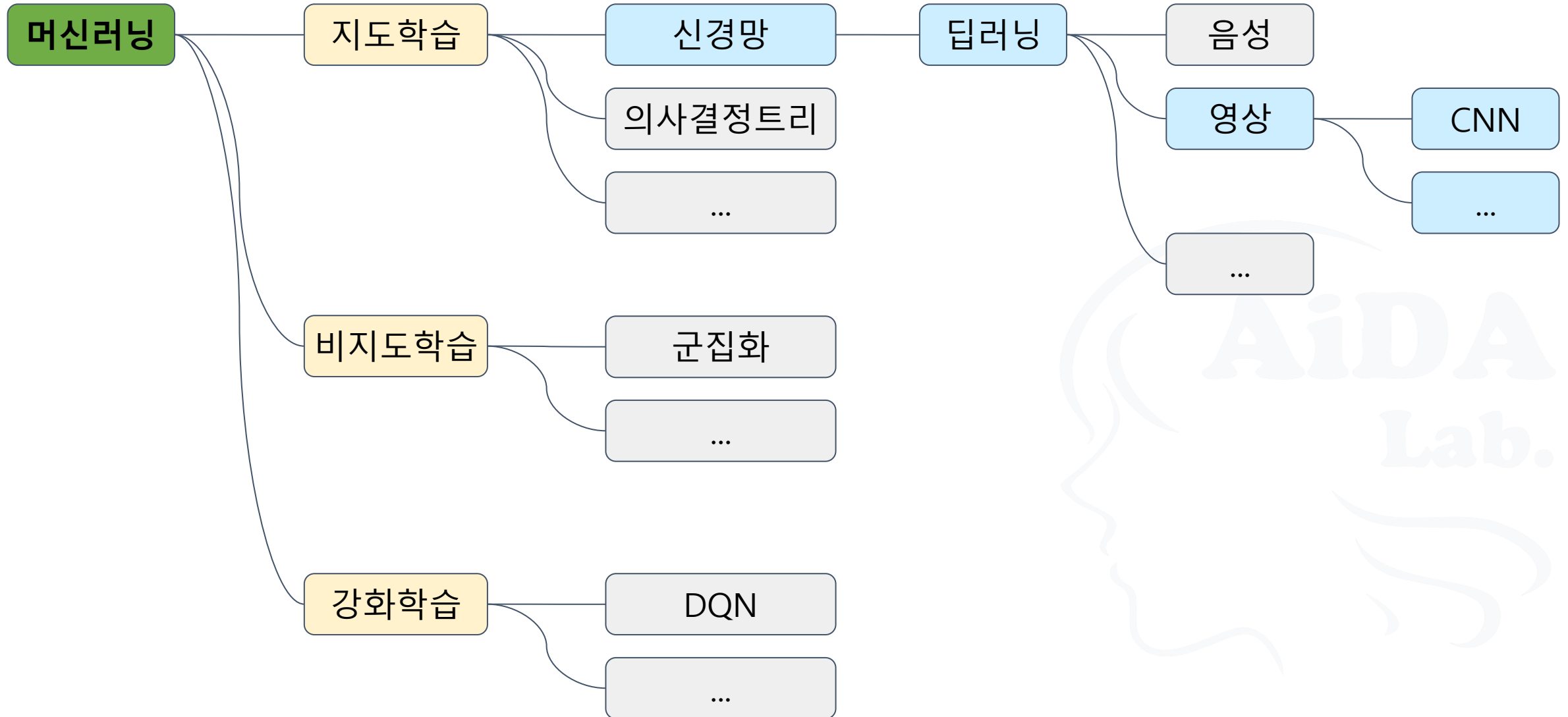


# GAN

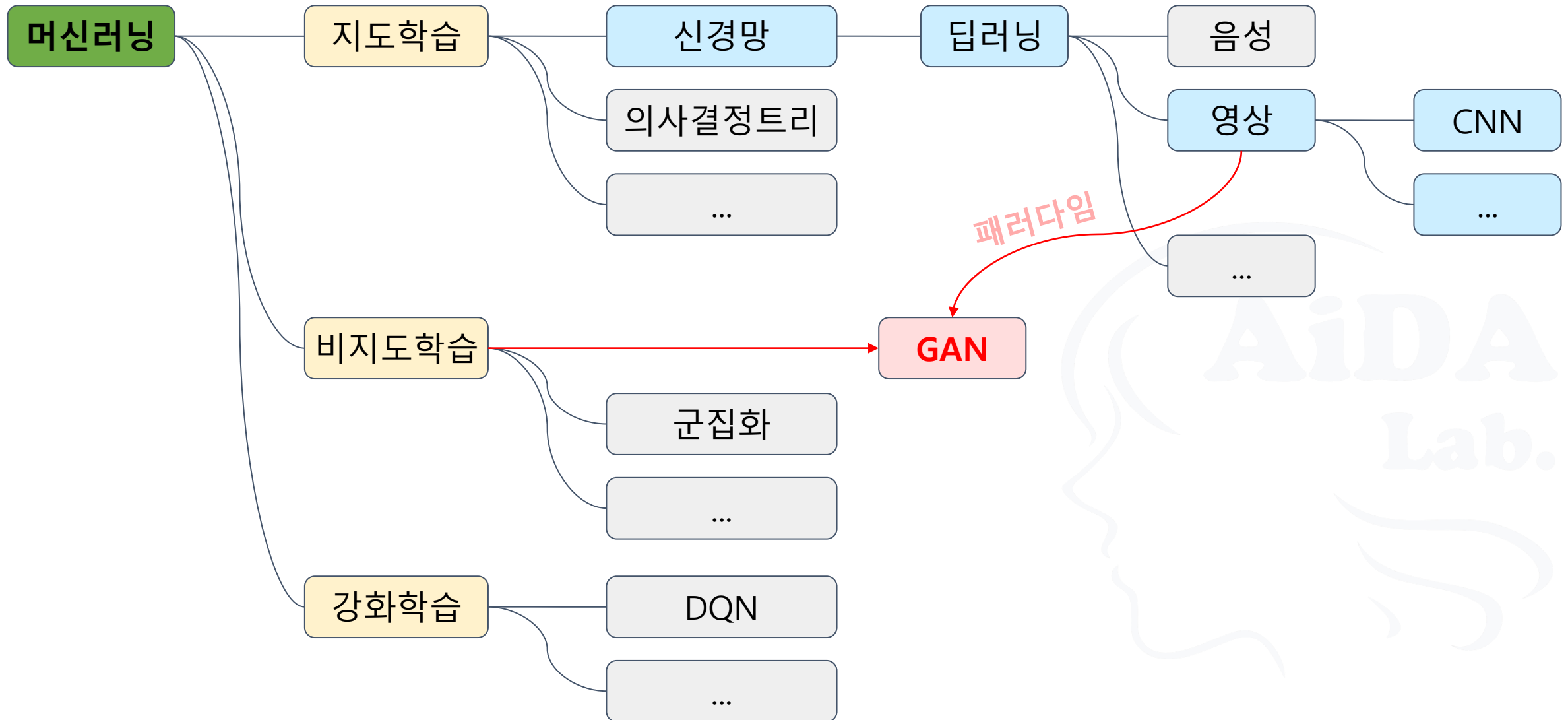
## (Generative Adversarial Networks)

생성적 적대 신경망

# 머신러닝의 종류



# GAN의 등장



# GAN (Generative Adversarial Networks)



- 동시에 두 개의 모델을 훈련하는 머신 러닝의 한 종류
- GAN : 생성적 적대 신경망 (Generative Adversarial Networks)
  - 2014년, 이안 굿펠로우가 NIPS에서 발표
  - 지도학습 중심의 딥러닝 패러다임을 비지도학습으로 전환시킴
  - 얀 르쿤, 최근 20년간 머신 러닝 연구 중 가장 혁신적인 아이디어로 꼽음
  - NIPS(Neural Information Processing Systems, 2017년 NeurIPS로 약자 변경)

# GAN (Generative Adversarial Networks)



- **생성적 (Generative)**
  - 모델의 목적 → 새로운 데이터를 생성하는 것
  - 어떤 결과를 생성할 것인가? → 학습을 위한 훈련 데이터 셋에 의해 결정
- **적대 (Adversarial)**
  - 적대하는 구조 → 경쟁 구조
- **신경망 (Network)**
  - ?



?

일반적인 신경망과는 관계없어 보이지만...

→ **판별자, 생성자** 네트워크의 내부 학습구조 /

오류 역전파 - 가중치 수정 모델은

기존의 신경망 모델을 따름

- 두 신경망 모델(생성자, 판별자)의 경쟁을 통해 학습, 결과물 도출
  - 생성자 (Generator)
    - 실제 데이터를 학습하고 이를 바탕으로 거짓 데이터 생성
    - 실제에 가까운 거짓 데이터를 생성하는 것이 목적
  - 판별자 (Discriminator)
    - 생성자가 제시한 데이터가 실제인지 거짓인지 판별하도록 학습
    - 생성자의 거짓 데이터에 속지 않는 것이 목적

- **훈련 데이터 셋: 목적에 관련된 데이터 셋 선택**

예: 레오나르도 다빈치의 작품처럼 보이는 그림을 그리고 싶다

→ 다빈치의 작품을 훈련 데이터 셋으로 선택

- **생성자의 목표는?**

→ 실제 데이터와 구분하기 어려울 정도의 유사한 샘플 만들기

- **판별자의 목표는?**

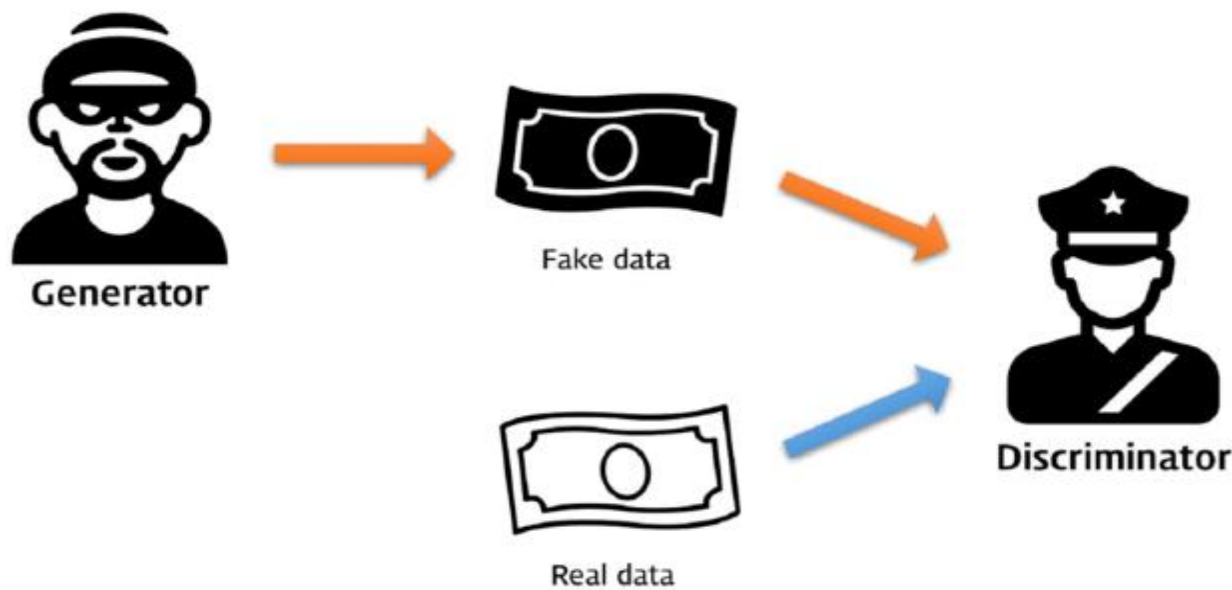
→ 생성자가 만든 샘플이 훈련 데이터 셋의 실제 데이터와 다르다고 판별하기

