Image Completion

Manikanta B, Geetha Charan Y
DEPARTMENT OF COMPUTER SCIENCE AND AUTOMATION

AIP 2023

Introduction

- Image completion(or image inpainting) is a technique that allows filling in target regions with alternative matching contents.
- The ability to complete the missing regions comes from the fact that natural images, despite their diversity, and are highly structured.
- One of the main motivations of image completion is being able to remove unwanted objects in images.
- The approaches need to be able generate fragments that may not appear else where in the image.

Original Image

Masked Image

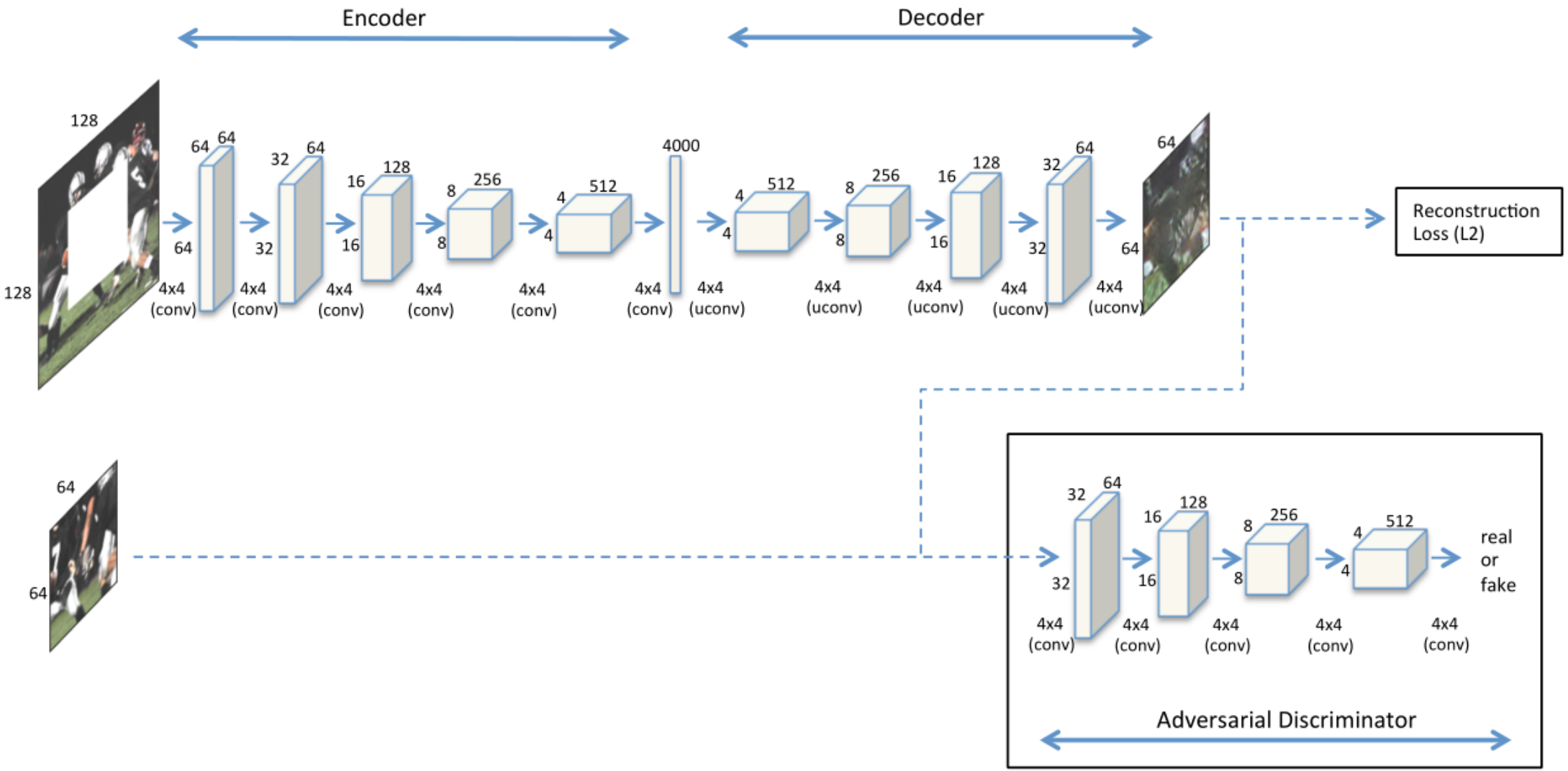
Reconstructed Image

Context Encoders - Introduction

Image completion using Context Encoders :

- In this approach, the model takes an image as input with a missing region and then it is trained on a convolutional neural network to regress to the missing pixel values.
- This model is called context encoder.
- It consists of an encoder capturing the context of an image into a compact latent feature representation and then a decoder which uses that representation to produce the missing image content.

