

# GAN 적대적 생성 신경망과 이미지 생성 및 변환 기술 동향

## Research Trends of Generative Adversarial Networks and Image Generation and Translation

조영주 (Y.J. Jo, run.youngjoo@etri.re.kr)

배강민 (K.M. Bae, kmbae@etri.re.kr)

박종열 (J.Y. Park, jongyoul@etri.re.kr)

시각지능연구실 연구원

시각지능연구실 연구원

시각지능연구실 책임연구원/실장

### ABSTRACT

Recently, generative adversarial networks (GANs) is a field of research that has rapidly emerged wherein many studies conducted shows overwhelming results. Initially, this was at the level of imitating the training dataset. However, the GAN is currently useful in many fields, such as transformation of data categories, restoration of erased parts of images, copying facial expressions of humans, and creation of artworks depicting a dead painter's style. Although many outstanding research achievements have been attracting attention recently, GANs have encountered many challenges. First, they require a large memory facility for research. Second, there are still technical limitations in processing high-resolution images over 4K. Third, many GAN learning methods have a problem of instability in the training stage. However, recent research results show images that are difficult to distinguish whether they are real or fake, even with the naked eye, and the resolution of 4K and above is being developed. With the increase in image quality and resolution, many applications in the field of design and image and video editing are now available, including those that draw a photorealistic image as a simple sketch or easily modify unnecessary parts of an image or a video. In this paper, we discuss how GANs started, including the base architecture and latest technologies of GANs used in high-resolution, high-quality image creation, image and video editing, style translation, content transfer, and technology.

**KEYWORDS** 딥러닝, 적대적 생성 신경망, 이미지 생성, 이미지 편집, 이미지 변환, GAN

\* DOI: <https://doi.org/10.22648/ETRI.2020.J.350409>

\* 본 연구는 과학기술정보통신부의 재원으로 정보통신기술진흥센터의 지원을 받아 수행한 연구임[No. B0101-15-0266, 실시간 대규모 영상 데이터 이해 예측을 위한 고성능 비주얼 디스커버리 플랫폼 개발].



본 저작물은 공공누리 제4유형

출처표시+상업적이용금지+변경금지 조건에 따라 이용할 수 있습니다.

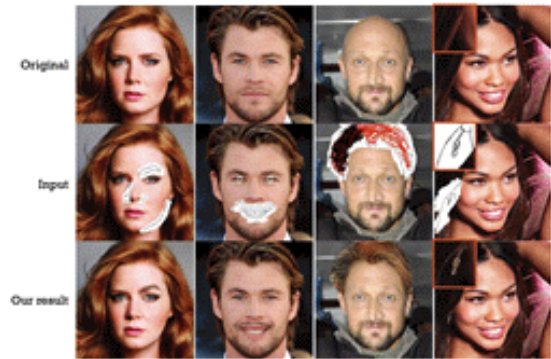
©2020 한국전자통신연구원

## I. 서론

GAN(Generative Adversarial Networks), 적대적 생성 신경망을 활용한 이미지 생성 및 변환 기술은 인공 신경망이 다양한 노이즈(Noise) 입력을 받아 원하는 카테고리의 기존에 존재하지 않는 새로운 이미지를 생성해내거나 입력 이미지나 비디오를 다른 형태나 정보를 지닌 이미지 또는 비디오로 변환하는 기술로 최근 몇 년간 급속도로 발전하여 많은 관심을 받고 활발하게 연구되고 있는 분야이다. 최근에는 이미지 편집을 많이 활용하는 디자인 분야를 겨냥하여 그림 1과 같이 이미지를 손쉽게 편집하거나 그릴 수 있는 프로그램으로 실제 서비스에 적용되고 있다.

GAN은 기존의 딥러닝(Deep learning) 기술에서 활용한 인공 신경망과는 다른 학습 방법을 활용하여 사람이 보기에 진짜와 구분하기 힘든 정도로 정교한 가짜 이미지를 만들어 내는 기술이다. 기존의 이미지 처리에서 많이 활용되는 딥러닝 기술은 학습데이터에 대하여 다층의 인공 신경망 하나를 학습시키는 방법을 활용하였지만, GAN은 2개의 인공 신경망의 상호작용을 활용하여, 최종적으로 사람이 보기에 진짜인지 구분하기 힘든 가짜 이미지를 만드는 1개의 생성 신경망을 실제 서비스에 활용한다. 하지만, 기존의 딥러닝 이미지 처리 방식과 다르게 성능을 객관적인 수치로 표시하는 방법이 부족하고, 4K와 같은 고화질 영상을 만들기에 고성능의 메모리가 필요하여 기술의 고화질 영상 생성 기술의 발전속도가 느린 것과 같은 문제가 있다. 따라서 앞으로 성능개선 및 성능평가 방법 개발 등 여전히 해결해야 할 문제가 많은 것으로 알려져 있다.

본 고에서는 최근 주목받고 있는 GAN의 시작부터 최신 기술의 연구 사례까지 살펴보고, 성능에



출처 Y. Jo et al., "SC-FEGAN: Face Editing Generative Adversarial Network with User's Sketch and Color," in CVPR, 2019.

그림 1 GAN의 이미지 편집/생성 활용사례

대해 살펴보도록 한다.

본 고의 II 장에서는 GAN에 대한 개괄적인 설명을 하고, III 장에서는 대표적으로 GAN의 활용 가능성을 보여준 연구 기술들에 대해 살펴본다. IV 장에서는 GAN을 활용한 이미지 생성/변화 기술의 최신 동향과 성능에 대해 살펴보고, V 장에서는 결론을 맺는다.

