SinGAN: Learning a Generative Model from a Single Natural Image

1^{st} Paper Study | 2022.07.06 Wed.

한 다 희 Han Dahee

STRADVISION

Outline

1. **Abstract**
2. **Introduction**
3. **Method**
4. **Result**
5. **Evaluation**
6. **Application**
7. **Conclusion**

Abstract

✓ **Single natural image**로 학습하는 **unconditional generative model**

✓ **Image 가 아닌 Image patch** 단위로 internal distribution을 학습

✓ **A pyramid of fully convolutional GAN**으로 구성 (GAN net이 여러 개)

✓ 다양한 application에 활용

Abstract

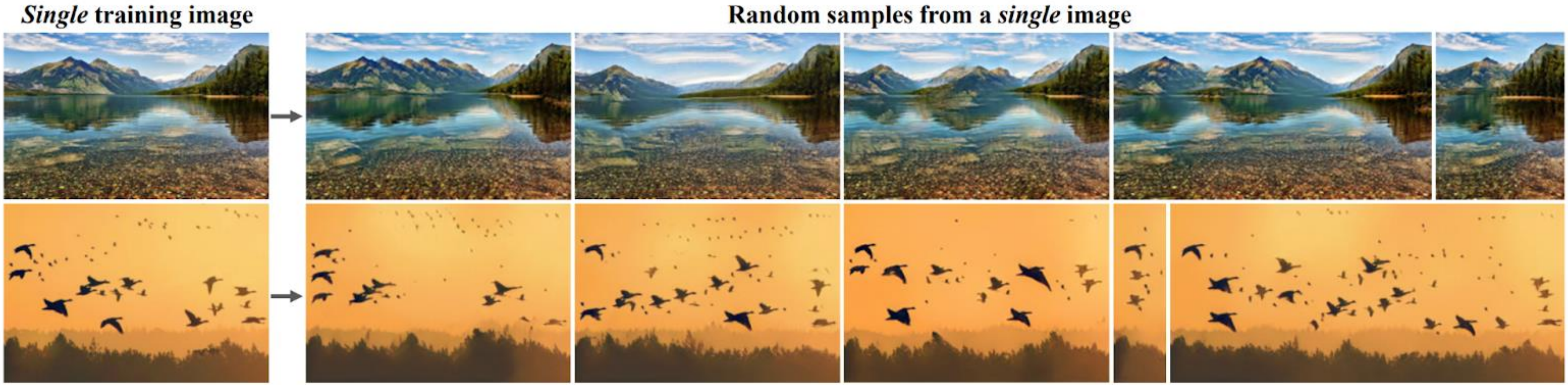


Figure 1: **Image generation learned from a single training image.** We propose *SinGAN*–a new unconditional generative model trained on a *single natural image*. Our model learns the image's patch statistics across multiple scales, using a dedicated multi-scale adversarial training scheme; it can then be used to generate new realistic image samples that preserve the original patch distribution while creating new object configurations and structures.

Introduction

✓ Face, bedrooms과 같은 Class specific dataset으로 학습한 unconditional GAN의 성공

✓ Multiple object class를 대한 unconditional GAN 실험은 major challenge

→ single natural image를 활용한 unconditional generation 모델 제안

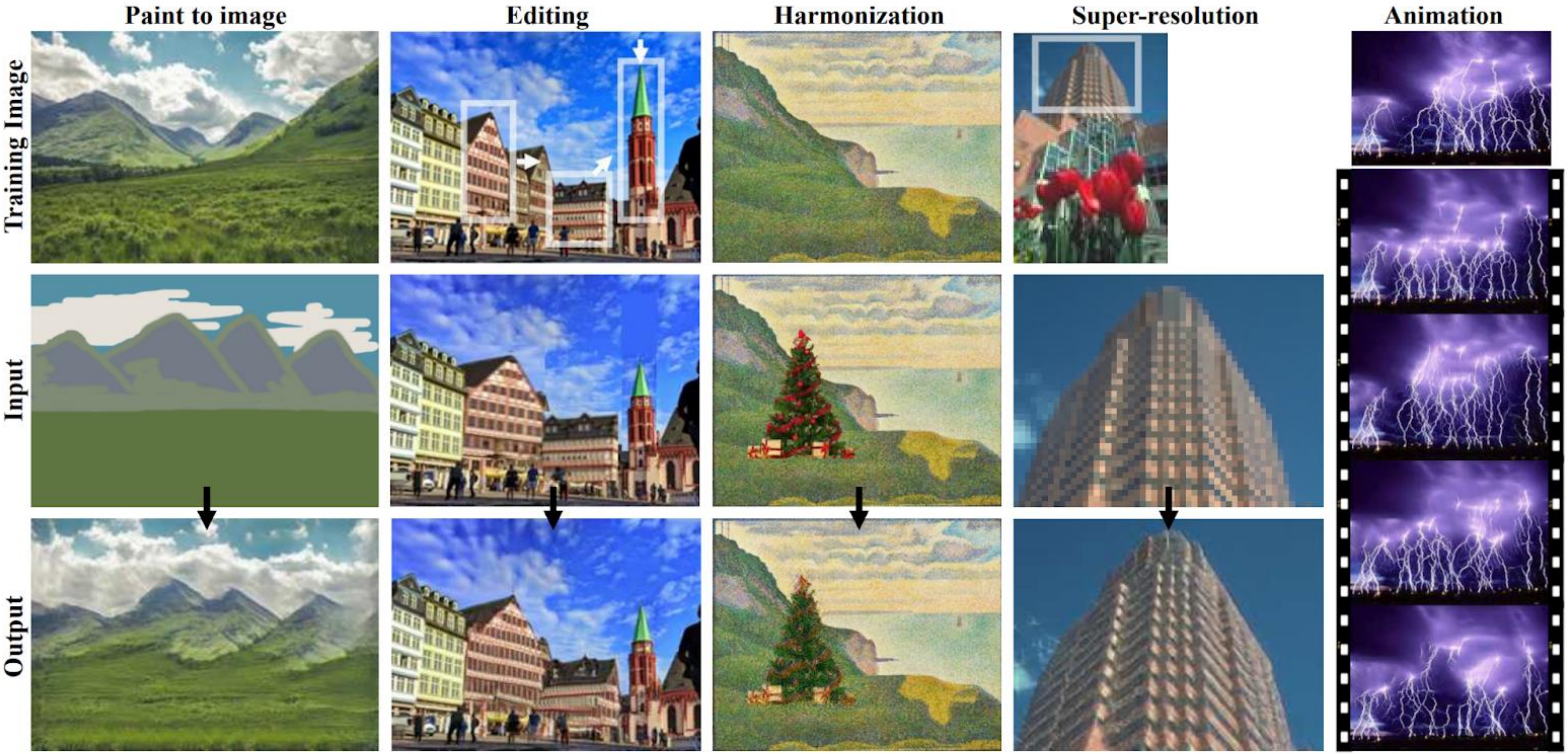
✓ **A pyramid of fully convolutional light-weight GAN** :

