Image Super Resolution

Guggilam Sai Prabhat
Department of Computer Science and Engineering
Amrita Vishwa Vidyapeetham
Amritapuri, Kerala, India
amenu4aie20032@am.students.amrita.edu

Natte Sai Bharath
Department of Computer Science and Engineering
Amrita Vishwa Vidyapeetham
Amritapuri, Kerala, India
amenu4aie20052@am.students.amrita.edu

Raja Pavan Karthik
Department of Computer Science and Engineering
Amrita Vishwa Vidyapeetham
Amritapuri, Kerala, India
amenu4aie20060@am.students.amrita.edu

Sista Sai Subramanya Mrinaal
Department of Computer Science and Engineering
Amrita Vishwa Vidyapeetham
Amritapuri, Kerala, India
amenu4aie20067@am.students.amrita.edu

Abstract

Image Super-Resolution (SR) is a significant family of image processing algorithms used in computer vision to improve the resolution of pictures and videos.In recent years, deep learning techniques to image super-resolution have advanced dramatically. The purpose of this article is to offer a detailed overview of current improvements in picture super-resolution utilizing deep learning algorithms. In general, extant studies on SR approaches may be divided into three broad categories: supervised SR, unsupervised SR, and domain-specific SR. We also discuss several other critical topics, such as publicly available benchmark datasets and performance evaluation measures. Finally, we end our study by outlining various prospective paths

Introduction

Image Super-Resolution (SR) is an essential course of picture refining methods to improve the resolution of pictures and video clips in computer system vision. In recent years we have seen amazing development of picture super-resolution utilizing deep discovering methods. This project targets picture super-resolution utilizing deep discovering methods. Generally, we can approximate the current research of SR methods into 3 significant classifications: SupervisedSR, unsupervisedSR, and Domain-Specific SR. There are lots of techniques utilized to refix the images. Pre-Upsampling Super Resolution, Post-Upsampling Super Resolution, Residual Networks, Multi-Stage Residual Networks, Recursive Networks, Progressive Reconstruction Networks, Multi-Branch Networks Attention-Based Networks, Generative Models are the different types of Super-Resolution Methods and Techniques. The project focuses on Post-Upsampling Super-Resolution and Generative Models.

The formula Ix = D(Iy;) can be used to convert low-resolution (LR) photos to high-resolution (HR) images. The output is Iy as the High-Resolution(HR) picture, where Ix is the low-resolution image and is the noise, and D is the degradation function. In this project, our main aim is to estimate a high-resolution, super-resolved image $I^{SR}$from a low-resolution input image $\boldsymbol{I^{LR}}$. Here $\boldsymbol{I^{LR}}$is the low-resolution version of its high-resolution counterpart I^{HR}. These images(H.R) are only available during training. In the training, we can get $I^{LR}$by applying a Gaussian filter toI^{HR}followed by a downsampling operation.

Review of literature

In this article, SRGAN is introduced. A Generative Adversarial Network (GAN) for super-resolution imaging. In this article, it is the first frame that can be used to derive a realistic natural image at 4x magnification. Work related to this topic-included super-resolution of images, design of convolutional neural networks, and loss functions. To achieve the results of this topic, a convolutional network architecture consisting of perceptual loss function (consists an adversarial loss and a content loss) have been proposed. In the experiment, the benchmark dataset was used to measure the data and similarity. BSD100, BSD300 test sets, and experiments were performed at 4x magnification between high and low-resolution images. In the training details and parameters, all networks were prepared on the NVIDIA Tesla M40 GPU utilizing an arbitrary example of 350,000 pictures from the Image Net information base. MOS tests were performed to measure the capacity of various ways to deal with recreate perceptually convincing pictures. Think about the presentation of SRResNet and SRGAN with NN, bicubic introduction, and four progressed strategies. This result confirms that SRGAN far surpasses all reference methods and establishes a new state-of-the-art for photorealistic SR imaging.

