

# 꼼꼼한 딥러닝 논문 리뷰와 코드 실습

Deep Learning Paper Review and Code Practice

나동빈([dongbinna@postech.ac.kr](mailto:dongbinna@postech.ac.kr))

Pohang University of Science and Technology

# SinGAN: Learning a Generative Model from a Single Natural Image (ICCV 2019)

- 본 논문은 오직 한 장의 이미지를 이용해 GAN 네트워크를 학습합니다.
- SinGAN은 다양한 애플리케이션(application)에서 활용 가능합니다.
  - 한 장의 이미지에 대한 이미지 수정(editing), 애니메이션(animation), 초해상도(super-resolution) 등

*Single* training image

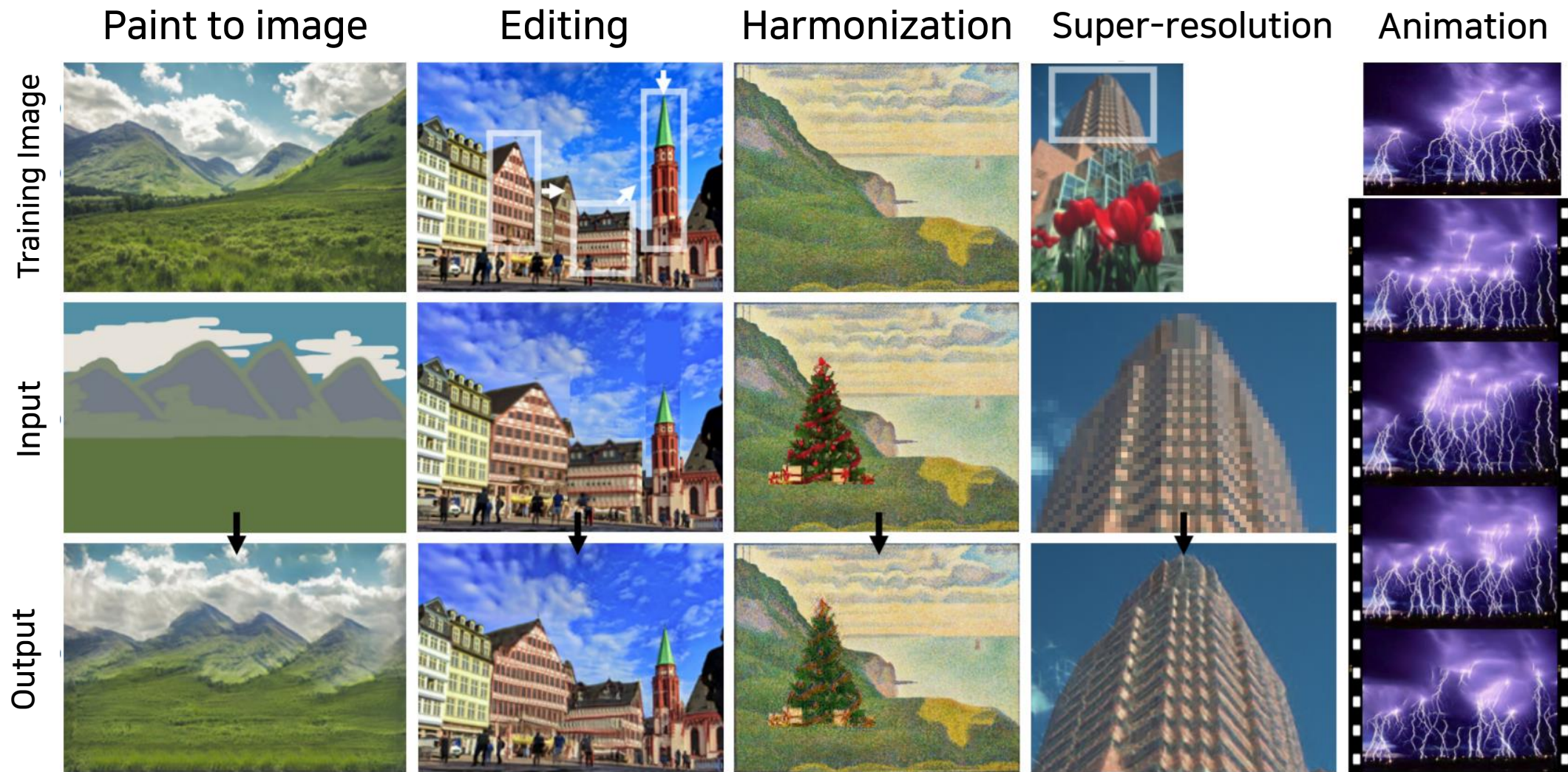


Random samples from a *single* image





# SinGAN을 활용한 다양한 애플리케이션 예시



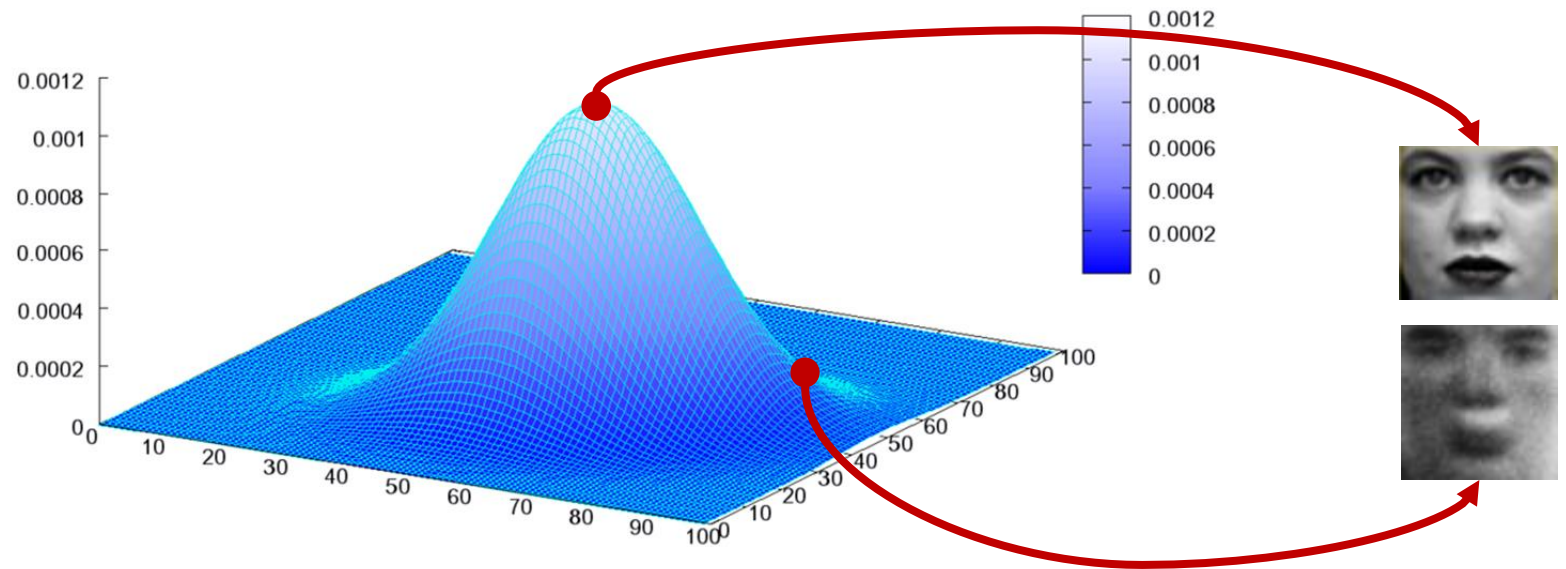
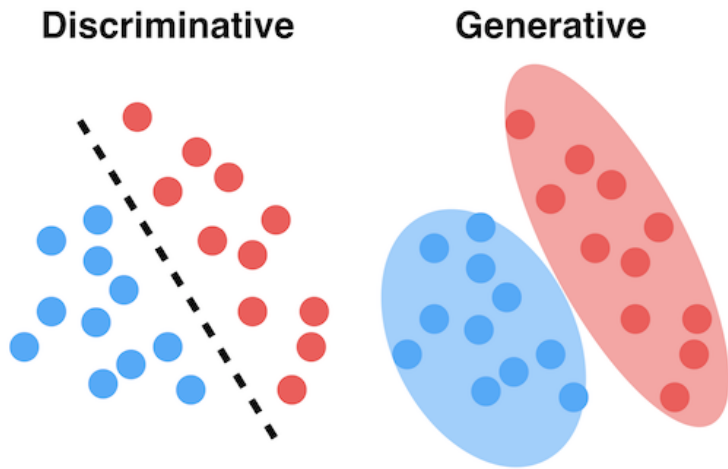
## 연구 배경: Generative Adversarial Network (GAN)

## 연구 배경: 생성 모델 (Generative Models)

- 생성 모델은 실존하지 않지만 있을 법한 이미지를 생성할 수 있는 모델을 의미합니다.

