Abstract

A Generative Adversarial Network uses two neural network models- Generator and Discriminator. Generator network generates new samples similar to original data. Discriminator network learns to differentiate generated samples from real samples.

SRGAN is an impressive application of GAN models. The model generates higher resolution images from lower resolution images by upscaling and improving the finer texture details in an image by calculating losses.

Introduction

The challenging task in the computer vision field is recovering higher resolution images from lower resolution images. The SRGAN model calculates different losses to generate higher resolution images that resemble original images- adversarial loss and feature losses constitute perspective loss and MSE loss.

Methods

The section describes the steps followed in building the model.

A. Dataset

DIV2K_train_HR and DIV2K_valid_HR datasets
are used for training and validating the data respectively.
The ground truth images are downsampled to acquire the lower resolution images.

B. Pre-train the Generator network

Using EPOCH value 2, the generator network is pre-trained to ensure right convergence of the generator and the discriminator loss functions. MSE loss is calculated by taking the pixel-wise difference between generated higher resolution images and the ground truth images.
A pre-trained VGG-16 model is used to extract the features of both generated high resolution images and the ground truth images. The difference between these features are considered as perspective loss.

C. Discriminator network

Discriminator network classifies real and generated images. The architecture is as shown in figure 1 and is as per the standards specified by Radford[2]. The discriminator uses residual blocks with skip connections. K3n64s1 signifies a kernel of size 3, 64 channel, stride of 1 and padding 1.

D. Generator network

The architecture implements a residual block network as shown in figure 1 and follows standards specified by Johnson[3]. K9n64s1 signifies a kernel of size 9, 64 channel, stride of 1 and padding 4. Pixel shuffling is used to upscale the image by scale of power 2.

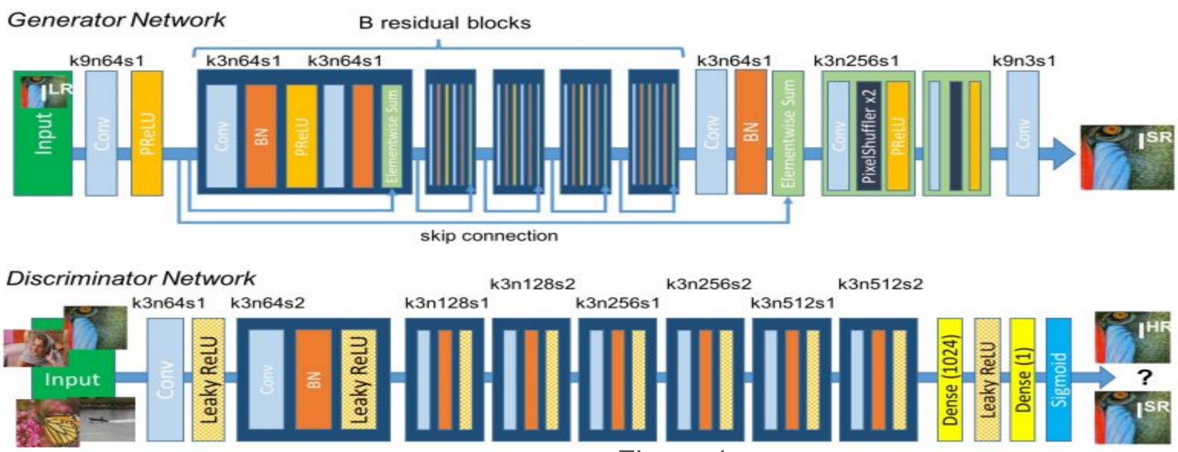

Figure 1

E. Losses

The discriminator loss calculates the BCE loss between high resolution original images labels and real labels(1) and generated high resolution images labels and fake labels(0).

The generator loss is calculated using the following formula.
totalGloss = image_loss + 0.001 * adversarial_loss + 0.006 * perception_loss.
image_loss is a MSE loss between generated high resolution images and original high resolution images.
adversarial_loss is a BCE loss between generated high resolution images labels and real labels.
perception_loss is a MSE loss between the extracted VGG-16 features of generated high resolution images and original high resolution images.

For both discriminator and generator, ADAM optimizer with learning rate of 0.0002 is used for optimization.

F. Result

Ground Truth | Low resolution Image | Generated high resolution image

Ground Truth | Low resolution Image | Generated high resolution image

References:

[1] Zhihao Wang, Jian Chen, Steven C.H. Hoi, Fello, IEEE “Deep Learning for Image Super-resolution: ASurvey”
[2] A. Radford, L. Metz, and S. Chintala. Unsupervised representation learning with deep convolutional generative adversarial networks. In International Conference on Learning Representations (ICLR), 2016. 3, 4
[3] J. Johnson, A. Alahi, and F. Li. Perceptual losses for real-time style transfer and super-resolution. In European Conference on Computer Vision (ECCV), pages 694–711. Springer, 2016. 2, 3, 4, 5, 7