SRGAN is a groundbreaking work that allows you to create realistic textures in a single frame of super-resolution. However, unpleasant artifacts often accompany hallucination details. we have read and understood the three components of the SRGAN architecture: adversary loss and perceptual loss so we can improve visual quality. Each of them will be upgraded to derive the extended SRGAN. RRDB (Residual-in-Residual Dense Block) was introduced as a basic network-building unit. By using the pre-activation features, the Perceptual loss is improved. The SR problems are solved by the deep neural network approach. We will first discuss the proposed network architecture and then the improvements and loss of awareness made by the discriminator. Finally, we examined through the network interpolation methodologies to track down balance perceptual quality and PSNR By limiting function prior to activation, a more effective loss of perception occurs.

Research Methods

SRGAN

SRGAN is a generative adversarial network which used for image super-resolution for single images. This method uses of a perceptual loss function in that which consists of an adversarial loss and a content loss. Here a discriminator network trained to distinguish between super-resolved images and original photo-realistic images which directs the solution to the natural image manifold via the adversarial loss. Furthermore, other than pixel space similarity, this method uses the content loss which is motivated by perceptual similarity.

Architecture:

Like other GAN architectures, here the Super Resolution GAN is divided into two parts: the generator and the discriminator. The generator which generates data based on a probability distribution, and the discriminator which attempts to predict whether data based on the input given dataset or came from the generator. After that, the generator attempts to optimize the generated data by generated to fool the discriminator. We will see architectural details of the generator and discriminator.

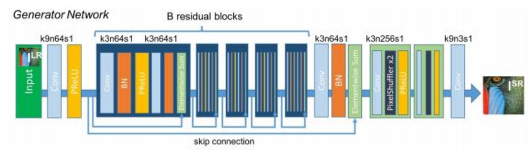
$$l_{\text{Gen}}^{SR} = \sum_{n=1}^{N} -\log D_{\theta_D}\left(G_{\theta_G}(I^{LR})\right)$$

$$\hat{\theta}^G = \arg \min_{\theta_G} \frac{1}{N} \sum_{n=1}^{N} \ell^{SR}\left(G_{\theta_G}(I_n^{LR}), I_n^{HR}\right)$$

→ loss function

The architecture of the Generator:

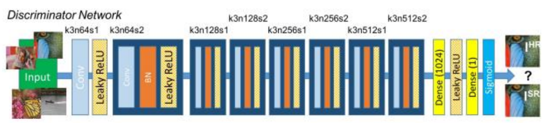
Here the generator architecture which employs residual networks rather than deep convolution networks because here for this they are easier to train and which allows them to be significantly deeper in order to produce better and good results. This is due to the residual network's use of a connection type known as skip connections. The ResNet creates the B residual blocks (16). Within the residual block, two convolutional layers with small 33 kernels and 64 feature maps will be used, followed by batch-normalization (BN) layers and as well as the activation function as ParametricReLU.



To improve the resolution of the provided input image, two trained sub-pixel convolution layers will be used. In addition, as opposed to LeakyReLU, this uses a fixed value for a rectifier parameter, this generator architecture uses parametric ReLU as an activation function (alpha). It adaptively learns the rectifier's parameters and improves accuracy at a negligible additional computational cost. During the training, a high-resolution image (HR) will be down sampled to a low-resolution image (LR). And by using the generator architecture, the image is then up sampled from low resolution to super-resolution. Then the image is then passed through the discriminator, which attempts to distinguish between a super-resolution and a high-resolution image and get produces an adversarial loss, which is then backpropagated into the generator architecture.

The architecture of Discriminator:

The discriminator's main job or aim is to distinguish between the genuine HR images and the fabricated SR images. This study's discriminator architecture is similar to a DC-GAN design with LeakyReLU as the activation. The network will be having eight convolutional layers, each of them consisting of 33 filter kernels, by increasing the factor of two from 64 to 512. And these Stripped convolutions are used here to reduce image resolution when the number of features is doubled. The 512 feature maps will be followed by thick two layers with a leakyReLU will be applied between them and with the final sigmoid activation function to obtain a probability for sample classification.