**Generative Model** (produce) → An image that does not exist but is likely to exist

- A statistical model of the joint probability distribution
- An architecture to generate new data instances



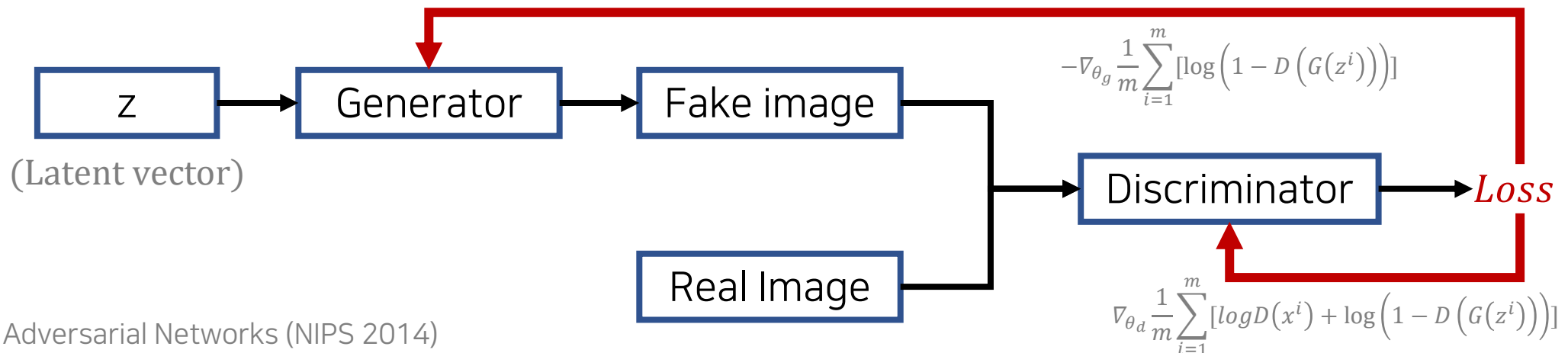
# 연구 배경: Generative Adversarial Networks (GAN)

- 생성자(generator)와 판별자(discriminator) 두 개의 네트워크를 활용한 생성 모델입니다.
- 다음의 목적 함수(objective function)를 통해 생성자는 이미지 분포를 학습할 수 있습니다.

$$\min_G \max_D V(D, G) = E_{x \sim p_{data}(x)} [\log D(x)] + E_{z \sim p_z(z)} [\log(1 - D(G(z)))]$$

Generator  $G(z)$ : new data instance

Discriminator  $D(x)$  = Probability: a sample came from the real distribution (Real: 1 ~ Fake: 0)



Generative Adversarial Networks (NIPS 2014)



## 연구 배경: DCGAN (ICLR 2016)

- Deep Convolutional Layers를 이용하여 이미지 도메인에서의 높은 성능을 보입니다.

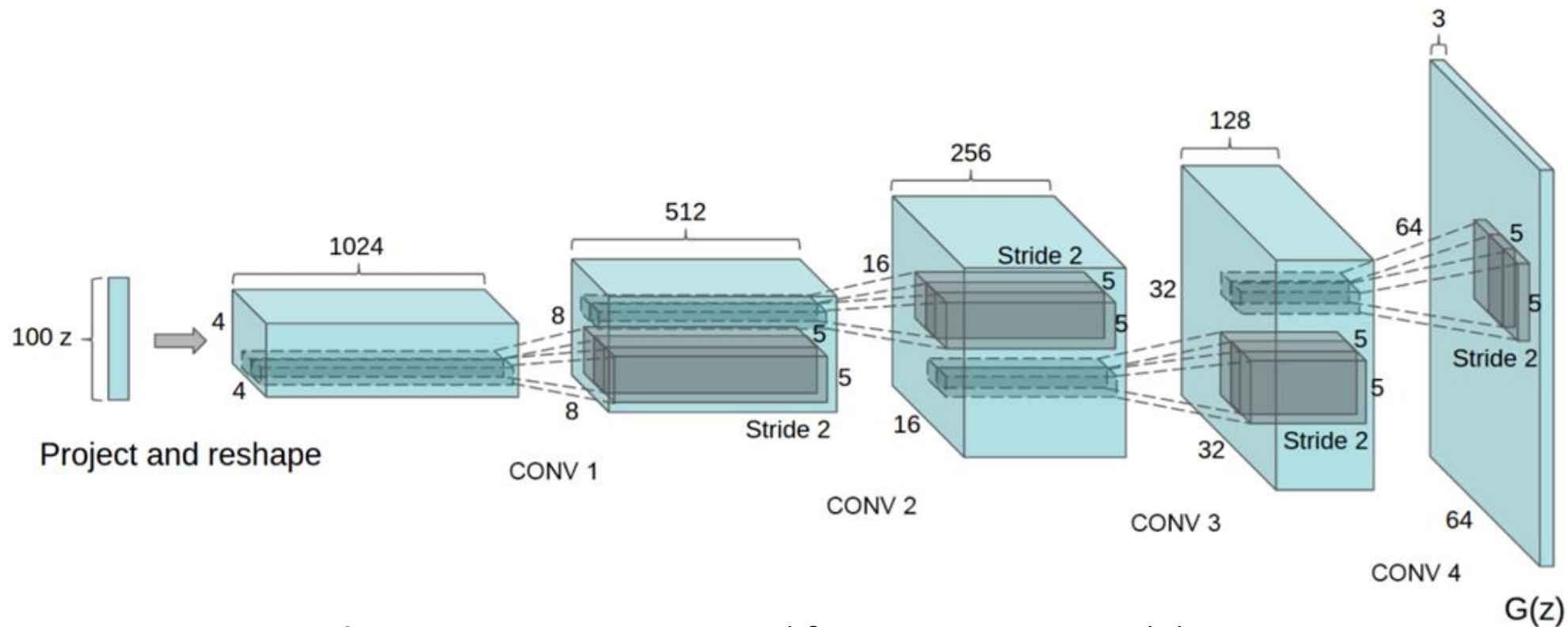
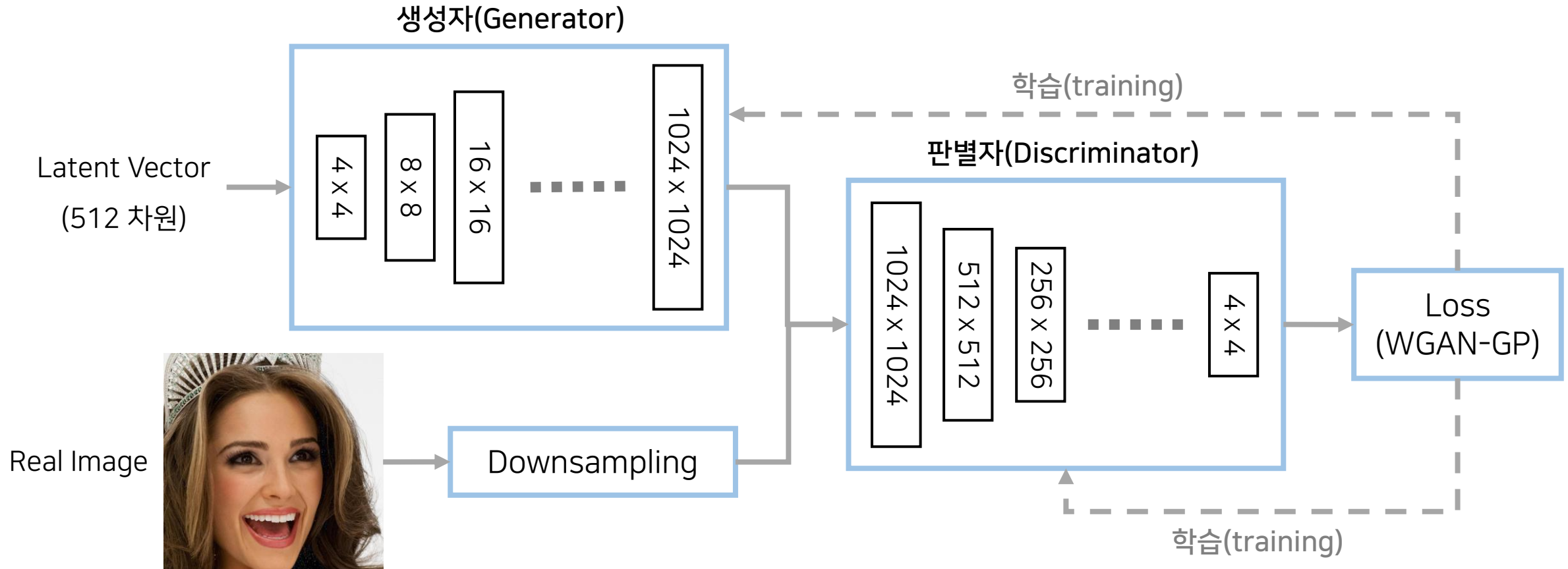


