

일반논문 (Regular Paper)

방송공학회논문지 제27권 제3호, 2022년 5월 (JBE Vol.27, No.3, May 2022)

<https://doi.org/10.5909/JBE.2022.27.3.437>

ISSN 2287-9137 (Online) ISSN 1226-7953 (Print)

## 딥러닝 기반 단일 이미지 생성적 적대 신경망 기법 비교 분석

정 성 훈<sup>a)</sup>, 공 경 보<sup>b)†</sup>

### Deep Learning-based Single Image Generative Adversarial Network: Performance Comparison and Trends

Seong-Hun Jeong<sup>a)</sup> and Kyeongbo Kong<sup>b)†</sup>

#### 요 약

생성적 적대 신경망(GAN, Generative Adversarial Networks)은 이미지 생성 분야에서 주목할 만한 발전을 이루었다. 하지만 큰 데이터 셋에서 불안정한 모습을 보인다는 한계 때문에 다양한 응용 분야에 쉽게 적용하기 어렵다. 단일 이미지 생성적 적대 신경망은 한 장의 이미지의 내부 분포를 잘 학습하여 다양한 영상을 생성하는 분야이다. 큰 데이터셋이 아닌 단 한장만 학습함으로써 안정적인 학습이 가능하며 이미지 리타겟팅, 이미지 조작, super resolution 등 다양한 분야에 활용 가능하다. 본 논문에서는 SinGAN, ConSinGAN, InGAN, DeepSIM, 그리고 One-Shot GAN 총 다섯 개의 단일 이미지 생성적 적대 신경망을 살펴본다. 우리는 각각의 단일 이미지 생성적 적대 신경망 모델들의 성능을 비교하고 장단점을 분석한다.

#### Abstract

Generative adversarial networks(GANs) have demonstrated remarkable success in image synthesis. However, since GANs show instability in the training stage on large datasets, it is difficult to apply to various application fields. A single image GAN is a field that generates various images by learning the internal distribution of a single image. In this paper, we investigate five Single Image GAN: SinGAN, ConSinGAN, InGAN, DeepSIM, and One-Shot GAN. We compare the performance of each model and analyze the pros and cons of a single image GAN.

Keyword: Generative adversarial network, single image, image manipulation, image retargeting

a) 부경대학교 신문방송학과(Department of Media Communication, Pukyong National University)

b) 부경대학교 미디어커뮤니케이션학부 휴먼ICT융합전공(Division of Media School, Pukyong National University)

† Corresponding Author : 공경보(Kyeongbo Kong)

E-mail: kbkong@pknu.ac.kr

Tel: +82-51-629-4072

ORCID: <https://orcid.org/0000-0002-1135-7502>

※ 이 논문은 부경대학교 자율창의학술연구비(2021년)에 의하여 연구되었음.

※ This work was supported by a Research Grant of Pukyong National University(2021).

· Manuscript April 18, 2022; Revised May 12, 2022; Accepted May 12, 2022.

## 1. 서론

2014년도에 Ian Goodfellow가 발표한 생성적 적대 신경망(GAN, Generative Adversarial Network)<sup>[1]</sup>은 지도 학습 중심의 딥러닝 체계를 비지도 학습으로 변화시킴으로써 인공지능 분야에 큰 영향을 미쳤다. 기존의 지도 학습은 정답이 있는 데이터 셋을 필요로 했기에, 학습하기 위해 데이터 셋을 만드는 데에 비효율적으로 많은 비용과 시간을 투자해야 했다. 하지만 생성적 적대 신경망은 비지도 학습이기 때문에 많은 데이터 셋을 필요로 하지 않고, 이전의 연구들과 비교하여 월등히 향상된 성능을 보였다.

생성적 적대 신경망에 대한 연구가 활발히 이루어지고 있지만, 매우 다양한 이미지를 가진 데이터 셋을 GAN으로 학습할 때, 전체 데이터셋 분포를 학습하지 못하고 일부만 학습하는 모드 붕괴(Mode Collapse) 현상이 종종 발생하였다. 따라서 많은 연구자들이 생성적 적대 신경망의 성능을 개선시키고 모드 붕괴 현상을 극복하기 위해 많은 연구를 진행했다.

GAN 모델의 올바른 학습과 더 나은 성능을 위해 많은 연구가 현재까지도 이어지고 있다. 개선된 GAN 모델들을 구별하고 분류하고자 할 때, 크게 '구조 변형', '손실 변형' 두 가지 접근법으로 나눌 수 있다. 구조 변형에서는 '네트워크 구조', '잠재 공간(latent space), '애플리케이션' 세 가지 요소로 나눌 수 있다. Progressive GAN(PROGAN)<sup>[2]</sup>은 기존 GAN의 네트워크 진행 구조에 대한 개선을 통해 모델의 성능을 이끌어 내었다. 먼저 낮은 해상도의 이미지로 학습을 진행하고, 점차 더 높은 해상도의 이미지를 학습하는 방식으로 기존 GAN 모델의 성능을 개선했다. 잠재 공간에 대해 변화를 준 InfoGAN<sup>[3]</sup>은 input noise에 잠재 변수(latent variable)를 추가하여 output과의 상호정보량이 커지도록 비지도 방식으로 학습함으로써 생성된 이미지와 지도 학습으로 생성된 결과물을 비교했을 때, 거의 유사한 이미지를 출력하는 성능을 보여준다. 애플리케이션 요소는 GAN이 단순히 새로운 이미지만을 생성하는 것이 아닌 다양한 기능을 할 수 있도록 네트워크 구조를 변형시키는 접근법으로, CycleGAN<sup>[4]</sup>은 이러한 애플리케이션 요소를 네트워크 구조에 적용하여 원본 이미지를 유명한 화가의 화풍으로 바꾸거나, 이미지 내부의 특정 object를 다른 스타일로 쉽고 비교

적 정확하게 변경할 수 있다.

또한 GAN을 이용하여 손상된 이미지에 대해 훼손된 영역을 추측하여 사실적인 이미지로 복원하고 질감 또한 복원할 수 있는 이미지 복원 기능<sup>[5]</sup>, 깊이 카메라로 이미지를 추출할 때 GAN의 학습 과정을 이용하여 손실된 부분을 복원하는 기능<sup>[6]</sup>, 주어진 이미지에 대한 맥락을 고려하여 이미지의 외부를 채우는 이미지 아웃페인팅 기능<sup>[7]</sup>을 하는 애플리케이션이 있다.

손실 변형에서는 '손실'과 '정규화' 두 가지 요소로 나눌 수 있는데, 손실 유형은 GAN 모델에 최적화될 수 있는 다양한 손실 함수(loss function)를 적용하는 것이다. Wasserstein GAN<sup>[8]</sup>은 output에 대한 예측 확률값이 아닌 확률 분포 간의 거리를 측정하는 방식인 wasserstein distance을 사용하는 비평가(Critic)를 판별자(Discriminator) 대신 사용하여, 모델이 최적점까지 학습할 수 있도록 한다. 정규화 유형은 손실 함수 또는 네트워크에 대해 수행되는 모든 유형의 정규화 작업에 설계된 추가적인 페널티를 의미한다. SN-GAN<sup>[9]</sup>은 보다 안정적인 방식으로 판별자(discriminator)를 학습시키기 위해 가중치 정규화(weight normalization)을 사용하고 이는, 모델이 학습하는데에 부담을 줄여주고 기존의 GAN에 쉽게 적용할 수 있다.

이러한 GAN의 성능을 발전시킨 모델 중, 학습의 안정성을 강화하고 더 나은 성능을 보이기 위해 대량의 데이터가 아닌 오직 하나의 원본 이미지만을 필요로 하는 단일 이미지 생성 모델의 연구 분야가 최근 대두되고 있다. 단일 이미지 생성 모델은 하나의 영상을 patch 단위로 분석하여 내부 분포를 학습하는 방법으로 불안정한 GAN 모델을 안정적으로 학습하여 다양한 애플리케이션에 활용할 수 있다는 장점이 있다. 대표적으로 생성적 적대 신경망의 모델을 개선하고, 단 하나의 원본 이미지만을 필요로 하는 SinGAN<sup>[10]</sup>, 그 SinGAN의 모델을 개선하고 성능을 향상시킨 ConSinGAN<sup>[11]</sup>, 원본 이미지의 크기나 형태, 비율을 변경할 때 이미지 내부 분포가 자연스럽게 변형되는 InGAN<sup>[12]</sup>, 특정 모델을 활용하여 학습에 필요한 데이터를 효율적으로 증식시켜 뛰어난 성능을 보인 DeepSIM<sup>[13]</sup>, 그리고 이미지 뿐만 아니라 비디오를 학습하는 One-Shot GAN<sup>[14]</sup> 등이 이에 해당한다.

