IJCSE International Journal of Computer Sciences and Engineering Open Access
Research Paper Vol.8, Issue.5, May 2020 E-ISSN: 2347-2693

Imagenics Super-Resolution Generative Adversarial Networks (ISRGAN)

Yash Bansal[1], Vishal Sharma[2*], Siddharth Singh[3], Vanshika Bhatt[4], Pankaj Agarwal[5]

[1,2,3,4,5]Dept. of Computer Science and Engineering, IMS Engineering College, Ghaziabad, Uttar Pradesh-201009, India

**Corresponding Author: vishal2910s@gmail.com, Mob.: +918439097019*

DOI: https://doi.org/10.26438/ijcse/v8i5.196200 | Available online at: www.ijcseonline.org

Received: 10/May/2020, Accepted: 20/May/2020, Published: 31/May/2020

Abstract— The Enhanced-Super Resolution Generative Adversarial Networks is an enhancement of Super-Resolution Generative Adversarial Networks by tweaking the model architecture to achieve high resolution. ISRGAN aims to further improve the quality of the image produced by the model by utilizing specially trained instances to upscale different portions of the image by enhancing each portion of the image by a model that is specially trained for such certain objects or classes. The idea is to divide and conquer the super-resolution problem utilizing the specialized models to up-scale sub-problems and improving the quality of generated images. Firstly image is passed through the Object detection phase which utilizes the Yolov3 structure to identify different classes present in the object, each class is then given to a generator that is specialized in that domain to further improve the quality. For the objects having unidentified classes or the base background image, we will have a generalized generator which will be trained on a combination of different domains. Also, to reduce the hardware requirement and improve the efficiency, we developed a way to split the images into sub-images to be enhanced individually and combined together to obtain the final image. These small images are in the form of squares which are enhanced and with the help of specialized generators and base models it is intended to convert low-resolution images into higher resolution models by up-scaling them to 4 times.

Keywords — Super Resolution, Generative Adversarial Networks, Image enhancement, Upscaling

I. INTRODUCTION

Single image super-resolution (SISR), is fundamentally a low resolution enhancement problem quite popular in research communities and Artificial Intelligence-based companies and communities. The original SRCNN paper proposed by Dong, a deep neural network (CNN) can be used to improve image quality and also the Peak Signal to Noise Ratio value [5]. ISRGAN is intended to improve the performance in the field of Super-Resolution. Instead of using a single Deep Neural Network, ISRGAN utilizes multiple Neural Networks all together and with the help of Object Detection algorithm Yolov3 utilizing base, it intends to integrate the functionality of both Networks to improve the quality of the image produced while up-scaling.

Bicubic ISRGAN

Fig. 1: The Bicubic Interpolation of 4x and the proposed ISRGAN

Several perceptually driven models have been proposed to enhance the resolution and visual quality. Such as perceptual loss proposed to be used which utilizes feature space instead of pixel space. Generative Adversarial Networks feature more realistic images as they are trained to fool the classifier network that differentiates images part of the data set from generated images by the generative network [9]. It provides more realistic images with the help of the GAN framework and the aggregation of loss functions being used [2]. Loss function consists of mean squared error which minimizes the intensity variation between the super-resolution generated image and the ground truth and also the perceptual loss or VGG loss to keep the images closer to the ground truth.

Instead of using the model trained upon all different sorts of images, it tackles this problem by first keeping the specially trained models for particular domains and then utilizing specially trained models to up-scale images for desired domains. With the combination of multiple models working together to achieve super-resolution, the quality of the image is improved. This method intends to reduce the gap that is between the results of ESRGAN generated images and the ground truth by utilizing the models to their fullest and more [1].

The perceptual quality and detailing are judged by the non-reference measure of Ma's score and NIQE that is perceptual index = 1⁄2 ((10-Ma) + NIQE) [1]. This provides good quality for perceptual detailing and ensures the generated images seem real and not generated or fake. Thanks to its parallel architecture, it is able to have different structural similarity indices as each specialized generator and its discriminator are deemed with freedom and provided with intensive training specialized in particular domains to outperform general models that are trained on all sorts of domains.