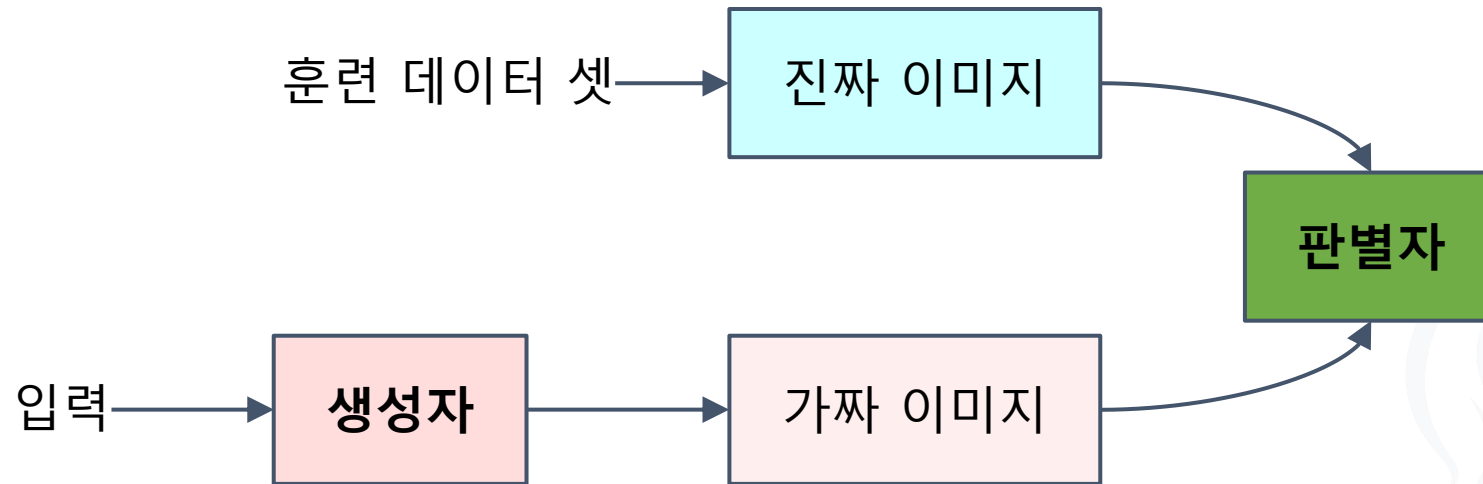


## • 위조지폐범과 경찰

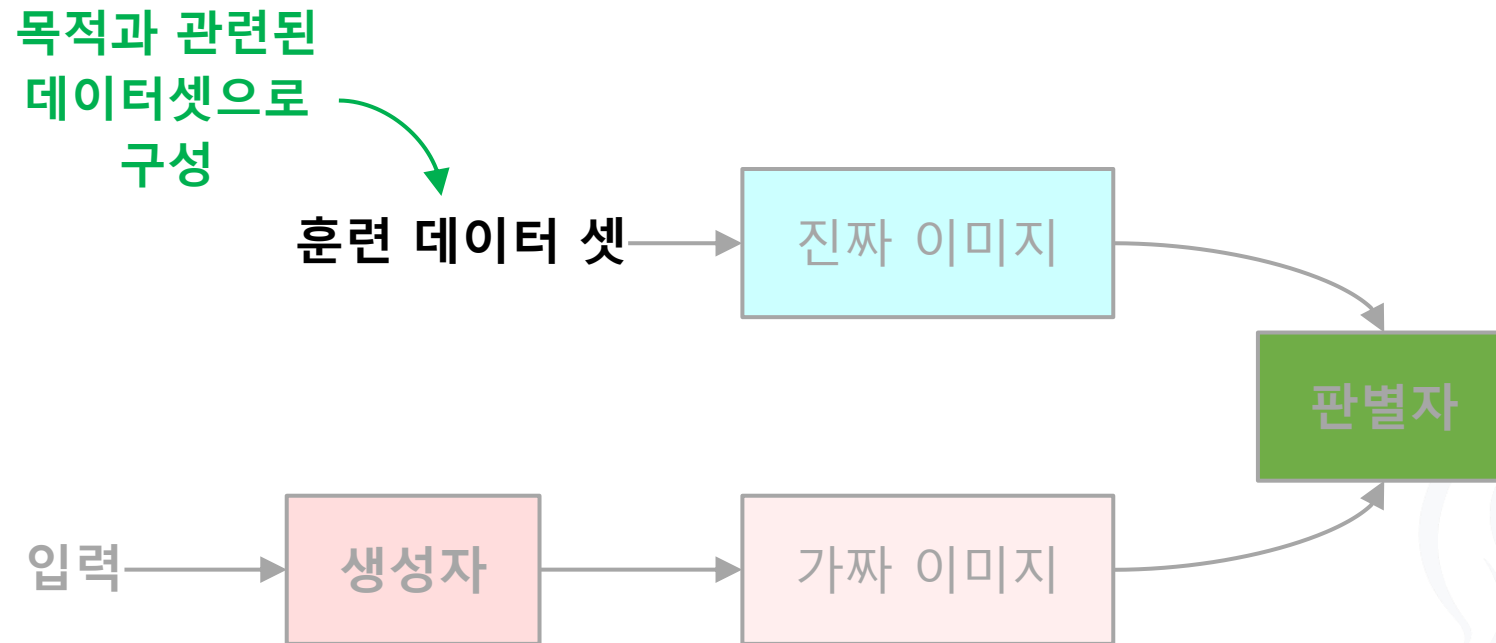
- 위조지폐범: 경찰을 속이기 위해 점점 지폐 위조 기술을 발전시킴
- 경찰: 위조지폐범을 잡기 위해 점점 위폐 판별 기술을 발전시킴

점점 완벽에  
가깝게 발전





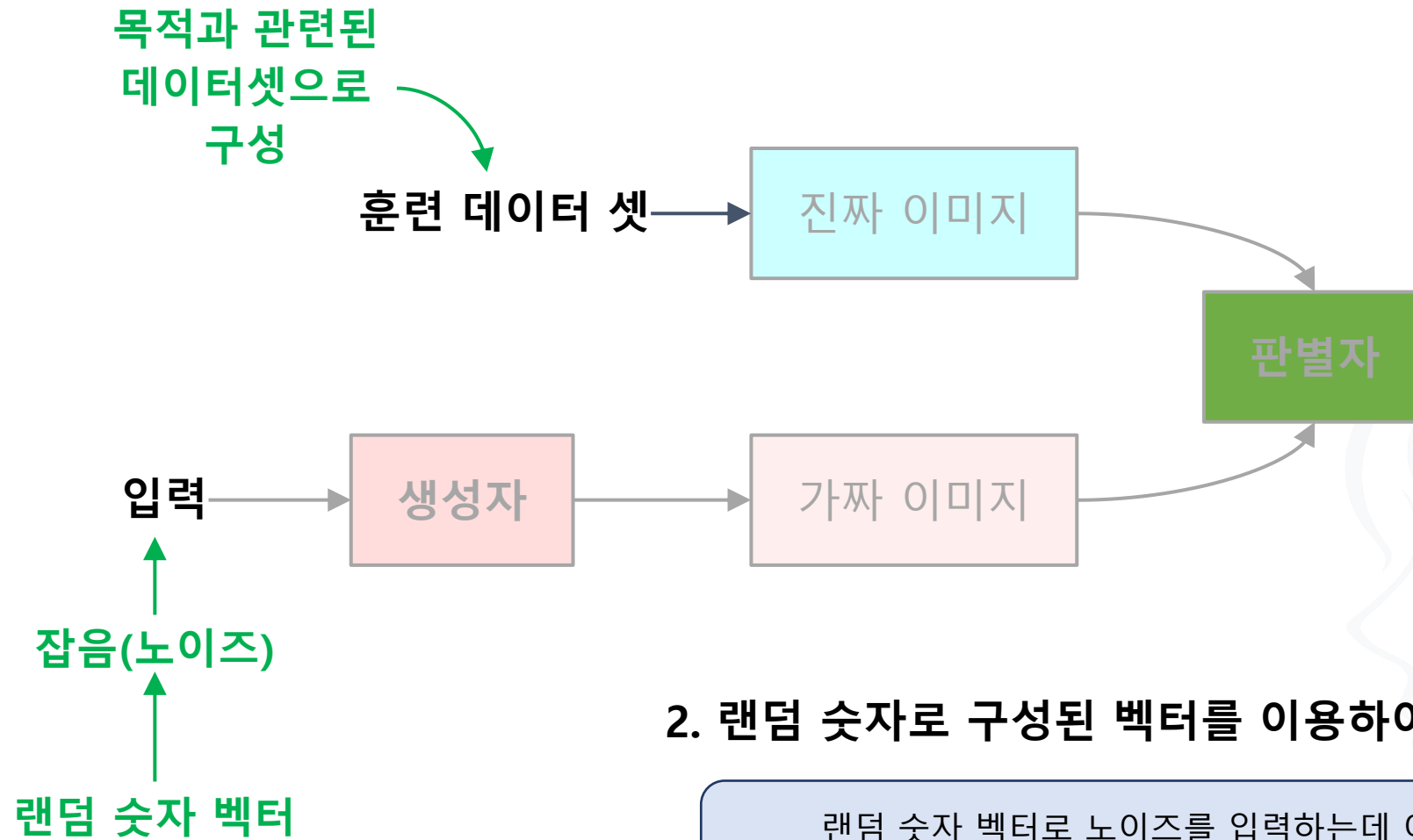
# GAN의 단계별 학습 과정



## 1. 해결하려는 목적과 관련된 데이터 셋을 준비한다.

기본적인 GAN은 범용성보다는 원하는 결과를 생성한다는 정확한 목적이 있으므로 목적에 부합하는 훈련데이터 셋을 준비하여야 함

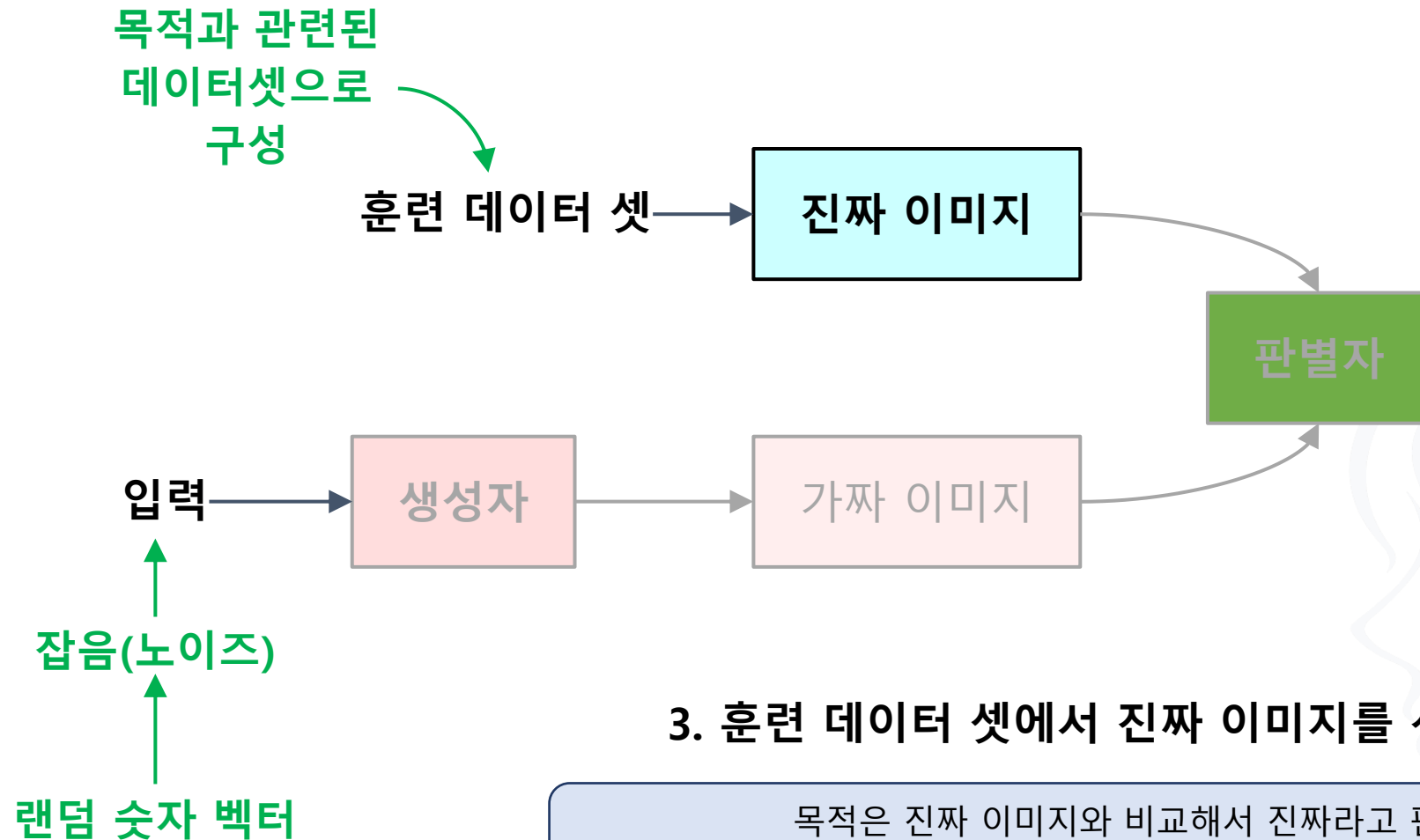
# GAN의 단계별 학습 과정



## 2. 랜덤 숫자로 구성된 벡터를 이용하여 잡음 데이터를 생성한다.

랜덤 숫자 벡터로 노이즈를 입력하는데 어떻게 학습을 진행할 것인가?  
: 처음에는 아무런 의미 없는 데이터 → 만들면서 오류가 적어지는 방향을 찾음