본 논문에서는 많은 분야에서 활발히 사용되고 있는 “딥

러닝 기반 단일 이미지 생성적 적대 신경망 기법”의 현황과 동향을 살펴보고자 한다. II장에서는 딥러닝 기반 단일 이미지 생성적 적대 신경망 기법에 대해 자세히 알아보고자 한다. III장에서는 다양한 이미지들에 대한 각 모델들의 학습 결과를 정성적으로 비교 분석하고, 분석 결과를 통해 각 모델들의 장단점을 알아볼 것이다. IV장에서는 분석한 결과를 바탕으로, 단일 이미지 생성적 적대 신경망 모델들의 성능과 기능을 어떻게 최적화했는지 알아보고 이러한 딥러닝 기반 단일 이미지 생성적 적대 신경망 기법의 장단점을 비교하고 앞으로의 연구 방향에 대해 제안한다.

## II. 딥러닝 기반 단일 이미지 생성적 적대 신경망 기법 비교 분석

### 1. 생성적 적대 신경망(GAN: Generative Adversarial Networks)

생성적 적대 신경망은 생성자와 판별자 두 모델의 경쟁하는 학습(adversarial training)을 활용한다. 생성자는 데이터 분포를 포착하여 원본과 유사한 이미지를 생성하고, 판별자가 원본 이미지와 생성된 이미지의 차이를 구분하지 못하도록 한다. 판별자는 생성자가 만들어낸 이미지와 원본 이미지를 비교하여 생성자가 만들어낸 이미지를 거짓으로 구별하고 원본 이미지를 참이라고 학습한다. 많은 양의 데이터를 사용하여 모델을 학습하면 원본과는 유사하지만 새로운 이미지를 만들어낸다. 최종적으로는 생성자가 만들어낸 이미지와 원본 이미지를 구분하는 확률이 50%에 이르게 되며, 어떤 이미지가 원본 이미지 혹은 생성된 이미지인지 판별자가 구분할 수 없게 되는 상황이 이상적인 학습 결과이다<sup>[1]</sup>.

생성적 적대 신경망은 이전의 이미지 생성 모델들에 비해 괄목할 만한 성능을 달성했지만, 다양한 문제점들이 발견되었다. Neural network가 수렴하지 못하거나 생성자와 판별자의 학습정도가 편향적으로 학습이 되는 등, 학습 과정이 불안정했다. 또한 생성자가 데이터 셋의 이미지들 내부의 분포를 제대로 모델링하지 못해 훈련된 데이터 셋의 일부가 누락되는 문제도 발견되었다.

위와 같은 다양한 문제점들을 개선하기 위해, 또 더 나은 생성적 적대 신경망 모델을 만들기 위해 많은 연구들이 진행되어 왔다. 손실 함수에 중점을 두어 모델을 개발하고 발전시킨 연구<sup>[15][16]</sup>도 있고, 모델 내부의 네트워크를 변형시키거나 새로운 개념을 접목시킨 연구들<sup>[17][18]</sup>도 있다. 본 연구에서는 생성적 적대 신경망을 개선하고 발전시킨 모델 중, 많은 데이터 셋을 사용하지 않고 단 하나의 이미지만을 사용하는 모델들에 대해 분석하고 그 동향이 어떻게 되는지에 대해 살펴보고자 한다. 먼저 단 하나의 이미지의 내부 분포를 학습하여 새로운 이미지를 생성하는 SinGAN<sup>[10]</sup>에 대해 알아보고자 한다. 그리고 SinGAN을 더 개선시키고 최적화한 ConSinGAN<sup>[11]</sup>에 대해 살펴보고자 한다. 다음으로 는 이미지가 가지고 있는 고유한 특징들의 분포를 포착하여 학습함으로써 이미지의 크기나 형태, 종횡비를 변형시켜도 내부 이미지의 비율이나 분포가 자연스럽게 변형되도록 학습하는 InGAN<sup>[12]</sup>에 대해 알아보고자 한다. 그리고 Thin-Plate-spline(TPS) 모델<sup>[19]</sup>을 활용하여 효율적으로 이미지를 증식함으로써 단일 이미지 생성적 적대 신경망의 성능을 높인 DeepSIM<sup>[13]</sup>에 대해 알아보고자 한다. 마지막으로 생성적 적대 신경망의 네트워크 구조를 개선시켜 하나의 이미지 또는 비디오를 학습함으로써 새로운 이미지 또는 비디오를 생성하는 One-Shot GAN<sup>[14]</sup>에 대해 알아보고자 한다.

### 2. SinGAN: Learning a Generative Model from a Single Natural Image

생성적 적대 신경망을 통해 새로운 이미지를 만들어낼 때, 모델은 많은 수의 이미지 데이터에서 이미지를 학습한 후 결과를 도출한다. 이전의 생성 모델들과 비교한다면 생성적 적대 신경망은 개선된 성능을 보여주지만, 이미지의 일부 영역에 대한 분포만을 학습하고(mode collapse), 또 위조범과 경찰의 역할을 하는 생성자와 판별자 간의 경쟁을 통해 학습하는 게임 이론에 근거를 두고 있기 때문에<sup>[1]</sup> 안정된 결과 값을 도출해 내기 위해서는 대량의 데이터 셋을 필요로 하는 단점을 가진다. 생성적 적대 신경망 연구에 대한 단점을 최소화하기 위해, 연구자들은 적은 분량의 데이터를 가지고도 효과적으로 학습할 수 있는 모델을 연구하고자 했다. 이러한 배경에서 나온 모델이 바로 SinGAN이다.

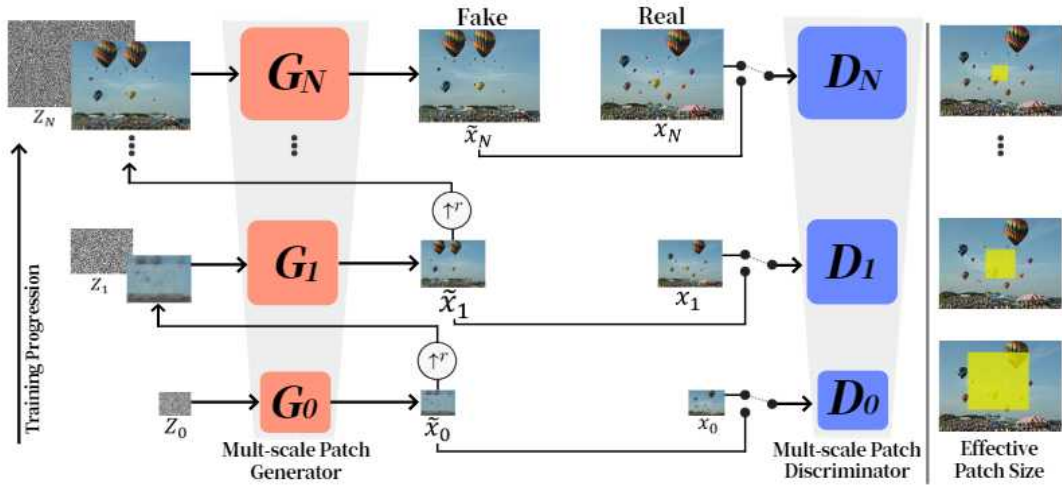


그림 1. SinGAN의 네트워크 구조. 단계가 높아질수록 세밀한 분포를 생성하는 것을 알 수 있다<sup>[10]</sup>

Fig. 1. Network architecture of SinGAN<sup>[10]</sup>

생성적 적대 신경망과 달리 SinGAN은 단 하나의 이미지만을 필요로 하고, 그 이미지 내부 분포를 완벽하게 학습하여 이미지를 생성하는 것이 목표이다. 이를 위해 다수의 생성자 네트워크와 판별자 네트워크를 쌓아 여러 스케일에서 결과를 누적한 이미지를 반복적으로 학습한다. 높은 단계의 스케일에서는 전체적인 형태, 즉 이미지의 배경 등을 학습하고 낮은 단계의 스케일에서는 출력된 이미지의 세부적인 형태를 만들어간다. SinGAN의 네트워크 구조는 그림 1과 같다. 먼저 G0에서부터 시작하여 가장 높은 단계의 스케일을 학습하고, 학습 이후 생성된 이미지를 G1로 보내는 방식이다. 스케일의 단계가 낮아질수록 이전 단계보다 더 세밀한 부분을 학습하며 동시에 판별자는 원본 이미지를 같은 스케일로 축소한 후 생성된 이미지와 비교하여 참, 거짓을 판단한다. G0은 이전 단계가 존재하지 않으므로 노이즈만 입력받아 판별자를 속이기 위한 학습을 진행한다. 이후의 생성자 네트워크 G1~GN은 이전 단계에서 생성된 이미지  $\text{output}(x)$ 에 새로운 노이즈를 추가로 입력받는다. 생성자 네트워크에 들어간  $\text{output}(x)$ 은 노이즈  $Z_n$ 과 더해져서 생성자 네트워크를 통과한다. 새로운 이미지를 출력하기 전에 이전 단계의  $\text{output}(x)$ 를 더해줌으로써 새로운 이미지를 출력하게 된다.

