Image Inpainting and Restoration

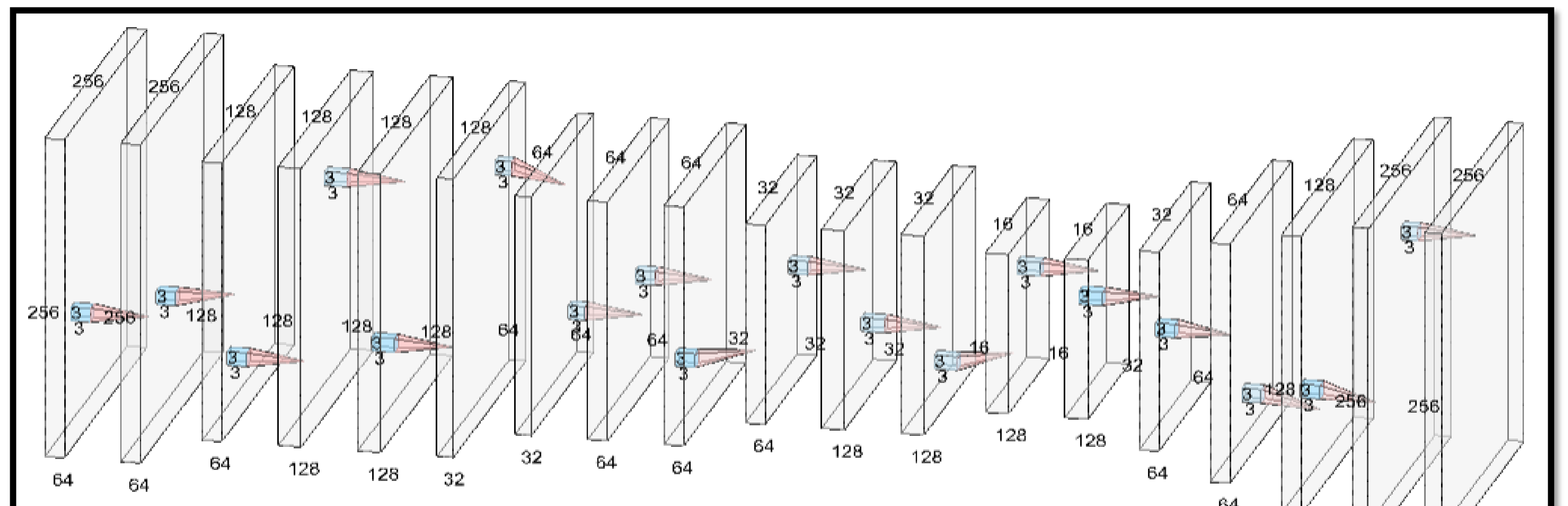
Meshwa Savalia (16BIT114), Guided By: Prof. Chirag Nathwani, Nirma University

1 Abstract

Image inpainting and restoration is the process to reconstruct the missing or damaged portions of the image, in order to make it semantically and visually believable. We have used GAN architecture for this problem. We propose a new training strategy to trigger the model to learn different types of corruptions. In addition, our model is lightweight and requires much less model parameters to achieve superior results.

2 Network Architecture

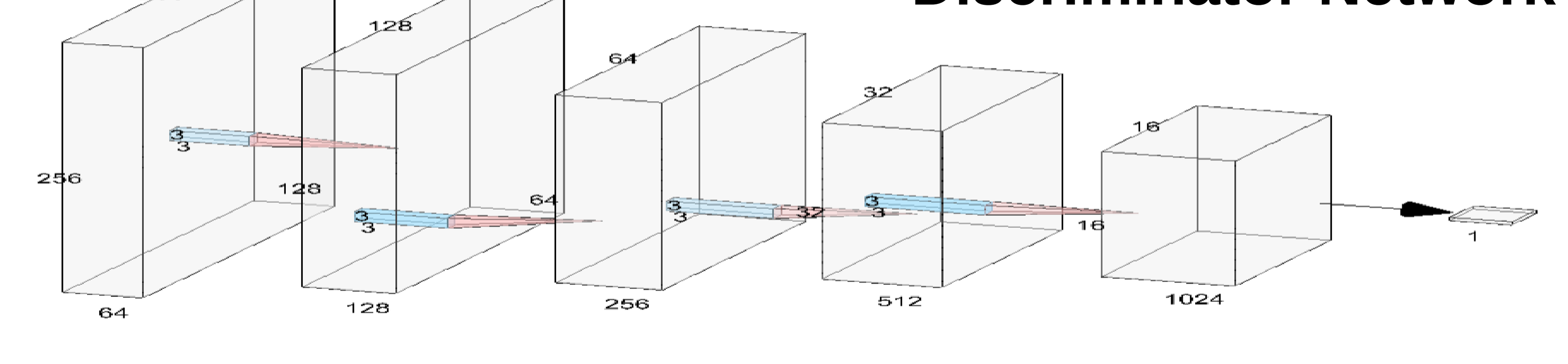
Our network architecture contains two main networks, Generator and Discriminator. Generator is used to generate novel images from the corrupted images. It encodes semantic features of the corrupted image and then generates a novel image using that encoded semantic patch. The discriminator aims evaluating whether the repaired result is correct and consistent or not.



