Study of SINGAN: Learning a Generative Model from a Single Natural Image

Project report for MVA-RecVis 2019-2020

KILANI Al Houceine, NAOUMI Salmane

***Abstract*—**
In this report, we are studying the SinGAN, an unconditional generative model that can be learned from a single natural image. This model can be used for multiple goals such as : Super-resolution, Harmonization, Editing.. (all presented in the paper). Our goal in this -short- study is to explain briefly its functioning and see how it can be extended/adapted -if possible- to be used for Denoising/Deblurring an image.

I. Principle - Architecture

A. The Architecture

This model takes a multi-scale progressively growing architecture similar to generative adversarial networks. The idea is that the generator starts off with a low resolution image (at the first same it is a random Gaussian sample) to be generated and it is compared against the original image that has been down-sampled to the same resolution. This is an interesting discriminator model in that it is not looking at the entire image but it is just looking at patches of the image (Markovian discriminator) to make the classification real/fake and then aggregate the decisions. This is done to avoid the generator generating exactly the same target image.
As we are going up in the pyramid, residual learning is performed : when we want to go to an up-sampled image, we take the previously generated image and propagate it ahead such that it is just mapping the Gaussian noise into a new image that is going to be added to the up-sampled previously generated image thus creating the next scale generated image. Then the discriminator is going to be looking for real/fake classification in the patches of this image and of the original one down-sampled to that same scale.

B. The training and output control

Rather than using a cross entropy loss on these predicted patches in the discriminator, the loss adopted was a Wesserstein Gann with a Gradient penalty forcing Lipschitz constraint. There was another loss ”The reconstruction loss” added so that the generator can still reconstruct the training image even given this multi scale process (starting with a Gaussian noise and then only inputing the upsampled generated image of the previous level at each level)

The control that we have over the variability of the generated images is performed by controlling which scale on the multi scale hierarchy we are choosing to ”vary” the image at. Concretely, we choose at what level we want to start the generation (random Gaussian at the bottom or intermediate generated down-sampled image). If we start at the bottom, we intend to have an image which varies a lot compared to the original one but have to same pixel distribution (example of the of the zebra with 6 legs in the paper) but as we go higher in the pyramid, we only change slight differences in the image (strides of the zebra in the same example).

Another way of controlling the output images is the scale in which we choose to down-sample the images that will give us the scale at which we are going to up-sample the generated image at each step. If more scales are used, we get a much more overall coherent image rather than just random spacial patches having the same texture as the training picture (example of the Van Gogh painting in the paper).

C. Sanity check : colusseum.png

We cloned the github repository of the SinGAN official implementation. In order to perform a sanity check, we tried to reproduce some results.
We trained the different Generators and Discriminators on the ”colusseum.png” image (Figure 1). Since after each level of training, the weights are fixed, we use the trained layers (the current plus the previous ones) to generate a sample from a white noise z. The figures below, show the different images generated after training each scale.

Fig. 1. Original picture used to train the Generators and Discriminators

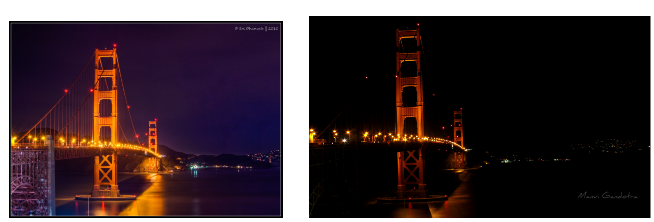

Fig. 2. Good image and darker (noisy) image used

We can observe that the coarsest levels generate general templates and shapes of the image (the Colusseum, the ground

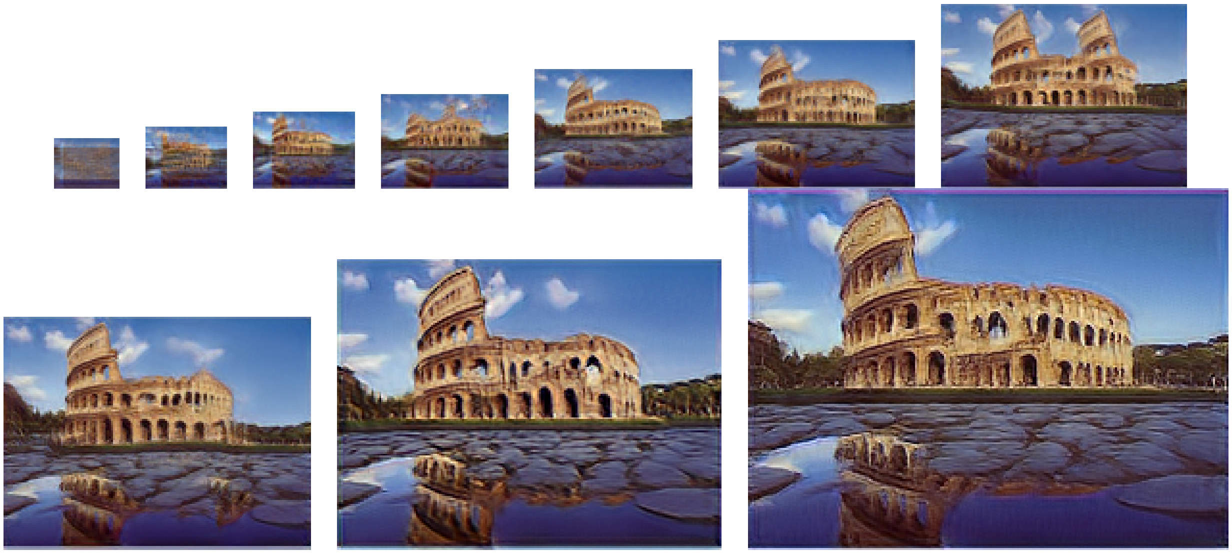

Fig. 3. 10 levels generation - The generation process at level n involves all generators $G_N, ..., G_n$ and all noise maps z_N,..., z_n up to this level

and the sky) and then it and adjusts more details in further levels. But for this exact case, we see that it suffers from a problem of naive duplication of templates (at the 6th scale), eventhough it corrected the image when it moved to the next levels.