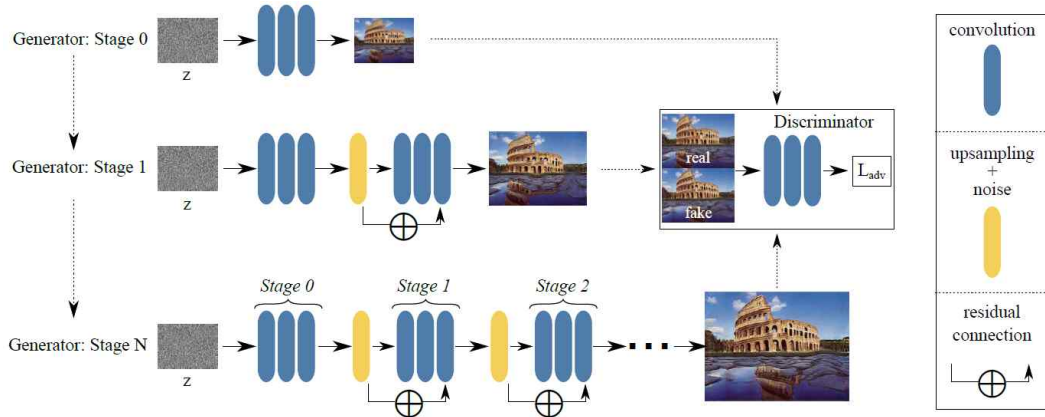
$$\tilde{x}_n = (\tilde{x}_{n+1}) \uparrow^r + \psi_n(z_n + (\tilde{x}_{n+1}) \uparrow^r) \quad (1)$$

이 과정을 통해 출력된 이미지는 원본 이미지와 상이하지 않은 출력 이미지를 생성한다. 생성자 네트워크는 내부적으로 합성곱 레이어로 구성되어 있고, filter의 수는 32개에서 시작해 스케일의 단계가 낮아질수록 더 많은 feature map을 생성하도록 구성되어 있다. 학습은 높은 단계의 스케일에서부터 순차적으로 진행된다.

Fréchet Inception Distance(FID)<sup>[20]</sup>는 생성된 이미지의 품질을 평가하는 지표 중 하나이고, 영상 집합 사이의 거리를 나타낸다. GAN을 이용하여 생성된 영상의 집합과 실제 만들어내고자 하는 데이터 셋의 분포에 대한 거리를 계산한다. 단일 이미지 FID(SIFID)는 단일 이미지 영역에 대해 FID를 적용시킨 것이며 생성된 이미지와 실제 이미지 간의 네트워크 활성화 통계를 비교한다. 거리가 가까울수록, 즉 낮은 점수를 기록할수록 생성된 이미지와 실제 이미지 간의 차이가 적다는 것을 의미한다. SinGAN은 SIFID 점수를 0.09점을 기록하여 생성 이미지를 실제 이미지와 거의 유사하게 만들어낼 수 있다는 것을 확인할 수 있다.

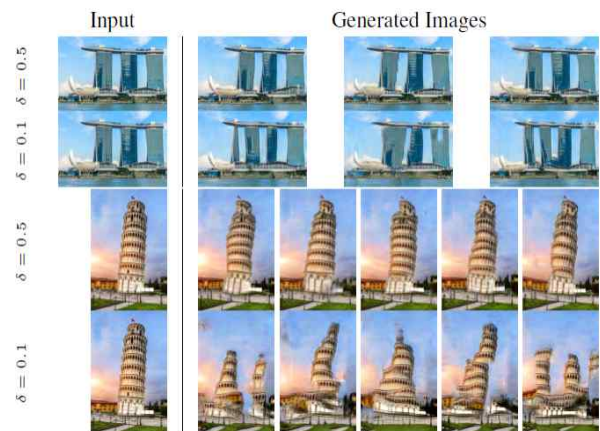
### 3. ConSinGAN: Improved Techniques for Training Single-Image GANs

SinGAN은 단 하나의 이미지만을 학습하여 원본 이미지와 유사한 새로운 이미지를 만들어 낼 수 있고, 학습된 모델

그림 2. ConSinGAN의 네트워크 구조<sup>[11]</sup>Fig. 2. Network architecture of ConSinGAN<sup>[11]</sup>

에 다른 이미지  $x$ 가 들어가도  $x$ 와 유사하지만 다른 이미지를 출력한다. 하지만 SinGAN은 하나의 이미지를 학습하는데 오랜 시간이 소요되고, 또 생성된 이미지의 해상도를 결정하는 부분에 있어서 최적화되지 못한 부분들이 있다. 또한 한 장의 이미지 내부에서 패치 단위로 이미지 분포를 학습하기 때문에 이미지 내부 분포가 복잡한 이미지를 학습하는 것은 한계가 있었다. 따라서 높은 성능에도 불구하고 다양한 application들을 범용적으로 활용하기에는 다소 부족한 부분들이 있었기에 연구자들은 SinGAN을 더 개선시키고 성능을 향상시키기 위한 연구를 진행하였다.

ConSinGAN은 위와 같은 배경에서 만들어진 딥러닝 기반 단일 이미지 생성적 적대 신경망이다. ConSinGAN은 그림 2에 나타나듯이 SinGAN을 기반으로 만들어진 모델이지만, 여러 부분에서 개선되거나 차이를 보인다. (i) SinGAN은 Gn 단계를 훈련할 때, 이전 단계의 생성자 네트워크는 고정되어 다른 단계의 훈련과 상호작용이 없다. 반면 ConSinGAN은 한 번에 여러 단계의 생성자 네트워크를 학습하는데, 한 번에 모든 단계의 생성자 네트워크를 학습하면 과적합의 가능성이 상승하기에 한 번에 3개의 단계를 학습한다. (ii) 또한 한 번에 3개의 생성자를 훈련하는데, 각 단계마다 다른 learning rate를 사용한다. 낮은 단계의 생성자 네트워크에서 높은 learning rate를 사용하면 학습 이미지에 더 유사한 이미지가 출력되고, 낮은 learning rate를 사용하면 퀄리티는 낮지만, 더 다양한 이미지가 그림 3과 같이 출력된다. (iii) SinGAN은 이전 단계에서 생성된 이미지

그림 3. ConSinGAN에서 Learning Rate의 변화에 따른 생성된 이미지의 변화<sup>[11]</sup>Fig. 3. Changes in the generated image according to the change of the learning rate in ConSinGAN<sup>[11]</sup>

를 다음 단계에 전달하지만, ConSinGAN은 이미지 대신, 이미지에서 추출한 특성만을 다음 단계에 전달한다. (iv) 사실적인 이미지를 생성하기 위해 필요한 학습 단계는 세밀한 부분의 분포를 학습하는 단계보다는 전체적인 배경에 대한 분포를 학습하는 단계이기 때문에, ConSinGAN은 SinGAN과 비교했을 때, 전체적인 배경 이미지 분포에 대한 부분의 가중치를 높이고 세밀한 이미지의 분포를 학습하는 단계는 줄였다. 이러한 개선된 부분들에 의해 ConSinGAN은 SinGAN에 비해 최대 6배 빠른 속도로 모델을 학습하고 더 적은 파라미터를 학습하며, 이미지의 전체적인 구조를 보다



적절하게 파악할 수 있다<sup>[11]</sup>.

ConSinGAN은 SIFID 점수를 0.06점을 기록했고, 이는 SinGAN이 0.09점을 받은 것에 비해 약간의 높은 점수를 기록했다. 하지만 ConSinGAN은 SinGAN에 비해 적은 양의 데이터와 적은 양의 학습단계, 그리고 1/6 정도의 소요시간을 기록했기 때문에 다방면에서 SinGAN의 성능을 개선한 것을 알 수 있다.