## II. 적대적 생성 신경망 개괄

GAN은 2014년에 발표된 논문 Generative Adversarial Networks[3]에서 처음 소개되었다. GAN은 한 개의 인공 신경망을 학습시켜서 활용하는 기존의 방법과는 다르게 두 개의 인공 신경망을 활용한다. 하나는 이미지를 생성하는 생성(Generative) 신경망, 다른 하나는 생성 신경망이 만든 이미지를 진짜인지, 가짜인지 판별하는 판별(Discriminative) 신경망이다. 생성 신경망은 판별 신경망이 자신이 만든 이미지를 진짜 이미지로 판별하도록 학습되어야 하고, 판별 신경망은 생성 신경망이 만든 이미지가 입력으로 들어오면 가짜라고 판별하고 실제 이미지가 입력으로 들어오면 진짜라고 판별하도록 학습되어야 한다. 따라서 이 두 개의 신경망

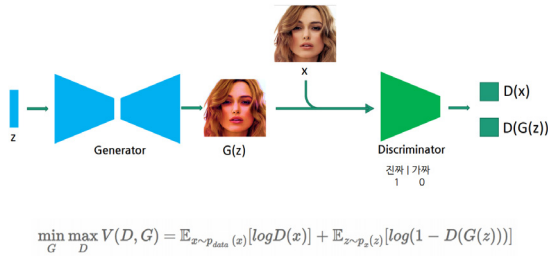


그림 2 GAN의 구조 및 학습 과정

은 서로 적대적인 관계에 있으므로 이를 학습시키는 방법을 이른바 adversarial training이라고 한다. Adversarial training은 기본적으로 학습하고자 하는 데이터가 레이블(Label)이 존재하는 지도학습이 아닌 레이블이 존재하지 않는 비 지도학습 방법이다. 생성 신경망은 노이즈를 입력으로 받아 다수의 층(Layer)을 통과하면서 특징맵(Feature map)을 확장시켜 나가는 구조로 되어 있다. 마지막 층을 통과하여 나오는 특징맵 크기가 실제 원하는 이미지의 크기와 같도록 만드는 것이다. 판별 신경망은 반대로 특징맵의 크기를 줄여나가는 기준에 많이 활용되는 전통적인 구조의 인공 신경망으로 구성되어 있다. 따라서 판별 신경망은 생성 신경망이 만들어진 이미지나 실제 이미지 데이터셋(Dataset)의 이미지를 입력으로 받아서 최종적으로 진짜 이미지인지 아니면 생성 신경망이 만든 가짜 이미지인지 판별한다(그림 2).

이때 학습의 방향은 생성 신경망과 판별 신경망이 서로 반대가 된다. 생성 신경망은 자신이 만든 최종 결과가 판별 신경망이 진짜 이미지라고 판별하도록 학습 방향이 진행되고, 판별 신경망은 반대로 생성 신경망이 만든 결과가 가짜 이미지라고 판별하도록 학습 방향이 진행되게 된다. 이러한 adversarial training이 충분히 진행되면 최종적으로는, 판별 신경망이 생성 신경망이 만든 이미지가 진짜인지 가짜인지를 알아맞히게 되는 확률이 약 50%

가 되고, 따라서 생성 신경망이 만든 이미지는 진짜와 거의 구분되지 않는 이미지가 된다.

이러한 구조에서 두 가지의 데이터셋을 준비하고 하나의 데이터셋을 생성 신경망의 입력으로 넣어주면 하나의 데이터셋 입력에 대하여 다른 데이터셋의 특징을 가지는 결과를 만들어 내는, 이른바 이미지 변환 기능을 하는 생성 신경망을 학습시킬 수 있게 된다.

다음 장에서는 대표적으로 이러한 GAN의 활용 가능성을 제시한 연구 사례에 대해 상세히 살펴본다.

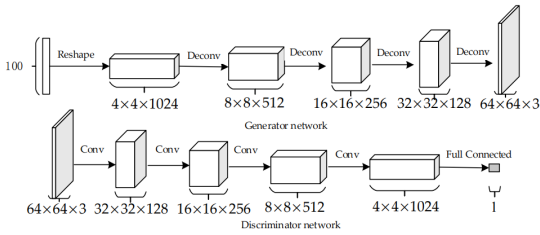
### III. 이미지 생성/변환 기술의 시작

본 장에서는 GAN의 실제 활용 가능성을 보여준 기술 사례들을 살펴보도록 한다. 먼저 대표적인 노이즈 입력을 받아서 이미지를 만드는 기술들을 살펴보고, 입력을 조절하여 결과물의 방향성을 조절하는 방법을 살펴본다. 다음으로, 이미지 입력을 받아서 다른 형태의 이미지로 변환하는 이미지 변환에 사용되는 대표적인 기술을 살펴본다.

#### 1. 노이즈를 이미지로 변환

##### 가. DCGAN

Ian Goodfellow가 2014년 처음으로 GAN을 발표하고 많은 연구가 이루어졌지만, 기존의 단일 네트워크를 학습시키던 방법과는 다른 구조 때문에 학습이 불안정하다는 문제가 야기되어왔다. 이에 많은 연구가 GAN의 불안정한 학습을 해결하고자 많은 노력을 하였다. 이러한 문제에 대하여 2016년 Google에서는 DCGAN[4]을 발표하여 이러한 불안정성 문제를 해결한 연구를 발표하였다. DCGAN은 현재 개발되고 있는 GAN 구조의 기초가 되는



출처 J. Jhang et al., "An Efficient Hyperspectral Image Retrieval Method: Deep Spectral-Spatial Feature Extraction with DCGAN and Dimensionality Reduction Using t-SNE-Based NM Hashing," Remote Sens., vol. 10, 2018, CC BY 4.0.