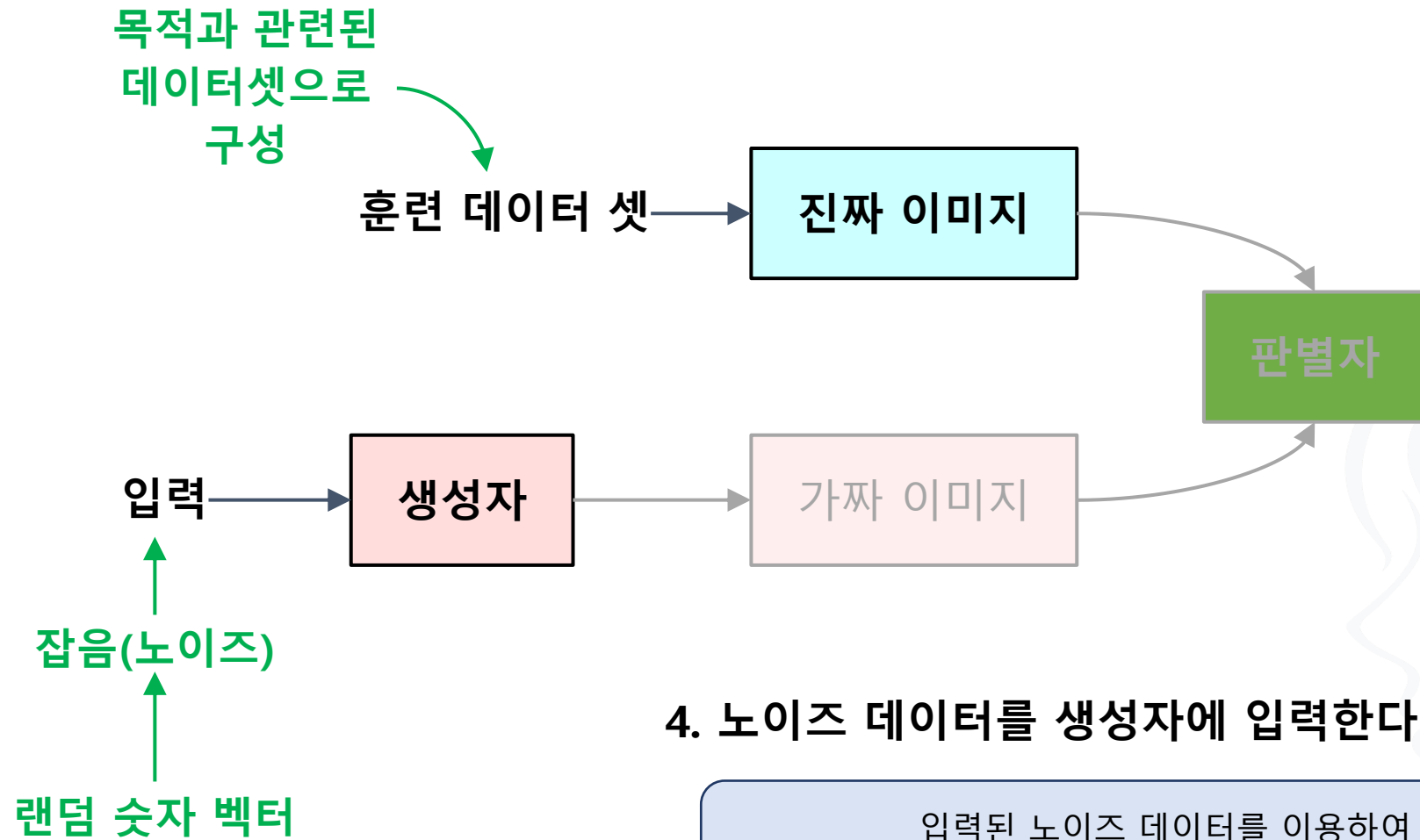
# GAN의 단계별 학습 과정



### 3. 훈련 데이터 셋에서 진짜 이미지를 선택한다.

목적은 진짜 이미지와 비교해서 진짜라고 판별하도록 하는 것임  
그러나 완전히 똑같은 이미지일 필요는 없음(단순히 똑같기만 하면 복사와 다른 점이 없음)

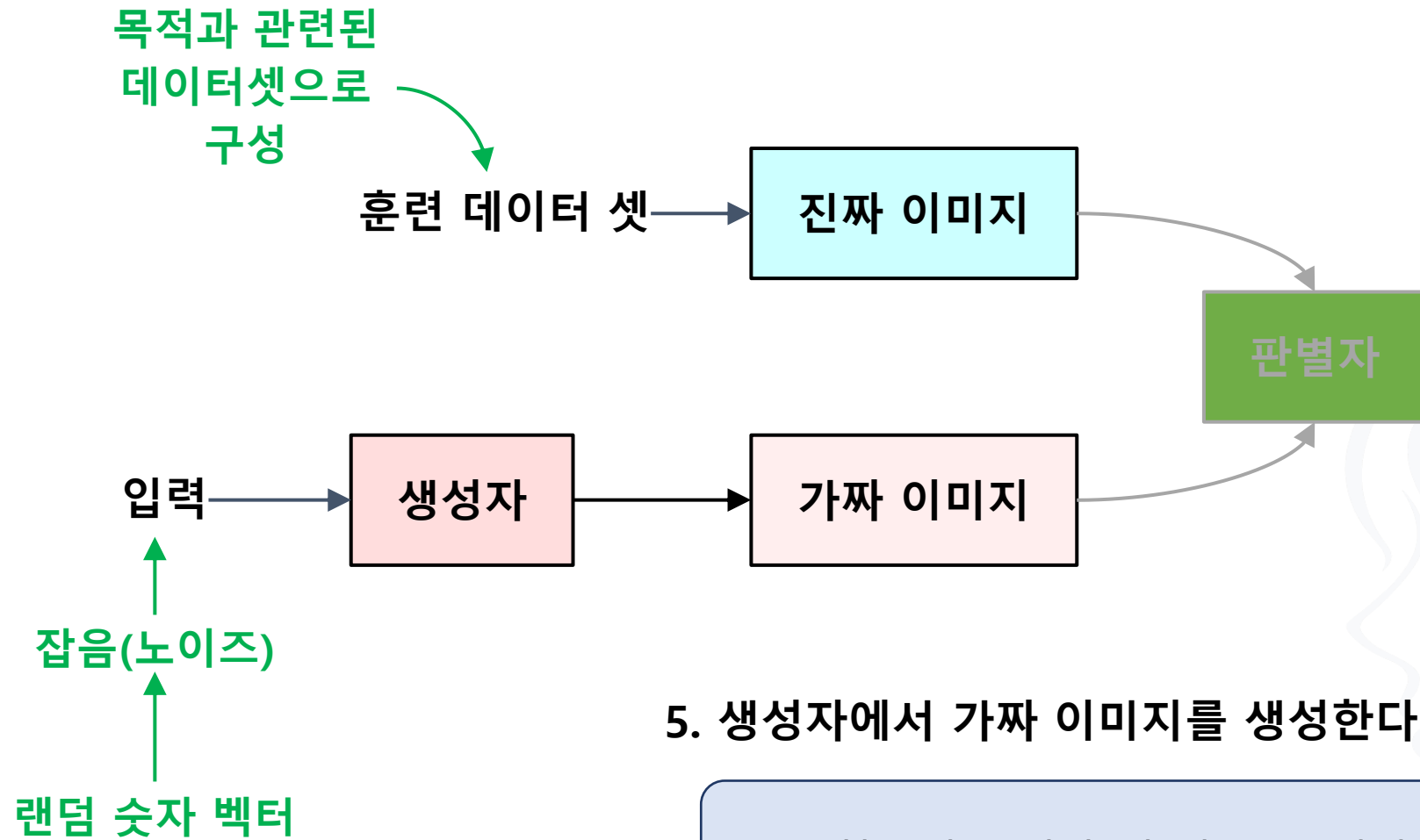
# GAN의 단계별 학습 과정



## 4. 노이즈 데이터를 생성자에 입력한다.

입력된 노이즈 데이터를 이용하여 이미지 생성을 준비함.  
후반에 이야기 하겠지만 생성목표가 이미지에 국한되지는 않음.

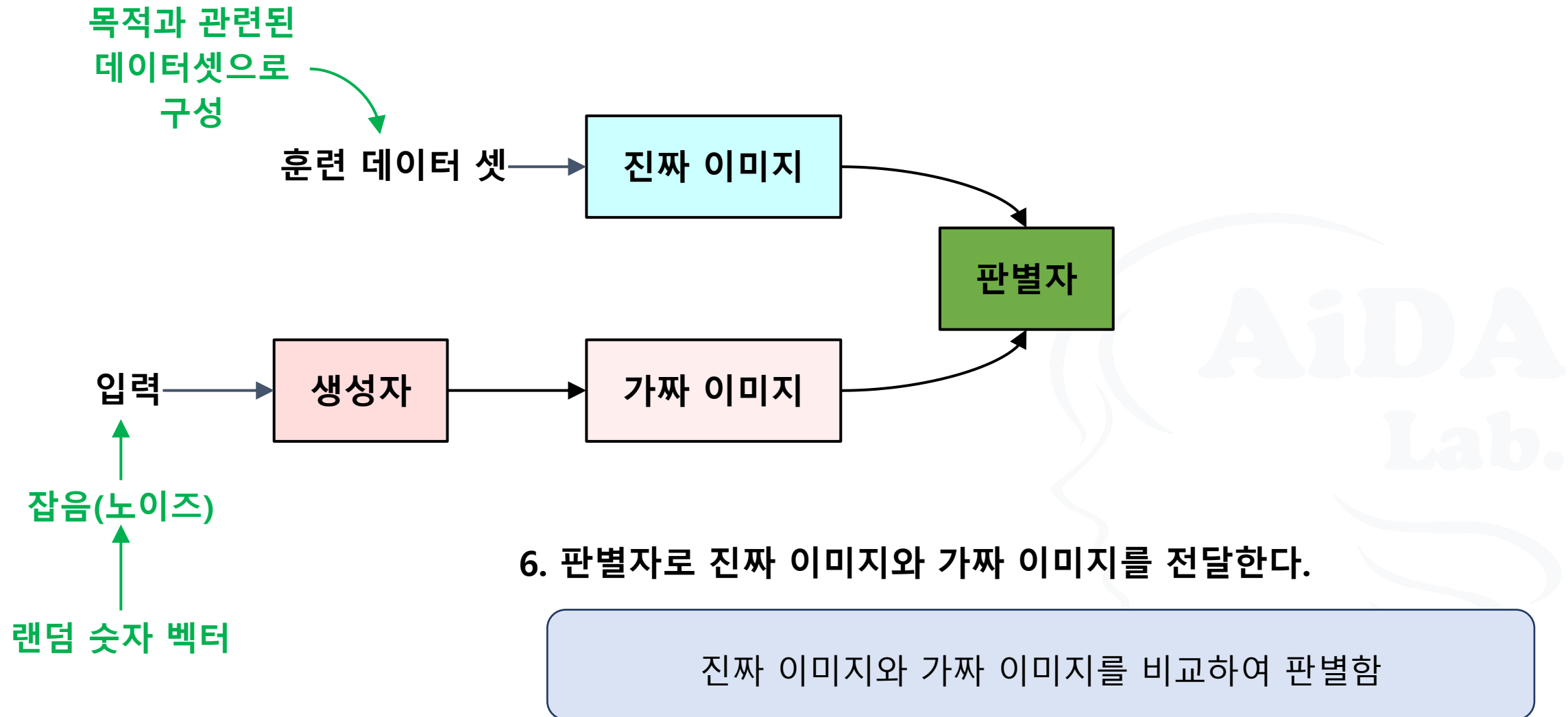
# GAN의 단계별 학습 과정



5. 생성자에서 가짜 이미지를 생성한다.