The SRGAN employs the perpetual Loss Function (LSR), two loss components weighted sum is called perceptual loss the two components are content loss and adversarial loss. This loss is critical to the Regarding the performance of the generator architecture this loss will be critical.

$$l^{SR} = \underbrace{\underbrace{l^{SR}_{\mathrm{X}}}_{\text{content loss}} + \underbrace{10^{-3} l^{SR}_{Gen}}_{\text{adversarial loss}}}_{\text{perceptual loss (for VGG based content losses)}}$$

In this project we use these two types of Content Loss and the pixelwise MSE loss will be using for the SRResnet architecture, which is the most commonly used MSE loss for Image Super-Resolution; and pixelwise MSE loss for the SRResnet architecture, which is the least commonly used MSE loss for Image Super-Resolution. MSE loss, on the other hand, is incapable of dealing with high-frequency content in the image, resulting in images that are overly smooth. As a result, the paper's authors decided to employ the loss of various VGG layers. The VGG loss is calculated by using the ReLU activation layers which are pre-trained 19 layer of the same aka VGG network.

$$l^{SR}_{\frac{VGG}{i}.j} = \frac{1}{W_{i,j}H_{i,j}} \sum_{x=1}^{W_{i,j}} \sum_{y=1}^{H_{i,j}} \left(\phi_{i,j}(I^{HR})_{x,y} - \phi_{i,j}\left(G_{\theta_G}(I^{LR})\right)_{x,y}\right)^2$$

$$l^{SR}_{MSE} = \frac{1}{r^2 WH} \sum_{x=1}^{rW} \sum_{y=1}^{rH} \left(I^{HR}_{x,y} - G_{\theta_G}(I^{LR})_{x,y}\right)^2$$

$\phi_{i,j}$ →feature map obtained by the j-th convolution
(after activation) before the i-th maxpooling layer within the VGG19 network

Adversarial loss, It is a loss function which forces the generator to generate images which are more similar to high-resolution images. And by using the discriminator finally which has been trained to distinguish between these high-resolution and super-resolution images. As a result, the entire content of this architecture will be lost.

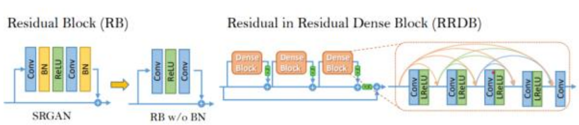
ESRGAN

SRGAN is a watershed moment in the search for visually pleasing results.The basic model is generated with residual blocks and optimized in a GAN framework using perceptual loss. SRGAN considerably enhances the overall visual quality of reconstruction over PSNR-oriented approaches using all of these techniques.

There is still a significant difference between SRGAN findings and ground-truth (GT) photos. SRGAN's core components are revisited, and the model is improved in three ways. First, To improve the network's topology, we include the Residual-in-Residual Dense Block. The RDDB (Residual-in-Residual Dense Block), which is more competent and easier to train.. We also remove the Batch Normalization (BN) layers used in and replace them with residual scaling and smaller initialization to make training a very deep network easier. Second, we train the discriminator with the Relativistic average GAN (RaGAN), which learns to assess "if one picture is more realistic than the other" rather than "whether one image is real or phoney."Experiments suggest that this enhancement aids the generator in recovering more realistic texture features. Third, we suggest a better perceptual loss by employing VGG features prior to activation rather than after activation as in SRGAN. As will be demonstrated, we experimentally discover that the modified perceptual loss produces sharper edges and more aesthetically appealing outcomes.

Network Architecture

To increase the retrieved picture quality of SRGAN even further, We primarily change the structure of generator G in two ways: 1) discard all BN layers; 2)The Residual-in-Residual Dense Block (RRDB) replaces the original basic block by combining a multi-level residual network with dense connections, as illustrated in Figure.