The background study in the field of super-resolution of images includes many types of research from normal image upscaling using the Nearest Neighbor technique, Bilinear and Bicubic interpolation upscaling to using deep convolutional neural networks for generating high-resolution images from the low-resolution images. Over time there have been many models developed like SRResNet, SRGAN, ESRGAN that use a single Generator and discriminator function of GANs and produce HR images.

In contrast to the previous research in the implementation of GANs for super-resolution of images, our research focuses on the specific Generator function trained on specific object class to produce 4 times upscaled HR image. A model specifically trained on one domain will perform better as compared to models that are trained on all domains as the difference is similar compared to "Specialization " and "Generalization".

In order to balance visual detailing and Peak Signal to Noise Ratio, we further propose the increase in specialized models to be co-existing with other models and also the Image Detection to be more thorough with high accuracy to fully utilize the image up-scaling with specialized models and appreciate the enhancement.

Rest of the paper is organized as follows, Section I contains the introduction of this paper (ISRGAN), Section II contains the related work done in the field of super-resolution of the image using GANs and other approaches, Section III contains the methodology behind the proposed approach in terms of network architecture, individual loss and working procedure of the essential steps involved in super-resolution of the image using our approach (ISRGAN), Section IV describes results and discussion along with the comparison of the results from previous techniques of image up-scaling, Section V concludes research work with future directions.

II. RELATED WORK

We focus on the Deep Neural Network to solve the Super Resolution problem [5]. Dong proposed SRCNN to learn from LR to HR images in an end-to-end manner, achieving superior performance against previous works, it utilized the mean squared error and penalized the model solely based on individual pixel intensities and their differences [5]. This leads to problems with clarity of image since the model can predict any image from the Hyper Plane of Mean Squared Error instead of generative images closer to the ground truth. B. Lim proposed the EDSR model by removing unnecessary Batch Normalization layers in the residual blocks and making the entire model a little bit deeper in order to get more complex equations and generate upscaled images with more overall clarity and quality [6],[7]. Other methods like reinforcement learning, Q-Learning, and even unsupervised learning are also introduced to solve general image restoration problems.

There have been multiple methods that have been proposed to stabilize training a very deep model. The residual path is a good example as it was developed to stabilize the training and improve performance. Residual scaling is first employed by Szegedy. and also used in EDSR. For general deep networks, V. Ferrari discussed a robust initialization method for VGG-style networks without BN [8]. To facilitate training a deeper network, we develop a compact and effective residual-in-residual dense block, which also helps to improve the perceptual quality [1].

To improve the visual quality of SR results, perceptual driven approaches have been proposed, and based on the idea of being closer to perceptual similarity, perceptual loss is proposed to enhance the visual quality which is done by minimizing the error in a feature space instead of the pixel space. Feature distribution is a better point of focus for achieving significant improvement so we expand contextual loss in order to generate images that look more natural. C. Ledig and others proposed the SRGAN model that uses perceptual loss along with adversarial loss in order to choose those image outputs that look more natural [2]. V. Ferrari and colleagues discussed a similar approach and looked further into the local texture matching loss [8]. Building upon these previous works, Wang and colleagues proposed a spatial feature transform to constructively incorporate semantic prior in an image and to also improve the reclaimed textures [8].

Photo-realism is attained by adversarial training with GAN as the generator network yields more and more realistic images [2],[7]. There has been some recent research that looks to make GANs more effective. One of these works is WGAN. WGAN the minimization and reasonable approximation of distance discriminator by performing weight clipping. A relativistic discriminator is developed to increase the probability that generated data are real and to also simultaneously decrease the probability that real data are real and so we look to enhance SRGAN by using a more effective relativistic average GAN [2],[7].