그림 3 DCGAN의 생성 신경망 구조

구조이다.

기존의 지도학습에서 이미지 처리를 위한 딥러닝에서 컨볼루션 구조를 이용하여 안정성과 성능을 잡았던 것과 비슷한 방법으로 그림 3과 같이 GAN의 생성 신경망 구조에서 노이즈를 확장시키는 데 있어서 미분이 불가능한 모든 층을(max pooling과 같은) 모두 미분이 가능한 컨볼루션으로 대체 구성하여 학습의 안정성과 성능을 높이는 방법을 제시하였다.

이러한 구조를 통하여 만든 이미지는 기존에 많은 시도를 했던 GAN보다 월등히 좋은 성능을 보여주었으며, 나아가 생성 신경망이 만들어 내는 이미지를 조절할 수 있음을 보여주었다. 예를 들어, 그림 4에서처럼 안경을 쓴 남자의 이미지에서 안경을 쓰지 않은 남자의 이미지 정보를 제거하고 안경을 쓰지 않은 여자의 이미지 정보를 추가하면, 안경을 쓴 여자의 이미지를 만들어 낼 수 있다는 이미지 연산이 가능함을 보였다.



그림 4 DCGAN의 이미지 연산을 통하여 생성된 이미지

DCGAN은 이러한 결과뿐만 아니라 GAN에서 학습된 특징들이 어떻게 신경망에 저장되어 있는지를 보여주고 이를 분석하여, 입력 노이즈의 latent space를 변화시켜서 원하는 방향으로 결과 이미지를 조절할 수 있다는 것을 보여주었다. 이러한 DCGAN은 GAN의 안정적인 학습과 결과분석에 대한 기초가 되는 연구가 되었다.

### 나. BEGAN

DCGAN에 이어 Google은 2017년 GAN을 학습시키는 손실 함수를 이미지 데이터 분포에 초점을 맞추지 않고 손실 함수 자체의 분포에 초점을 맞추어 변형시키는 방법으로 다시 한번 GAN의 성능을 끌어올리는 BEGAN[5]을 발표하였다. 신경망의 구조가 단순하여도 생성 신경망과 판별 신경망끼리의 학습 진행 정도의 차이가 생성 신경망이 만들어 내는 이미지의 성능에 얼마나 큰 영향을 미치는지를 보여주었다. 또한, 처음으로 그림 5와 같이 128x128 사이즈의 이미지를 만들어 내었으며, 당

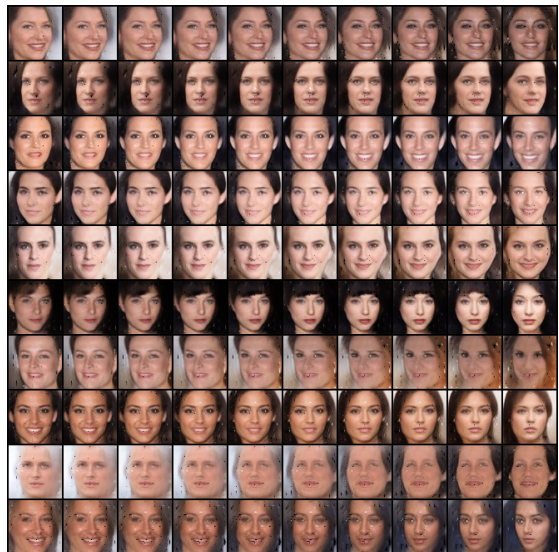


그림 5 BEGAN의 생성 이미지 결과들



시의 GAN의 이미지 성능 중 가장 뛰어난 모습을 보여주었다.

이를 기반으로 생성 신경망과 판별 신경망 사이의 학습 균형을 조절하고자 하는 연구가 많이 진행되었으며, 최신 연구들은 이를 해결하여 더욱 높은 해상도와 높은 품질의 이미지를 만들어 내고 있다.

## 2. 이미지를 이미지로 변환

### 가. Pix2Pix

Pix2Pix[6]는 기존의 노이즈 입력을 이미지로 변환하여 학습 데이터셋의 분포를 가진 이미지를 생성해내는 GAN을 넘어서 이미지를 입력으로 받아 다른 형태의 이미지로 변환하는 GAN을 제시하였다. 이미지를 생성하는 생성 신경망은 U-net 구조를 활용하고, 판별 신경망은 입력 이미지와 다른 콘텐츠의 생성 이미지를 쌍으로 입력으로 받아 판별하게 된다. 판별 신경망도 전체 이미지를 한 번에 판별하는 방법이 아닌 Patch별로 판별하는 PatchGAN 형태의 판별 신경망을 활용하였다(그림 6).

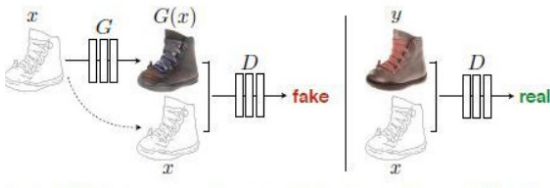
생성된 이미지를 픽셀 단위로 판별하는 L1 loss와 GAN 학습에 사용되는 loss를 함께 사용함으로써 이미지의 변환을 자연스럽게 할 수 있음을 보여주었다. Pix2Pix의 이러한 구조는 그림 6과 같이 입력 이미지와 타겟 이미지가 쌍으로 존재하는

거의 모든 데이터셋에 대하여 활용 가능하며, 이미지 변환을 활용하는 거의 모든 연구에 기초가 되고 있다.

