SinGAN for Inpainting
Final Project - Object Recognition and Computer Vision

Julia Linhart and Roman Castagné
ENPC - ENS Paris-Saclay
Master MVA 2020
julia.linhart@eleves.enpc.fr roman.castagne@gmail.com

Abstract

SinGAN [Shaham et al., 2019] is a special kind of Generative Adversarial Network trained to model the internal distribution of patches within a single natural image. It allows to solve several types of image manipulation tasks, such as Super-resolution, Harmonization or Editing. The focus of this project will be its use to perform Inpainting, where one wishes to fill-in missing parts using statistical information of the rest of the image. After a quick introduction to the SinGAN architecture, we will explain our proposed extension with two main contributions: "smart" hole initialization and the use of Partial Convolutions introduced by [Liu et al., 2018].

1. Image Manipulation with SinGAN

Generative Adversarial Networks (GANs) constitute one of the most succesful subfields in deep learning, capable of generating realistic image data. But this ability of learning high dimensional distributions of visual data often requires large, class specific datasets or conditionning the generation on an other input signal. SinGAN [Shaham et al., 2019] on the contrary is a model that relies and is trained only on a single image.

The SinGAN architecture takes the form of a multi-scale pyramid of fully convolutional GANs, that each learn the internal distribution of patches within the given image at a different scale (cf. Figure 1). For the adversarial training of every GAN at each scale, the generation of a sample follows a coarse to fine method: starting with a gaussian noise, the information of the coarser level is used as a prior to train the finer scale (i.e. adding the missing details).

This rich multi-scale information makes SinGANs unconditional (i.e. no additional information or training needed) and not task specific (cf. Figure 3), allowing us to deal with general natural images containing complex structures and textures.

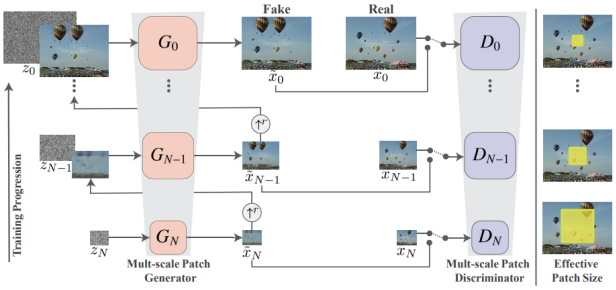


Figure 1: SinGAN multi-scale architecture of fully convolutional GANs [Shaham et al., 2019].

2. SinGAN for Inpainting

This project focuses on the task of image inpainting, where we consider two settings: object removal (Figure 2a) and image reconstruction of damaged images (Figure 2b). In both cases the objective is clear: fill in missing parts of an image using statistical information of the *rest of the image*.

2.1. Problem Definition

The general idea is to fill-in the holes with a well chosen initialization method and then use the internal image distributions learned by SinGAN to reconstruct a realistic visual content.

This idea is largely influenced by the harmonization method presented in Figure 3, where the SinGAN works on the details and fine textures of the input image to *harmonize* the added object into the original training image, without transforming its general structure (injection of a down-sampled version of the input image into one of the finer scales). The main difference here is that we will need to adapt the injection scale for the generation pipeline (see Figure 1): large holes that have been initialized at a coarse scale will require the SinGAN to reconstruct general shapes and objects of the training image, before moving on to finer details. However, we heavily rely on the assumption that the

(a) Object Removal.

(b) Realistic Image Reconstruction.

Figure 2: Inpainting tasks.

.

SinGAN is actually capable of learning the "true" internal image distribution (even for damaged images).

The main contributions of this project therefore aim at answering the following two questions:

- How do we initialize the holes? (cf. section 2.2)
- How can we generalize this method to heavily damaged images where the learned distribution might be unrepresentative of the global image? (cf. section 2.3)

2.2. Initialization Methods

Hole-initialization is the first and probably most important step of our method. The SinGAN will indeed use this information as a prior to its generation process. Below we explain the three proposed methods shown in Figure 4. Numerical results and the influence of those different initialization techniques will be presented in section 3.1.

Mean Value : As we wish to only use the "outside-hole"-information of the image, the first intuition was to use a simple estimate of this distribution: the average value of all the pixels *outside* the hole. We denote it *Mean Value*.

