Image Super-Resolution using Generative Adversarial Networks

DInesh Adhithya
IISER B

Introduction

The aim of this work is to use deep learning techniques to improve the resolution of images used for semantic segmentation tasks. CNN (Convolutional neural network) based models outperform previously developed models and are a good choice for this work as well.Recently GAN based models have been proposed for this task and make use of conventional RMSE reconstruction error and adversarial loss. GAN's make use of a discriminator and Generator approach to construct high resolution images from low resolution images.This work makes use of residual networks , GAN's and transfer learning approaches to upsample images of high resolution from low resolution images.

Initial work on such tasks was carried out but minimizing the root mean square error between upscaled low resolution image and high resolution image , this in turn maximizes the PSNR metric. The PSNR metric inturn tries to over smooth images which makes it unusable for such tasks as the loss function.So we introduce an adversarial loss which is a min-max game between generator and discriminator.

Approach

In this problem we have a low resolution image I^{LR} and a high resolution Image I^{HR} , the generator estimates I^{HR} from I^{LR} by using a parameterized function θ_G which is constructed by CNN'S with residual layers to ensure feature transfer through the network , makes use of pixel shuffling to upsample the image to obtain the high resolution image namely called I^{HR} , we define a loss function namely l^{SR} , which can be minimized by using backpropagation.

$$\hat{\theta}_G \equiv argmin \frac{1}{N} \sum_{n=1}^{N} l^{SR}(G_{\theta_G}(I^{LR}), I^{HR'})$$

Where $I^{HR'}$ is the high resolution image , I^{HR} is the predicted high resolution image.

The discriminator network D_{θ_D} is optimized along with G_{θ_G} as a min-max problem

$$min\, max\, E\, I^{HR} \sim ptrain(I^{HR})\, [\, log\, D_{\theta_D}(I^{HR}) \; + \; E\, I^{LR} \sim ptrain(I^{HR})\, [\, log\, 1 \; - \; D_{\theta_D}(G_{\theta_G}(I^{LR}))\;]$$

The discriminator’s task is to differentiate the high resolution image produced by the generator from the high resolution image used in training.THis approach makes sure that the image produced by generator is similar to real world images.

The Generator B residual blocks each consisting of 64 convolution layers with kernel size 3*3 , then batch normalization is done followed by a ReLU activation function. Image resolution is increased by using one or two sub-pixel layers.

The discriminator has 8 layers each with kernel size 3*3 , with layer depth increasing from beginning to end , from 64 to 512.

The loss function is defined the sum of adversarial loss and content loss:

$$l^{SR} = \; l_{Content}^{SR} + 10^{-3} * \; l_{Adversarial}^{SR}$$

The content loss is pixel-wise between the super resolved image and the original image.

Experiments and Results

Data and Metrics

The training data will comprise of the RELLIS-3D dataset and it contains about 17779 images and for model testing , we shall use the RUGD dataset which is of low resolution and we want to upgrade it to a higher resolution , we shall downsample using bilinear interpolation for obtaining low resolution images , the PSNR metric shall be used to quantify the performance of various models.

