

# Image To Image Translation 기술 동향 분석

신지혜 (21110689@sju.ac.kr) 세종대학교 인공지능학과 석사과정

## ABSTRACT

본 리포트는 GAN(Generative Adversarial Networks)을 이용한 Image to Image Translation에 관한 연구이다. GAN은 생성 모델과 판별 모델이라는 두 개의 심층 네트워크로 구성되어 있으며, 경쟁적인 방식으로 훈련된다. 심층 네트워크 구조와 경쟁적인 훈련으로 인해 GAN은 합리적이고 현실적인 이미지를 생성할 수 있으며 많은 이미지 합성 및 편집 응용 프로그램에서 뛰어난 기능을 보여준다. 본 리포트에서는 Image To Image Translation 주제에 관한 최근 GAN의 기술 동향을 분석한다.

## 1. 서론

인터넷과 영상 기술의 발전으로 많은 양의 이미지를 쉽게 사용할 수 있게 되었다. 영상에서 원치 않는 물체를 제거하고, 영상의 색을 조절하고, 영상을 예술품으로 바꾸거나 예술품을 영상으로 바꾸는 것과 같이 영상을 합성 및 편집하는 것이 필요한 작업의 필요가 점점 증가하고 있다. 이러한 작업들 중 가장 많이 사용되는 것은 입력 영상을 이에 상응하는 다른 이미지로 변환하는 "Image to Image Translation (I2IT)"이다. I2IT 작업을 간단하게 설명하자면, A 도메인의 영상을 B 도메인의 영상으로 변환시키는 작업을 의미하며 이 작업을 위해 많은 방법이 제안되었지만, 여전히 어려운 문제로 남아 있다.

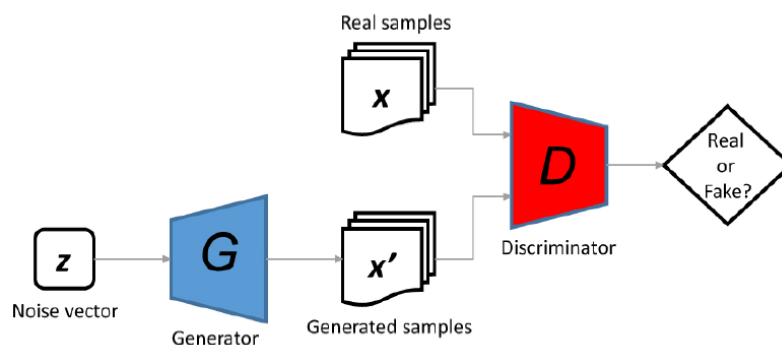
GAN의 개념과 GAN의 확장된 형태에 대해서 알아보고, GAN을 Image to Image Translation에 적용한 방법들에 대해서 살펴본다. 설명된 방법들 중 Pix2pix, BicycleGAN, SPA-GAN, CE-GAN은 Supervised Image to Image Translation에 속하며, CycleGAN, AIGN, DualGAN, UNIT, MUNIT은 Unsupervised Image to Image Translation에 속한다. Supervised Image to Image Translation의 학습에 경우, align이 맞는 한 쌍의 이미지 셋을 사용하여 입력 이미지와 출력 이미지 간의 매핑을 학습하는 반면 Unsupervised Image to Image Translation의 학습에서는 source 및 target 이미지 셋이 쌍으로 이루어지지 않은 것을 사용합니다.

## 2. GAN (Generative Adversarial Network)

2014년에 Ian J. Goodfellow 외 연구진[1]에 의해 GAN이 제안되었다. GAN은 출력 이미지가 진짜인지 가짜인지 분류하는 판별 모델을 학습하는 동시에 가짜를 진짜로 만들기 위한 생성 모델을 학습을 진행한다. 처음 이 방법을 제안한 연구진은 GAN의 구조를 경찰과 위조지폐범

사이의 게임으로 비유하였다. 생성 모델은 가짜 화폐를 생산해 적발되지 않고 사용하려는 위조지폐범과 유사하고, 판별 모델은 진짜 화폐와 가짜 화폐를 완벽히 판별하여 위조지폐범을 검거하는 것을 목표로 하는 경찰과 유사하다. 경찰과 위조지폐범은 모두 성공적으로 생성 및 판별을 하기 위해 경쟁적으로 각자의 방법을 개선하려고 하는 이러한 대립 프로세스가 게임과 같다고 할 수 있다.

