Final project: Transforming image into pictures and vice-versa using generative adversarial networks (SinGANs)

Mathieu Rita
Ecole Polytechnique
mathieu.rita@polytechnique.edu

Quentin Le Lidec
Ecole Polytechnique
quentin.le-lidec@polytechnique.edu

Abstract

Our project is an extension of "Image manipulation with GANs". Our work will be based on the article "SinGAN: Learning a generative model from a single natural image" from Tamar Rott Shaham et al. [7] and will investigate several artistic applications of SinGAN, in particular the link between paintings and real images. In the first part, we will focus on the photo to painting task which appears to be adapted for SinGAN. Then, we will address the more complex task of going from a painting to the picture that inspired the artist. We will show why this task is challenging and how we can try to improve the results of SinGAN.

1. Introduction

Adversarial networks have proven their effectiveness in understanding the distribution of an image dataset and in generating new samples from this distribution. For instance, given a whole dataset of paintings from Van Gogh, a GAN would be able to understand the underlying structure of these images and then to produce paintings in Van Gogh's style.
Unlike GANs, the SinGAN are able to learn the structure of an image from a unique sample. Indeed, extracting patches of different scales allows SinGANs to discover the internal statistics of patches in an image. We will use this architecture, introduced in [7] by Tamar Rott et al., to build a model able to go from pictures to paintings and vice-versa.

1.1. The architecture of SinGAN

SinGANs are purely generative as they map a noise to image samples. The main idea of SinGAN is to build a pyramid of convolutional GANs (Figure [1]). Thus, once an image is generated by a GAN, this image is upsampled and add to some noise to be the input of the next GAN of the pyramid. Each one is made to learn the distribution of patches at one scale and the scale goes finer as we go up in the pyramid. Tamar Rott et al. [7] show that this multi-scale architecture allows to capture more global consistency. Indeed, when the number of scales is increased, SinGAN manages to understand the global arrangement of an image and not only the fine textures. This is why the coarsest scales are determinant in the architecture.
When it comes to the photo to painting tasks, this architecture allows to control the degree of stylization we want for the photo we are modifying. Indeed, by starting from the bottom of the pyramid the SinGAN will modify numerous scales by introducing characteristics of the original painting and the style of the painter will be more visible.

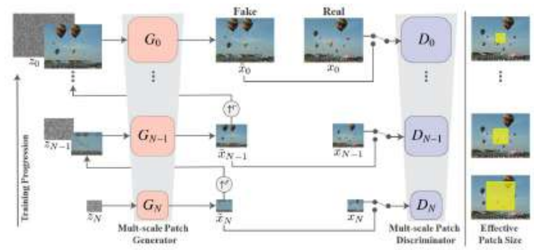


Figure 1. Architecture of a SinGAN

1.2. Evaluation

For such a task, the most appropriate evaluation is **qualitative**. When we will transform photo into paintings we will evaluate how well we recover the style of the painter from which we trained our model while keeping the structure of teh original photo.

For the painting to photo task, we will merely take into account the realism of the obtain photo. In general, surveys are done to realize this task. However, in our case it would be artificial to create polls of a few persons. Therefore, we will only bring qualitative comments.

In addition, we will use the Single Image Fréchet Inception Distance (SIFID) as a **quantification** metric. The Fréchet Inception Distance (FID) is introduce in [5] by Heusel et al. and allows to compare the distributions of two sets of images. The FID corresponds to the Fréchet Distance between the distributions of the last activation layer of an inception layer of two distributions of photo. In [7], Tamar Rott et al. adapted the FID to compare the internal statistics of two images with the Single Image Fréchet Inception Distance (SIFID). Thus, SIFID is the FID between the distribution of patches of two photos.

2. From photos to paintings

2.1. Advantages of SinGAN

SinGAN seems particularly adapted to this task for at least two reasons:

- Only a single painting is needed to capture the style of a painter. Indeed, one can easily make the difference between a painting from Van Gogh or Picasso. Therefore, we may think that a single painting is enough to capture patches in the style of a painter.

- The style of a painter appears at different scales in a painting. Some painters (eg. Picasso, Van Gogh) manage to deform the main lines of the paintings (large scale) while other painters (eg. Seurat, Monet) mostly work on textures. The ability of SinGAN to work at different scales is therefore a strong advantage.

2.2. Methodology

In order to perform this task we trained SiNGAN on paintings of painters that present different style characteristic. Then, we implemented a script [1] that allows to generate a painting from an original picture fed in the trained model at a given starting scale N. We ran the script for different values of N (typically in $\{1, ..., 8\}$). Then, we compared the N paintings and chose the most appropriate one. This choice was a trade-off between the style of the painter and the original image. The resulting starting scale is therefore highly linked to the characteristic scale of the style of the painter and the shapes of the objects present in the original image. In **Appendix [1]**, we show how an image evolved when generated by a model trained on a painting from Van Gogh.

2.3. Results

Following this methodology, we trained models of 5 different painters chosen because of the characteristic shape of their style (Seurat: small scale to Wharol: large scale). Moreover, the main colors of the paintings vary between the photos. These paintings will therefore constitute a strong base for comments. We present the results and comment them in **appendix [2]**.

2.4. Comparison with the state of the art

The results we got are very correct. We both recognize the original picture and the style of the painter. However, we have to compare our results with the state of the art in order to evaluate the quality of SinGAN for this task. Neural Style Transfer (NST) [3] is the reference for algorithms trained on a single image that perform style transfer. The comparison is shown in **Appendix [3]**. Here are some comments about the images:

- In overall, the results obtained with NST can be considered as better: **SinGAN does not improve the state of the art**

- We see that NST is very good in capturing textures and apply them at a bigger scale. Indeed, the results for Seurat, Van Gogh and Picasso are impressive. On the contrary, when the algorithm is trained on patterns that are larger that the characteristical scale of photo to transform, the results are not convincing. See the result when training on Andy Warhol. Moreover, we can notice that the algorithm does not adapt the colors to the input photo. It merely apply the set of colors present in the training painting (which is, depending on the taste, good or no).

- The application of texture is not as impressing when using SinGAN. Nevertheless, the style transfer is done. In comparison with NST, we can observe that SinGAN tend to blur the image. This can be explain by the upscaling step that is done when the image is "transfered" from a GAN cell to the other. In addition, even if the algorithm adapts slightly the colors, it does not transform much the histogram of colors. This point is interesting compared to NST.

3. From paints to photo

Now let's move on to a way more complicated task: transforming paintings into photos. This task seems very challenging since the algorithm is trained only on a single image and therefore is limited in the number of patches it learns. Moreover, if artefacts or blurry effects could be accepted in the previous artistic task, we want them to be totally removed for this new task. To address this new problem, we consider a progressive protocol:

- Reproduce a result from the paper: converting cliparrt into photo

- Convert a painting to a photo where SinGAN is trained on the target distribution of patch

- Convert a painting when the pair training photo/painting are very similar

- Convert a painting when the pair training photo/painting are slightly more different

3.1. Preliminary experiment: from clipart to photo

We addressed the slightly easier task of going from a clipart to a photo in order to understand the advantages of SinGAN over other state of the art algorithms but also the main challenges it induces when going from a painting to a picture. The results are available in **Appendix [4]**.

