IMAGE UPSAMPLING USING SR-GAN

Ayush -20MIM10033

Animesh -20MIM10111

Divyansh -20MIM10112

Kartik -20MIM10028

INTRODUCTION

The acronym "image super-resolution" defines the process of creating high-resolution pictures from low-resolution ones. High resolution corresponds to denser, more flexible pixels that can display more minute details in the image. The aforementioned details are extremely valuable in everyday circumstances, such as satellite imaging, medical imaging, etc., where they can assist in identifying essential characteristics and identify targets in high-resolution pictures. Fine anatomical information may be found in high-resolution (HR) MRI scans, which is useful for making precise assessments and conclusions. The gathering of picture data is complicated since it necessitates not just expensive equipment but also a lengthy scanning period. Due to the sluggish acquisition of data and imaging speed, more applications cannot be developed. The super-resolution (SR) approach may enhance the visibility of image features or restore image details through the generation of a high-resolution (HR) picture from one or more low-resolution (LR) images. SR approaches may substantially improve MRI's spatial resolution without altering its hardware or scanning components. The implementation of image SR in MRI typically uses one of three approaches: interpolation-based, construction-based, or machine learning-based.

The interpolation-based SR approaches entail that either a polynomial or an interpolation function with a priori smoothness may be used to expand the area in the LR picture to the comparable region. The interpolation-based super-resolution reconstruction approach has amazing real-time performance and is simple, but it has the drawback of being too basic to fully use the previous knowledge in MR images. Particularly since the single-MR picture super-resolution restoration approach contains prominent shortcomings that result in a blurry rendition of the associated HR reference image. For the purpose of trying to solve an optimization challenge incorporating two terms, reconstruction-based SR methods are introduced. These terms are the regularisation term, which encourages sparsity, and the fidelity term, which penalizes the difference between a degraded SR image and an observed LR image. When the input data is too sparse or the model is even somewhat wrong, the performance of these approaches deviates from ideal, especially in the high-frequency range. Due to these flaws, reconstruction-based SR techniques may only be used at high magnifications, however, they may be effective at smaller magnifications up to 4. Due to their cutting-edge performance in SR for natural photos, machine learning techniques, in particular deep learning

(DL)-based SR approaches, have lately gained a lot of interest. The majority of contemporary techniques use data-driven deep learning models to reassemble the necessary features for precise super-resolution. Input and output relationships are intended to be automatically learned from training samples using deep learning-based methodologies. The reconstruction of CT/PET images has also greatly benefited from deep learning, as demonstrated by PET Image Reconstruction from Sinogram Domain.

The Generative Adversarial Network (GAN), introduced by Goodfellow et al., has recently been shown to perform well in image processing and super-resolution imaging thanks to the advancement of deep learning. The standard super-resolution GAN (SRGAN) framework was suggested by Sanchez et al. for creating super-resolution pictures of the brain. Convolutional layers are used in the majority of GAN-based picture creation models. Convolutions handle data in local neighborhoods, however creating distant relationships in pictures using solely convolutional layers is ineffective. With a tiny convolution kernel, learning the relationships between pictures is challenging. The performance of the model will suffer due to the convolution kernel's excessive size. Additionally, enlarging the convolution kernel's size can also enhance the receptive field, although doing so necessarily makes the system more complicated.

