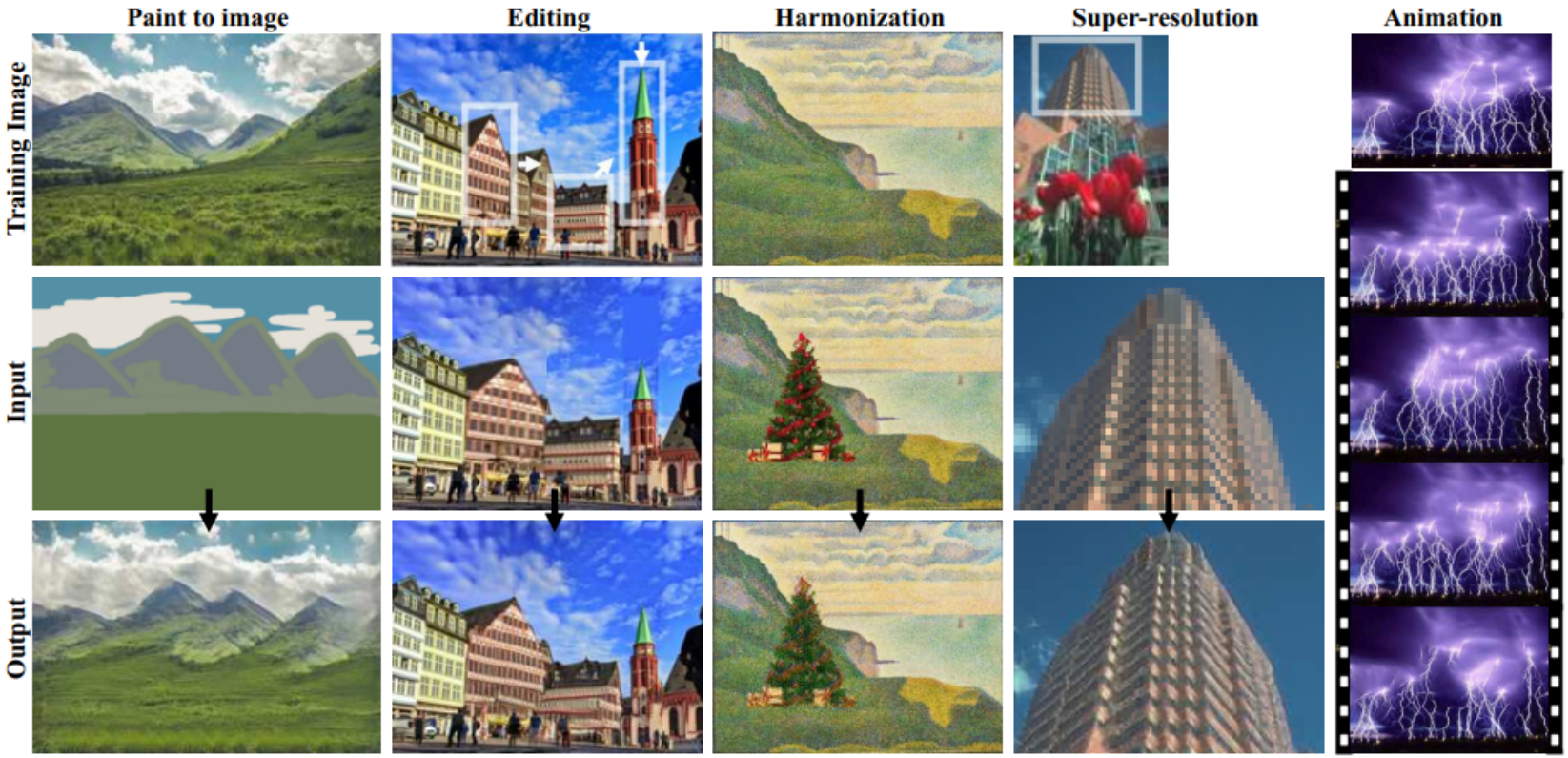
SinGAN: Learning a Generative Model from a Single Natural Image