3.1.1 The advantages of SinGANs

In the case of the clipart to photo tasks, as the entry contains very little information, the algorithm must be able to make important changes on the entry by modifying its style but also partially the content (**Appendix [4] Figure [1]**). This is something that can be handled by SinGAN due to its generative characteristics but it is not the case for other methods like NST which try, by definition, to be as close as possible to the content of the entry by transforming only the style.

	SinGAN	Neural Style Transfer	Contextual Loss
SIFID	0.634	0.638	0.635

Table 1. Evaluation of Single Image Fréchet Inception Distance from the style image on the clipart to photo task

This ability is very useful when the entry is perturbed like in Figure [3] as it allows to obtain a consistent result. More generally, SinGANs benefits from being generative models when it comes to tasks where creativity is needed (assistive design etc...).

3.1.2 Transferring different color histograms

To use the information contained in a clipart, which generally is the main shapes at a large scale, we use the clipart as an entry of a generator in the pyramid (the lower the clipart is introduced, the more it is modified and the closer it is to the target picture). Then, the clipart replace what should be the output of the generator of the precedent scale of the pyramid and they needs to have approximately the same characteristics to make the SinGAN works properly. Moreover, the characteristics of an output from a generator of a SinGAN are close to those of the training picture because it is exactly what SinGAN is trying to learn. Then, the main characteristic that needs to be matched between the clipart

and the training image is the histogram of colors. In fact, it appears that SinGAN is identifying objects with their color and can be conused if they don't match (**Appendix [4] Figure [2]**).
This challenge is also important when it comes to transforming a complex painting into a picture and it explains why we will need to go through a preprocessing step to equalize the histogram of colors for the painting.

3.1.3 Learning object of different scales

When we wanted to achieve the photo to painting task, the SinGAN had to learn the style of painter which often consists in characteristics of a similar size. To transform a painting into a picture, the SinGAN needs to learn objects of very different scales in the training picture and this is the main reason why this task is more complicated. For instance, to learn the specificity of a portrait, the SinGAN needs to understand the spatial consistency of a nose but also of the head which are objects of very different size.
We studied the ability of SinGAN to learn object of different scales by artificially introducing objects in the previous picture of a sky. To do that, we added some periodic noise ζ to the original picture: $\zeta_{ij} = A\cos(2\pi f i)$ where i is the abscissa of the pixel, A the amplitude and f the frequency of the noise. This noise appears in the form of horizontal strips. Thus, by varying the frequency of the noise we can control the size of the stripes and then observe how the SinGAN learned these objects.
Finally, we observed that the SinGAN was able to learn the stripes for the low frequency noise, which corresponds to the scale of the clouds, but not when this frequency was too high (**Appendix [4] Figures [4] [5]**). Thus, understanding objects of different scales in a single picture is a complex task for SinGANs and we will address it for the painting to photo task.

3.2. SinGAN trained on the target distribution of patches

The work with clipart allows us to understand some of the main challenges that are involved to succeed in the Painting2photo task. As a first set of experiments, let's try to transform a painting into a photo when SinGAN is trained on the target distribution of patches. To do so, we consider an original photo $\mathcal{O}$. Then we use one of our previous models trained on a painting to transform the photo $\mathcal{O}$ into the photo $\mathcal{I}$. We train SinGAN with $\mathcal{O}$ and feed it with $\mathcal{I}$ in order to recover the photo $\mathcal{O}$.

3.2.1 Result without preprocessing steps

The results are presented in **Appendix [5]** for a portrait of Alain Delon that has been generated in the Van Gogh style. We observe that the result for high values of N are bad and that they get better and better when N is decreasing. From $N = 6$ to $N = 4$, we see that SinGAN starts to be able to recover the high frequency part of the image (hair, top of the hear), then from $N = 4$ to $N = 2$, the flat parts of the image (skin) are becoming neater. This result was expected since for high values of N, SinGAN begins to work directly at small scales. On the contrary, when we start to work at $N = 2$, SinGAN starts with a large scale, and therefore is able to reconstruct larger patches. We therefore see how important the scales are in the reconstruction. A natural improvement is to train a model on a larger number of scales as it is done in **Appendix [6]**.
Eventually, the result is promising but still imperfect both for the models with 8 scales and 14 scales. Even though the results are better with the 14 scales model, there are still artefacts.

3.2.2 Results with preprocessing steps

During the experiments with the cliparts, we observed that the colors in the input image were very important. It seems that SinGAN tends to associate to a given patch of a certain color, the corresponding patch in the training set that has the same color. Therefore, if the color histograms are very different, we cannot expect to obtain a good results (as in **Fig.8 Appendix 8**)
To face this issue, we add a preprocessing steps in the process. We perform a color histogram transfer from the input image $\mathcal{I}$ to the training image $\prime$ according to the method presented in [2]. Then we apply our SinGAN model to this resulting image. The final result is almost perfect and looks like the original image. Thus, the distibution of colors of the inputs seem to be a key issue when using SinGAN. See the comparison of the results and the whole process in **Appendix [7]/**

3.3. Result on similar training image/input pairs

We see in the previous example, that using a small preprocessing step allows to get a nice result. Nevertheless, the experience in itself is a kind of overfitting: the model is trained on the image we want to get. Therefore, let's try a slightly more complex problem: converting a painting into a photo when the training photo is close to the input.
The result we get by merely applying a trained SinGAN is very bad (see **Appendix [9]**). The input is another portrait of Alain Delon and SinGAN is not able to output a correct result. Therefore, once again, we tried to add some pre-processing steps. These pre-processing steps are motivated by some additional experiments we made (see **Appendix [8]**). We observed that SinGAN was very sensitive to color transfer, image rescaling, rotations (even small ones) and that small "painting" details lead to artefacts. To face all these issues, we added three preprocessing steps: first we align the images (for a few examples, we did it by hand but it could be easily generalized), then transfer the colors as explained before, and finally smooth the image by cutting the high frequencies using the DCT denoising algorithm [4].
The whole process is shown in ANNEXE ????. After the preprocessing, we applied SinGAN twice. The final result is quite nice but we see some small artefacts.

3.4. Conclusion: Generalization to unsimilar input / training photo pairs

The previous example has been performed on a very similar pair (training photo/input). We therefore have to be careful with the final results. By testing (ANNEXE ????) our previous process on a slightly more difficult examples, we see that the result is not very good. Thus, we see that it is difficult to generalize our process: the two previous cases are kind of idealistic.
In conclusion, we can consider that SinGAN is not adapted to such a task. It is no surprise, since SinGAN is trained on a single image and thus a very limited number of patches. If we look compare our result to those obtained with NST [3], we see that they are comparable. Our model {preprocessing steps + SinGAN} is performing better than NST on the two portrait tasks and is is comparable for the landscape (NST learns very well the texture but put some texture at the wrong place).
Transforming a painting into a photo is thus still a challenge for networks trained of a single image. For this task, we have to use GANs that are trained on a way larger number of images, such as CycleGAN [6]

References

[1] https://github.com/mathieurita/singan$_styletransfer$.