EXISTING WORK

Nasrollahi and Moeslund or Yang et al. have recently written overview papers on image SR. Here, we'll concentrate on single-image super-resolution (SISR) and skip discussions of methods for extracting HR pictures from multiple photos. One of the earliest techniques to address SISR was prediction-based. The SISR problem is oversimplified by certain filtering techniques, such as linear, bicubic, or Lanczos filtering, which often results in solutions with excessively smooth textures. Edge preservation has been the focus of certain methods that have been proposed. More effective methods often rely on training data and try to create a complicated mapping between low- and high-resolution picture information. Many example-pair-based techniques rely on LR training patches that are known to have known HR equivalents. Freeman et al. presented their earlier work. Compressed sensing is where related methods to the SR problem first appear. Patch redundancies between scales in the picture are used by Glasner et al. to drive the SR. This self-similarity paradigm is also used by Huang et al., who enhance self-dictionaries by further allowing for minor modifications and shape variances. Convolutional sparse coding, which Gu et al. suggested, increases consistency by analyzing the entire picture rather than overlapping portions. Tai et al. combine an edge-directed SR technique based on a gradient profile prior with the advantages of learning-based detail synthesis to restore realistic texture detail while avoiding edge artefacts. A multi-scale dictionary is suggested by Zhang et al. to identify redundant information in comparable distinct. Yue et al. obtain related HR photos from the web with comparable content in order to super-resolve landmark images, and they then offer an alignment criterion that takes into account structure. By locating comparable LR training patches in a low dimensional manifold and integrating their related HR patches for reconstruction, neighbourhood embedding algorithms up sample an LR image patch. By utilising kernel ridge regression, Kim and Kwon's authors highlight the propensity of neighbourhood techniques to overfit and create a map of example pairings that is more inclusive. Gaussian process regression or Random Forests can also be used to address the regression issue. In Dai et al., the most suitable regressors are chosen during testing after a large number of patch-specific regressors are learnt. Recently, SR methods based on convolutional neural networks (CNNs) have demonstrated exceptional performance. In Dai et al., the most suitable regressors are chosen during testing after a large number of patch-specific regressors are learnt. Recently, SR methods based on convolutional neural networks have demonstrated exceptional performance. In Wang et al., the authors use a learnt iterative shrinkage and thresholding approach to embed a sparse representation prior into their feed-forward network design. In order to obtain cutting-edge SR performance, Dong et al. employed bicubic interpolation to upscale an input picture and trained a three-layer deep fully convolutional network end-to-end. It was then demonstrated that allowing the network to directly train the upscaling filters may significantly improve performance in terms of both accuracy and speed. Kim et al. established a very effective architecture that enables long-range pixel processing using their deeply recursive convolutional network.

ARCHITECTURE

A Generative Adversarial Network (GAN) Super Resolution architecture is a particular variety of GAN created to improve the resolution of low-quality photographs. A discriminator network and a generator network make up the architecture's two primary parts. A low-resolution image is sent into the generator network, which creates a high-resolution image. It generally consists of many convolutional layers and up samples the picture using a method like reversed convolution or sub-pixel convolution.

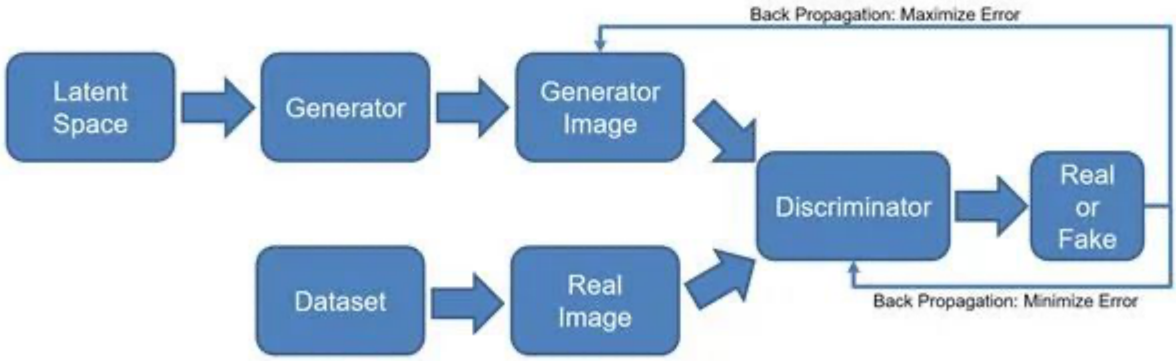
The produced high-resolution photos and the actual high-resolution images are separated using the discriminator network. It is trained to distinguish between false and real photos, classifying the produced images as fake.

