Report - SinGAN for transferring drawings to photo-realistic images

Philippe Ganshof
ENS Paris-Saclay
92220 Cachan
philippe.ganshof@hotmail.com

Abstract

We use SinGAN, an unconditional generative model introduced by T.R.Shaham and al in [1], for transferring drawings to photo-realistic images. We trained SinGAN on several pictures of buildings. For each training example, we designed one or more paintings in order to test the ability of SinGAN to reproduce realistic images. We demonstrated that SinGAN preserves the main structure of the drawing but often fails to recreate all details in a realistic way unless the painting is an accurate representation of the input. SinGAN has also been compared to another well known model, the Neural Style transfer method. Two quantitative tools have been used to distinguish them and we showed that the outputs share great similarities.

1. Introduction

SinGan is an unconditional generative model introduced by T.R.Shaham and al in [1]. As explained in the paper, the model is trained on a single training example and captures the internal distribution of patches, within the image. The authors propose several applications such as super-resolution or harmonisation. We focus our attention on the *paint-to-image* application. Three different landscape images are presented in the paper to show the efficiency of SinGAN in this specific task. The model shows great results but it is fairly easy to say that the contrast and the structure play a less important role in the design of a realistic landscape than a building for instance.

In this project, we extended the application of SinGAN by using images of buildings as the training example. Two types of images were considered, one with the building at the forefront and the other in the background of a landscape. In both cases, we tested the ability of the model to reproduce the geometry of the house as well as all its decorations. We compared our results with another well-known model, the Neural Style Transfer method taken from this article [5]. A deep neural network pre-trained on ImageNet, classifying our generated samples, was used as a quantitative tool to compare both methods. We also created a survey reporting the confusion rate between our outputs.

In a first time, we introduce SinGAN and its particularities. Our results are then commented and we conclude by recalling the project highlights.

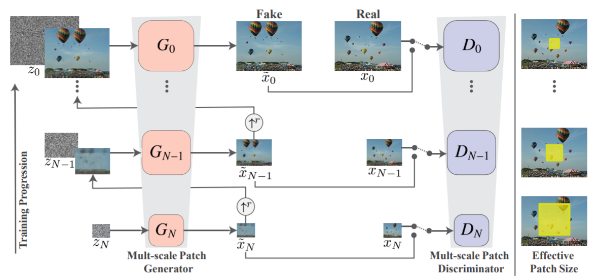

Figure 1. **SinGAN's framework**. Illustration taken from [1]

2. SinGAN

Unlike other generative models, SinGAN proposes a new training method for generating images from random noise by using a single image. Instead of modelling the generator's distribution over many data, the different patches within the image are used. It certainly reduces the diversity of the outcomes produced by the network but allows style transfer from a single target image. SinGAN also differs from the well-known style transfer methods as they are usually limited to texture images.

We can see in figure 1 that the model follows a pyramid structure. At each level, a GAN takes as input a random noise and a different scale of the training patch, downsampled with factor r^n. For our experiments, we set r to 4/3 as recommended in the paper [1]. In order to get a photo-realistic image from a drawing, one feeds the painting at one of the coarse scale N or $N-1$ replacing $\tilde{x}_N$ or $\tilde{x}_{N-1}$ respectively from figure 1. This means that the painting is supposed to represent the coarse features of the training example. At each level the generation process will add finer details till the level with the original image size is reached.

