Image manipulation with generative adversarial networks (GANs)

El Mennaoui Mohamed, Kerroumi Mohamed

Object recognition and computer vision 2019-2020, Final Project

Outline

Outline

Introduction

- The paper [4] proposes a new unconditional generative model trained on a single natural image.
- The model learns the internal distribution of patches within the image
- Then the model generate new realistic and diverse image samples that preserve the original patch distribution while creating new object configurations and structures.
- SinGAN can also be used to achieve a variety of image manipulation tasks.(e.g Paint-to-image , harmonization...).

Outline

Model Description - SinGAN

- **Goal :** Capture the internal statistics of a single training image I
- Unlike conventional GAN setting, the training samples here are patches of a single image
- We need to capture the image statistics at many different scales
- The architecture consists of a hierarchy of patch-GANs, each level of the pyramid capture the patch distrubtion at a specific scale.

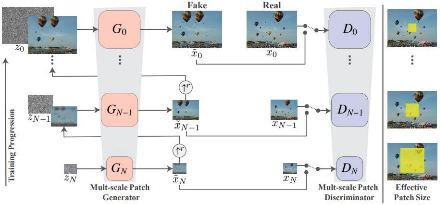


Figure: SinGAN's multi-scale pipeline [4]

- The model consists of a pyramid of generators $\{G_0, ..., G_N\}$ train against a pyramid of images of I $\{I_0, ..., I_N\}$, I_n is a downsampled version of I by a factor $r > 1$
- Generation starts at the coarsest scale up to the finest with noise injection at each scale.
- $$n < N, \tilde{I}_n = (\tilde{I}_{n+1}) \uparrow^r + \Psi_n(z_n + (\tilde{I}_{n+1}) \uparrow^r)$$

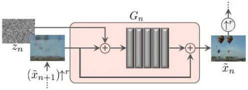


Figure: Single scale generation [4]

Model Description - SinGAN

- The loss used for training the GANs comprises an adversarial term and a reconstruction term.
- $$\min_{G_n} \max_{D_n} L_{adv}(G_n, D_n) + \alpha L_{rec}(G_n)$$
- The reconstruction loss ensures the existence of a specific input noise maps, which generates the original image I.
- $$L_{rec} = ||G_n(0, (\tilde{I}_{n+1})^{rec} \uparrow^r) - I_n||^2$$

Outline

Random Samples

- We trained the SinGAN model on two natural images, and we generated random image samples [3]
- The generated images preserve the visual content of the training image, and generate new combinations of patches that do not exist in the training image.

Training image

Random Sample 1

Random Sample 2

Random Samples

Training image

Random Sample 1

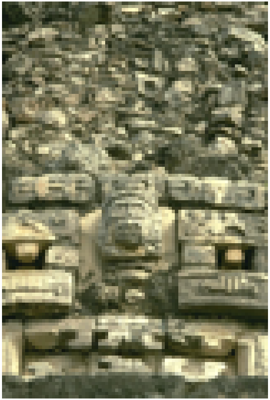
Random Sample 2

Random Sample 3

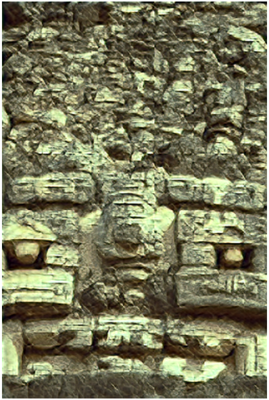
Super-Resolution

- We first train the SinGAN on the LR image , with a reconstruction loss weight $\alpha = 100$
- Then we upsample the LR image by **r** and feed it to the last generator $\mathbf{G}_0$.