✓ 각 scale에서 patch distribution을 capture

✓ 다양한 image manipulation task

→ paint-to-image, editing, harmonization, super-resolution, animation

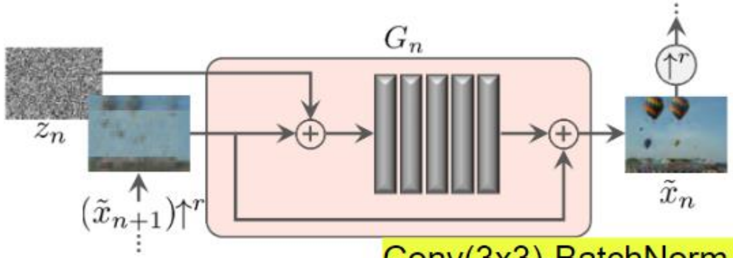
Introduction



Method

- ✓ 한 장의 원본 이미지로부터 **overlapping되는 patch의 internal statistics를 capture**하는 unconditional generative model
- ✓ 각 scale에서 다른 특성을 capture
  - ✓ arrangement, shape (global property) → fine detail, texture
- ✓ Patch-GAN → 계층구조

Method



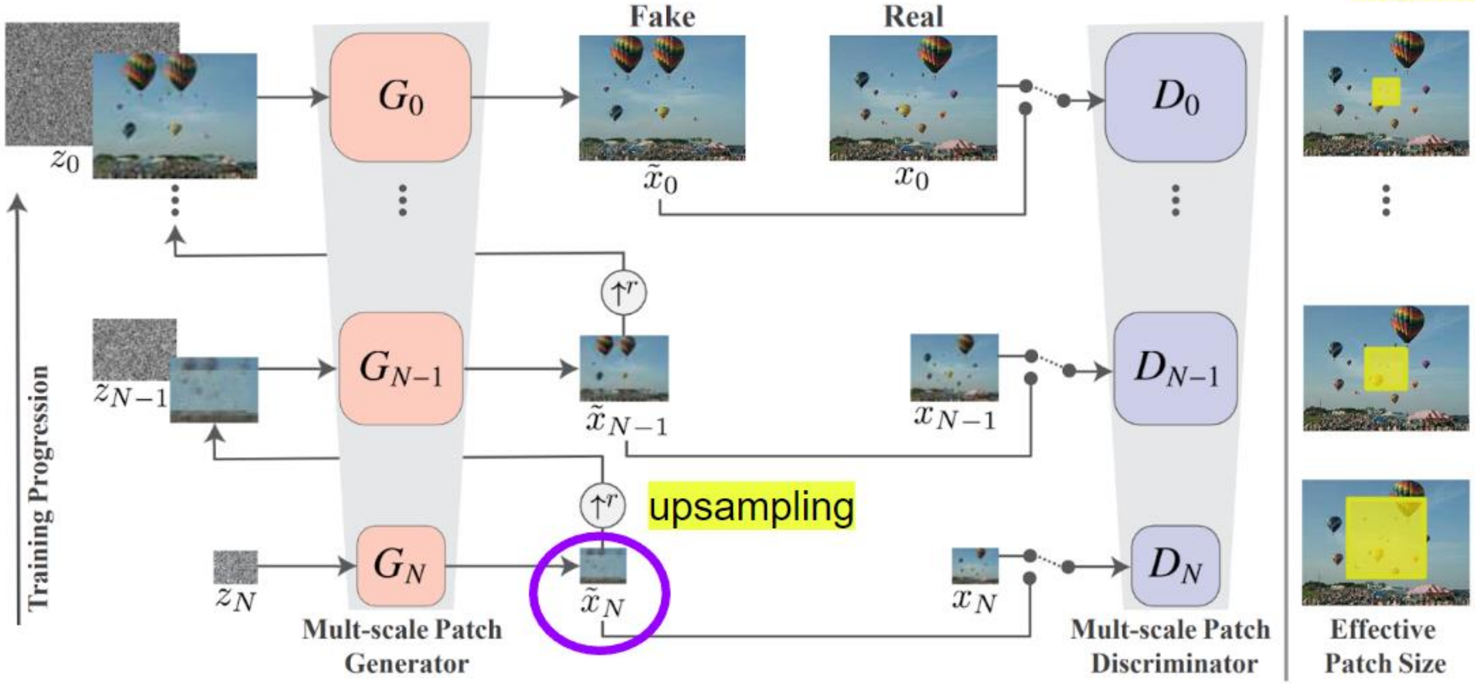
Conv(3x3)-BatchNorm-LeakyReLU

$$\tilde{x}_N = G_N(z_N). \tag{1}$$

$$\tilde{x}_n = G_n\left(z_n, (\tilde{x}_{n+1})\uparrow^r\right), \qquad n < N. \tag{2}$$

$$\tilde{x}_n = (\tilde{x}_{n+1})\uparrow^r + \psi_n\left(z_n + (\tilde{x}_{n+1})\uparrow^r\right), \tag{3}$$

fully conv layer



upsampling

Method

✓ **Training (real : 1, fake : 0)**

$$\min_{G_n} \max_{D_n} \mathcal{L}_{\text{adv}}(G_n, D_n) + \alpha \mathcal{L}_{\text{rec}}(G_n). \qquad (4)$$

✓ adversarial loss :

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

✓ patch 단위로 학습 진행 (patch 내의 distribution 분포가 같도록)

✓ reconstruction loss :

$$\mathcal{L}_{\text{rec}} = \|G_n(0, (\tilde{x}_{n+1}^{\text{rec}}) \uparrow^r) - x_n\|^2, \qquad (5)$$

for $n = N$, we use $\mathcal{L}_{\text{rec}} = \|G_N(z^*) - x_N\|^2$

Result

- ✓ 데이터셋 : Berkeley Segmentation Database, Places, and the web
- ✓ 25px(at the coarsest) ~ 250px(at the finest), scale N은 지정