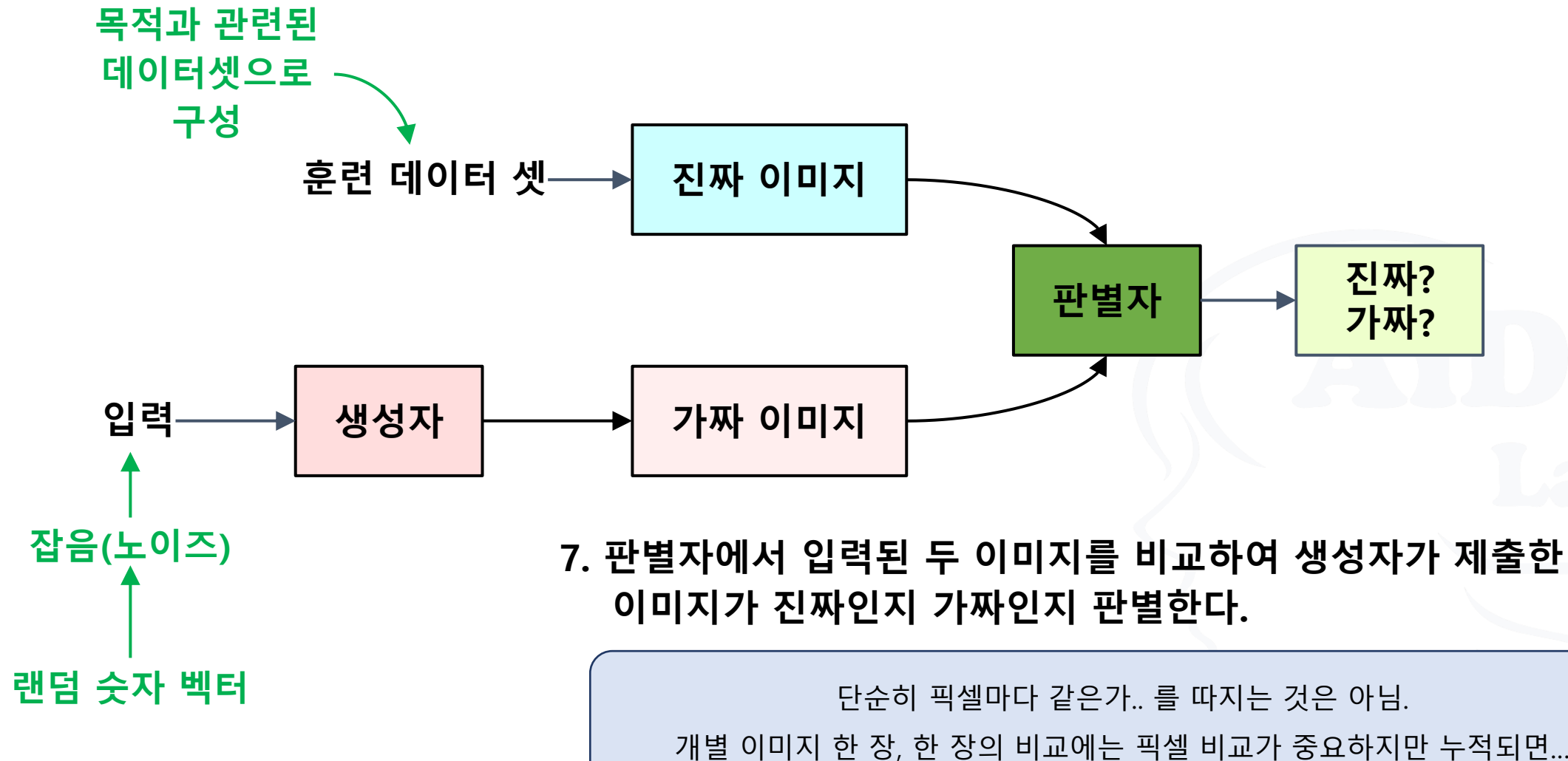
처음에는 뭐가 뭔지 모를 이상한 것들만 생성될 것임

# GAN의 단계별 학습 과정

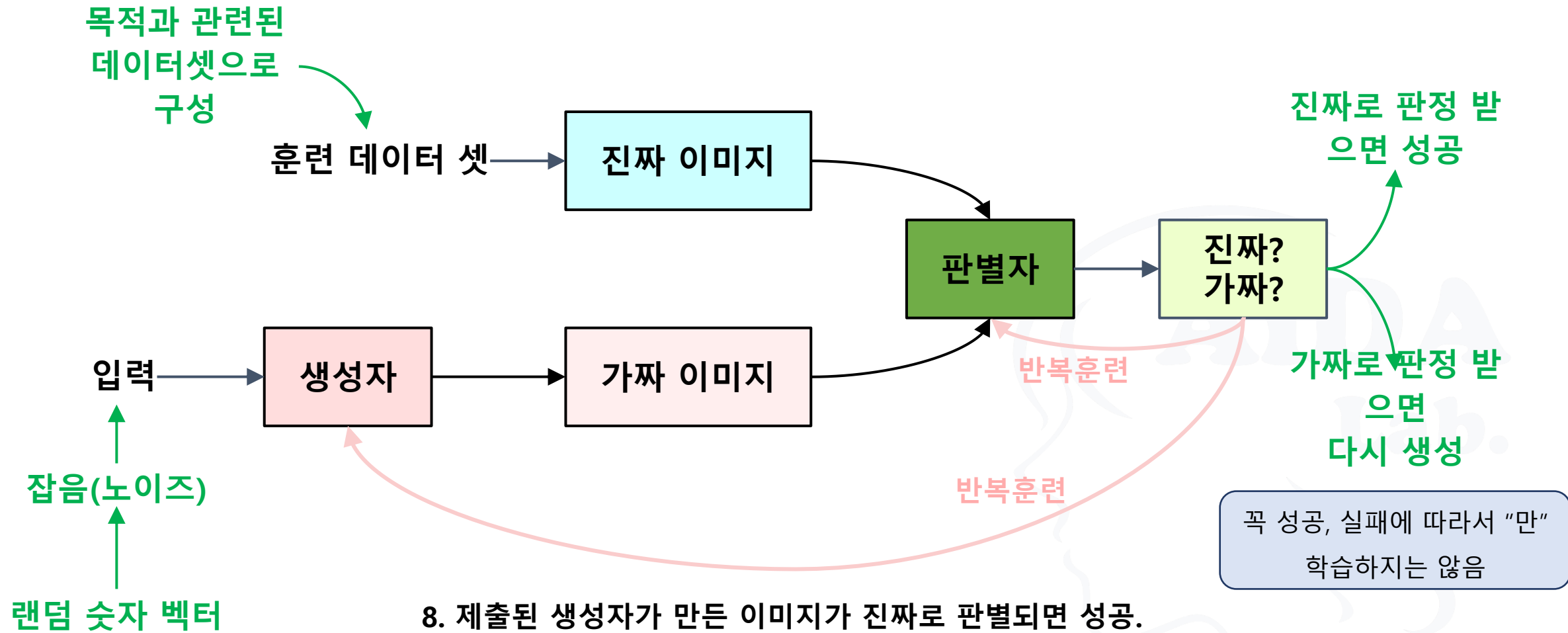




# GAN의 단계별 학습 과정



# GAN의 단계별 학습 과정



8. 제출된 생성자가 만든 이미지가 진짜로 판별되면 성공.  
성공한 경우에는 판별자에게 더 정확하게 판별하도록 훈련을 지시한다.  
가짜로 판별되면(실패) 생성자에게는 다시 생성할 것을 지시한다.

- 판별자는 무엇을 비교하여 판별하는가?

- 입력: 생성자가 만들어 낸 가짜 이미지, 훈련 데이터셋에서 선택된 진짜 이미지
- 출력: 입력된 각 이미지가 진짜일 확률을 계산하여 출력 → 예측이라고 볼 수 있음
- 판별자의 예측이 얼마나 정확한가 평가, 판별
- 단순히 픽셀의 1:1 비교가 아니라 진짜일 확률을 계산하여 판별하므로
- 완전히 동일한... 단순한 이미지 복사와는 다르다

## • 무엇을 학습하는가?

- 판별자의 예측에 대한 정확도를 이용하여 성공, 실패 구분
- 오류 역전파(Error Backpropagation) 방식으로 생성자, 판별자를 구성하는 네트워크의 노드에 대하여 훈련 가능한 파라미터들을 갱신
- 판별자의 학습: 분류 정확도를 최대화 하도록 가중치와 편향(Bias) 갱신  
(=분류오차 최소화)
- 생성자의 학습: 분류 오차율(잘못 분류할 확률)을 최대화하도록 가중치와 편향(Bias) 갱신  
(=분류오차 최대화)

# GAN은 언제까지 학습하여야 하나?

- 일반적인 신경망

- 훈련용 데이터 셋, 검증용 데이터 셋을 사용하면서
- 검증 오류가 향상되다가 떨어지기 시작하는 시점부터 훈련 중단
- 과적합(Overfitting) 회피

- GAN

- 두 개의 상반된 목적의 네트워크로 구성
- 하나의 네트워크가 좋아지면 → 다른 하나의 네트워크는 나빠짐 (Trade Off)
- 게임 이론(Game Theory)의 내시 균형(Nash Equilibrium)에 도달하면 중단  
→ But!!! 실무에서는 현실적으로 불가능!!!

- 게임 이론

- 상호 의존적인 의사 결정에 관한 이론

- 내용

- 개인 또는 기업이 어떠한 행위를 했을 때,
    - 결과가 게임과 같이 자신 뿐만 아니라 다른 참가자의 행동에 의해서도 결정되는 상황에서,
    - 자신의 최대 이익에 부합하는 행동을 추구한다는 수학적 이론

- GAN에서의 생성자, 판별자 간의 학습은 제로섬 게임과 유사

- 제로섬 게임에서는 한 사람이 얻는 이득만큼 다른 사람이 손해를 봄

## • 내시 균형(Nash Equilibrium)

- 경쟁자 대응에 따라 최선의 선택을 하면 서로가 자신의 선택을 바꾸지 않게 되는 균형상태
- 상대방이 현재 전략을 유지한다는 전제 하에 나 자신도 현재 전략을 바꿀 유인이 없는 상태
- 모든 제로섬 게임은 참가자 모두 자신의 상황을 더 이상 개선할 수 없거나, 자신의 행위를 변경함으로써 이익을 얻을 수 없는 시점에서 내시 균형에 도달함
- GAN에서는 내시 균형에 도달하면 학습을 중단하는 것을 권장함

내시 균형: 미국 경제학자 & 수학자인 존 포브스 내시의 이름을 따서 명명된 이론  
"뷰티풀 마인드" (전기/영화)

- GAN은 언제 내시 균형에 도달하는가?

- 생성자가 훈련 데이터 셋의 실제 데이터와 구별이 되지 않는 데이터를 생성할 때 내시 균형 도달
- 어떻게 확인할 수 있는가?
  - **판별자**가 할 수 있는 최선의 방법이
  - 특정 샘플이 진짜인지 가짜인지 랜덤으로 추측할 수 밖에 없을 때
  - 이런 경우 **샘플의 진위여부가 50:50의 확률로 추측됨**

- GAN이 내시 균형에 도달함 = GAN이 수렴했다

실무에서는 GAN의 수렴이 **현실적으로 불가능에 가까움**  
상세한 설명, 상황의 분류는 수학적으로 매우 어려운..  
GAN 연구에서 가장 중요한 **미결 문제의 하나** 이므로 Pass!!

