SinGAN: Learning a Generative Model from a Single Natural Image

Tamar Rott Shaham
Technion

Tali Dekel
Google Research

Tomer Michaeli
Technion

ICCV 19 Best Paper Award

演讲人：顾嘉辉

日期：2019.11.15

作者介绍

第一作者来自以色列工学院

Tamar Rott Shaham

PhD Candidate, Computer Vision

TECHNION
Israel Institute
of Technology

HOME ABOUT NEWS PUBLICATIONS

About

I am a PhD candidate, supervised by Prof. Tomer Michaeli at the Electrical Engineering faculty of the Technion, where I also received my BSc. My research interests include image processing, computer vision and machine learning.

Email: stamarot at campus dot technion.ac.il

Office: Mayer 514

作者介绍

第二作者来自谷歌研究院

Tali Dekel

tdekel@google.com | talidekel@csail.mit.edu | talibasha@gmail.com

[Curriculum Vitae]

I'm a Research Scientist at Google, Cambridge. Before that I was a Postdoctoral Associate with Prof. Bill Freeman, at CSAIL, MIT. I completed my Ph.D. in the school of Electrical Engineering of Tel-Aviv University, where I was supervised by Prof. Shai Avidan (TAU) and Prof. Yael Moses (IDC). My main research interests include images and videos analysis, multi-view systems, 3D structure and motion estimation, image synthesize and rendering.

作者介绍

第三作者同样来自以色列工学院

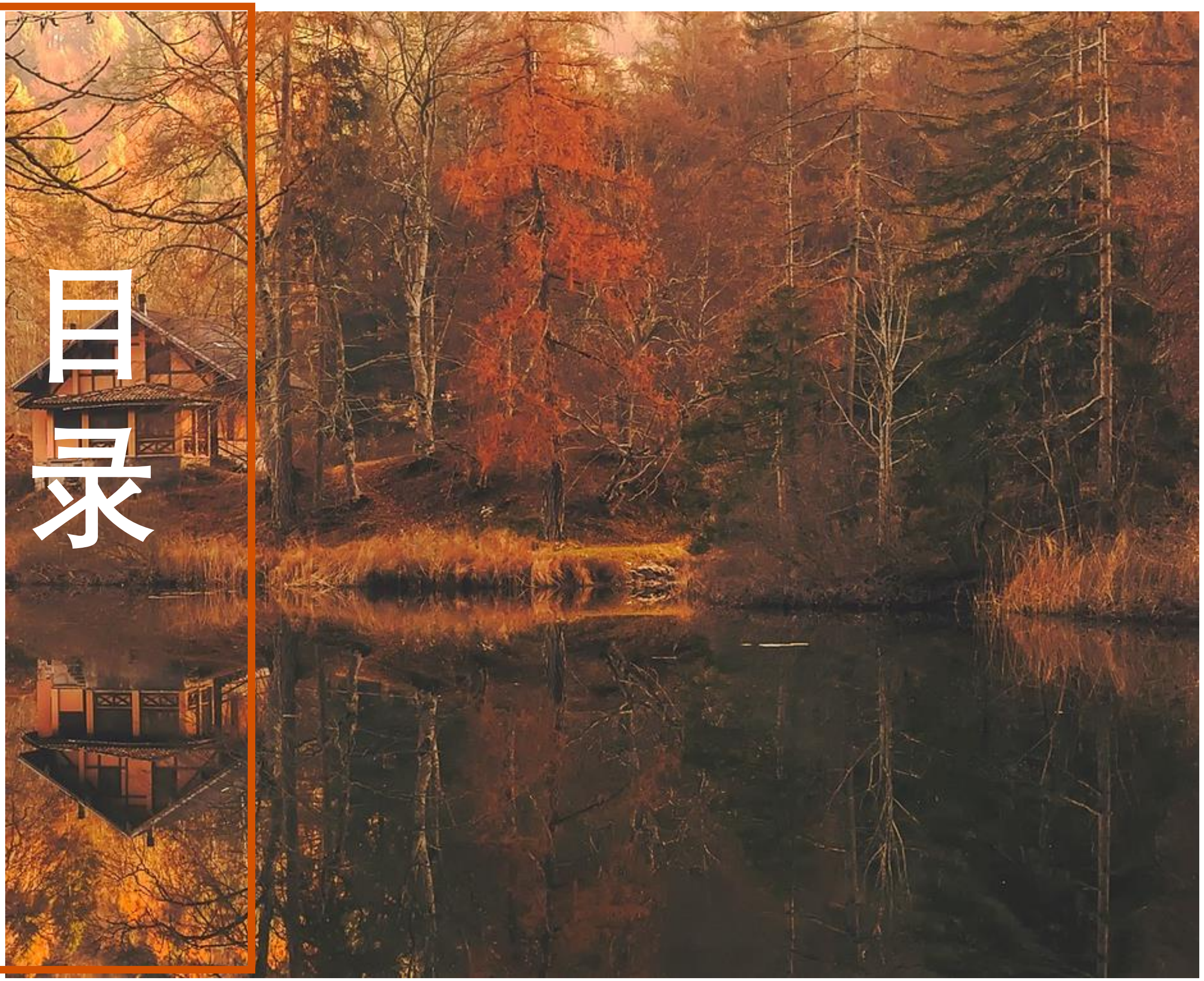
I am an Assistant Professor at the Faculty of Electrical Engineering at the Technion – Israel Institute of Technology