## IV. 이미지 생성/변환 기술의 동향

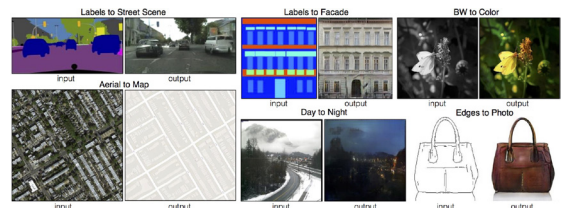
최근 적대적 생성 신경망을 활용한 이미지 생성 및 변환 기술에 관한 연구가 더욱 활발해지고 있다. 초기의 GAN은 noise와 같이 의미 없는 정보에서 다양한 이미지를 생성하는 기술이 대부분이었다. 이러한 초기 GAN 모델들은 2가지의 문제점이 있었다. 첫 번째는 생성한 이미지의 해상도가  $128 \times 128$ 에 불과했다는 것이다. 두 번째는 noise로부터 생성된 이미지는 생성이 완료되기 전까지 어떤 모습일지 전혀 예측할 수 없었다.

최근의 GAN은 이러한 두 가지 문제를 해결하는 방향으로 연구가 활발히 진행되고 있다. 과거 사람 얼굴 데이터셋(CelebA[7])에 대해 학습하여 저해상도( $128 \times 128$ )의 사람 얼굴 이미지를 만들어 냈다면 이제 고해상도( $1,024 \times 1,024$ )의 얼굴 이미지를 만들어 내고 있다. 그리고 과거 생성될 사람의 얼굴이 남자일지, 여자일지 또는 검정 머리일지 노란 머리일지 전혀 예측할 수 없었지만 현재는 이것을 조절할 수 있는 Conditional image generation 방법들이 연구되고 있다. 이러한 연구는 CelebA 데이터셋에 한정되지 않고 다른 여러 데



출처 <https://github.com/phillipi/pix2pix>, BSD License.

그림 6 Pix2Pix의 학습 방법



출처 <https://github.com/phillipi/pix2pix>, BSD License.

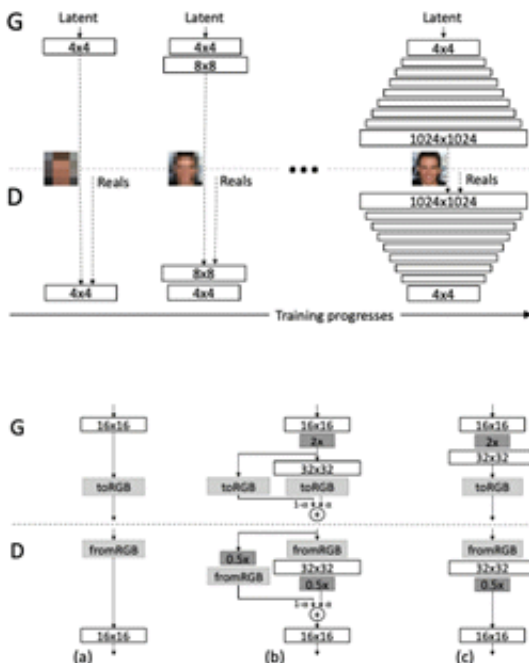
그림 7 Pix2Pix의 이미지 변환 결과들

이터셋으로 확장하려는 노력이 계속되고 있다. 그래서 이번 기술 동향 분석에서는 GAN을 이용해 고해상도의 이미지를 만들어내는 연구와 조건에 따라 이미지를 생성하고 변환하는 이미지 변환 기술을 소개한다.

## 1. 이미지 생성 기술 동향

### 가. ProGAN

고해상도에 대해서는 명확하게 기준이 정해져 있지는 않다. 하지만 GAN 연구에서 일반적으로 최소 512×512 이상의 이미지에 대해서 고해상도라고 이야기하고 있다. ProGAN[8]에서는 고해상도의 얼굴 이미지를 생성하기 위해 초저해상도(4×4)에서부터 차근차근 이미지를 생성하는 방법



출처 [https://github.com/tkarras/progressive\\_growing\\_of\\_gans](https://github.com/tkarras/progressive_growing_of_gans), CC BY-NC 4.0.

그림 8 ProGAN의 네트워크 및 학습 구조



출처 [https://github.com/tkarras/progressive\\_growing\\_of\\_gans](https://github.com/tkarras/progressive_growing_of_gans), CC BY-NC 4.0.

그림 9 ProGAN의 실험 결과(1,024×1,024)

을 학습하여 고해상도 이미지를 생성하는 progressive growing이라는 방법을 제시하였다.

ProGAN[2]에서는 그림 8의 위에서처럼 4×4의 해상도에서 이미지를 생성하여 차근차근 생성하는 이미지의 크기를 키워 간다. 이때 Generator(G)와 Discriminator(D)에 은닉 계층을 새롭게 추가하여 커진 해상도에 대처할 수 있도록 네트워크의 복잡도를 높여준다. 최종적으로 목표하는 1,024×1,024의 해상도에 도달하면 학습이 종료된다.