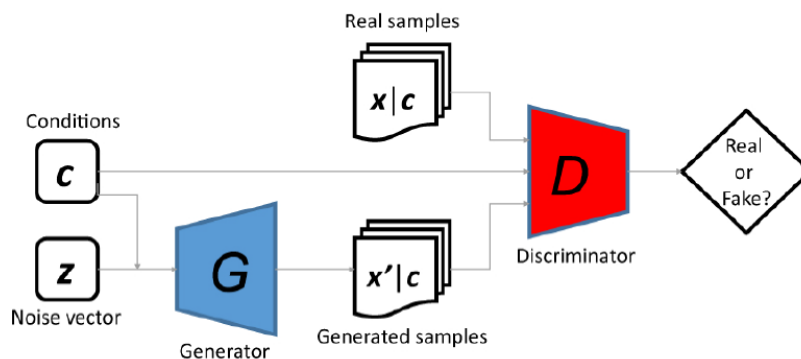
Variational AutoEncoders (VAEs)[2]와 같은 다른 생성 모델과 비교할 때 GAN에 의해 생성된 이미지는 일반적으로 덜 흐리고 더 사실적이다. 또한 최적의 GANsexist, 즉 생성 모델이 실제 이미지의 분포와 잘 일치하는 이미지를 완벽하게 생성한다는 것이 이론적으로 입증되었으며 판별 모델은 항상 1/2을 생성한다. 네트워크의 수렴이 이루어지기 어려우며, 종종 생성 모델이 다른 noise vector  $z$ 에 대한 똑같거나 비슷한 이미지를 생성하는 "mode collapse"에 빠지는 등의 이유로 인해 GAN을 실제로 학습시키는 것은 어렵다. 이러한 학습의 안정성을 향상시키기 위해 GAN의 다양한 확장이 제안되었다.



[그림 1] GANs의 구조

## 2.1 cGANs (conditional Generative Adversarial Networks)

Mirza and Osindero 연구진[3]은 GANs를 conditional model로 확장한 cGANs를 소개한다. cGANs에서는 생성 모델  $G$ 와 판별 모델  $D$ 는 몇 가지 추가 정보  $c$ 에 따라 조절되는데, 추가 입력으로  $c$ 를  $G$ 와  $D$  모두에 입력한다. 추가 정보는 class label과 text 그리고 sketches 등이 될 수 있다. cGANs는 생성 중인 데이터의 종류에 대한 추가 제어 기능을 제공하는 반면 원래 GAN에는 이러한 제어 기능이 없다. cGANs는 이미지 합성 및 이미지 편집 애플리케이션에 널리 사용된다.



[그림 2] cGANs의 구조

## 2.2 DCGANs (Deep Convolutional Generative Adversarial Networks)

Radford와 연구진[4]은 DCGANs를 제안하였다. 생성 모델과 판별 모델 모두에 대해 구조적으로 제한된 convolution networks를 제안하였다. 구조적으로 다음과 같은 내용을 포함하고 있다. (1) 모든 pooling layers를 strided convolutions 와 fractional-strided convolutions를 대체한다. (2) batchnorm layers를 사용한다. (3) fully connected hidden layers를 제거한다. (4) 생성 모델에서는 활성화 함수로 tanh를 사용하고 다른 layers에서는 ReLU 함수를 사용한다. (5) 판별 모델에서는 모든 layers에서 LeakyReLU 함수를 사용한다. DCGANs는 모든 학습에서 보다 안정적이며 고품질의 이미지를 생성할 수 있어 많은 응용 분야에서 널리 사용되고 있다.

## 2.3 LAPGAN (Laplacian Generative Adversarial Networks)

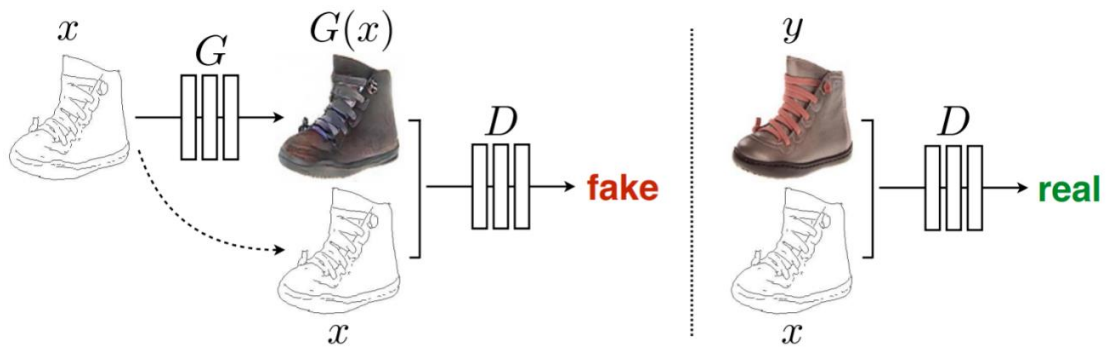
LAPGAN[5]은 K 레벨을 갖는 Laplacian 피라미드의 프레임워크를 갖는 계단식 convolutional GANs로 구성된다. K 레벨은 가장 거친 레벨로 GAN은 가장 거친 해상도의 이미지에 noise vector를 매핑하는 학습을 합니다. 가장 거친 k 레벨을 제외한 피라미드의 각 레벨에서 별도의 cGAN이 학습되며, 이 레벨에서 잔여 이미지를 생성하기 위해 k레벨에서 출력된 이미지를 조건 변수로 사용한다. 이러한 거칠고 미세한 방식으로 인해 LAPGAN은 더 높은 해상도의 이미지를 생성할 수 있다.