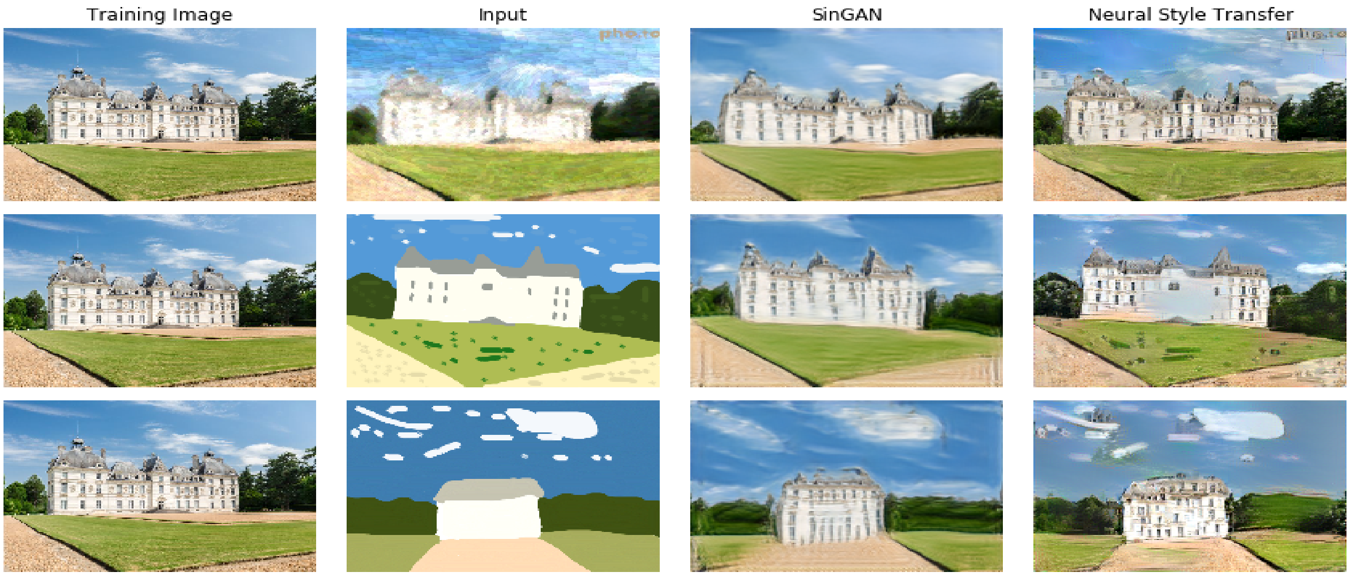

Figure 2. **Paint-to-image**. We train SinGAN and the Neural Style Transfer method on an image of a palace and test them on three different paintings. From top to bottom, we start from an accurate representation (painting 1) to arrive at a row approximation (painting 3) of the training example.

3. Results

We trained with both SinGAN and the Neural Style Transfer method on a large variety of building images taken from the website *Unspalshed*, see the link [2]. Paintings were made using *YouIDdraw* (e.g., painting 1 and 2 from fig 2) and also *pho.to* (e.g., painting 3 from fig 2) [3], [6].

We investigated in a first time what kind of drawings could lead to realistic outputs. Figure 2 shows that the general features of the original image should be preserved in the painting if one were to hope to get interesting results. In the case where the house is not at the forefront, finer points are treated as texture by SinGAN and the model often fails to reproduce all details attached to the building in a realistic way. We can see this in figure 2 where the presence or absence of windows in the painting does not affect the output. SinGAN is also sometimes confused when two objects in the training image are similar in terms of colour and shape and reproduces the wrong item. Indeed, we observe in figure 2 that SinGAN confuses the different roof shapes. The Neural Style Transfer method shares similar problems and our opinion suggests that SinGAN tends to produce smoother outputs which makes them more realistic. However, the difference is not clear and we compare them using two different quantitative tools explained in the next section.

Despite these common mistakes, a drawing representing well the original image often leads to a realistic output as shown in figure 3. We then explored how both methods were behaving when modifying these drawings. In contrast with the Neural Style Transfer method, adding or removing decorations on the house(e.g., windows, statues) from the painting is successfully handled by SinGAN. An example is shown in figure 3 where we see that the window is replaced by synthetic texture in a realistic way. However, both models failed to preserve more complex modifications such as changing the shape of an object for instance.

Model	Input	Prediction	Probabilities
	Training example	Palace	90.4%
SinGAN	Painting 1	Castle	88.9%
	Painting 2	Castle	70.7%
	Painting 3	Castle	49.58%
Neural style transfer	Painting 1	Castle	48.6%
	Painting 2	Castle	57.3%
	Painting 3	Castle	80%

Table 1. **Image classification.** We used our outputs from figure 2 as inputs of the pre-trained network ResNext50.

3.1. Quantitative Results

In order to quantify the realism of our outputs and compare both models, we used two different metrics: (i) a user study in the form of a survey, and (ii) an image classifier [7].

Poll We presented to twenty people of our surroundings the outputs produced by both models shown in figure 4. For each training example, we ask them to pick the outcome that looked the most real. Then people were asked