Result

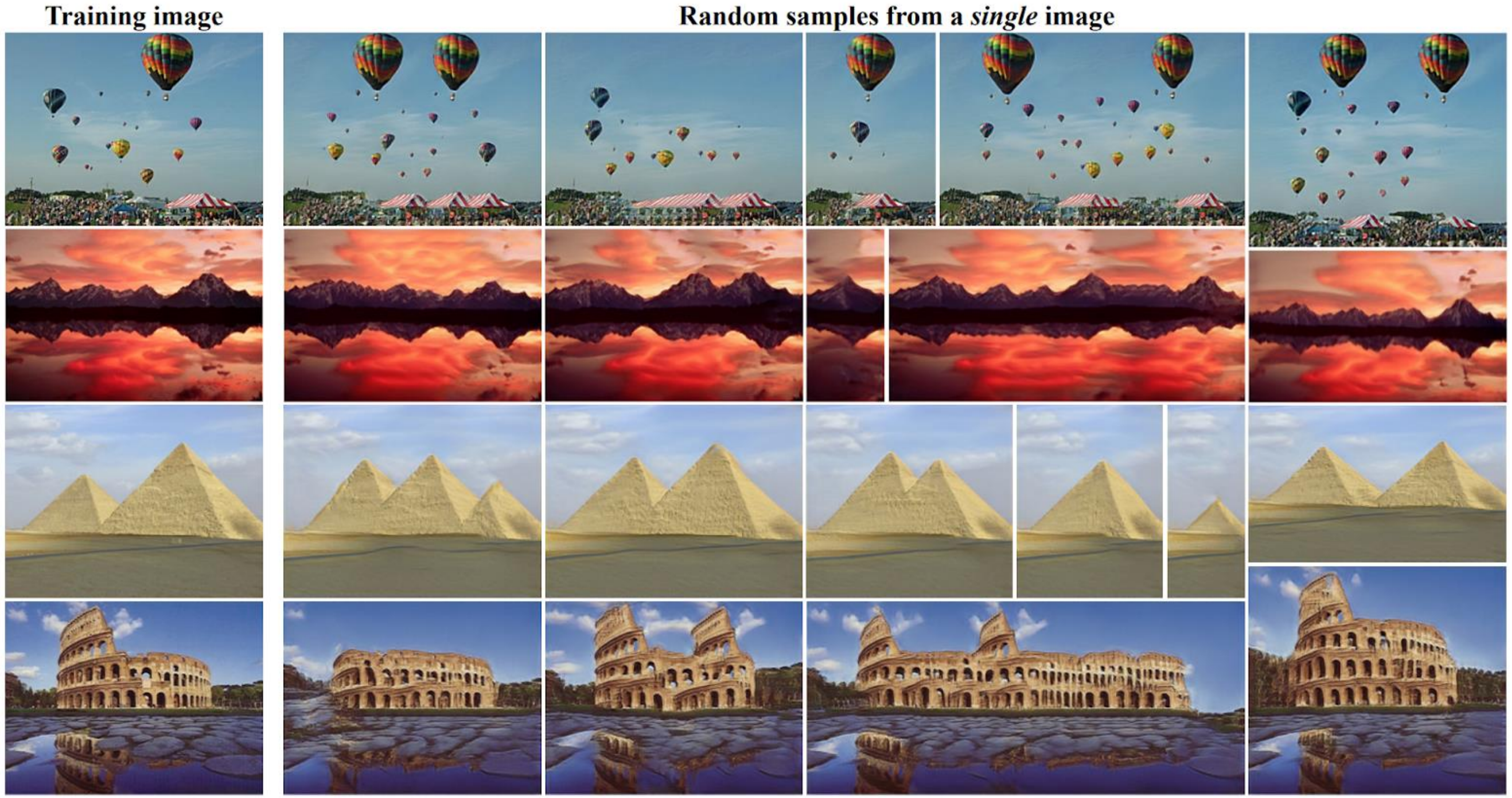


Figure 6: **Random image samples.** After training SinGAN on a single image, our model can generate realistic random image samples that depict new structures and object configurations, yet preserve the patch distribution of the training image. Because our model is fully convolutional, the generated images may have arbitrary sizes and aspect ratios. Note that our goal is not image retargeting – our image samples are random and optimized to maintain the patch statistics, rather than preserving salient objects. See SM for more results and qualitative comparison to image retargeting methods.

Result

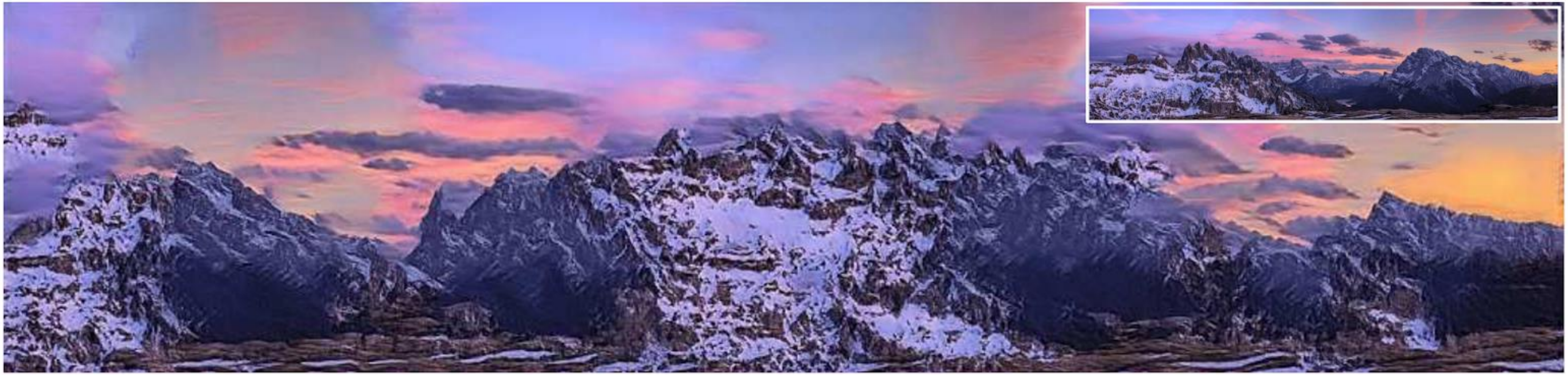


