Super-Resolution Image Generation using GANs

1st Ramtin Asgarianamiri
Electrical and Computer Engineering Department
Toronto Metropolitan University
Toronto, Canada
ramtin.asgarianamiri@torontomu.com

***Abstract*—This report outlines the Super-Resolution Image Generation using GANs. Single-image super-resolution facilitates the enhancement of low-resolution images to higher resolutions. Many recent advancements in generative adversarial networks (GANs) exhibit promising results with limited data. This report, by reviewing related previous works on this specific subject, not only implements a model but it also introduces the elements of the model. The dataset contains both low-resolution and high-resolution pictures to train the generative part of the model and can give the discriminator of the model, the perspective on how to punish the model in terms of having better accuracy in training.**

***Index Terms*—Super-Resolution, GAN, Discriminator, Deep learning, SRGAN**

I. Introduction

Image super-resolution is a relevant problem in image processing[1]. Super-resolution image generation addresses the reproducing of high-resolution (HR) images from low-resolution (LR) ones, founded important among various domains like computer vision and medical imaging. Traditional methods involve complex tasks such as image registration and interpolation, demanding significant computational resources and time. Generative Adversarial Networks (GANs) offer a promising solution, capturing patterns and structures in data. GANs address this challenge through the utilization of perceptual loss, guiding the image reconstruction process towards the natural image manifold, thereby generating solutions that are perceptually more realistic and convincing. Presently, CNN and GAN-based techniques are widely utilized for image super-resolution (SR) reconstruction. Key CNN-based methods for SRCNN (super-resolution convolutional neural network), VDSR (very deep convolutional networks for super-resolution), and EDSR (enhanced deep residual networks for super-resolution).[2,3] While these methods outperform conventional bicubic interpolation techniques, they still exhibit room for improvement, resulting in less pronounced reconstruction effects. Introduced by Goodfellow et al. [4] in 2014, the generative adversarial network (GAN) has emerged as a powerful deep learning model. In recent years, GANs have demonstrated significant potential for unsupervised learning tasks involving complex distributions. Since its inception, GANs have attracted considerable attention from academia and industry alike, leading to rapid advancements in both theoretical understanding and practical model construction. With diverse applications in fields such as computer vision and human-computer interaction, the technology continues to evolve, promising innovative solutions for various challenges. The super-resolution process is not only up sampling the images but additional debasing causes such as noise, blur, or decimation can also be important for different applications. In this line of work, this article presents a comparison of GANs for the image super-resolution problem. Successful GANs architectures are selected from the review of relevant related works. The quality of images was evaluated for a standard dataset.The main results indicate that SRGAN models are able to compute accurate results, within the range of state-of-the-art results, and good transfer capabilities.

II. Literature Survey

In 2009, The study [5] employed Brain MRI images of both (32 x 32) and (128 x 128) resolutions, splitting the dataset with 90 percent for model training and 10 percent for testing. Evaluations occur every 5 epochs, focusing on performance metrics such as Peak Signal-to-noise ratio (PSNR), Mean Squared Error (MSE), and Structural Similarity Index (SSIM). As the epochs progress, an improvement in SSIM is noted, culminating in the successful production of desired (128 x 128) high-resolution images from their low-resolution counterparts. This comprehensive approach demonstrates the efficacy of the proposed GAN architecture, supported by the evaluation metrics and the practicality of the accompanying web application.

Image Super-Resolution using an Improved Generative Adversarial Network [6], published in 2019, focuses on Super-Resolution (SR) problems, utilizing the DIV2K dataset of 800 2K HD images ideal for GAN-based SR and training a GAN network to recover 4x down-sampled images, cropping inputs to 384*384 and down-sampled sizes to 96*96 with a batch size of 16. By removing batch normalization layers, training stability, and speed were enhanced and the new encoder generator structure improved image detail realism, using 3x3 kernel convolutional layers with ParametricReLU activation. Overall, their model generated sharper, clearer images with more natural structures than existing methods.

Generative Adversarial Network with Residual Dense Generator for Remote Sensing Image Super Resolution [7] introduces RDGAN which combines residual dense networks with generative adversarial training, outperforming SRGAN with a 1.67 dB increase in PSNR and 0.04 in SSIM, though decreasing by 2.065 on NIQE due to MSE limitations in capturing perceptual differences. This method enhances super-resolved image quality in perceptual quality and traditional metrics.

