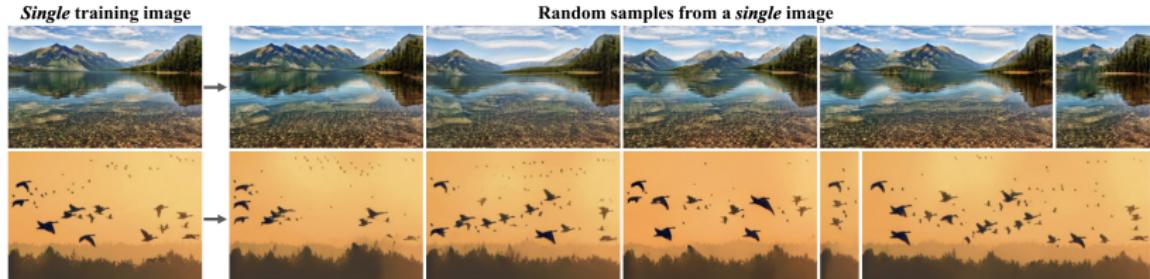


SinGAN: Learning a Generative Model from a Single Natural Image¹

April 9, 2020

¹ "SinGAN: Learning a Generative Model From a Single Natural Image" by Shaham, Dekel, and Michaeli

Overview

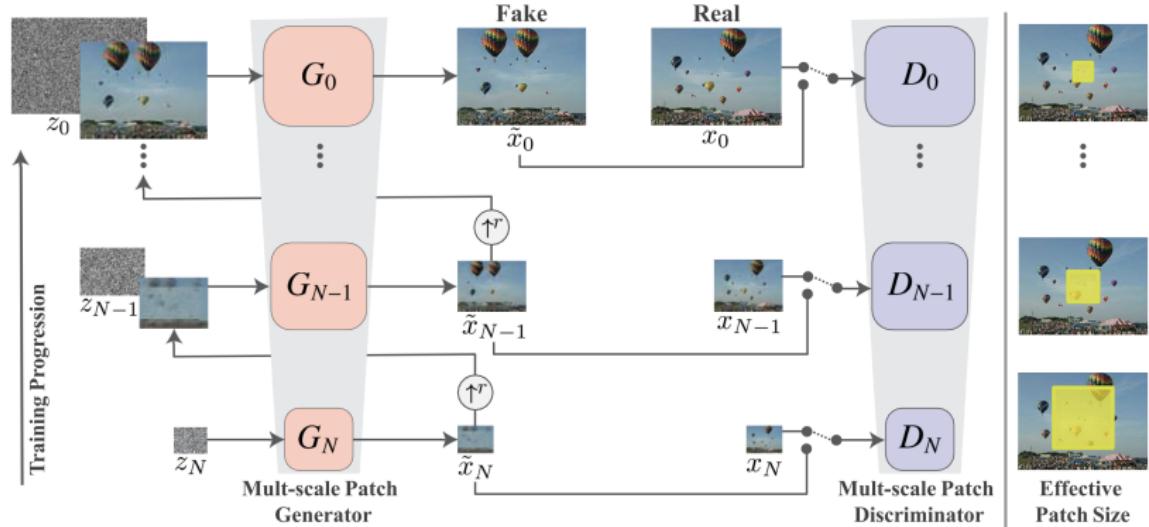


- ▶ **SinGAN** is a GAN model which is trained on a *single* image and can generate similar ones.
- ▶ It does so by using a Pyramid of GAN models, starting with a small resolution and gradually increasing it.
- ▶ Authors obtained *very* good results on other generative tasks: super-resolution, paint to image, editing, etc.

SinGAN in short

- ▶ Take an image and downsample it several times to obtain a sequence $x_0 > x_1 > \dots > x_N$ of its downsampled versions.
- ▶ Initialize N GANs $(G_0, D_0), \dots, (G_N, D_N)$, where (G_n, D_n) operate on image resolution of x_n .
- ▶ On each level n , generator takes an image \tilde{x}_{n-1} from the previous level and produces an image for the next level: $\tilde{x}_n = G_n(z_n, \tilde{x}_{n-1})$.
- ▶ In addition to an adversarial loss, we also use a reconstruction loss to reconstruct original image.

Architecture



- ▶ On each level, we take noise z_n and image from the previous level \tilde{x}_{n+1} and generate image $\tilde{x}_n = G_n(z_n, \tilde{x}_{n+1})$.
- ▶ Discriminator D_n works on patches instead of the full image scale.