Generator Network

Input Size: 256 x 256
Output Size: 256 x 256
Layers: 15 Con2D, 7 Con2DT
Activation: Leaky ReLU
Parameters: 8 M

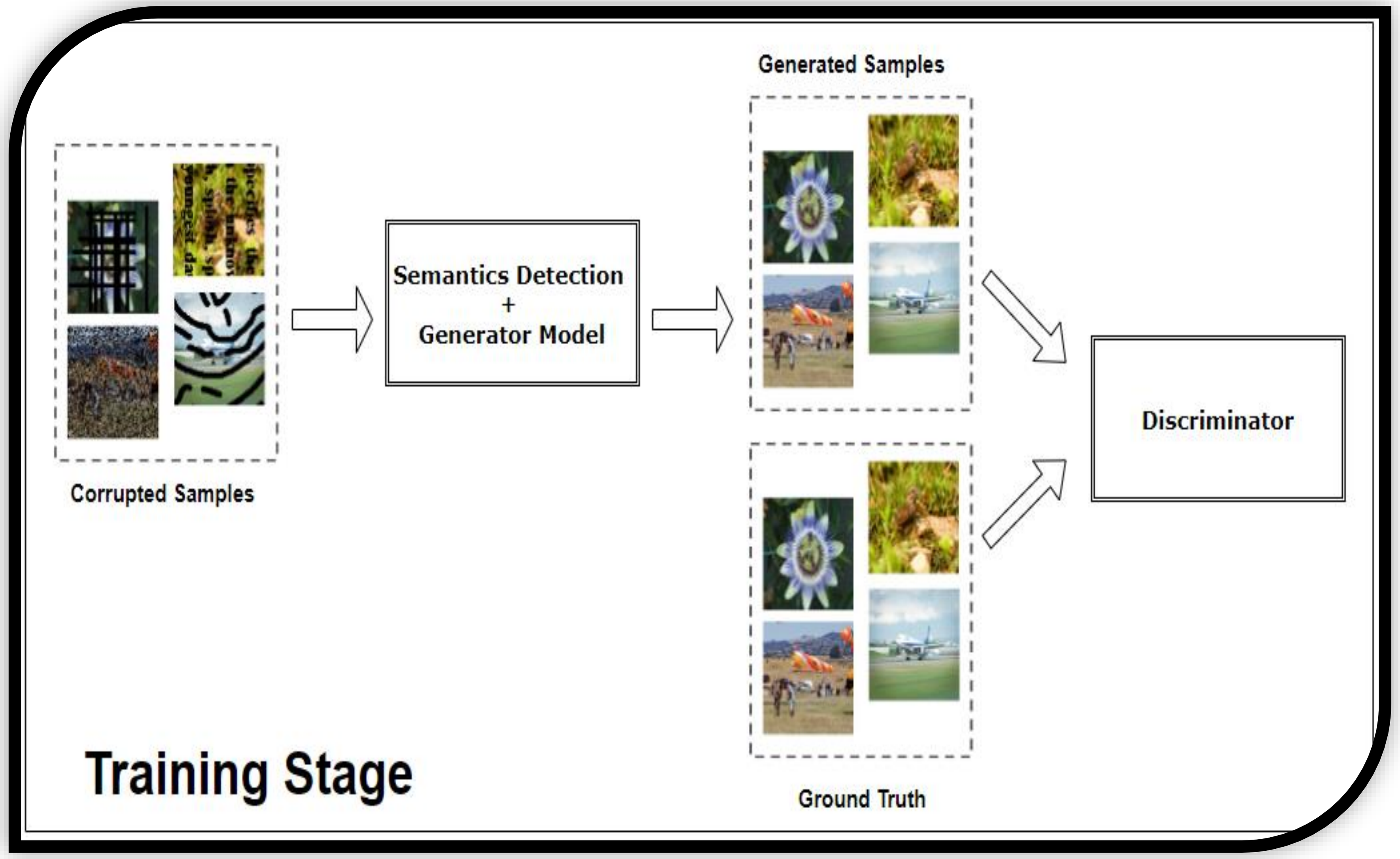
Input Size: 256 x 256
Output Size: 1 (F/R)
Layers: 5 Con2D, 1 FC
Activation: Leaky ReLU
Parameters: 6.5 M



