉 아날로그 사진과 디지털 사진, 알고리즘 이미지 모두 지표의 속성을 공유함을 뜻하지는 않는다. GAN의 ‘지시대상’은 어디에도 없기 때문이다. 따라서 알고리즘 이미지의 ‘지위’를 규정할 새로운 관점이 요청된다.

34. 박상우, 『인간 유형의 발명: 골틴의 합성사진』, 『박상우의 포토폴로지: 베르티옹에서 마레까지 19세기 과학사진사』(문학동네 2019), pp. 150-151.

3. 확률게임과 블랙박스

기술 이미지 개념을 처음 제안한 플루서는 그것을 손으로 만든 전통 이미지와 구분하면서 장치 프로그램의 산물로 규정했다. 그에 따르면 카메라는 최초의 장치고, 사진은 최초의 기술 이미지다. 한편 알고리즘은 입력 값으로부터 일정한 결과 값을 얻어내기 위한 규칙들의 집합, 요컨대 프로그램의 일종이다. 이 관점을 도식화 시키면 알고리즘에 따라 생산된 GAN 이미지 또한 기술 이미지의 범주에 속한다고 할 수 있다. 그렇다면 기술 이미지로서의 사진과 GAN 알고리즘 이미지는 어떤 속성을 공유할까? 혹은 어떤 차별성을 가질까?

가상 이미지를 생산하는 아날로그 사진과 디지털 사진, GAN 알고리즘 이미지를 비교해보자. 골턴의 아날로그 합성초상에서 최종 이미지의 생산을 주도하는 자는 여전히 인간이다. 먼저 ‘사람이’ 8장의 사진 크기를 동일하게 조절해야 한다. 다음은 얼굴의 구성요소인 눈, 코, 입이 정확히 겹치도록 수평축은 동공에, 수직축은 사진의 중앙에 맞추고 노출 시간에 따라 8번의 촬영을 반복해야 한다.³⁵ 요컨대 각각의 원본 사진은 바쟁의 표현처럼 ‘자동생성’ 이미지일지라도 합성사진을 얻어내기 위해서는 여전히 인간의 ‘손’에 의존해야 한다는 뜻이다. 포토샵으로 디지털 사진을 수정할 때도 수작업은 필수적이다. 디지털 합성이나 수정을 위해서는 픽셀 값을 ‘수동으로’ 조절해야 하고, 형태를 바꾸려면 ‘마우스’를 정교히 움직여야 한다. 여기에서 사람의 노동은 한결 줄어들었고 형태의 변화를 위한 픽셀 값의 조절은 특정 알고리즘에 따라 ‘자동으로’ 이루어진다. 수동과 자동이 교차하는 셈이다. GAN 알고리즘에서 자동성은 한층 강화된다. 이미지는 알고리즘에 따라 자동 생성되기 때문이다. 즉 원본 사진들은 프로그램의 ‘의도대로’ 변형되어 인간의 선택을 기다릴 뿐이다. 이 ‘사진’들의 생산에 인간이 ‘기여’ 한 것은 거의 없다. 물론 알고리즘을 ‘기획’한 프로그래머가 있다. 그러나 그는 이미지의 생산에 직접 관여한 바가 없다. 그리고 어떤 이미지가 산출될지 전혀 예측할 수 없다. StyleGAN은 앞에서 언급했듯 이미지 합성을 ‘통제’하기 위해 고안됐다. 그래서 각각의 단계, 즉 거친 해상도, 중간 해상도, 미세 해상도 레이어에서 학습용 사진의 스타일을 입힐 수 있는 새로운 구조를 추가했다. 그러나 여기서도 인간의 개입은 최소화되어 있으며 실제 이미지 합성을 주도하는 것은 합수다. 말하자면 ‘정확히’ 어떤 이미지가 산출될지 누구도 알 수 없다는 것이다. 결국 GAN 이미지의 생성과정은 확률게임인 셈이다.

사진이 최초의 기술 이미지라는 플루서의 규정은 사진과 더불어 이미지 생산의 ‘자동화’가 시작됐다는 뜻을 포함하고 있다. 그리고 ‘자동성’의 초기 단계에서 인간의 개입 여지, 즉 ‘수동

35. 골턴이 합성사진을 제작하기 위해 고안한 ‘기술적’ 방법에 대해서는 박상우의 글에 상세히 기술되어 있다. 박상우, 『인간 유형의 발명』 pp. 148-150.