Figure 7: **High resolution image generation.** A random sample produced by our model, trained on the 243 × 1024 image (upper right corner); new global structures as well as fine details are realistically generated. See 4Mpix examples in SM.

Result

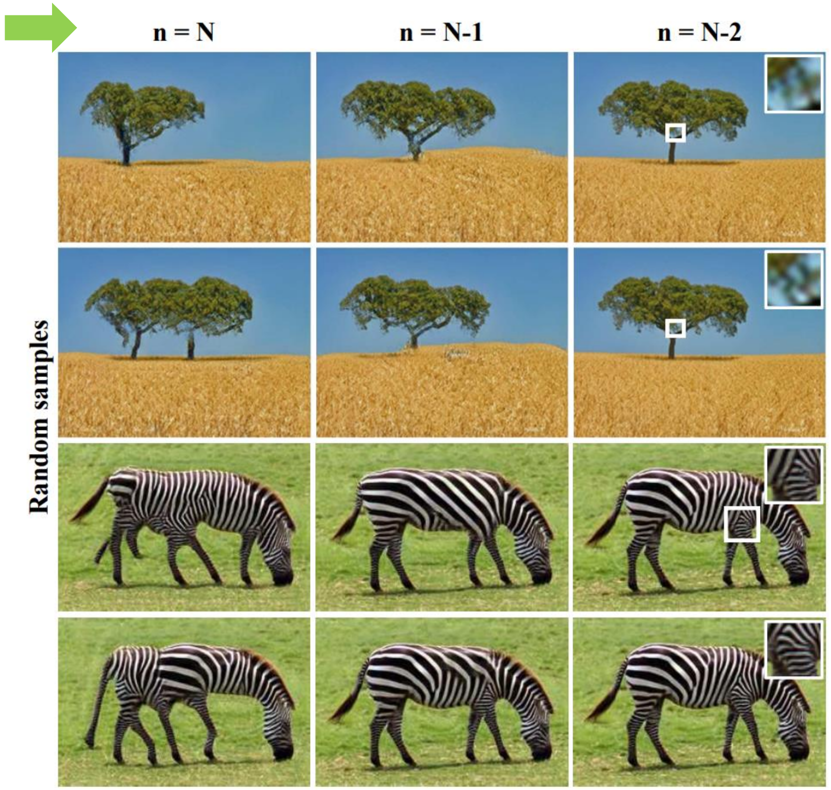


Figure 8: **Generation from different scales (at inference).** We show the effect of starting our hierarchical generation from a given level n. For our full generation scheme ($n = N$), the input at the coarsest level is random noise. For generation from a finer scale n, we plug in the downsampled original image, x_n, as input to that scale. This allows us to control the scale of the generated structures, *e.g.*, we can preserve the shape and pose of the Zebra and only change its stripe texture by starting the generation from $n = N - 1$.

