Introduction to Computer Visualization

SinGAN: Learning a Generative Model from a Single Natural Image

Marcin Hamerlik   
*Faculty of Electronics*  
*Wrocław University of Science and Technology*   
Wrocław, Poland  
[209854@student.pwr.edu.pl](mailto:209854@student.pwr.edu.pl)

Osman Velat Kabak  
*Faculty of Electronics*  
*Wrocław University of Science and Technology*   
Wrocław, Poland  
[248673@student.pwr.edu.pl](mailto:248673@student.pwr.edu.pl)

# Introduction

In the recent years the application of programming solutions and algorithms to images has been a growing trend. Thanks to new solutions it is possible to utilize the computing power of machines to learn and recognize images and apply them to a variety of tasks and business venues. However, most of those solutions come with limitations or costs that impede their usage – such as processing time and need of multiple datasets to properly learn the model required for processing a specific image, thus lengthening the time spent on a given task or creating a highly-specialized solution that may not be applicable anywhere else.

The SinGAN is a new proposed model that attempts to overcome these obstacles via a model generated from just a single image. An extension of the Generative Adversarial Nets (GANs) technology, it uses a pyramid of conventional GANs to analyze the single image and gain training data not just from objects of interest but also “noise”. Additionally, SinGAN offers not only data recognition but also dynamic image generation and manipulation using the same model.

This paper will attempt to analyze and test the functionality and applicability of SinGAN in order to determine its potential real-life applications both by studying it’s open-source code and documentation but also through practical testing of its capabilities on varied samples. The end result will also be compared with other options currently available to properly gauge how SinGAN fits in the business and private sector of image processing.

# Related Works

A number of articles and publications has been gathered to achieve the goals of this article and offer a better understanding of not only the technology behind SinGAN but also image processing in general and its current real-world applications.

**SinGAN: Learning a Generative Model from a Single Natural Image [1]** is the main article describing the idea behind SinGAN and its specifications. The article focuses both on the motivations and inspirations behind the model but also the mechanics behind it’s unique approach – complete with diagrams, mathematical background and data. Furthermore if offers its own exhaustive testing of various capabilities and comparison of their results as well as it’s mechanical advantages and disadvantages. For anyone interested in the SinGAN technology, this is the go-to source of easily-understandable information.

**Image Generation with GANs-based Techniques: A Survey [2]** offers an extended summary of the currently-available image generation techniques and how they work. This will be crucial in determining the image-generation functionality of SinGAN and can serve as a base for experiment techniques, result analysis, and applicability criteria. Notably, it does contain information on SinGAN and its category of image generation which can be used for this article.

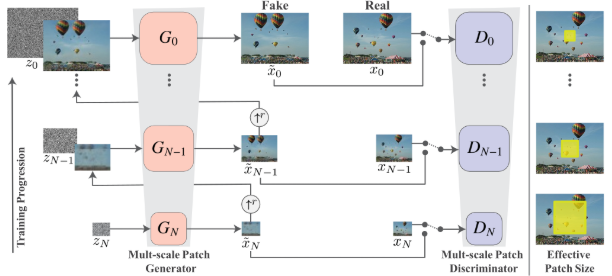
**Generative Adversarial Nets [3]** has been selected as the main source of information on GANs and how they work. This will be crucial to understanding how SinGAN works and determining whether it’s application of technology furthers the technique itself. Additionally it is a good summary of the advantages and disadvantages of the technology itself which further lets us analyze how SinGAN performs.

**Improved Techniques for Single-Image GANs [4]** covers the methodology and implementation of Single-Image GANs. This publication offers the most recent (as of March 2020) summary of the single-image training technology and will be used as a basis for determining how well SinGAN performs in compared to other implementations of the technique. Additionally, it contains a massive base of testing samples and how different solutions perform which will make the testing phase of this work much faster.

**Program-Guided Image Manipulation [5]** concerns the area of dynamic, algorithm-based image manipulation, chosen to be the main source of information needed to accurately gauge the capabilities of SinGAN in this area. The publication contains code samples, diagrams and examples of various techniques used for program-guided image manipulation, including GAN-based models and the techniques used by the SinGAN model itself.