Removing Batch-Normalization layers has proven to improve performance and minimize computational complexity in a variety of PSNR-related applications, such as SR and deblurring.BN layers normalize the features using mean and variance in a batch during training and use the estimated mean and variance of the entire training dataset during testing. When the training and testing datasets' statistics are considerably different, BN layers tend to create undesirable artifacts and impair the capacity to generalise.When the network is deeper and trained using a GAN framework, we discover that BN layers are more likely to introduce artifacts.These artifacts arise on occasion amid iterations and varied settings, contradicting the necessity for consistent performance overtraining .As a result, BN layers are no longer required to ensure consistent training and performance.Additionally, removing BN layers enhances generalization while lowering computational cost and memory use.

Perceptual Loss

In ESRGAN, the features preceding the activation layers are used**.**There would be downsides if features after activation layers are employed. First, as illustrated above, the active features are quite scarce, particularly after a very deep network. For example, the average percentage of active neurons for the picture 'baboon' following the VGG19–546 layer is just 11.17 percent. Because of the sparse activation, oversight is inadequate, resulting in poor performance. Second, when comparing the reconstructed brightness to the ground-truth picture, using features after activation results in an uneven reconstructed brightness.

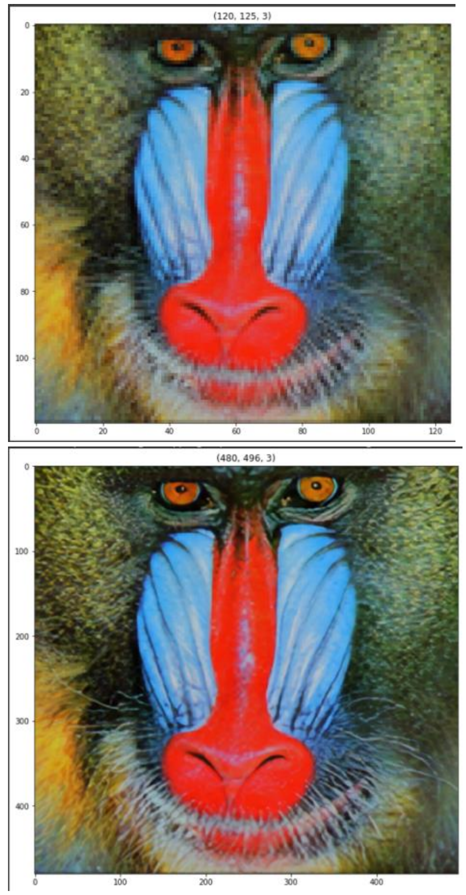
The total loss is :

$$L_G = L_{percep} + \lambda L_G^{Ra} + \eta L_1$$

where Lpercep is the VGG loss, LGRA is the adversarial loss, and L1 is the standard L1 loss, which is act as a content loss.

Results & Discussion

We have tried out the 2 most effective methods regarding image super-resolution which are SRGAN and ESRGAN.In SRGAN and ESRGAN methods we will first take a model as a reference then after their respective processes, we will get the Super Resolution images. In SRGAN we have used the shape Function from matplot model to determine the count of horizontal and vertical pixels of the image taken and output. So that we can compare the images and know how much the image is super resolute. And as for ESRGAN, we have used the PSNR function to define the percentage compared with the input image and the processed image.The ESRGAN model achieves a continuous better visual quality than the previous SR methods. When it comes to time taken we have calculated wall time which is the time taken for the process. We have used a time module for ESRGAN to calculate the time taken by the process. And if compare them both with respect to time and resolution ESRGAN is a better method compared with SRGAN. Moreover, we can say that the ESRGAN is the next extension or developed version of SRGAN. This implies that ESRGAN is more effective than SRGAN. Test results show that with our hybrid quantization strategy, the accuracy of the two standard SR models is kept almost unchanged while the model size decreases significantly. This combination of multiple measurement methods makes the performance of the model greatly improved, i.e. the PI value of the image you receive is reduced from 2.4731 to 2.1049 when using the SRGAN model, and from 2.688 to 2.2075 when using the ESRGAN model.