$$PSNR = 20 * log_{10}(255) - 10 * log_{10}(MSE)$$

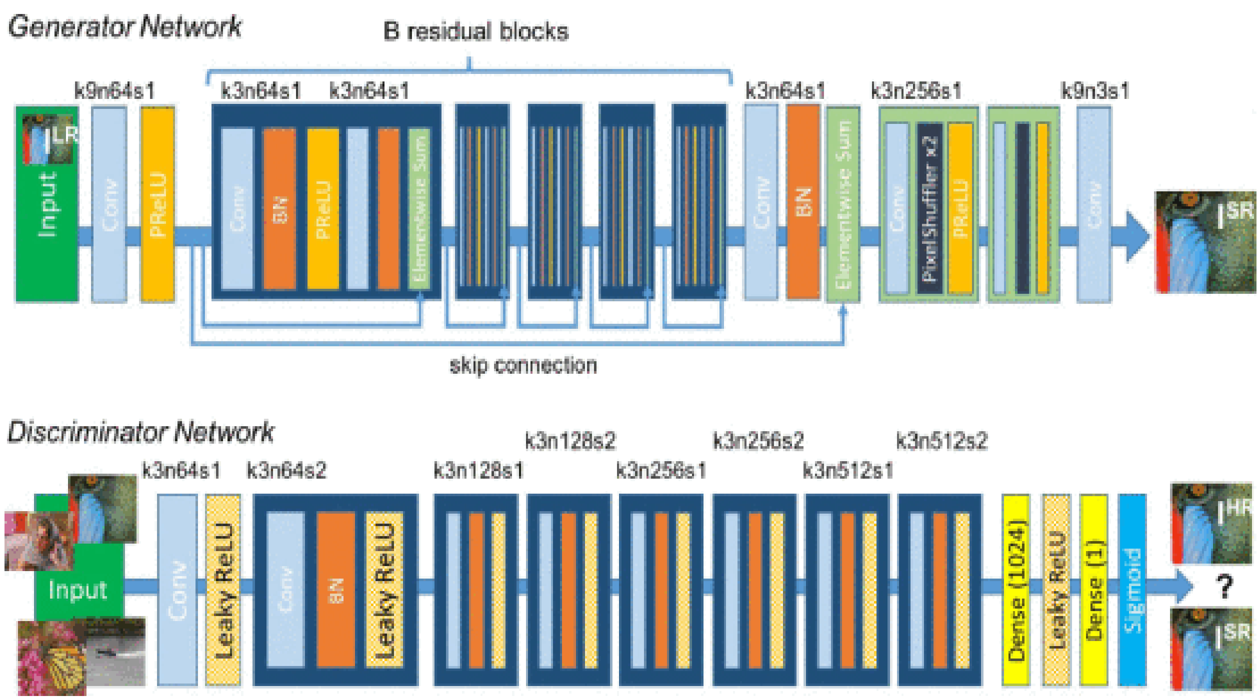

Figure 1 : SRGAN model architecture

Model Training

The models were trained on a NVIDIA GEFORCE RTX 3060. For optimization both for generator and discriminator Adam was used. Due to hardware constraints the images were trained with batch size as 1 In each batch training the generator was done first then the generator was trained . Initially images of low resolution size 100*100 were used to resolve them into 400*400 size high resolution image.Later low resolution images of size 900*700 was used for high resolution images of size 1920*1400.

Model Performance

This approach makes use of the SResNet model for model prediction with an PSNR score of 25.04 and final loss at 0.054 during training.

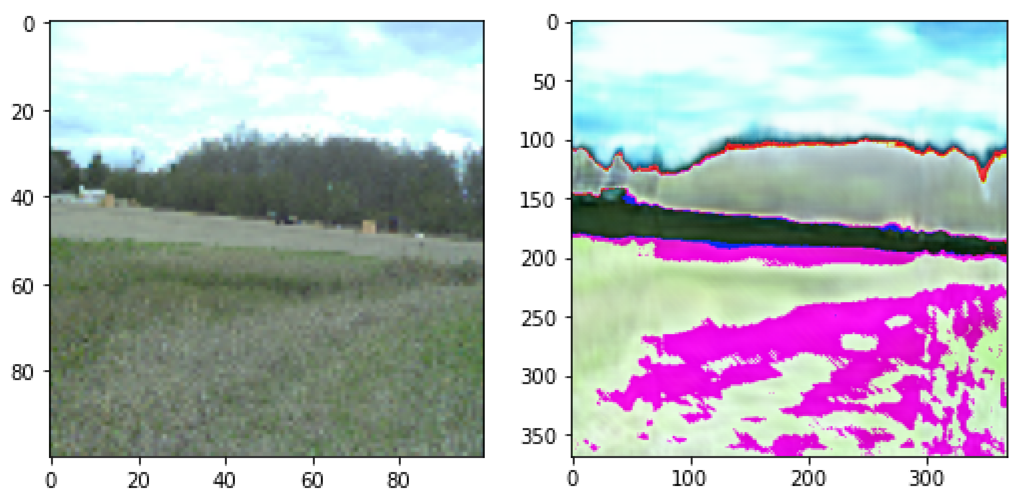

Figure 2: Image showing low resolution image(left) and high resolution output obtained from the generator.