#### 4. InGAN: Capturing and Remapping the “DNA” of a Natural Image

생성적 적대 네트워크는 일반적으로 대규모의 이미지 데이터 셋 내의 이미지 분포를 학습하고, 이 분포에서 새로운 이미지를 생성할 수 있다. 그러나 각 자연 이미지에는 고유의 내부 특성이 있으며 이를 고려하여 이미지의 크기나 비율을 자연스럽게 조정하는 것은 어려운 작업이었다. 크기나 모양, 중형비를 조절할 때, 이미지 내부의 분포가 자연스럽게 바뀌지 않고 왜곡되거나 부자연스럽게 변경되는 것을 볼 수 있다. 이러한 문제점들을 개선하기 위해 InGAN은 하나의 원본 이미지를 사용하여 원본 이미지의 내부 분포와 출력된 이미지의 내부 분포를 일치하도록 하고 그 반대의 과정도 마찬가지로 하도록 모델을 구성함으로써 출력되는 이미지에 의도하지 않은 오브젝트가 추가되지 않고, 중요한 오브젝트 또한 남아 있게 할 수 있다. 또한 생성된 이미지에서의 오브젝트 위치를 원본 이미지 내부에 있는 오브젝트 위치와 동일하게 출력하도록 모델이 구성되어 있다<sup>[12]</sup>.

그림 4는 InGAN의 구조를 나타내는 그림으로, 원본 이미지  $x$ 를 출력된 이미지  $y$ 로 리타겟팅하는 생성자  $G$ 로 구성되며, 그 크기와 형상은 기하학적 변환의 역할을 하는 레이어  $T$ 에 의해 결정된다. Multi-scale의 판별기  $D$ 는 출력된 이미지  $y$ 의 패치 특성을 원본 이미지의 패치 특성과 구별하는 방법을 학습한다. 또한  $G$ 의 automorphism을 이용하여  $G$ 와 정반대되는 구조를 가진  $T$ 를 사용하고 그것에 대한 input 이미지로  $y$ 를 받아 입력 이미지를 복원한다. InGAN은 multi-scale 레이어를 가진 판별자를 사용하는데, 단일 스케일 판별자는 특정 크기의 패치를 이용해 이미지 내부의 특성을 포착하지만, multi-scale 판별자를 사용하면 사용하는 모든

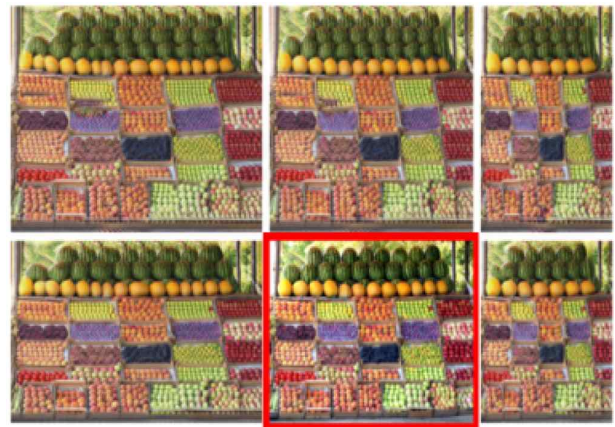


그림 5. InGAN을 통해 생성된 이미지. 화면의 비율과 크기가 변해도 이미지 내부 오브젝트가 자연스럽게 변하는 것을 확인할 수 있다<sup>[12]</sup>

Fig. 5. InGAN can naturally synthesize a new images of significantly different size/shape/aspect ratios<sup>[12]</sup>

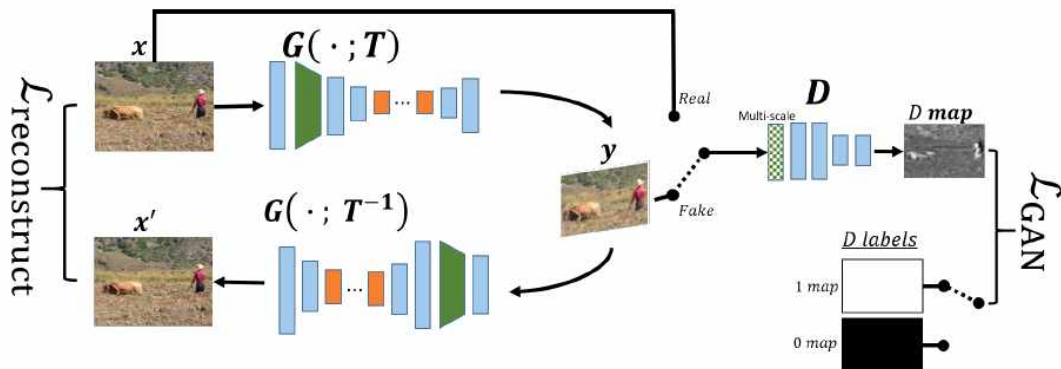


그림 4. InGAN의 네트워크 구조<sup>[12]</sup>

Fig. 4. Network architecture of InGAN<sup>[12]</sup>

패치 크기에서 패치 분포를 일치시킬 수 있고, 이미지의 미세한 구조와 대략적인 구조 모두를 학습한다. 이러한 InGAN의 모델의 특성으로 인해 그림 5와 같이 이미지의 크기와 비율이 조정되어도 자연스러운 이미지를 형성하는 것을 볼 수 있다<sup>[12]</sup>.

## 5. DeepSIM: Image Shape Manipulation from a Single Augmentation Training Sample

DeepSIM은 많은 수의 이미지 데이터를 사용해야 하는 딥러닝 모델들과는 달리 하나의 이미지와 그 이미지를 단순화시킨 **primitive** 이미지를 한 쌍을 원본 데이터로 하여 학습하는 모델이다<sup>[13]</sup>. 원본 이미지가 모델에 주어지면 먼저 원본 이미지에 대한 **primitive** 이미지가 생성된다. 모델이 원본 이미지와 생성된 이미지 내부의 분포들을 학습한 후에, 사용자는 **primitive** 이미지를 조작하여 원본 이미지를 어떻게 변경할 것인지 명시적으로 설계할 수 있다. 조작된 **primitive** 이미지를 학습된 모델에 넣어주면 모델은 그림 6과 같이 조작된 **primitive** 이미지를 실제 이미지로 변환한다. DeepSIM과 유사한 학습 과정을 가진 TuiGAN<sup>[21]</sup>에서는 학습을 위해 한 쌍의 이미지 데이터 원본 이미지로 사용하지만, 한 번 학습을 하고 난 후에 다른 어플리케이션을 적용시키기 위해서는 다시 학습을 해야하는 단점이 있다. 하지만 DeepSIM은 한 번 학습을 하고난 후 다양한 어플리케이션에 적용시키는 것이 가능한, 더 향상된 성능을 보인다.

딥러닝 기반 단일 이미지 생성적 적대 신경망 기법들에서

는 하나의 이미지에서 얼마나 많은 것을 배울 수 있는지가 중요한 연구 주제이다. DeepSIM은 **thin-plate spline(TPS)** 기법에 기반을 둔 데이터 증식(**data augmentation**)을 활용하여 충분한 학습 데이터를 확보한다. TPS 기법은 여러 컴퓨터 비전 응용 프로그램에서 성공적으로 적용된 좌표 변환을 모델링하는 효과적인 기법이다<sup>[19]</sup>. DeepSIM에서 이용한 TPS 기법은, 이미지 위에 같은 간격의 3X3 격자를 배치하여 이미지 내부 픽셀에 **control points (i, j)**를 지정함으로써 무작위로 이미지를 왜곡시킨다. 학습하는 이미지에 대한 과도한 왜곡은 모델의 성능에 부정적인 영향을 미치기 때문에, 왜곡되는 정도는 이미지 너비와 높이의 최대 10%로 제한했다.

**Learned Perceptual Image Patch Similarity(LPIPS)**<sup>[22]</sup>는 신경망 모델에서 추출되는 특성을 이용하여, 학습을 하면서 사람의 인지적 특성에 얼마나 적합한지 유사도를 평가하는 모델이다. LPIPS의 점수가 낮으면, 생성된 이미지가 실제 이미지와 유사하다는 의미이다. DeepSIM은 LPIPS와 SIFID 각각의 지표에서 원본으로 삼았던 모델인 Pix2PixHD<sup>[23]</sup> 모델보다 더 낮은 점수를 기록했고, 이는 기존의 모델보다 더 나은 성능을 보인 것을 의미한다.