---

Research interests:

- Image Processing
- Computer Vision
- Signal Processing
- Machine Learning

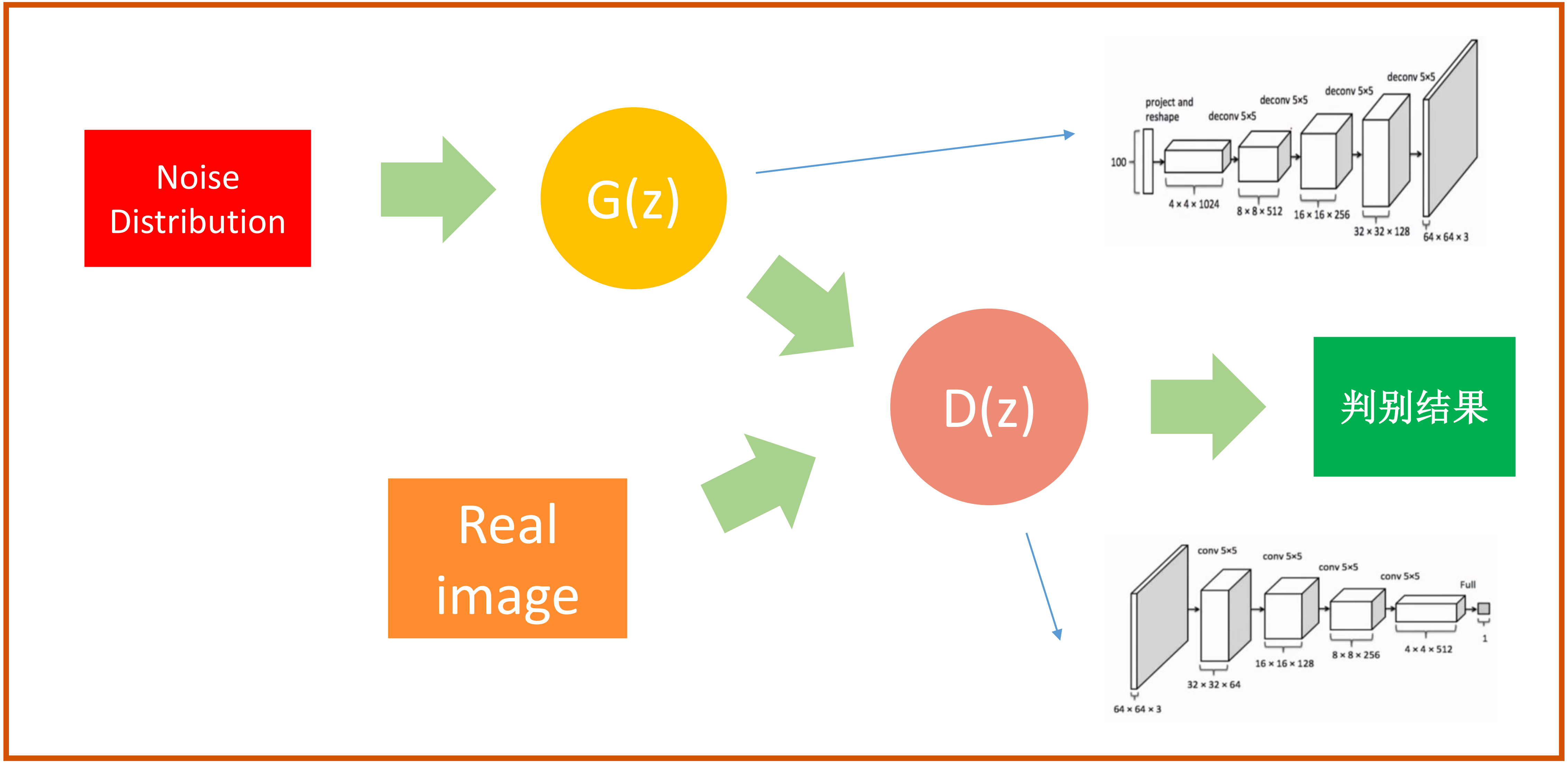
目录



01

原始GAN的介绍

原始GAN的架构



Noise Distribution
G(z)
project and reshape
deconv 5×5
deconv 5×5
deconv 5×5
deconv 5×5
100
4 × 4 × 1024
8 × 8 × 512
16 × 16 × 256
32 × 32 × 128
64 × 64 × 3
D(z)
判别结果
Real image
conv 5×5
conv 5×5
conv 5×5
conv 5×5
Full
64 × 64 × 3
32 × 32 × 64
16 × 16 × 128
8 × 8 × 256
4 × 4 × 512
1

原始GAN的损失函数

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

G(z)

D(z)