## 3. Image to image translation (I2IT)

Image to image translation (I2IT)은 A 도메인에 위치한 영상을 B 도메인에 위치한 영상으로 변환시키는 작업을 의미한다. 예를 들어, 열화상 영상(Thermal image)를 입력으로 하여 RGB 영상(Visible image)로 만드는 것과 같은 작업이다. I2IT 분야에서 GAN을 활용한 방법에 대해서 살펴본다.

### 3.1 Pix2pix

Phillp Isola 외 연구진[6]은 cGAN을 사용하는 일반적인 I2IT framework인 pix2pix를 제안한다. 이들의 네트워크 아키텍처는 DCGANs[4]의 지침을 따르며, 몇 가지 추가적인 변경 사항이 있다: (1) convolution-BatchNorm-ReLU 형식의 모듈을 적용한다. (2) 생성 모델 G에 deep layers와 shallow layers 사이에 skip connections를 추가하였고, U-Net 기반의 Encoder와 Decoder를 사용하였다. 판별 모델 D에는 패치 스케일에서 더 빠르게 실행되고 비현실적인 구조에 불이익을 주는 PatchGAN을 사용한다. 판별 모델을 속일 수 있는 현실적인 이미지를 생성하는 것뿐만 아니라 생성된 이미지가 실제 결과에 가까워지도록 요구하는 것이 목적이다. 따라서, 적대적 손실 외에도 객관적인 기능에 콘텐츠 손실을 추가로 포함한다. Pix2pix는 도시 풍경에 대한 레이블, 풍경에 대한 레이블, 사진에 대한 가장자리, 낮에서 밤까지 등 다양한 이미지 대 이미지 변환 작업에 효과적이라는 것이 입증했다. 이 방법의 가장 큰 contribution은 다양한 문제에 대해 cGAN이 합리적인 결과를 산출한다는 것을 보여준다는 것이다. 또한, 좋은 결과를 얻기에 충분한 간단한 아키텍처를 제시하고 몇 가지 중요한 아키텍처 선택의 영향을 분석하는 것이다. 그러나 cGAN은 지상 실측 자료와 같이 한 눈에 선명한 이미지를 생성하지만 실제로는 작은 물체를 많이 포함한다.



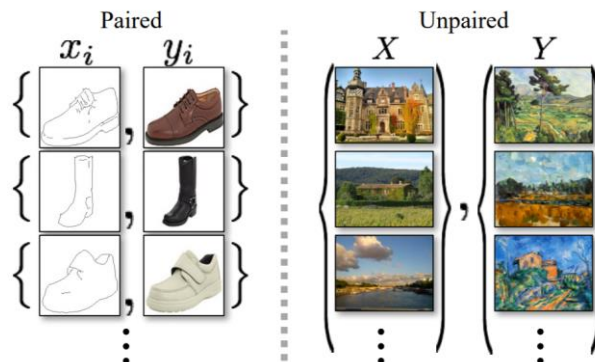
[그림 3] map edge를 photo로 변환하는 훈련을 하는 cGAN. 판별 모델 D는 생성 모델 G에 의해 만들어진 가짜와 진짜를 분류하는 방법을 학습하고, 생성 모델 G는 판별 모델 D를 속이는 방법을 학습한다. unconditional GAN과는 다르게 생성 모델과 판별 모델 모두 input edge map을 관찰한다.

### 3.2 CycleGAN (cycle-consistent Generative Adversarial Networks)

Pix2pix[6]는 학습 데이터로 paired한 이미지(translation 전의 이미지와 translation 후의 대응 이미지)를 요구하지만, 이러한 paired한 이미지 셋이 존재하는 경우가 많지 않다. 이러한 문제를 해결하기 위해 Jun-Yan Zhu 외 연구진[7]은 unpaired한 I2IT 프레임 워크인 CycleGAN을 제안했다. CycleGAN은 두 개의 분리된 GAN으로 구성되며, 하나는 하나의 도메인에서 다른 도메인으로 이미지를 변환한다. 예를 들어 설명하자면, 말에서 얼룩말로의 변환을 위한 GAN  $x_{trans} = G(x)$ 와 얼룩말에서 말로의 변환을 위한 GAN  $x = G_{inv}(x_{trans})$ , 이와 같이 두 개의 분리된 GAN으로 구성되어 다른 도메인으로 이미지를 변환하며, 두 개의 GAN은 연대적으로 훈련된다.

pix2pix와 비슷하게 판별 모델로는 PatchGAN을 사용한다. LSGAN[25]에 따르면, 적대적 손실은 더 안정적인 훈련을 위해 로그 함수 대신 최소 제곱 함수를 사용한다. 두 GAN의 두 적대적 손실 외에도 목적 함수에는 L1주기 일관성 손실이 추가로 포함되어 있어  $x \approx G_{inv}(G(x))$  와  $x_{trans} \approx G(G_{inv}(x_{trans}))$  와 같이 변환 주기 후에 이미지가 변환되도록 한다.