[2] Bruce Gooch Erik Reinhard, Michael Ashikhmin and Peter Shirley. Color transfer between images. *IEEE Computer Graphics*, pages 34–41, 2001.

[3] Leon Gatys, Alexander Ecker, and Matthias Bethge. Image style transfer using convolutional neural networks. pages 2414–2423, 06 2016.

[4] Guillermo Sapiro Guoshen Yu. Dct image denoising: a simple and effective image denoising algorithm. *IPOL*, 2011.

[5] Martin Heusel, Hubert Ramsauer, Thomas Unterthiner, Bernhard Nessler, and Sepp Hochreiter. Gans trained by a two time-scale update rule converge to a local nash equilibrium. 2017.

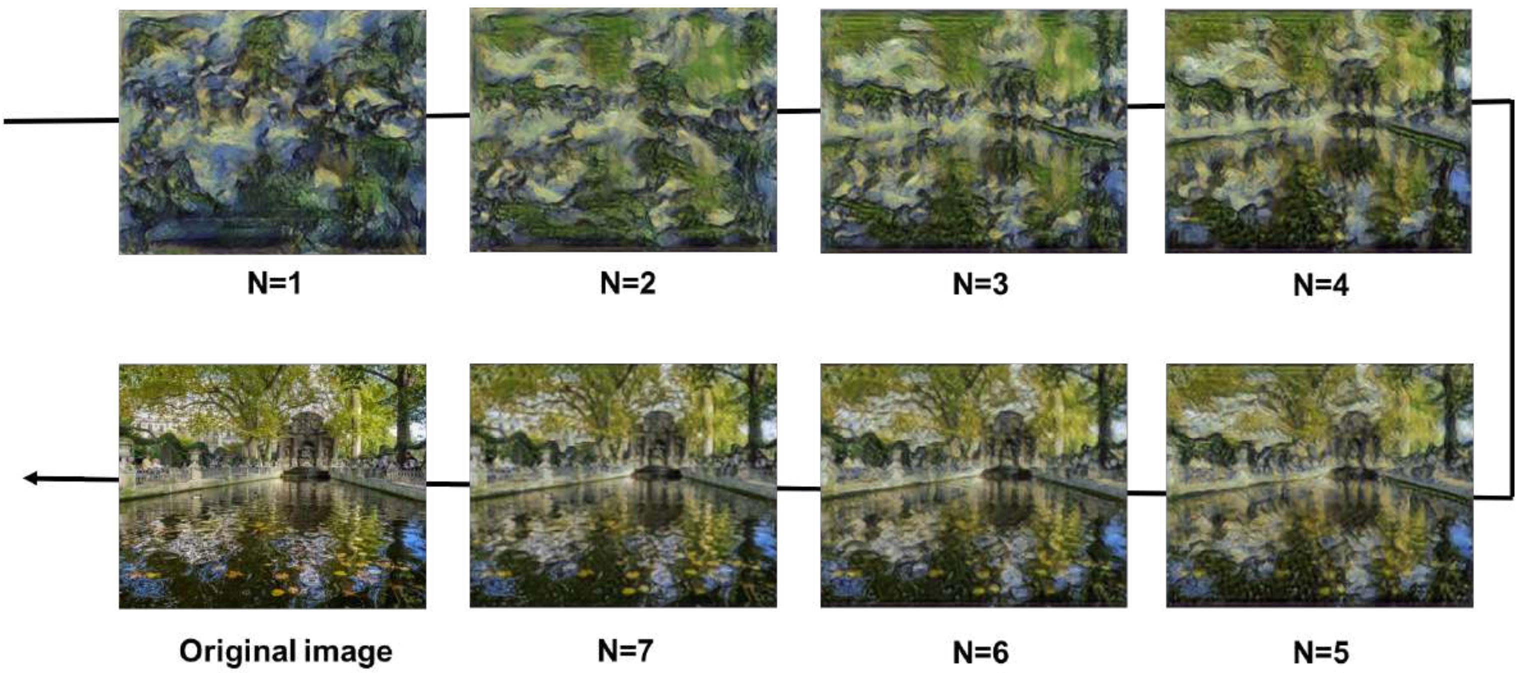
[6] Phillip Isola Alexei A. Efros Jun-Yan Zhu, Taesung Park. Unpaired image-to-image translation using cycle-consistent adversarial networks. *ICCV*, 2017.

[7] Tomer Michaeli Tamar Rott Shaham, Tali Dekel. Singan: Learning a generative model from a single natural image. *ICCV*, 2019.