이 때는 **생성자**를 조금이라도 수정하면  
실패하게 되므로 성능은 더 나빠짐



# GAN을 공부해야 하는 이유

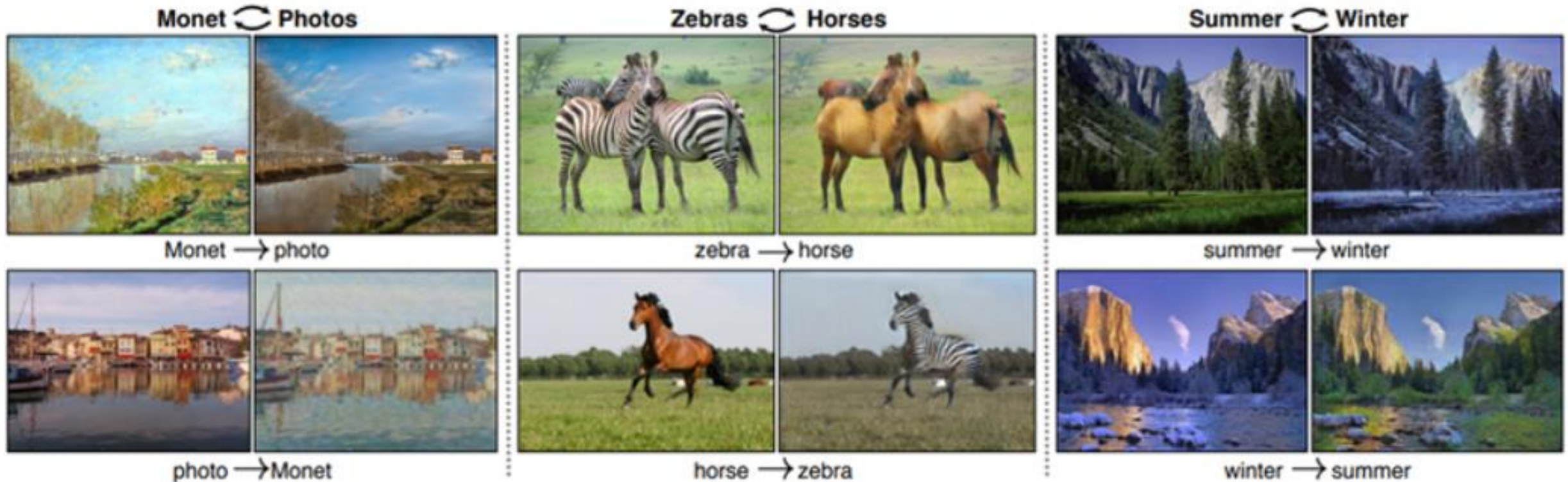
- 안 르쿤, 최근 20년간 머신 러닝 연구 중 가장 혁신적인 아이디어!!
- GAN의 가장 중요한 능력
  - 초 현실적인 이미지를 생성해 내는 능력
  - Image to Image 변환 능력
  - 일반 인공지능의 완성에 다가가는 중요한 디딤돌로 평가 받음

# GAN의 이미지 생성 능력



ProGAN을 통해 유명인 사진을 바탕으로 만들어진 허구의 인물 <출처: 엔비디아>

# GAN의 이미지 변환 능력



[그림4] cycleGAN을 통한 Image Translation

그림 출처(논문): Unpaired Image-to-Image Translation using Cycle-consistent Adversarial Networks.  
Jun-Yan Zhu, Teasung Park, Phillip Isola, Alexei A. Efros.

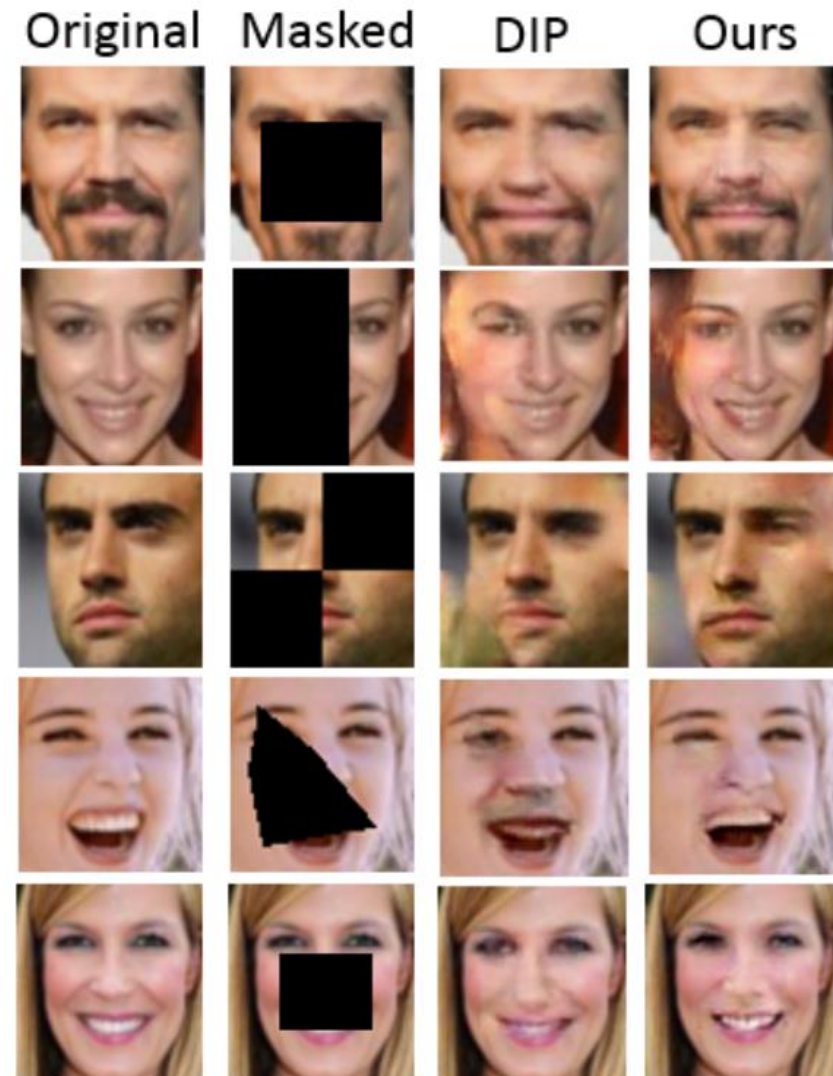
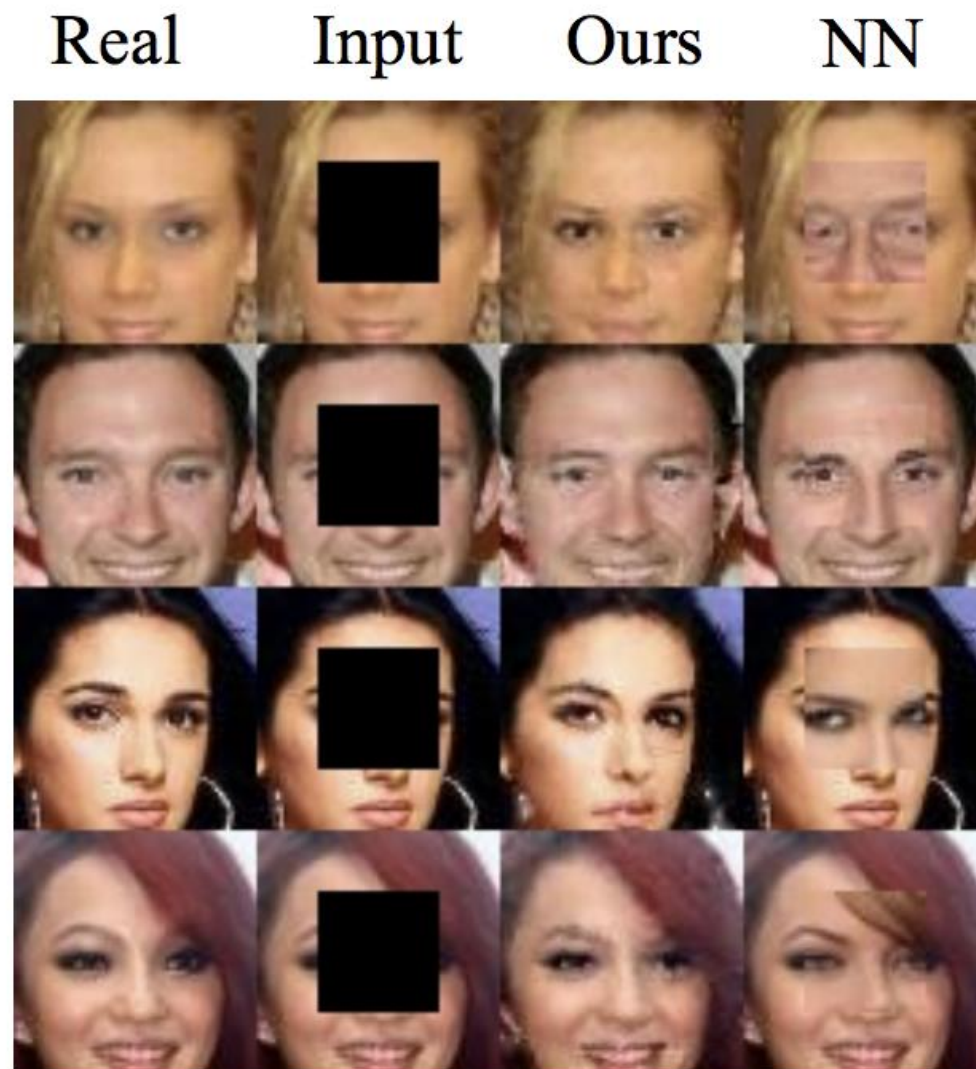


# GAN을 이용하여 유명 화가의 화풍을 흉내 낸 그림 생성

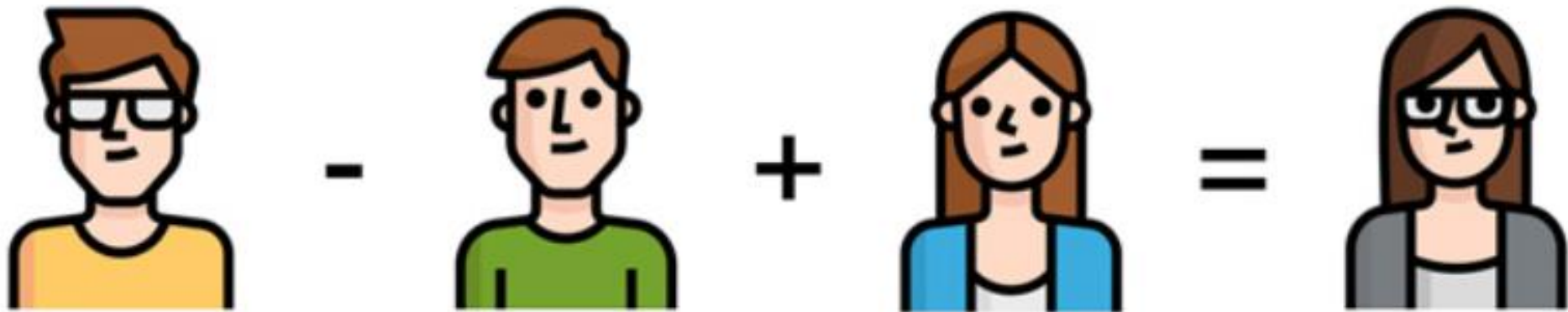


| 캠브리지 컨설턴트가 엔비디아의 기술을 활용해 만든 빈센트 AI <출처: 엔비디아>

# GAN을 이용한 손상된 이미지의 복원

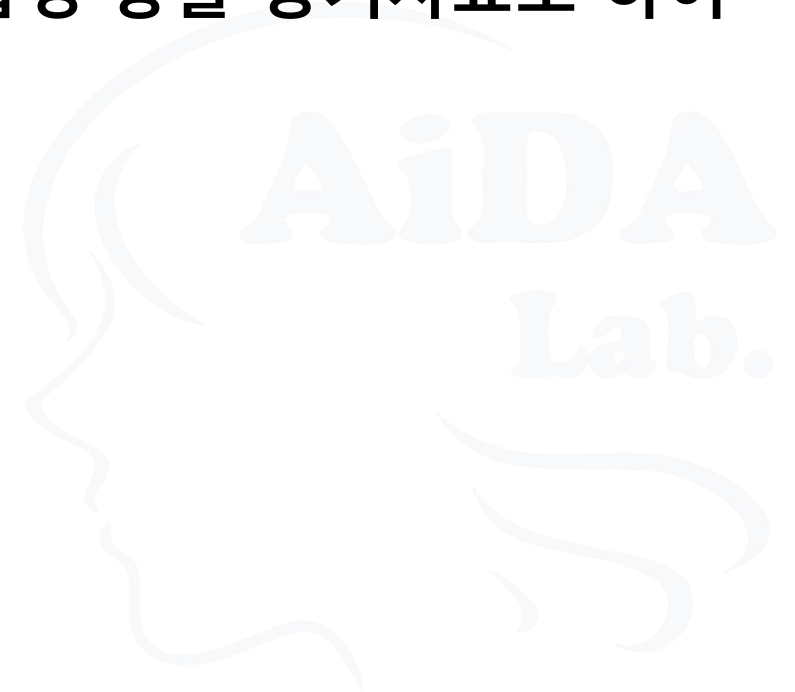


- '안경을 쓴 남자' 이미지를 생성하는  $z$  에서 '안경을 쓰지 않은 남자' 이미지의 입력인  $z$  를 빼고 '안경을 쓰지 않은 여자' 이미지에 해당하는  $z$  를 생성자  $G$  에 넣어주면 '안경을 쓴 여자' 이미지가 아래 그림처럼 생성됨