LR Image



HR Image



Outline

Problematic

- Inpainting is a process of restorative conservation, where damaged, deteriorating, or missing parts of an artwork are reconstructed, ultimately with the goal of presenting the work as it was originally created (wikipedia)
- There exists in the Litterature many attempts to perform inpainting using GAN, mainly:
  - Free-Form Image Inpainting with Gated Convolution, 2018
  - Generative Image Inpainting with Contextual Attention, 2018

Outline

Approach

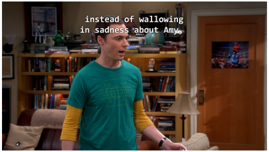
- Assume that we already have a reference image which is quite similar to the image to be inpainted.
- Train on the Image with subtitle or occluding object using the SinGAN with a reconstruction coefficient of 70.
- Use the editing script providing it with mask of the occluding object and the reference image.

Outline

Inpainting

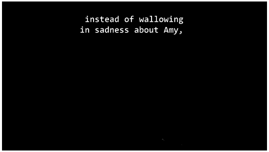
We follow the approach described before and try to apply inpaintaing to some input images with occluding objects, and try to return the clean images.

Occluded image



Reference Image

Mask



Output Image

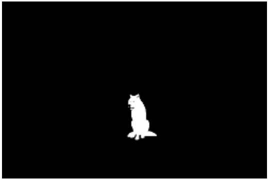
Results

We tried to extend this approach of inpainting to other types of object (not only subtitles). The pictures below show the output we got.

Occluded image

Reference Image

Mask



Output Image

Results

Occluded image

Reference Image

Mask

Output Image

Outline

Quantitative Evaluation

- Traditionnaly the Inpainting process is evaluated by humans subjectively.
- We focus on two metrics [1]:
  - Structural Similarity Index (SSIM), $ssim(x, y) = \frac{(2\mu_x\mu_y + c_1)(2\sigma_{xy} + c_2)}{(\mu_x^2 + \mu_y^2 + c_1)(\sigma_x^2 + \sigma_y^2 + c_1)}$
  - Average Squared Visual Salience (ASVS), is a non reference (NR) metric.

$$ASVS = \frac{\sum_{p \in \Omega} |S^{'}(p)|^2}{||\Omega||}$$

Quantitative Evaluation

Image	SinGAN	NVIDIA[2]
Movie	0.93	0.92
Forest	0.97	0.90
Mountains	0.98	0.73

Table: SSIM

Image	SinGAN	NVIDIA
Movie	0.012	0.02
Forest	0.05	0.06
Mountains	0.06	0.12

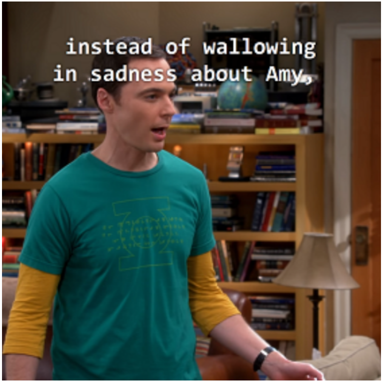
Table: ASVS

Qualitative Evaluation

We compared our model with the Nvidia model for inpainting.

https://www.nvidia.com/research/inpainting/

Input Image



Output Image

Qualitative Evaluation

Input Image

Output Image

Qualitative Evaluation

Input Image

Output Image

Conclusion

- Inpainting using SinGAN works great when given a good reference Image
- It would be interesting to investigate usage of a contextual layer to do inpainting during training.

Bibliographie I

Azeddine Beghdadi Muhammad Ali Qureshi, Mohamed Deriche and Asjad Amin.
A critical survey of state-of-the-art image inpainting quality assessment metrics.
In *Journal of Visual Communication and Image Representation*, pages 177–191, 2017.

NVIDIA.
https://www.nvidia.com/research/inpainting/.

Tamar Rott Shaham.
https://github.com/tamarott/singan.
2019.

Bibliographie II

Tali Dekel Tamar Rott Shaham and Tomer Michaeli.
Singan: Learning a generative model from a single natural image.
In *Advances in neural information processing systems*, pages 5998–6008, 2019.