Ngày 16 tháng 11 năm 2019

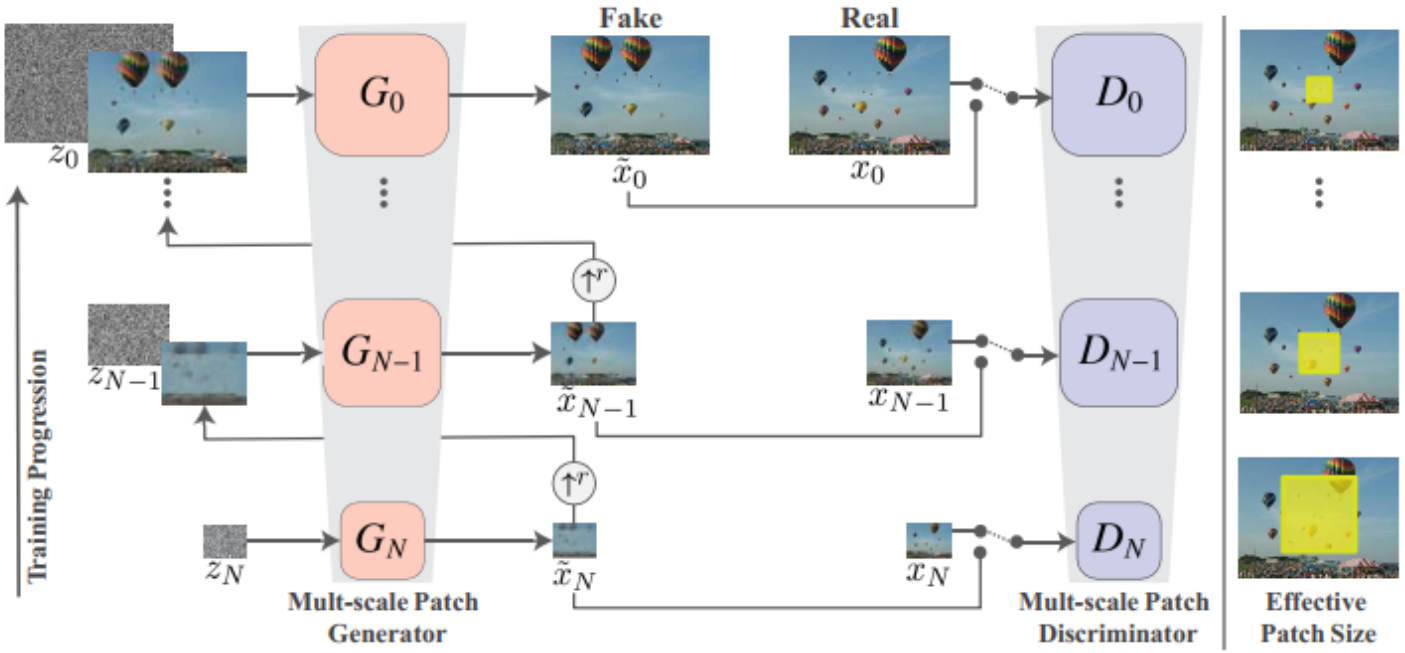
Introduction

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



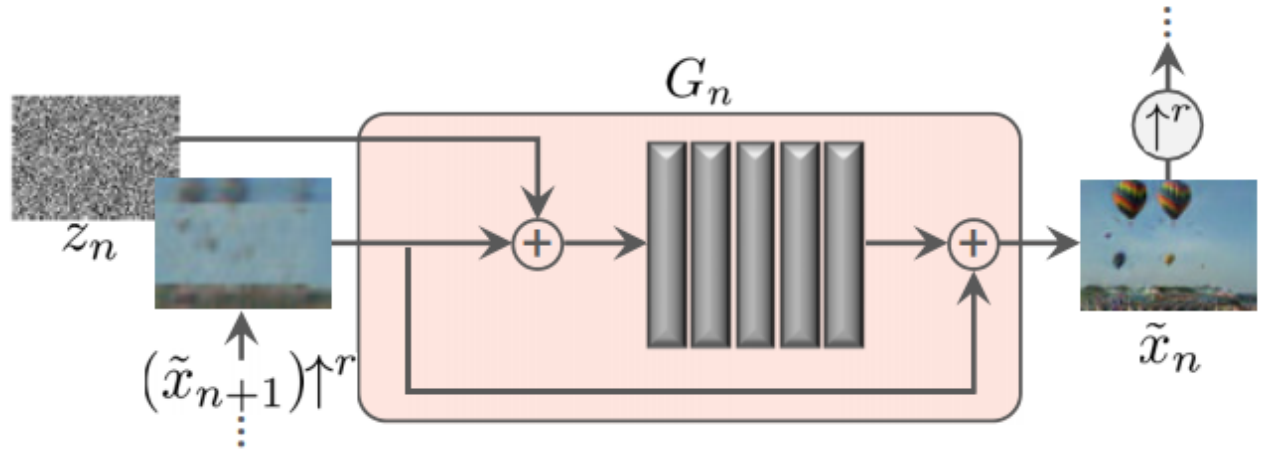
Multi-scale architecture

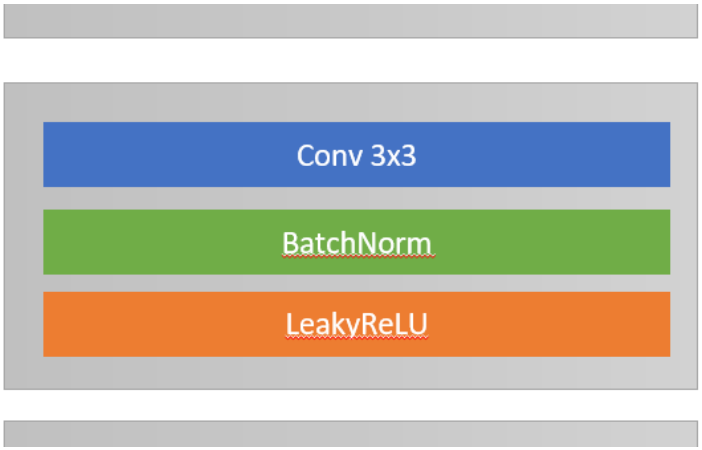


$$\tilde{x}_n = G_n\left(z_n, (\tilde{x}_{n+1})\right)$$

Discriminator phân biệt theo từng patch bị overlap (patchGAN)

Residual Learining in Multi-Scale Generation





$$\widetilde{x}_n = (\widetilde{x}_{n+1})\uparrow^r + \phi_n\,(z_n + (\widetilde{x}_{n+1})\uparrow^r)$$