- GAN 생성자의 결과물을 우리가 원하는 데로 마음껏 조작할 수 있다는 가능성 확인
- 단순한 데이터의 분류로서의 이해가 아닌 새로운 것을 창조할 능력을 가지게 된 것을 의미

- 실제와 구분되지 않는 거짓이 현실 왜곡 가능
  - GAN을 활용한 딥페이크 프로노영상 유통
  - 가짜 뉴스의 범람: 가짜 뉴스에 맞는 영상, 음성 합성 등을 증거자료로 하여 진짜 뉴스처럼 속임





- 영상 합성



| 버락 오바마 전 미국 대통령의 가짜 영상 <출처: 워싱턴대학교>

그림출처(논문): Synthesizing Obama: Learning Lip Sync from Audio.  
Supasorn Suwajanakorn, Steven M. Seitz, Ira Kemelmacher-Shlizerman

- 가짜 뉴스, 가짜 영상을 이용한 사회적 혼란 야기 가능성
- GAN을 이용한 대규모의 시스템 취약점 악용 범죄 초래 가능성



## • 미국

- Microsoft: 2017년 자사 AI 연구인력을 위한 'AI 디자인 원칙', 'AI 윤리 디자인 가이드' 소개
- AI가 효율성을 극대화하되 인류를 위협하지 않고 인류 발전에 기여해야 하며 투명성을 갖추고 기술이 신뢰에 기반해야 한다는 내용
- 아실로마 AI 원칙 발표: 2017년 1월, 테슬라 CEO, 알파고 개발자 데미스 허사비스 등

## • 한국

- 카카오: 2018년 1월, 알고리즘 윤리 헌장 발표

## • 중국 - 사이버공간청(정부)

- 2019년 11월, AI로 음성, 영상을 만들 경우 반드시 사실 공개. 미공개 시 형사처벌

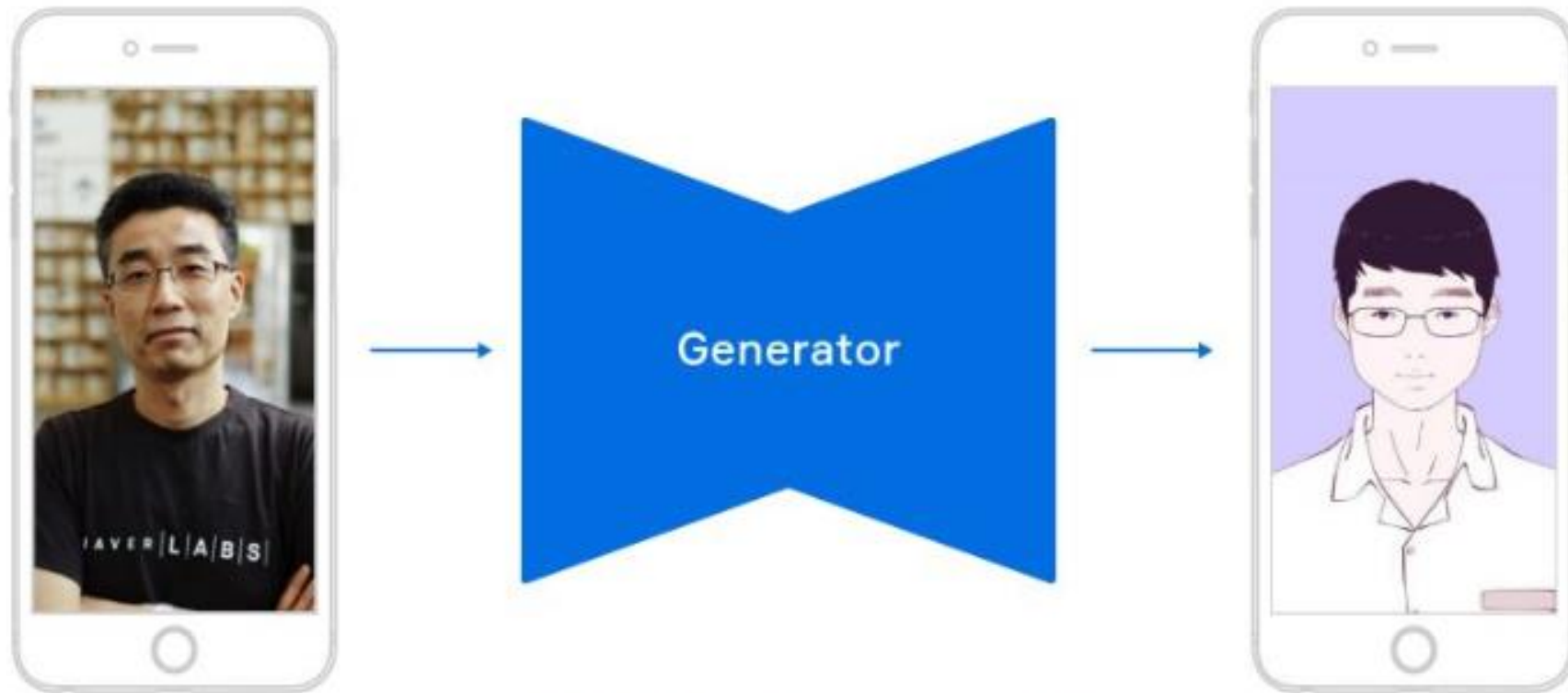
- 시제품 디자인 하기



GAN을 활용해 간단한 스케치 만으로 시제품을 디자인 할 수 있다. (출처: Berkeley AI Research(BAIR))

## • 몰입형 웹툰 제작

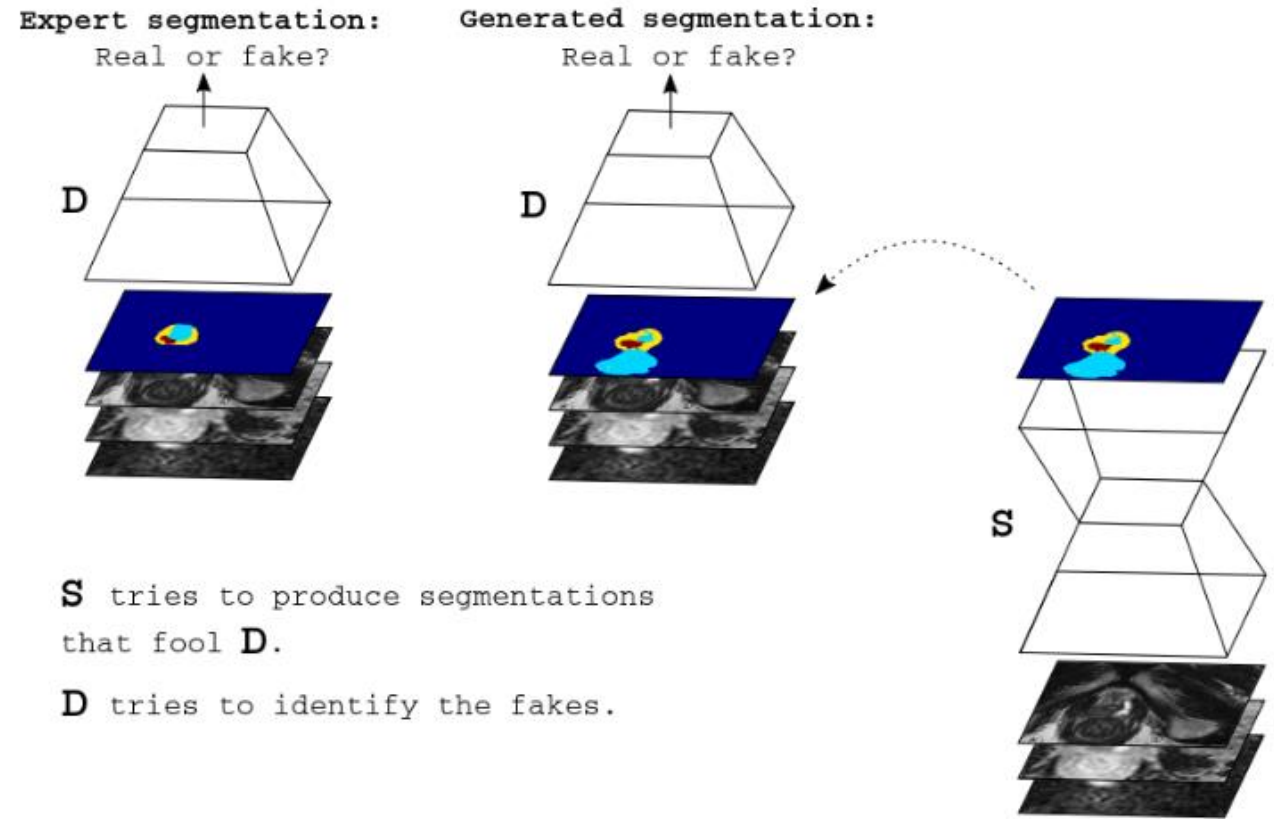
- 독자의 사진을 웹툰 이미지로 생성하여 독자가 웹툰 속 주인공이 되는 경험을 제공함



| GAN을 활용한 네이버 웹툰 '마주쳤다' (출처: 네이버랩스)

## • 의료 영상에 응용

- 의료 영상의 경우 영상 생성보다는 **병변 분할, 영상 변환** 등에 적용되고 있음
- 독일 암센터 연구팀
  - 자기공명영상(MRI)에서 진행성 전립선 암 병변 검출을 위한 **GAN 기반 영상 분할 방법** 제안
  - 전문가의 병변 표식과 분할모델이 생성한 병변 표식 구분 모델 학습 후 결과를 다시 분할모델 학습에 반영하는 것을 반복



[그림 4] GAN을 이용한 진행성 전립선암 병변 분할

- 텍스트 생성

- MIT 연구진

- 수 천개의 이미지와 시를 쌍으로 학습시켜 AI가 이미지를 보고 시를 만들도록 하는 연구
    - 30명의 영문학 전문가를 포함한 500명에게 AI가 만든 시와 인간이 쓴 시를 구별하는 실험
    - 영문학 전문가: 60%만이 AI가 쓴 시를 찾아냄
    - 기타 심사자: 훨씬 못한 결과

- 기타 사례

- 스타트업 알레시오

- 태아의 입체 초음파 사진을 GAN을 응용하여 생후 아기 얼굴로 변환해 주는 서비스 준비

- 그 외에 음성신호, 자연어 처리 등으로 영역 확대 중



- CNN

- 학습: 이미지를 입력받아 확률을 예측하기 위해 입력받은 이미지를 다운 샘플링하는 과정
- 판별: 학습된 데이터는 MaxPooling과 같은 다운샘플링 기술로 처리하여 Label에 매칭

- GAN

- 생성자: 랜덤 노이즈 벡터를 입력받아 이미지를 만드는 업 샘플링 수행 과정
- 판별자: 학습(생성)과정에서 생성된 새로운 데이터를 실제 이미지와 비교하여 0~1 사이의 확률값으로 반환
  - 이미지를 고차원 확률 분포의 샘플로 해석하고 다변량 정규분포 등에 적용하여 확률값 도출