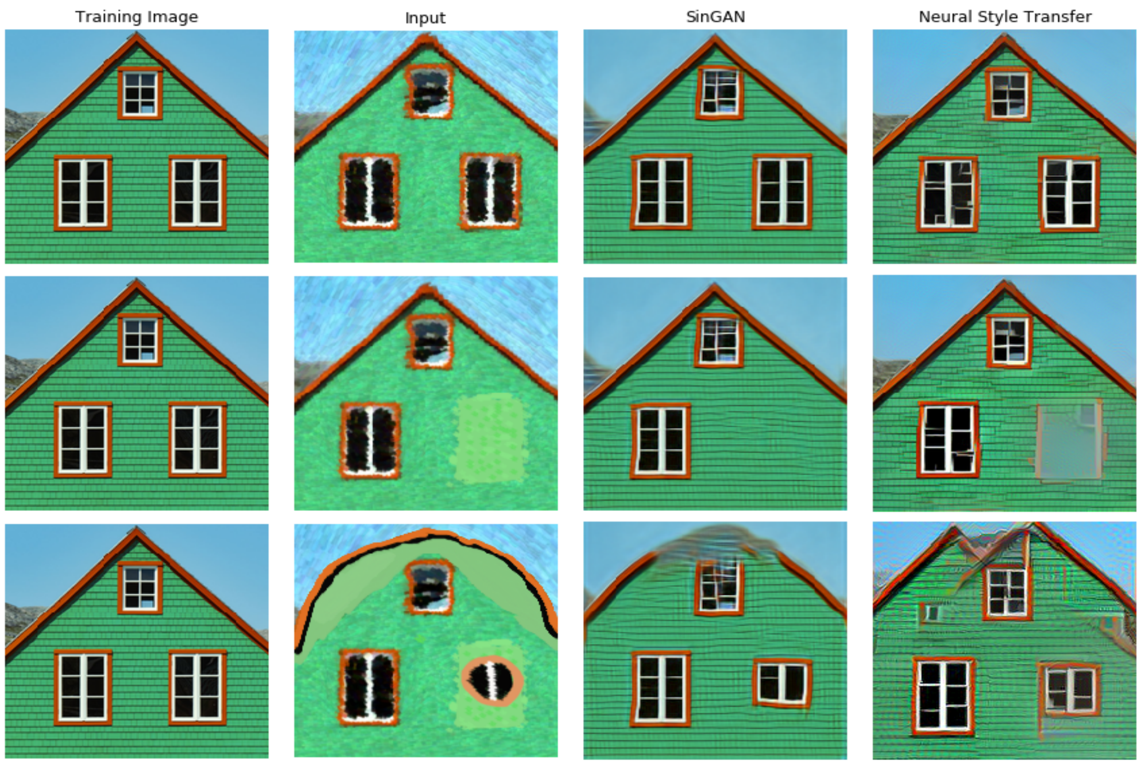

Figure 3. **Paint-to-image**. We train both SinGAN and the Neural Style Transfer method on an image of a house at the forefront and we test them on three different paintings. The first one is an accurate approximation of the original image and we modified it to obtain the two other drawings.

Number of misclassified images	0	1	2	3	4
Ratio	15%	60%	25%	0%	0%

Table 2. **Confusion rate.**

Picture number	SinGAN	Neural Style Transfer
1	0%	100%
2	35%	65%
3	0%	100%
4	80%	20%
5	50%	50%

Table 3. **Poll.** Columns 2 and 3 report the ratio of people who have chosen the output of the model for each image. Pictures are numbered from top to bottom.

to separate the images into two groups, one consisting of the images produced by SinGAN and the other by the Neural Style Transfer method. This way, we not only compare the realism of our outputs but also quantify how similar they look.

Surprisingly, Neural Style Transfer outputs were chosen more often on average. This is mainly due to images number 1 and 3 because in both cases, SinGAN unlike the other model fails to reproduce the structure of the building in some areas. For pictures 1 and 3, all people thought Neural Style Transfer outputs looked more realistic. People were more indecisive about the other images, especially image 4 and 85% of the people misclassified at least one image. This leads us to the conclusion that none of the methods really stand out and that there is a lot of variance in the outputs.

Image classifier Next, we use an image classifier ResNext50 pre-trained on ImageNet to quantify how realistic the outputs from figure 2 look.

As expected, the quality of the painting and how well it represents the original image plays an important role in the success of the outcome. Higher confidence probabilities are obtained with SinGAN in general which may lead us to believe that it performs better. However, in opposition to our opinion, the neural network gives high probability to the last output reducing the reliability of this quantitative measure. We also note that the predicted label of the training example differs from the other images. We believe that this is due to the mistake mentioned earlier, that the round shapes of the roof are replaced by triangular shapes which makes the picture look like a castle.