This approach makes use of SR GAN without the adversarial loss
On the 100*100 low resolution images dataset the model performed with PSNR of 28.98 , with final loss at 0.00877.The below image shows the results of the model.

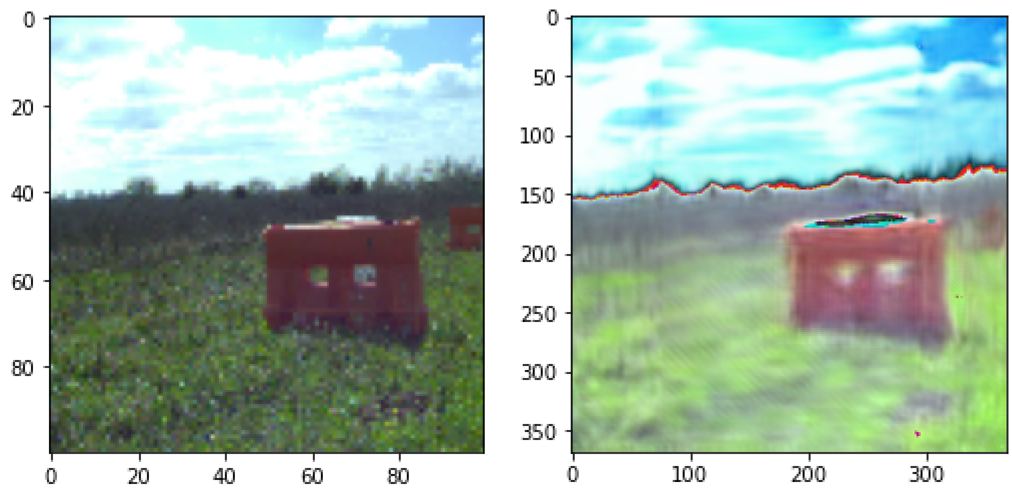

Figure 3: Image showing low-resolution image(left) and high-resolution output obtained from the generator.

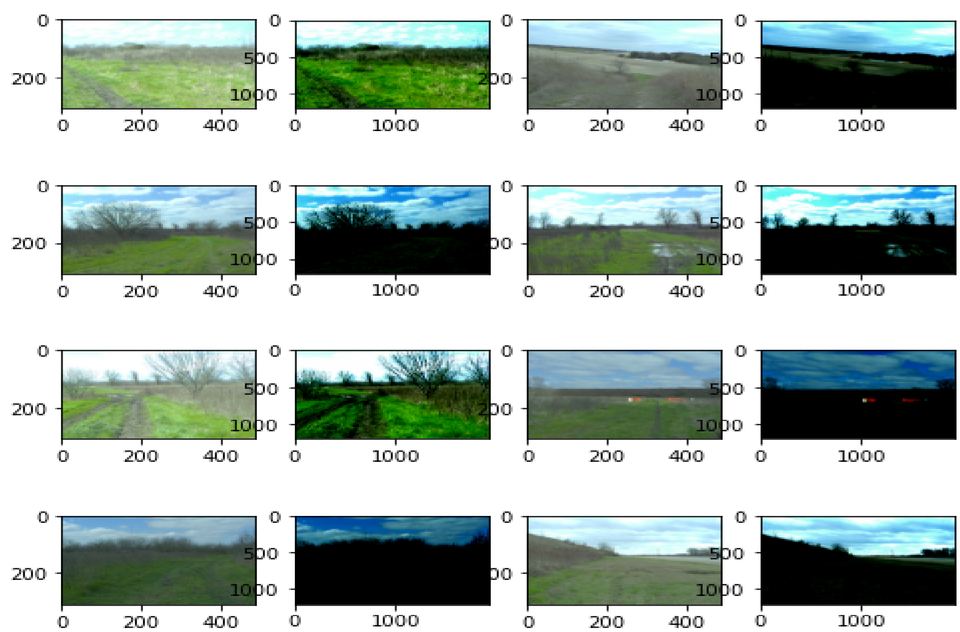

Figure 4: Image showing low-resolution image(left) and high-resolution output obtained from the generator for image input size 288*450

The transfer learning approach trained on the Rellis-3D dataset yielded images with a loss of 0.00012 with a PSNR value of 35.5 and can be easily seen to be predicted well. Images were upsampled to the size 1920*1400.