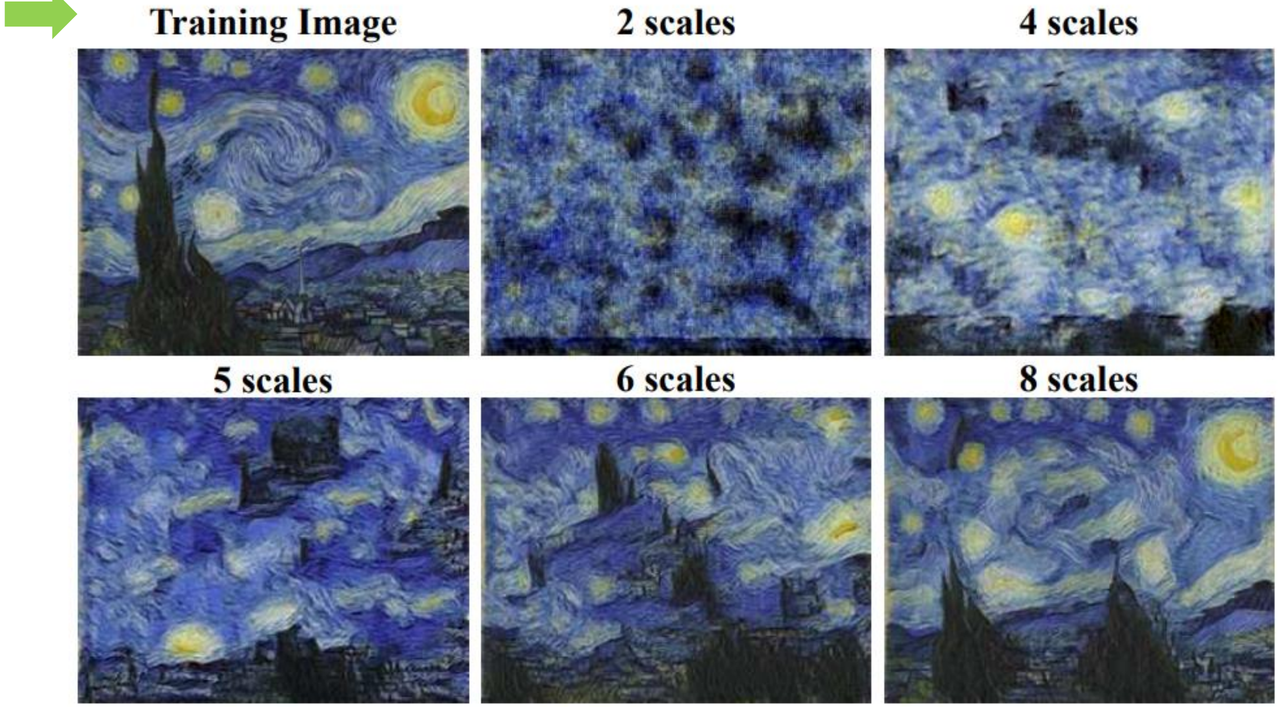
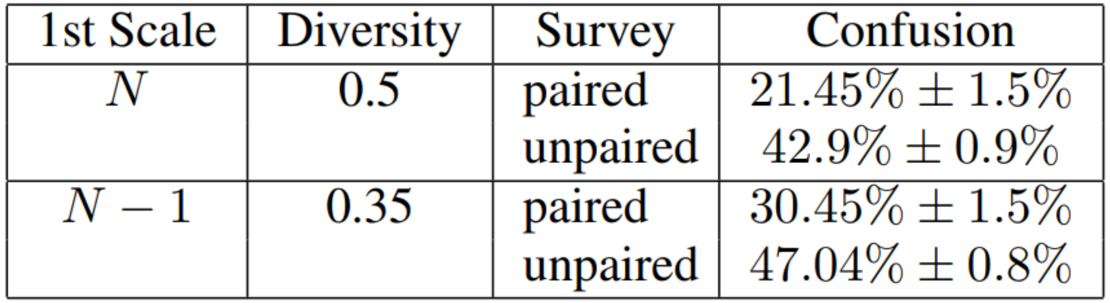


Figure 9: **The effect of training with a different number of scales.** The number of scales in SinGAN's architecture strongly influences the results. A model with a small number of scales only captures textures. As the number of scales increases, SinGAN manages to capture larger structures as well as the global arrangement of objects in the scene.

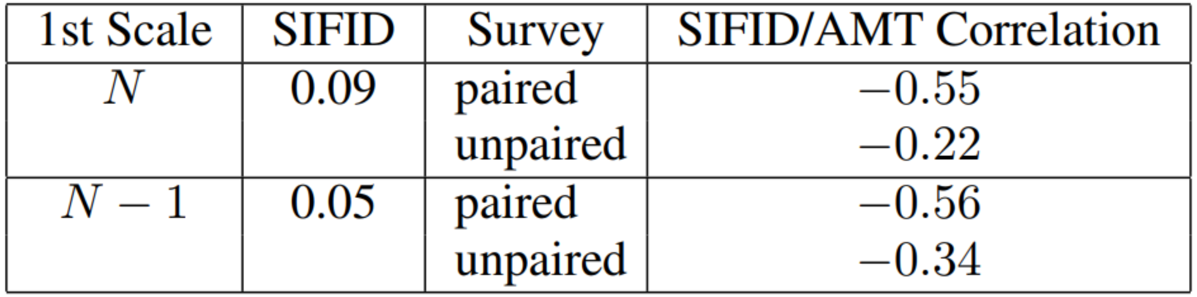
Evaluation

- ✓ Quantitative Evaluation
  - ✓ AMT (Amazon Mechanical Turk) : 50에 가까울수록 이미지와 유사
  - ✓ FID(Frechet Inception Distance) : a common metric for GAN