Appendix: Additional figures

January 20, 2020

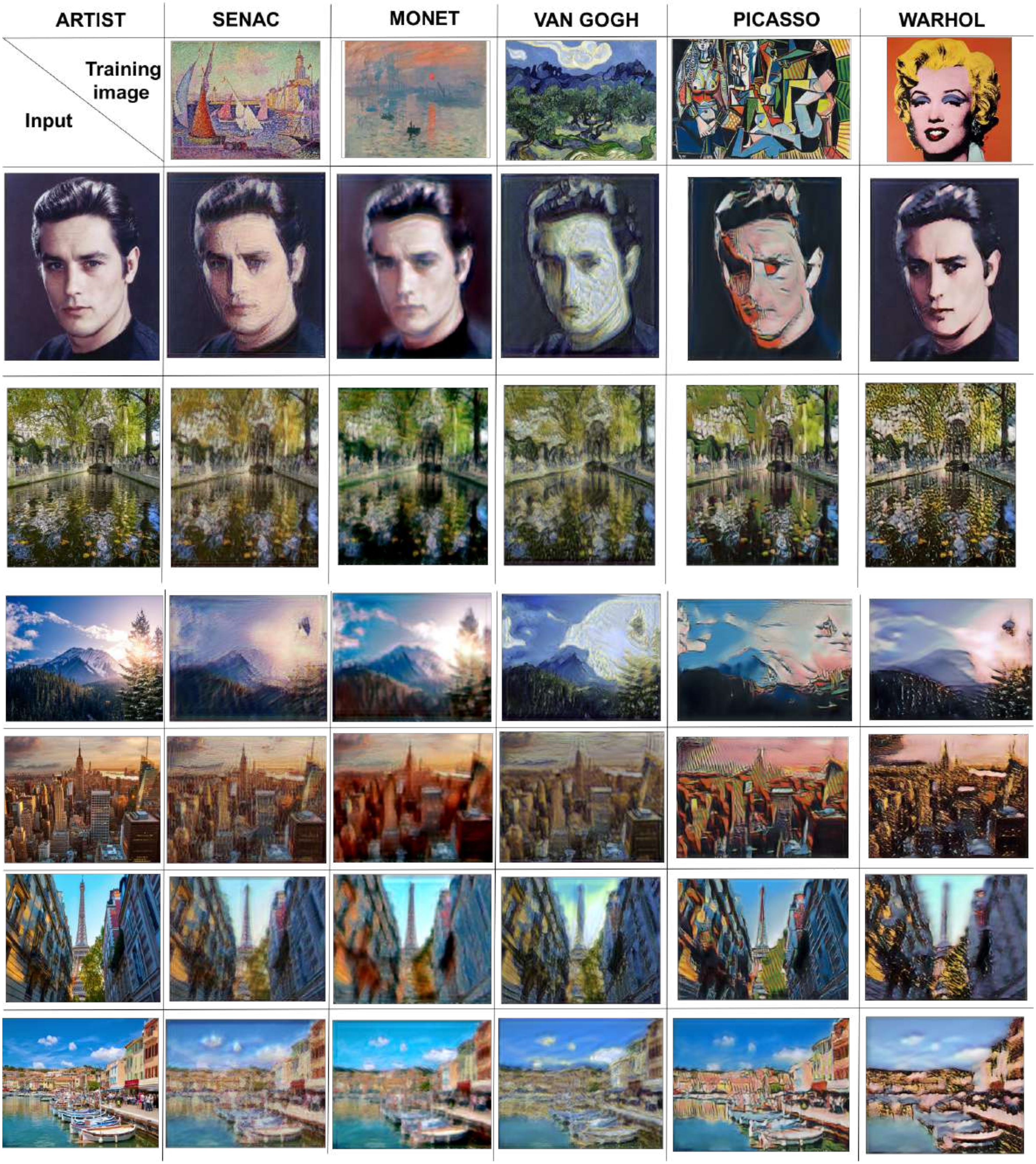
A Appendix 1



Appendix 1: *Outputs of SinGAN when it is trained on a photo of Van Gogh and trained on different strating scales N.*

Based on these outputs, we select the painting that better satisfy the trade-off between being in the style of the painter and respecting the original image

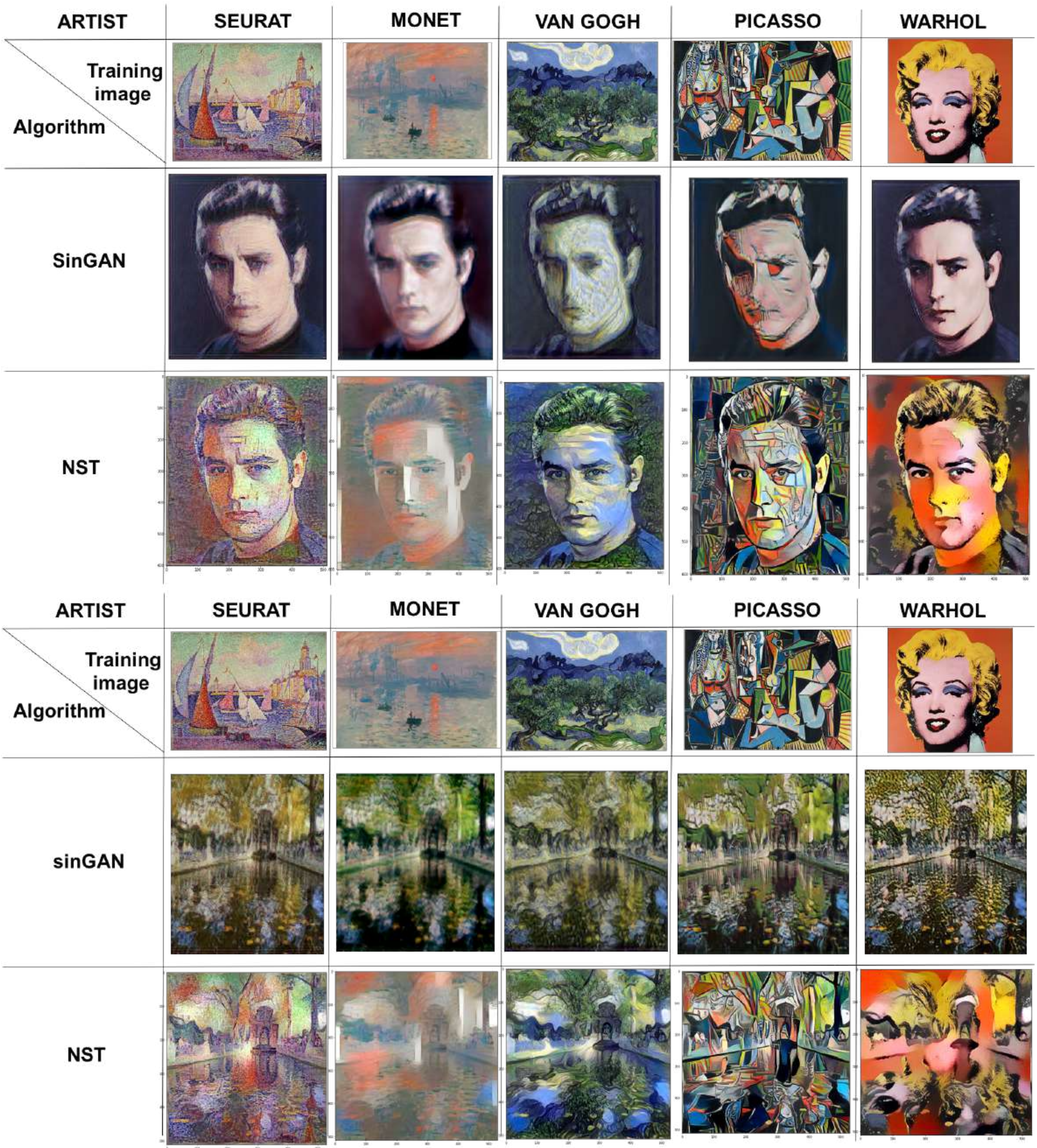
B Appendix 2



Appendix 2: *Result for the photo2paint task for 5 models and several inputs*

- *We trained 5 models of 5 painters. We can see the results for different outputs. The 5 painters have been selected because their characteristic styles appear at different scales.*
- *We see that the results are quite good. Indeed, we well observe the style transfer.*
- *The results for the landscapes are very interesting. We mainly recover the the original image while seing the style of the painter*
- *Since the portrait is at a large scale, the style of Picasso and Warhol has a nice effect. On the contrary, when the style is present at a smaller scale (Seurat, Van Gogh, Monet), there is a blurry effect. This effect can be explained by the succession of upscaling. It is more and more important when the starting scale is low. The problem is that, if we want to capture small patches of the painter, we have to choose a small starting scale. Therefore, we see that there are two antagonists trends. In the end, we choose a scale that give not too blurry paintings while capturing a certain style. The result is correct for Van Gogh and Monet but in the end we do not capture the style of Seurat*

C Appendix 3



Appendix 3: *Comparison between SinGAN and NST on two examples*

D Appendix 4

Figure 1: Results obtained on the clipart to photo task using SinGAN, NST and Contextual Loss. Top: Training image and entry, Bottom: Results from SinGAN, NST and Contextual Loss