성'은 부분적으로 자동성과 섞여 있었다. 이 때 사진은 '자동생성'의 원리를 따르지만 실제 그 '자동성'에는 부분적으로 인간의 개입과 통제가 수반된다는 것이다. 프로그램이 고도화되면서 수동성은 빠르게 줄어들고 자동성은 점차 강화된다. 이제 GAN 알고리즘의 발전과 더불어 이 '자동성'은 한층 고도화됐으며 인간이 개입할 수 있는 여지는 더욱 줄어들었다. 인간은 '확률게임'에 따라 자동으로 산출된 이미지를 선택할 따름이다. GP의 경우 수요자는 몇 가지 파라미터를 통해 이미지 선택에 개입할 수 있다. 성별과 나이, 머리카락 색깔 등이 그것이다. 그러나 그 파라미터는 실제 이미지 생산에 영향을 주는 것이 아니라 알고리즘의 제안 범위를 좁히기 위한 종속변수일 뿐이다. 그리고 우리는 그 알고리즘이 왜 그 사진을 제안하는지 알 수 없다.

결국 GAN의 생성자는 '블랙박스'며, 그 안에서 어떤 일이 벌어지는지는 베일에 싸여있다. 이는 비록 사진 촬영 과정에 사진가가 '부분적으로' 개입한다 할지라도 이미지의 생성 과정에 대해 전혀 알지 못한다는 점에서 카메라가 '블랙박스'라는 플루서의 주장과 일맥상통한다.³⁶ 문제는 GAN의 경우 심지어 이미지 생산을 '기획'하고 그 프로그램을 설계했던 '조물주'에게조차 생성자가 블랙박스라는 점이다. 기술 이미지의 본질은 프로그램의 자동성에 있다. 그런 점에서 GAN은 진화한 형태의 기술 이미지다. 그런데 자동성이 고도화되면서 인간은 자신이 향유하고 소비할 이미지의 생산과정으로부터 추방당했으며, 생성자의 불투명성 또한 함께 강화됐다. 이제 인간은 영문도 모른 채 '블랙박스'에서 쏟아져 나오는 이미지의 소비자로 전락했다. 이미지의 생산력은 향상됐으나 생산 과정으로부터 스스로를 축출한 셈이다. 그것이 기술 이미지의 진화가 야기한 명과 암이다.

IV. 결론

알고리즘이 이미지 생산에 개입하면서 몰고 온 변화는 크게 두 축을 중심으로 전개되고 있다. 특히 GAN 알고리즘은 이미지의 기계학습에 최적화 된 프로그램으로 발전 속도는 매우 빠르고 종류도 다양하여 이 변화를 주도적으로 이끌고 있다. 첫째는 생성 이미지의 지위와 성격이다. GAN은 원본 사진에 대한 기계학습을 통해 스스로 가상 이미지를 생성한다. 그 이미지는 가상이지만 사진처럼 보여 실제로 착각할 만큼 진짜에 가까운 가짜다. 요컨대 '유사'의 차원에서 사진과 다를 바 없으나 이미지의 지시 대상이 존재하지 않는다는 점에서 '지표'의 속성을 지닌 사진과는 명확히 갈린다. 물론 최종 이미지의 재료는 사진이었다. 그리고 그 이미지를 구성

36. Vilém Flusser, *Towards a philosophy of photography* (London: Reaktion Books Ltd, 1983), p. 27.

하는 ‘원래’ 픽셀들은 지표의 자격을 갖고 있었다. 그러나 픽셀 값 변환을 통해 산출된 최종 이미지에 그 흔적은 남아있지 않다. 그럼에도 불구하고 그 이미지가 사진과 무관하다고 할 수 없다. 원본 사진에 새로 ‘그린’ 픽셀 지도의 원형은 그대로이기 때문이다. 따라서 ‘고전적인’ 사진 이론의 관점, 특히 ‘지표이론’으로 알고리즘 이미지의 특성을 설명할 수 없다. 디지털 사진도 생성원리는 지표이론을 따르지만 그 이후는 전혀 다른 문제다. 픽셀 지도 위에 무한히 새로운 이미지를 만들어 나갈 수 있기 때문이다.