    → SIFID (Single Image FID) : 작을수록 좋은 값



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



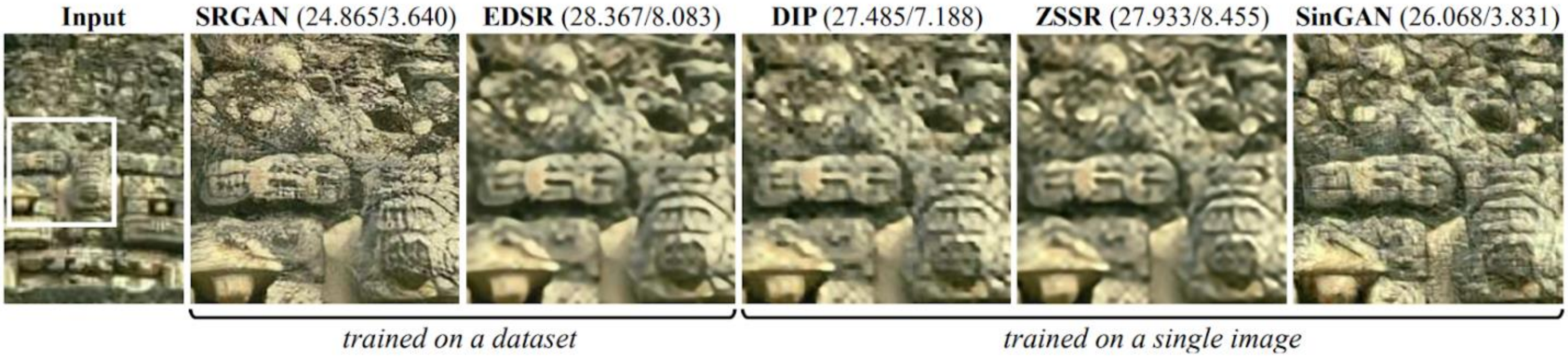
1st Scale	SIFID	Survey	SIFID/AMT Correlation
N	0.09	paired unpaired	-0.55 -0.22
$N-1$	0.05	paired unpaired	-0.56 -0.34

Application

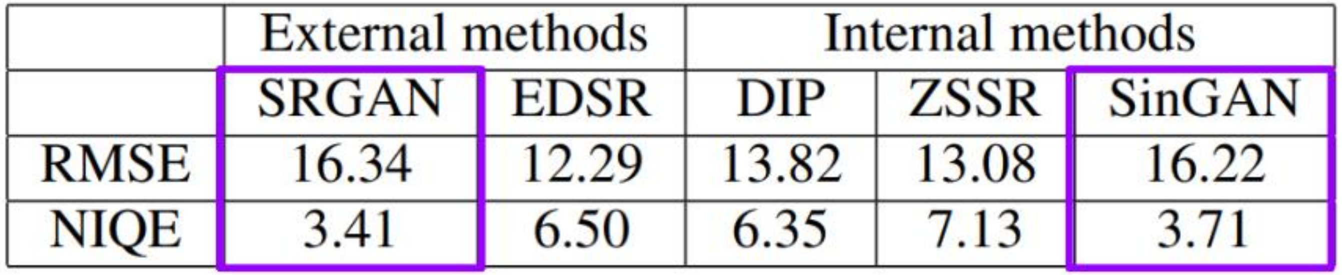
1. Super-Resolution
2. Paint-to-Image
3. Harmonization
4. Editing
5. Single Image Animation

Application “Super-Resolution”

PSNR/NIQE



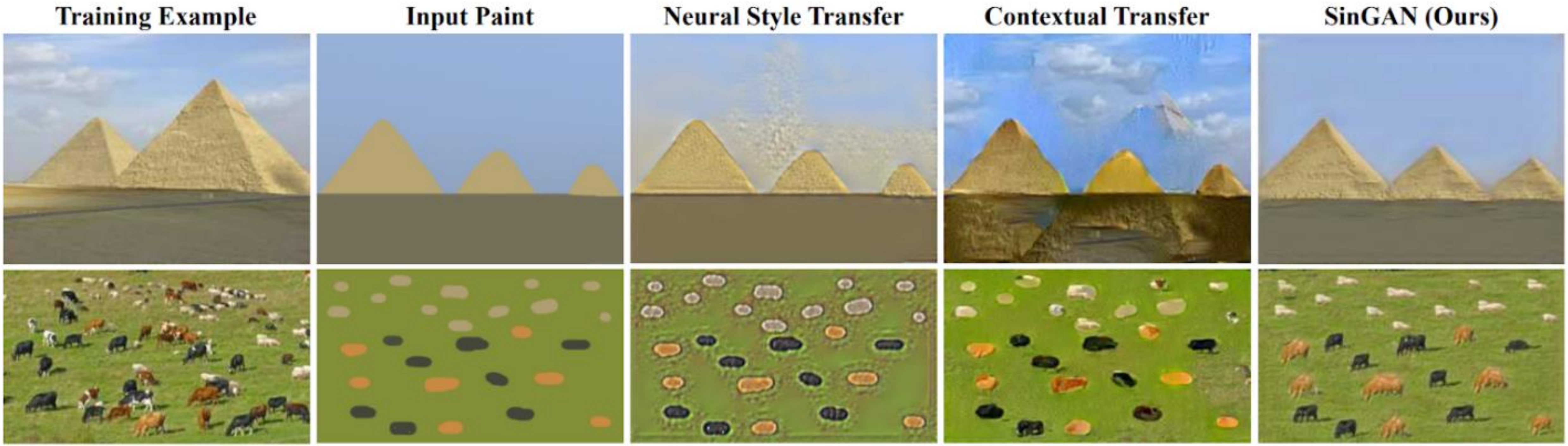
BSD100



	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

Application “Paint-to-Image”

well-known style transfer method



Application “Harmonization”, “Editing”