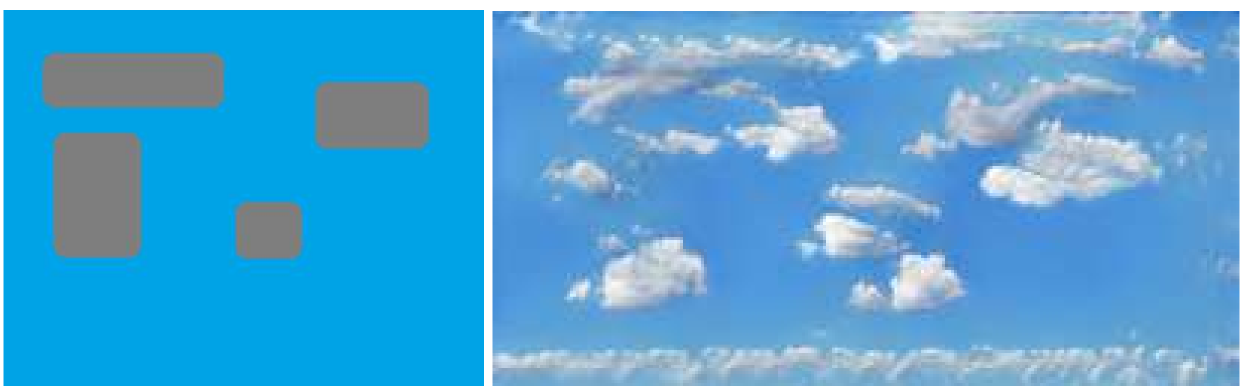


Figure 2: Left: Clipart with a wrong color histogram, Right: The result obtained from the Singan is deteriorated

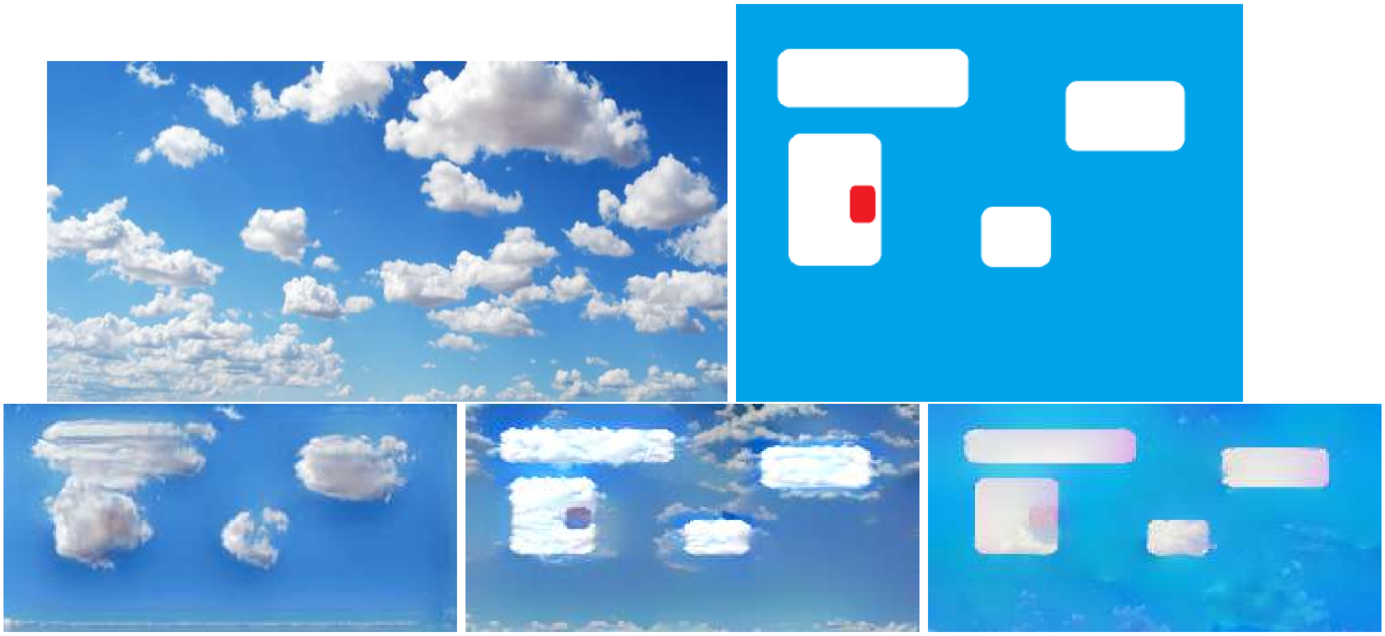


Figure 3: Results obtained on the clipart to photo task using SinGAN, NST and Contextual Loss when the entry is perturbed Top: Training image and entry, Bottom: Results from SinGAN, NST and Contextual Loss

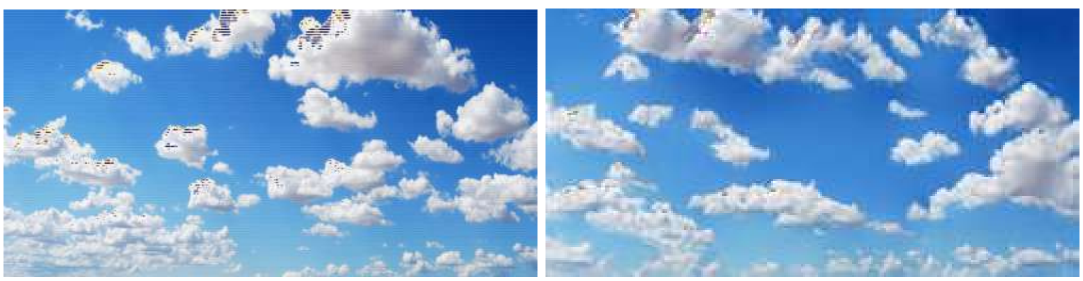


Figure 4: Left: Training image of a sky with high frequence noise, Right: Random sample generated with the SinGAN