두 번째는 프로그램과 인간의 관계에 대한 문제다. GAN은 기술 이미지의 고도화가 어떤 방향으로 진행될 지를 알려주는 징후다. 이는 이미 플루서의 진단을 통해 부분적으로 예견됐던 바이기도 하다. 기술 이미지가 프로그램의 산물이고 그 본질이 자동성에 있다면 인간 배제는 필연적이다. 기술 이미지의 ‘기획’ 단계에 인간 배제는 내재되어 있었던 셈이다. GAN 알고리즘은 빅데이터와 딥러닝, 학습의 안정성을 통해 이런 결과를 가속화하고 있다. 한편 배제와 더불어 혹은 그 결과 인간은 자신이 구축한 이미지 생산 메커니즘에 대해 무지한 상태다. 이미지를 생산하는 인공지능은 인간에게 블랙박스며, 심지어 프로그램의 설계자에게조차 그렇다. 이 블랙박스가 해체될 가능성은 거의 없어 보인다. 기술 이미지의 고도화는 이 구조를 유지한 채 진행될 것이며 결국 인간은 이미지 생산을 통제하지 못하는 상황을 맞게 될 가능성이 높다. 소위 ‘딥페이크’의 위험성에 대한 경고가 수많은 가능성들 중의 하나다.

새롭게 등장한 문제에 대한 해답은 본문에서 언급한 로체스터의 제안처럼 프로그램, 즉 사전에 주어진 규칙들 속에 있을 수 있다. 다만 우리는 어떤 규칙이 문제 해결에 적합한지 알지 못할 뿐이다. 예를 들어 알고리즘이 블랙박스라면, 그리고 인간이 그 내부를 들여다 볼 수 없다면 대항 알고리즘을 만드는 것이 대안 중의 하나가 될 수 있다.

논문투고일: 2020년 12월 10일

심사기간: 2020년 12월 16일-2021년 1월 8일

최종게재확정일: 2021년 1월 9일

참고문헌

- 박상우, 「인간 유형의 발명: 골턴의 합성사진」, 『박상우의 포토로지: 베르티옹에서 마레까지 19세기 과학사진사』, 문학동네 2019.
- Bazin, André, "Ontologie de l'image photographique", *Qu'est-ce que le cinéma?*, Paris: Cerf 1945.
- Berthelot, David, Thomas Schumm & Luke Metz, "BEGAN: Boundary Equilibrium Generative Adversarial Networks", arXiv:1703.10717v4, May 31, 2017.
- Choi, Yunjeon et al., "StarGAN: Unified Generative Adversarial Networks for Multi-Domain Image-to-Image Translation", arXiv:1711.09020v3, Sep. 21, 2018.
- Flusser, Vilém, *Towards a philosophy of photography*, London: Reaktion Books Ltd 1983.
- Goodfellow, Ian J. et al., "Generative Adversarial Nets", arXiv:1406.266.1v1, Jun. 10, 2014.
- Gunning, Tom, "La retouche numérique à l'index", in: *Etudes photographiques*, no. 19, 2006, pp. 96-119.
- Isola, Phillip et al., "Image-to-Image Translation with Conditional Adversarial Networks", arXiv:1611.07004v3, Nov. 26, 2018.
- Karras, Tero et al., "Progressive Growing of GANs for improved quality, stability, and variation", arXiv:1710.10196v3, Feb. 26, 2018.
- Karras, Tero, Samuli Laine & Timo Aila, "A Style-Based Generator Architecture for Generative Adversarial Networks", arXiv:1812.04948v3, Mar. 29, 2019.
- Ledig, Christian et al., "Photo-Realistic single image Super-Resolution using a Generative Adversarial Network", arXiv:1609.04802v5, May 25, 2017.
- McCarthy, John et al., "A Proposal for the Dartmouth Summer Research Project on Artificial Intelligence", in: *AI Magazine*, vol. 27, no. 4, August 31, 1955, pp. 12-14 (DOI: 10.1609/aimag.v27i4.1904).
- Parker, James R., *Generative art: algorithms as artistic tool*, Calgary: Alberta 2020.
- Radford, Alec & Luke Metz, "Unsupervised Representation Learning with Deep Convolutional Generative Adversarial Networks", arXiv:1511.06434v2, Jan. 7, 2016.
- Zhu, Jun-Yan et al., "Unpaired Image-to-Image Translation using Cycle-Consistent Adversarial Networks", arXiv:1703.10593v6, Nov. 15, 2018.
- <https://generated.photos/>

도판목록

[도 1] PGGAN의 학습 구조와 생성된 고해상도 이미지

[도 2] Pix2Pix가 수행한 과업의 예

[도 3] CycleGAN이 수행한 이미지 생성의 예

[도 4] StarGAN이 수행한 생성 이미지의 예

[도 5] StyleGAN이 생성한 이미지의 예