SR algorithms are usually evaluated by several common distortion measures such as PSNR and SSIM. However, these metrics don't necessarily agree with the subjective evaluation of human observers. Non-reference measures are used for perceptual quality evaluation, including Ma's score and NIQE, both of which are used to calculate the perceptual index in the PIRM- SR Challenge. In a recent study, Blau and others found that the distortion and perceptual quality are clashing with each other [8].

III. METHODOLOGY

Our main goal is to improve the overall perceptual detailing and quality for Super-Resolution [5]. In this section, we first describe our proposed neural network architecture and then discuss the improvements for both the super-resolution and image quality [5]. We use different specialized perceptual loss and discriminators and at last, we describe the PSNR and overall quality with the use of proposed architecture.

1. Network Architecture

Network architecture includes a number of specialized containers for each domain, called classes, which will be used and these individual generators will be given work for their own domains to be fulfilled by each and every working generator to achieve Super-resolution [5]. With the help of object detection, each generator is assigned its work corresponding to its domain.

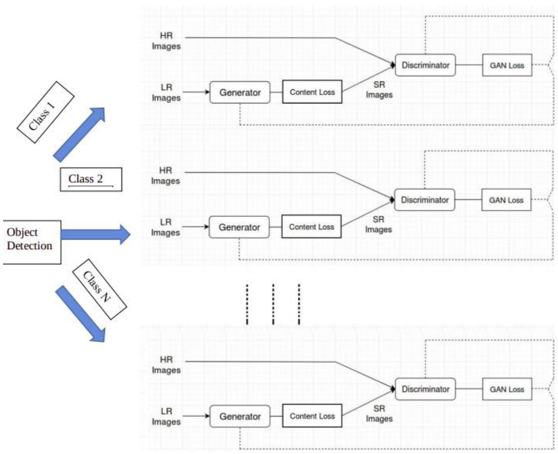

Fig.1

In order to reduce the computational complexity, eliminating BN layers is an efficient solution that also increases the performance of PSNR oriented tasks. BN layers normalize the features with the help of metrics such as mean and variance in a batch during the training process and also levy estimated mean and variance of the entire training data set during testing. If there occurs a conflict in the statistics of training and testing data sets, it may result in artifacts when the network is deeper and trained using a GAN framework. These Artifacts sometimes occur among iterations and various settings, infringing the need for a stable performance overtraining. To overcome this we chose to eliminate BN layers in order to achieve consistent performance. This further helps us in bettering the overall generalization ability and also reduces computational complexity with lesser memory usage.

The proposed RRDB employs a deeper and more complex structure than the original residual block in SRGAN [2],[7], as it has been known that more layers and connections result in better performance, which is the residual-in-residual structure where residual learning is used in different levels. A similar network structure is proposed which also applies a multi-level residual network. However, our RRDB differs from that in that we use dense blocks in the main path, where the network capacity becomes higher benefiting from the dense connections.

In addition to the improved architecture, we also exploit several techniques to facilitate training a very deep network such as residual scaling which is the scaling down of residuals by multiplying a constant between 0 and 1 before adding them to the main path to prevent instability. We also used smaller initialization, as we experimentally find that residual architecture is easier to train when the initial parameter variance becomes smaller.

2. Individual Loss

We develop individual loss which is an expression used for different loss for each individual model, for instance, a model trained on domain 1 has a different set of properties like different structures and hence different structural similarity index loss of perceptual loss. ISRGAN utilizes two losses that follow common properties for each GAN, which are a distance from reality and distance from fake-generated images. This summarizes how far the image is on the real axis and how real it actually looks which ignores how close it is to the ground truth.

Therefore, the total loss for the generator is:

$$LG = LPERCEP + + \eta L1,$$

Where L1 = EXI ||G() − y||1 is the content loss that evaluates the 1-norm distance between recovered image G() and the ground-truth y, and, are the coefficients to balance different loss terms. This loss is to be calculated for each individual GAN since there is a collection of used in ISRGAN.