Architectural details

- ▶ Given noise z_n and upsampled previous image \tilde{x}_{n+1}^\uparrow , generator G_n predicts only “missing details” which we should add to an image, i.e. we do not predict the image from scratch.
- ▶ To incorporate noise, we just sum it up: $z_n + \tilde{x}_{n+1}^\uparrow$ (this forces G_n not to discard it)

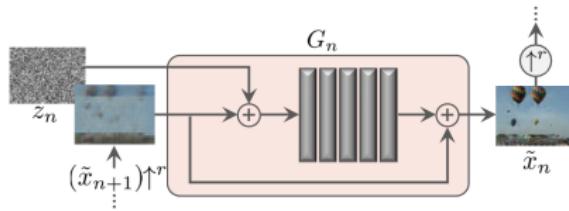


Figure: Generator architecture

- ▶ Discriminator D_n operates on patches of images instead of their full scales
- ▶ G_n, D_n are kept simple so not to overfit.

Training details

- ▶ The whole architecture is trained sequentially, i.e. once the (G_n, D_n) is trained, it is kept fixed.
- ▶ Adversarial loss:
 - ▶ Adversarial loss \mathcal{L}_{adv} is WGAN-GP loss
 - ▶ They used PatchGAN (or “markovian discriminator”), which predicts scores for image patches independently and the final score for an image is an average of individual patch scores.
- ▶ Reconstruction loss:
 - ▶ We want to ensure that our model can produce an original image for some noise vector.
 - ▶ Pick “golden” noise vectors $z_N^{\text{rec}}, z_{N-1}^{\text{rec}}, \dots, z_0^{\text{rec}}$ to be $z^*, 0, \dots, 0$, where z^* is random and fixed.
 - ▶ Force G_n to produce original (downsampled) image x_n from the given noise z_n^{rec} and previous reconstruction $\tilde{x}_{n+1}^{\text{rec}}$ by optimizing:

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

- ▶ Total loss:

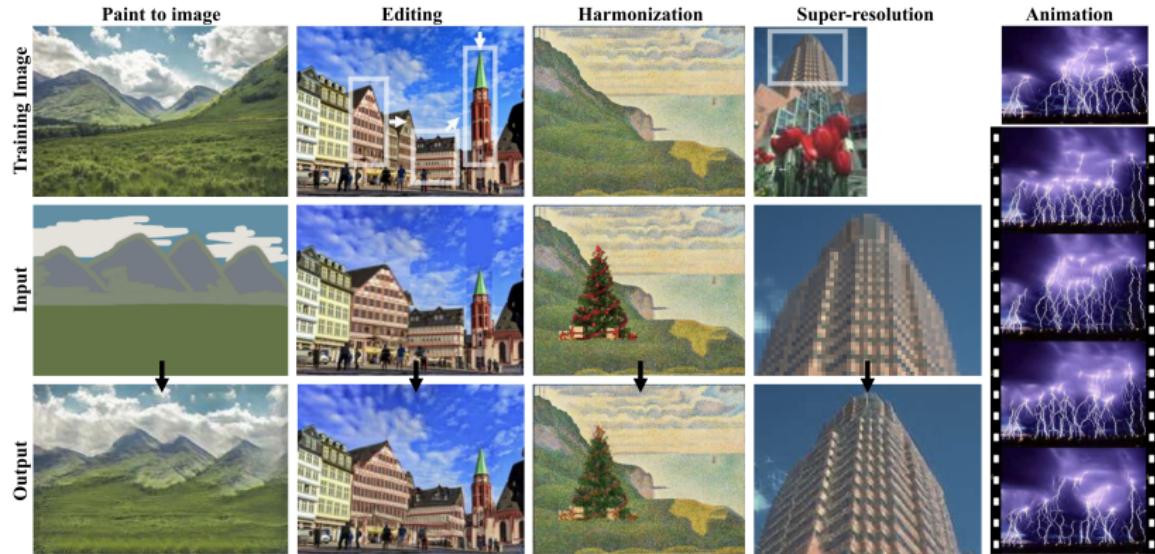
$$\min_{G_n} \max_{D_n} \mathcal{L}_{\text{adv}}(G_n, D_n) + \alpha \mathcal{L}_{\text{rec}}(G_n) \quad (2)$$

Image manipulation with SinGANs

SinGAN gives a straightforward way to do different image manipulations after training *without any additional tuning*:

- ▶ Super-resolution: just apply G_0 several times to a low-resolution image.
- ▶ Paint to image (create an image from a clip art): downscale clip-art image to the coarsest scale and feed it to G_N .
- ▶ Harmonization (blend a pasted object with a background): downscale the image to a factor $n \approx 3$ and feed it to G_n .
- ▶ Editing (replace image regions): same as for harmonization.

Image manipulation samples



Harmonization samples

Training Image



Input Image



Deep Paint. Harmonization



SinGAN (Ours)



Conclusion

- ▶ Good quantitative results on standard metrics and human evaluation
- ▶ A possibility to train a good GAN model from a single image gives a lot of opportunities.
- ▶ Previous works on single-image GANs were only about producing texture
- ▶ Good results on image manipulation
- ▶ Ablation study is missing
- ▶ How will we benefit from using several images?