## 6. One-Shot GAN: Learning to Generate Samples from Single Images and Videos

충분한 양의 데이터가 없으면 생성적 적대 신경망은 예상되는 이미지에서 크게 벗어난 이미지를 생성하거나 전혀 변화가 없는 이미지를 생성하기도 한다. 데이터의 양이 제한

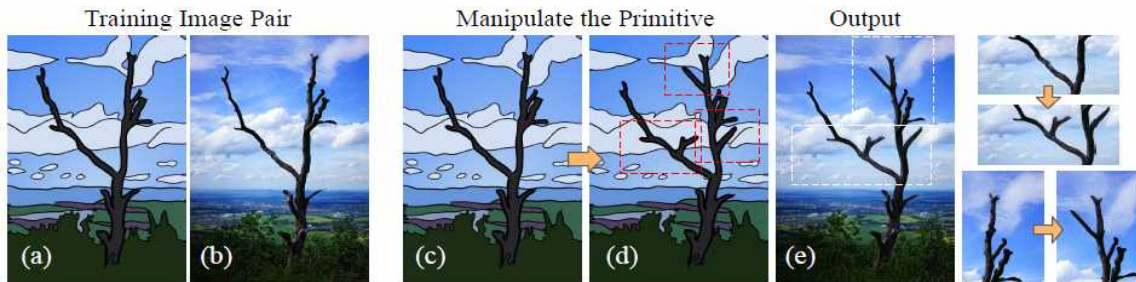


그림 6. DeepSIM의 이미지 생성과정과 결과. 원본 이미지와 **primitive** 이미지 한 쌍을 학습하여 자연스러운 이미지를 생성한 것을 확인할 수 있다<sup>[13]</sup>

Fig. 6. Image manipulation process of DeepSIM. A natural image is generated by learning a pair of original image and primitive image<sup>[13]</sup>

된 상황에서는 이러한 모델들의 성능은 크게 저하될 수밖에 없고, 현실의 상황에서는 데이터를 충분히 확보하는 것이 매우 어려운 상황들이 종종 있기에<sup>[14]</sup> 연구자들은 생성적 적대 신경망을 꾸준히 개선했다. 이러한 배경에서 나온 모델들이 SinGAN과 InGAN 등 이지만, 성능은 이미지 데이터에서만 개선되었고 비디오 데이터에 대해서는 적용되지 않아 비디오 데이터를 학습할 수는 없다. 따라서 이미지 내부 특성을 포착하는 패치가 일관성이 없이 특성들을 포착하고, 생성된 이미지가 왜곡 되거나 비현실적이기 때문에 활용하기에는 무리가 있었다. One-Shot GAN은 이전 모델들의 문제점을 개선하여 원본 이미지의 내부 분포가 유지되면서 원본과는 크게 다른 이미지를 생성해내도록 모델의 성능을 끌어올렸다.

One-Shot GAN 모델의 판별자는 원본 이미지 내부의 오브젝트 분포와 이미지의 현실성을 결정하는 두 가지 방식의 네트워크로 구성된다. 이와 같은 네트워크 기능의 분리는 **overfitting**을 방지하며 생성자로 하여금 더 많은 데이터를 학습할 수 있도록 한다<sup>[14]</sup>. 구체적으로 그림 7을 보면, 판별

자 D는 낮은 단계에서 이미지 내부 분포를 파악하는 **Dlow-level**, 공간 정보에 상관없이 이미지 내부 오브젝트를 파악하는 **Dcontent** 및 이미지의 공간 정보만을 파악하여 사실적인 이미지인지 판별하는 **Dlayout**으로 구성된다. **Dcontent**와 **Dlayout**의 분리는 모델이 기본적인 정보를 파악한 후인 중간 단계에서 시행되고, 이전 단계에서 제한된 정보만 받기 때문에 **overfitting**을 방지한다. 결과적으로 다양한 이미지를 생성하는데에 긍정적인 영향을 미친다. 이를 통해 원본 이미지의 내부 오브젝트와 배경을 유지하면서 다양한 이미지를 생성할 수 있고, 비디오 데이터 또한 학습할 수 있으며, 시각적으로 새로운 장면들을 합성해 낸 것을 그림 8에서 확인할 수 있다.

### III. 네트워크 성능 비교

이번 장에서는 딥러닝 기반 단일 이미지 생성적적대 신경망에서 출력된 이미지들의 결과를 살펴보고 각 모델들의 성

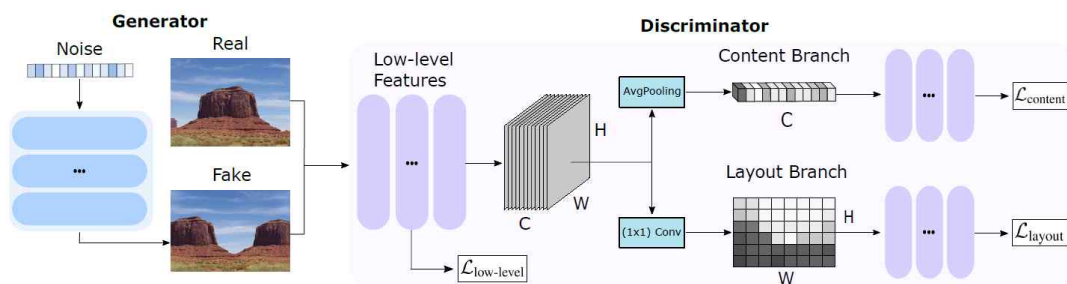


그림 7. One-Shot GAN의 네트워크 구조. 판별자 네트워크가 두 기능으로 분리된 것을 확인할 수 있다<sup>[14]</sup>

Fig. 7. Network architecture of One-Shot GAN. The discriminator of One-Shot GAN has two different branches<sup>[14]</sup>



그림 8. One-Shot GAN으로 이미지나 비디오 프레임을 생성한 결과. 빨간색 테두리 이미지는 입력, 나머지는 One-Shot GAN 생성 결과<sup>[14]</sup>

Fig. 8. The result of generating images or video frames with One-Shot GAN. The red border are the input images, and the rest is the generation results of One-Shot GAN<sup>[14]</sup>



능을 평가하고자 한다. 성능 평가를 위해 각 논문들의 코드 및 모델은 각 논문의 저자들이 공개한 것을 바탕으로 하였다<sup>[24][25][26][27]</sup>. 실제 사진에서의 효용성을 평가하기 위해서, 기존 논문에서 사용하지 않은 광안대교의 배경 사진과 불꽃놀이 사진을 학습했다. DeepSIM의 경우 모델의 기능이 이미지 크기 변형이 아닌 이미지 조작에 집중되어 있기 때문에, 광안대교 사진의 이미지 크기 변형 결과와 모델의 특징을 통해 어떠한 이유에서 이미지 크기 변형의 성능이 좋지 못한 이유를 살펴보고, DeepSIM의 원래 목적인 이미지 조작 기능은 나무 사진에 대한 학습 결과를 통해 알아보하고자 한다. 비디오 데이터 학습에 특화된 One-Shot GAN은 이미지를 학습하는 다른 4개의 모델과 직접적인 비교가 어려워 평가에서 제외하였다.

그림 9는 원본 이미지에 대한 SinGAN과 ConSinGAN의 모델에 학습시킨 후 너비를 두 배로 늘리는 리타겟팅 작업을 수행한 것이다. SinGAN은 multi-scale로 학습하는 과정에서 이미지의 전체적인 특성을 학습하다가 점차 세세한 부분을 학습한다. SinGAN은 전체적인 하늘, 바다의 모습과 왼쪽에 있는 건물들같이 배경과 큰 오브젝트에 대해서는 크게 변화하는 모습이 없지만, 이미지 상에서 광안대교 다리의 기둥 부분의 개수가 늘어나고 무작위한 위치에 생성된 것을 볼 수 있다. ConSinGAN의 학습 과정은 한 번에 3개의 생성자 네트워크를 학습하고, 동시에 학습 단계가 진행될수록 점진적으로 더 높은 learning rate를 적용한다. 즉 전체적인 배경을 학습하는 단계에서 원본 이미지를 학습하는 비율이 SinGAN보다 더 작기

때문에 원본 이미지에 비해 생성된 이미지의 전체적인 구조가 크게 변화한다. 이러한 ConSinGAN의 학습 단계의 특징 때문에 더 다양한 결과가 출력되지만 동시에 왜곡도 심해지는 결과를 볼 수 있다. ConSinGAN의 광안대교 사진 학습 결과는 원본 이미지와 유사하지만 SinGAN에 비해 더 왜곡된 배경과 내부 오브젝트를 출력했다.

SinGAN의 학습과정을 염두에 두고 생성된 불꽃놀이 사진에서 원형으로 퍼지는 큰 불꽃들이 다양하게 나타날 것으로 예상했으나, 실제 결과에서는 원형으로 퍼지는 불꽃들을 오브젝트로 인식한 것이 아닌 각각의 작은 불꽃들을 오브젝트로 인식하여 이미지 전반에 걸쳐 복제되어 나타나는 것을 확인했고 ConSinGAN 또한 유사한 학습 결과를 나타냈다.