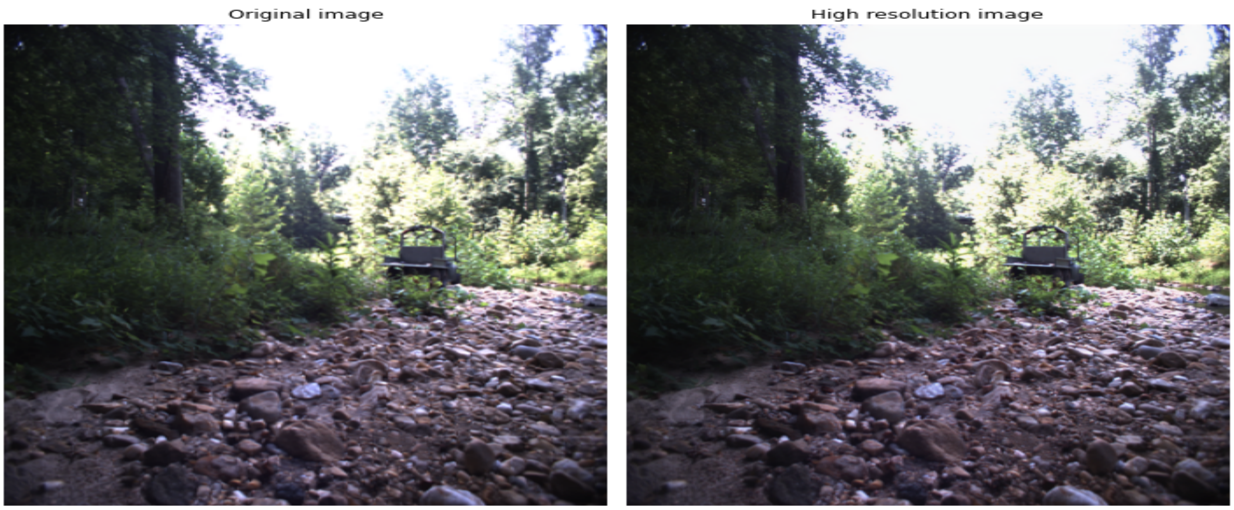

Figure 4: Image showing low-resolution image(left) and high-resolution output obtained from the generator.

Conclusion

Although we achieve considerable results using GAN, training and fine-tuning such models is a difficult task and the divergence of such models makes them difficult to use. The predictions of the SRGAN can later be used to have a single unified dataset of images of both RUGD and RELLIS-3D image datasets for image segmentation tasks.

References

[1] K. S. Viswanath, Kartikeya Singh, Peng Jiang, P. B. Sujit, and Srikanth Saripalli. Offseg: A semantic segmentation framework for off-road driving. 2021 IEEE 17th International Conference on Automation Science and Engineering (CASE), pages 354–359, 2021.

[2] Christian Ledig, Lucas Theis, Ferenc Husz´ar, Jose Caballero, Andrew Cunningham, Alejandro Acosta, Andrew Aitken, Alykhan Tejani, Johannes Totz, Zehan Wang, and Wenzhe Shi. Photo-realistic single image super-resolution using a generative adversarial network. In 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pages 105–114, 2017.

[3] C. Dong, C. C. Loy, K. He, and X. Tang. Image super-resolution using deep convolutional networks, 2014.

[4] J. Bruna, P. Sprechmann, and Y. LeCun. Super-resolution with deep convolutional sufficient statistics. CoRR, abs/1511.05666, 2015.