Kaggle Photo Reconstruction Challenge Final-Report

Group-32

Sombit Bose

Neeraj Boddeda

Unnikrishnan M

Sri Bharat

Introduction:

Image inpainting is the process of filling in missing or corrupted parts of an image based on the surrounding context. It has many practical applications in fields such as image editing, computer vision, and medical imaging. One of the most effective methods for image inpainting is the Pix2pix GAN, which uses a combination of adversarial loss and pixel-wise loss to generate realistic and accurate completion of the missing or corrupted parts.

Our Methodology:

We first tried using a pre-trained stable diffusion model from stabilityai/stable-diffusion-2-inpainting. Since this is a pretrained model and performed very little training with the dataset provided. When we tried to predict the masks using this Model, we surprisingly got some good images as output but limited only to few. We then used the pix2pix GAN (Generative Adversarial Networks) model. For discriminator in the model we used a deep CNN and trained it by taking cross entropy loss and for Generator model we tried U-net architecture.

Pix2pix GAN is an effective and efficient method for image inpainting. It uses a combination of adversarial loss and pixel-wise loss to generate realistic and accurate completion of missing or corrupted parts of images. Pix2pix GAN has shown to produce high-quality results and outperform other state-of-the-art methods for image inpainting.

To use Pix2pix GAN for image inpainting, the generator network is trained on pairs of input images with missing or corrupted parts and ground truth images without any missing or corrupted parts. The generator network is shown the input image and asked to generate an output image that matches the ground truth image. The generator's output is then compared to the ground truth image using the adversarial loss and pixel-wise loss, and the generator's parameters are updated to minimize these losses.

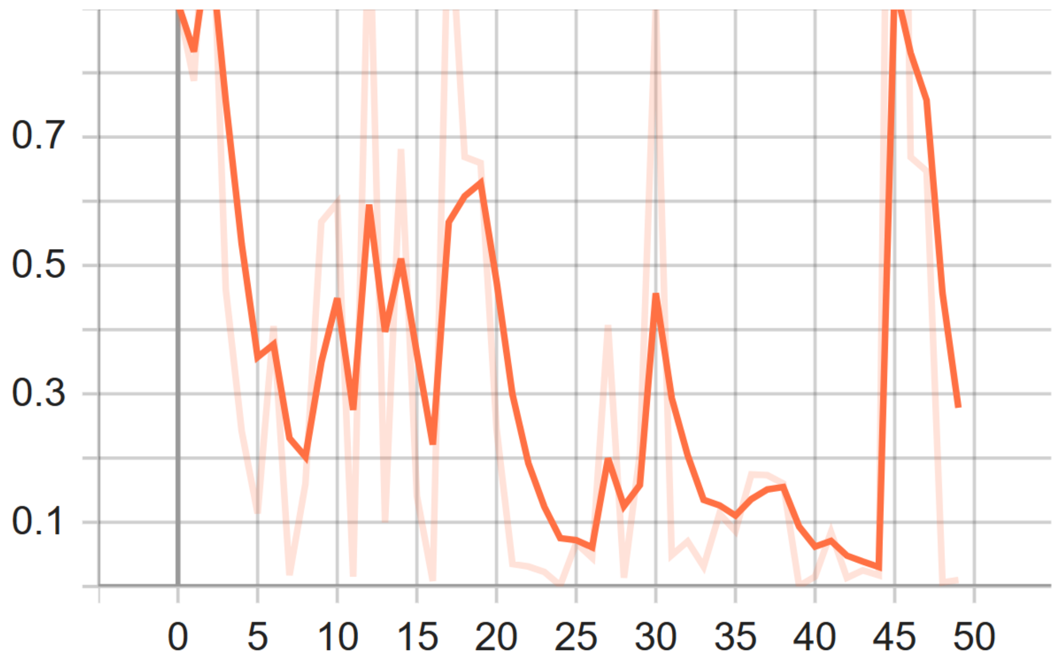
Pix2pix GAN is a type of generative adversarial network that consists of two main components: a generator network and a discriminator network. The generator network is responsible for generating a completed version of the input image, while the discriminator network tries to distinguish between the generated images and the ground truth images. During training, the generator network is optimized to produce output images that are as close as possible to the ground truth images, while the discriminator network is optimized to correctly classify the generated images as fake.

The generator takes an input image of size (256, 256, 3) and passes it through a series of downsample and upsample blocks. The downsample blocks reduce the spatial resolution of the image while increasing the number of feature channels. The upsample blocks increase the spatial resolution of the image while decreasing the number of feature channels. The downsample and upsample blocks are connected by skip connections. These connections allow the generator to access high-resolution details from the input image while also incorporating information from the downsampled feature maps. The generator outputs an image of size (256, 256, 3) with pixel values in the range [-1, 1]. The activation function used in the last layer is hyperbolic tangent (tanh), which maps the pixel values to the desired range.