Progressive Local Means : However, this first approximation is rather rough, especially for large holes. To avoid being sensible to values of pixels located far from the hole, we propose a progressive fill-in method using "local means": for a given radius r, we consider the patch $P_{r,p}$ of size $2r+1$ centered in the pixel p. The new value $u(p)$ for the considered pixel p will then be:

$$u(p) = \sum_{q \in P_{r,p}} u(q) \tag{1}$$

We iterate over all the pixels p inside the hole and use, at each iteration, the previously updated values of $u(q)$ to update $u(p)$ as in (1). The radius r should be tuned and adapted to the size and more or less complex color distributions of the image.

Nearest Neighbor Patches : Finally, we also chose to implement a coarse-scale and inpainting-adapted version of the *PatchMatch* algorithm by [Barnes et al., 2009] based on the code from [Guo, 2019]: "Nearest-Neighbour"-patches from the non occluded part of the image are used as an approximate solution to the missing patches.

2.3. Partial Convolutions

While initialization ensures that the SinGAN will base its sample generation only on the "outside-hole" part of the image, the model actually learns the internal distribution of the entire image, including holes. This "shift" can be problematic if the part we wish to reconstruct has a big impact on the image distribution: the SinGAN now learns a "false" distribution. The reconstruction will thus be biased and can lead to unwanted results (see section 3.1). So how can we make the SinGAN learn only the "outside-hole" distribution? Partial Convolutions [Liu et al., 2018] give us a way to enforce what part of the image the model should train on.

Architecture : Partial Convolutions are a building block for models similar to classic convolutions, but working only on the pixels indicated by a mask (same size as the image). Formally, the output x' can be written using the network weights W, biases b, a binary matrix M representing the mask and the features X:

$$x' = \begin{cases} W(X \odot M)\frac{1}{sum(M)} + b & \text{if } sum(M) > 0 \\ 0 & \text{otherwise} \end{cases} \tag{2}$$

This method has the advantage of requiring minimal architectural modifications[1]: we transformed the original SinGAN model by replacing all classic convolutions with Partial Convolutions. However, we still need to provide a mask at each level of the SinGAN's multiscale-architecture.

Down-sampling the masks : Indeed, it is not straightforward how one should down-sample the *binary* mask, since the down-sampling method returns continuous values. Furthermore, a bad mask (i.e. not covering the entire hole) would make Partial Convolutions useless.

We first implemented bounding boxes covering localized holes, which would be easy to down-size. A better way that can be applied to unstructured holes, is to down-sample the masks directly in the same way as the images and then threshold this new version to obtain binary values. A sufficiently high threshold (we chose 0.99) ensures that the mask covers the entire hole. Figure 5 displays an example of the method.

3. Results and Evaluation

Our code can be found at https://github.com/RomanCast/SinGAN_for_Inpainting.

[1] https://github.com/NVIDIA/partialconv

3.1. Results

We here present the results of our proposed inpainting method using SinGAN and compare it to a few recent inpainting techniques (cf. Appendices).

Influence of Initialization : The results presented in Figure 6 show that SinGAN is very sensible to the chosen initialization method. Indeed, leaving the hole blank (Figure 6c) makes SinGAN reconstruct an empty space. Using the *Mean Value* shows an improvement (Figure 6d), but the reconstruction remains unrealistic, with a large hole in the ground appearing (see Figure 6a). Using local techniques such as *Local Means* (Figure 6e) or *NN-patches* (Figure 6f) helps to avoid any confusion with the sky region of the image. However, the *NN-patches* method leads to visually unsatisfying results. The SinGAN attempts to reconstruct ground-truth information of the NN-patches: global shapes that are not necessarily at the "right" place become visible if the hole is large enough. A finer *PatchMatch* version could solve the problem, but would be computationally expensive, thus not fit as a *fast* initialization method. Overall, the progressive *Local Means* filling method appears to be the best solution.

Impact of Partial Convolutions : In Figure 7, we compare the vanilla SinGAN model against the SinGAN implemented with Partial Convolutions (PC-SinGAN). We argue that PC-SinGAN can be especially helpful in the case of very damaged images, where only a small part of the image is preserved. The vanilla model has trouble reconstructing the geometry of the image (vertical and horizontal bars are sometimes missing or deformed), while PC-SinGAN succeeds in maintaining these global shapes. The vanilla SinGAN also reproduces a "false" color distribution (some red on the right side of figure 7b), while the "true" color-distribution is maintained by PC-SinGAN (Figure 7c). Vanilla SinGAN also reconstructs a part of the hole (white part in figure 7b).

Both models fail to properly fill in the holes in the dark parts of the image (extreme left, extreme right in figures 7b and 7c). Those border problems are probably due to the initialization method.