Tại bước đầu tiên (coarsest scale) có 32 kernel ở mỗi block, và cứ 4 scale số kernel sẽ được tăng gấp 2 lần lên

Objective Function

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

Adversial loss

$$\mathcal{L}_{adv} = \underbrace{D_n\left(\widetilde{x}_n\right)}_{\text{Fake data}} - \underbrace{D_n\left(x_n\right)}_{\text{Real data}} + \lambda \underbrace{\left(\left\|\nabla_{\widehat{x}_n} D\left(\widehat{x}_n\right)\right\|_2 - 1\right)^2}_{\text{gradient penalty}}$$

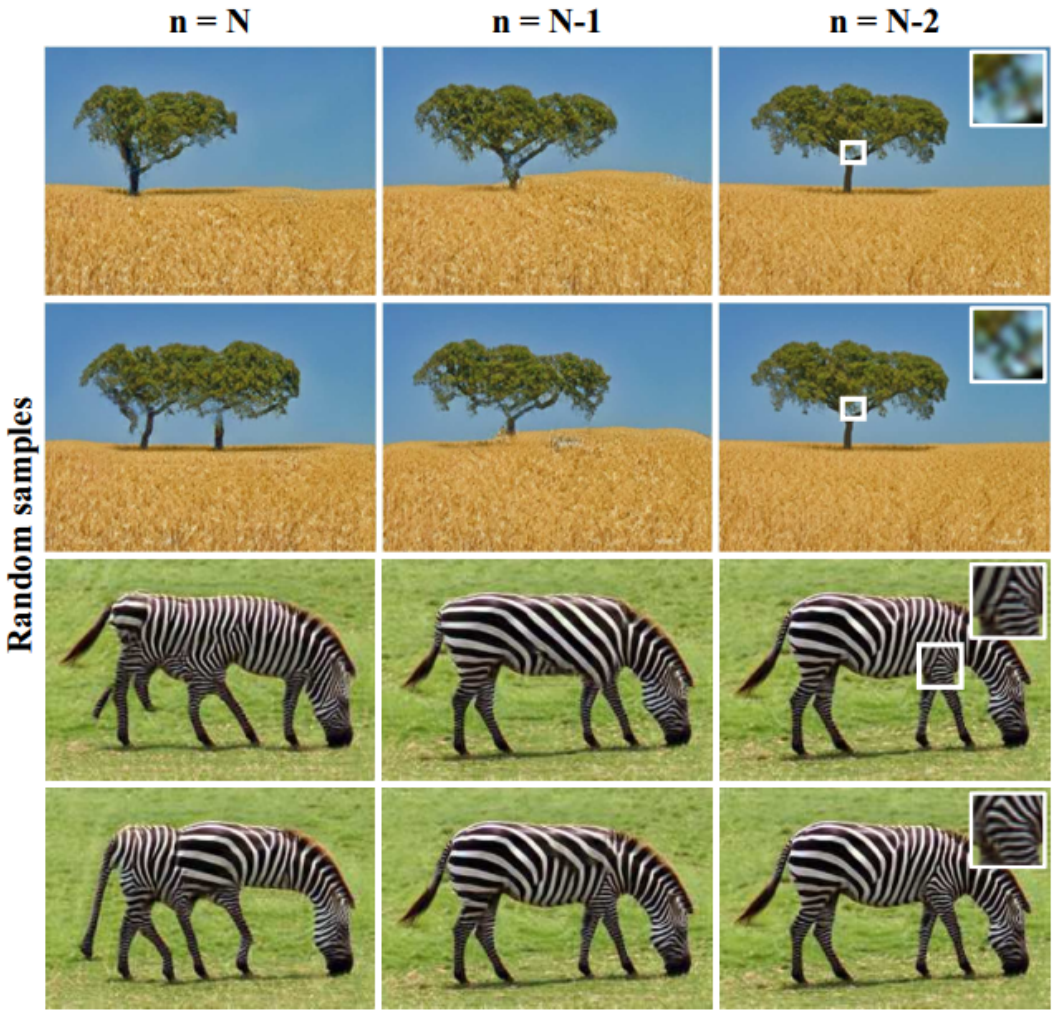
$$\widehat{x}_n = (1-t)\,\widetilde{x}_n + t x_n$$