We also explore a variant of the perceptual loss in the PIRM-SR. As opposed to the commonly used perceptual loss that adopts a VGG network trained for image classification, we develop a more suitable perceptual loss for SR MINC loss which is based on a fine-tuned VGG network for material recognition. This focuses on textures more rather than the object. Although the gain of the perceptual index brought by MINC loss is marginal, we still believe that exploring perceptual loss that focuses on texture is critical for SR.

3. Working Procedure

First Phase is object recognition and detection phase:
Image is passed through an image detection algorithm (yolov3 [3],[4] from) which is utilized to find different objects within the given input image with a high amount of confidence, these coordinates are recorded along with their class names to which they belong. This piece of information is crucial since it is responsible for choosing which generator to select for the up-scaling of that particular sub-portion of image.

Fig.2

Above is a demonstration of object detection which identifies objects with an almost high level of confidence and those sub-images or portions of the image are then utilized in the next phase.

Second Phase is utilization of individual :
In this phase, information received after the first phase is processed which includes coordinates or sub-images within the images to be processed using different generator models. These generator models are specialized in that particular domain and hence perform superior to the base model. Each sub-image or portion of the image is extracted and enhanced separately using individual Super-resolution [5] generators and after all separate enhancements are done, the final result includes these images to be combined and superimposed together to form the final result.

IV. RESULTS AND DISCUSSION

We analyzed the performance of our model based on two well-known objective image quality metrics, the peak signal-to-noise ratio (PSNR) as well as the structural similarity index measure (SSIM).

Results from previous up-scaling techniques

- Result of Nearest Neighbor Up-Scaling (Previous technique):

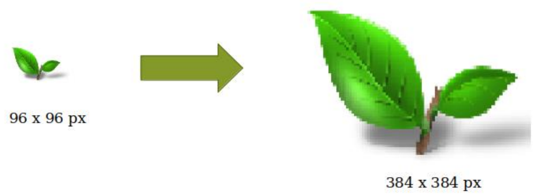

Fig.3
PSNR/SSIM = 22.98 dB / 0.7952

- Result of Bi-Cubic Interpolation Up-Scaling (Previous Technique):

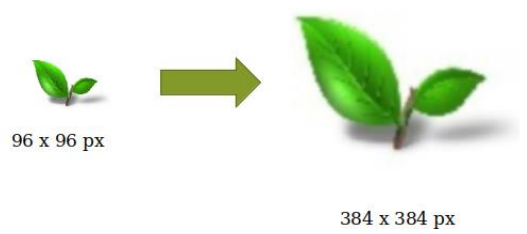

Fig.4
PSNR/SSIM = 22.36 dB / 0.8052

Below is the comparison of the result obtained from our model (ISRGAN) and the output from some previous models. For comparison of the results, we evaluated the quality of the image using the above-mentioned performance metrics which are PSNR values and SSIM values.

Table 1 Performance Comparison table:

Method	*Dataset*	*PSNR*	*SSIM*
SRGAN	Div2k	23.59 dB	0.8578
Bicubic	Div2k	22.66 dB	0.8025
SRCNN	Div2k	23.14 dB	0.8280
ISRGAN (ours)	**Div2k**	**24.72 dB**	**0.8982**

Using the above two performance parameters, the quality of the image which is obtained from our model is calculated and are as follows:
PSNR-value : 24.72 dB
SSIM-value : 0.8982
So our model (ISRGAN) is performing 4.8% better than the previous SRGAN model.

- Result of ISRGAN Up-Scaling (our Model):

Fig.5
PSNR/SSIM = 24.54 dB / 0.8915

V. CONCLUSION AND FUTURE SCOPE

We have presented an ISRGAN model that achieves considerably better results when compared to previous SR methods in terms of perceptual quality and detailing. This model utilizes the previous ESRGAN [1] model which performs without batch normalization layers and in-addition this model is having parallel architecture to perform better when recovering lost details and textures from low-resolution images that were supposed to be present in the high-resolution image. Using multiple generators for each object class present in the image helps to map the low-resolution image to the 4X the upscaled image by providing more detailing and textures in the upscaled image. Moreover, we've reduced the hardware consumption and increased hardware efficiency by utilizing the split image method to process the image in sub-problem fashion. The model is feasible as the improvement in the quality of upscaled images is substantial enough to put out in the work of the upscaling of images.