# Technology

Structure of SinGan is made up of multi-scale pipeline, consisting of a generator and a discriminator pair that it uses for multi-scale representation learning. The lower the scale of the before mentioned pair, the less detailed the are learned ( i.e. in the lowest scale only background is processed but in the highest scale finer details like the wings of a bird are processed.



Structure of SinGAN

When an image is being processed, it gets reduced to different sizes according to the pairs it will be sent to, including the result of the generated image from the pair below, with the lowest pair only receiving the reduced image.

The generated image received by the generator part of a pair is made up by all the pairs below that one, put together. And when that specific pair outputs its generated image, it does so by changing the one it received with the one it produced, so the cumulative process goes on.

The process ran by the generator can be explained like so : **b.noise + g.image -> conv + g.image -> result.**

Which means, the background noise specifically made for that pair, put together with the pair’s own reduced image, processed with the data sent by the lower pairs to achieve a result.

The whole network is designed to work sequentially, starting from the lowest pair going up to the highest pair. When a certain GAN is done, it is kept and so on.

The loss functions used are for adversarial and reconstruction. Adversarial loss function is used to match the distribution of samples while the reconstruction loss function is used to generate images that look similar to the original one, which is a requirement.

# Experiment Design

To truly test the performance of SinGAN, an experiment was designed that will measure how well it’s features work. For each feature a source image will be taken and tested then compared with the original to see how well the given function was performed.

Due to the limited scope of this article, the testing of image generation realism will be measured by itself and by providing the resulting images without context to small groups of people to measure their response in how believable they are. Since generated images need to have a basis in reality, the testers will also be provided the source image and then asked to compare them and point out any specific differences that would make the image stand out as computer-generated.

The measure of image editing is much harder to perform and must be judged by comparing the result within the context of the intent and how close it came to realizing the function that the user wanted to perform. In order to get a more quantifiable measure of editing, an image differencing software will be used. We have chosen an open-source, free program ImageMagic due to its ease of use and clearly readable output.



Fig. Results of Image Differencing by using ImageMagic

Additionally, SinGAN provides its own measure of differentiation between real and fake images using the Single Image Fr´echet Inception Distance (SIFID) technique, a measure of evaluation GANs that relies on calculating the distance between feature vectors for real and simulated images. Using this option will not only allow us to get an additional set of data on the generated/edited image but also see how well this function of SinGAN performs.

# Results

The first noticeable thing about SinGAN is the long time needed to train a model. This is of course the cost of using a single image as a source. Training a model on a mid-end PC until the 8th scale takes about two hours per image, thus limiting the personal usage of SinGAN – it is more suited to industrial-strength computers.

## Image Generation

To test the capabilities of SinGAN image generation we have picked a source image that can be easily interpreted by anyone and serve as a good base for generation – a view of the Sudety mountain chain.



Fig. Input image for generation

This picture was then trained in SinGAN using the main training process then ran with the random sample generation function on scales 1 and 8 to test how well they perform and how much time it’ll take on both ends of the spectrum. The images were then presented to a number of people to test how believable the result is.



Fig. Output of Image generation at scale = 1



Fig. Output of Image Generation at scale = 8

In the end the resulting images are quite believable compared to the original but do not display much other changes. The problem comes from the lack of resolution that causes the image to have low quality and therefore cannot be used in many scenarios. We have provided both samples to a group of n=15 people to get their input on the matter. As seen in **Table 1**, most of the results are believable with the bigger scale providing a more realistic output.

|  |  |  |  |
| --- | --- | --- | --- |
| Scale | Response | | |
| Real | Generated | Don’t know |
| Scale = 1 | 73% | 26% | 0% |
| Scale = 5 | 86% | 7% | 7% |

Table Believability of the output image

It is hard to judge just how accurate the generation process is by empiric data. SinGAN offers the SIFID judgement which we have used on the real and fake image and gathered the end results.

|  |  |  |  |
| --- | --- | --- | --- |
| Scale | SIFID | Survey | SIFID/AMT Correlation |
| Scale = 1 | 0.02 | Paired  Unpaired | -0.22  -0.05 |
| Scale = 5 | 0.01 | Paired  Unpaired | -0.19  -0.01 |

Table SIFID results

As with the previous results we can see that both images are fairly close to the source image and can be seen as quite believable due to the small SIFID result – which implies that in small-scale images the image generation provides satisfactory output for this kind of image.