- 머신 러닝을 처음 접할 때는 주로 이미지 분류 문제를 학습  
→ 직관적이므로!!
- 더욱 AI적인 느낌을 주는 생성 모델은 이해하기 어려워서 GAN 등장  
이전에는 크게 주목받지 못함
- 오토 인코더 (Auto Encoder)
  - GAN 이전에 제안된 생성 모델, GAN의 원조 격에 가까움
  - 자료와 연구 결과가 풍부함
  - CycleGAN 등 GAN의 연구에도 활용되며 많은 영향을 미침

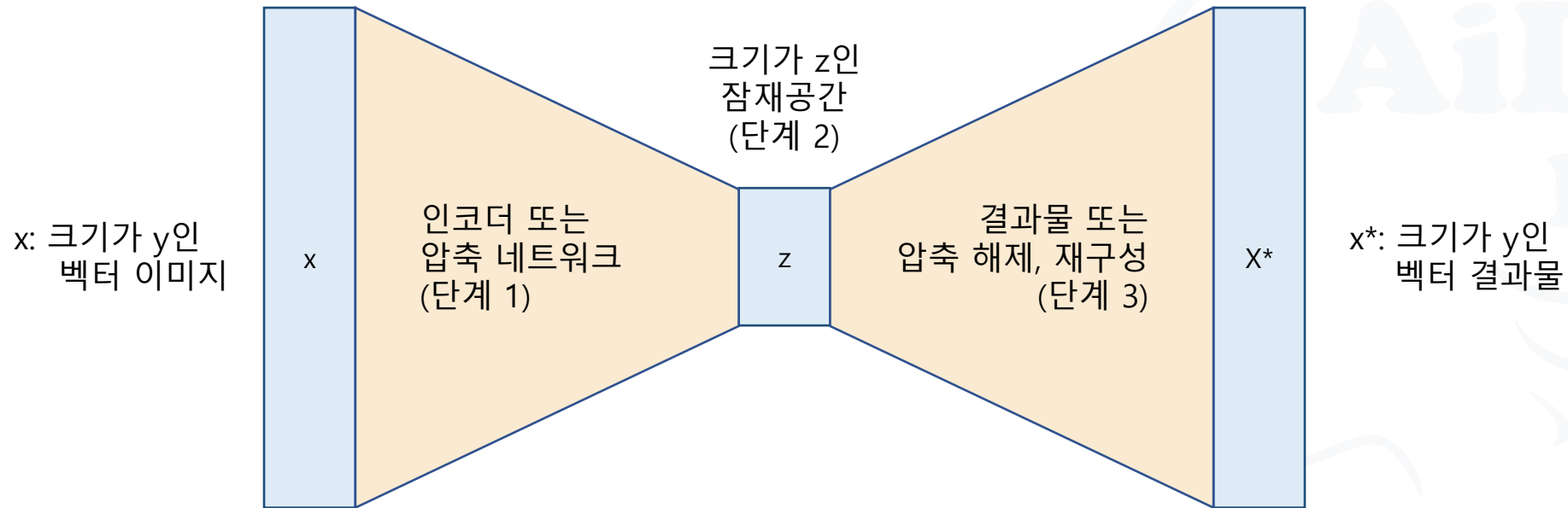


## • 사람의 뇌의 동작

- 알고 있는 개념 설명에 많은 시간을 투자하지 않기 위하여 데이터, 정보를 압축
- 의사 소통의 경우:
  - 다양한 개념, 지식 등이 자동 압축된 오토 인코더로 가득 차 있음
  - 그러나 실제 대화, 소통은 문맥에 종속적
  - 사용자가 처한 환경에 따라 어떤 내용은 설명이 필요하고 어떤 내용은 불필요한지 다름 (개발 관련 내용은 개발자에게 설명이 불필요하지만 일반인에게 필요한 것처럼)
  - 반복되는 개념 등은 사전에 동의한 추상적인 표현으로 압축, 필요한 내용은 상세하게  
→ 정보의 처리량은 증대 시키고 전송을 위한 대역폭은 축소 시킴

- 오토 인코더

- 데이터를 자동으로 인코딩 할 수 있게 도와주는 모델
- 인코더와 디코더로 구성됨

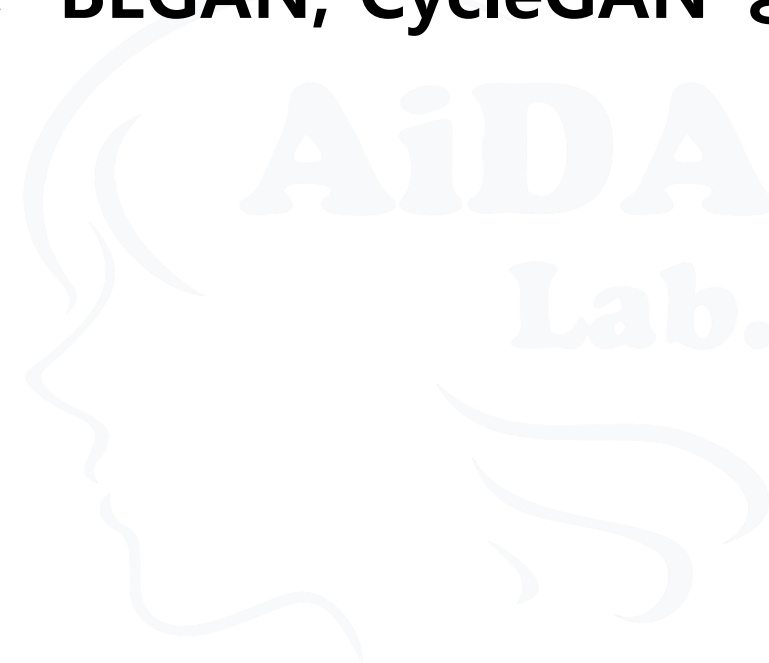


- 오토 인코더의 동작

- 이미지  $x$ 를 오토 인코더에 입력
- 재구성된 이미지  $x^*$ 를 획득
- $x$ 와  $x^*$ 의 차이인 재구성 손실 측정
  - $x$ 와  $x^*$ 의 픽셀 간 거리(예, 평균 제곱 오차)를 함수에 사용
  - 경사 하강법 적용

- 축소된 잠재 공간에서 아이템과 타겟 클래스의 유사도를 빠르게 확인할 수 있는 1클래스 분류기에 활용 → 정보 검색, 이상치 탐지(잠재 공간 안에서 거리 비교) 분야 등에 활용
- 데이터의 노이즈 제거
- 흑백 이미지의 채색 (GAN도 좋은 성능을 보임)
- 새로운 이미지의 생성
  - 이미지 데이터 셋을 통해 훈련된 오토 인코더에
  - 입력된 데이터를 기반으로 기존에 저장된 개념(학습된 내용)이 잠재공간에 놓인 위치를 찾아서
  - 이미 본 적이 있는 유사 데이터를 꺼내어 활용 → 생성

- 레이블된 데이터가 필요하지 않음 (비지도학습, 자기 훈련)
- 단순한 구조와 이미 보유한 데이터에서 새로운 표현을 찾는 능력  
→ 다른 모델에 적용, 활용, 응용이 용이함(예, BEGAN, CycleGAN 등)



- DCGAN (Deep Convolutional GAN)

- 2016년 소개됨
- GAN 모델에서 더욱 강력한 신경망 구조를 적용하면 어떨까? 라는 아이디어로 시작
- GAN이 보유한 간단한 2개 층(생성자, 판별자)의 순방향 신경망 대신 CNN으로 생성자와 판별자를 구현한 모델
- GAN과 CNN을 함께 사용하는 많은 연구에서 불안정성, 경사가 너무 작아서 학습이 느려지는 Gradient 포화문제로 GAN의 훈련이 어려움을 겪음
- 각 층의 입력을 정규화하여 안정적으로 훈련하게 도와주는 배치 정규화 기법을 적용, CNN으로 생성자, 판별자를 구현하여 GAN과 CNN을 완전하게 통합시킴

- ProGAN

- Full HD 화질로 실제 사진같은 이미지를 생성하는 최신 기법(2018)
- 4가지 혁신적인 기법 도입
  - 고해상도 층으로 점진적 증가와 단계적 도입
  - 미니배치 표준편차
  - 균등 학습률
  - 픽셀별 특성 정규화

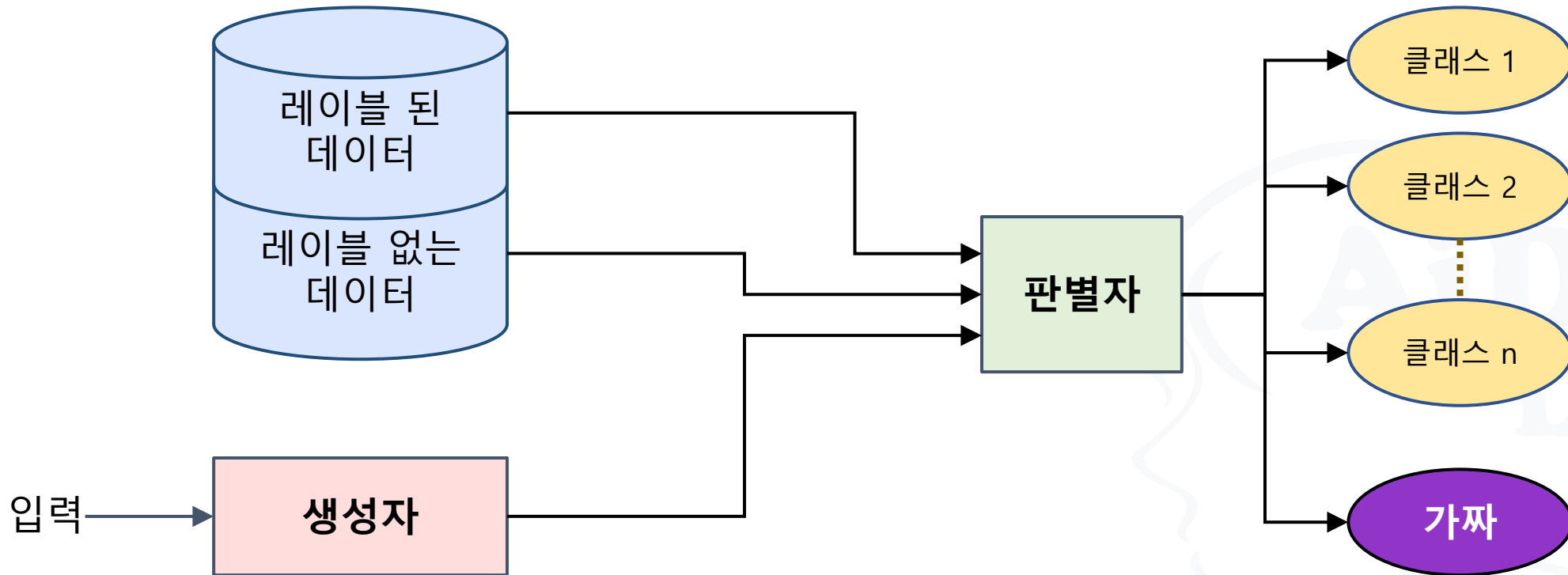


- **SGAN (Semi-supervised GAN)**

- GAN은 지도학습인 딥러닝 모델 중심의 패러다임을 비지도학습의 영역으로 끌어당김
- SGAN은 다시 일부 특성을 지도학습의 방향으로 끌어당겨 준지도학습을 기반으로 하는 GAN
  - 레이블 된 데이터와 레이블이 없는 데이터를 동일한 분포로 수집하여 활용
  - 데이터에 감춰진 내부 구조를 사용하여 레이블 된 데이터 포인트를 일반화
  - 이전에 본 적 없는 새로운 샘플을 분류하는데 활용



