Advancements in Image Super-Resolution: A Deep Learning Approach

K.G.S.Venkatesan1*

1Department of Computer Science & Engineering, MEGHA Institute of Engineering & Technology for Women, Edulabad, Hyderabad, Telangana, India.

Received: 02.08.2023 • Accepted: 15.08.2023 • Published Online: 31.08.2023

Abstract: The goal of this study is to address the problem of picture super-resolution (SR) by using a deep learning strategy to the opposite problem. Super-resolution (SR) imaging has garnered attention for a long time. Since then, several different branches of machine learning and deep learning have made significant strides in solving problems like these in medical imaging. Both deep learning and machine learning belong to this category. We explore the many applications of the Generative Adversarial Network (GAN) method, and talk about the deep learning strategies that may be utilised to address the problem of picture super-resolution. The Enhanced Super-Resolution Generative Adversarial Network (ESRGAN) and the Residual in Residual Dense Network (RRDN) are two state-of-the-art systems that have been created for picture super-resolution by the ‘idealo’ group. In this study, we analyse the effectiveness of both of these networks. The ESRGAN and RRDN teams are responsible for the development of both of these methodologies. The proposed model will be retrained with new parameters, and its results will be compared to those of two state-of-the-art techniques, which are adversarial networks for super-resolution of a single picture and enhanced super-resolution adversarial networks. Both of these approaches are aimed at improving the resolution of a single image. It can be good to understand the specifics of the required model as well as how our parameter choices compare to those of the other team in this particular circumstance.

Key words: Data Preprocessing, Image Super-Resolution, Convolutional Neural Networks, Loss Function.

*Correspondence: Professor, Department of Computer Science & Engineering, MEGHA Institute of Engineering & Technology for Women, Edulabad, Hyderabad, Telangana, India. Email:drkgsvenkatese@meghaengg.ac.in
https://doi.org/10.58599/IJSMIEN.2023.1801
Volume-1, Issue-8, PP:1-10 (2023)

This work is licensed under a Creative Commons Attribution 4.0 International License CC BY-NC-ND 4.0.
1. INTRODUCTION

Concepts at the core of artificial intelligence, such as machine learning also deep learning, have been efficiently implemented in a wide variety of contexts, with encouraging outcomes in many of those cases. Deep Neural Networks, often known as DNNs, are computer programmes that analyse enormous volumes of data in order to discover previously undiscovered answers to inverse problems. On the other hand, analytical approaches need a crystal clear articulation of the problem before any domain-specific particulars may be included into the solution [1] described the inverse problem, which is a subject that has been researched for a considerable amount of time. Image restoration, deconvolution, inpainting, projection-based reconstruction, compressive sensing, and Super-Resolution (SR), among other procedures, are all susceptible to experiencing inverse difficulties. Many scholars have invested a significant amount of time and effort into attempting to solve inverse problems utilising a diverse array of methodologies. However, nowadays we may gain assistance from neural networks thanks to DNNs, and this has led to the development of solutions that are both more focused and more effective. These modules, which are sometimes referred to as "layers" in the research literature pertaining to deep learning, are used by the model in order to assist it in extracting characteristics from the input that are essential to resolution the data that correlate to the crushed truth. By doing so, the model will become more accurate representation of the world. It is possible to produce a High Resolution (HR) version of the identical picture by first creating it in Low Resolution (LR) and then using it as a template to make the HR version.