II. Application on Denoising:

In this section, we will try to see if SinGAN can have great performance for image Denoising. Indeed we can suppose that if the models learns to generate templates from a learned image, it can manage to "paste" those templates on another noisy image.

In our case, we will use a picture of the Golden Gate Bridge [1]. (Figure 2)

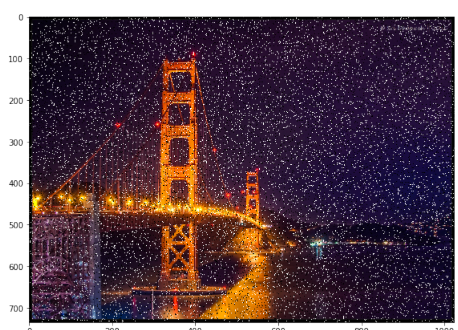

Fig. 4. Noisy : Salt and pepper noise added picture

A. Case 1: Train on good picture - Treat bad picture

We began by training on the original image and learning its underlying patch distribution keeping the same default parameters of $r = 0.8$ and $\alpha = 10$. We proceed to test the model on a denoising task on an artificially noisy image. (Figure 4). In the end we got a 10 levels Pyramid of generators. We also set the noise Z_0 equal to 0 as there is no need for it.

This task is kind of similar to the Super-Resolution task presented in the original paper. Since we want to replace the pixels that have been perturbed and keep the overall structure of the forms displayed in the image, we injected the noisy picture at a higher scale of the pyramid : the 10th scale. Then we keep re injecting the result for a number of times. Here were the results we got in (Figure 5)

Fig. 5. 3 successive injections in the G_0

We see that the image obtained by this process is not in a great quality. Indeed, It blurs and darkens the image even though it manages decrease the noise.

Trying to denoise the darker image (Figure 4 right) by training on the good image (Figure 4 left) yield to the results shown Figure 6 : an even darker image.

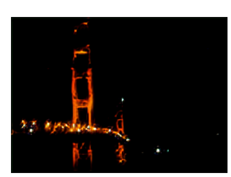

Fig. 6. Denoising darker image

Another attempt was to use the Super resolution mode of the SinGAN to denoise it. We obtained better results as shown in Figure 8. This is why we kept on using SR for denoising.

Fig. 7. Denoising using super resolution

B. Case 2: Train on bad picture - Treat same bad picture

Fig. 8. Output of SR SinGAN trained on Figure 4

Visually we get to see that there is a slight increase in the quality of the image but that the result is still highly improvable.

C. Comparison and Metrics

In order to quantify the quality of the denoising, we implement PSNR and SSIM of the pair of images. We chose PSNR because it is an approximation to human perception of reconstruction quality. As for the Structural Similarity Index (SSIM) that is a perceptual metric that quantifies image quality degradation which is , unlike PSNR, based on visible structures in the image.

In fact, quantifying the quality of denoising is an ill-posed problem since even clean patches of images may contain some noise at its original scale which makes the use of those metrics controversial.

First, we measure these metrics for the original image and the noisy one and then between the original one and the denoised ones. If there is an increase in PSNR, that means that the method increased the image quality. An SSIM of 1 indicates perfect structural similarity.

Metric	noisy image	Proposed method	SR case 1	SR case 2
PSNR	12.15	12.51	22.28	13.47
SSIM	0.147	0.126	0.794	0.169

Comparison with the original image

III. Discussion and Summary

We saw that the approach we proposed (re-injecting the image at the higher scale over and over again) did not give good results on a denoising task. We could not explain why it darkens the image (We see that the PSNR does not increase by a lot and the SSIM is even lower) nor how to solve the technical problem of having an output having a smaller size than the input. This is why we moved to using the super resolution and began to see satisfying outputs. Here is the summary of what we got :

- Training SR on good image (Figure 2 left) and treating noisy (Figure 4) : Very good results as shown in the improvement of PSNR and SSIM.
- Training on noisy image(Figure 4) ad treating the same image : We see that the improvement is not that great (slight inscrease in PSNR as well as SSIM).

We also saw that taking a whole new image and denoising it using our method did not give good results as it is merging 2 problems : being noisy as well as having a change in the viewpoint of the represented object Figure 6.

Some other ideas we got and did not have time to test are :

- As there is a phenomena for noisy images having their image patches benefit significantly from going down in scale (Indeed these patches will have a very good representative clean patch directly below them in a coarser scale of the noisy image). We thought that this directional pyramidal structure of SinGAN can strengthen this behaviour and serve deleting the noise. Concretely, we take the noisy image, downscale it and then pass it through an intermediate level of SinGAN to get the denoised output.
- An other solution could be applying super resolution to small parts of the image (for instance subdivide an image using a grid an apply super resolution to each of the grids) and then merge them all together to have to constitute a denoised overall image.
- Perform some standard denoising methods first, then apply SinGAN (using SR, downscaling then passing them through some levels of the generators, retry the initially proposed method)