이 방법에서는 쌍을 이루는 데이터가 존재하지 않는 광범위한 응용 프로그램에서의 일반성을 보여줍니다. 이 방법을 통해 설득력 있는 결과를 많이 얻을 수 있지만, 그 결과는 균일한 것과는 거리가 멀다고 할 수 있다. 색상 및 텍스처 변환 작업에서는 성공적인 결과를 얻지만, 기하학적인 변경이 필요한 경우에는 성공적인 결과를 얻을 수 없다.

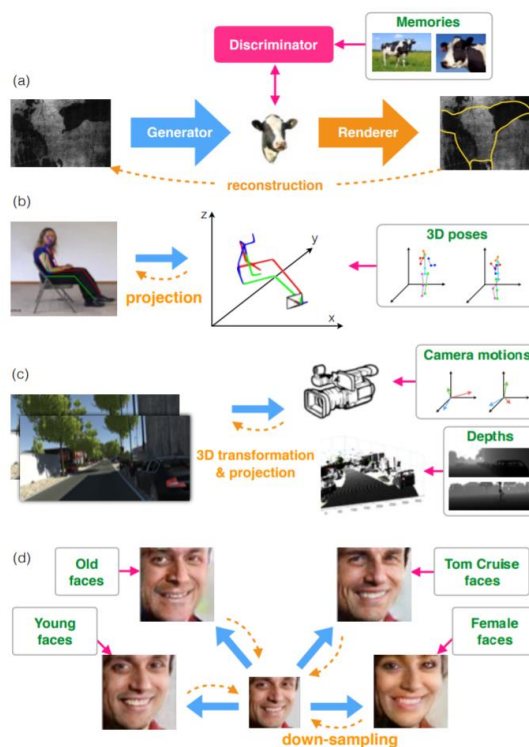


[그림 4] Paired한 이미지 셋과 Unpaired한 이미지 셋

### 3.3 AIGN (Adversarial Inverse Graphics Network)

AIGN[8]는 또한 Paired 하지 않은 이미지 셋으로 훈련한다. AIGN은 생성 모델 G와 판별 모델 D 그리고 Renderer P로 구성된다. 생성 모델 G는 입력 이미지  $x$ 를 출력 이미지  $G(x)$ 에 매핑하고 Renderer는 생성 모델의 출력을 입력으로 다시 매핑한다. 훈련을 위한 목적 함수는 적대적 손실 및  $x \approx P(G(x))$ 를 시행하는 재구성 손실을 포함한다.

AIGN은 예측과 “memories” 사이에서 distribution matching을 통해 부과된 예측의 다른 renderer로부터 오는 피드백을 이전과 결합한다. 일반적인 아키텍처는 그림 5의 (a)와 같이 cow illusion와 함께 제시되며, 이어서 (b)의 3D 사람의 포즈를 추정하는 것과 (c)의 한 쌍의 프레임에서 3D 깊이와 카메라 모션을 추출하는 것, (d)의 창의적이고 제어 가능한 이미지 생성과 같은 실험 과제를 수행한다. 각 작업은 주황색 점선 화살표로 표시된 고유 도메인별 렌더러를 사용한다.



[그림 5] AIGN 아키텍처

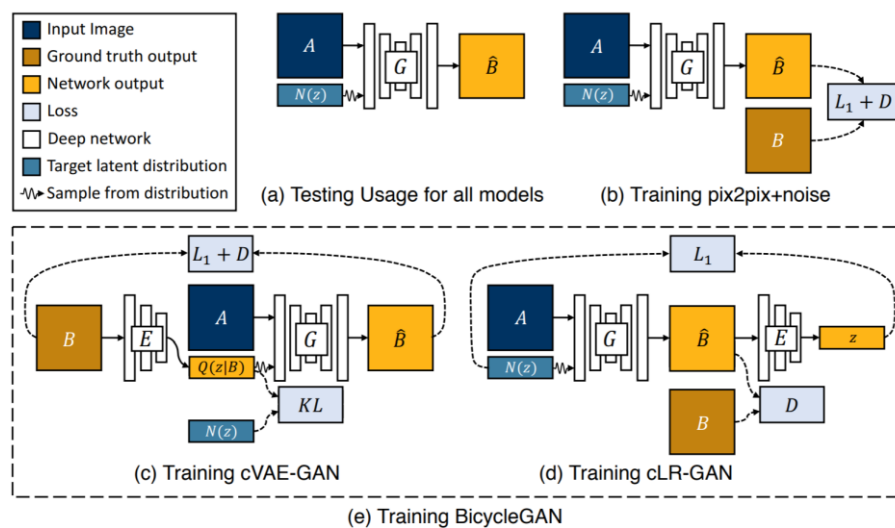
### 3.4 BicycleGAN

Jun-Yan Zhu 외 연구진[9]은 기존의 I2IT 연구에서 여러 개의 latent code에서 하나의 출력으로 many-to-one mapping 되는 mode collapse 문제를 해결하기 위해 출력에서 latent code로 가는 mapping 네트워크를 학습시켰다. 또한, 다양한 목적 함수와 네트워크 아키텍처, latent code 들을 제시하여 현실적이고 다양한 이미지를 생성하였다. BicycleGAN은 cVAE-GAN과 aLR-GAN을 결합하여 잠재 인코딩과 출력을 양방향으로 연결하고 향상된 성능을 구현한다.