The phrase "lower resolution" refers to a fall in the spatial resolution of a picture, which may be the result of a range of different conditions. The reasons for this drop in resolution might vary. The degradation function is applied to the HR data in order to extract the LR data, which ultimately results in a decrease in the overall image resolution. As a direct result of this, the overall picture quality is diminished. In order to find a solution to the inverse problem, one has to devise a method that converts LR to HR as precisely as is humanly likely. Deep learning algorithms have been found towards be real for super resolution through using an approach that was qualified several times through utilizing the dataset to create HR also LR values that were economized since HR. Many studies using a diverse array of research methods have contributed to the significant progress that has been accomplished in this area of study over the course of the last few decades [2]. The vast amounts of data that can be extracted from high-resolution photos may be used to a variety of scientific fields, which is one reason why these images are in such high demand. The better the picture’s resolution, the more information that may be gleaned from it. For example, the fact that every single one of the TVs that we use on a daily basis have a resolution of 4K or higher is one factor that has contributed to the desire for improved image quality. The better the visual quality, the lower the likelihood that the viewer would acquire myopia, and the greater the viewer’s capacity to enjoy what they are seeing. One of the ways in which picture super-resolution contributes to improvements in medical visualisation is by enhancing a physician’s capacity to precisely estimate the size of a cancer or another sickness. You will need access to a variety of different kinds of data in order to formulate a reasonable conclusion. Not only
in these particular industries, but also in others, such as the transportation and military industries, among others. Relatively recently, a satellite picture of 195 gigapixels was obtained in China utilising quantum technology. This made it possible to collect an unprecedented quantity of information, all the way down to the street level in the central business district of Shanghai. Hanoi, the capital of Vietnam, was the subject of an image taken by [3], which had a resolution of 13 gigapixels. Because it is of such critical importance to acquire photos with a higher resolution, a variety of methods for producing SR images have been devised. Single-image super-resolution (SISR), or single-frame pictures, are one kind of image processing, whereas multiple-image or multi-frame pictures are another. The purpose of this study is to offer a detailed account of the SISR, focusing especially on the DL Network methods that were used. The concept of sparse representation-based image SR reconstruction was developed by[4]. It’s possible to use this strategy in future picture super-resolution techniques. The SISR experiments yielded better PSNR also SSIM index than the bicubic also NCSR ones. By improving the resolution of a single CT picture [5] shown the usefulness of their technology in the medical imaging industry [6] looked examined how well different types of interpolation may be used to reconstruct high-resolution (HR) photographs in 2007. In 2009, Thuong Le- Tien and colleagues [7] used a SISR to recreate super-resolution images by fusing the frequency domain with the wavelet domain. On the other hand, in terms of SR image reconstruction, the Deep Learning technique has been investigated and put into practise, which has resulted in some exciting discoveries [8]. introduced the SRCNN in an effort to improve the upscaling of low-resolution (LR) images to high-resolution (HR) images, and they reported positive results; however, they also acknowledged that accelerating deep models and having an in-depth understanding of deep models, as well as the criteria for designing and evaluating the objective functions, are challenges for optimising their model [9]. found that using Deep Learning to execute Multi-frame Image SR yielded improvements on par with those obtained using Multi-frame Image SR [10] introduced and tweaked generative adversarial networks (GANs), several new and improved GAN techniques have been developed, yielding better results for supervised information fusion and re-use (SISR). To infer natural pictures with photorealistic 4x scale introduced SRGAN in 2017. Microsoft Research is responsible for creating SRGAN. After finding that their model worked as intended across several datasets [11, 12], they compared it to others using common performance indicators like PSNR and SSIM.

2. METHODOLOGY
The flowchart for the method is seen in Figure 1, and it states that low-resolution photographs will be put into a generator in order to make high-resolution pictures. The HR pictures and the SR images will be compared with one another so that it can be ascertained whether or not the SR picture was successfully made in close proximity to the ground truth. The feature maps of the SR image that are utilized to calculate the loss are the ones that are successful in getting through the discriminator. The feature maps of the HR image are generated by applying the VGG model. After that, the generator’s loss will be adjusted in such a way that the look of the SR photos will be brought as closely as feasible
to that of the HR images. Back propagation is then used by the mechanism in order to bring itself up to date in accordance with the generator and discriminator losses. The fact that these losses have begun to converge towards the value of the median shows that the generator and the discriminator have finally attained convergence.