InGAN의 실험 결과는 그림 10과 같다. InGAN은 기하학적 변환의 기능을 하는 T 레이어로 인해 원본 이미지의 비



그림 10. InGAN으로 원본 이미지와 다른 비율로 이미지를 생성한 결과  
Fig. 10. The generated images of different ratios from the original via InGAN

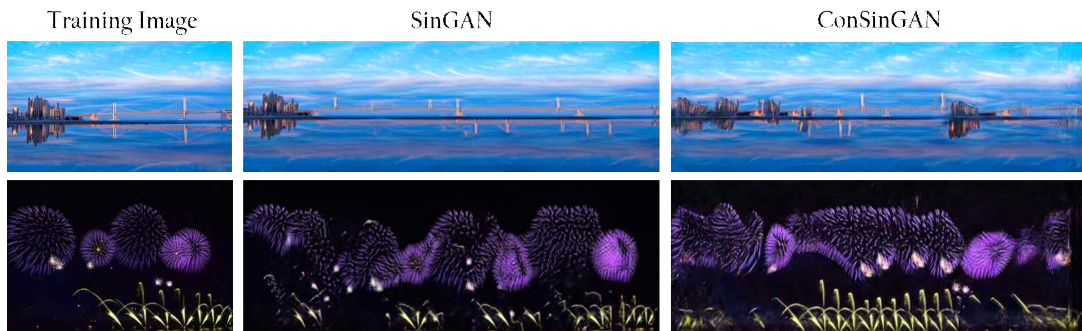


그림 9. 광안대교와 불꽃놀이 이미지에 대해 SinGAN, ConSinGAN으로 각각 학습하여 다른 비율의 이미지를 생성한 결과  
Fig. 9. The result of generating images of different ratios by learning each Gwangangan Bridge and fireworks images with SinGAN and ConSinGAN

울을 조정하면 내부 오브젝트의 배치를 자연스럽게 조정한다. InGAN 모델의 이러한 특징 때문에 SinGAN과 ConSinGAN의 광안대교 사진의 학습 결과와는 달리, 원본 이미지 내부 오브젝트를 복제하여 새로운 이미지를 생성한 것을 확인할 수 있다. 불꽃놀이 사진에서 InGAN은 다른 모델과는 달리 원형으로 퍼지는 불꽃들을 복제한 부분도 있지만, 전체적으로는 SinGAN, ConSinGAN과 마찬가지로 원형으로 퍼지는 불꽃이 아닌 작은 불꽃들을 복제한 것을 확인할 수 있다.

그림 11은 DeepSIM 모델의 광안대교 이미지 조작 및 이미지 크기 변형 결과이다. 원본 이미지와 segmentation을 적용시킨 이미지를 한 쌍으로 하여 학습했고, 이후에 segmentation 이미지에서 광안대교의 기둥 사이에 기둥을 하나 더 늘려 테스트를 진행했을 때, 자연스러운 이미지를 출력했다. 그러나 원래의 segmentation 이미지 우측에 추가적으로 기둥을 하나 더 늘려서 이미지 크기 변형을 한 결과, SinGAN과 ConSinGAN, InGAN에 비해 사실적인 이미지를 생성하지 못했다. DeepSIM 모델은 학습 과정에서 인식한 이미지를 바탕으로 확장된 부분에 대해 추가적으로 이미지를 생성하는 것이 아닌, 확장된 부분에 이미 학습한 이미지 분포를 적용시키는 것에 집중하기에 이미지 내부 오브젝트의 분포를 고려하지 않고 이미지 크기만을 늘린 결과물을 생성했다. 그림 11의 Image resize 결과에서, 왼쪽에 건물물이 있는 부분은 상대적으로 생성이 잘 된 반면, segmentation 이미지에서 확장된 오른쪽의 다리 부분은 형체를 거의

알아볼 수 없다. 이는 근본적으로 DeepSIM 모델이 이미지 크기 변형 기능에 초점을 맞춘 모델이 아닌, 이미지 조작 기능에 더 집중한 모델이기 때문에 InGAN과 같이 내부 오브젝트를 복제하면서 이미지를 확장시키는 기능을 수행하기에는 부적합하다.

그림 12는 DeepSIM의 이미지 조작에 대한 성능을 SinGAN 모델과 비교한 결과이다. 원본 이미지와 segmentation 이미지 한 쌍을 학습하여, 조작된 segmentation 이미지를 학습된 모델에 적용시킨 결과는 DeepSIM 모델의 성능이 SinGAN 모델에 비해 확연히 좋은 모습을 보였다. 이는 DeepSIM 모델의 기능이 이미지 크기 변형이 아닌, 이미지 조작에 더 집중하고 있는 것을 알 수 있는 결과이다.

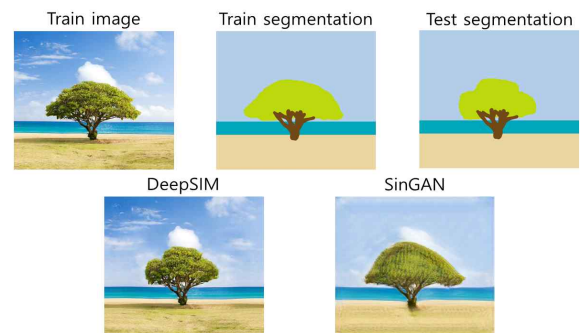


그림 12. SinGAN과 DeepSIM의 이미지 조작 성능 비교  
Fig. 12. Comparison of image manipulation performance of SinGAN and DeepSIM

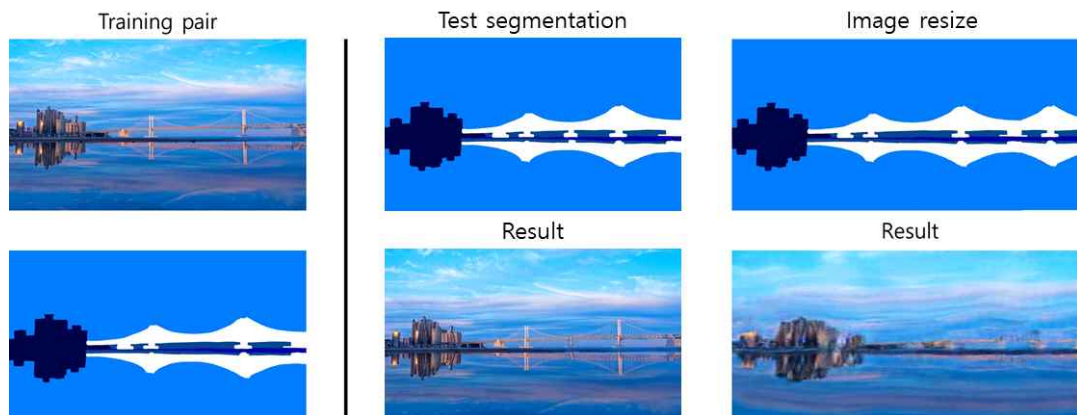


그림 11. DeepSIM 모델을 이용한 이미지 조작 및 이미지 크기 변형 결과  
Fig. 11. Image manipulation and image resize results using DeepSIM model

표 1. 각 모델들의 장단점 비교

Table 1. Comparison of advantages and disadvantages of each network

	Detail	Diversity	Manipulability	Training time	Memory usage
SinGAN	medium	low	medium	30 minutes	9139 MB
ConSinGAN	low	high	medium	15 minutes	2131 MB
InGAN	high	low	medium	216 minutes	3019 MB
DeepSIM	high	medium	high	106 minutes	6349 MB
One-Shot GAN	medium	high	medium	-	-

표 2. 정량적 비교 측정

Table 2. Quantitative evaluation

Method	SinGAN <sup>[10]</sup>	ConSinGAN <sup>[11]</sup>	InGAN <sup>[12]</sup>	DeepSIM <sup>[13]</sup>
SIFID <sup>[10]</sup> ↓	0.218	0.219	0.354	0.040
LPIPS <sup>[22]</sup> ↓	0.347	0.373	0.376	0.108