그림 6에서 훈련 모델에 대해 살펴볼 수 있다. (a) 모든 모델에 대한 테스트 방법이다. 샘플 출력을 생성하기 위해 잠재 코드  $z$ 는 먼저 알려진 분포에서 무작위로 샘플링된다. 생성 모델 (G)는 입력 이미지(A)(파란색)와 latent sample( $z$ )을 매핑하여 출력샘플( $\hat{B}$ )(노란색)을 생성한다. (b)

pix2pix와 noise를 결합하여 훈련하는 방법이다. A에 대응되는 추가적인 Ground truth 이미지(B)(갈색)가 있다. (c) cVAE-GAN(및 cAE-GAN)은 ground truth 이미지(B)로부터 시작하여 latent space로 인코딩한다. 그런 다음 생성 모델은 샘플링된  $z$ 와 함께 입력 이미지(A)를 원래 이미지(B)로 다시 매핑하려고 시도한다. (d) cLR-GAN은 알려진 분포로부터 latent code를 무작위로 샘플링하고, 이를 사용하여 A를 출력 B에 매핑한 다음, 출력에서 latent code로 재구성하려고 한다. (e) 하이브리드 BicycleGAN 방법은 양방향으로 제약 조건을 결합한다.

결론적으로, cVAE-GAN과 cLR-GAN을 결합하여 만든 하이브리드 모델이 목표인 것이다. Latent space와 output space 사이의 전체적인 매핑을 장려하기 위한 여러 목표를 결합함으로써 보다 현실적이고 다양한 결과를 얻는다.



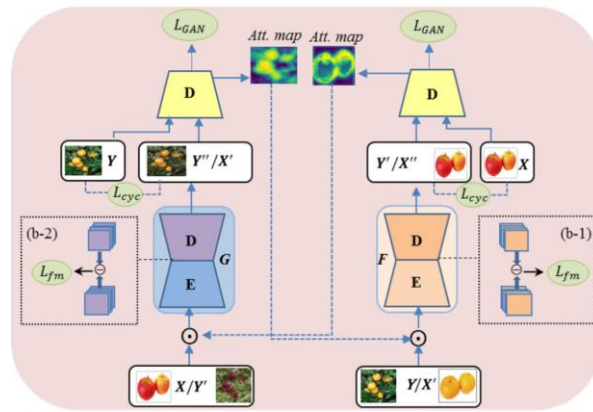
[그림 6] BicycleGAN의 훈련 모델

### 3.5 SPA-GAN (SPatial Attention Generative Adversarial Network)

Hajar Emami 외 연구진[10]은 GAN 아키텍처에 직접적으로 Attention mechanism을 도입하여 I2IT를 위한 새로운 SPA-GAN을 제안했다. SPA-GAN은 판별 모델에서 attention을 계산하고 이를 사용하여 생성 모델이 source 도메인과 target 도메인 사이의 가장 차별적인 영역에 더 집중할 수 있도록 하여 보다 사실적인 이미지가 출력될 수 있도록 한다. 또한 I2IT 작업 중에 도메인 특성을 보존하기 위해 SPA-GAN 훈련에서 추가적인 feature map loss를 도입하는 것이 도움이 된다는 것을 확인하였다. SPA-GAN은 기존 Attention 방식의 GAN 모델과 비교해 추가적인 Attention 네트워크나 supervision이 필요 없는 경량화 모델이다.

그림 7에서 볼 수 있듯이, 판별 모델 D는 입력 이미지를 분류하는 것 외에도 생성 모델 G에 공급되는 spatial attention maps를 생성하여 가장 차별적인 영역에 집중할 수 있도록 한다. 또한 feature map loss는 decoder에서 계산된 실제 및 생성된 이미지들의 feature maps 사이의 차이인 점선 블록들(b-1 및 b-2)에 도시된다. Feature map loss는 Image Translation에서 도메인 특정 feature를 보존하는 데 사용된다.

즉, Unsupervised한 환경에서 I2IT를 위한 SPA-GAN을 통해 판별 모델에서 spatial attention map을 계산하고, 생성 모델에 지식을 전달하여 두 도메인 사이의 차별적 영역에 집중하여 보다 사실적인 이미지가 생성되도록 하여 품질을 향상시킨다. SPA-GAN은 경량화 모델로 현재의 최첨단 기술에 비해 질적, 양적으로 우수한 성능을 달성했다.