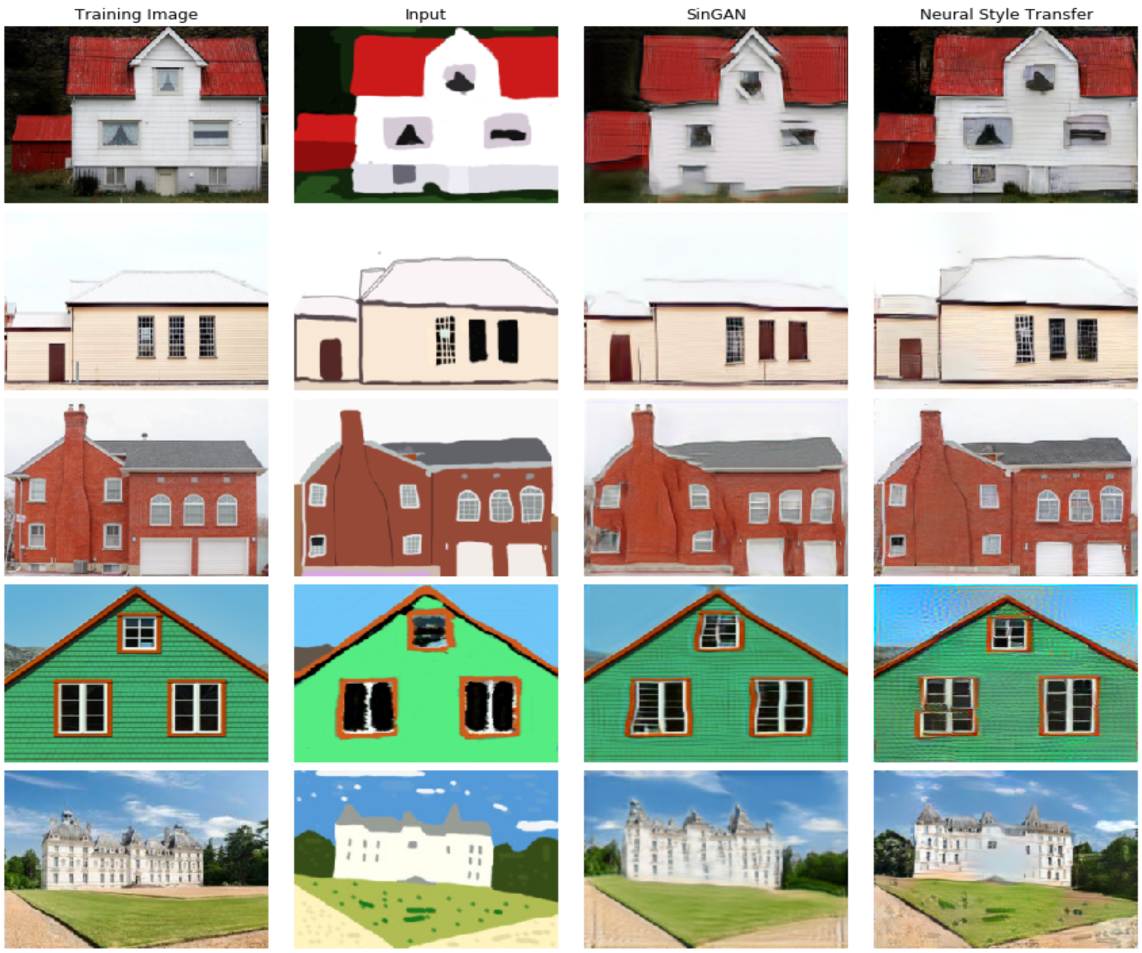

Figure 4. **Paint-to-image**. We train SinGan and the Neural Style Transfer method on five images selected randomly and we test them using paintings made by us.

4. Conclusion

Our analysis shows that with SinGAN, the drawings leading to photo-realistic images are limited. The quality of the outputs varies widely depending on the image and is also very dependent on the drawing. We also demonstrated through two experiments that SinGAN does not necessarily give better results than the Neural Style Transfer method and that both models often produce similar outputs.

References

[1] T.R.Shaham, T.Dekel, T.Michaeli, SinGan:Learning a Generative Model from a Single Natural Image, LarXiv:1905.01164v2 [cs.CV] 4 Sep 2019

[2] Unsplahed, https://unsplash.com/s/photos/modern-architecture

[3] Youidraw Painter, https://www.youidraw.com/apps/painter/

[4] Implementation of SinGan, https://github.com/tamarott/SinGAN

[5] Leon A Gatys, Alexander S Ecker, and Matthias Bethge, Image style transfer using convolutional neural networks. *In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pages* 2414-2423, 2016. 7, 8.

[6] Pho.to, http://funny.pho.to/fr/antique-oil-painting/

[7] S.Xie, R.Girshick, P.Dollar, Z.Tu, K.He, Aggregated Residual Transformations for Deep Neural , arXiv:1611.05431

[8] Implementation of Neural Style Transfer https://github.com/cysmith/neural-style-tf