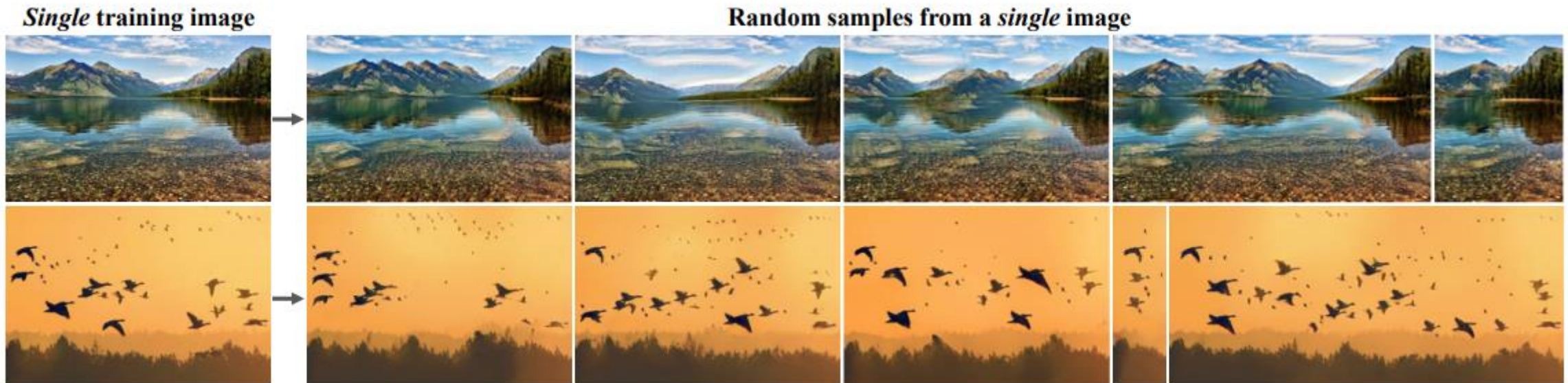


SinGAN : Learning a Generative Model from a Single Natural Image

Tamar Rott Shaham
Tali Dekel
Tomer Michaeli

Presentation by 김진성



목차

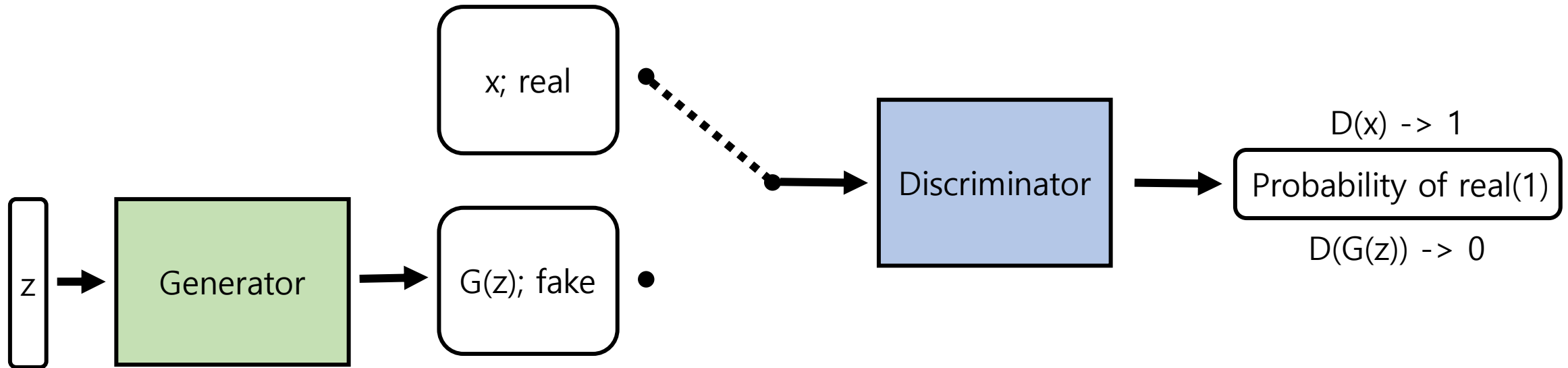
- Abstract
- GAN
- Related work
- SinGAN architecture
- Training
- Result
- Applications

Abstract

- SinGAN : a unconditional generative model that can be learned from Single natural image
- SinGAN 특징
 - SinGAN은 이미지의 patch 단위로 분포를 학습하고, 그로 인해 높은 퀄리티의 이미지를 다양하게 만들어 낼 수 있다.
 - SinGAN은 fully convolutional GAN들이 쌓인 피라미드 형태의 구조를 가지고, 각 층은 다른 스케일의 이미지를 patch 단위로 학습한다. 이로 인해, 다양한 크기의 이미지를 만들어 낼 수 있고, 전체적인 모습부터, 세밀한(fine)부분까지 만들어 낼 수 있다.
- 이전의 single image deep model들은 제한적인 task를 수행했거나, conditional 한 모델이었으나, SinGAN은 여러 task를 수행할 수 있고, noise로 부터 image를 생성할 수 있는 unconditional한 모델이다.
- SinGAN은 하나의 모델로, paint to image, editing, harmonization, super-resolution, animation 등의 task를 실제 이미지와 같은 이미지를 만들어 낼 수 있다.

GAN; Generative Adversarial Networks

- GAN is a framework for estimating generative model via an adversarial process of Generator and Discriminator
- Discriminator estimates the probability that a sample came from the training data rather than G.
- Generator trained to generate fake realistic image samples.



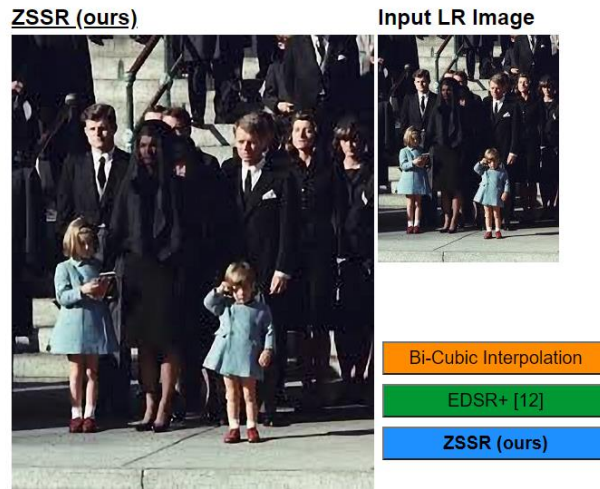
- Objective function $V(D, G)$; minmax problem

$$\min_G \max_D V(D, G) = \mathbb{E}_{\mathbf{x} \sim p_{\text{data}}(\mathbf{x})} [\log D(\mathbf{x})] + \mathbb{E}_{\mathbf{z} \sim p_{\mathbf{z}}(\mathbf{z})} [\log(1 - D(G(\mathbf{z})))].$$

Related Work

Previous Single image deep models

- Super resolution : "Zero-Shot" Super-Resolution using Deep Internal Learning [Shocher et al.]
- Texture expansion : Non-Stationary Texture Synthesis by Adversarial Expansion [Yang et al.]
- InGAN: Capturing and Remapping the "DNA" of a Natural Image (Shocher et al.' 2018)
 - single natural image를 internal GAN Based model로 학습하는 것을 처음으로 제시.
 - Image retargeting을 위한 모델.
 - Conditonal model