$$E_{x \sim p_{data}(x)} \log(D(x))$$

$$D(x) = 1 \text{ when } x \sim p_{data}(x)$$

$$E_{z \sim p_z(z)} \log(1 - D(G(z)))$$

02

SinGAN的主要方法

Abstract

1、这篇文章提出了一个非条件生成模型架构SinGAN，可以生成多样化的样本

2、该网络架构采用了一个金字塔形的组合，由粗糙到精细逐渐训练图片的patch

3、所有的训练数据仅仅来自一张单一的图片，并且实现了图像处理中的多种用途

Introduction

1、此前的GAN网络架构也可以生成多样化的图片，但都使用的是包含多类对象的数据集

2、为了实现某一类对象的多样化甚至需要添加另外的输入信号（其实就是条件GAN）

3、该网络架构将GAN这项技术引入了一个新的领域，仅从一张单一的自然图像生成多样化的样本，且不需要来自该类的数据集

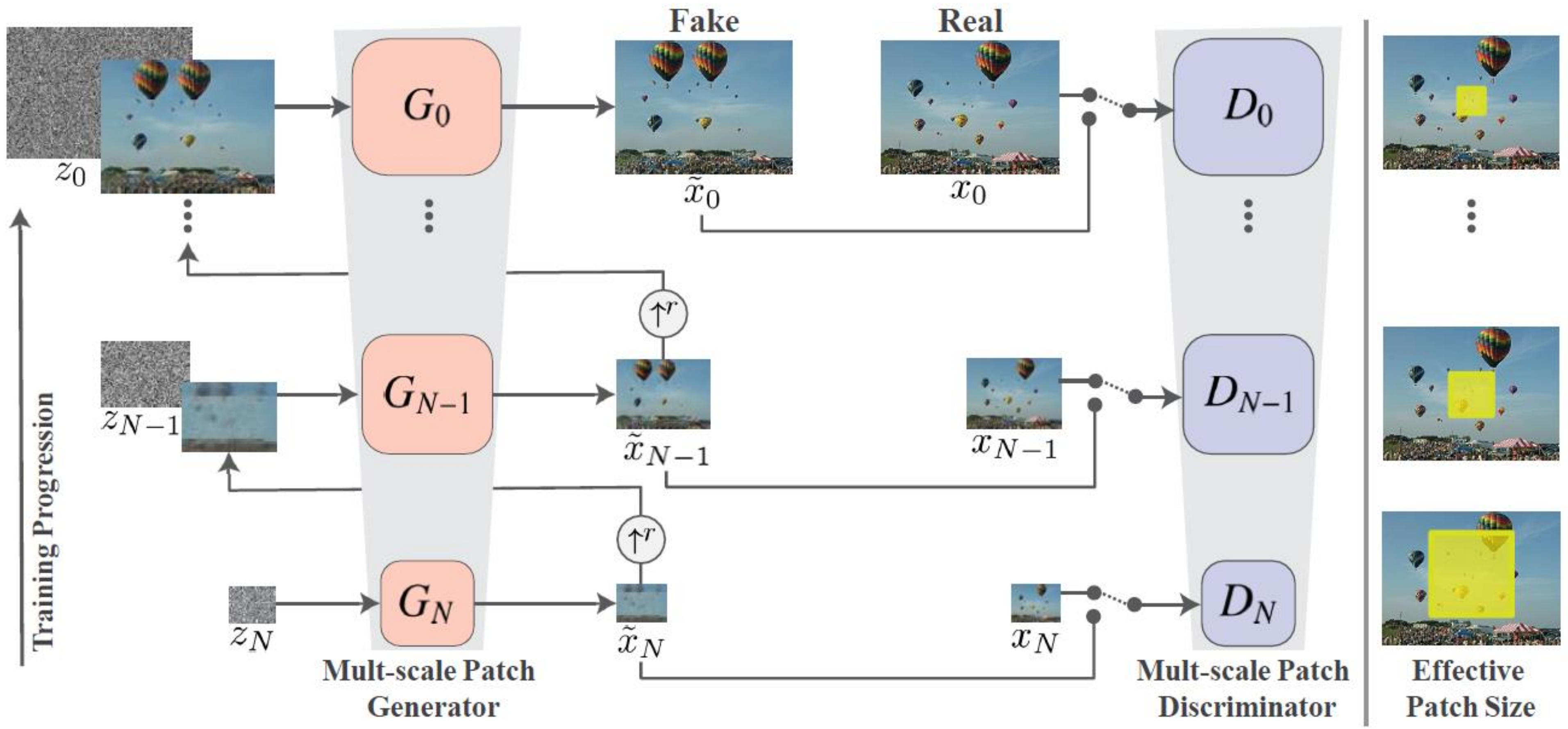
Method

1、和传统的GAN网络架构一致，该网络架构也使用了一个生成器（Generative）和一个判别器（Discriminator）

2、但是训练的样本并非是传统的方法那样使用整张图片，而是使用图片的patch

3、网络架构采用了金字塔形的结构，由粗糙到精细对图片的patch进行训练

Method



Fake
Real
G_0
$\tilde{x}_0$
x_0
D_0
z_0
z_{N-1}
G_{N-1}
$\tilde{x}_{N-1}$
x_{N-1}
D_{N-1}
$\uparrow^r$
z_N
G_N
$\tilde{x}_N$
x_N
D_N
Training Progression
Mult-scale Patch Generator
Mult-scale Patch Discriminator
Effective Patch Size