Architecture



Architecture

- The overall architecture is a simple encoder-decoder pipeline.

Encoder :

- The proposed architecture of Encoder takes an input image of size 128 × 128 with missing central region of size 64 x 64, we use five convolutional layers, wherein each layer contains a convolution operation, LeakyReLU and Batch Normalization in the same order.
- The final output of encoder is a bottleneck of 4000 units.

Decoder :

- The bottleneck layer is followed by a series of five up-convolutional layers with learned filters, each with a rectified linear unit (ReLU) activation function.

Adversarial Discriminator :

- GANs are incorporated into the image completion model for quality image generation.
- Generator is modelled as a context encoder and discriminator tries to distinguish ground truth central region and inpainted central region.
- We conditioned only the Generator on the image containing missing region.

Architecture

Loss Functions :

- The overall loss function is a weighted combination of Reconstruction Loss and Adversarial Loss.
- Reconstruction Loss : The reconstruction (L2) loss is responsible for capturing the overall structure of the missing region.

$$\mathcal{L}_{rec}(x) = \|\hat{M} \odot (x - F((1 - \hat{M}) \odot x))\|_2^2,$$

- Adversarial Loss : The adversarial loss tries to make prediction look real, and ensures to have continuity within the missing region.

$$\mathcal{L}_{adv} = \max_D \mathbb{E}_{x \in \mathcal{X}}[\log(D(x)) + \log(1 - D(F((1 - \hat{M}) \odot x)))]$$

- Overall Loss : $\mathcal{L} = \lambda_{rec}\mathcal{L}_{rec} + \lambda_{adv}\mathcal{L}_{adv}.$