ACKNOWLEDGMENT

This work is done by team Imagenics under the mentorship of Pankaj Agarwal, Professor and Head, Department of Computer Science and Engineering, IMS Engineering College, Ghaziabad, India.

REFERENCES

[1] *X. Wang, K. Yu, S. Wu, J. Gu, Y. Liu, C. Dong, Y. Qiao, C.C. Loy*, "ESRGAN: Enhanced Super-Resolution Generative Adversarial Networks", In the Proceedings of the 2018 European Conference on Computer Vision, Munich, Germany.

[2] C. Ledig, L. Theis, F. Husz´ar, J. Caballero, A. Cunningham, A. Acosta, A. Aitken, A. Tejani, J. Totz, Z. Wang, S. Wenzhe, "Photo-realistic single image super-resolution using a generative adversarial network", In the Proceedings of the 2017 Conference on Computer Vision and Pattern Recognition.

[3] J. Redmon, A. Farhadi, "YOLOv3: An Incremental Improvement ", In the Proceedings of the 2018 IEEE Conference on Computer Vision and Pattern Recognition Workshops.

[4] J. Redmon, S. Divvala, R. Girshick, A. Farhadi, "You Only Look Once: Unified, Real-Time Object Detection", In the Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops.

[5] C. Dong, C.C. Loy, K. He, X. Tang, "Learning a deep convolutional network for image super-resolution", In the Proceedings of the 2017 European Conference on Computer Vision.

[6] B. Lim, S. Son, H. Kim, S. Nah, K.M. Lee, "Enhanced deep residual networks for single image super-resolution", In the Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops.

[7] Y. Wang, F. Perazzi, B. McWilliams, A. Sorkine-Hornung, O. Sorkine-Hornung, C. Schroers, "A Fully Progressive Approach to Single-Image Super-Resolution", In the Proceedings of the 2018 IEEE Conference on Computer Vision and Pattern Recognition Workshops.

[8] V. Ferrari, M. Hebert, C. Sminchisescu, Y. Weiss, "Computer Vision" In the Proceedings of the 2018 European Conference on Computer Vision, Munich, Germany.

[9] N.S. Lele, "Image Classification Using Convolutional Neural Network", International Journal of Scientific Research in Computer Science and Engineering, Vol.6, Issue.3, pp.22-26, 2018

Authors Profile

Mr. Yash Bansal is currently pursuing his bachelor in Computer Science and Engineering from IMS Engineering College, Ghaziabad, Uttar Pradesh, India since 2016.

Mr. Vishal Sharma is currently pursuing his bachelor in Computer Science and Engineering from IMS Engineering College, Ghaziabad, Uttar Pradesh, India since 2016.

Mr. Siddharth Singh is currently pursuing his bachelor in Computer Science and Engineering from IMS Engineering College, Ghaziabad, Uttar Pradesh, India since 2016.

Mrs. Vanshika Bhatt is currently pursuing his bachelor in Computer Science and Engineering from IMS Engineering College, Ghaziabad, Uttar Pradesh, India since 2016

Dr. Pankaj Agarwal, Professor, and Head, Department of Computer Science and Engineering, IMS Engineering College, Ghaziabad, Uttar Pradesh, India. He completed his Ph.D. in Computer Science from Jamia Millia Islamia Central University, Delhi in 2009.

He has a teaching experience of more than 19 years in engineering colleges and research experience of more than 14 years. He has more than 45 research publications in international journals/conferences of repute to his credit. His areas of interest & research include Algorithms in Computational Biology, Soft computing Techniques, Parallel Algorithms, Machine Learning using R & Python