은닉 계층을 추가할 때는 네트워크가 천천히 적응할 수 있도록 하는 것이 중요하다. 그래서 새로운 은닉 계층을 통과하기 전의 값과 통과한 이후 값에 그림 8의 아래처럼  $\alpha$ 라는 가중치를 두어 합하는 방식으로 새로운 은닉 계층을 학습시켰다. 이때  $\alpha$ 는 학습 과정에 따라 계속 값을 키워 그 값이 1이 되면 그 이전 새로운 은닉 계층에 대한 적응이 끝나게 된다. 최종적으로 그림 9와 같이 1,024×1,024 사이즈의 얼굴 이미지 생성 결과를 얻을 수 있게 된다.

과거 BEGAN[5], EBGAN[9], DCGAN[4] 등에서 겨우 128×128 크기 해상도의 얼굴 이미지 생성이 겨우 가능했는데, 최근 생성되는 얼굴 이미지의 해상도가 1,024×1,024인 것을 보면 그 크기가 비약적으로 증가하였음을 알 수 있다.

## 나. SinGAN

ProGAN[8]에서의 progressive growing을 응용하여 SinGAN[10]에서는 좀 더 다양한 것들을 수행하고자 하였다. SinGAN[10]에서는 그림 10의 위와 같이 단계별로 다른 G를 사용하였다.

SinGAN[10] 그림 10의 아래와 같이 각 스케일에서 이미지를 생성할 때 noise를 추가하여 artifact를 제거하는 효과를 얻었다. 그리고 가중치를 없애고 각 step에서 이미지 크기를 2배로 키우지 않고 4/3 배씩 이미지 크기를 증가시키는 방법을 선택하였다. 그 결과 고해상도의 이미지도 성공적으로 생성할 수 있다.

SinGAN[10]은 ProGAN[8]에서 제안한 방법과 유사하게 progressive growing을 활용하지만, 이미지 생성 이외에 변형, 조작, Super resolution, 동영상 생성 등으로 활용했다는 점에서 차이가 있다. 그림 11은 SinGAN[10]을 이용하여 여러 활용을 하였을 때 그 결과를 나타낸다. 그림 11의 위는 고해상도 이미지 생성이 가능하다는 것을 보여주고 그림 11의 아래는 이미지 생성, 변형, 조작, super



출처 <https://github.com/tamarott/SinGAN>, MIT License.

그림 11 SinGAN에서의 실험 결과

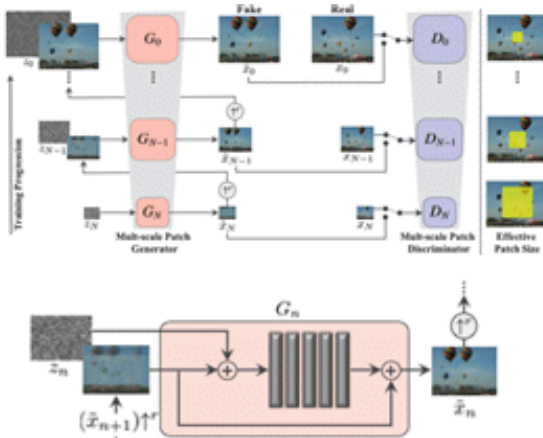
resolution 및 동영상 클립 생성까지도 가능하다는 것을 보여준다.

## 2. 이미지 변환 기술 동향

최근 GAN을 이용한 이미지 변환 기술이 많이 연구되고 있다. 초기의 이미지 변환 기술과 달리 생성된 이미지에 조건을 주고자 하는 연구들이 많이 발표되고 있다. 이때 사용되는 조건으로는 이미지, 글자 등이 있다. 여기서는 GAN을 이용한 이미지 변환 기술의 연구 동향에 대해 알아본다. 이미지 변환 기술에서는 주로 조건을 이미지와 같은 3차원 텐서로 받는 경우가 많다. 이때 3차원 텐서로는 이미지, segmentation map, edge 등이 있을 수 있으면 차례로 최신 이미지 변환 기술 동향을 알아본다.

### 가. Pix2PixHD

Pix2PixHD[11]는 대표적인 이미지 변환 연구로 edge 정보 또는 segmentation map을 이용해 이미지를 생성하는 연구이다. 이때 segmentation map이란 그림 12의 좌측 하단과 같이 각 픽셀의 label이 무엇



출처 <https://github.com/tamarott/SinGAN>, MIT License.

그림 10 SinGAN의 네트워크 구조도

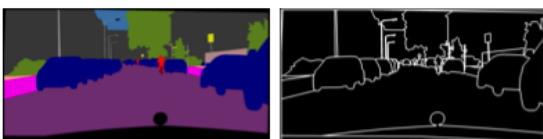


출처 <https://github.com/NVIDIA/pix2pixHD>, BSD License.

그림 12 Pix2PixHD 실험 결과

인지 표기해둔 이미지를 의미한다. 그리고 edge 정보란 이미지의 색 정보를 제거하고 남은 edge를 의미하며, 그림 12의 상단과 같이 사람의 얼굴 또는 고양이의 edge 정보만을 이용하여 이미지를 복원할 수 있다.

Pix2PixHD[11]는 기존의 Pix2Pix[6]와 다르게 instance 정보를 추가로 사용하였다. 기존의 segmentation map은 두 개의 동일한 물체가 겹쳐져 있을 경우 그 차이를 표시할 수 없다는 문제가 있었다. 그래서 그림 13에서와 같이 좌측의 segmentation map뿐만 아니라 우측의 boundary map을 같이 입력으로 주었다.