Architecture

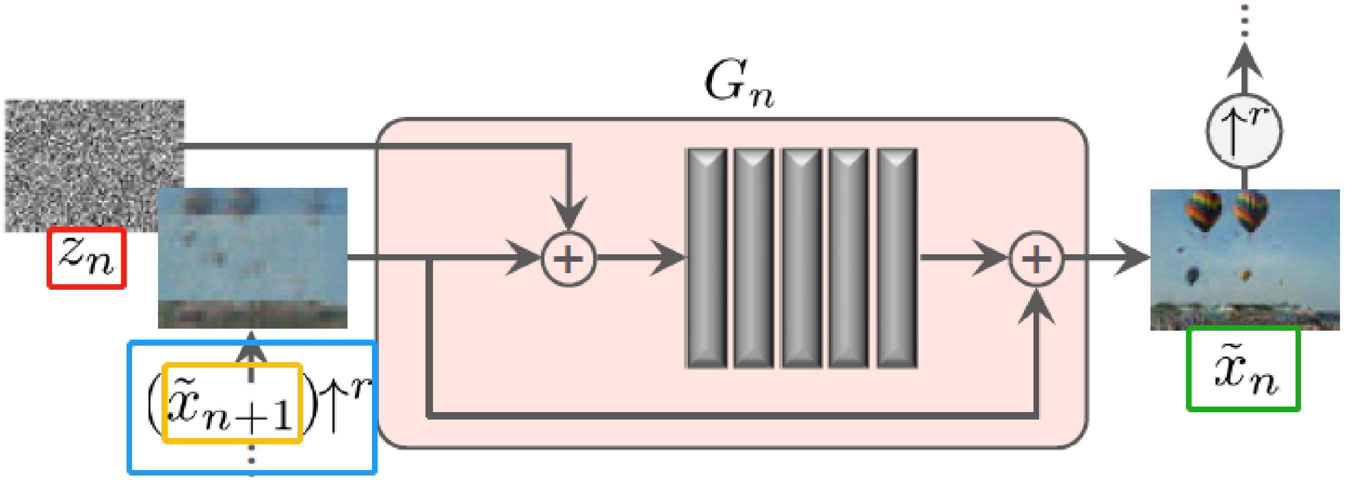


Figure 5: **Single scale generation.** At each scale n, the image from the previous scale, $\tilde{x}_{n+1}$, is upsampled and added to the input noise map, z_n. The result is fed into 5 conv layers, whose output is a residual image that is added back to $(\tilde{x}_{n+1})\uparrow^r$. This is the output $\tilde{x}_n$ of G_n.

Architecture

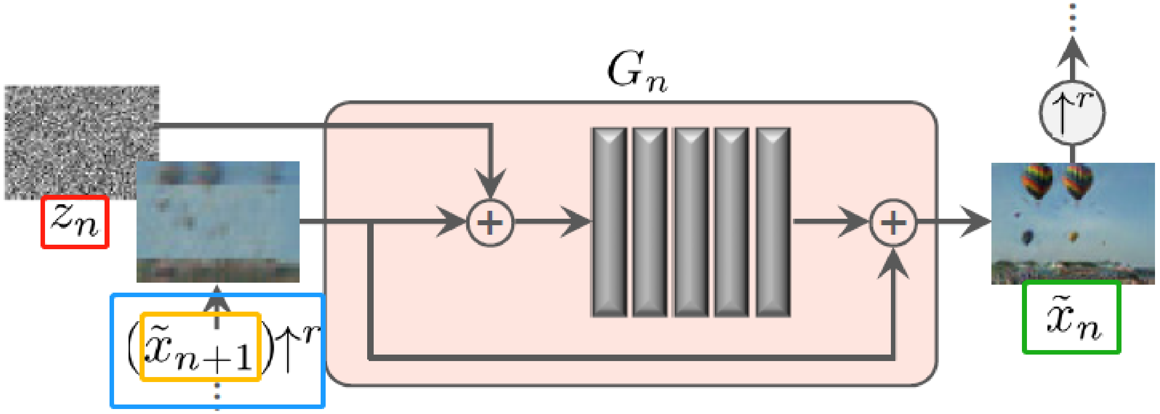


Figure 5: **Single scale generation.** At each scale n, the image from the previous scale, $\tilde{x}_{n+1}$, is upsampled and added to the input noise map, z_n. The result is fed into 5 conv layers, whose output is a residual image that is added back to $(\tilde{x}_{n+1})\uparrow^r$. This is the output $\tilde{x}_n$ of G_n.

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

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

Architecture

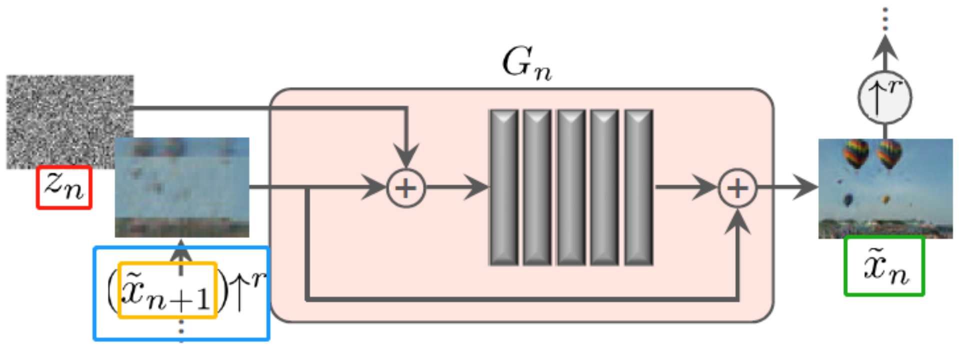


Figure 5: **Single scale generation.** At each scale n, the image from the previous scale, $\tilde{x}_{n+1}$, is upsampled and added to the input noise map, z_n. The result is fed into 5 conv layers, whose output is a residual image that is added back to $(\tilde{x}_{n+1})\uparrow^r$. This is the output $\tilde{x}_n$ of G_n.