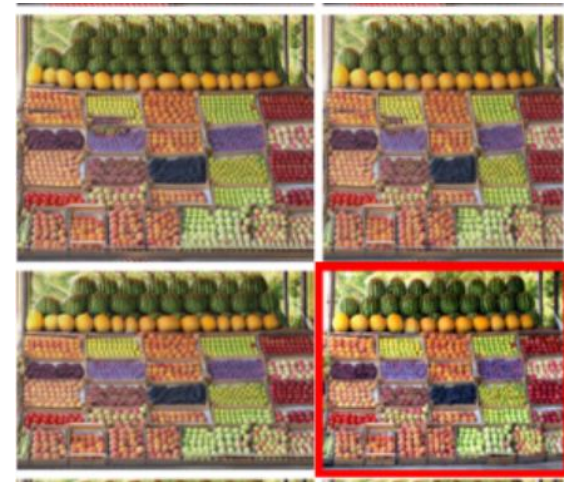
Zero-shot SR

출처 : <http://www.wisdom.weizmann.ac.il/~vision/zssr/>



Texture expansion

출처 : <https://arxiv.org/pdf/1805.04487.pdf>

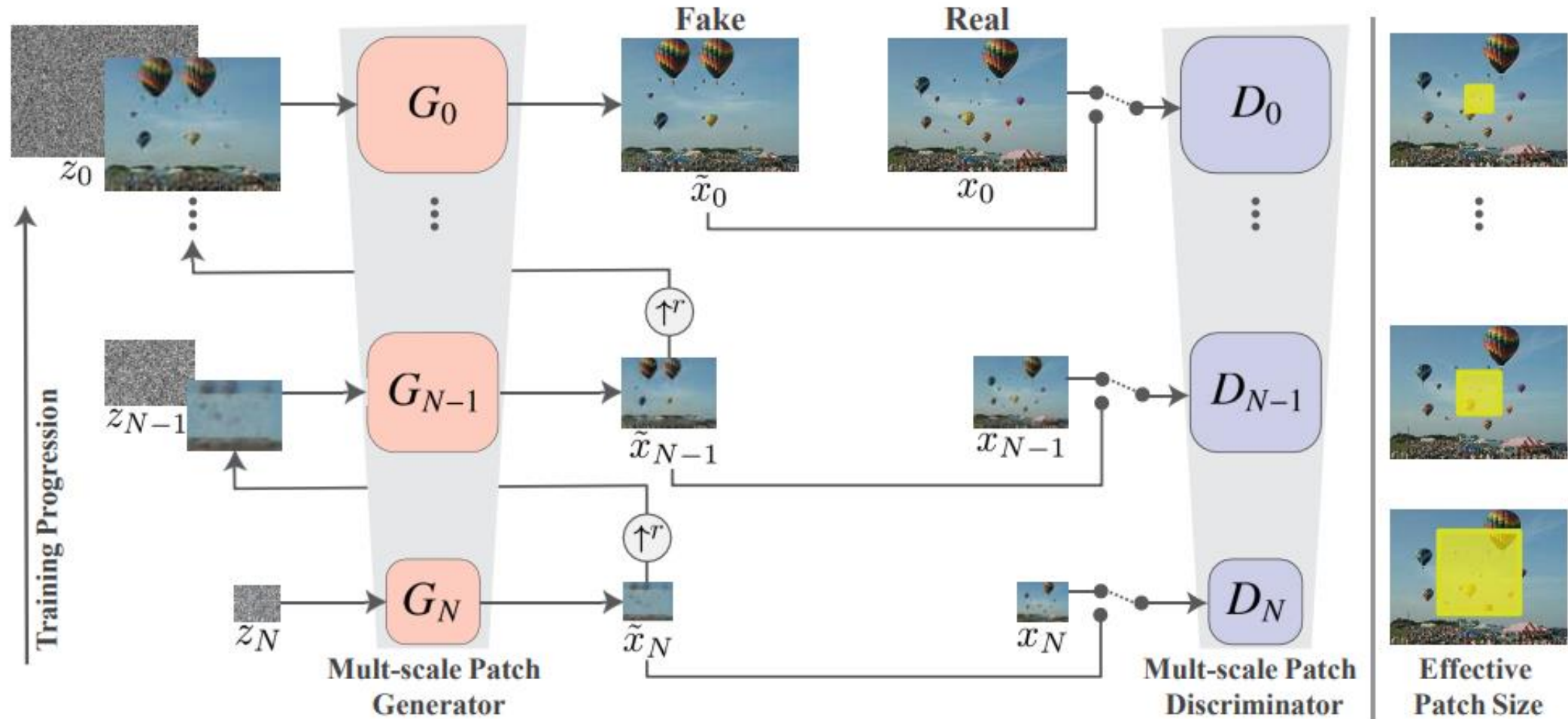


InGAN

출처 : InGAN paper (<https://arxiv.org/abs/1812.00231>)

SinGAN architecture:

SinGAN architecture:

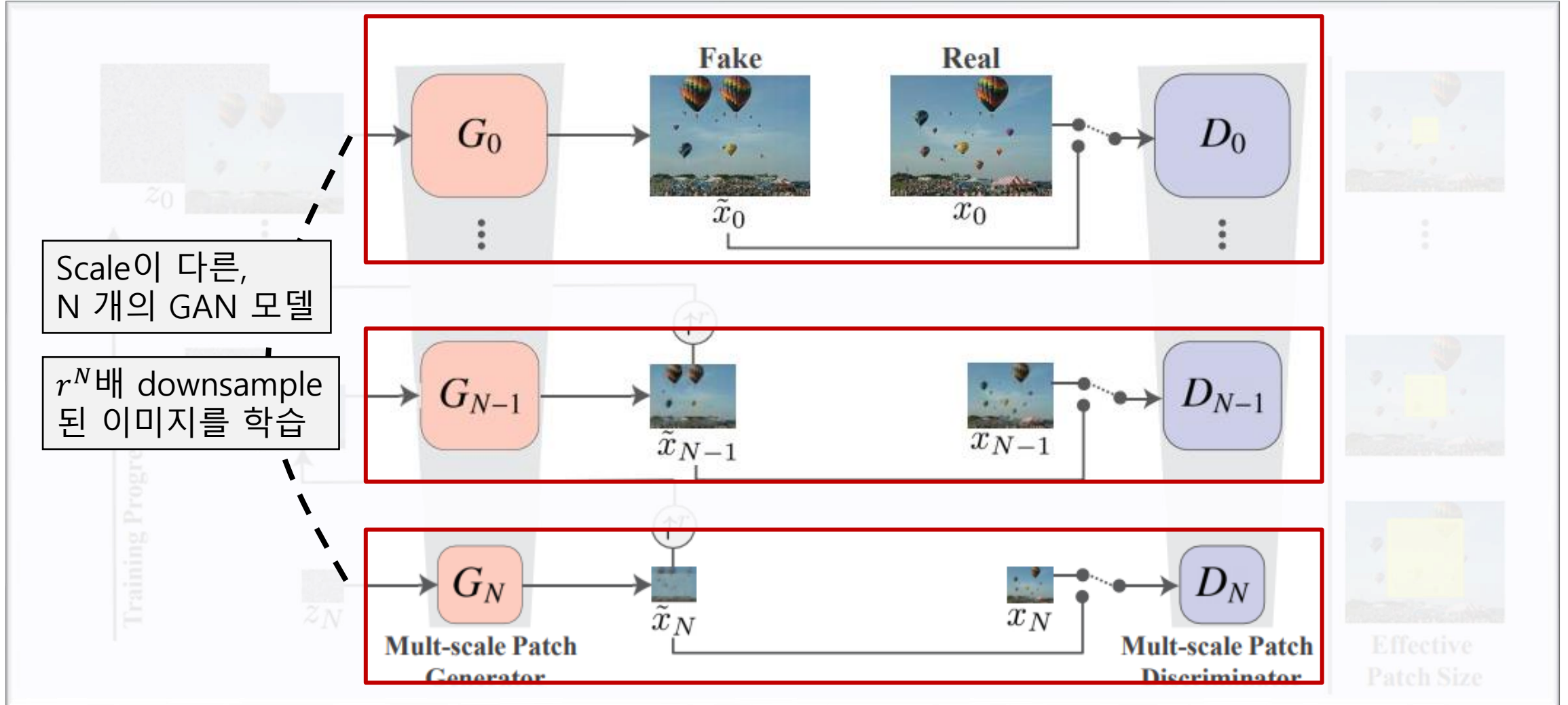


SinGAN architecture

출처 : SinGAN paper(<https://arxiv.org/abs/1905.01164>)

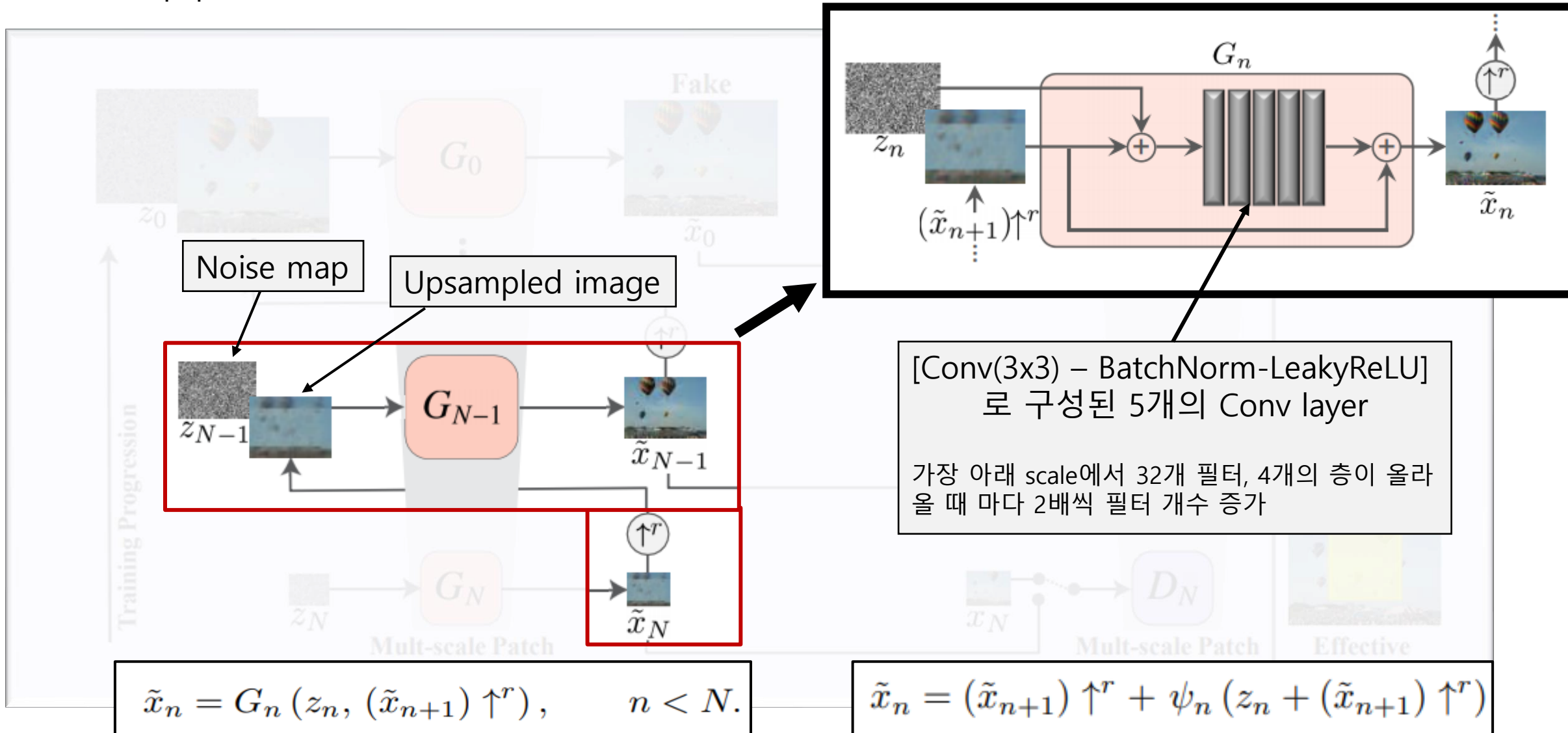
SinGAN architecture: pyramidic Multi-scale GANs