출처 <https://github.com/NVIDIA/pix2pixHD>, BSD License.

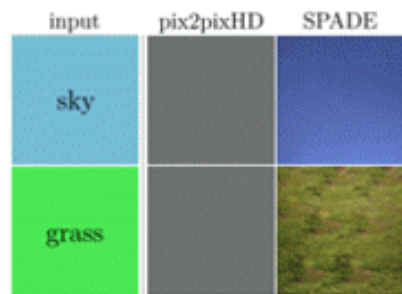
그림 13 Pix2PixHD의 입력

## 나. SPADE

Pix2PixHD[11]에서는 겹쳐진 물체에 대해서도 비교적 정확하게 이미지를 생성할 수 있게 되었다. 하지만 SPADE[2]에서는 segmentation map에서 배치 정규화 계층(Normalization layer)이 잘못되었다고 주장했다.

딥네트워크에서 일반적으로 배치 정규화 계층은 중요한 것으로 여겨지며 배치 정규화 계층 유무에 따라 성능의 차이가 크게 나는 경우도 있다. 이때 딥네트워크에서의 배치 정규화 계층이란 이전 은닉 계층의 특징 벡터의 평균과 분산을 계산한 다음 정규화 시켜주는 과정을 의미한다. 이때 그림 14의 위와 같이 동일한 uniform segmentation map이 입력으로 들어오게 될 경우 기존의 배치 정규화 계층은 모두 동일한 결과를 만들어 하늘과 풀의 texture 정보를 전혀 살리지 못하게 된다. 그래서 그림 15와 같이 새로운 배치 정규화 계층을 만들었다.

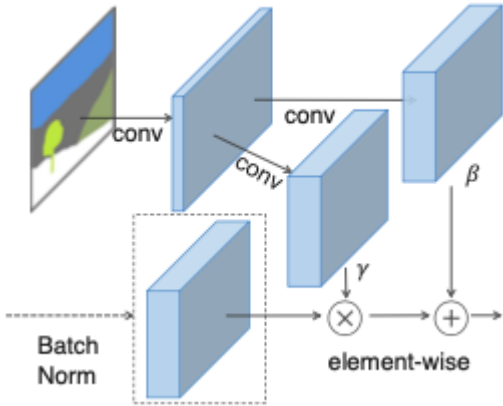
SPADE[2]는 이러한 새로운 배치 정규화 계층을 통해 segmentation map에서 큰 부분이 동일한 label을 가지고 있을 때 texture 정보를 잃어버리는 문제를 보완하였다. 그 결과 그림 7과 같이 다른 알고리



출처 <https://github.com/NVlabs/SPADE>, CC BY-NC 4.0.

그림 14 SPADE 실험 결과





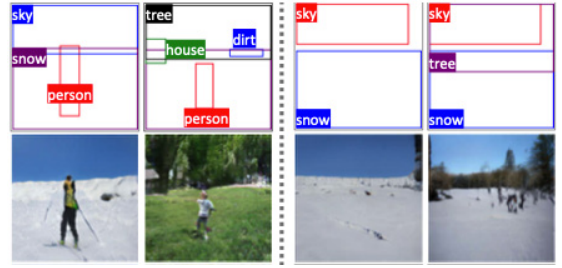
출처 <https://github.com/NVlabs/SPADE>, CC BY-NC 4.0.

그림 15 SPADE 구조도

즘에 비해 좋은 결과를 얻을 수 있었다.

#### 다. LostGAN

최근 이미지 변환 연구에서 layout을 이용해 이미지를 생성하는 연구도 수행되고 있다. 여기서 layout이란 segmentation map과 다르게 bounding box만을 이용해 물체가 어디에 있는지 명시만 해 준다. 즉, 물체의 모양에 대한 정보는 전혀 제공되지 않는다. 따라서 학습하는 네트워크는 layout을 이용해 이미지를 만들어내는 것과 동시에 물체의 모양도 동시에 예측해야 하기 때문에 seg-



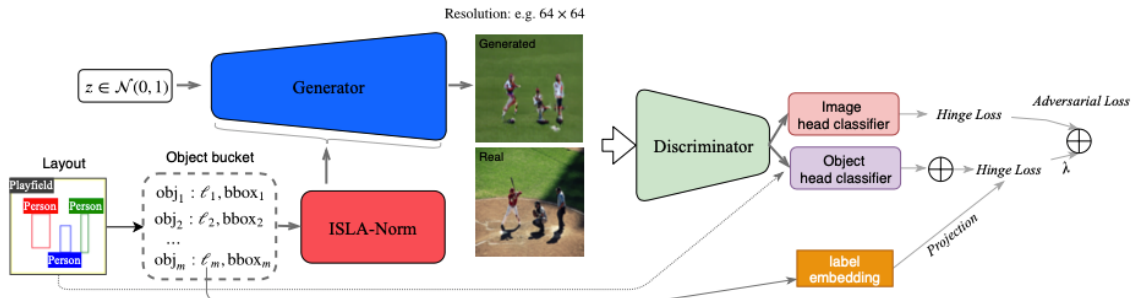
출처 Reprinted with author's permission from W. Sun et al., "Image Synthesis From Reconfigurable Layout and Style," in ICCV, 2019.

그림 16 LostGAN 실험 결과

mentation map을 이용한 이미지 생성보다 어려운 task로 여겨진다. LostGAN[12]에서는 layout으로부터 이미지를 생성하는 연구를 수행하였다. 그림 16은 LostGAN[12]의 실험에 관한 결과이다.