Fig. DCGAN generator used for LSUN scene modeling.

Unsupervised Representation Learning with Deep Convolutional Generative Adversarial Networks (ICLR 2016)

## 연구 배경: Progressive Growing of GANs (PGGAN = ProGAN)

- 학습을 진행하는 과정에서 점진적으로(progressively) 네트워크의 레이어를 붙여 나갑니다.



<https://towardsdatascience.com/explained-a-style-based-generator-architecture-for-gans-generating-and-tuning-realistic-6cb2be0f431>

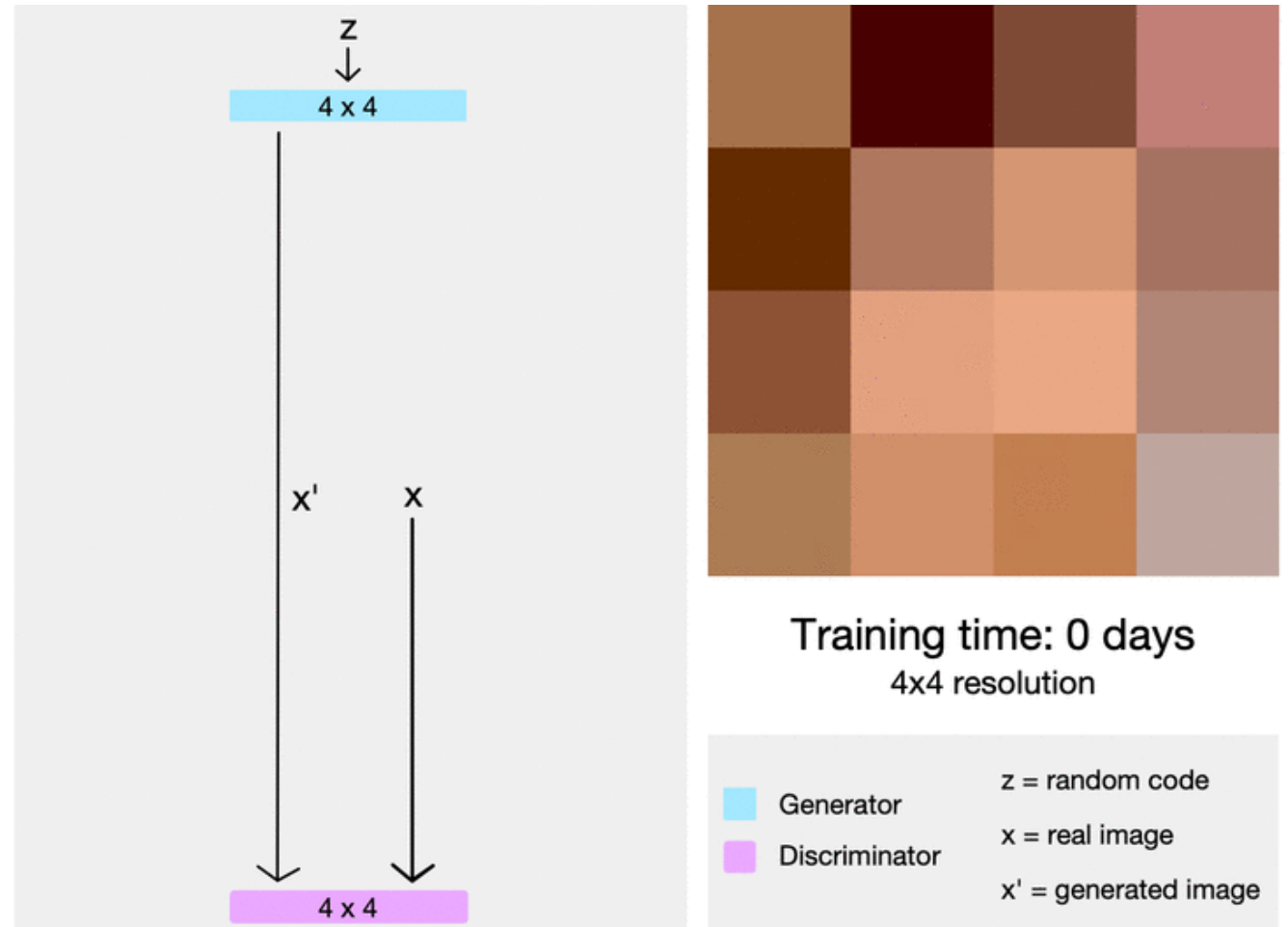


## 연구 배경: Progressive Growing of GANs (PGGAN = ProGAN)

- 메인 아이디어
  - 학습 과정에서 레이어를 추가
  - 고해상도 이미지 학습 성공
- 한계점
  - 이미지의 특징 제어가 어려움



추후에 StyleGAN 논문에서 개선됩니다.

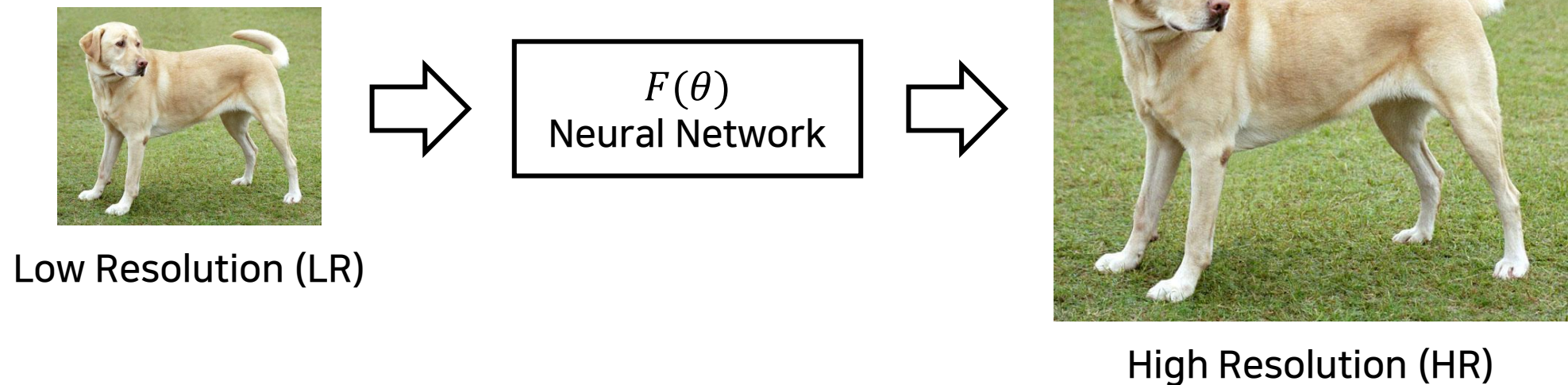


<https://towardsdatascience.com/progan-how-nvidia-generated-images-of-unprecedented-quality-51c98ec2cbd2>