SRGAN Compared by using pixels (120,125,3) to (480,496,3)

Limitations

Super-resolution GANs do not learn to keep a person's identity, but they do learn to construct better resolution without inferring and synthesizing face characteristics such as lips and eyes. Super-resolution GAN models frequently generate distortion. For landscape syntheses, such as trees and buildings, the distortion is acceptable. Without viewing the ground truth image, a closer look at the reconstructed image of SRGAN appears pretty lifelike. However, when the original image is compared to the recovered SRGAN image, the texture is not accurately reconstructed..GAN adds something to the texture section that was not there in the original image. In the 'text' regions, the observation becomes more pronounced. SRGAN retrieves completely different things if the text in the source image is too tiny or distorted.In SRGAN, it is no longer text or readable. It's due to the loss function. It helps the image appear more realistic in general, which favors natural scenes over unnatural ones such as text or manmade structures. The same was true for the super-resolution of a human face. It might result in a completely different individual. This is risky if you do classification tasks relating to these regions, such as texture classification, license plate recognition, and so on. Real-ESRGAN can recover most real-world photos, but it has certain limitations. Due to aliasing concerns, some restored photos (particularly those of buildings and interiors) show distorted lines. On some samples, GAN training introduces unfavorable artifacts. In the actual world, it was unable to eliminate sophisticated out-of-distribution degradations. On some samples, GAN training introduces unfavorable artifacts. In the actual world, it was unable to eliminate sophisticated out-of-distribution degradations. Worse, it may exaggerate these artifacts. These shortcomings have a significant influence on the actual implementation of Real-ESRGAN and must be addressed in future research.

Conclusion

We managed to get PSNR up to 28.92 in the database we published and managed to score 26.097330 PSNR points out very high compared to almost all GAN networks currently available for Image Recovery purposes. images found online about the lost cultural history of India, and give a good idea of the field of finding lost data in the surrounding regions of the world.

The scope of the project can be further expanded with the simple use of the proposed ESRGAN as an anaconda. a backend that provides a web frontend that includes a web forum that sends blurry images as apps to the background and background model and sends a snapshot that can be used by a wide range of applications from forensics to historical rethinking. Provides a modern way to restore the image using an easy-to-use platform for historians, photographers, and web application developers to use Machine Learning and in-depth responsive web learning methods application development. An important idea is to introduce the Ranger to learn behavioral behavior detection by learning to measure approach. In addition, our proposed approach may include the power of different SR methods and better productivity results.

References

Cen, zhuo. “An Overview of ESPCN: An Efficient Sub-Pixel Convolutional Neural Network.” *Medium*, Medium, 17 Apr. 2020, medium.com/@zhuocen93/an-overview-of-espcn-an-efficient-sub-pixel-convolutional-neural-network-b76d0a6c875e.

“Image Super Resolution: Deep Learning for Image Super Resolution.” *Analytics Vidhya*, 27 May 2021, www.analyticsvidhya.com/blog/2021/05/deep-learning-for-image-super-resolution/.

Lichtman, Amit. “An Introduction to Super-Resolution Using Deep Learning.” *BeyondMinds*, 26 June 2021, beyondminds.ai/blog/an-introduction-to-super-resolution-using-deep-learning/.

Maddrell-Mander, Sam. “SRGAN, a TensorFlow Implementation.” *Medium*, Towards Data Science, 29 Oct. 2018, towardsdatascience.com/srgan-a-tensorflow-implementation-49b959267c60.

“Super Resolution GAN (SRGAN).” *GeeksforGeeks*, 29 May 2021, www.geeksforgeeks.org/super-resolution-gan-srgan/.

https://arxiv.org/pdf/1609.05158.pdf

https://arxiv.org/pdf/1609.04802v5.pdf

https://arxiv.org/pdf/1902.06068.pdf

https://arxiv.org/pdf/1809.00219v2.pdf

https://arxiv.org/pdf/1501.00092.pdf