Muti-scale GAN 구조를 사용



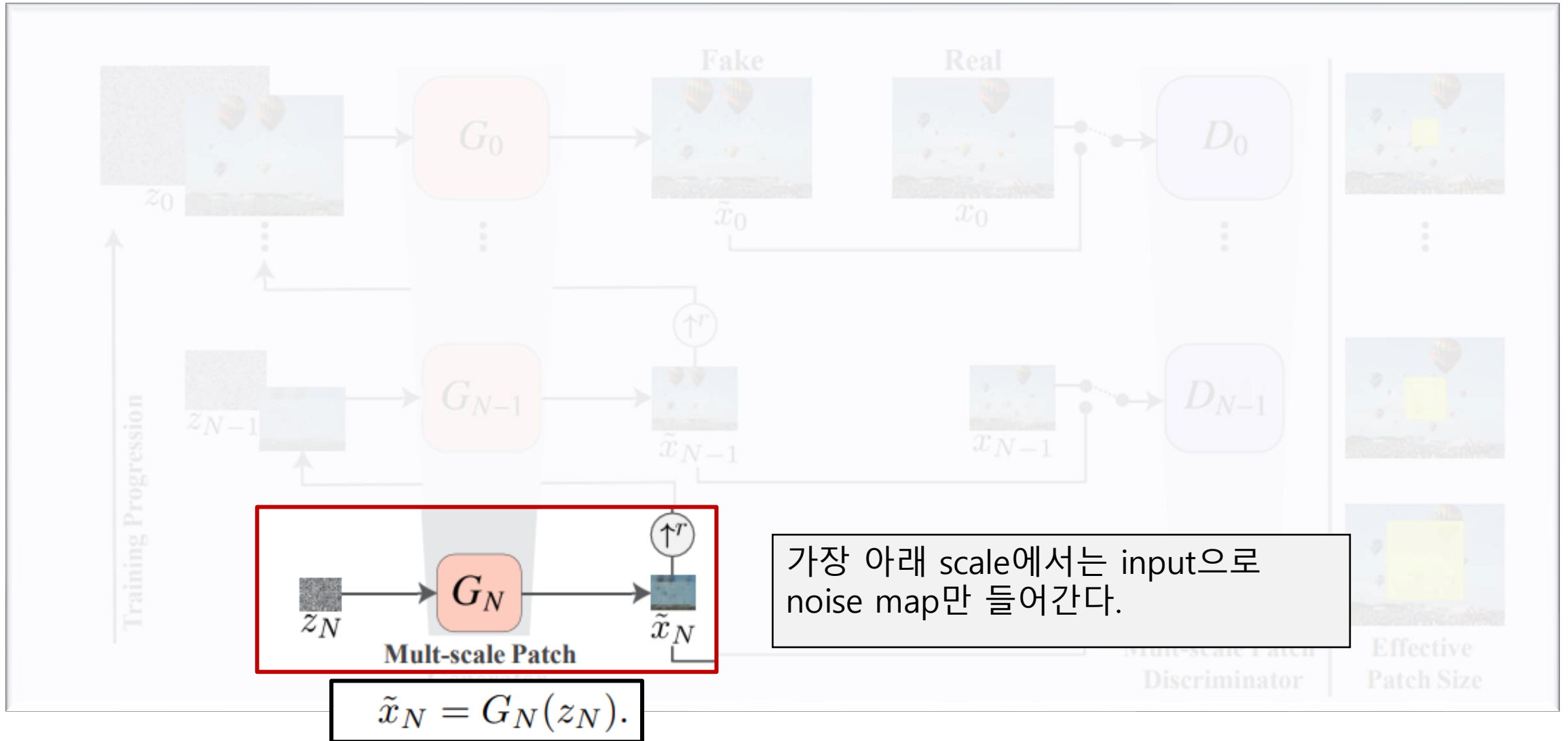
SinGAN architecture: Generator

Generator 의 구조



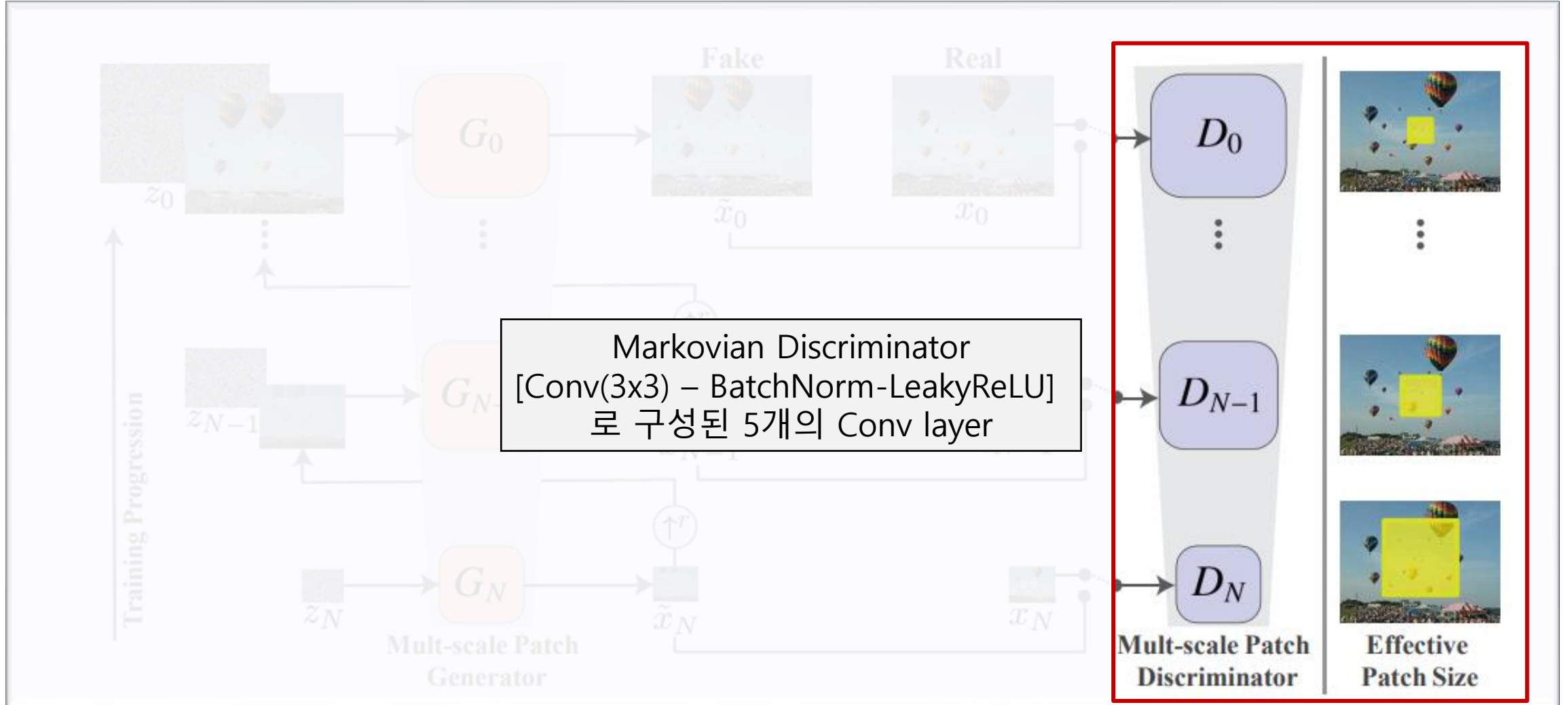
SinGAN architecture: Generator

Coarsetset scale 에서의 Generator



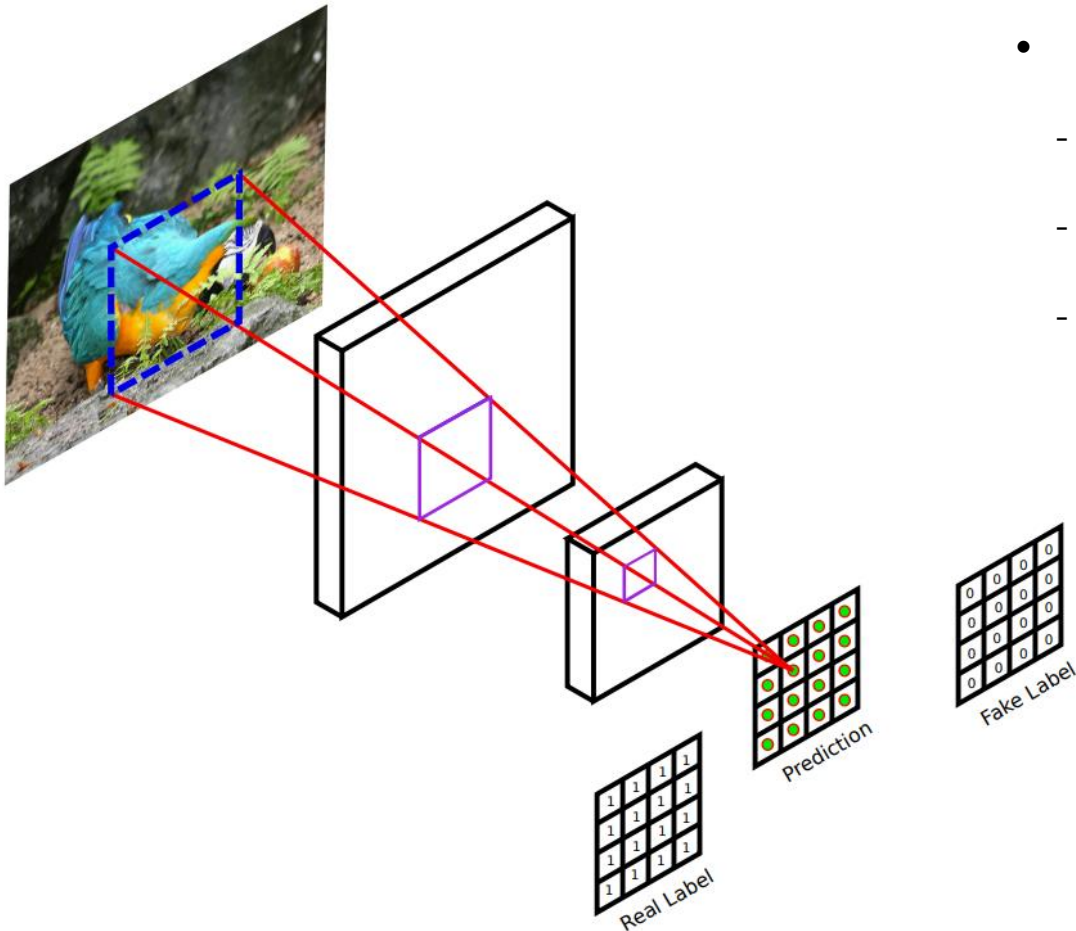
SinGAN architecture: Discriminator(patch GAN)