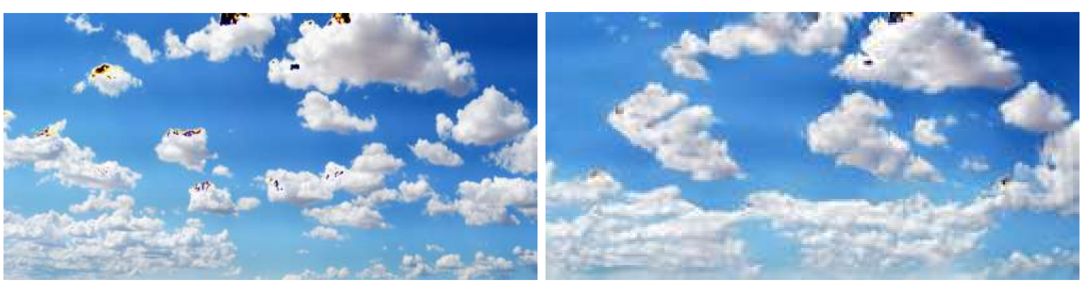


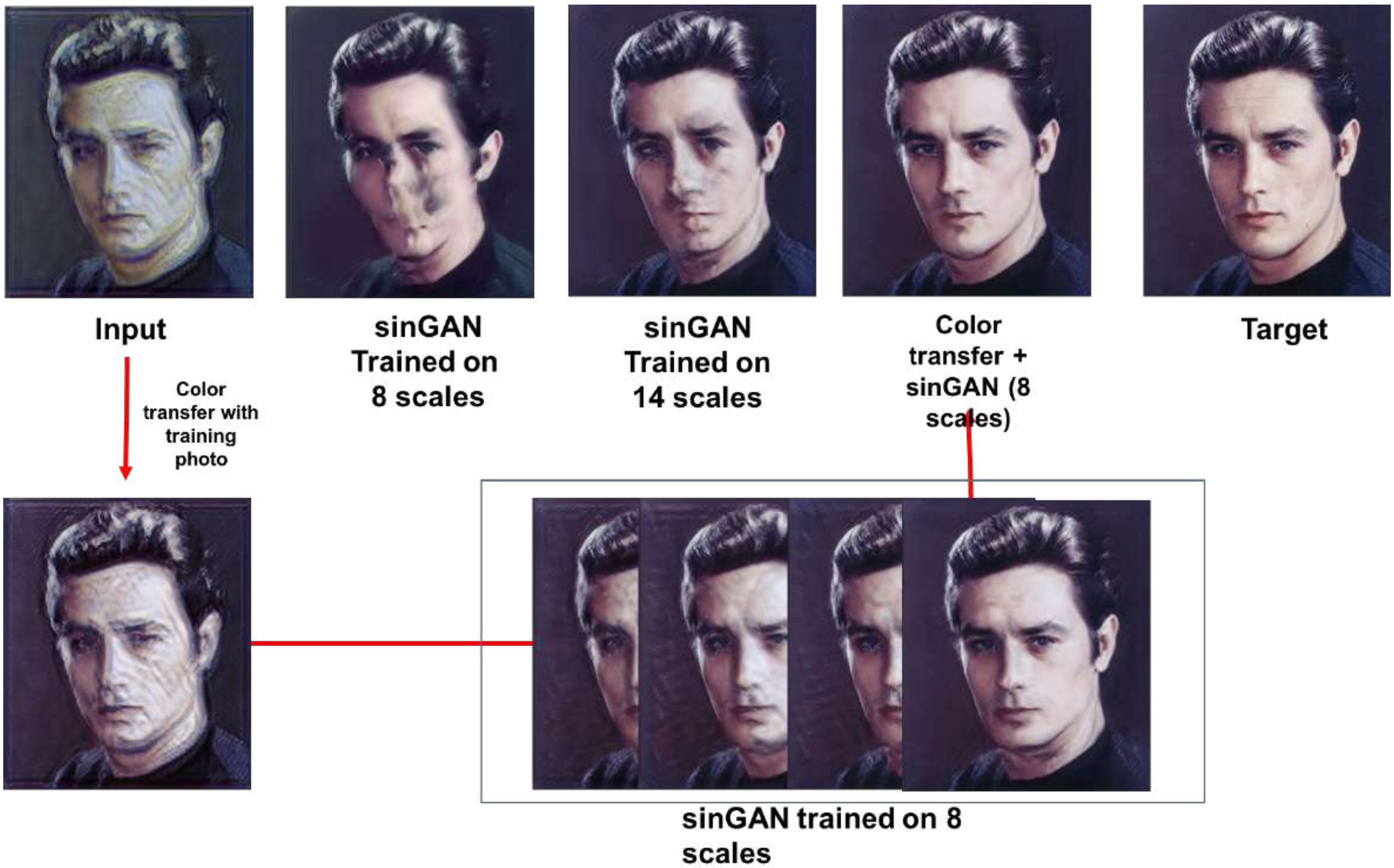
Figure 5: Left: Training image of a sky with low frequence noise, Right: Random sample generated with the SinGAN

Figure 6: From paint to image. Comparisons of results obtained with a SinGAN, a Neural Style Transfer of **?** and a Contextual Loss **?** Top: training image and entry paint, Bottom: Results from SinGAN, NST and Contextual Loss

E Appendix 5

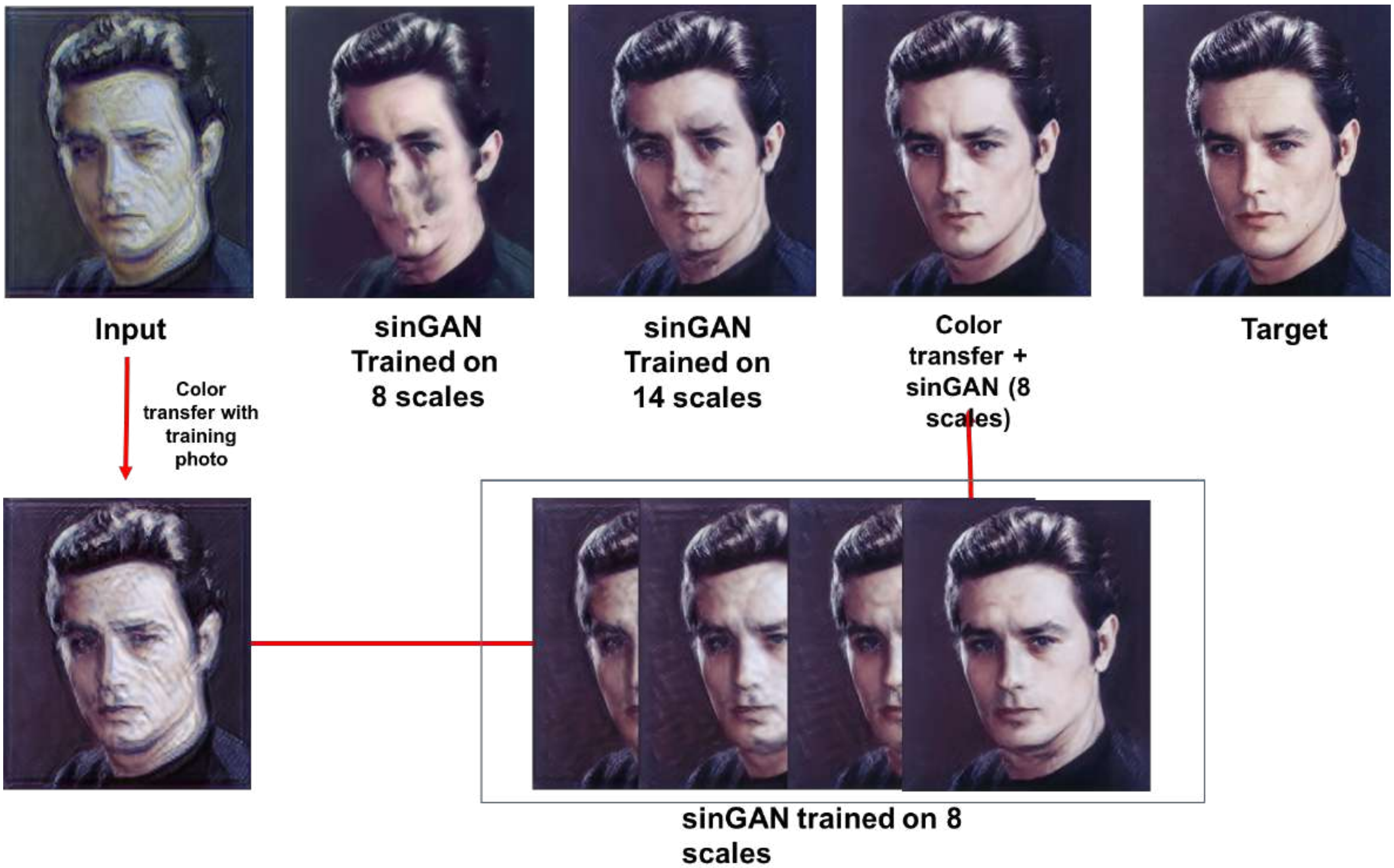
Appendix : *Result of SinGAN for the paint2photo task without preprocessing*