## 연구 배경: 초해상도(Super-Resolution) 기술

## 연구 배경: Single Image Super-Resolution (SISR)

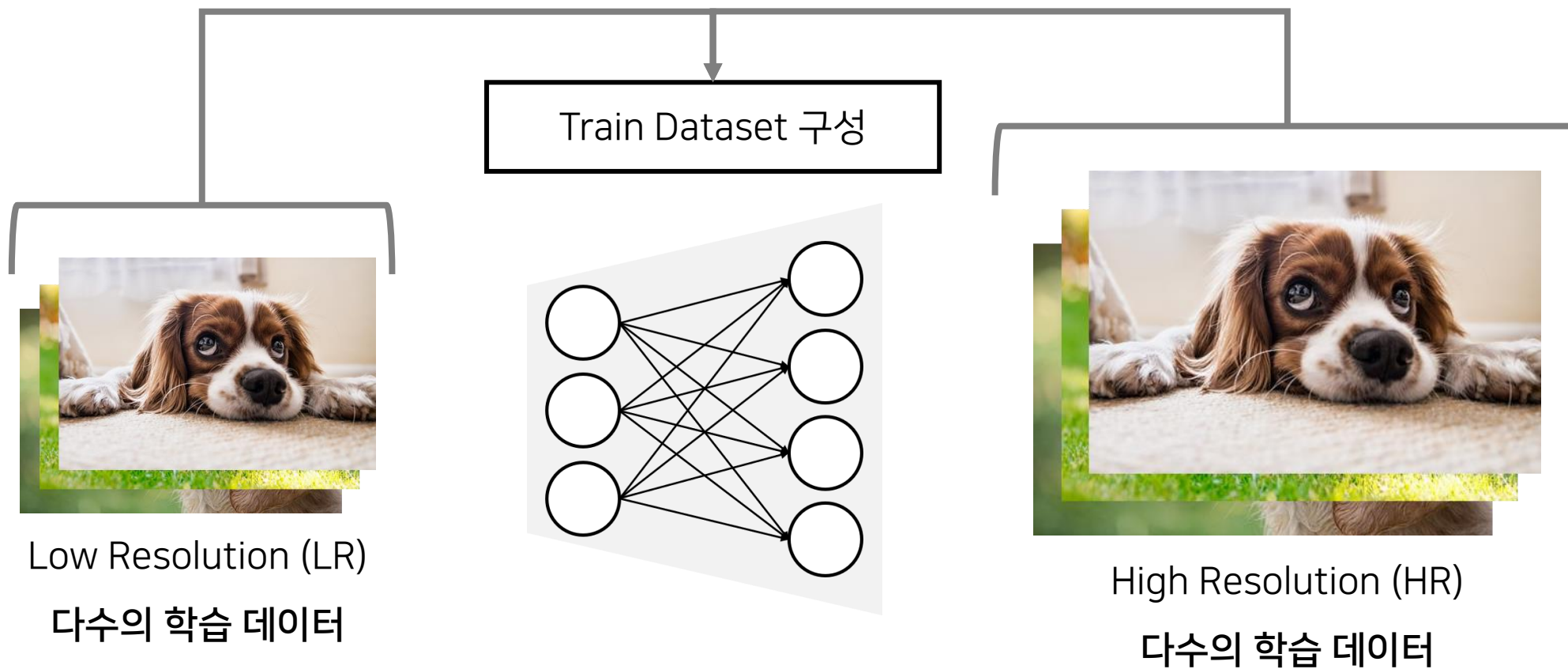
- 한 장의 저해상도 이미지(LR)를 고해상도 이미지(HR)로 변환하는 방법을 연구하는 분야입니다.
  - 최근 CNN 기반의 접근 방식이 높은 성능을 보이고 있습니다.



## 학습 방법에 따른 분류: **Externally** Trained Network (Supervised SISR)

- 학습 시기: 다수의 HR-LR 쌍에 대하여 학습을 진행합니다.

$$\mathbf{I}_{LR}^k = (\mathbf{I}_{HR} * \mathbf{k}) \downarrow_s + \mathbf{n}$$



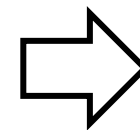
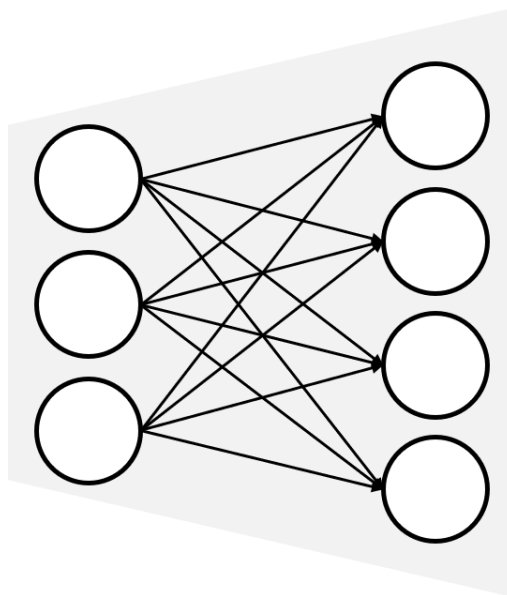
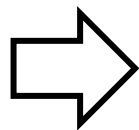
## 학습 방법에 따른 분류: **Externally** Trained Network (Supervised SISR)

- 테스트 시기: 다수의 데이터로 학습된 정보를 토대로 현재 테스트 데이터에 대한 고해상도 결과를 예측합니다.



Low Resolution (LR)

테스트 데이터



High Resolution (HR)