Discriminator 의 구조



SinGAN architecture; Markovian discriminator

- 특정 크기의 patch 단위로 이미지의 진위 여부를 판단하는 Discriminator를 가지는 GAN



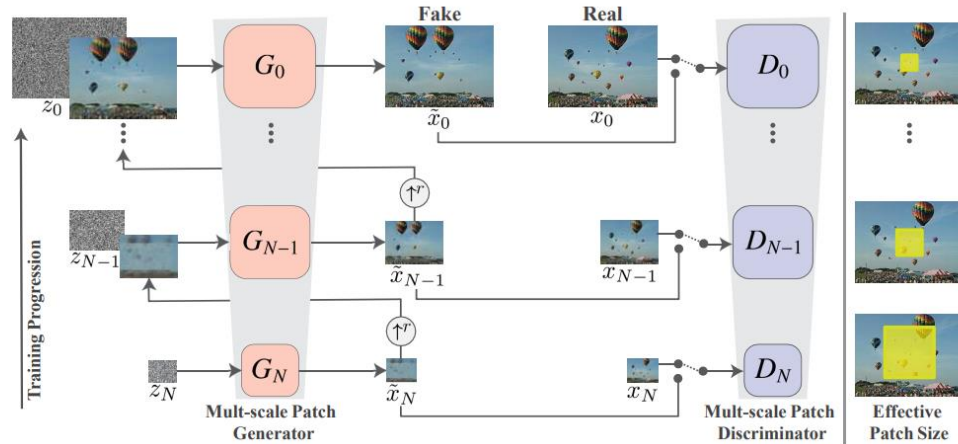
- Patch GAN을 사용하는 이유

- Patch 단위로 학습이 진행 되므로, 각 effective field의 특징이 학습됨.
- 위와 같은 이유로, PatchGAN은 High frequency 성분의 학습이 이루어짐.
- 파라미터의 개수가 작아지고, 이미지의 크기에 영향을 받지 않아, 유연해짐.

SinGAN architecture: Summary

1. Architecture

- Multi-scale GAN 구조를 사용하여 coarset 부터 fine까지 이미지를 잘 생성 할 수 있게 하였다.



2. Generator

- 다음의 수식을 따른다.

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

- Fully convolutional net 으로 구성되어 있어, input 크기에 유연하다.
- Fine scale 을 학습할 때,

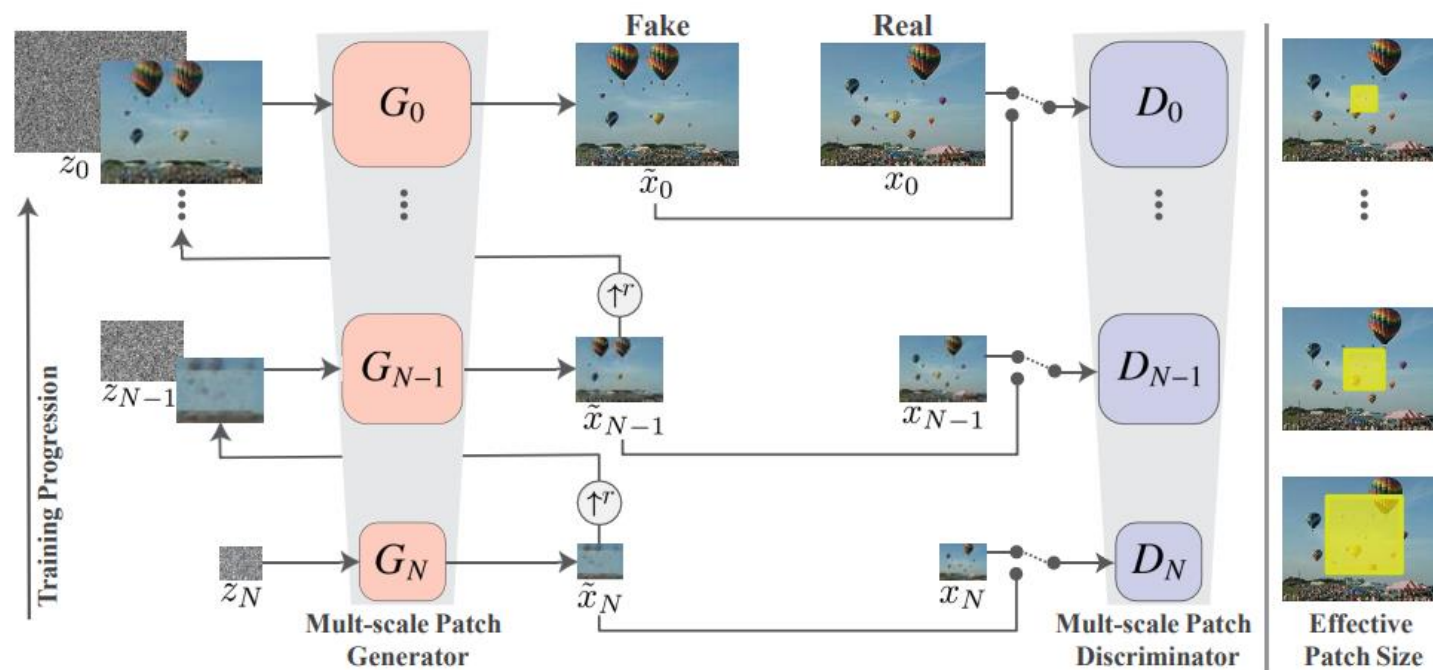
3. Discriminator

- Markovian discriminator를 사용하여, 각 scale 마다 다른 effective size를 학습하게 한다.

Training

Training; Sequence

- Multi-scale architecture를 가지므로, Coarset scale 부터 finest 까지 순차적으로 학습을 진행한다.
- 학습이 완료된 GAN은 fixed 된다.
- 다음 G,D를 생성할 때, 이전의 학습이 완료된 G,D와 convolution filter 개수가 같다면, 학습된 G,D값을 initial parameter로 사용한다.
 - 학습 수렴속도를 증가시킬 수 있다.



Training; Loss

- SinGAN은 Adversarial loss와 Reconstruction loss로 구성된다.

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

- Adversarial loss
 - Adversarial loss로 Wasserstein GAN-GP Loss를 사용한다.
 - 학습에 사용되는 z_n 는 standard deviation(σ_n)이 $(\tilde{x}_{n+1}^{\text{rec}})^r$ 과 x_n 의 RMSE(root mean square error) 값인 noise를 사용한다.
- Reconstruction loss
 - Reconstruction loss는 G가 생성한 이미지와, 그 단계의 real image(x_n)의 pixel 간의 차이를 줄이기 위해 사용한다.

$$\mathcal{L}_{\text{rec}} = ||\tilde{x}_n^{\text{rec}} - x_n||^2$$

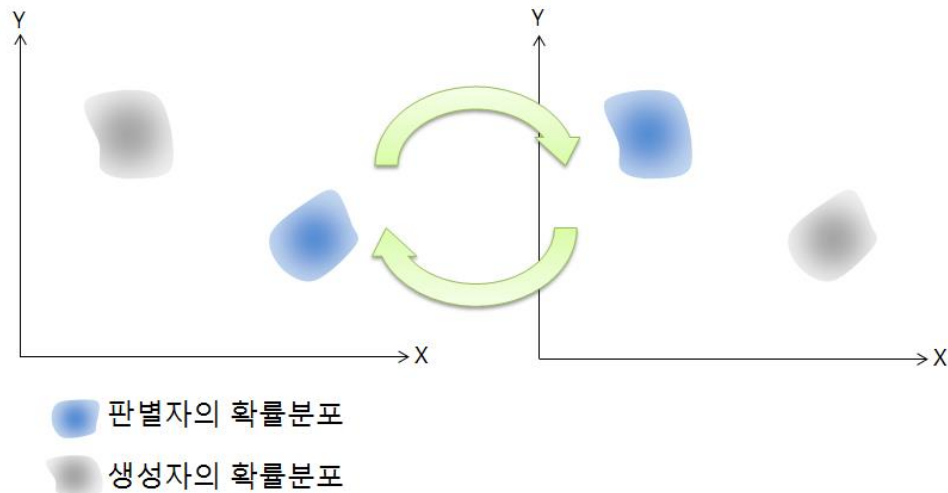
- 여기서 \tilde{x}_n^{rec} 는 fixed된 noise map, $\{z_N^{\text{rec}}, z_{N-1}^{\text{rec}}, \dots, z_0^{\text{rec}}\} = \{z^*, 0, \dots, 0\}$ 을 통해 만들어진 이미지를 뜻한다.