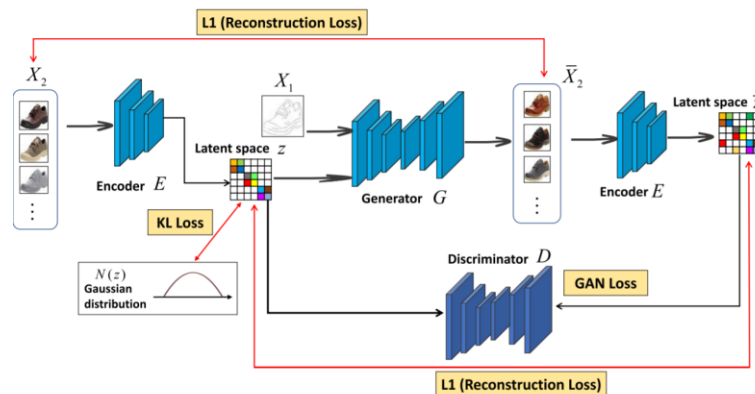
[그림 7] SPA-GAN 아키텍처

### 3.6 CE-GAN (Consistent Embedded Generative Adversarial Networks)

CEGAN[11]은 대상 영역에서 시각적으로 사실적인 출력을 생성하기 위한 조건부 생성 모델을 학습하고 실제 이미지 공간과 latent 공간 모두에서 긴밀한 연결을 수행하여 latent 다중 모드 결과의 전체 분포를 포착하는 것을 목표로 한다. 이 모델은 mode collapse로 인한 문제를 피하고 보다 다양하고 현실적인 결과를 도출한다.

그림 8에서 모델 CE-GAN은 생성 모델 G와 판별 모델 D 그리고 자동 인코더 E(그림의 두 개의 Encoder E는 본질적으로 동일함)로 구성된다. 먼저, ground truth인  $x_2$ 가 latent 공간으로 인코딩하여 latent code  $z$ 를 얻는다.  $x_1$ 은 생성 모델 G의 입력으로 latent code  $z$ 와 결합한다. 생성 모델은 입력 이미지  $x_1$ 을 latent code  $z$ 와 함께 원래 이미지  $x_2$ 로 다시 매핑하려고 시도한다. 생성된 샘플 축적  $x_2$ 를 E에 의해 latent code  $z$ 로 인코딩한다. 정리하자면, 대상 영역에서 사실적이고 다양한 이미지를 생성하기 위해 ground truth의 latent code와 이미지를 재구성하려고 노력한다.

결론적으로, 이 방법은 latent 공간과 실제 이미지 공간 사이의 긴밀한 연결을 강화함으로써 잠재적인 다중 모드 결과의 전체 분포를 포착한다. 특히, 생성된 이미지의 중복성과 노이즈의 영향을 완화하기 위해 다른 GAN과 달리 모델의 판별 모델은 latent 공간에서 진짜 이미지와 가짜 이미지를 구별한다. 이러한 경험적 실험 결과는 이 방법이 잘 정립된 여러 이미지 생성 방법보다 훨씬 우수하다는 것을 보여준다.