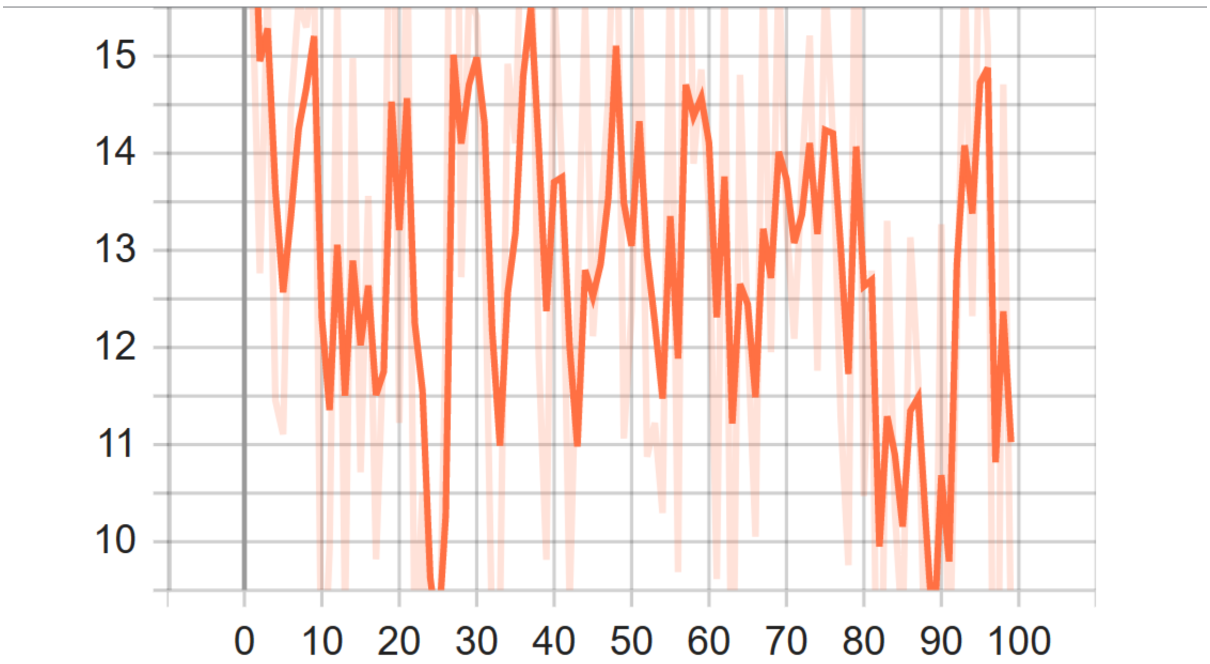
The discriminator takes two inputs: the original masked image and the predicted image (the output of the generator). These two images are concatenated along the channel axis and fed into the discriminator network. The discriminator network consists of a series of convolutional layers followed by batch normalization and LeakyReLU activation. The input images are downsampled through these convolutional layers until the output has a resolution of 30x30. The output of the final convolutional layer is a single-channel image with a resolution of 30x30. This output represents the discriminator's prediction of whether the input image pair (original masked image and predicted image) is real or fake. The discriminator is trained to distinguish between real and fake image pairs. The generator is trained to create fake images that are realistic enough to fool the discriminator into thinking they are real. The two networks are trained together in an adversarial manner, with the goal of improving the generator's ability to generate realistic images and the discriminator's ability to distinguish between real and fake images.

Plots from Tensorboard:

Dicriminator Loss vs no of epochs:



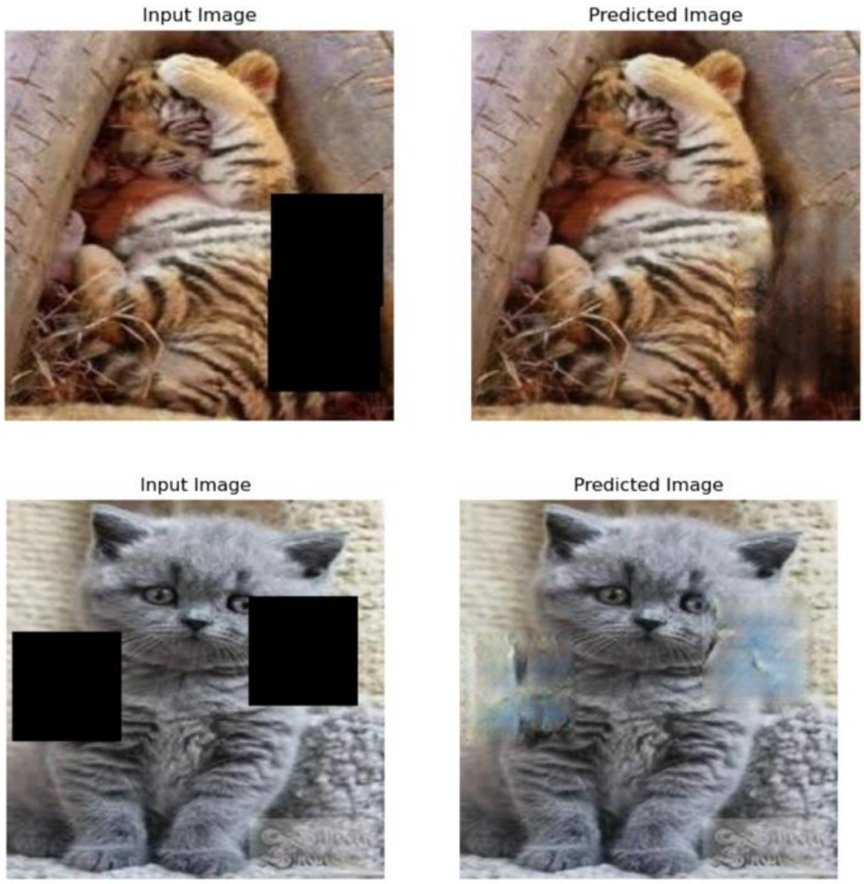
Generator Total Loss vs no of epochs:



Results:

Pix2pix GAN has been shown to produce high-quality results for image inpainting tasks. In a study conducted by researchers at NVIDIA, the Pix2pix GAN was able to generate realistic and visually pleasing completion of missing or corrupted parts of images. The study also demonstrated that Pix2pix GAN outperformed other state-of-the-art methods for image inpainting in terms of both visual quality and quantitative metrics.

Our model performed very well and that can be deduced from our outputs.



The above predicted images are generated by our model and are very accurate. Our score in Kaggle is 0.23463.