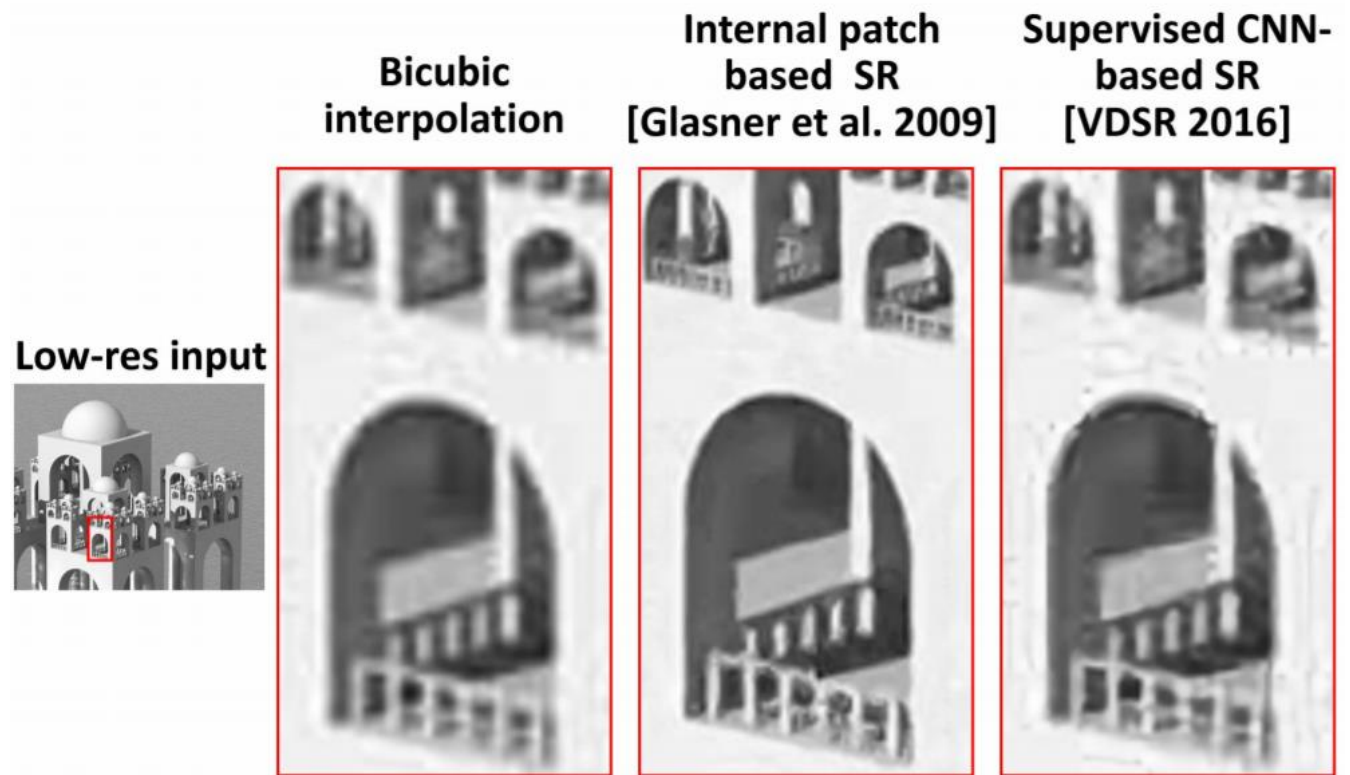
예측 결과



# 이미지의 자기 반복성

- 한 장의 이미지에도 자기 반복성(internal recurrence)이 존재합니다.
  - 많은 자연의 이미지들은 내재된 데이터 반복(internal data repetition)을 특징으로 가집니다.

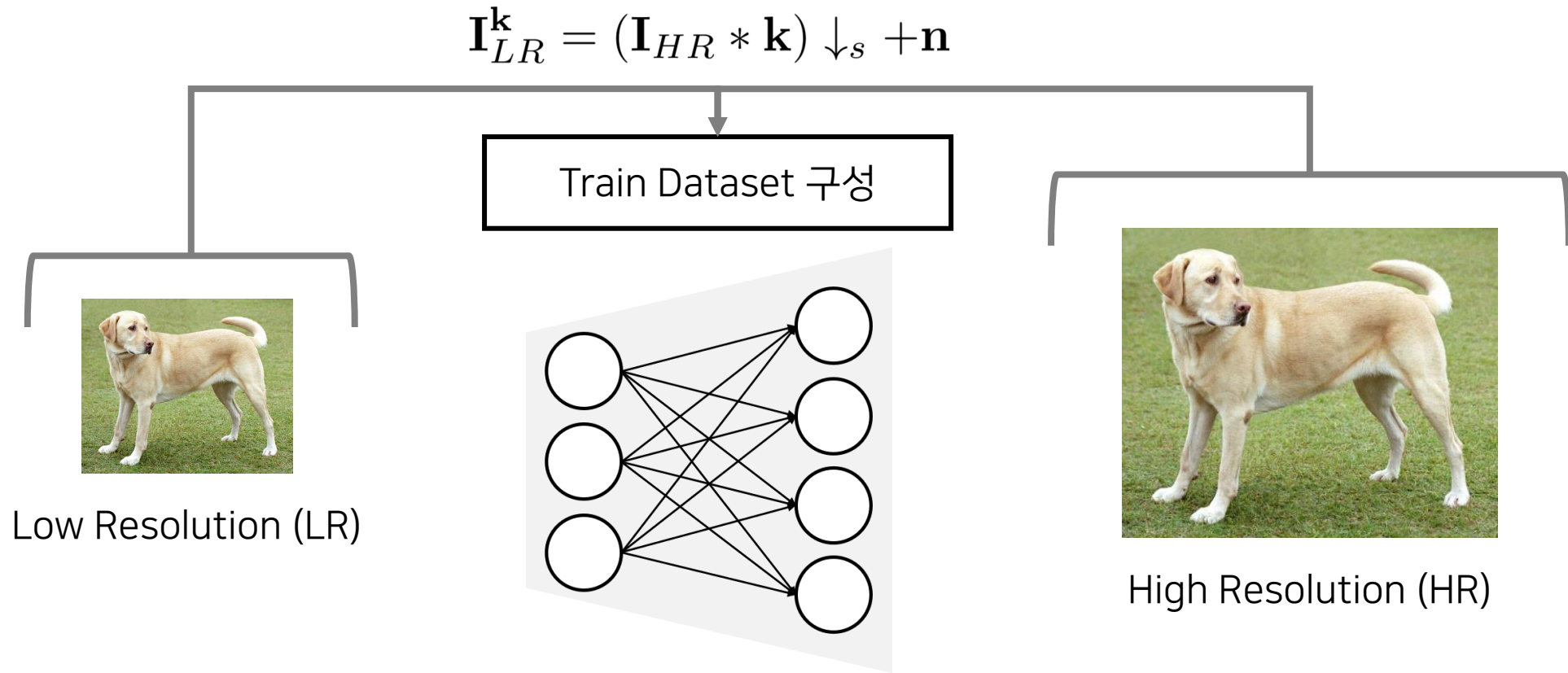
"The only evidence to the existence of these tiny handrails exists internally, inside this image, at a different location and different scale. It cannot be found in any external database of examples, no matter how large this dataset is!"



"Zero-Shot" Super-Resolution using Deep Internal Learning (CVPR 2018)

## 학습 방법에 따른 분류: Internally Trained Network (Unsupervised “Zero-Shot” SISR)

- 학습 시기: **테스트 이미지 자기 자신**으로부터 추출된 HR-LR 쌍에 대하여 학습을 진행합니다.
  - 결과적으로 한 장의 이미지에 특화된 CNN이 학습됩니다.



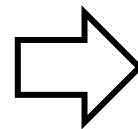
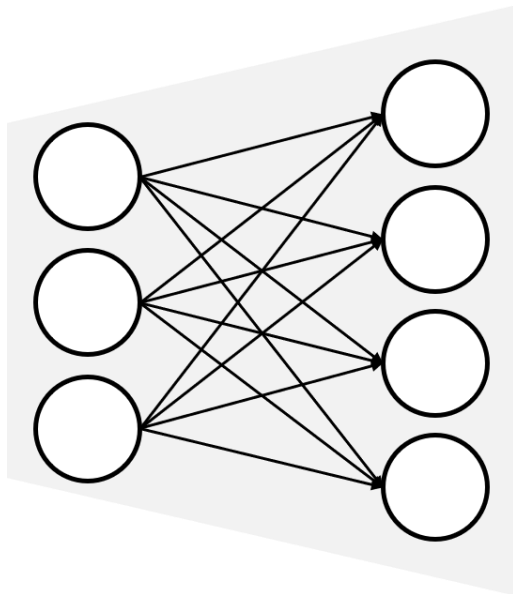
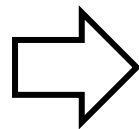
## 학습 방법에 따른 분류: Internally Trained Network (Unsupervised “Zero-Shot” SISR)

- 테스트 시기: 한 장의 이미지에 내재된 특징 정보를 토대로 고해상도 결과를 예측합니다.



