Methods:

Our task is to implement image super resolution, a technique which enhances the resolution of low-resolution images. Generative Adversarial Networks (GAN) are used for super image resolution. Our model is based on GAN. In the GAN model there is generator and discriminator. The Generator generates fake high-resolution images from low-resolution images whereas discriminator’s task is to identify whether the generated images from generator are real or fake. Our model is based on the srgan.

Base Model Architecture:

The architecture of the based model is inspired from the srgan architecture. Our aim is to build a model based on to srgan by reducing the complexity in the model and the model’s performance should be as competitive as the srgan model.

In the Generator, we will be using 8 residual networks instead of 16 residual networks from the srgan generator architecture, the reason is to have the as good performance as of the srgan with the less complex architecture. The discriminator architecture is based on the srgan discriminator architecture.

A diagram of a block diagram

Description automatically generated

Final Model Architecture:

We have considered to tweak the base model performance. This is done by adding second discriminator for the base mode. The first discriminator is referred as Global Discriminator, and the second discriminator is referred as Local Discriminator. The Global discriminator works the same as the based model discriminator and for the whole image it detects whether the image is fake or real, whereas the local discriminator is used for chunks of the images to detect whether the chunks are real or fake. The size is 24 x 24 and when its processed then it outputs 96 x 96, we are dividing the 96 x 96 image into 4 parts. With this we can improve the image quality further by improving the chunks of the images. The complexity of the local discriminator is reduced as the image is much lesser compared to the image in the global discriminator.

A diagram of a computer process

Description automatically generated with medium confidence

Experiments:

For training the model, we have used the Hugging face’s DIV2K dataset. The dataset contains 800 images which has low resolution images and its ground truth image. But we have only used 200 images due to the system constraints. All the images from this dataset is pre-processed, where the low resolution images are converted to the size 96 x 96 and the high resolution images are converted to the size 384 x 384 as the generator outputs the same size. In order to have the comparable and valid calculations we need to have the same size.

For both our models, single discriminator and double discriminator we have used the same loss functions such as L1 loss, VGG loss, BCE, MSE. The BCE loss is the adversarial loss which encourages the generator of the model to output images that are similar to the real images. The VGG loss is a perceptual loss which uses VGG19 pretrained deep neural network to compare high-level feature of the generated and real images. The MSE is a pixel-wise loss which measures the difference between generated and real images. We have also used the same Adam optimizer for both the models.

The input size of our image is 96 x 96 low resolution image and the model outputs image size of 384 x 384 high resolution image which should be as good as its ground truth image. The evaluations will be done on three datasets: DIV2K, CIFAKE, Kwenter blur dataset. The CIFAKE just has low resolution images, so after getting the trained weights of the model we just evaluate on CIFAKE. We compare our model results with the models SRGAN, SRCNN which are tested on the similar dataset. We ran our base model for 58 epochs and the final model for 40 epochs. For the base model the output was getting good after the 10 epochs and for the final model the good images were seen after 6 epochs itself

The results are presented in the below table.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Dataset = DIV2K | Base Model  (Single Discriminator) | Base Model\_V2  ( Double Discriminator) | SRGAN | SRCNN |
| Peak Signal to Noise Ratio(PSNR) | 22.97 | 23.52 | 29.4 | 30.07 |
| Structural Similarity Index(SSIM) | 0.63 | 0.6974 | 0.84 | 0.86 |

|  |  |  |
| --- | --- | --- |
| Dataset = Kwenter blur | Base Model  (Single Discriminator) | Base Model\_V2  ( Double Discriminator) |
| Peak Signal to Noise Ratio(PSNR) | 22.74 | 23.49 |
| Structural Similarity Index(SSIM) | 0.63 | 0.67 |

A blurry image of a cat

Description automatically generatedA blurry cat sitting on a table

Description automatically generatedA blurry image of a cat

Description automatically generated

Low resolution resized Single discriminator double discriminator.