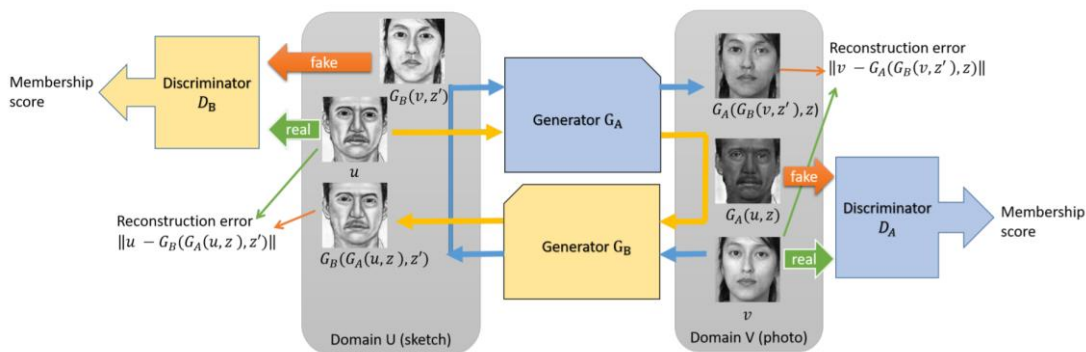


[그림 8] CE-GAN 아키텍처



### 3.7 DualGAN

Zili Yi 외 연구진[12]은 Image Translation을 두 도메인의 레이블이 없는 이미지 셋을 이용하여 훈련할 수 있는 새로운 DualGAN을 개발했다. 아키텍처에서 기존의 GAN에서 도메인 U에서 도메인 V의 이미지로 이미지를 변환하는 방법을 배우는 반면 DualGAN에서는 작업을 반전하는 방법을 배운다. 원시 및 이중 작업으로 만들어진 closed-loop를 사용하면 두 도메인의 이미지를 변환한 다음 재구성할 수 있다. 따라서, 이미지의 구성 오류를 설명하는 손실 함수를 사용하여 Translation을 훈련할 수 있다. 그림 9에서와 같이 도메인 U에 있는 이미지  $u$ 는  $G_A$ 를 사용하여 도메인 V (photo)로 변환한다. 도메인 V에서 훈련된  $G_A(u, z)$ 는 얼마나 잘 변환되었는지 판별 모델  $D_A$ 로 평가한다. 그리고 나서  $G_A(u, z)$ 는  $G_B$ 를 사용하여 다시 도메인 U로 변환하여 출력으로  $u$ 를 재구조화한 버전인  $G_B(G_A(u, z), z')$ 를 얻는다. 유사하게, 도메인 V에 있는 이미지  $v$ 는 도메인 U에 있는  $G_B(v, z')$ 로 변환되고 그 후,  $G_A(G_B(v, z'), z)$ 로 재구성된다. 판별 모델  $D_A$ 는 positive samples인 이미지  $v$ 와 negative samples인  $G_A(u, z)$ 로 훈련되고, 판별 모델  $D_B$ 는 positive samples인 이미지  $u$ 와 negative samples인  $G_B(v, z')$ 로 훈련된다. 생성 모델  $G_A$ 와  $G_B$ 는 판별 모델  $D_A$ 와  $D_B$ 를 속일 수 있는 가짜 이미지를 출력하기 위해 최적화한다. 결론적으로 라벨링되지 않은 데이터만으로 DualGAN은 레이블이 있는 데이터로 훈련된 cGAN보다 비교 가능하거나 더 나은 출력을 생성할 수 있다.

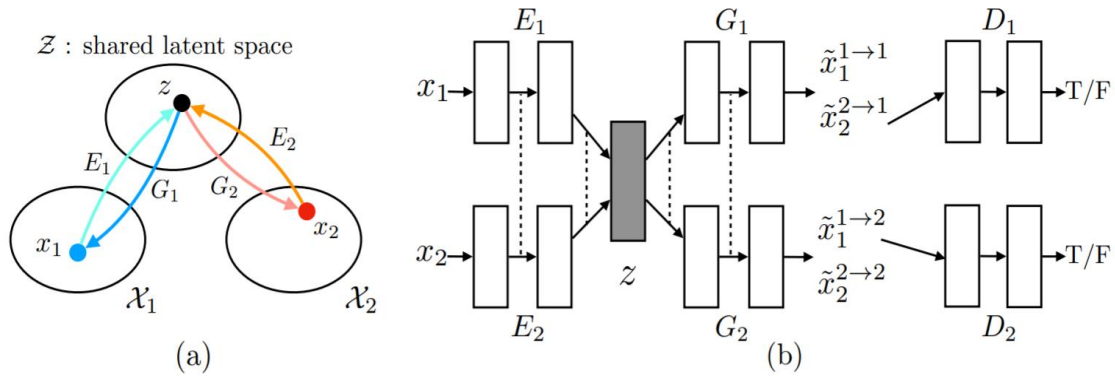


[그림 9] DualGAN 아키텍처

### 3.8 UNIT (UNsupervised Image-to-image Translation network)

Ming-Yu Liu 외 연구진[13]은 확률적 모델링 관점에서 I2IT의 문제점을 분석하였고, 서로 다른 도메인에서 이미지의 공동 분포를 배우는 것이 핵심이라는 것을 파악하여 이 문제를 해결하기 위해 Coupled GANs와 shared-latent space 가정에 기초한 UNIT를 제안하였다. 이 방법은 서로 다른 영역에서 한 쌍의 대응하는 이미지가 shared-latent space에서 동일한 latent 표현에 매핑될 수 있다고 가정한다. 그림 10의 (a)는 shared-latent space 가정에 대한 그림이다. 한 쌍의 이미지 ( $x_1, x_2$ )에 대응되는 두 개의 다른 도메인인  $X_1$ 과  $X_2$ 는 shared-latent space  $Z$ 에 의해 동일한 latent code  $z$ 로 매핑될 수 있다.  $E_1$ 과  $E_2$ 는 두 개의 인코딩 함수이고, latent codes로 매핑한다.  $G_1$ 과  $G_2$ 는 두 개의 생성 함수이고, 이미지로 latent codes를 매핑한다. (b)는 제안된 UNIT 프레임워크다. 결론적으로 제안된 프레임워크를 다양한 Unsupervised I2IT 문제에 적용하여 고품질 이미지 변환 결과를 얻었다. 그러나 Gaussian latent 공간을 가정으로 인한 단일 모드이며 saddle point 검색 문제로 인해 훈련이 불안정하다는 한계점을 가지고 있다.