The generator network seeks to produce realistic, high-resolution pictures that are identical to the actual pictures during training. On the other hand, the discriminator network seeks to reliably identify whether the produced pictures are phony or real.

Minimizing the adversarial loss among the generator and discriminator nodes is a goal of the iterative training process. The generator loss and discriminator loss, which assess how effectively the generator produces realistic pictures and the way the discriminator distinguishes between actual and created images, are combined to form the adversarial loss.

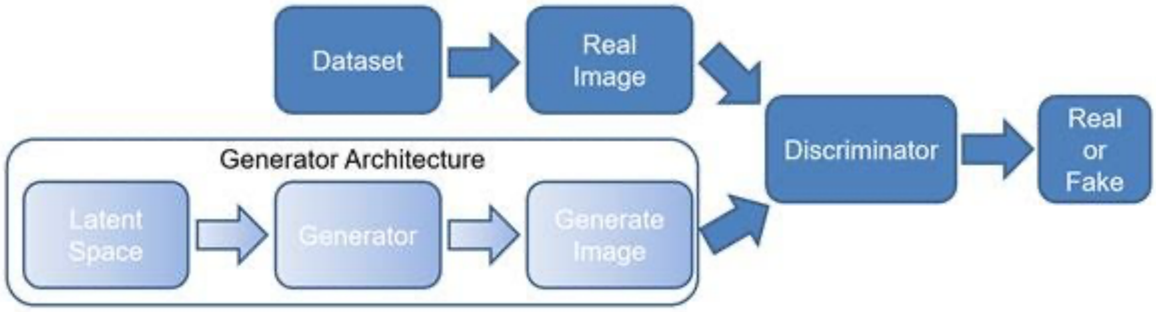
The resolution of surveillance films, satellite photos, and medical imaging have all been improved with the help of the GAN Super Resolution architecture.

The discriminator must train for a few epochs before beginning the adversarial training since the discriminator will need to be able to actually categorize pictures. The generator will begin training alongside the discriminator in the GAN framework. The loss function is the last component of this structure. The halting conditions for the Generator and Discriminator training procedures are provided by the loss function. How can we organize all of these components into something we can train with? Look at the illustration below:



Generator architecture

The following diagram represents the important pieces of the generator:



The generator components in the architecture diagram: latent space, generator, and image generation by the generator. The focus in the diagram ensures that you see the core piece of code that you'll be developing in the generator section.

- Here are a few steps to describe how we create a generator conceptually:
- First, the generator samples from a latent space and creates a relationship between the latent space and the output.
- We then create a neural network that goes from an input latent space to an output
- We'll train the generator in an adversarial mode where we connect the generator and discriminator together in a model
- The generator can then be used for inference after training

METHOD

In SISR the aim is to estimate a high-resolution, super- resolved image ISR from a low-resolution input image I LR . Here I LR is the low-resolution version of its high- resolution counterpart I H R . The high-resolution images are only available during training. In training, ILR is obtained by applying a Gaussian filter to I H R followed by a downsampling operation with downsampling factor r. For an image with C color channels, we describe ILR by a real-valued tensor of size W × H × C and IHR, ISR by rW × rH × C respectively.

Our ultimate goal is to train a generating function G that estimates for a given LR input image its corresponding HR counterpart. To achieve this, we train a generator network as a feed-forward CNN GθG parametrized by θG. Here θG = {W1:L ; b1:L } denotes the weights and biases of a L-layer deep network and is obtained by optimizing a SR-specific

loss function lSR. For training images IHR , n = 1, . . . , N n

withcorrespondingILR,n=1,...,N,wesolve: n

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

In this work we will specifically design a perceptual loss lSR as a weighted combination of several loss components that model distinct desirable characteristics of the recovered SR image. The individual loss functions are described in more detail in Section 2.