$$\tilde{x}_n^{\text{rec}} = G_n(0, (\tilde{x}_{n+1}^{\text{rec}} \uparrow^r))$$

Training : WGAN and WGAN-gp

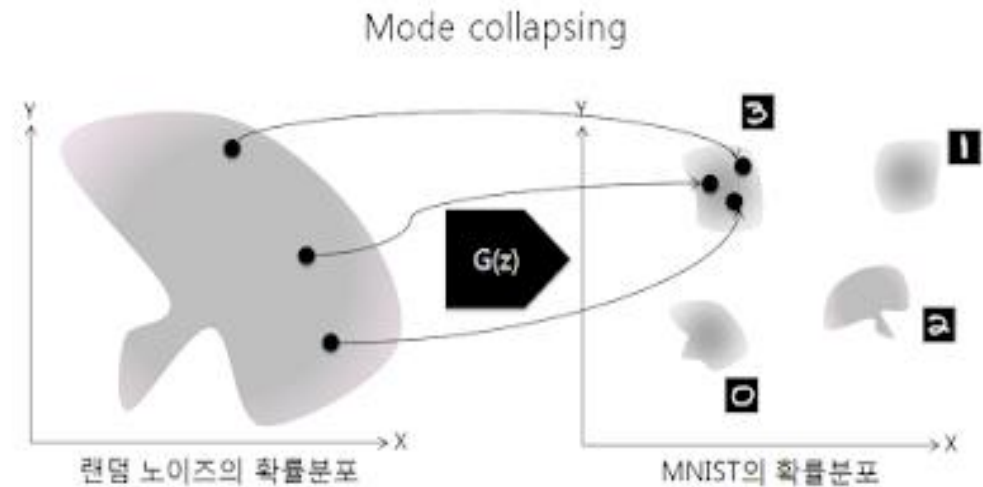
기존 GAN의 문제

- Oscillating : Generator와 Discriminator 간의 학습이 서로를 상쇄시켜 학습이 진행 되지 않는 현상
- Mode collapsing or dropping : Generator가 일부 분포만을 따라하도록 학습되는 현상



GAN Oscillation 현상

출처 : <http://dl-ai.blogspot.com/2017/08/gan-problems.html>



GAN mode collapsing 현상

출처 : <http://dl-ai.blogspot.com/2017/08/gan-problems.html>

Training : WGAN and WGAN-gp

WGAN

- Discriminator 대신 Critic을 사용.
- Distance metric으로 Earth-Mover(EM) Distance를 사용.

$$W(\mathbb{P}_r, \mathbb{P}_g) = \inf_{\gamma \in \Pi(\mathbb{P}_r, \mathbb{P}_g)} \mathbb{E}_{(x,y) \sim \gamma} [\|x - y\|]$$

EM Distance or Wasserstein-1

- 단, EM Distance를 loss로 사용하기 위해서, 미분이 가능해야 하는데, 연속성 제약 조건(Lipschitz continuity)을 만족하기 위해, clipping이 필요.

Training : WGAN and WGAN-gp

WGAN

- 연속성 제약 조건을 만족하는 f 에 대해, WGAN의 목적함수는 다음과 같다.

$$\max_{w \in \mathcal{W}} \mathbb{E}_{x \sim \mathbb{P}_r} [f_w(x)] - \mathbb{E}_{z \sim p(z)} [f_w(g_\theta(z))]$$

Algorithm 1 WGAN, our proposed algorithm. All experiments in the paper used the default values $\alpha = 0.00005$, $c = 0.01$, $m = 64$, $n_{\text{critic}} = 5$.

Require: : α , the learning rate. c , the clipping parameter. m , the batch size. n_{critic} , the number of iterations of the critic per generator iteration.

Require: : w_0 , initial critic parameters. θ_0 , initial generator's parameters.

```
1: while  $\theta$  has not converged do
2:   for  $t = 0, \dots, n_{\text{critic}}$  do
3:     Sample  $\{x^{(i)}\}_{i=1}^m \sim \mathbb{P}_r$  a batch from the real data.
4:     Sample  $\{z^{(i)}\}_{i=1}^m \sim p(z)$  a batch of prior samples.
5:      $g_w \leftarrow \nabla_w [\frac{1}{m} \sum_{i=1}^m f_w(x^{(i)}) - \frac{1}{m} \sum_{i=1}^m f_w(g_\theta(z^{(i)}))]$ 
6:      $w \leftarrow w + \alpha \cdot \text{RMSPProp}(w, g_w)$ 
7:      $w \leftarrow \text{clip}(w, -c, c)$ 
8:   end for
9:   Sample  $\{z^{(i)}\}_{i=1}^m \sim p(z)$  a batch of prior samples.
10:   $g_\theta \leftarrow -\nabla_\theta \frac{1}{m} \sum_{i=1}^m f_w(g_\theta(z^{(i)}))$ 
11:   $\theta \leftarrow \theta - \alpha \cdot \text{RMSPProp}(\theta, g_\theta)$ 
12: end while
```

Training : WGAN and WGAN-gp

WGAN-gradient penalty

- WGAN에서 clipping parameter c 가 크면 학습이 오래 걸리고, c 가 작으면 gradient vanishing 현상이 발생.
- 이러한 clipping 문제를 해결하기 위해, clipping 대신 gradient penalty Loss를 사용

Algorithm 1 WGAN with gradient penalty. We use default values of $\lambda = 10$, $n_{\text{critic}} = 5$, $\alpha = 0.0001$, $\beta_1 = 0$, $\beta_2 = 0.9$.

Require: The gradient penalty coefficient λ , the number of critic iterations per generator iteration n_{critic} , the batch size m , Adam hyperparameters α, β_1, β_2 .

Require: initial critic parameters w_0 , initial generator parameters θ_0 .

```
1: while  $\theta$  has not converged do
2:   for  $t = 1, \dots, n_{\text{critic}}$  do
3:     for  $i = 1, \dots, m$  do
4:       Sample real data  $\mathbf{x} \sim \mathbb{P}_r$ , latent variable  $\mathbf{z} \sim p(\mathbf{z})$ , a random number  $\epsilon \sim U[0, 1]$ .
5:        $\tilde{\mathbf{x}} \leftarrow G_{\theta}(\mathbf{z})$ 
6:        $\hat{\mathbf{x}} \leftarrow \epsilon \mathbf{x} + (1 - \epsilon) \tilde{\mathbf{x}}$ 
7:        $L^{(i)} \leftarrow D_w(\tilde{\mathbf{x}}) - D_w(\mathbf{x}) + \lambda(\|\nabla_{\hat{\mathbf{x}}} D_w(\hat{\mathbf{x}})\|_2 - 1)^2$ 
8:     end for
9:      $w \leftarrow \text{Adam}(\nabla_w \frac{1}{m} \sum_{i=1}^m L^{(i)}, w, \alpha, \beta_1, \beta_2)$ 
10:   end for
11:   Sample a batch of latent variables  $\{\mathbf{z}^{(i)}\}_{i=1}^m \sim p(\mathbf{z})$ .
12:    $\theta \leftarrow \text{Adam}(\nabla_{\theta} \frac{1}{m} \sum_{i=1}^m -D_w(G_{\theta}(\mathbf{z}^{(i)})), \theta, \alpha, \beta_1, \beta_2)$ 
13: end while
```

WGAN-gp Algorithm

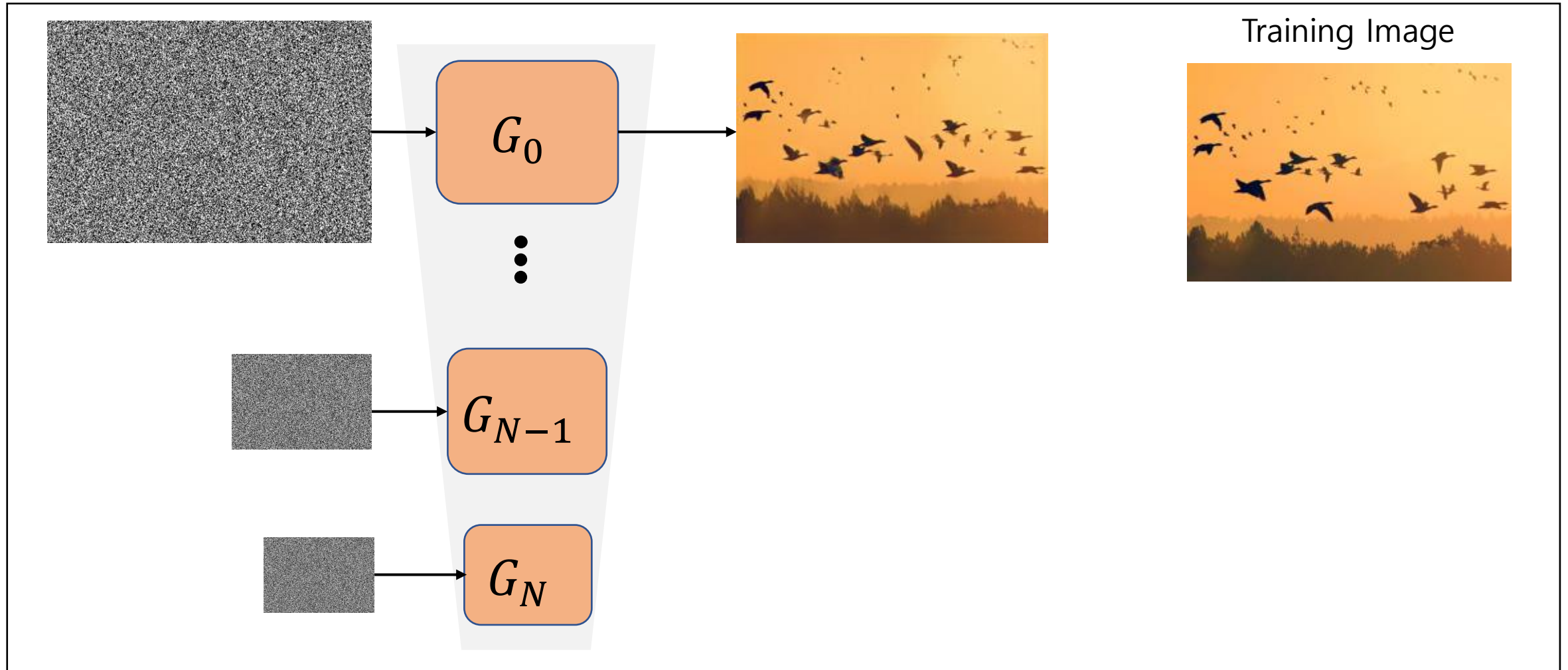
출처 : Improved Training of Wasserstein GANs (<https://arxiv.org/pdf/1704.00028.pdf>)