표 1은 위의 실험 결과를 종합하여 각 모델들 특징 및 장단점을 세밀함(Detail), 다양성(Diversity), 조작성(Manipulability), 학습 시간(Training time), 메모리 사용량(Memory usage)으로 비교 정리하였다. SinGAN은 원본 이미지와 비슷한 형태의 이미지를 출력하여 다양성이 떨어지며, 입력 노이즈에 따라 영상을 출력하기 때문에 조작성도 떨어진다. 또한 각 stage를 순차적으로 학습하기 때문에 학습 속도도 느리다는 단점이 있다. 이에 비해 ConSinGAN은 여러 stage를 동시에 학습하기 때문에 학습 속도가 SinGAN에 비해 많이 단축되었으며, 다양성도 증가하였다. 하지만 그림 9와 같이 결과 이미지가 많이 왜곡되기 때문에 Detail 떨어진다. SinGAN과 ConSinGAN의 사용되는 메모리를 비교했을 때, InGAN은 기하학적 변형에 초점을 맞춘 논문이기 때문에 원본 이미지와 매우 유사한 영상이 출력되어 Detail은 높고, Diversity는 떨어진다. InGAN에 사용되는 메모리는 다른 모델들에 비해 상대적으로 작은 편이나, 소요되는 시간은 5개의 모델 중 가장 많은 것을 확인할 수 있다. DeepSIM 네트워크는 이미지 조작을 타겟으로 하는 논문이기 때문에 하나의 pair를 학습하면 다양한 조작이 가능하다는 장점이 있다. One-Shot GAN은 이미지 뿐만 아니라 비디오도 생성할 수 있기 때문에 다양성이 매우 크다는 장점이 있으나 조작성은 떨어진다는 단점이 있다.

표 2는 SIFID와 LPIPS를 통해 각 모델들이 생성한 광안

대교 이미지를 원본 이미지와 비교하여 측정한 점수이다. Single Image FID(SIFID)는 생성된 이미지와 실제 이미지 간의 feature space에서의 거리를 나타낸다. 거리가 가까울수록, 즉 낮은 점수를 기록할수록 생성된 이미지와 실제 이미지 간의 유사도가 높다는 것을 의미한다. Learned Perceptual Image Patch Similarity(LPIPS)<sup>[22]</sup>는 신경망 모델에서 추출되는 특성을 이용하여, 학습을 하면서 사람의 인지적 특성에 얼마나 적합한지 유사도를 평가하는 모델이다. 즉, LPIPS의 점수가 낮으면, 생성된 이미지가 실제 이미지와 인지적으로 유사하다는 의미이다. SinGAN과 ConSinGAN의 SIFID 점수는 크게 차이가 나지 않았다. 또한 ConSinGAN은 학습할 때 원본 이미지와는 다른 다양한 이미지를 생성하기 때문에 LPIPS 점수가 SinGAN에 비해 높았다. DeepSIM은 feature space에서나 인지적으로도 원본 이미지와 생성된 이미지 사이에 큰 차이가 없기 때문에 SIFID와 LPIPS 점수가 상대적으로 낮게 측정된 것을 확인할 수 있다.

## IV. 결 론

본 논문에서는 단일 이미지만 사용하여 학습하는 생성적 적대 신경망 기법의 현황 및 동향에 대해 살펴보았다. 단일

이미지를 다양한 스케일로 학습하여 내부 분포를 분석하고 이를 다양한 어플리케이션에 적용시킬 수 있는 SinGAN, 이 모델의 구조를 개선하고 학습 과정을 수정하여 더 빠른 학습 속도와 향상된 성능을 보인 ConSinGAN, multi-scale 판별자를 사용하고 생성자 네트워크와 역방향 생성자 네트워크를 연결함으로써 원본 이미지의 크기나 비율을 자연스럽게

게 리타겟팅하는 InGAN, 원본 이미지와 primitive 이미지 한 쌍을 이용한 학습을 통해 자연스러운 이미지 조작이 가능하고, TPS모델을 이용한 이미지 데이터 증식으로 뛰어난 성능을 보인 DeepSIM, 판별자 네트워크를 두 가지의 기능으로 분리하여 원본 이미지 내부 오브젝트와 배경을 따로 구분함으로써 이미지뿐만 아니라 비디오에서도 새로운 장

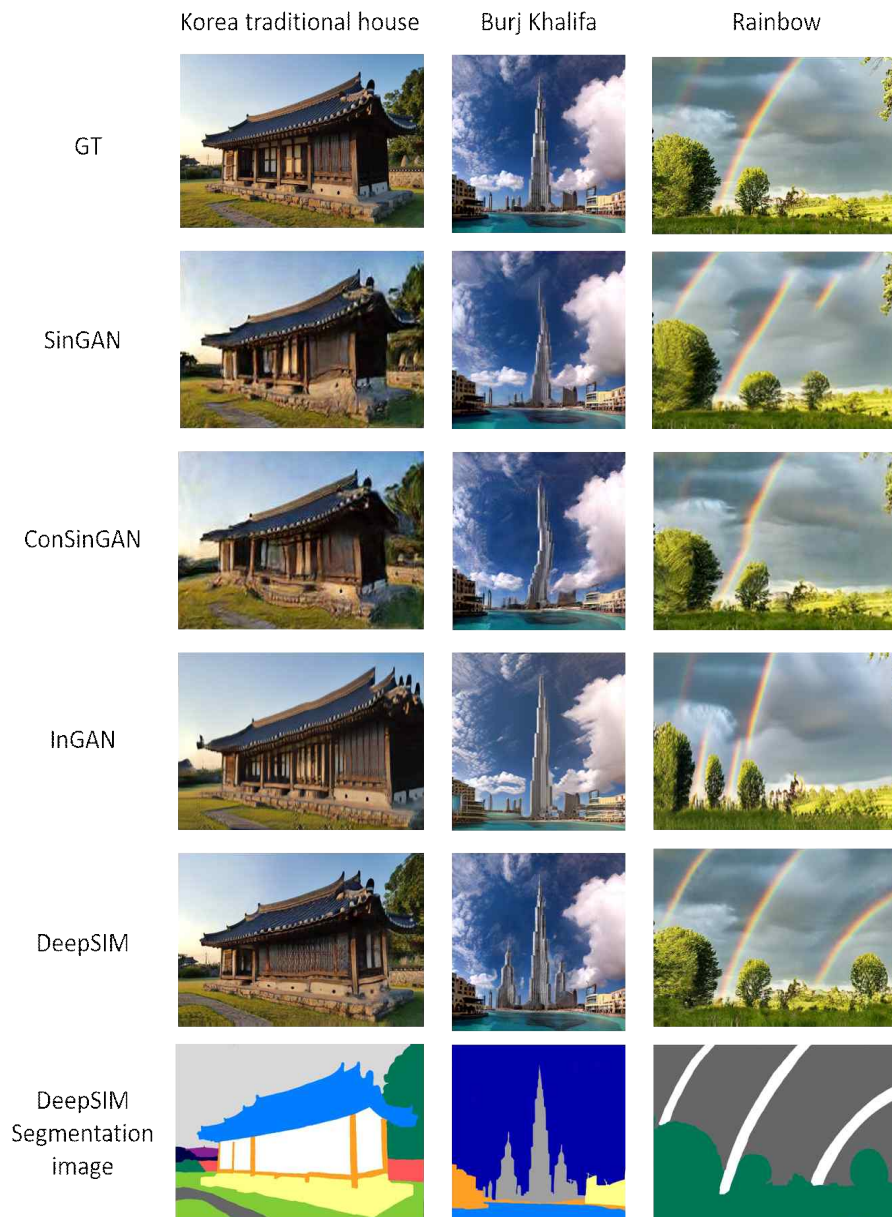


그림 13. 다양한 이미지에 대해 각 모델이 생성한 이미지  
Fig 13. Images generated by each network for different images



면들을 학습할 수 있는 One-Shot GAN 등, 단일 이미지 생성적 적대 네트워크 연구들은 꾸준히 모델을 개선하고, 학습 방법을 최적화하는 방식으로 연구를 진행해왔다.

각 모델들은 단일 이미지를 학습한다는 공통점이 있지만, 네트워크 구조에 따라 수행할 수 있는 애플리케이션이 다르다. SinGAN은 하나의 이미지만을 학습하여 여러 애플리케이션에 활용할 수 있지만 그 성능이 만족스럽지 못하다. ConSinGAN은 더 빠르고 개선된 성능을 보였지만 생성하는 이미지가 현실적이지 못하고 왜곡이 심한 모습을 보였다. InGAN은 리타겟팅에 대해 훌륭한 성능을 보이지만 다양한 애플리케이션으로 활용하지 못하고, DeepSIM 또한 이미지 조작에 초점을 두어 좋은 성능을 가졌지만, 그 외의 기능에 대한 성능은 좋지 못한 결과를 확인했다.

