Team #2 - Daniel Yohan & Lilach Cohen

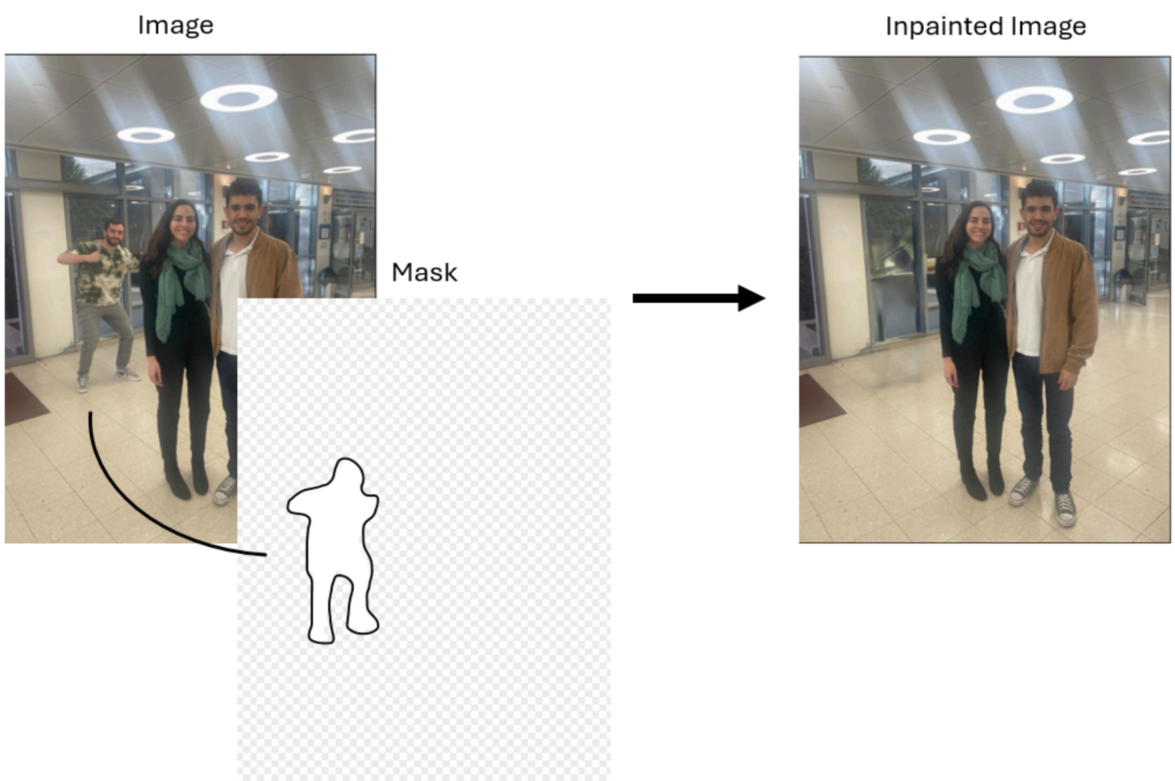
Image Inpainting

I/O Framework:

Input: Original images and corresponding masks.

Output: Inpainted images with alternative synthesized content.

Example (using our model) :



Technology Utilized:

A pretrained model on “places2” and “celebA” datasets as published in the paper

“Free-Form Image Inpainting with Gated Convolution”

[https://paperswithcode.com/paper/free-form-image-inpainting-with-gated]

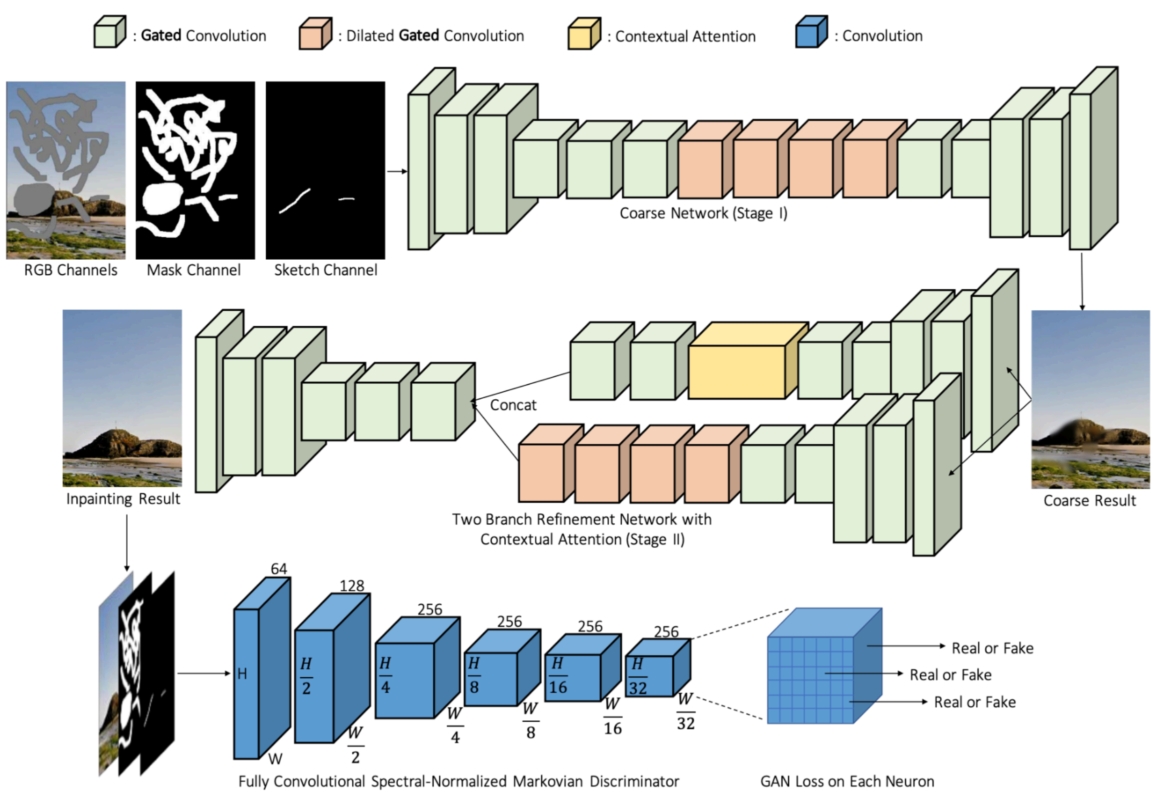
Functionality:

Image inpainting - Replacing an “empty mask” with a realistic generated background that mimics the natural appearance of the original image.

Model Type:

Spectral-Normalized Patch GAN. GAN consists of two neural networks - a generator and a discriminator that are trained simultaneously through adversarial training. The generator creates synthetic data samples, while the discriminator evaluates whether the generated samples are real or fake.

Model overview:



The generator is made of Coarse and Refinement networks, each constructed as a simple encoder-decoder utilizing Gated Convolutions. In these convolutions, the missing pixels (of the mask) are considered invalid, whereas the rest of the image is valid. The convolution is masked and applies only on valid pixels. The mask isn’t necessarily a binary matrix, and is learned in iterations. Some layers consist of Dilated Gated Convolutions. It learns a dynamic feature gating mechanism for each channel and each spatial location like inside or outside masks and RGB channels.

The discriminator is a convolutional network. Its input is an output image from the generator, alongside the mask. It operates on Markovian patches within the input data that differ in locations and semantics. This formulation enables the discriminator to simultaneously evaluate numerous GANs, each targeting distinct locations and semantics within the input data. The approach enhances the discriminator's ability to provide precise feedback to the generator.

Loss Function: A balanced ratio of 1:1 between two types of loss is maintained: Hinge loss is computed for each point on the output map for the discriminator, and pixel-wise $\ell 1$ reconstruction loss is applied.

Contextual Attention: A layer that captures long-range spatial dependencies. Partial convolution is integrated in it to better handle free-form masks and valid pixels only.

Workflow:

We embarked on creating a custom generative model but then realized that the computational complexity was not feasible. We later transitioned to using a pretrained model.

We selected a pretrained model ranked #3 on the Places2 dataset, specifically tailored for inpainting free-form patches. Ensuring alignment with our project's objectives, we reviewed the paper the model is based on.

Later we conducted testing using images and masks sourced from the COCO dataset. After receiving team #1's masks we adjusted their format and improved them using classical computer vision techniques, and ran their output through our pipeline.

Results:

Inpainting Results Using the Original Dataset Masks:

Overall, the model produced satisfactory results in this study. While the majority of the inpainted images were good quality, there were a few repeated problems:

1. In some instances the inpainted image failed to maintain consistency, because certain objects that were attached to the person in the picture were excluded from the mask. For instance, objects such as person shadows or remnants of skates were left behind in the image, inadvertently revealing the absence of the removed person.

2. The model demonstrated superior performance when dealing with natural backgrounds devoid of distinct lines or shapes. In such scenarios, the inpainting process appeared more realistic and seamless.
3. However, when applied to more complex backgrounds with intricate details or distinct patterns, the model's performance was notably diminished. In these

cases, the inpainting results were less satisfactory and often struggled to seamlessly integrate the inpainted regions.

4. A few images exhibited issues related to color reproduction in the inpainted areas, highlighting a potential limitation of the model in accurately replicating colors.
5. many images exhibited smudged inpainted sections, indicating that the model struggled to accurately mimic textures and details. This aspect detracted from the overall visual quality of the results.

Inpainting Results Using team #1 masks :

When applied with the output received from Team #1, the model yielded mediocre results. However, significant improvements were achieved by applying code that thickened and smoothed the masks. The primary issue stemmed from incomplete masks, where parts of the pixels holding the person segment remained untouched.