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

Training

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

1、SinGAN的损失函数有两个部分组成：Adversial term和Reconstruction term

2、Adversial loss用于衡量生成图片和真实图片的distance

3、Reconstruction loss主要用于确保存在一组噪声输入是一定可以生成一张真假难辨的图片

Adversial loss

$$\mathcal{L}_{\text{adv}}(G_n, D_n)$$

1、Adversial loss使用的是WGAN-GP loss（是对Wasserstein GAN改进了的损失函数）

2、使用该损失函数显著提高了训练的稳定性

3、在计算最终的discriminator score的时候是求所有patch分数的平均值

Reconstruction loss

[illegible] Then for $n < N$,

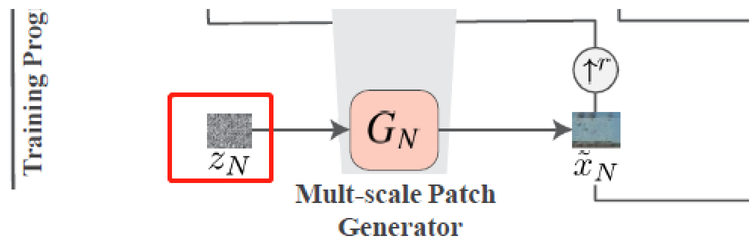
$$\mathcal{L}_{\text{rec}} = \|G_n(0, (\tilde{x}^{\text{rec}}_{n+1}) \uparrow^r) - x_n\|^2,$$