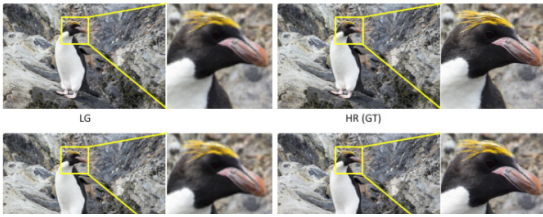
Layout에서 이미지를 생성하는 연구는 객체의 모양이 주어지지 않아 모양을 먼저 예측할 필요가 있었다. 따라서 LostGAN[12]은 ISLA Norm이라는 새로운 정규화 계층을 이용하여 모양을 예측하고, 이를 사용하여 이미지 생성에 활용하였다.

LostGAN은 그림 17에서와 같이 구성된다. ISLA Norm은 그림 17에서와 같이 layout의 bounding box 좌표, label 등의 정보를 이용하는 정규화해 주는 계층을 제안하였다. 그 결과 그림 16처럼 layout만을 이용하여 이미지를 생성할 수 있게 되었다. 이



출처 Reprinted with author's permission from W. Sun et al., "Image Synthesis From Reconfigurable Layout and Style," in ICCV, 2019.

그림 17 LostGAN 구조도



출처 <https://github.com/tensorlayer/srgan>

그림 18 SRGAN 실험 결과

외에도 LSTM을 이용하여 물체들 간의 관계를 파악하려고 시도한 layout2image[13]와 같은 연구들이 있다.

## 라. SRGAN

Super resolution은 과거부터 오랫동안 연구되어 오던 task다. 최근 GAN 알고리즘의 발전으로 super resolution에 이를 적용하려는 연구도 많아지고 있다. Super resolution이란 저해상도의 이미지를 고해상도로 복구하는 작업이다. 과거부터 super resolution에서는 fine texture를 복원하는 데 어려움이 있었다. 따라서 SRGAN[14]은 GAN 알고리즘을 이용하여 fine texture 정보를 유지하려고 노력하였다.

SRGAN[14]은 texture 정보를 복원하기 위해 기존의 방법들과 다르게 perceptual loss라는 것을 제안하였다. Perceptual loss는 super resolution 후 이미지

의 픽셀 값을 단순히 비교하지 않고 딥네트워크가 추출한 feature space에서 비교하는 loss이다. 그 결과 그림 19처럼 PSNR은 낮더라도 사람이 보기에 이상하지 않게 fine texture 정보를 잘 복원할 수 있게 되었다. 이는 과거 딥네트워크만을 사용한 참고문헌 [15]보다 시각적으로 좋은 결과를 얻었다.

## 마. Deepfill

Deepfill[16]은 GAN을 이용하여 image inpainting을 수행하기 위해 연구된 방법이다. Image inpainting은 이미지에서 사라진 부분을 복원하는 task를 의미하는데, 이때 사라진 부분은 단순히 유실된 것일 수도 있지만, 사용자가 이미지 편집을 위해 의도적으로 제거할 수도 있다. 만약 사용자가 사진에서 자신 이외에 다른 행인이 같이 찍혔을 경우 image inpainting을 이용해 불필요한 배경을 제거할 수 있다. 즉, 포토샵과 같은 이미지 편집 툴을 익혀 전문가가 되지 않아도 누구나 손쉽게 이미지를 편집할 수 있게 된다. 그림 19의 경우 Deepfill[16]의 실험 결과를 나타낸다. 하얀색 사각형을 이용하여 사용자가 이미지의 지우고 싶은 부분들에 대해 표시하였고 학습된 네트워크가 이를 복원한다.

Deepfill[16]는 최근 gated convolution을 사용하여 image inpainting 결과를 발전시켜 Deepfillv2[17]를 발표하였다. 그 이외에 사용자의 의도를 더 명확하게 반영하여 image inpainting을 수행하는 연구도 발표되었다. 특히 스케치나 색에 맞춰 이미지를 복원하는 SC-FEGAN[1]이나 edge 정보를 이용하려고 한 edge-connect[18]와 같은 연구들도 최근 많이 발표되고 있다.

## 바. StyleGAN

StyleGAN[19]는 ProGAN[8]의 발전된 기술이다. 특히 사람의 머리 색, 나이, 성별 등을 네트워크가



출처 Reprinted with permission from [https://github.com/JiahuiYu/generative\\_inpainting](https://github.com/JiahuiYu/generative_inpainting)

그림 19 Deepfill 실험 결과

출처 Reprinted with permission from <https://github.com/NVlabs/stylegan>

그림 20 StyleGAN 실험 결과

분석하여 자유롭게 변형할 수 있다는 장점이 있다. 특히 그림 20에서는 각 열에서 비슷한 성별과 나이를 가지고 각 행에서 비슷한 머리색과 피부색을 가지도록 한 실험 결과이다.

StyleGAN[19]에서는 AdaIN[20]에서 영향을 받은 정규화 레이어를 이용해 사람의 성별, 머리색, 나이대와 같은 style 정보를 네트워크에 알려줄 수 있도록 하였다. 하지만 StyleGAN[19]에서는 이미지를 생성할 때 결과에 이상한 무늬가 발생하는 현상이 있었다. 이를 보완하고자 StyleGAN2[21] 정보를 발표하여 과거 이미지 생성에서 발생하던 문제들을 보완하였다.