## Image Editing

In order to test the capabilities of the image editing software we have taken a more complex image – a view of the street of the Wrocław main square. The image was then edited to change the proportions of the building and the changes were noted on the edit mask. The edited image was then treated with SinGAN on both ends of the scale (1 and 8) to test the capabilities of the program.



Fig. Input image for editing



Fig. Naively edited image



Fig. Naive edit mask



Fig. Output of SinGAN editing with scale = 1

Fig. Resolution upscaling result



Fig. Output of SinGAN editing with scale = 8



Fig. Image differencing between the original image and SinGAN scale 8 edit (left) and naive edit(right)

As one can see while the scale 1 edit is blurry and doesn’t look realistic, the scale 8 edit is very close to reality and fulfils the desired edit parameters. Looking at the image (**Fig. 10**) differencing results created by applying ImageMagic, the original naïve edit has clear-cut differences that make it appear clearly edited compared to the original picture, while the SinGAN treated image outputs a more fuzzy result. This actually works to its advantage as usage of the trained model allows SinGAN to output a more believable result by adjusting the edited parts to fit the model.

The main disadvantagof this process is the low output resolution of the edited file – this of course means the process takes little time but ends up having quite low quality and may be unusable for anything else but image differencing applications as-is. 

The other problem is the need of mask files that highlight the edited areas (**Fig.7**) – this is a tedious process that needs to be done by hand to point out the edited areas to SinGAN, which limits how quickly this process can be applied to multiple images with multiple edits.

## Resolution upscaling

SinGAN comes with a resolution upscaling tool that can be used to improve the quality (and therefore usability beyond testing) of its own output images along with any other image. This however requires SinGAN to learn the model of the pre-upscaled image just as it did with previous images, making it last a lot longer than just the simple upscale function itself.

As we can see on **Fig.11** the result does indeed come in higher resolution but much of the detail is lost compared to the source image and therefore the function remains usable only to upscale certain parts that do not have many complex details. This means that it can be applied in certain sectors of image detection and processing where a higher resolution is needed to detect certain data but the level of detail is of no or minimal concern.

# Conclusions

SinGAN is a tool whose advantages serve as disadvantages at the same time – learning from a single image does indeed make the learning process quite easy but it suffers from long training times and reduced quality of output. Whether the positives outweigh the negatives depends on the context of the model itself and the intended result the user wants to receive.

Given the ease of access however, it still retains many applications in the business sector as learning from a single image saves the time needed to gather different images for the training process and can actually mean a single instance of an object can be used for image generation which fits certain technical tasks in the engineering sector.

Overall, even with its many downside SinGAN is a stellar tool that manages to fill its niche – a self-learning image processing tool that requires only one source image – quite well and provides a breadth of functions. It’s application is limited by an unintuitive interface but these disadvantages can be accepted due to the free nature of the program. There has been attempts to improve the SinGAN technology which gives hope that single-image learning can be improved to be more reliable.

# References

Shaham, Tamar Rott et al. “*SinGAN: Learning a Generative Model From a Single Natural Image*.” 2019 IEEE/CVF International Conference on Computer Vision (ICCV) (2019): 4569-4579.

1. Esfahani, Shirin Nasr and Shahram Latifi. “*Image Generation with GANs-based Techniques: A Survey*.” International Journal of Computer Science and Information Technology 11 (2019): 33-50.
2. Goodfellow, Ian J. et al. “*Generative Adversarial Nets*.” NIPS (2014).
3. Hinz, Tobias et al. “*Improved Techniques for Training Single-Image GANs*.” ArXiv abs/2003.11512 (2020): n. pag.
4. Mao, Jiayuan et al. “*Program-Guided Image Manipulators*.” 2019 IEEE/CVF International Conference on Computer Vision (ICCV) (2019): 4029-4038.