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

1、Reconstruction loss是为了确定存在一组具体的输入噪声，能够生成原始图片

2、reconstructed image还有另一个功能：在每一个训练的scale中确定输入噪声Zn的标准差

Reconstruction loss



Then for $n < N$,

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

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

1、由于该模型的最底层是直接从一堆噪声中生成一张图片（而且必须是真假难辨的图片），那么如何保证这一堆噪声一定可以生成这样的图片呢，就采用这个损失函数

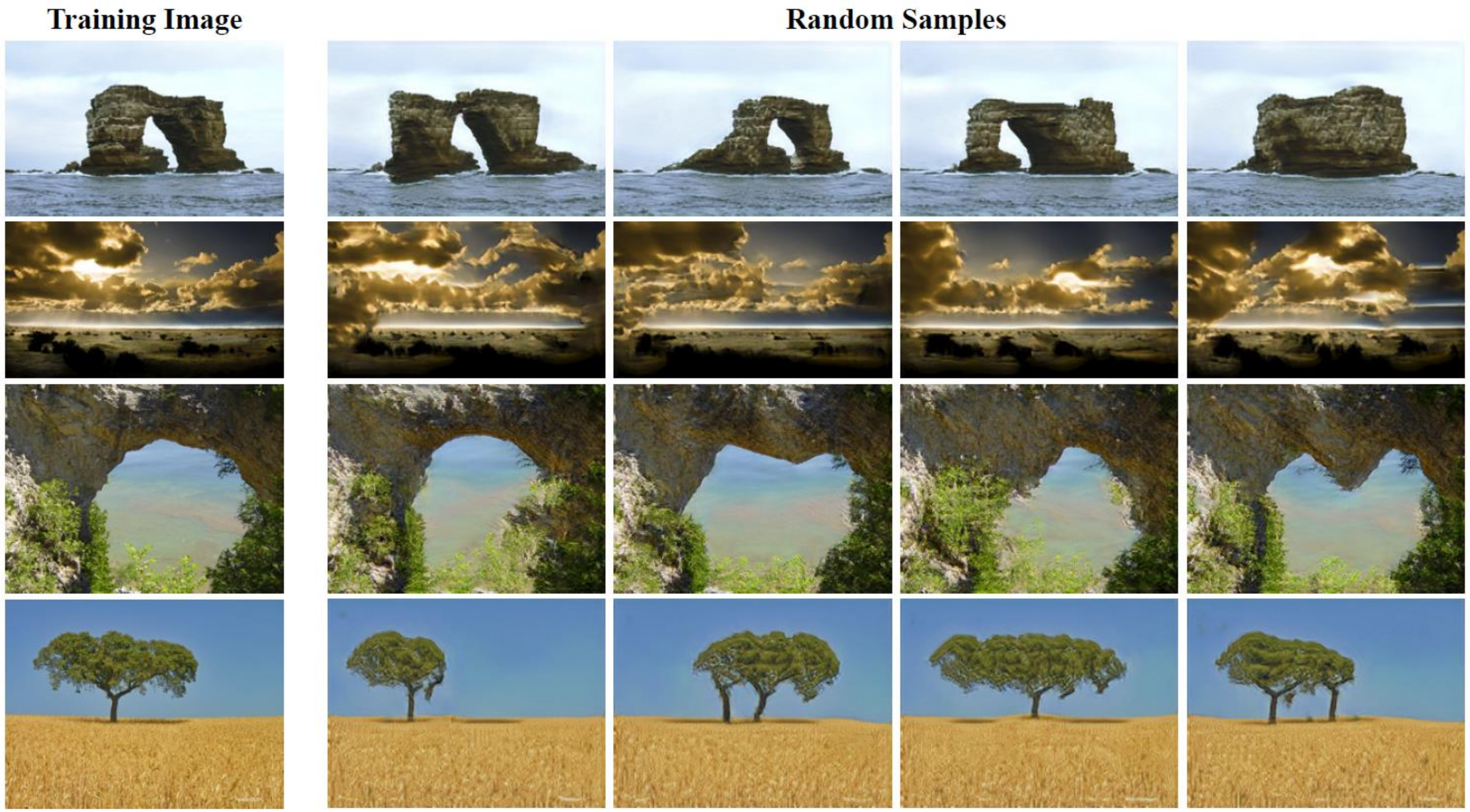
2、计算生成器生成的图片和原始图片衡量L2距离，训练时使这个值尽量小