Blur blurry image of a table and chairs

Description automatically generatedA blurry image of a table

Description automatically generated

Single discriminator double discriminator

ThenumberofAIgeneratedimageareincreasingdayby day. Therearearound15billionAIgeneratedimagesand mostof themaretext toimagegenerated. Butasper the resolutionof thedataset, thegeneratedimagesmightnot beofhighqualityandhighresolution. Thisproblemcan betackledusingdifferentmethodsandthenthelowresolu tionimagecanbetransferredintothehighrelationimage. Therearefewmethodtoconvert thelowresolutionimage tothehighresolutionsuchasinterpolation,reconstruction, anddeeplearning. Inthisprojectwechosetousethedeep learningmethodtodesignourarchitectureforconstructing thehighresolutionimage.Mostofthemodelinthemarket rightnowarebasedontheGenerativeadversarialnetworks. GANmodel consistsof twolayerssuchasgeneratorand discriminator.Thediscriminatorlayerisresponsibletocon structthefakeimageandthenthediscriminatorneedtover ifyiftheimageisrealorfake. Thegeneratorhavetocon structtheimageinsuchawaythatthediscriminatorwon’t beabletoidentifyiftheimageisfakeornot.Anddiscrimi natorneedtomakesuretoidentifyallthefakeimage.Both thegeneratoranddiscriminator layerconsistsof theneu ralnetworks,activationfunction,normalizationlayer.Alto gethertheyarecalledastheresidualblock. Intheoriginal SRGAN(Superresolutiongenerativeadversarialnetworks) thereare16residualblocks. Thediscriminator layergen eratetheprobabilitybetween0and1determinedhowreal isthegeneratedimage.Buttherearefewproblemwiththe GANmodels, thererequiremorecomputationalpowerand theyarehardtotrain. Todecreasethecomplexityandin creasetheefficiencyofthemodel,wedesignedthearchitec turewithlessnumberofresidualblocks.Originalthereare 16residualblocksusedintheGANmodebutinourmodel, weusedonly8residualblocks.Also,thereisamajordiffer enceinourmodelandotherGANmodels,originallythere isonly1discriminatorusedintheGANmodelbut inour architectureweused2layerofdiscriminator.Thefirstdis criminator takes thewholegeneratedimageanddetects if it is real or fake, and theseconddiscriminator takes the chunksoftheimageandthendothesameoperation. Inour model,theimageisdividedinto4chunksandthenthe2nd layerofdiscriminatordoesitsoperation.Thisdecreasethe complexityofthemodelandincreasesitsefficiency.There arefewothermodelalsoavailableforthesametaskinthe market.Butallthemodelshavefewshortcomingandsingle imagesuperresolutionwasnotperformedontheAIgener atedimages. Inthisprojectwearetryingtoincreasethe resolutionofsuchimages.

Image super-resolution has become increasingly important in various applications because of their demand for producing high quality images from the low-quality images. Earlier for the image enhancements techniques like deblurring were performed to get the quality image. with the advancements in the Generative Adversarial Networks (GAN), the generating of high-quality image from the low-quality image has been outstanding. The models like SRGAN, ESRGAN are the competitive models which make the Image-Resolution look good because of their performance on the images. But the architecture of the SRGAN which is a state-of-art model is complex and ESRGAN is built on the SRGAN, but by observing the results of the SRGAN the image quality looks good. we try to build a Super-Image Resolution by having the less complex architecture which is faster than SRGAN and the results aren't compromising even after reducing the architecture. We have built our base model based on the SRGAN by reducing the complexity in the architecture. In our final model we added another discriminator which enhances the sub-parts of the images to improve the image quality. Our aim is to build an efficient model where the architecture of our model is less complex than srgan and give as competitive results as srgan. Our results for the final model compared to our base model shows that there were significant improvements in the image quality. The code for our project can be found here:

Over the years there has been extensive research done for the high image resolution tasks. Traditional methods, such as fusion, deblurring, aimed to enhance image quality. However, these techniques just do the deblurring. That’s the reason Image super-resolution techniques advanced further where if we input low quality image, it results in the high-quality image. These techniques have real-world applications and are used in defence, medical fields etc.

Deep learning techniques have contributed with these tasks effectively. The rise of the GAN models also handled the super-image resolution task effectively. The advent of GAN has revolutionized the field of image super-resolution. The models like SRGAN, ESRGAN has set new benchmarks by producing high quality outputs from the low-quality image. Despite its good results, the architecture is complex and needs computational demands. The SRGAN model is a GAN based model where it has Generator and Discriminator. The generators task is to generate high resolution images to fool the discriminator and the discriminators task is to identify the real/fake images.