Low Resolution (LR)

한 장의 테스트 데이터 (원본)



High Resolution (HR)

예측 결과



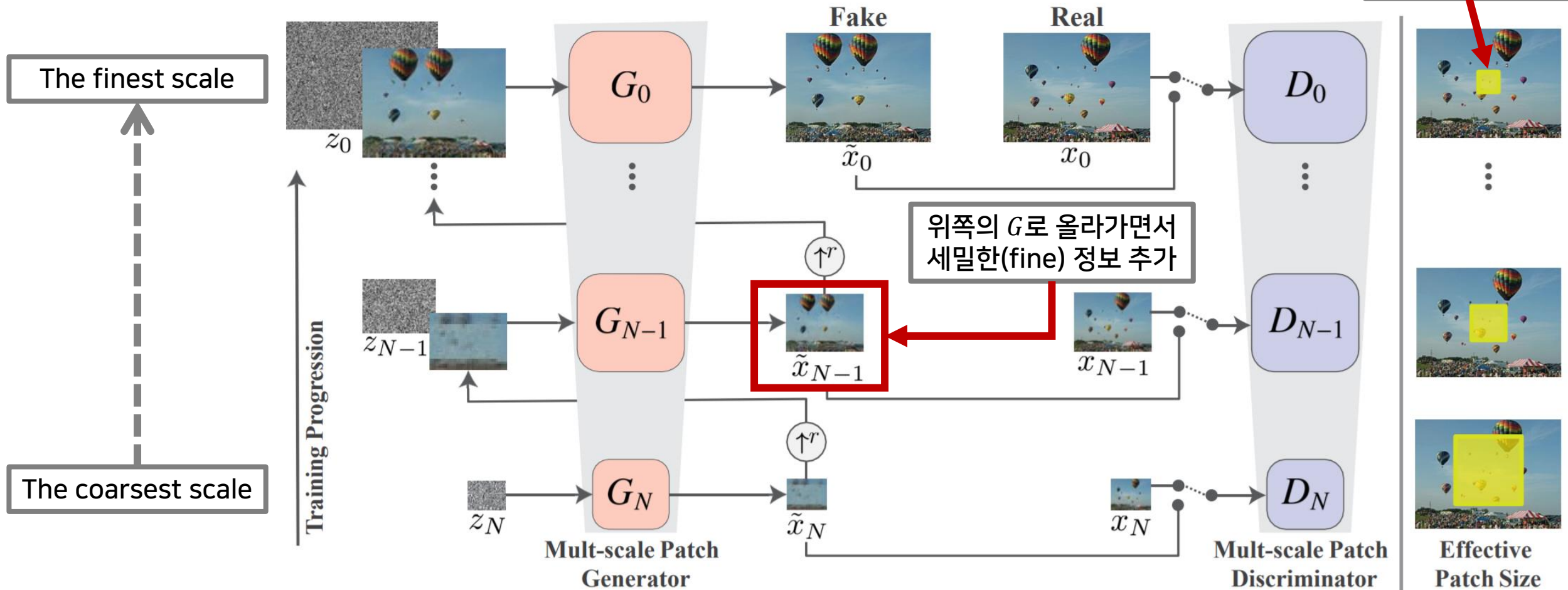
어떻게 더 효과적인 아키텍처를 만들 수 있을까요?

## 논문 소개: SinGAN (ICCV 2019)

# SinGAN 전체 파이프라인 설명

- 총  $N + 1$ 개의 가벼운 개별 GAN 네트워크(G와 D)를 학습합니다. ( $N$  번의 업샘플링 진행)

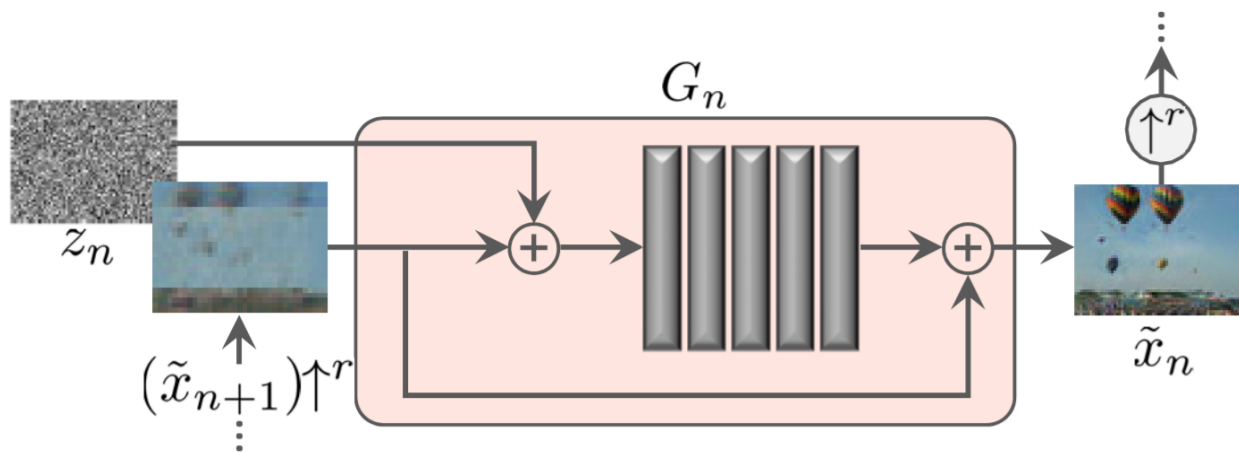
패치 단위 판별  
(discriminator)





## SinGAN 아키텍처: Single Scale Generation

- 하나의 스케일(scale)에서 사용되는 생성자(generator)는 다음의 아키텍처로 구성됩니다.
  - 잔여 학습(residual learning)을 이용하여 세밀한(fine) 정보를 추가하는 방식을 사용합니다.
  - 생성자는 5개의 블록으로 구성됩니다. (각 블록은  $Conv(3 \times 3)$ - $BatchNorm$ - $LeakyReLU$  형식)



- 각각의 생성자  $G_n$  ( $n < N$ )는 다음의 연산을 수행합니다.

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

## SinGAN의 목적 함수(Objective Function)

- 각 GAN 모델은 다음의 목적 함수(objective function)를 이용해 학습을 진행합니다.
  - 한 번 학습이 완료되면 coarse scale의 생성자는 고정되어 업데이트되지 않습니다.