3、有一个点要注意：除了最粗糙层（最底层）是用噪声放进生成器，其余每层都是用上一层上采样后的图片作为输入进入生成器

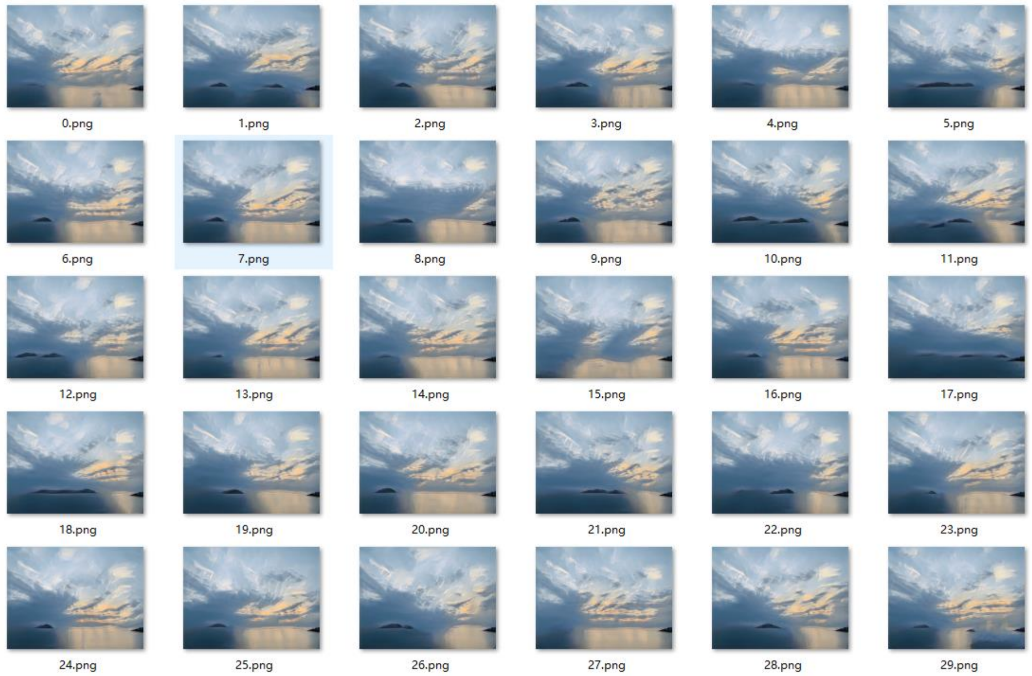
03

SinGAN的实验展示

Random Samples



Random Samples



Super-Resolution

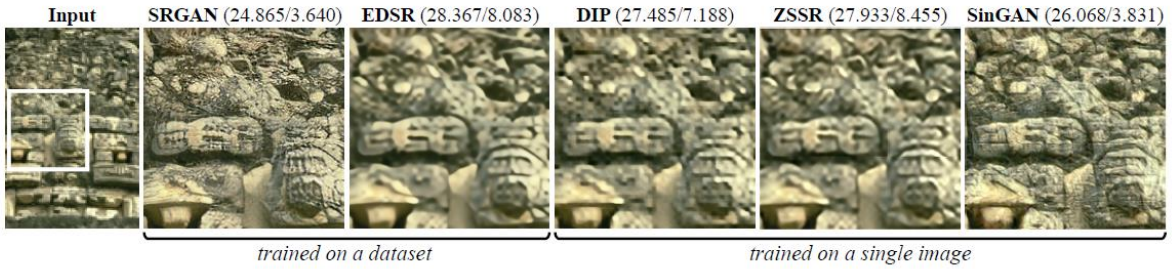
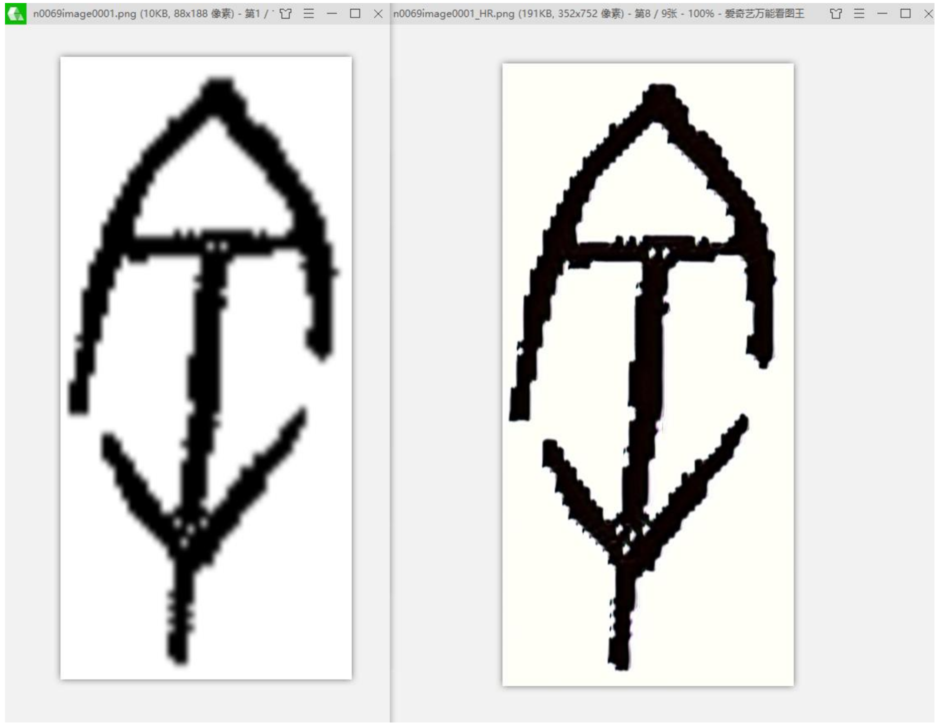


Figure 10: **Super-Resolution.** When SinGAN is trained on a low resolution image, we are able to super resolve. This is done by iteratively upsampling the image and feeding it to SinGAN's finest scale generator. As can be seen, SinGAN's visual quality is better than the SOTA internal methods ZSSR [46] and DIP [51]. It is also better than EDSR [32] and comparable to SRGAN [30], external methods trained on large collections. Corresponding PSNR and NIQE [40] are shown in parentheses.



Single Image Animation

Single Image Animation

04

SinGAN总结

Conclusion

1、提出了一个新颖的非条件生成网络架构，仅使用一张图片就可以实现多样化样本的生成

2、这个模型生成的图片不仅可以生成纹理不同的样本图片，还可以生成自然复杂的“真实”样本

3、SinGAN为图像处理任务提供了一个强有力的工具

4、https://www.youtube.com/watch?v=xk8bWLZk4DU&feature=youtu.be

谢谢聆听

演讲人：顾嘉辉