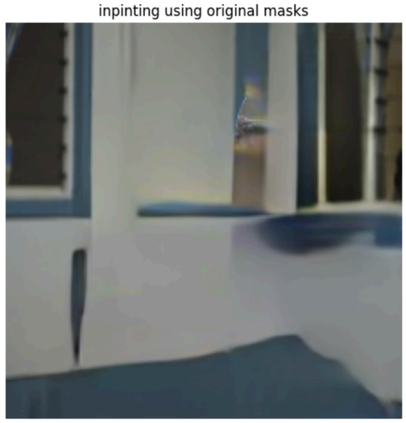
original image

inpinting using original masks

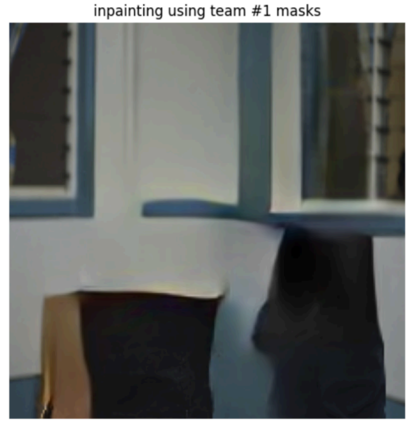
inpainting using team #1 masks

original image

inpinting using original masks



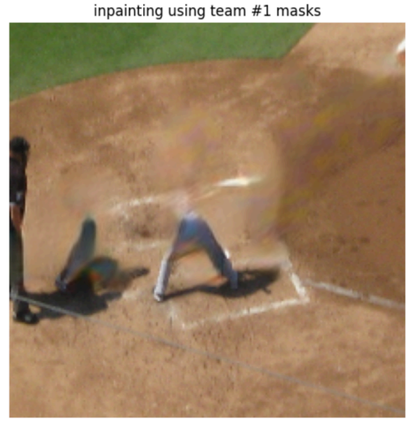
inpainting using team #1 masks



original image

inpinting using original masks

inpainting using team #1 masks

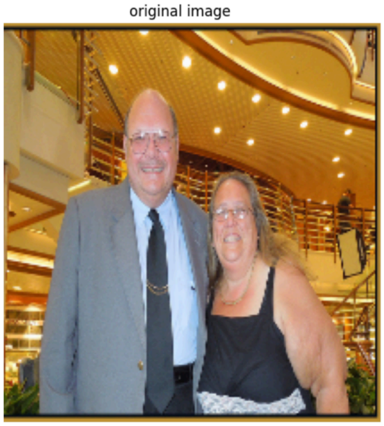


original image

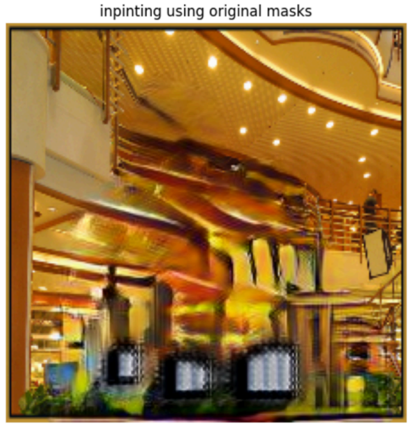
inpinting using original masks

inpainting using team #1 masks

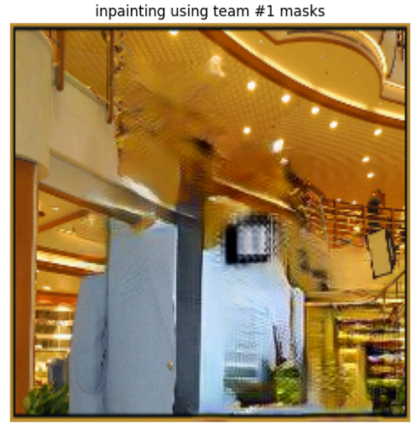
original image



inpinting using original masks

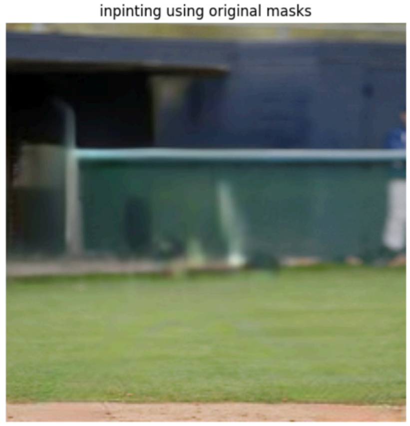


inpainting using team #1 masks



original image

inpinting using original masks



inpainting using team #1 masks

Consequences:

The results observed with Team #1's masks can be attributed to several factors. It's evident that the model's performance is influenced by the quality and resolution of the masks used during training. Team #1's masks may have been less accurate or detailed compared to those used in our training dataset. Additionally, the person segment size (the model works better as the person segment is smaller) may have contributed to the disparity in results.

Considering the relatively simple architecture of our model, the obtained results are considered very good. However, there are evident areas for improvement, particularly in addressing the detailed issues such as texture replication and maintaining consistency in complex scenes.

Moving forward, the next steps for our project involve making adjustments to the model to address these detailed issues. Furthermore, exploring alternative models, such as those based on stable diffusion, could offer promising avenues for achieving enhanced results. By incorporating these adjustments and exploring alternative methodologies, we aim to further improve the performance and versatility of our inpainting system.