Result

- Random image samples
- Generation from different scales

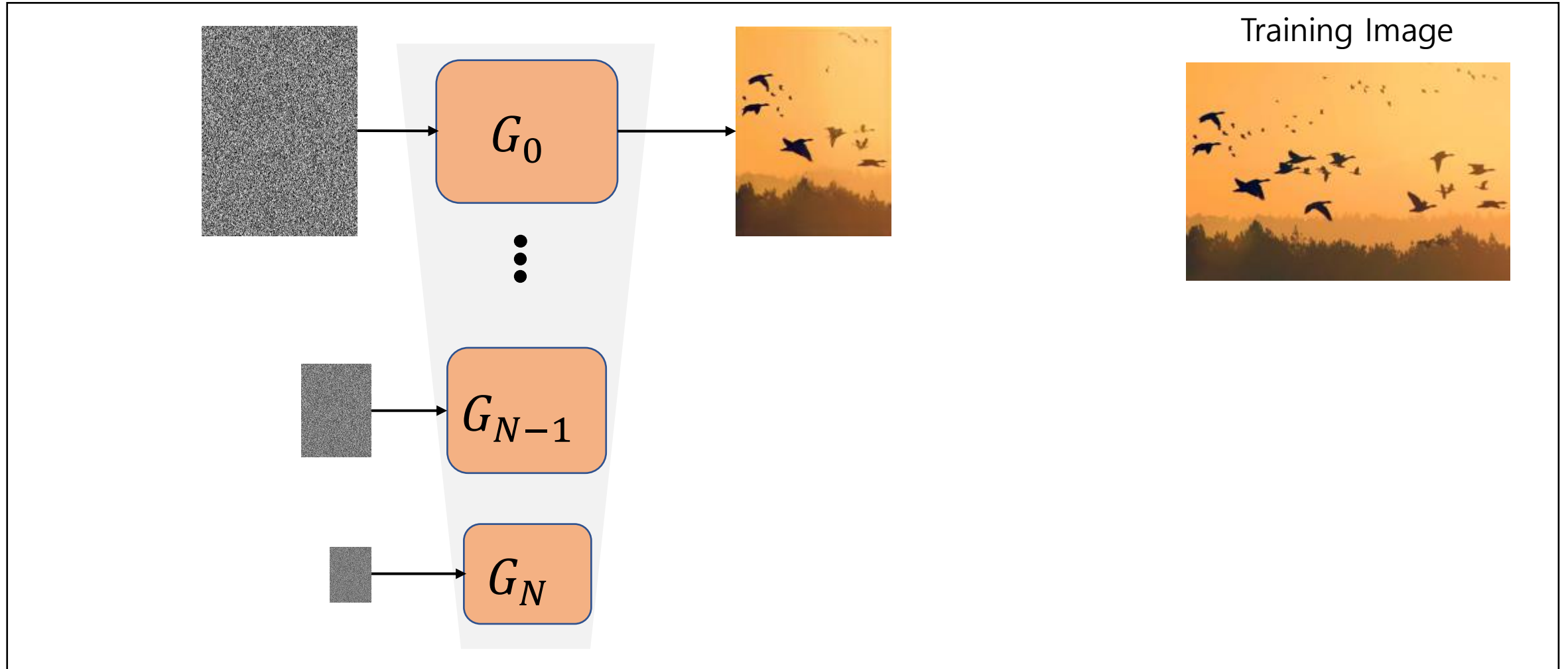
Result : Random image samples

- SinGAN의 가장 큰 특징으로, G가 fully convolutional nets 이므로, input size을 조절 할 수 있다.

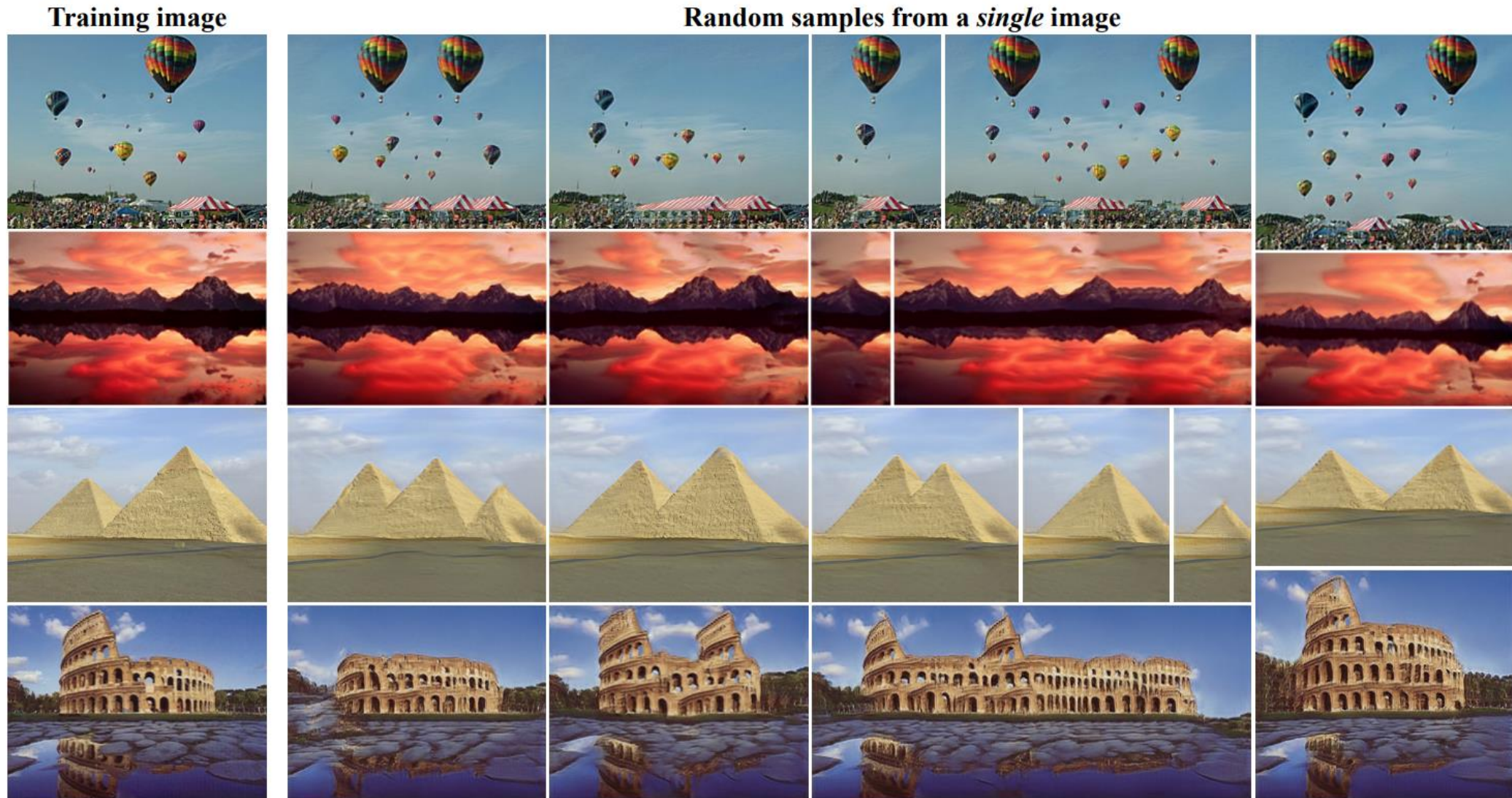


Result : Random image samples

- SinGAN의 가장 큰 특징으로, G가 fully convolutional nets 이므로, input size을 조절 할 수 있다.



Result : Random image samples



Random image samples

출처 : SinGAN paper(<https://arxiv.org/abs/1905.01164>)