Results

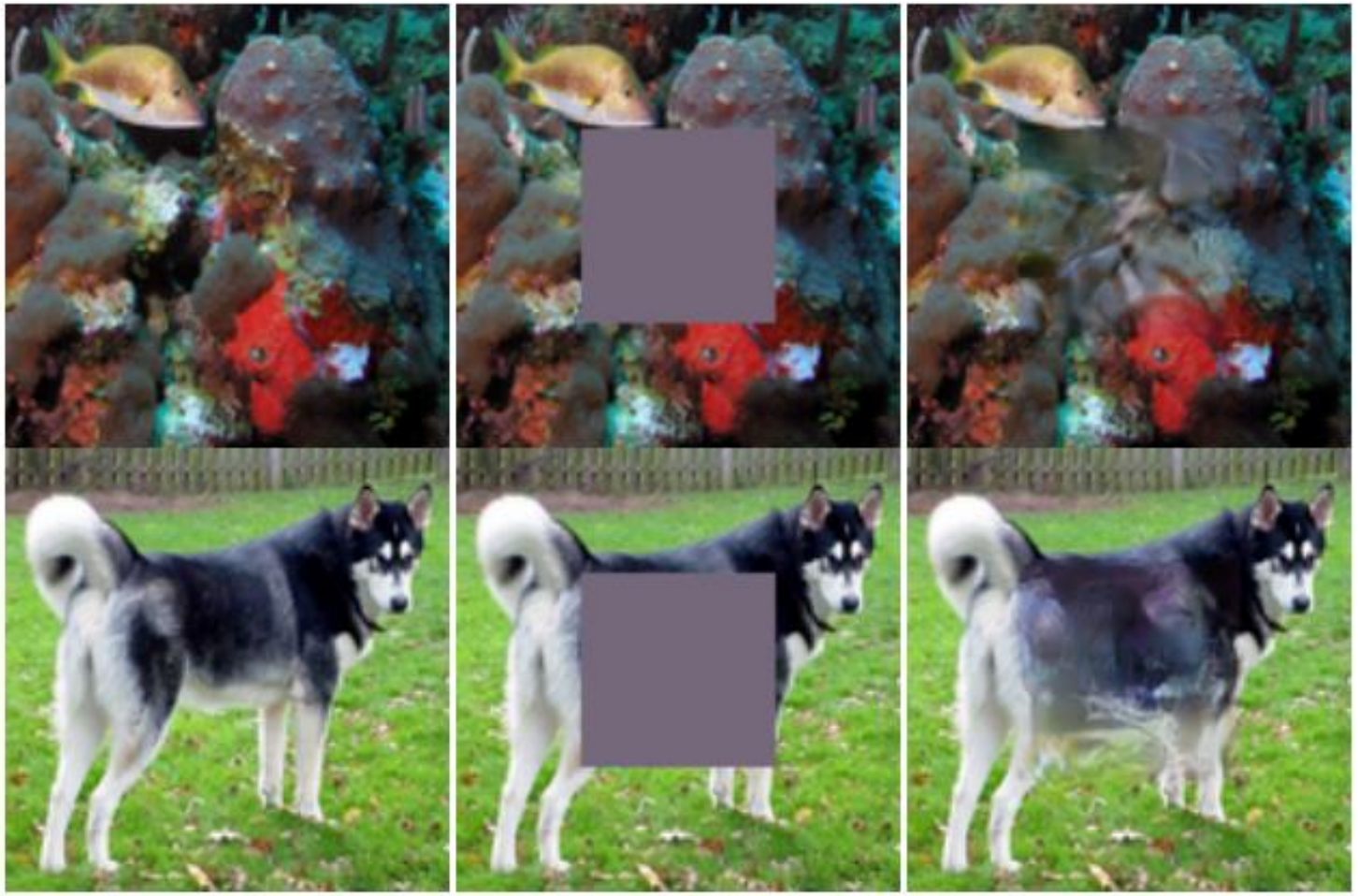
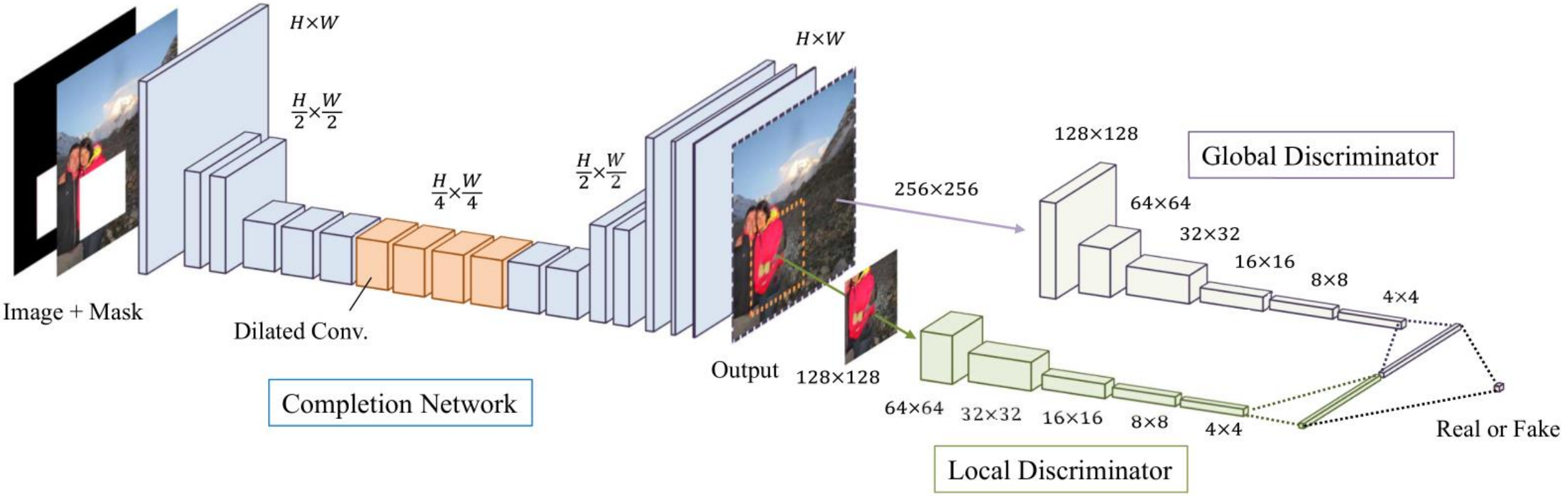


Figure 5: (a) Ground Truth Image (b) Masked Image (c) Reconstructed Image generated from trained Context Encoder Model

Globally and Locally Consistent Image Completion

- This approach builds upon the Context Encoder approach and addresses completing arbitrary inpainting regions.
- It also addresses how to complete missing regions of high resolution images.
- The proposed architecture is composed of three networks:
  - Completion network
  - Global context discriminator and
  - Local context discriminator.

Architecture



Architecture

Completion Network :

- The completion network is fully convolutional and used to complete the image.
- Dilated Convolution layers are incorporated into completion network to increase the receptive field.

Global and Local Discriminators:

- The global discriminator takes the full image as input to recognize global consistency of the scene.
- The local discriminator looks only at a small region around the completed area in order to judge the quality of more detailed appearance.