Recontrusction loss : chọn $\left\{z_N^{rec}, z_{N-1}^{rec}, ..., z_0^{rec}\right\} = \{z^*, 0, ..., 0\}$

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

khi $n = N$, ta sử dụng $L_{rec} = \left\| G_N\left(z^*\right) - x_N \right\|^2$

Generation from different scales (at inference)



The effect of training with a different number of scales.

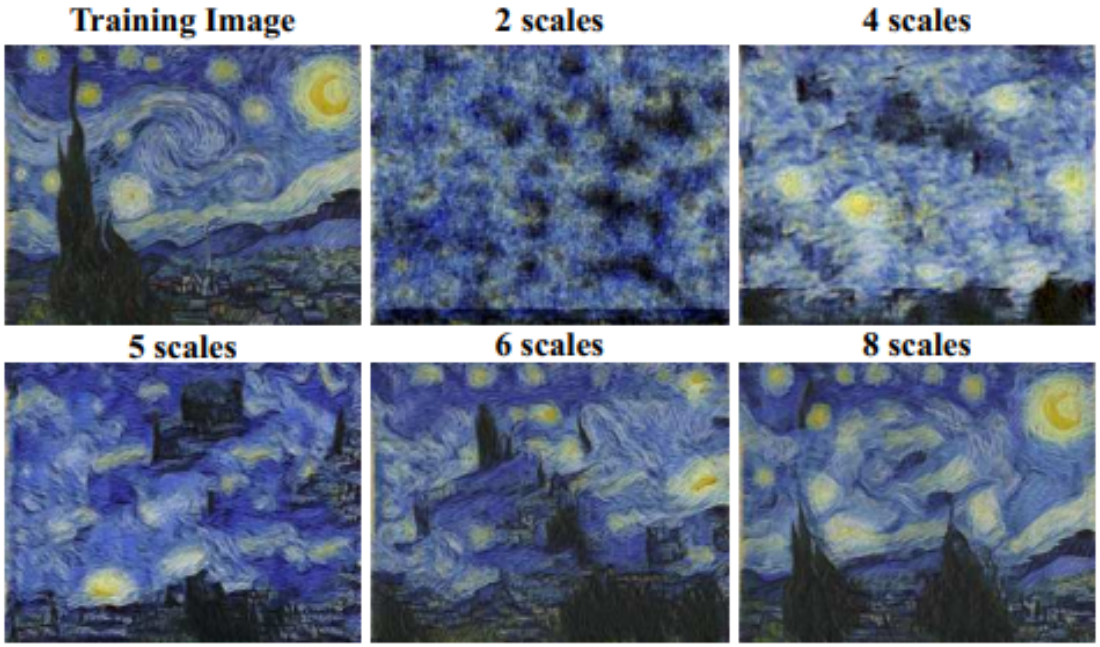


Figure 9: **The effect of training with a different number of scales.** The number of scales in SinGAN's architecture strongly influences the results. A model with a small number of scales only captures textures. As the number of scales increases, SinGAN manages to capture larger structures as well as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