### 사. StarGAN

StarGAN[22]에서는 한 개의 모델을 이용하여 내가 원하는 부분에 대해서만 이미지 변환을 수행할 수 있게 된다. 그림 21에서 만약 사용자가 사람의 머리색 또는 피부색과 같이 단 한 개만 변화시키고 싶을 때 사용할 수 있는 기술이다.

이처럼 단 한 개의 네트워크만을 이용해 여러 이미지 변환을 시도할 수 있어 기존의 방법들과 비교하여 효과적이다. 최근 사람의 얼굴뿐만 아니라 동물 얼굴 사진을 이용해 실험한 StarGAN v2[23]가

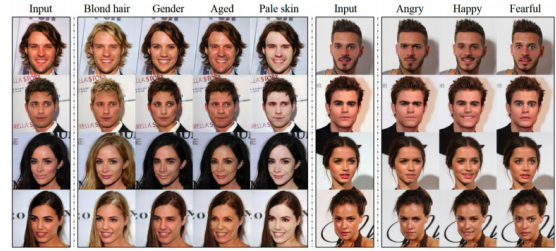
출처 <https://github.com/yunjey/stargan/blob/master/LICENSE>, MIT License.

그림 21 StarGAN 실험 결과

최근에 발표되었다.

## V. 결론

본 고에서는 최근 GAN을 이용하여 연구되고 있는 많은 기술들과 활용 방안들의 시작부터 최신 기술까지의 동향에 대해 살펴보았다. 기존의 데이터를 분석 및 분류하는 딥러닝을 넘어서 데이터를 만들고 조작하는 부분에서 GAN은 높은 성능을 보여주어 다양한 분야에서 딥러닝이 활용될 수 있음을 보여주었다. 특히 오늘날, 사용자의 의도를 파악하고 고해상도의 이미지를 생성하는 기술들은 보는 사람을 놀라게 한다. 최근 딥러닝을 이용한 이미지 처리 학회들을 참석해 보면 GAN을 이용한 이미지 기술의 발전속도가 매우 빠르다. 사람이 구분하기 힘들 정도의 성능으로 인하여 기술 악용에 대한 문제도 제기되고 있으며, 따라서 향후에는 지금보다 더 높은 해상도의 이미지를 GAN을 통해 만들어 내고 이를 실제 이미지와 구분하는 새로운 구조와 기술이 등장할 것이다.

### 용어해설

**Fine texture** 이미지에서 고주파 영역으로 사람의 머리카락이나 동물의 털과 같이 매우 섬세한 표현이 필요한 부분들을 의미

## 약어 정리

GAN	Generative Adversarial Networks
SR	Super Resolution

## 참고문헌

- [1] Y. Jo et al., "SC-FEGAN: Face Editing Generative Adversarial Network with User's Sketch and Color," in CVPR, 2019.
- [2] T. Park et al., "Semantic Image Synthesis with Spatially-Adaptive Normalization," in CVPR, 2019.
- [3] I. Goodfellow et al., "Generative adversarial nets," in NIPS, 2014.
- [4] A. Radford et al., "Unsupervised Representation Learning With Deep Convolutional Generative Adversarial Networks," in ICLR, 2016.
- [5] D. Berthelot et al., "BEGAN: Boundary Equilibrium Generative Adversarial Networks," in arXiv, 2017.
- [6] P. Isola et al., "Image-to-Image Translation with Conditional Adversarial Nets," in CVPR, 2017.
- [7] Z. Liu et al., "Deep Learning Face Attributes in the Wild," in ICCV, 2015.
- [8] T. Karras et al., "Progressive Growing of GANs for Improved Quality, Stability, and Variation," in ICLR, 2018.
- [9] J. Zhao et al., "EBGAN: Energy-Based Generative Adversarial Networks," in ICLR, 2017.
- [10] T. R. Shaham et al., "SinGAN: Learning a Generative Model from a Single Natural Image," in ICCV, 2019.
- [11] T. Wang et al., "High-Resolution Image Synthesis and Semantic Manipulation with Conditional GANs," in CVPR, 2019.
- [12] W. Sun et al., "Image Synthesis From Reconfigurable Layout and Style," in ICCV, 2019.
- [13] Z. Bo et al., "Layout2Image: Image Generation from Layout," in IJCV, 2020.
- [14] C. Ledig et al., "Photo-Realistic Single Image Super-Resolution Using a Generative Adversarial Network," in CVPR, 2017.
- [15] C. Dong et al., "Image Super-Resolution Using Deep Convolutional Networks," in TPAMI, 2015.
- [16] J. Yu et al., "Generative Image Inpainting with Contextual Attention," in CVPR, 2018.
- [17] J. Yu et al., "Free-Form Image Inpainting with Gated Convolution," in CVPR, 2019.
- [18] K. Nazeri et al., "EdgeConnect: Structure Guided Image Inpainting using Edge Prediction," in ICCVW, 2019.
- [19] T. Karras et al., "A Style-Based Generator Architecture for Generative Adversarial Networks," in CVPR, 2019.
- [20] X. Huang et al., "Arbitrary Style Transfer in Real-time with Adaptive Instance Normalization," in ICCV, 2017.
- [21] T. Karras et al., "Analyzing and Improving the Image Quality of StyleGAN," in arXiv, 2019.
- [22] Y. Choi et al., "StarGAN: Unified Generative Adversarial Networks for Multi-Domain Image-to-Image Translation," in CVPR, 2018.
- [23] Y. Choi et al., "StarGAN v2: Diverse Image Synthesis for Multiple Domains," in CVPR, 2020.