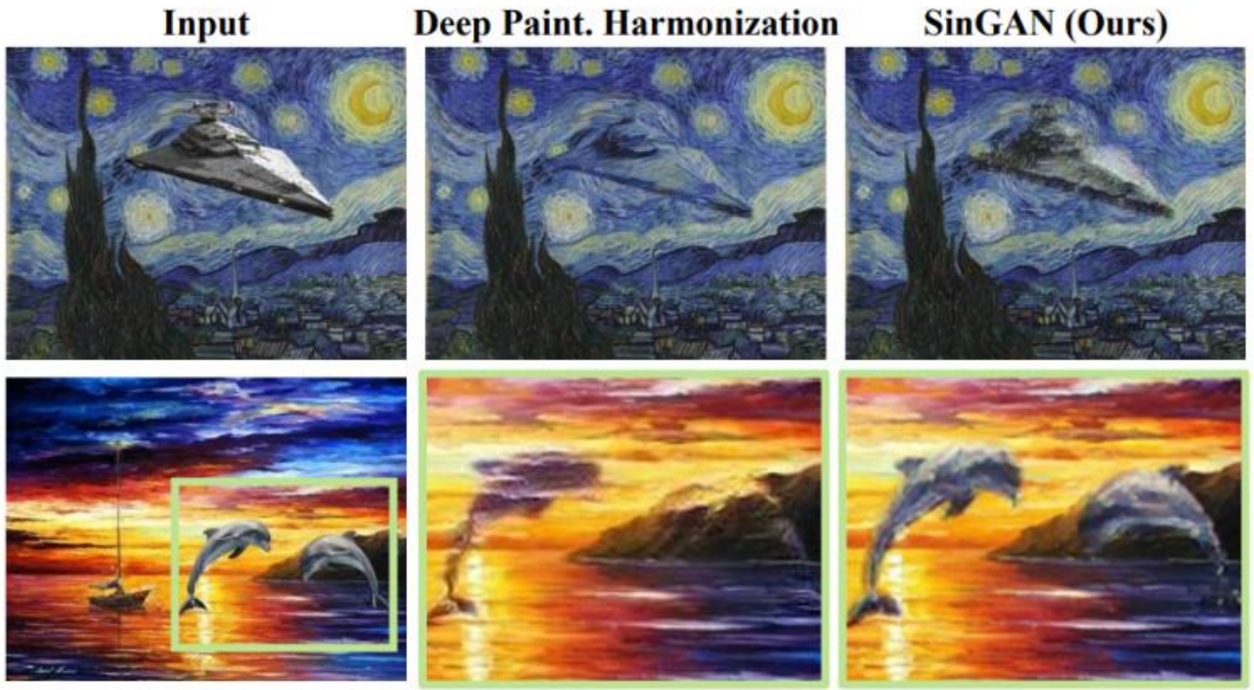


Figure 13: **Harmonization.** Our model is able to preserve the structure of the pasted object, while adjusting its appearance and texture. The dedicated harmonization method [34] overly blends the object with the background.