In our work we aim to develop a feasible architecture model which will be as good as the state-of-art model SRGAN. The primary objective of the project is to contribute in the super-image resolution techniques which can beat the state-of-art methods

Having the high resolution of the single image from low quality to high resolution is done in many ways. Previously there has been some work done in the same domain. In this section we will discuss some of the work which was referred in order to complete this project. \cite{CHEN2022124}Chen et al in the their paper gave a review about the single image super resolution, how its is very important now a days for image processing application. There are many applications for the single image super resolution like in the medical field, entertainment, surveillance and many more. They also gave the testimony for the super resolution. Super resolution is divided: video super resolution and image super resolution. The image super-resolution process is then divided into two parts: single image super resolution and multi frame image super resolution. Single Image Super-Resolution is then classified into many other forms such as: regression based method, deep learning based method, etc. Several datasets were discussed by the authors which have been used to train and test different models by the various types of super-resolution available in the market. \cite{8723565}Yang et al give a survey about employing deep learning for single image super-resolution task in their paper. They focused on two things in this survey; how effective neural network was and how optimization had been performed for the neural networks designed for single image super-resolution operation. The major challenges faced in this paper were divided into 3 components that include Advancements of Deep Models.\cite{10.1007/978-3-319-10593-2\_25}Yang et al described systematic benchmark evaluation for Single Image Super Resolution (SISR). Afterwards they conducted user study involving human to verify the quality of finally created images. Many different metrics like Bicubic interpolation, IP, SLJT etc., are used here to evaluate different models.\cite{10.1007/978-3-030-00563-4\_11}Ha et al illustrates about using convolutional neural network (CNN) in order to achieve single image video sequences resolution." The CNNs were first applied to SISR by simple three-layer neural network called SRCNN. It was referred as SRCNN. In case of GAN approach while training the discriminator already knows about dataset and its representation so it can discriminate between real or fake images.” This paper surveys various methodologies that could be employed to SISR, most importantly how should they design their NN relating HR and LR. \cite{Wang\_2018\_CVPR\_Workshops} Wang et al in t their peer discussed about the method they used by designing the GAN architecture. They mentioned that it is difficult to have more clarity in case of the large upsampling factor. The architecture they designed is called is ProGanSR which uses the gradual increase of resolution in the image. The upsampling is done in the steps where in each step the resolution is increased by some percentage. They also mentioned that in terms of the SSIM, their architecture ranks 2nd and in case of the PSNR score, it ranked 4th. \cite{8253599} Creswell et al, in their paper discussed the overview of the Generative Adversarial network. They discussed how two layer of network is required to make the GAN model and they are in competition with each other. The discriminator in the architecture have access to the both real images and then fake images which is generated by the generator. It need to determine if the output image is real or fake. There are different kinds of GAN architecture present like fully connected GAN, Convolution GAN, conditional GAN, GAN with inference model, and adversarial encoders.\cite{8039016} Wang et al, in their paper discussed about the two player zero sum game. They studied the GANs U+02BC in detail about its implementation and application. The important thing they discussed was about the parallel intelligence and hoe GAN model are really good for the parallel intelligence application.\cite{AGGARWAL2021100004} Aggarwal et al in their paper discussed the survey about the GAN models. GAN can be used for the applications like image identification, speech synthesis, text mining in which data can be represented in the form of probability distribution.\cite{10.1145/3422622} Goodfellow in their paper discuss abut the GAN models. GAN model are hard to train on and system failure is also one of the limitations when comes to the GAN model. The quality of the images which are getting generated by the generator layer of the GAM model is increasing day by day but still it is not perfect. It can be easily determined of the image ins real or fake.\cite{YI2019101552} Yi et al in their paper discussed the application of the GAN model in medical field. There are more application of the GAN models like domain adaption, data augmentation, and image to image translation. In case of medical field, it is widely used in the image reconstruction, segmentation, detection, classification and cross modality synthesis. All this work was really essential for us to understand the GAN in depth and designing our our model which consists of two discriminator layers.