Result : Random image samples

Training Image

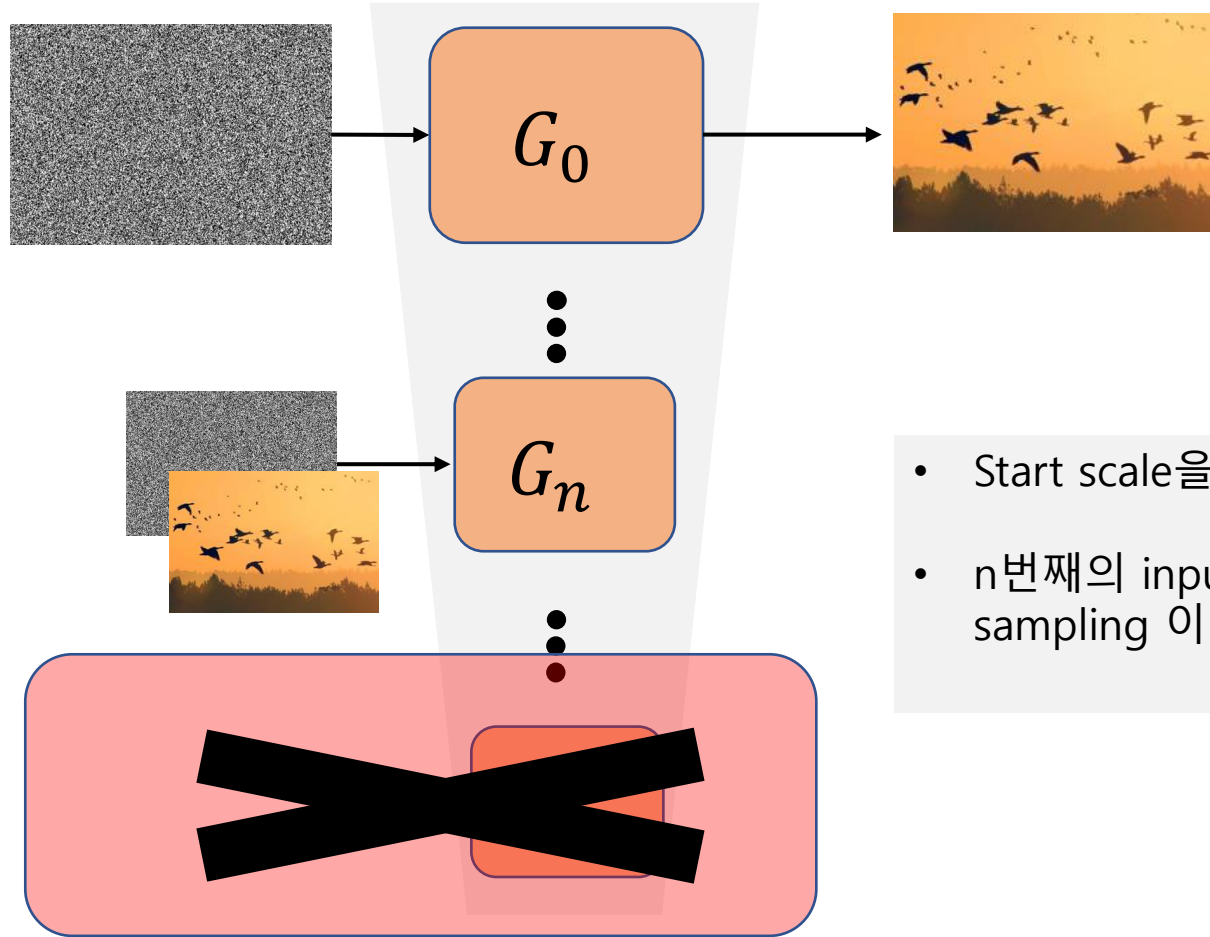


Random samples for single image



Result : Generation from different scales

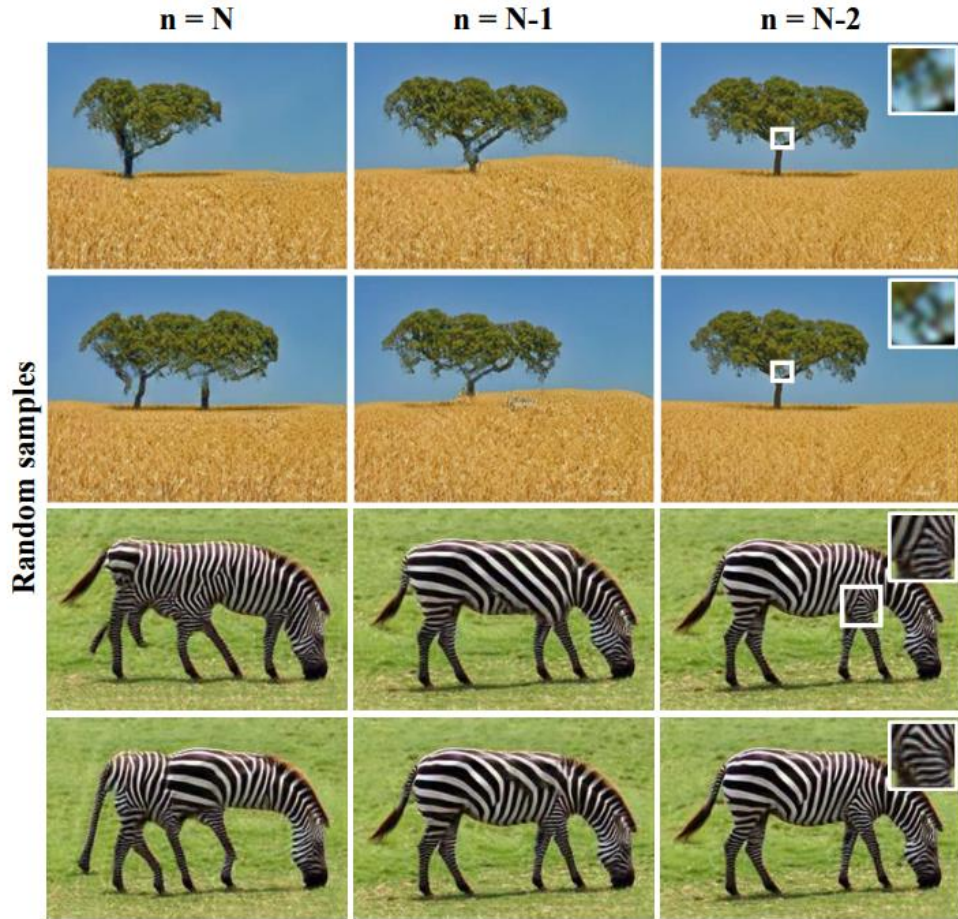
- SinGAN은 random generation을 시작하는 start scale을 선택할 수 있다.



- Start scale을 n 이라고 할 때.
- n 번째의 input으로 random noise와, real data의 down sampling 이미지를 사용하여 이미지를 생성한다.

Result : Generation from different scales

- Start-scale에 따른 이미지 비교와 정량적 평가



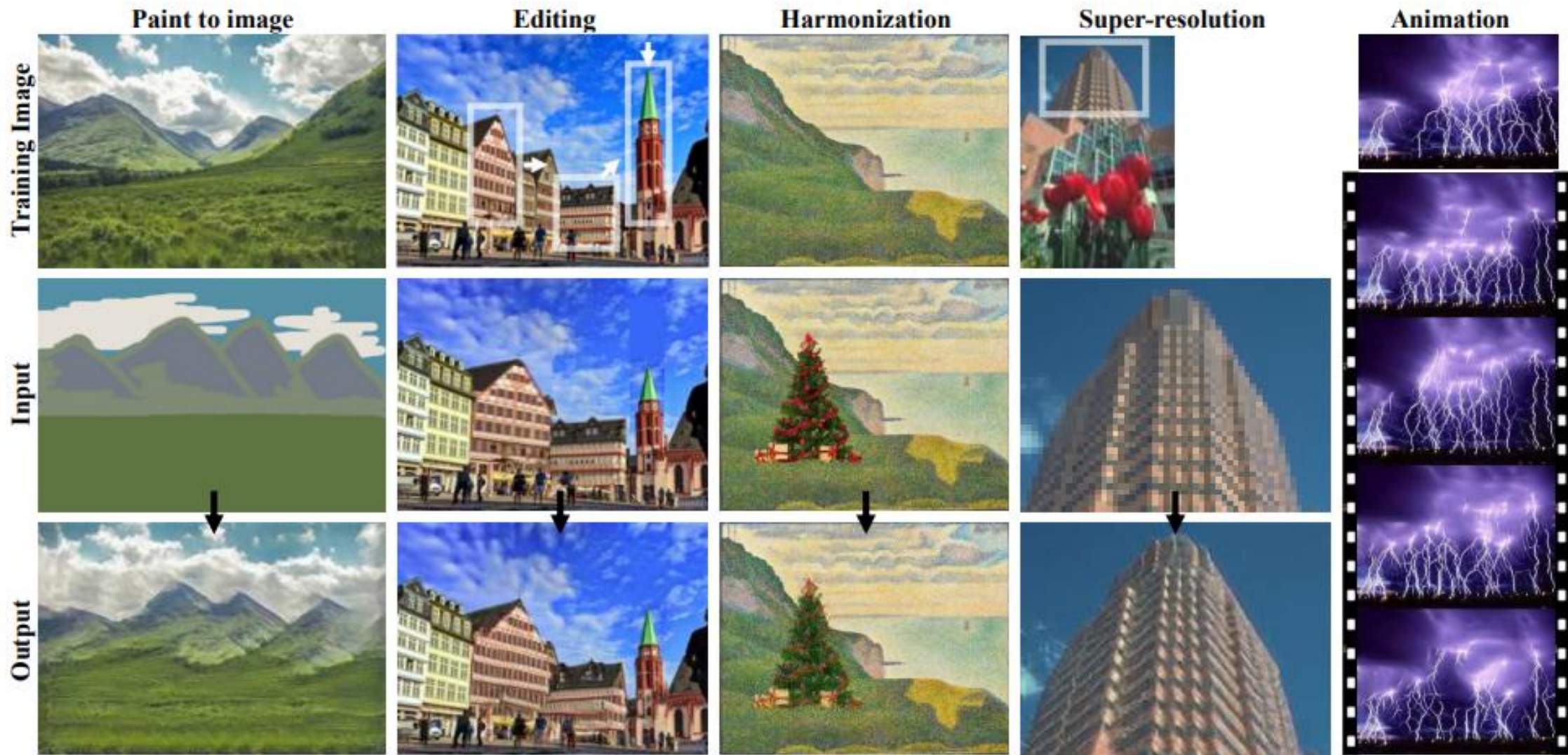
- AMT perceptual study

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	$30.45\% \pm 1.5\%$ $47.04\% \pm 0.8\%$

- Single Image FID(Frechet Inception Distance)