Figure 1. The process flowchart for the GAN algorithm

2.1. Network Architecture
To enhance the superiority of the picture recovered by SRGAN, devised a technique called Residual in Residual Dense Block. The BN layers are stripped away in this process. It was shown that by using this method, performance in a number of PSNR-centric applications could be enhanced while computational complexity was decreased. In order to standardise a set of features, BN layers perform computations to find their mean and variance during training. However, testing is when we learn the true mean and standard deviation of the whole training dataset. Figure 2 shows that the Shallow Feature Extraction network, the Residual Dense Blocks network, the Dense Feature Fusion network, and the up-sampling network all play significant roles in the RRDN’s construction. Their method achieves this by replacing the RDB block with Wang’s [10] RRDB block. In the RRDB, memory consistency is preserved by the integration of local residual learning and local feature fusion (LFF).
2.2. Data Preparation

Collect a big number of pictures with a high resolution, and then produce a huge number of images with a low resolution so that they may be used in conjunction with the collection. In order to create replicas with a lower resolution than the originals, methods such as bicubic interpolation are used to scale down the high-resolution originals.

2.3. Loss Function

Develop a loss function in order to quantify the degree to which the real high-resolution image differs from the one that was anticipated. Common loss functions include mean squared error (MSE), mean absolute error (MAE), and perceptual loss (which combines content and style loss). MSE and MAE are both forms of mean squared error.

2.4. Training

Train the deep learning model using the prepared dataset and the chosen loss function. Training is typically done on powerful GPUs or TPUs to speed up the process.

2.5. Inference

To perform super-resolution on a new low-resolution image, input the image into the trained deep learning model. The model will produce a high-resolution output image.
2.6. Post-processing
Optionally, apply post-processing techniques to further enhance the output, such as denoising or sharpening.

2.7. Deployment
Integrate the trained model into an application or system where super-resolution is required, such as image editing software, video enhancement tools, or medical imaging systems. It’s important to note that the success of a deep learning-based super-resolution approach depends on factors like the choice of architecture, the quality and quantity of training data, the loss function used, and the hyperparameters selected. Experimentation and fine-tuning are often necessary to achieve the best results for a specific application. Additionally, advances in deep learning and computer vision continue to drive improvements in image super-resolution techniques, so staying up to date with the latest research is essential for achieving state-of-the-art results.

3. RESULTS AND DISCUSSION
For our purposes, the DIV2K dataset is the primary data source. More than 800 images, from close-ups of pets and people to panoramic vistas of the outdoors, make up this massive collection. In general, the bicubic scale-down approach may be used to a wide range of scaling factors over the whole dataset. Multiplication by two, four, three, or eight are all instances of such factors. In addition, there is a validation task that makes inferences about the method using a validation dataset of around 200 images. The dataset is similar to the one used for training. The model that Wang and his colleagues developed may have its parameters changed in the following ways: The coefficients in this equation are as follows: \( C = 5 \), \( D = 4 \), \( G = 65 \), \( G0 = 65 \), \( T = 12 \), and \( 2x \) (the scaling factor).

![Figure 3. Train PSNR-Y](image)

Training takes a long period because of the \( 5e^{-4} \) learning rate. Decomposition happens at a rate of 30 times per hour, with a decay factor of 50 percent. We trained for 80 epochs overall, with
500 iterations each epoch serving as the step size. We had 16 people in our batch. While optimising, the PSNR is the statistic of choice. Figures 3 and 4 show that after 80 epochs, the validation PSNR has improved to an acceptable 32, while the train PSNR-Y peaked at 44.

![Figure 4. Valid PSNR-Y](image)

By increasing the smooth weight to 0.82 in the tensor-board tool of the tensor-flow framework, we are able to achieve a more precise depiction of the PSNR. An improved PSNR-Y indicated that the produced SR pictures were starting to look more like the HR versions. We were able to extract maximum value from the model thanks to its convergence without resorting to overfitting. When compared to current best practises in picture super-resolution, the end training PSNR-Y increase of 34.22 is sufficient. Figures 4 and 5 depict the various training and validation stages at which generator loss might become apparent. The generated images are sent to the discriminator, and if the loss value shifts, it means the discriminator has identified the fakes in comparison to the real ones. As a result, the generator must adapt the quality of the pictures it produces to account for the shifts in loss.