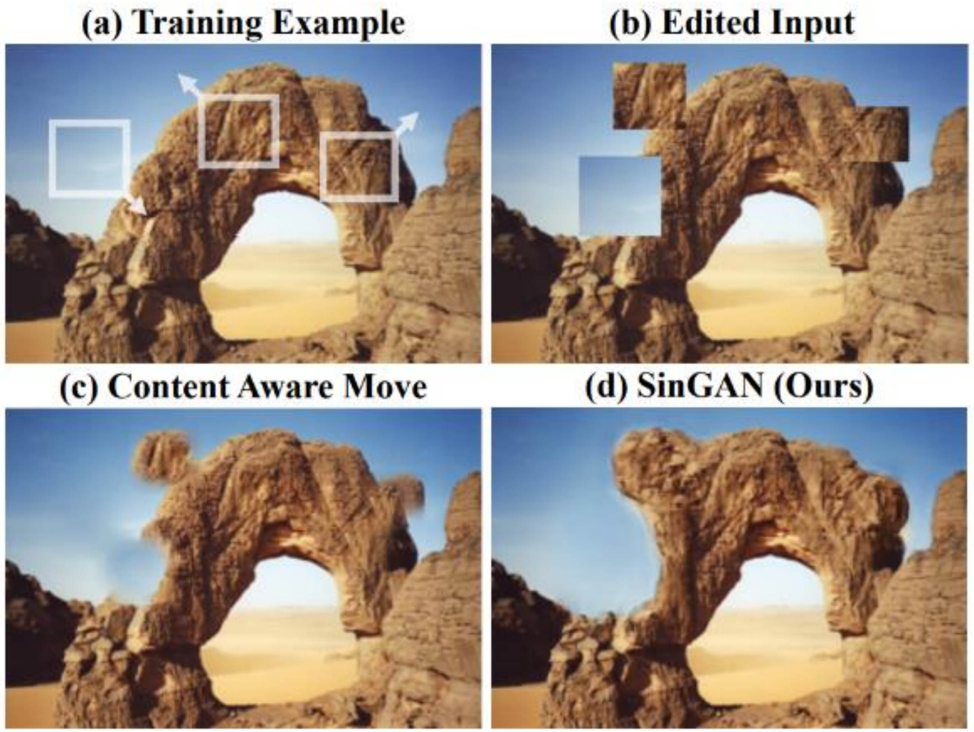
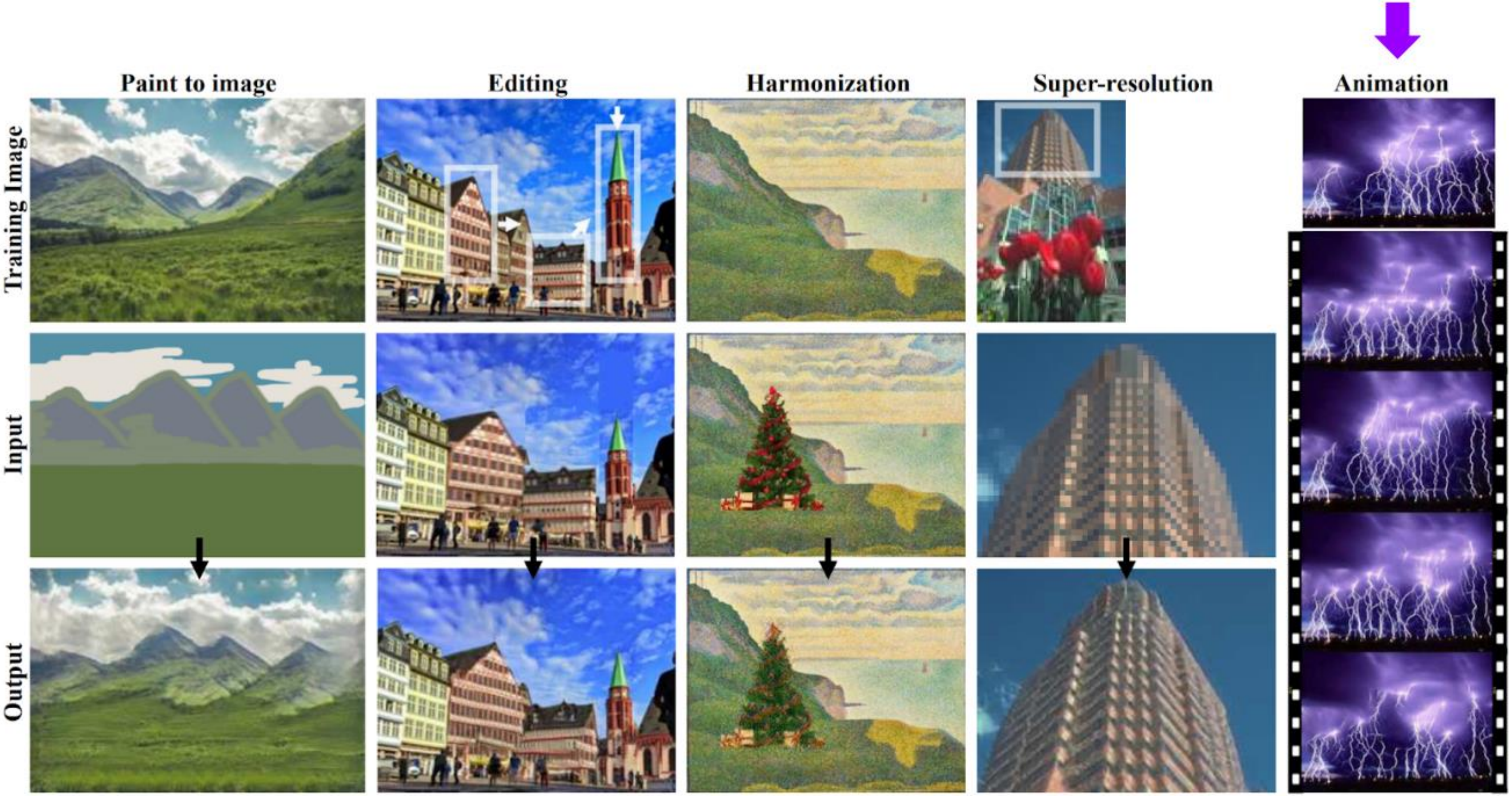


Figure 12: **Editing.** We copy and paste a few patches from the original image (a), and input a downsampled version of the edited image (b) to an intermediate level of our model (pretrained on (a)). In the generated image (d), these local edits are translated into coherent and photo-realistic structures. (c) comparison to Photoshop content aware move.

Application "Animation"



Conclusion

- ✓ **Single natural image** 사용하는 새로운 **unconditional generative schema** 제안
- ✓ Internal learning → **inherently 한계**
  - ✓ single dog → will not generate sample of different dog breeds
- ✓ 그럼에도, SinGAN은 다양한 image manipulation task에 쓰이는 강력한 tool

References

[Paper]

Blau, Y., & Michaeli, T. (2018). The perception-distortion tradeoff. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 6228-6237).

[Lecture]

https://www.youtube.com/watch?v=WxVvSXtRJqA

Q & A

Thank you ☺