Finally, the PC-SinGAN reconstruction appears much more blurry, with lines fading even in non-occluded parts of the image.

3.2. Qualitative Evaluation

In this section we compare our SinGAN results to those of 2 different inpainting methods, both trained on the Places2 dataset [Zhou et al., 2017]: SigGraph [Iizuka et al., 2017] and Generative Inpainting [Yu et al., 2018]. We used the pretrained models and Python implementations [2][3] and evaluated their performance on different occluded images. Our evaluation is strictly qualitative. Indeed, metrics like PSNR are mostly adapted for noise.

Object Removal : The comparison for object removal can be seen in figures 8,9 and 10. SinGAN performs better than SigGraph, which creates artifacts. For instance, we can see small objects in place of the birds in the top image, and some kind of deformation in the middle image. SigGraph has been created for rectangular shaped holes, probably making it unsuitable for general forms of occlusions.

However, SinGAN produces worse results than Generative Inpainting. Although both methods are on par when dealing with small holes (in the top image), Generative Inpainting produces much cleaner results for the bottom image, where it succeeds in reconstructing bars with precision. SinGAN on the other side produces blurry results in the occluded part. Still, this image belongs to the Places2 dataset which could introduce some positive bias.

Damaged Image Reconstruction: Results for damaged image reconstruction are on figure 11. SigGraph fails completely, probably due to it being unsuitable to non-rectangular holes. For the top image, SinGAN and Generative Inpainting perform comparably. SinGAN reconstructs the image fully but gives blurry results. Generative Inpainting has finer details but the mask is still perceptible. It performs much better on the beach image where the results of PC-SinGAN are again blurry. Still, this image belongs to the Places2 dataset which could explain the better results.

4. Conclusion and Perspectives

Overall, the ability of SinGAN to learn the distribution from a single image makes it suitable for image inpainting. In our study, we demonstrate that by carefully choosing how to initialize occluded parts and by modifying the model with Partial Convolutions so that it avoids modelling holes, we end up with a good inpainting method that is extremely sample efficient.

However, results become blurry when evaluating on large holes. In general, SinGAN has trouble reconstructing finer details.

Future work could improve the method of reconstruction, for example by working progressively in order to reconstruct large holes in several steps. Based on the presented results on initialization, one could try to use "smart" priors by precisely selecting the parts of the image that are wished to be used for reconstruction (e.g. detecting land or sky, before reconstruction) to then attribute high weights for important parts vs. low weight for unwanted ones.

[2] https://github.com/akmtn/pytorch-siggraph2017-inpainting

[3] https://github.com/JiahuiYu/generative_inpainting

References

[Barnes et al., 2009] Barnes, C., Shechtman, E., Finkelstein, A., and Goldman, D. B. (2009). PatchMatch: A randomized correspondence algorithm for structural image editing. *ACM Transactions on Graphics (Proc. SIGGRAPH)*, 28(3). 2

[Guo, 2019] Guo, M. T. (2019). Patchmatch. https://github.com/MingtaoGuo/PatchMatch. 2

[Iizuka et al., 2017] Iizuka, S., Simo-Serra, E., and Ishikawa, H. (2017). Globally and Locally Consistent Image Completion. *ACM Transactions on Graphics (Proc. of SIGGRAPH 2017)*, 36(4):107. 3

[Liu et al., 2018] Liu, G., Reda, F. A., Shih, K. J., Wang, T.-C., Tao, A., and Catanzaro, B. (2018). Image inpainting for irregular holes using partial convolutions. In *Proceedings of the European Conference on Computer Vision (ECCV)*, pages 85–100. 1, 2

[Shaham et al., 2019] Shaham, T. R., Dekel, T., and Michaeli, T. (2019). Singan: Learning a generative model from a single natural image. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 4570–4580. 1, 5

[Yu et al., 2018] Yu, J., Lin, Z., Yang, J., Shen, X., Lu, X., and Huang, T. S. (2018). Generative image inpainting with contextual attention. *arXiv preprint arXiv:1801.07892*. 3

[Zhou et al., 2017] Zhou, B., Lapedriza, A., Khosla, A., Oliva, A., and Torralba, A. (2017). Places: A 10 million image database for scene recognition. *IEEE transactions on pattern analysis and machine intelligence*, 40(6):1452–1464. 3