F Appendix 6



Appendix : *Result of SinGAN for the paint2photo task with preprocessing. Comparison with SinGAN alone trained on 8 and 14 scales*

G Appendix 7



Appendix : *Result of SinGAN for the paint2photo task with preprocessing. Comparison with SinGAN alone trained on 8 and 14 scales*

H Appendix 8

Figure 7: Result with the original picture

Figure 8: Color histogram transformation

Figure 9: Rescaling and translation

Figure 10: Small rotation

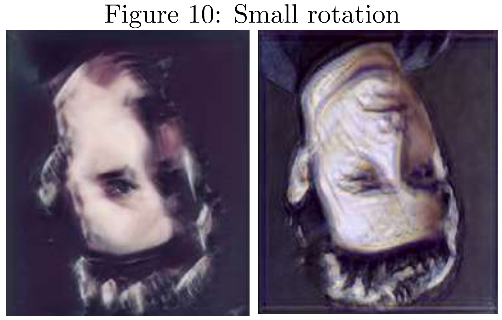
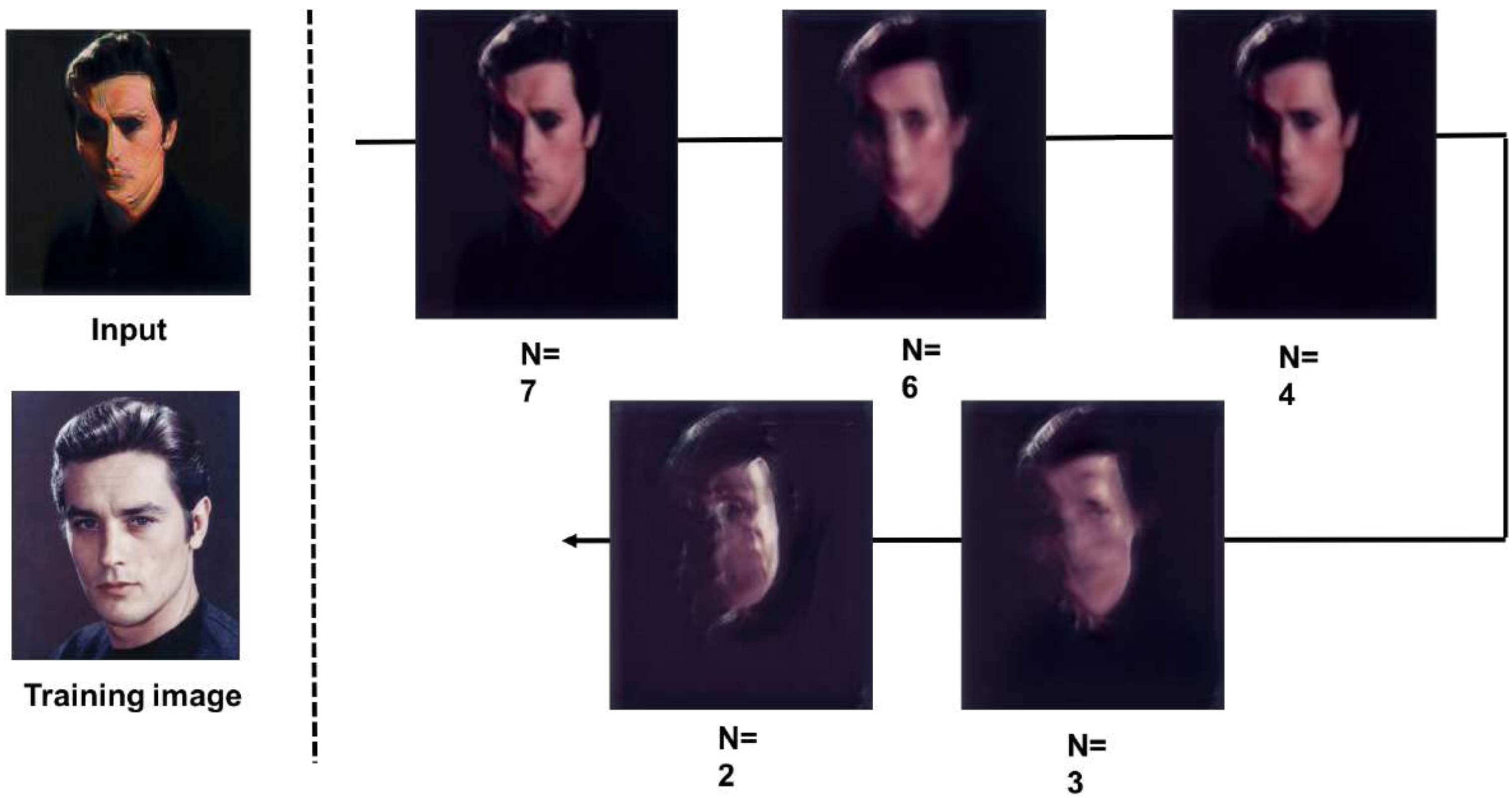