![Figure 5. Train generator loss](image)

The generator loss is seen in Figure 6 at a size four times its normal size. The loss line will
<table>
<thead>
<tr>
<th>Method</th>
<th>Scale</th>
<th>PSNR-x2</th>
<th>PSNR-x4</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Bicubic down sampling</strong></td>
<td>Bicubic</td>
<td>28.93</td>
<td>26.85</td>
</tr>
<tr>
<td></td>
<td>SRGAN</td>
<td>28.64</td>
<td>22.72</td>
</tr>
<tr>
<td></td>
<td>ESRGAN</td>
<td>31.23</td>
<td>25.74</td>
</tr>
<tr>
<td><strong>Unknown down sampling</strong></td>
<td>Bicubic</td>
<td>25.94</td>
<td>22.983</td>
</tr>
<tr>
<td></td>
<td>SRGAN</td>
<td>25.58</td>
<td>21.29</td>
</tr>
<tr>
<td></td>
<td>ESRGAN</td>
<td>25.72</td>
<td>21.94</td>
</tr>
</tbody>
</table>

always point in the right direction up to the 40k iteration. This ensures that the model will function as intended. Our prediction model is trained using data from a large and varied training set, both in terms of the people it represents and the contexts in which those people were collected.

A photo from the validation set measuring 192 by 255 by 3 was used using this method, and after being scaled up by a factor of 2, we obtained a picture measuring 384 by 510 by 3 that had all of the information from the original. The data are doubled in size using the bicubic scaling technique, and then compared to the final result. We expect that the bicubic method will provide less precise and less well-defined outcomes compared to the model’s predictions. After that, we tried to build the model using data that wasn’t included in the main dataset, all in an effort to prove that it could provide trustworthy results.

On the validation data set, the results of a comparison between the Mean PSNR of 100 Low-Resolution Images down sampled using the "Bicubic down sampling method" and the "Unknown down sampling method" are shown in Table I. This comparison was performed to determine which down sampling approach produced the best results for the Mean PSNR. The "Bicubic down sampling"
dataset indicates that in terms of average PSNR, ESRGAN performs better than both the bicubic up sample and the bicubic up sample with x2 up sampling. On the other hand, SRGAN’s performance is subpar in the most of our testing, but it is able to generate the scaled-up version of the Super-Resolution image. Because SRGAN performs badly in the majority of the evaluation procedures, we may draw the conclusion that an effective loss function is essential for the proper operation of a neural network. The x4 portion is where the contrasts between the three methods become most obvious, both in terms of the calculated numbers and the demo pictures.

4. CONCLUSION
In relation to the current situation, the elimination of BN ensures that everything remains stable and reliable without the appearance of any strange artefacts. It lessens the need for storage space and random access memory (RAM) while keeping the same degree of productivity. The utilisation of a vast.ai server equipped with 1080Ti and 16GB RAM imposed limitations on the batch size as well as the training phase throughout this investigation. Our findings are congruent with Wang’s since they both achieved a PSNR of 30 on the validation test and equally excellent results for images that were not included in either the dataset or the validation set. The RRDN model is used, and our findings indicate that more sophisticated models provide more accurate outcomes. However, since the model is so cutting edge, we recommend data in order to improve the findings. Denoising a dataset before training it is essential in order to prevent the dataset from generating more noise or being upscaled by mistake. We propose that data on the underwater environment into more generic categories, such as ”underwater,” ”land,” ”animals,” and so on. YOLO is only one of several algorithms that may be used to categories and segment photos; the other techniques are listed here. In order to run tests using RRDN, we first utilised classes to train additional weight files, and then we categorized the input data in order to choose the appropriate weight file to employ. We discover that the pictures that are produced contain improved information since they are more similar to the input that was trained using classes. Because of this, we need as much information as we can get for each category, which will cut down on the total amount of datasets.

References


