High-Res-SinGAN
-SinGAN Resolution Improvements-

Sungwoo Son
DGIST
sungw7@dgist.ac.kr

Abdur Rehman
DGIST
abdurrehman@dgist.ac.kr

Rene Solzbacher
DGIST
solzbacher@dgist.ac.kr

Abstract

There is a growing interest to train deep learning models with minimum amount of data. Single image generative models take this idea to an extreme and learn from single images. SinGAN was the first single image GAN that can generate realistic images for tasks other than texture generation. In this paper we take a deep dive to understand the shortcomings of SinGAN and propose improvements based on our experimental results. We were able to generate higher resolution samples and managed to reduce training time by 43 percent.

1. Introduction

Generative Adversarial Networks (GANs) are capable of generating realistic images [1]. However most GANs are trained on large datasets which can be time-consuming and expensive. So, SinGAN [2] is useful if we want to obtain variation of a given image, work with a very specific image or style, or only have access to little training data. This SinGAN can be trained on a single image for tasks such as unconditional image generation, paint-to-image, editing [3, 4], harmonization, super-resolution [5] and animation.

SinGAN is trained in a multi-stage, multi-resolution approach. Training starts at a low resolution (e.g., 25x25 pixels) at the first stage. The training progresses through several stages, a pyramid of fully convolutional lightweight GANs, each is responsible for capturing the distribution of patches at a different scale of the image.

Here we introduce an amplified version of SinGAN. The previous paper can produce output images of up to 250px. But High-Res-SinGAN can generate output images of up to 500px. We propose two methods to perform this task:

1. SinGAN optimization
2. Image splicing

In SinGAN optimization, we optimized the receptive field size, the number of kernels and the number of scales. For image splicing, the source image is divided into four equal sections. Each section is trained separately to be later merged into a single high-resolution image.

2. Methodology

2.1. SinGAN Optimization

SinGAN limits random sample generation to 250px at its longest side. Most practical applications require samples of higher image dimensions, so we explored high resolution random sample generation. Initial testing showed that SinGAN does not perform well with high-resolution sample generation. Extensive hyperparameter changes are required to get realistic results. We concluded that the most important hyperparameters are:

1. Receptive field size:
   For smaller values, the model fails to generate meaningful structures, as it can only look at a small region. For higher values, there is no variation in global structure of image.
2. Number of kernels:
   Adds fine detail to the generated image. The higher the number, the better the result.
3. Number of scales:
   Lower scales affect the global structure, while higher scales affect the fine texture.

As receptive field size is important in generating high level structures and kernels add fine level detail, these two parameters need to be manipulated based on the output resolution training the GAN model.

We empirically found that receptive field size scales linearly as we increase image resolution. However, the number of kernels follows a nonlinear trend. When the image size is doubled, the receptive field size must be doubled as well, and the number of filters needs to be squared. It is to be noted that, as we increased the values for different hyperparameters, the memory requirements increase rapidly as SinGAN requires multiple stage of GANs to generate samples. Care should be taken while changing these hyperparameters.

High resolution samples can be found in appendix 1.

Figure 1: *Input image (b) with method 1 generated output (a) and method 2 generated output (c)*

2.2. Time Reduction

Furthermore, with higher resolution sample generation, the training time increases significantly. This limits the usefulness of random image generation. To decrease training time, we propose the following solution:

2.2.1 Epoch Decay

As higher scales contribute to textural details and have little effect on structural changes, we can reduce training time by decreasing their number of epochs.

SinGAN uses a pyramid structure of stacked GANs. The first four networks each use the previous scale's weights. The model architecture changes by increasing the number of channels after the fourth scale. It remains the same for all following scales. As a result, we are not able to reduce the number of epochs without negatively affecting the quality. By fixing the architecture, each scale can benefit from the pretrained weights of the previous scale. Thus, linear epoch decay can be introduced without compromising quality.

After each scale we reduce the total training iterations by 10% of the initialized amount. Using this approach, the training times were reduced by up to 43 percent almost halved the training time.

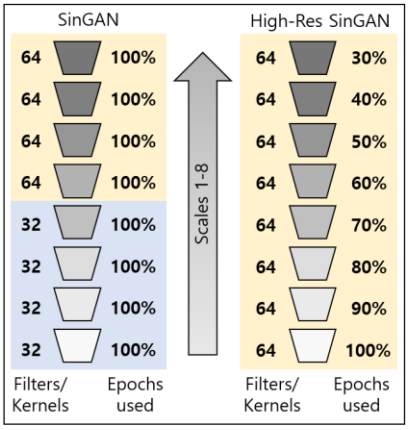

Figure 2: *Visualization of epoch decay. SinGAN on the left, epoch decay on the right.*

2.3. Image Slicing

The second proposed method for high-resolution image generation takes the input image and divides it into four. For each image, a separate model is trained, and random images are generated. Additionally, a base model is trained without any optimizations on the high-resolution input image. A new image is created by combining the generated slices and the editing function is used on the base model to stitch the sliced images.

The resulting stitch lines are not always satisfactorily removed, and some artifacts remain. We believe that traditional image stitching techniques would be better suited for the last step.

3. Conclusion

We introduced two methods of high-resolution image generation using High-Res-SinGAN. We can generate image dimensions up to 500 pixels. Comparing both methods, such as in figure 1, it can be observed that the image splicing method leaves artifacts at the image seams (c). These seams are unique to image splicing and do not occur in method 1. The overall quality of the generated image of method 1 (a) is better than that of method 2 (c).

High-Res-SinGAN is capable of generating high resolution images based on the SinGAN model.

References

[1] Yuki M Asano, Christian Rupprecht, and Andrea Vedaldi. A critical analysis of self-supervision, or what we can learn from a single image. In International Conference on Learning Representations, 2020.

[2] Tamar Rott Shaham, Tali Dekel, and Tomer Michaeli. Singan: Learning a generative model from a single natural image. In Proceedings of the IEEE International Conference on Computer Vision, pages 4570–4580, 2019.

[3] Roey Mechrez, Eli Shechtman, and Lihi Zelnik-Manor. Saliency driven image manipulation. In 2018 IEEE Winter Conference on Applications of Computer Vision (WACV), pages 1368–1376. IEEE, 2018.

[4] Kaiming He and Jian Sun. Statistics of patch offsets for image completion. In European Conference on Computer Vision, pages 16–29. Springer, 2012.

[5] Daniel Glasner, Shai Bagon, and Michal Irani. Super resolution from a single image. In 2009 IEEE 12th International Conference on Computer Vision (ICCV), pages 349– 356. IEEE, 2009.

4. Appendix 1

Input image (900 x 600)

Random output samples (500 x 334)

Input image (900 x 520)

Random output samples (400 x 232)