Appendices

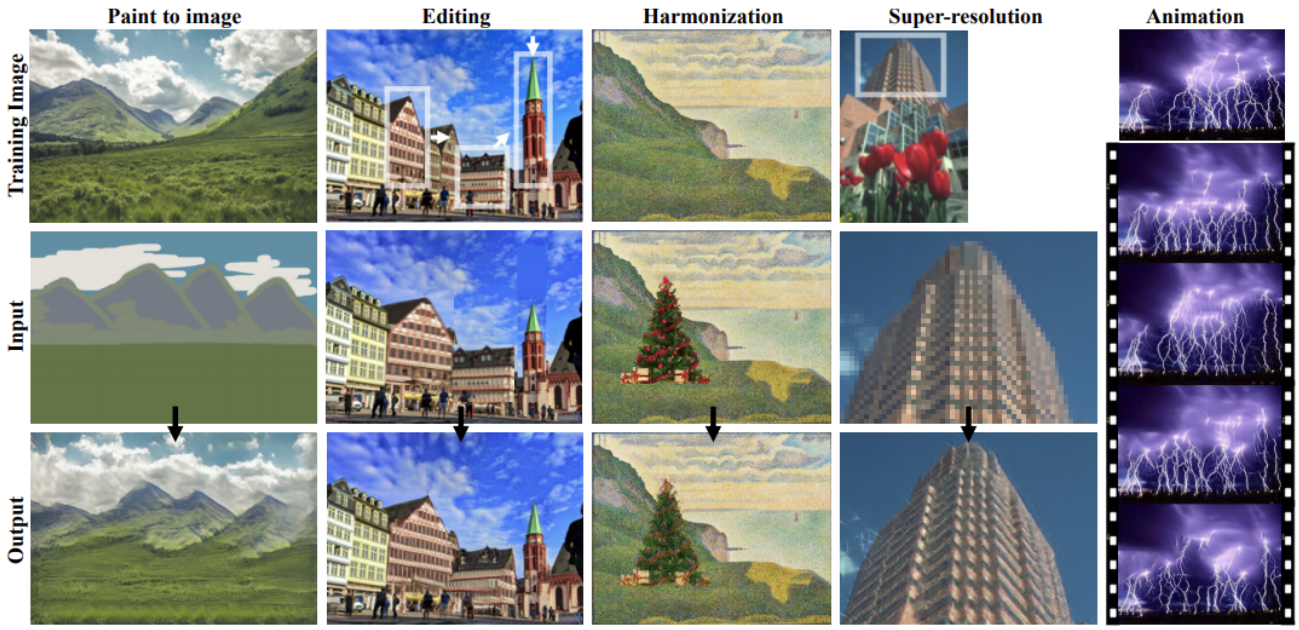
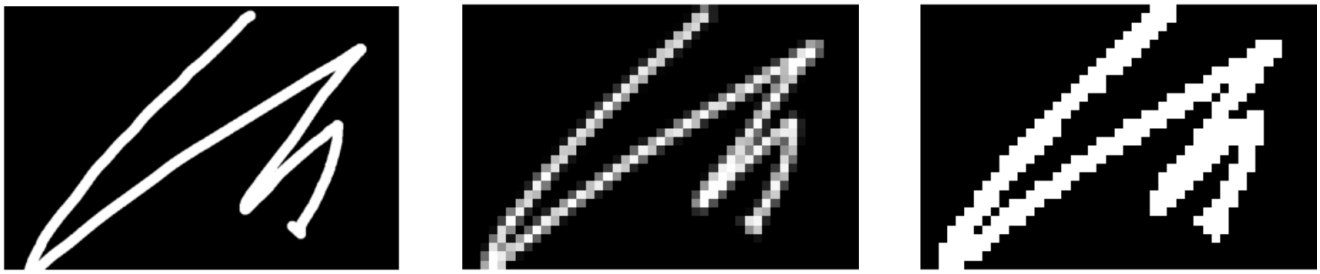


Figure 3: Image manipulation using SinGAN [Shaham et al., 2019].

Figure 4: Initialization Methods. Using the mean (left); local means (middle) of the "outside-hole" pixel values; using "nearest-neighbor"-patches (right) from the "outside-hole" image.



(a) Original mask. (b) Mask downsampled. (c) Binary mask.

Figure 5: Mask downsampling process.

(a) Damaged image.

(b) Mask of damaged image.

(c) No initialization.