위의 모델들이 생성한 이미지들은 공통적으로 완벽하게 모방하여 새로운 이미지를 만들어내는 것이 아닌, 왜곡되고 비현실적인 이미지들이 생성되는 것을 확인할 수 있었다. 물론 모델을 통해서 완벽에 가까운 이미지를 만들어낼 수 있지만 제한되고 통제적인 환경에서만 가능하기 때문에 아직 상용화 단계에 도달하기에는 불완전하다. 또한 아직까지 다양한 task를 처리하는 범용적인 모델은 성능이 많이 떨어지기 때문에 DeepSIM과 같이 특정 task에 최적화 된 모델이 연구되고 있다는 한계점이 있다. 이러한 방법들의 한계점을 극복하기 위해, 단일 이미지 생성적 적대 신경망의 연구는 다양한 기능들에 대해 범용적으로 좋은 성능을 가질 수 있는 방향으로 연구되어야 할 것이다.

## 참 고 문 헌 (References)

- [1] I.J. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A.C. Courville, Y. Bengio, "Generative adversarial nets," In Advances in Neural Information Processing Systems, pp.2672-2680, 2014.  
doi: <https://doi.org/10.48550/arXiv.1406.2661>
- [2] T. Karras, T. Aila, S. Laine, J. Lehtinen, "Progressive Growing of GANs for Improved Quality, Stability, and Variation," arXiv preprint arXiv:1710.10196, 2017.  
doi: <https://doi.org/10.48550/arXiv.1710.10196>
- [3] X. Chen, Y. Duan, R. Houthoof, J. Schulman, I. Sutskever, P. Abbeel, "InfoGAN: Interpretable Representation Learning by Information Maximizing Generative Adversarial Nets," 30th Conference on Neural Information Processing Systems, NIPS, 2016.  
doi: <https://doi.org/10.48550/arXiv.1606.03657>
- [4] J. Y. Zhu, T. Park, P. Isola, A. A. Efros, "Unpaired Image-To-Image Translation Using Cycle-Consistent Adversarial Networks," Proceedings of the IEEE International Conference on Computer Vision (ICCV), pp.2223-2232, 2017  
doi: <http://dx.doi.org/10.1109/ICCV.2017.244>
- [5] Moon, C. K., Uh, Y. J., Byun, H. R., "Image Restoration using GAN," JOURNAL OF BROADCAST ENGINEERING, pp.503-510, 2018  
doi: <https://doi.org/10.5909/JBE.2018.23.4.503>
- [6] Nah, J. Y., Sim, C. H., Park, I. K., "Depth Image Restoration Using Generative Adversarial Network," JOURNAL OF BROADCAST ENGINEERING, pp.614-621, 2018  
doi: <https://doi.org/10.5909/JBE.2018.23.5.614>
- [7] Kim, k. h., Kong, K. B., Kang, S. J., "A Review on Deep Learning-based Image Outpainting," JOURNAL OF BROADCAST ENGINEERING, pp.61-69, 2021  
doi: <https://doi.org/10.5909/JBE.2021.26.1.61>
- [8] M. Arjovsky, S. Chintala, L. Bottou, "Wasserstein Generative Adversarial Networks," Proceedings of the 34th International Conference on Machine Learning, pp. 214-223, 2017.  
doi: <https://doi.org/10.48550/arXiv.1701.07875>
- [9] T. Miyato, T. Kataoka, M. Koyama, Y. Yoshida, "Spectral normalization for generative adversarial networks," arXiv preprint arXiv:1802.05957, 2018.  
doi: <https://doi.org/10.48550/arXiv.1802.05957>
- [10] T.R. Shaham, T. Dekel, T. Michaeli, "Singan: Learning a generative model from a single natural image," Proceedings of the IEEE International Conference on Computer Vision, pp.4570-4580, 2019.  
doi: <https://doi.org/10.48550/arXiv.1905.01164>
- [11] T.Hinz, M. Fisher, O. Wang, S. Wermter, "Improved techniques for training single-image gans," In Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision, pp.1300 - 1309, 2021.  
doi: <http://dx.doi.org/10.1109/WACV48630.2021.00134>
- [12] A. Shocher, S. Bagon, P. Isola, M. Irani, "Ingan: Capturing and remapping the "dna" of a natural image," In IEEE International Conference on Computer Vision (ICCV) 2019.  
doi: <http://dx.doi.org/10.1109/ICCV.2019.00459>
- [13] Y. Vinker, E. Horwitz, N. Zabari, "Image Shape Manipulation from a Single Augmented Training Sample," Proceedings of the ICCV, 2021.  
doi: <http://dx.doi.org/10.1109/ICCV48922.2021.01351>
- [14] V. Sushko, J. Gall, A. Khoreva, "One-shot gan: Learning to generate samples from single images and videos," in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp.2596 - 2600. 2021.  
doi: <http://dx.doi.org/10.1109/CVPRW53098.2021.00293>
- [15] X. Mao, Q. Li, H. Xie, R.Y.K. Lau, Z. Wang, "Least squares generative adversarial networks," Proceedings of the ICCV, 2016.  
doi: <http://dx.doi.org/10.1109/ICCV.2017.304>
- [16] D. Berthelot, T. Schumm, L. Metz, "BEGAN: Boundary Equilibrium Generative Adversarial Networks," arXiv preprint arXiv:1703.10717, 2017.



- doi: <https://doi.org/10.48550/arXiv.1703.10717>
- [17] A. Radford, L. Metz, S. Chintala, "Unsupervised representation learning with deep convolutional generative adversarial networks," arXiv preprint arXiv:1511.06434, 2015.  
doi: <https://doi.org/10.48550/arXiv.1511.06434>
- [18] H. Zhang, I. Goodfellow, D. Metaxas, "Self-Attention Generative Adversarial Networks," Proceedings of the 36th International Conference on Machine Learning, pp.7354-7363, 2019.  
doi: <https://doi.org/10.48550/arXiv.1805.08318>
- [19] G. Donato, S. Belongie, "Approximate thin plate spline mappings," In European conference on computer vision, pp.21-31, 2002.  
doi: [http://dx.doi.org/10.1007/3-540-47977-5\\_2](http://dx.doi.org/10.1007/3-540-47977-5_2)
- [20] M. Heusel, H. Ramsauer, T. Unterthiner, B. Nessler, S. Hochreiter, "Gans trained by a two time-scale update rule converge to a local nash equilibrium," In Advances in Neural Information Processing Systems, pp.6626 - 6637, 2017.  
doi: <https://dl.acm.org/doi/10.5555/3295222.3295408>
- [21] J. Lin, Y. Pang, Y. Xia, Z. Chen, J. Luo, "Tuigan: Learning versatile image-to-image translation with two unpaired images," European Conference on Computer Vision. pp.18-35. 2020.  
doi: [http://dx.doi.org/10.1007/978-3-030-58548-8\\_2](http://dx.doi.org/10.1007/978-3-030-58548-8_2)
- [22] R. Zhang, P. Isola, A. A. Efros, E. Shechtman, O. Wang, "The unreasonable effectiveness of deep features as a perceptual metric," arXiv preprint arXiv:1801.03924, 2018.  
doi: <http://dx.doi.org/10.1109/CVPR.2018.00068>
- [23] T. C. Wang, M. Y. Liu, J. Y. Zhu, A. Tao, J. Kautz, B. Catanzaro, "High-resolution image synthesis and semantic manipulation with conditional gans," Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp.8798-8807, 2018  
doi: <http://dx.doi.org/10.1109/CVPR.2018.00917>
- [24] <https://github.com/tamarott/SinGAN>
- [25] <https://github.com/tohinz/ConSinGAN>
- [26] <https://github.com/assafshocher/InGAN>
- [27] <https://github.com/eliahuhorwitz/DeepSIM>

## 저 자 소 개



### 정 성 훈

- 2021년 : 부경대학교 신문방송학과 학사
- 현재 : 부경대학교 신문방송학과 대학원 석사과정
- ORCID : <https://orcid.org/0000-0002-5505-6189>
- 관심분야 : 인공지능, 딥러닝, 이미지 처리



### 공 경 보

- 2015년 : 서강대학교 전자공학과 학사
- 2017년 : 포항공과대학교 전자전기공학과 석사
- 2020년 : 포항공과대학교 전자전기공학과 공학박사
- 2021년 : 포항공과대학교 전자전기공학과 박사후연구원
- 현재 : 부경대학교 미디어커뮤니케이션학부 휴먼ICT융합전공 조교수
- ORCID : <https://orcid.org/0000-0002-1135-7502>
- 관심분야 : 멀티미디어 영상신호처리, 컴퓨터 비전, 딥러닝 시스템 설계