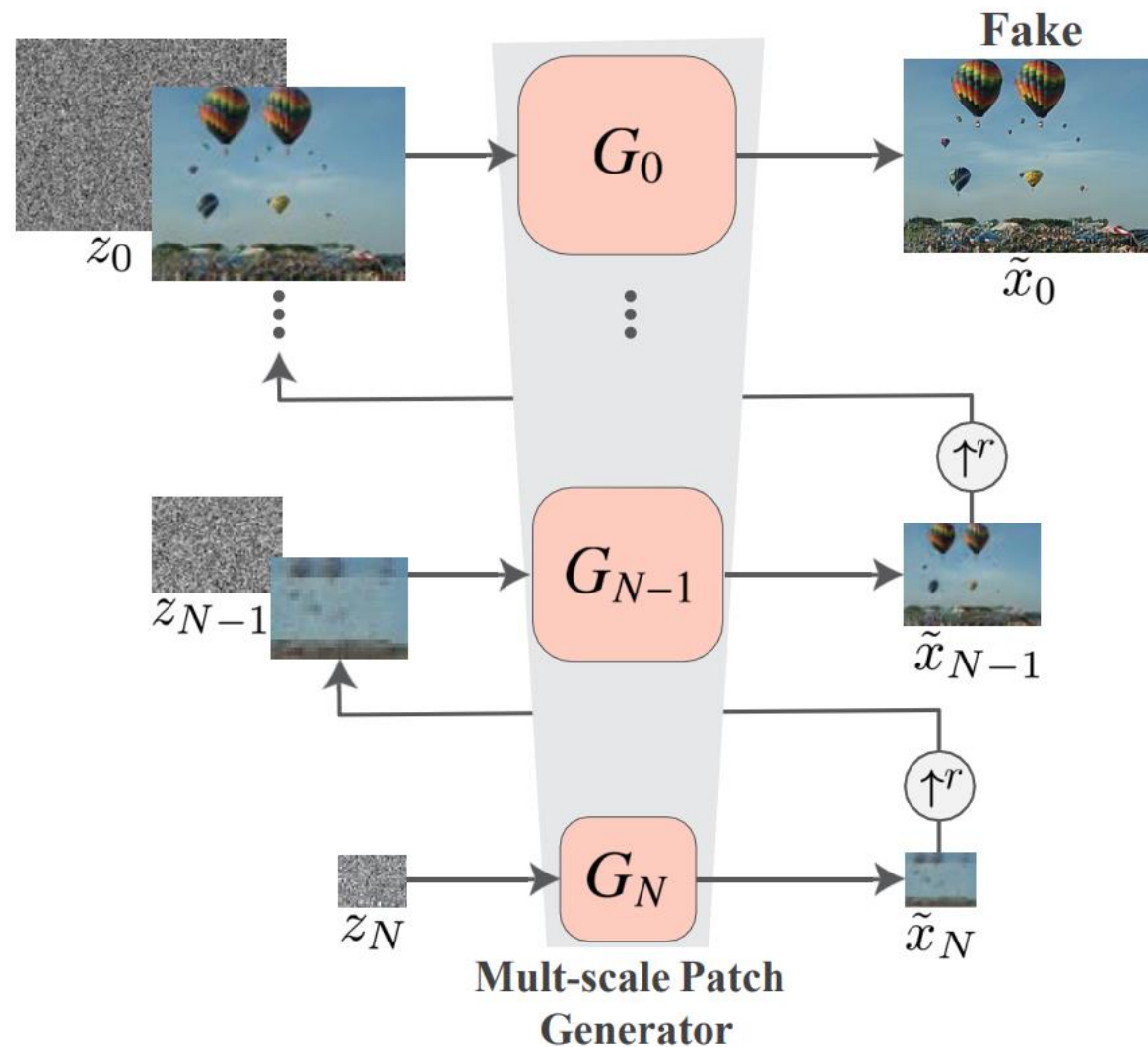
$$\min_{G_n} \max_{D_n} \mathcal{L}_{\text{adv}}(G_n, D_n) + \alpha \mathcal{L}_{\text{rec}}(G_n)$$

- **Adversarial loss**  $\mathcal{L}_{\text{adv}}(G_n, D_n)$ : 실제 이미지  $x_n$  내 패치와 가짜 이미지  $\tilde{x}_n$  내 패치의 분포가 같도록 학습
  - Markovian discriminator: 입력의 각 overlapping된 패치들에 대해서 진위 여부를 판별합니다.
  - WGAN-GP: GAN 네트워크를 안정적으로 학습하도록 도와줍니다.
- **Reconstruction loss**:  $\mathcal{L}_{\text{rec}}(G_n)$ : 실제 이미지  $x_n$ 를 정확히 생성할 수 있도록 학습
  - 노이즈의 값으로 0을 넣었을 때, 실제 이미지와 동일한 이미지를 생성합니다.

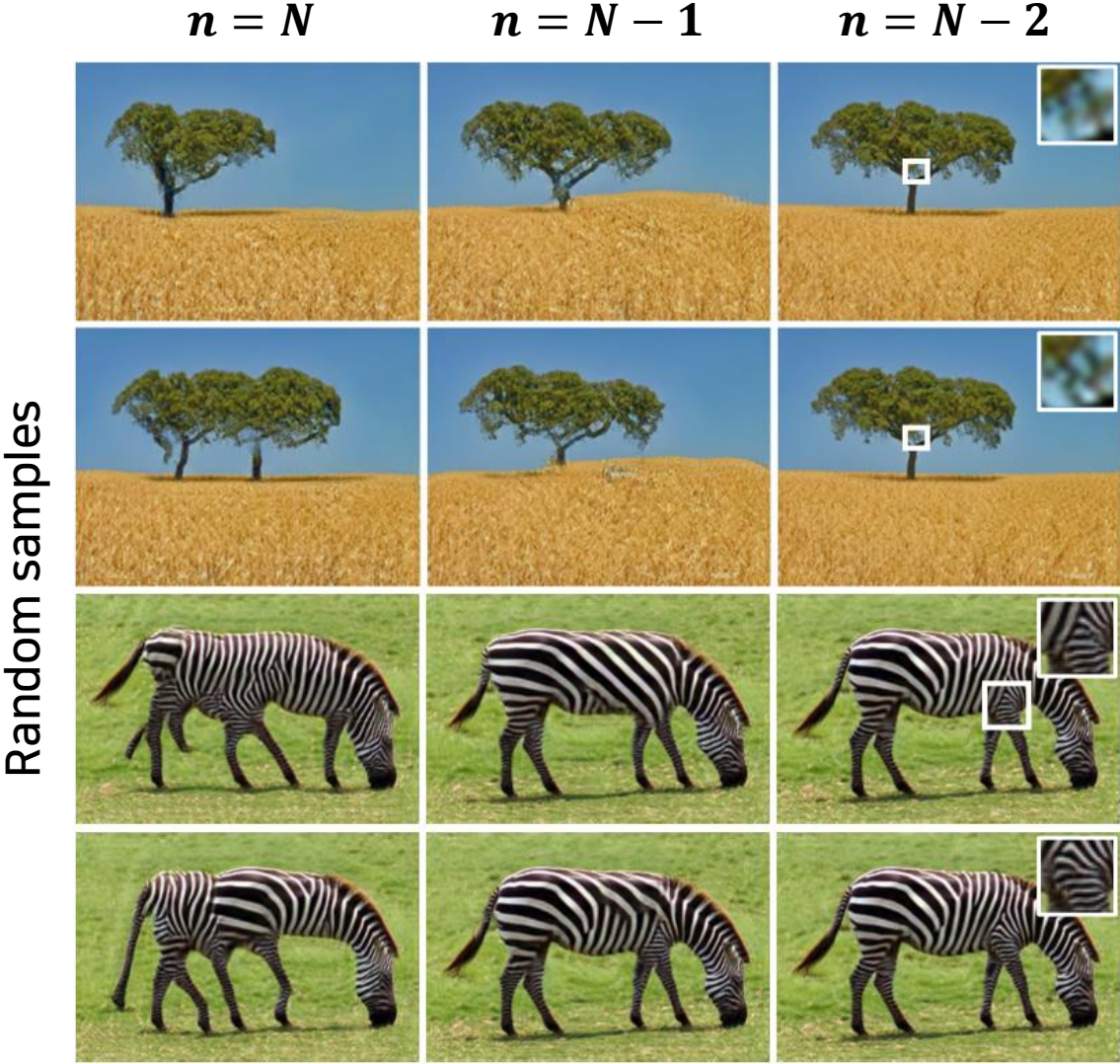
$$\mathcal{L}_{\text{rec}} = \|G_n(0, (\tilde{x}_{n+1}^{\text{rec}}) \uparrow^r) - x_n\|^2$$

## SinGAN의 기본적인 사용 방법

- 학습된 SinGAN의 생성자를 사용할 때는 원하는 scale부터 시작하도록 설정할 수 있습니다.
- $G_N$ 부터 이용한다면, 다양한(diverse) 이미지를 만들 수 있으나 원본 이미지와 매우 다른 형태의 이미지가 생성될 수 있습니다.
- $\tilde{x}_N$ 에 다운샘플링된 실제 이미지를 넣고,  $G_{N-1}$ 부터 이용한다면, 원본 이미지의 가장 coarse한 내용은 유지한 상태로 세밀한 정보를 다양하게 변경해 볼 수 있습니다. (많이 사용되는 방법)



# SinGAN의 성능 평가 및 결과 이미지 분석



- 첫 번째 스케일( $n$ )을  $N$ 으로 설정하면 다양성이 높아 지지만 그럴싸하지 않은 이미지가 나올 수 있습니다.