Architecture

Loss Functions :

- The overall loss function is a weighted combination of MSE and GAN Loss.
- MSE Loss : $\mathcal{L}(x, M_c) = \|M_c \odot (C(x, M_c) - x\|^2$
- GAN Loss : $\min_C \max_D \mathbb{E}[\log(D(x, M_d)) + \log(1 - D(C(x, M_c), M_c)]$
- Overall Loss :

$$\min_C \max_D \mathbb{E}[L(x, M_c) + \alpha \log(D(x, M_d)) + \alpha \log(1 - D(C(x, M_c), M_c)]$$

Results

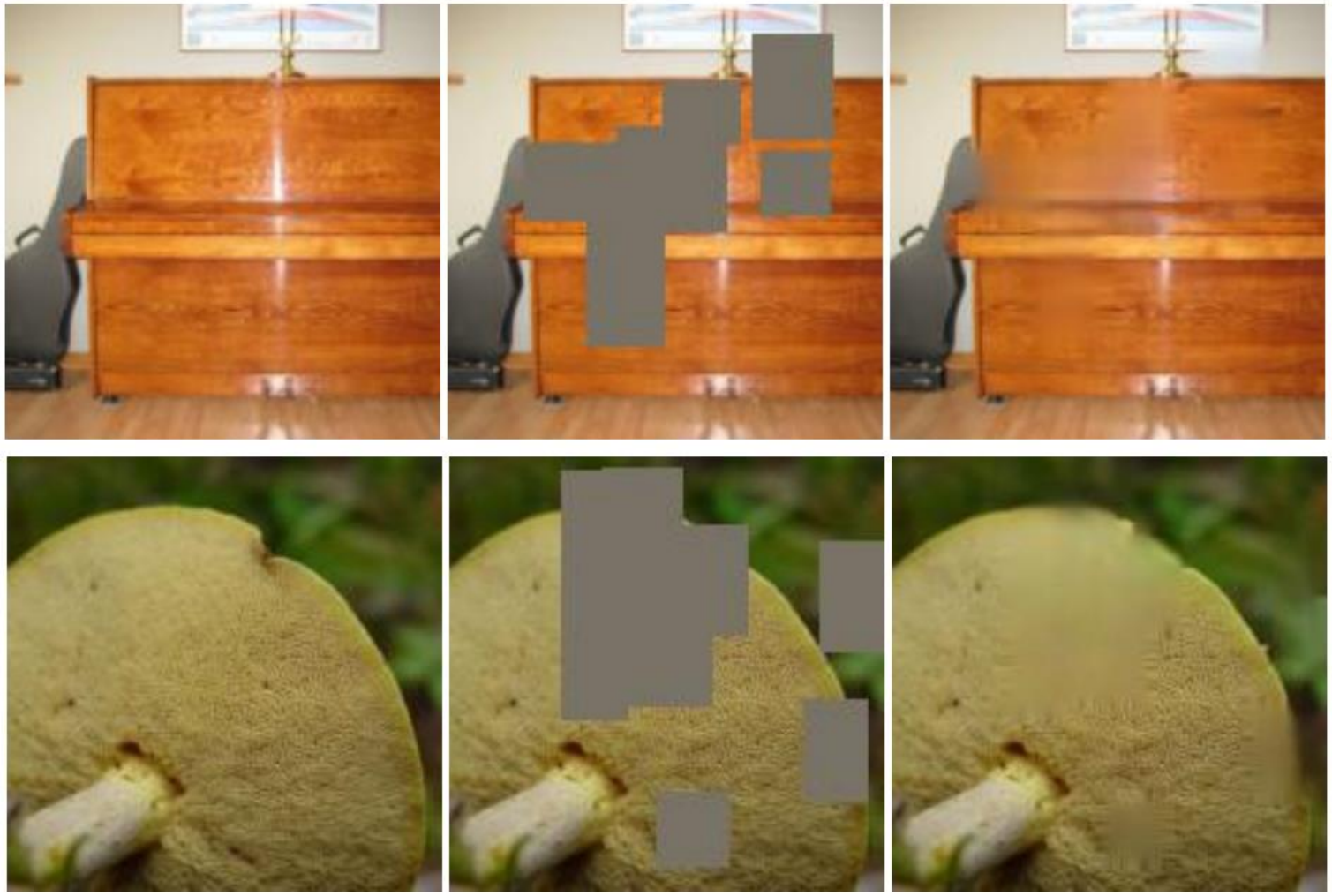


Figure 7: (a) Ground Truth Image (b) Masked Image (c) Reconstructed Image generated from trained Globally and Locally Consistent Image Completion Model

Comparison

a) Ground Truth
b) Masked Image
c) Reconstruction (CE)
d) Reconstruction (GL)

Observations

- Context Encoder is performing better when the neighborhood of the missing region is smooth.
- Completion Network is able to figure out complex missing regions and fill out in a visually plausible way due to the combination of local and global discriminator in action.
- Significantly large holes cannot be filled-in by both the approaches due to the spatial support of the model.