Figure 11: Big rotation

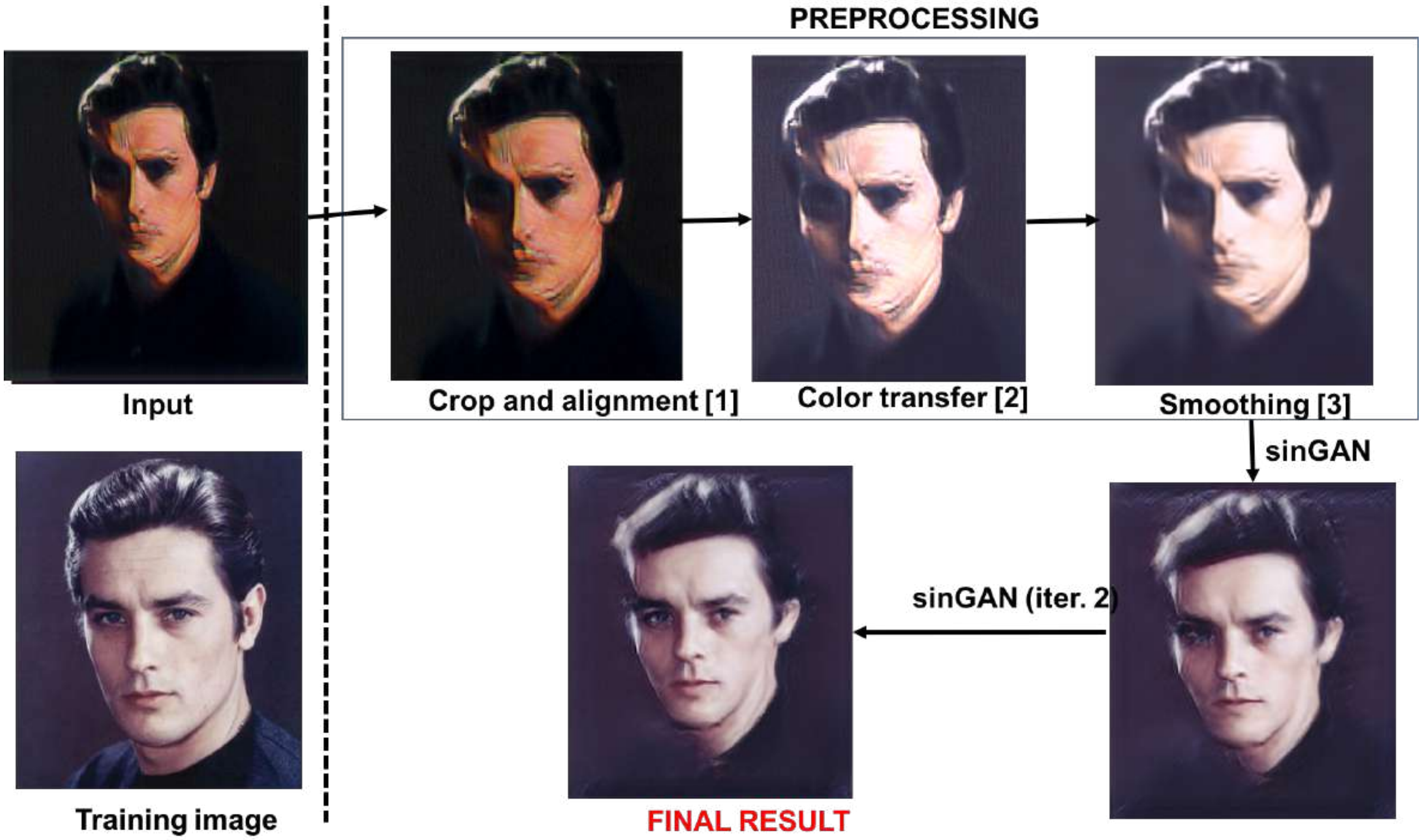
Appendix : *Result of sinGAN in paint2photo of Alain Delon's portrait after different transformations*

I Appendix 9



Appendix : *Result of SinGAN for the paint2photo task without preprocessing. It is trained on Alain Delon's portrait and with another portrait of Alain Delon as input*

J Appendix 10

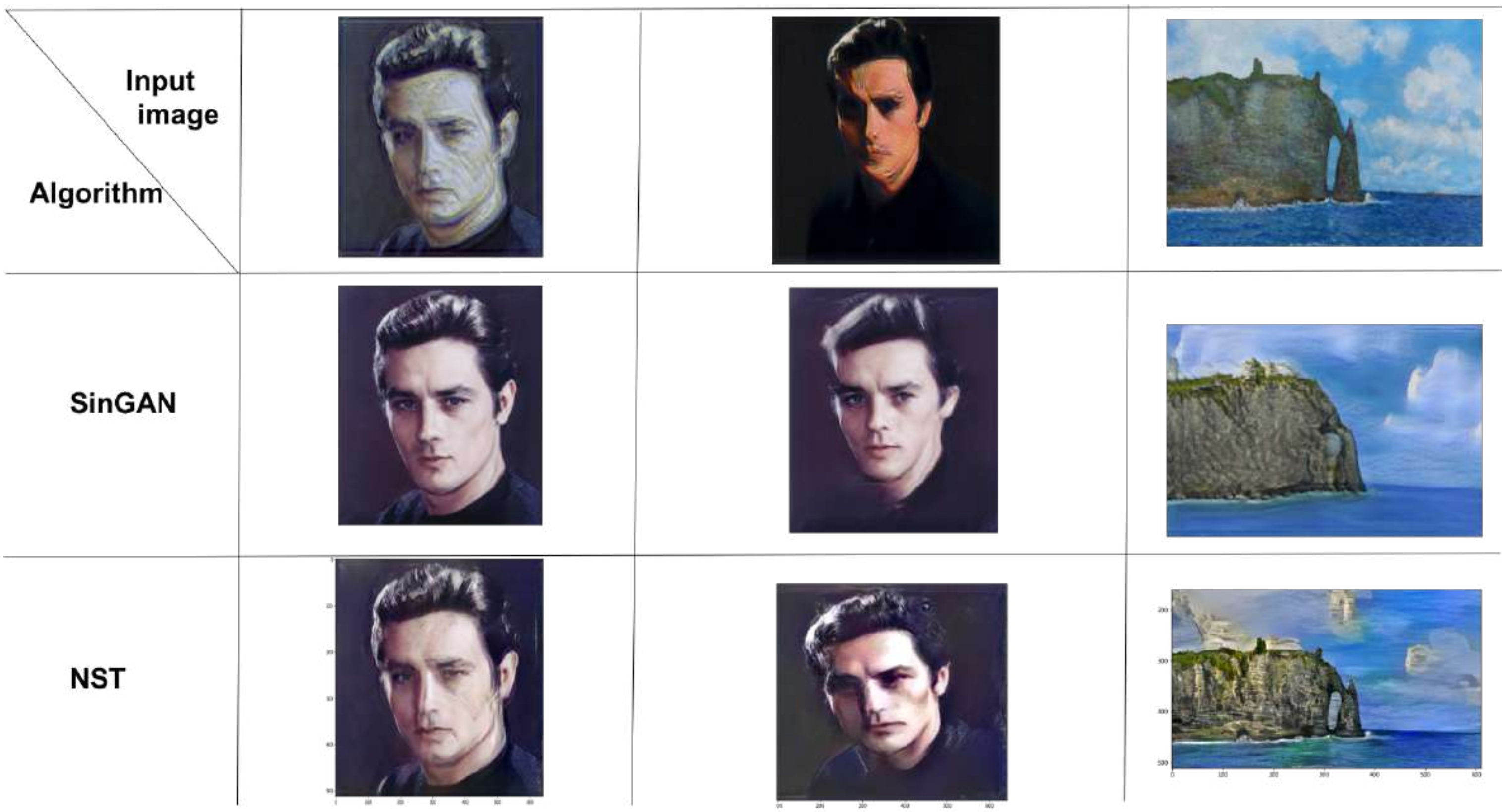


Appendix : *Result of SinGAN for the paint2photo task when trained on Alain Delon's portrait and with another portrait of Alain Delon as input*

K Appendix 11

Appendix : *Result of SinGAN for the paint2photo task for a common landscape (city of Etretat)*

L Appendix 12



Appendix : *Comparison between SinGAN and NST for the results of the Paint2Photo task*