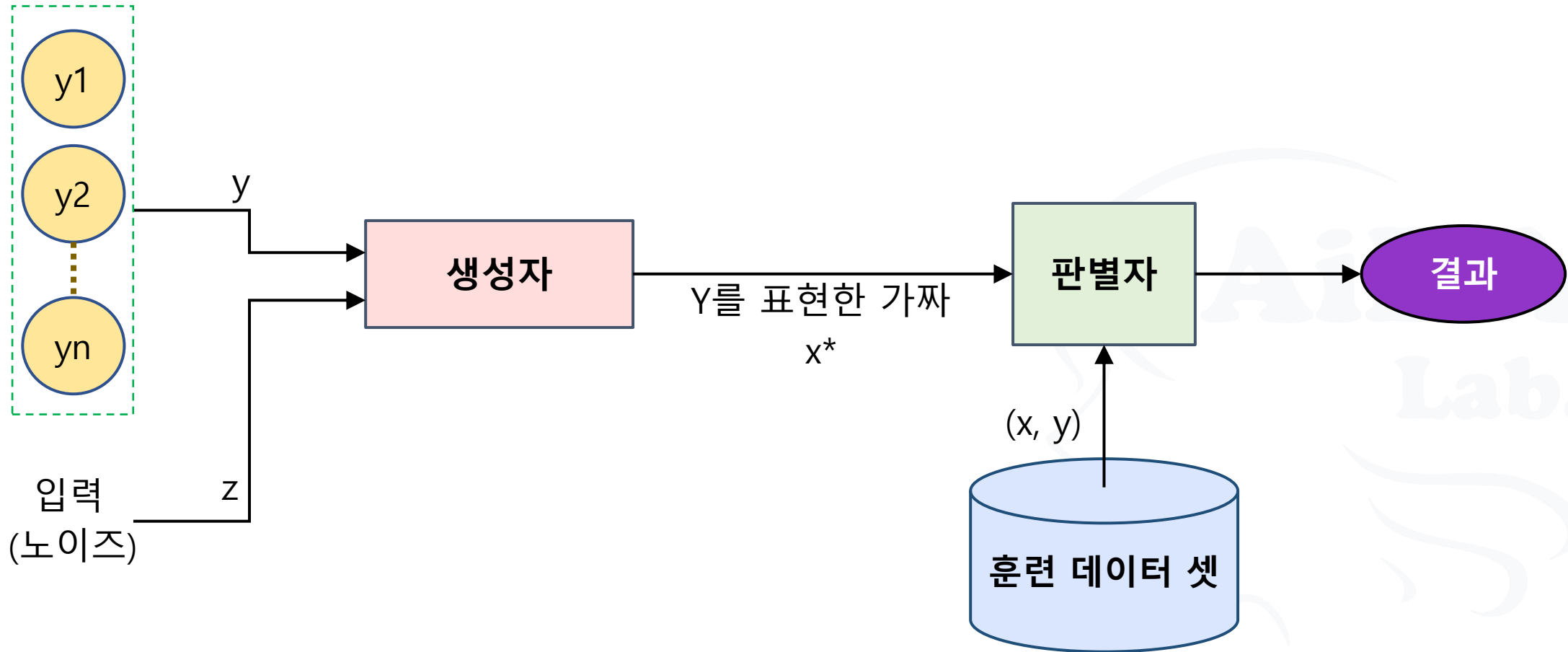
- SGAN (Semi-supervised GAN)



- **CGAN (Conditional GAN)**

- 생성자와 판별자를 훈련할 때 모두 레이블을 사용하는 GAN
- 생성자는 원하는 가짜 샘플을 합성할 수 있게 됨
  - 기존의 GAN, DCGAN 등은 훈련 데이터 셋을 바꾸어 학습하는 샘플의 종류는 조정 가능
  - 그러나 생성하는 샘플의 특징은 지정 불가능
    - 예: DCGAN으로 손글씨 숫자를 합성할 수 있음. 그러나 숫자 9가 아니라 숫자 7을 생성하도록 명령할 수 없음

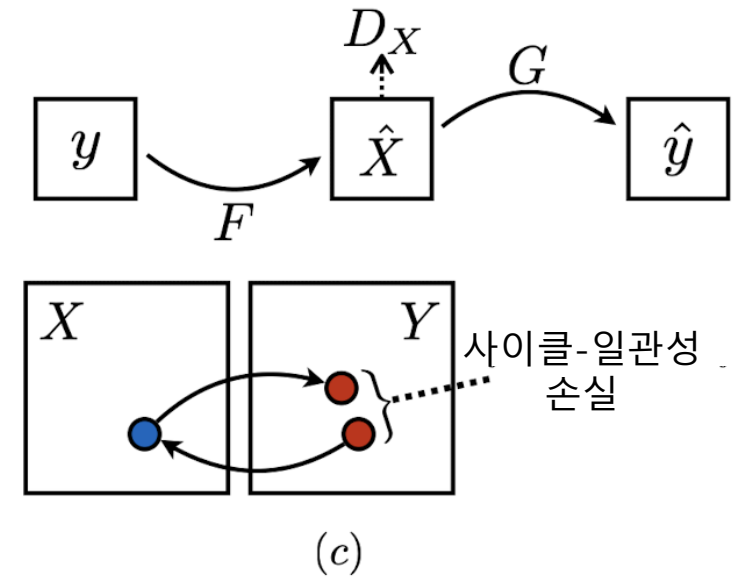
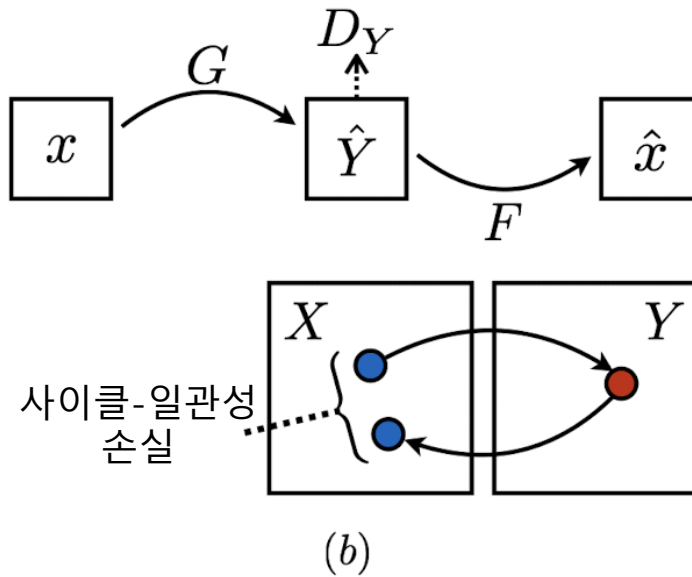
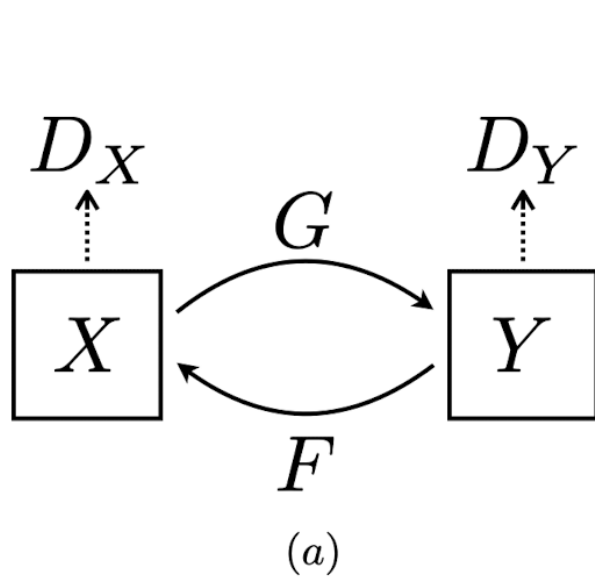
- CGAN (Conditional GAN)



- CycleGAN

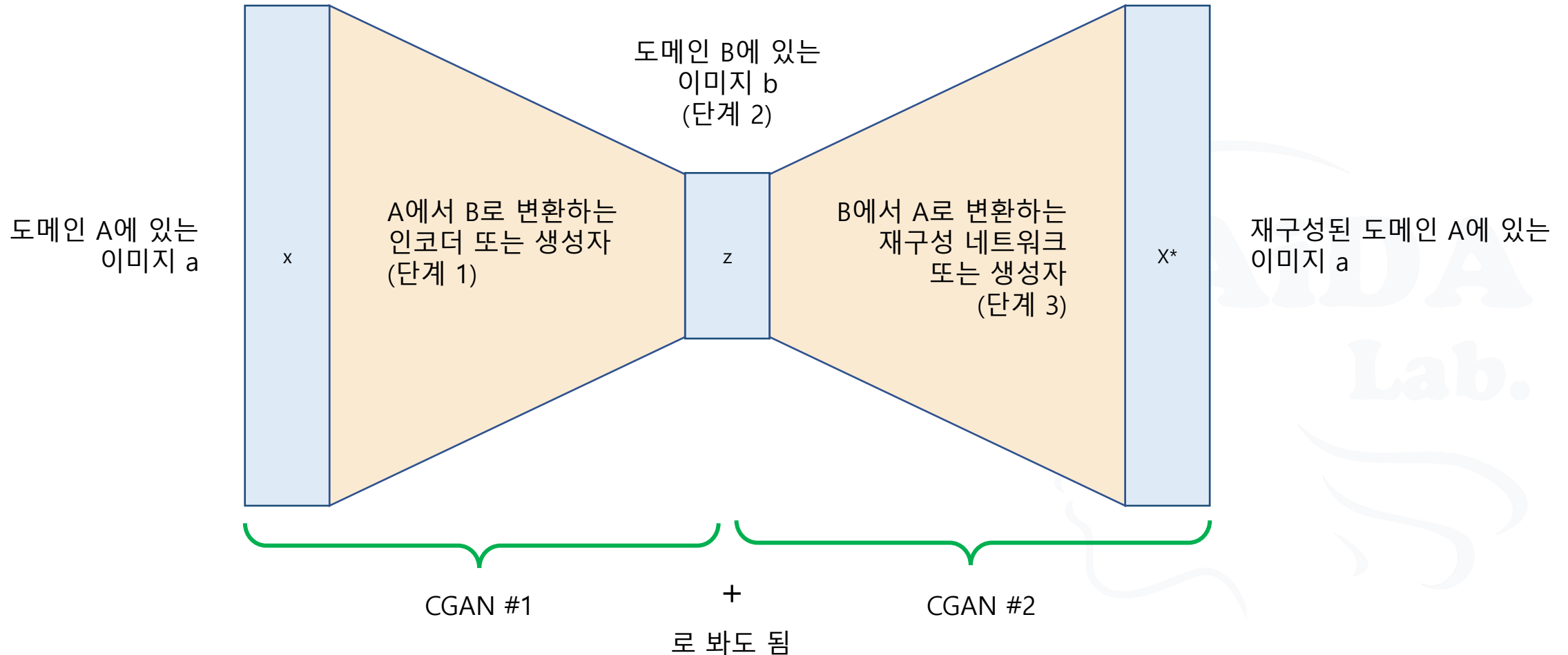
- 이미지 변환에 강점을 가진 GAN
- CGAN의 특별한 형태로 보기도 함
- 입력 이미지 자체에 레이블의 역할을 부여함
  - 다른 도메인에 정확히 같은 이미지로 매핑해야만 사용 가능
  - 캘리포니아 대학교 버클리 그룹의 연구: Cycle을 구성하면 완벽한 쌍이 아니어도 적용 가능
  - CycleGAN

## • CycleGAN



- 손실이 양방향으로 발생  $\rightarrow$  양방향으로 변환 가능  
(예, 여름사진  $\rightarrow$  겨울사진  $\rightarrow$  여름사진..)

## • CycleGAN



- GAN은 훈련이 복잡하고 어렵다

- 모드 붕괴

- 실제로 제공된 훈련 데이터 셋에 포함되어 있음에도 불구하고 일부 모드(클래스 등)가 생성된 샘플에서 잘 나타나지 않는다

- 예, 숫자를 생성하는 모델에서 생성된 MNIST 데이터 셋에 "8" 이 존재하지 않음

- 느린 수렴:

- GAN 및 비지도학습에서의 심각한 문제 (훈련 종료 시점 문제)

- GAN은 훈련이 복잡하고 어렵다

- 과잉 일반화

- 지원할 필요가 없는(존재하지 않는) 모드(잠재적인 데이터 샘플)가 발생한다

- 신경망 깊이 늘리기

- 설정 바꾸기

- 다양한 훈련기법 적용

등으로 보완 가능함





# GAN (Generative Adversarial Networks)



- 다양한 문제점에도 불구하고 뛰어난 성능으로 인해 각광받는 모델
- 일반 인공지능으로 가는 디딤돌이라고 평가받을 정도로 전망이 밝음
- GAN은 과학보다 예술에 가깝다는 평가처럼 다양한 가능성을 보유
- 이미지 변환 등의 능력에 따른 위험성에 따라 AI 윤리, 도덕적 가치관에 대한 인식의 필요성이 있음