Discriminator Network

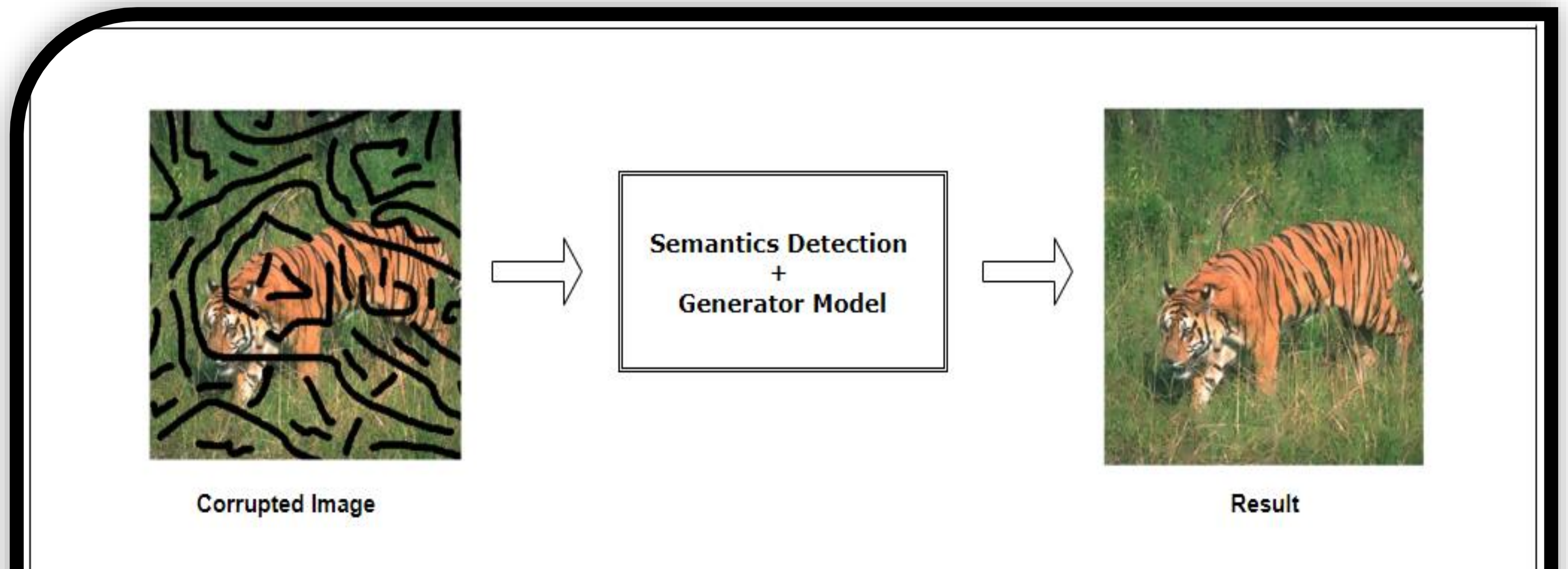
3 Proposed Approach

We have implemented GAN architecture for our project. We have used images from three datasets (CelebA, Places2, MS-COCO) for training. In order to increase the generality of our model to recover various types of missing regions from the images, we have prepared a diverse mask dataset for training which includes four common types of corruption masks i.e. random noise, textual matter, strokes and scrawls.



Training Stage

For training stage, G and D models are trained using adversarial settings. We apply corruption masks to images and G tries to inpaint it. Generated samples and original samples are fed to discriminator to calculate loss.



Testing Stage

For testing, we provide a corrupted image to the G for inpainting the image

4 Our Results

We demonstrate qualitative and quantitative results on our different generated training masks which are textual matters, strokes, scrabble and random noise. We use two of the image quality analysis metrics:

1. PSNR (Peak Signal to Noise Ratio)
2. SSIM (Structured Similarity Index)

We aim for higher PSNR and SSIM value as it will suggest good image quality. The below table depicts the results of our model.

Testing Mask Type	PSNR Value	SSIM Value
Textual Matter	30.75	0.844
Random Noise	29.09	0.931
Strokes	28.34	0.909
Scrawls	27.77	0.811
Average	**28.98**	**0.873**

We compare our model with other deep model in below table. Our model contains a lesser number of training parameters hence it is computationally more efficient.

Methods / Parameters	RED-Net	CE	Pix2pix	Ours
Adv. Training	No	Yes	Yes	Yes
No. of Layers in G	10	12	16	22
No. of Layers in D	--	4	4	5
Trainable Parameters	0.3M	71M	57M	15M

5 Conclusion

In this work, we show that our deep completion model and the proposed training strategy can provide superior image completion performance quantitatively and qualitatively on different dataset. The proposed lightweight deep networks can successfully generate stable and semantic image completion results and outperform previous methods.