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

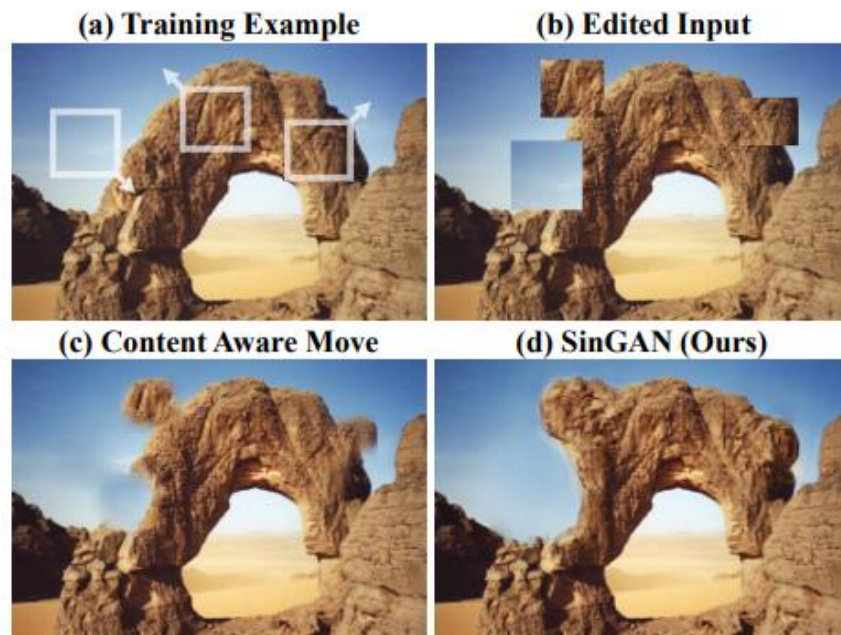
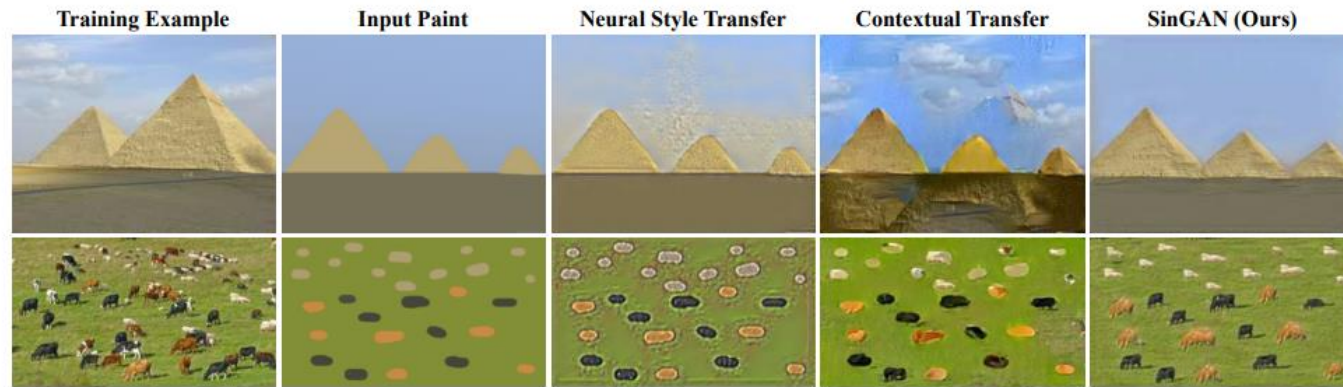
Applications



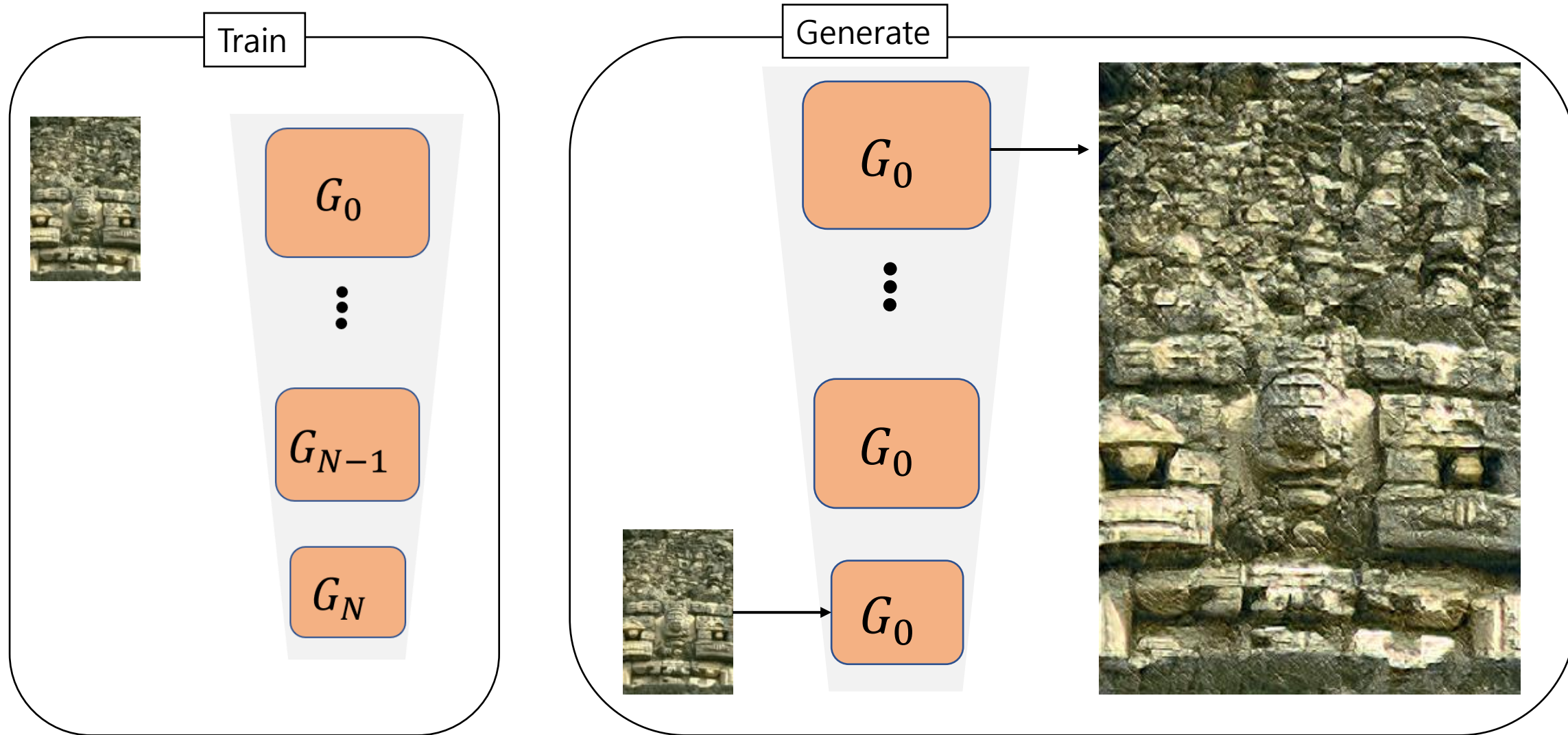
SinGAN applications
출처 : SinGAN paper(<https://arxiv.org/abs/1905.01164>)

Applications : Paint to image, Editing, Harmonization

- Coarset 에 가까운 scale에 input 이미지를 넣어서 생성

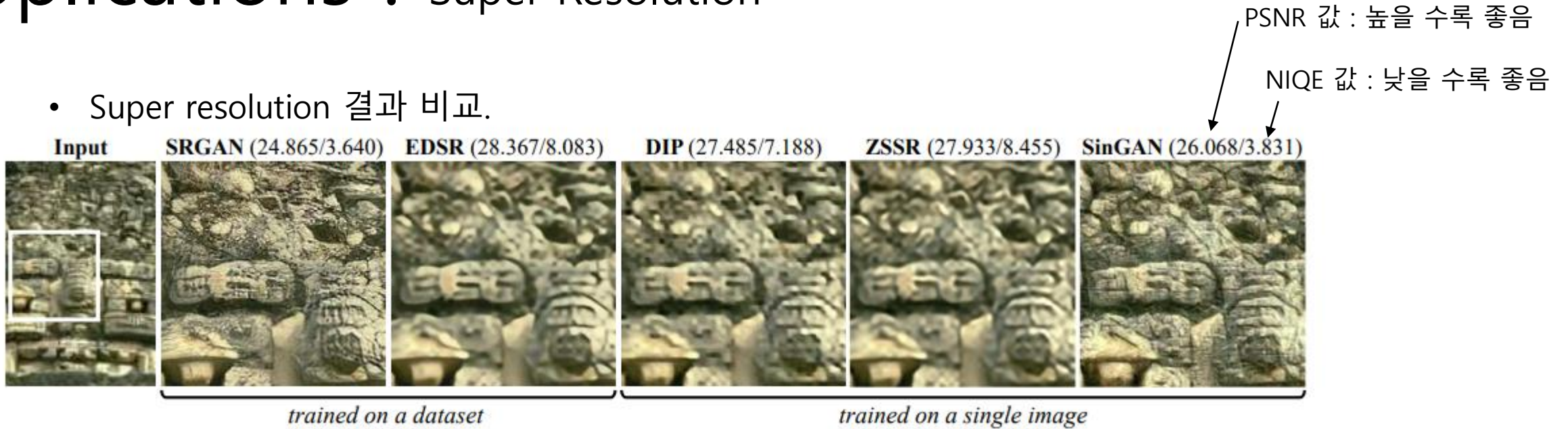


Applications : Super-Resolution



Applications : Super-Resolution

- Super resolution 결과 비교.



- BSD100 dataset의 Super resolution 결과 비교.

	External methods		Internal methods		
	SRGAN	EDSR	DIP	ZSSR	SinGAN
RMSE	16.34	12.29	13.82	13.08	16.22
NIQE	3.41	6.50	6.35	7.13	3.71

Applications : Animation <https://youtu.be/xk8bWLZk4DU>.

- Create a short video clip with realistic object motion, from a single input image

자연스러운 흔들림과 같은 움직임은, 이미지에서 약간만 변형된 이미지들의 모임이다.
따라서, 첫 프레임 이미지에서 출발하여 이미지의 manifold를 travel 하는 것과 같다.

SinGAN을 통해 학습된 이미지의 manifold에 접근할 수 있고 z^{rec} 을 통해 첫 프레임(학습 이미지)로 갈 수 있다.
따라서 z^{rec} 에서 시작하여 z-space를 random walk 하여 sinGAN을 통해 이미지를 생성하면 해당 이미지의 manifold를 random walk하는 것과 비슷한 효과를 낼 수 있다.

Ref

- GAN paper
- Patch-Based Image Inpainting with Generative Adversarial Networks
(<https://arxiv.org/pdf/1803.07422.pdf>)
- WGAN,
(https://medium.com/@jonathan_hui/gan-spectral-normalization-893b6a4e8f53)
- Lipschitz continuity
(https://medium.com/@jonathan_hui/gan-wasserstein-gan-wgan-gp-6a1a2aa1b490)
- SinGAN paper
(<https://arxiv.org/abs/1905.01164>)

Thank you