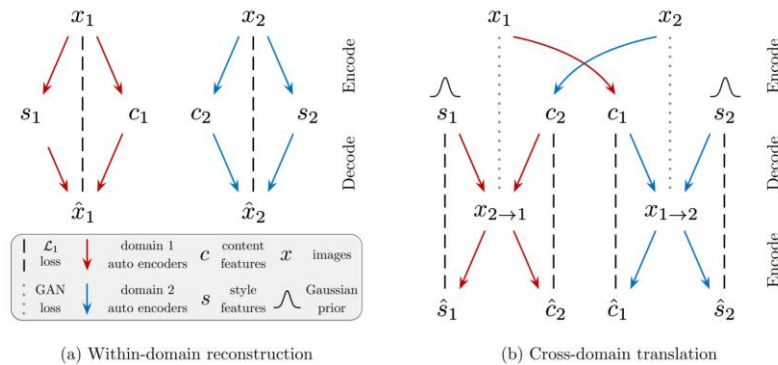


[그림 10] UNIT 아키텍처

### 3.9 MUNIT (Multimodal UNsupervised Image-to-image Translation)

Xun Huang 외 연구진[14]은 영상의 latent 공간을 content 공간과 style 공간으로 분해하여 multimodal 출력을 가능하게 하는 MUNIT을 제안했다.

그림 11에서 볼 수 있듯이 모델은 두 개의 자동 인코더(각각 빨간색과 파란색 화살표로 표시)로 구성되며, 각 도메인마다 하나씩 있다. 각 자동 인코더의 latent code는 content code  $c$ 와 style code  $s$ 로 구성된다. 변환된 이미지가 대상 도메인의 실제 이미지와 구별할 수 없도록 하는 적대적인 목표(점선)와 이미지와 latent code를 재구성하는 양방향 재구성 목표(점선)로 모델을 훈련한다. 공동으로 인코더와 디코더 및 판별 모델을 훈련시켜 적의 손실과 양방향 재구성 손실 조건의 가중치 합으로 최종 목표 값을 최적화한다. 결론적으로 다중 모드의 Unsupervised I2IT 방법은 기존의 Unsupervised 방법보다 우수한 품질과 다양성을 달성하며 최첨단 감독 접근법과 비교할 수 있다.



[그림 11] MUNIT 아키텍처

## 4. 결론

Image To Image Translation (I2IT) 작업을 위한 GAN의 연구는 최근 많은 진전을 이루었으며 Unsupervised 학습 알고리즘은 Supervised 학습 알고리즘과 비교할 만한 성능을 가지게 되었다. Paired한 이미지 셋을 수집하여 훈련하는 것은 비용이 많이 들거나 불가능하기 때문에 향후에는 Unsupervised된 환경에서의 방법이 활용될 가능성이 높다.

## 5. 참고문헌

- [1] Ian J. Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio, "Generative Adversarial Networks", 2014
- [2] Diederik P Kingma, and Max Welling, "Auto-Encoding Variational Bayes", 2013
- [3] Mehdi Mirza, and Simon Osindero, "Conditional Generative Adversarial Nets", 2014
- [4] Alec Radford, Luke Metz, and Soumith Chintala, "Unsupervised Representation Learning with Deep Convolutional Generative Adversarial Networks", 2015
- [5] Emily Denton, Soumith Chintala, Arthur Szlam, and Rob Fergus, "Deep Generative Image Models using a Laplacian Pyramid of Adversarial Networks", 2015
- [6] Phillip Isola, Jun-Yan Zhu, Tinghui Zhou, and Alexei A. Efros, "Image-to-Image Translation with Conditional Adversarial Networks", 2017
- [7] Jun-Yan Zhu, Taesung Park, Phillip Isola, and Alexei A. Efros, "Unpaired image-to-image translation using cycle-consistent adversarial networks", 2017
- [8] Hsiao-Yu Fish Tung, Adam W. Harley, William Seto, and Katerina Fragkiadaki, "Adversarial Inverse Graphics Networks: Learning 2D-to-3D Lifting and Image-to-Image Translation from Unpaired Supervision", 2017.
- [9] Jun-Yan Zhu, Richard Zhang, Deepak Pathak, Trevor Darrell, Alexei A. Efros, Oliver Wang, and Eli Shechtman, "Toward multimodal image-to-image translation", 2017
- [10] Hajar Emami, Majid Moradi Aliabadi, Ming Dong, and Ratna Babu Chinnam, "SPA-GAN: Spatial attention GAN for image-to-image translation", 2019
- [11] FENG XIONG, QIANQIAN WANG, and QUANXUE GAO, "Consistent Embedded GAN for Image-to-Image Translation", 2019
- [12] Zili Yi, Hao Zhang, Ping Tan, and Minglun Gong, "Dualgan: Unsupervised dual learning for image-to-image translation", 2017
- [13] Ming-Yu Liu, Thomas Breuel, and Jan Kautz, "Unsupervised image-to-image translation networks", 2017
- [14] Xun Huang, Ming-Yu Liu, Serge Belongie, and Jan Kautz, "Multimodal unsupervised image-to-image translation", 2018
- [15] Xian Wu, Kun Xu, and Peter Hall, "A Survey of Image Synthesis and Editing with Generative Adversarial Networks", 2017
- [16] Zhang Yongxin, "A Survey of Image to Image Translation with GANs", 2020