ADVERSARIAL LOSS

In addition to the content losses described so far, we also add the generative component of our GAN to the perceptual loss. This encourages our network to favor solutions that reside on the manifold of natural images, by trying tofool the discriminator network. The generative loss lSR Gen is defined based on the probabilities of the discriminator DθD (GθG (ILR)) over all training samples as:

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

Here, DθD (GθG (I LR)) is the probability that the recon- structed image GθG (I LR) is a natural HR image. For better gradient behavior we minimize − log DθD (GθG (I LR)) in- stead of log[1 − DθD (GθG (ILR))]

DISCUSSION

We confirmed the superior perceptual performance of SRGAN using MOS testing. We have further shown that standard quantitative measures such as PSNR and SSIM fail to capture and accurately assess image quality with respect to the human visual system [56]. The focus of this work was the perceptual quality of super-resolved images rather than computational efficiency. The presented model is, in contrast to Shi et al. [48], not optimized for video SR in real-time. However, preliminary experiments on the network architecture suggest that shallower networks have the potential to provide very efficient alternatives at a small reduction of qualitative performance. In contrast to Dong et al. [10], we found deeper network architectures to be beneficial. We speculate that the ResNet design has a substantial impact on the performance of deeper networks. We found that even deeper networks (B > 16) can further increase the performance of SRResNet, however, come at the cost of longer training and testing times (c.f.

supple- mentary material). We further found SRGAN variants of deeper networks are increasingly difficult to train due to the appearance of high-frequency artifacts.

In this work we found to yield the perceptually most convincing results, which we attribute to the potential of deeper network layers to represent features of higher abstraction [68, 65, 40] away from pixel space. We speculate that feature maps of these deeper layers focus purely on the content while leaving the adversarial loss focusing on texture details which are the main difference between the super-resolved images without the adversarial loss and photo-realistic images. We also note that the ideal loss function depends on the application. For example, approaches that hallucinate finer detail might be less suited for medical applications or surveillance. The perceptually convincing reconstruction of text or structured scenes [31] is challenging and part of future work. The development of content loss functions that describe image spatial content, but more invariant to changes in pixel space will further improve photo-realistic image SR results.

PERFORMANCE OF THE FINAL NETWORKS

We compare the performance of SRResNet and SR- GAN to NN, bicubic interpolation, and four state-of-the- art methods. Quantitative results are summarized and confirm that SRResNet (in terms of PSNR/SSIM) sets a new state of the art on three benchmark datasets. Please note that we used a publicly available framework for evaluation , reported values might thus slightly deviate from those reported in the original papers.

We further obtained MOS ratings for SRGAN and all reference methods on BSD100. Examples of images super-resolved with SRResNet and SRGAN are depicted in the supplementary material. The results shown in confirm that SRGAN outperforms all reference methods by a large margin and sets a new state of the art for photo- realistic image SR.

CONCLUSION

We have described a deep residual network SRRes- Net that sets a new state of the art on public benchmark datasets when evaluated with the widely used PSNR mea- sure. We have highlighted some limitations of this PSNR-focused image super-resolution and introduced SRGAN, which augments the content loss function with an adversar- ial loss by training a GAN. Using extensive MOS testing, we have confirmed that SRGAN reconstructions for large upscaling factors (4×) are, by a considerable margin, more photo-realistic than reconstructions obtained with state-of- the-art reference methods.

REFERENCES

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

https://www.sciencedirect.com/science/article/pii/S1110982322

00120X?via%3Dihub

https://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=9496642

https://link.springer.com/article/10.1007/s11831-019-09388-y https://link.springer.com/article/10.1007/s11042-020-09642-

https://link.springer.com/article/10.1007/s11276-020-02496-9

https://link.springer.com/article/10.1007/s11045-021-00806-8

https://link.springer.com/chapter/10.1007/978-981-15-9735

-

0 5https://www.sciencedirect.com/science/article/pii/S23529148
22001721

http://proceedings.mlr.press/v97/zhang19d/zhang19d.pdf

https://link.springer.com/chapter/10.1007/978-981-16-0882-7 64