1st Scale	Diversity	Survey	Confusion
$N$	0.5	paired unpaired	$21.45\% \pm 1.5\%$ $42.9\% \pm 0.9\%$
$N - 1$	0.35	paired unpaired	$30.45\% \pm 1.5\%$ $47.04\% \pm 0.8\%$

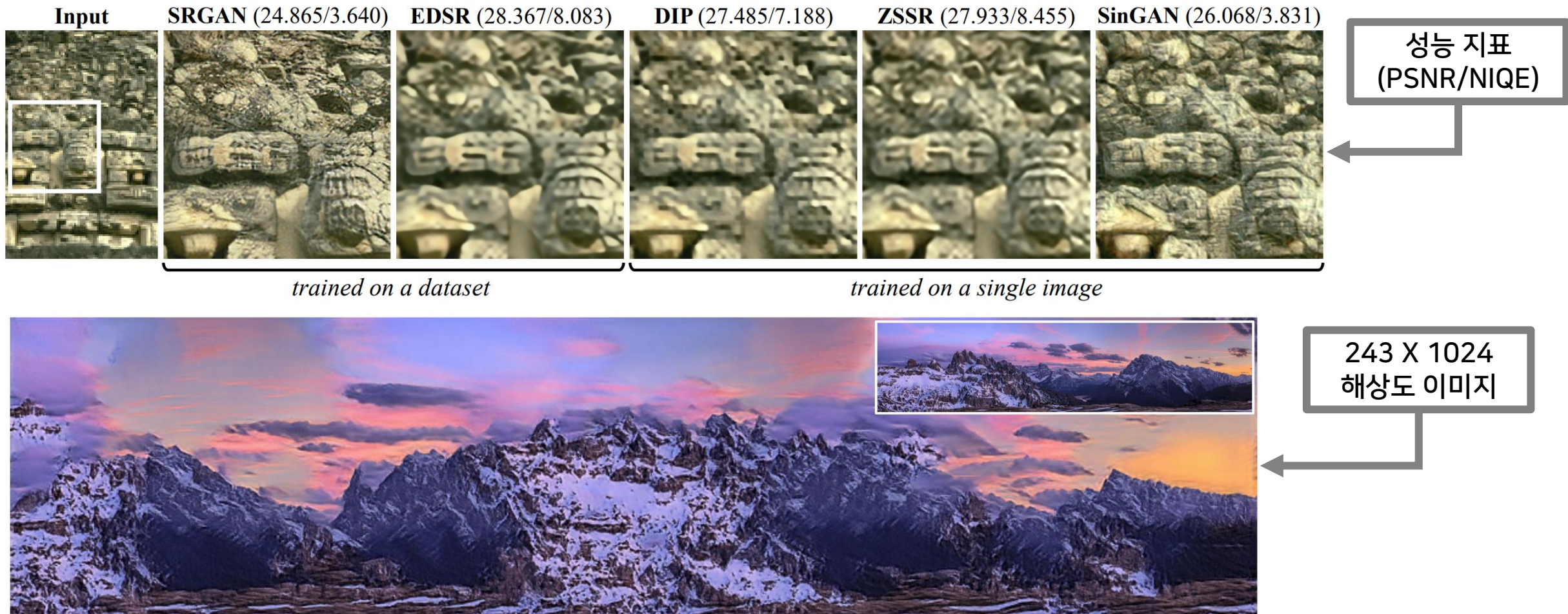
- 본 논문은 SIFID를 제안하였으며, AMT 결과와의 비교를 통해 SIFID가 효과적인 평가 지표라는 것을 보입니다.

1st Scale	SIFID	Survey	SIFID/AMT Correlation
$N$	0.09	paired unpaired	$-0.55$ $-0.22$
$N - 1$	0.05	paired unpaired	$-0.56$ $-0.34$



# SinGAN을 활용한 다양한 애플리케이션: 초해상도(Super-Resolution)

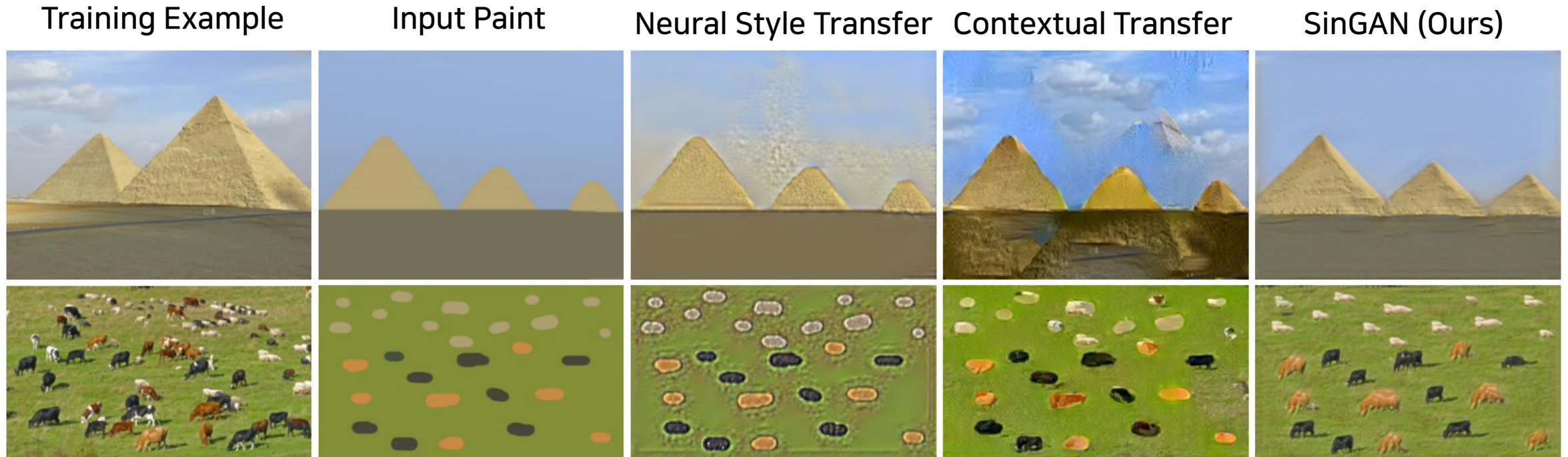
- SRGAN과 같이 외부 데이터셋으로 학습하는 방법(external methods)과 비교해도 좋은 성능을 보입니다.





# SinGAN을 활용한 다양한 애플리케이션: 그림-이미지 변환(Paint-to-Image)

- Coarse level 중 ( $N - 1$  혹은  $N - 2$ ) 하나의 입력으로 클립아트(clipart)를 넣어 그림을 이미지로 변환합니다.
  - 클립아트의 global structure는 유지한 상태로 세밀한 텍스처(texture)가 추가되는 것을 알 수 있습니다.



# SinGAN을 활용한 다양한 애플리케이션: 두 이미지 간의 조화(Harmonization)

- 배경(background) 이미지에 적절한 물체(object)를 적절히 섞을(blending) 수 있습니다.

Input



Deep Paint. Harmonization



SinGAN (Ours)



섞인 이미지를 다운샘플링한 뒤에 중간 스케일(scale)에 넣어 이미지를 생성합니다.



# SinGAN을 활용한 다양한 애플리케이션: 이미지 편집(Image Editing)

(a) Training Example



(b) Edited Input



(c) Content Aware Move



(d) SinGAN (Ours)



- 이미지 편집 과정은 다음과 같습니다.
  - (a) 학습 이미지를 원하는 대로 수정합니다.
  - (b) 다운샘플링 이후에 중간 스케일의 입력으로 이미지를 삽입합니다.
- 포토샵(Photoshop)의 Content Aware Move과 비교하여 우수한 성능을 보입니다.

본 논문의 메서드



## SinGAN을 활용한 다양한 애플리케이션: 이미지 애니메이션(Image Animation)



- 한 장의 이미지로부터 현실적인 모션(motion)을 생성할 수 있습니다.
- Z-space에서 랜덤 워크(random walk)를 수행합니다.

랜덤 워크를 수행하며 애니메이션 생성