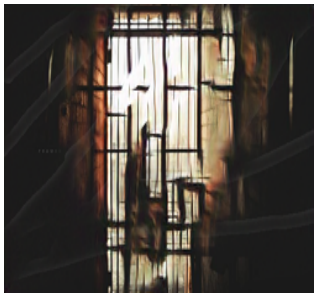
(d) Mean Value.

(e) Progressive Local Means.

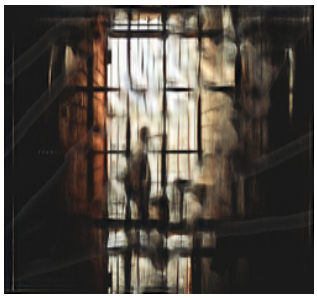
(f) NN-Patches

Figure 6: Results for different initialization methods using the same SinGAN and the same damaged image (6a) injected at scale 3/8. Partial Convolutions were used.

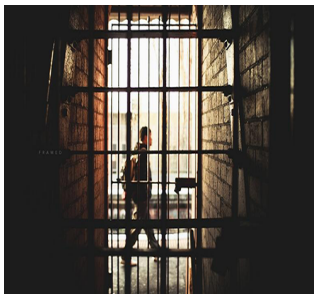
(a) Damaged image.



(b) Vanilla SinGAN.



(c) PC-SinGAN.



(d) Ground Truth.

Figure 7: Damaged Image Reconstruction. SinGAN trained on damaged image (7a) with (7c) or without the use of Partial Convolutions (7b).

Figure 8: Inpainting results for Object Removal. Comparison of our SinGAN model to SigGraph and Generative Inpainting (GntIpt). The SinGAN was trained with partial convolutions and initialized with local means. Injection scale is 3 for the first and last row, 4 for the beach image (middle row).

(a) Original image. (b) Occluded image. (c) PC-SinGAN (ours). (d) Generative Inpainting.

Figure 9: Inpainting for Object Removal using SinGAN and Generative Inpainting. For an image that does not belong to the Places2 dataset, SinGAN is much more consistent with previous results while the hole is poorly reconstructed using Generative Inpainting.

(a) Original image. (b) Occluded image. (c) SigGraph. (d) PC-SinGAN (ours). (e) Generative Inpainting.

Figure 10: Inpainting for Object Removal using SigGraph, SinGAN and Generative Inpainting. The borders of the SigGraph method remain visible here, and Generative Inpainting creates dark artifacts in the middle of the image.

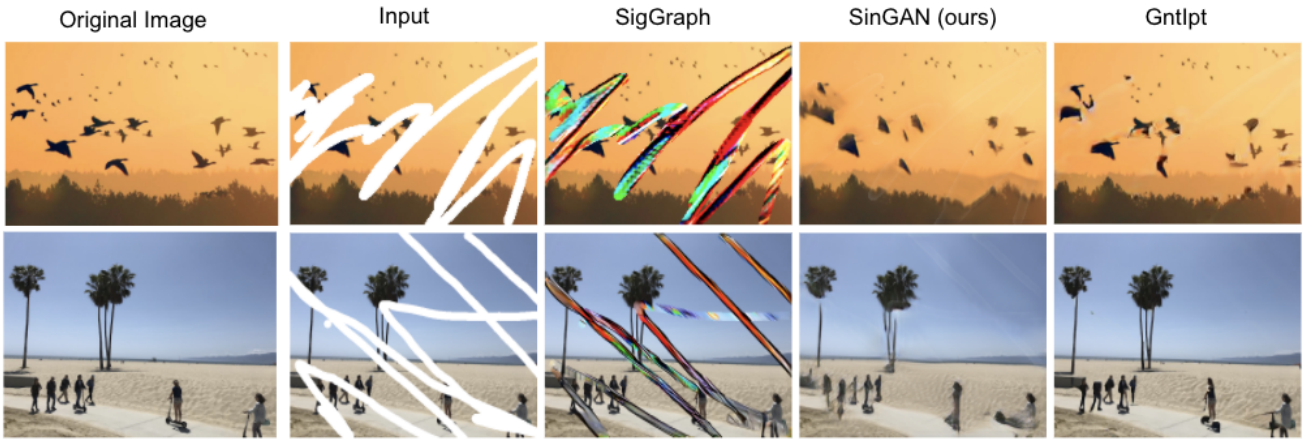


Figure 11: Inpainting results for Damaged Image Reconstruction. Comparison of our SinGAN model to SigGraph and Generative Inpainting (GntIpt). The SinGAN was trained with Partial Convolutions and initialized with local means and different radius values (7 and 15 respectively). Injection scale is 3.