Generative Adversarial Networks for Image Super-Resolution [8] suggests enhancing network architectures, reducing complexity. Proposed solutions include enhancing stability through attention mechanisms and lightweight architectures, utilizing self-supervised methods for reference image generation, decomposing complex tasks, and employing image quality assessment metrics.

In 2013, A paper[9] brought up the comparison between SR-GAN and ESRGAN.In detailed assessments, ESRGAN shines in specific image features, such as water droplets and sports outfits, surpassing even the impressive results of SRGAN and SRResNet in certain instances. These results indicate the promising capabilities of both SRGAN and ESRGAN in super-resolution tasks, with ESRGAN particularly standing out for its attention to nuanced details and superior image quality.

In Improving Image Quality Using Deep Learning Based Super Resolution [10], it conducted a comparative experiment between the SRGAN, an established model, and the proposed BSRGAN.BSRGAN exhibited stable decreases in generator loss and pixel-related loss, with a significantly larger discriminator loss than SRGAN, indicating BSRGAN's ability to generate more plausible images.

Super-resolution fundus images using GENs [11] introduces RFISRGAN, a retinal fundus image super-resolution technique employing GANs for medical applications. RFISRGAN focuses on reconstructing high-resolution retinal images from low-resolution inputs, achieving a 4x super-resolution.

In 2023, Xuan Wang et al [12] provided an overview of GAN techniques for super-resolution reconstruction, emphasizing remote sensing images. It mentions ISRGAN and TE-SAGAN, which address issues like blurred edges and artifacts. ISRGAN enhances stability and generalization, while TE-SAGAN incorporates self-attention and texture refinement for improved feature extraction and training stability.

Optimal Design of Color Laparoscopic Super-Resolution Image Quality Based on Generative Adversarial Networks [13] discussed the relationship between image quality and network learning by analyzing loss functions, indicating the effectiveness of SRGAN in generating laparoscopic super-resolution images.

Structure-Aware Deep Networks and Pixel-Level Generative Adversarial Training for Single Image Super-Resolution [14] introduces a novel image super-resolution network comprising two branches: a structure-aware gradient generation branch and a structure-aware super-resolution branch. It proposes a hyperparameter-free non-local self-similarity module for global feature acquisition. Addressing the limitations of image-level generative adversarial training, the paper introduces pixel-level adversarial loss alongside reconstruction and perceptual losses.

III. Problem Statement

In the modern world, high-resolution images are of great importance across various aspects of life. They work as powerful tools for communication, conveying complex ideas and emotions in a visually appealing and easily understandable format. From marketing and branding to education and research, high-quality images play a crucial role in capturing attention, enhancing understanding, and creating memorable experiences. In the digital age, they are essential for engaging social media audiences, showcasing products in e-commerce, aiding learning in education, and documenting historical moments for future generations. High-resolution images are a crucial aspect of technological advancements, ranging from medical imaging and space exploration to engineering and design. They facilitate precise analysis, accurate diagnosis, and detailed documentation, driving progress in scientific fields. In the field of journalism and storytelling, these images serve as visual anchors, capturing the essence of events and preserving cultural heritage. Image super-resolution using advanced deep learning techniques such as Generative Adversarial Networks (GANs) and Convolutional Neural Networks (CNNs) presents a significant opportunity for educational, academic, and social contributions. This project addresses the challenge of enhancing the resolution of low-quality images, aiming to impact these domains positively.

IV. Proposed models

Super Resolution Generative Adversarial Neural Networks (SRGANs) are a type of deep learning network that falls under the family of GANs. They are designed specifically to improve the resolution of images. Using SRGANs instead of traditional methods like bilinear interpolation leads to higher-quality images with enhanced features.

When the resolution of an image is increased using the SRGAN model, it adds the necessary information to the image to correctly enhance its features. This information is learned by the SRGAN model during training. The SRGAN consists of two main components: the Generator and the Discriminator, which are similar to other GAN designs.

The generator network takes small, low-resolution images and creates larger, high-resolution images, which we call the "generated image." On the other hand, the discriminator block acts like a yes-or-no judge, deciding whether an image is real or fake. It works with the generator's output as input.

To help in this process, we use the VGG-19 architecture with pre-set weights. This architecture helps by extracting important features from the images. Both the generator and discriminator networks work together during training to achieve the goal of turning low-resolution images into high-resolution ones.

A. The generator

The generator architecture includes an initial convolutional layer followed by ParametricReLU activation. It then incorporates residual blocks from the ResNet architecture. Each residual block consists of multiple convolutional layers with

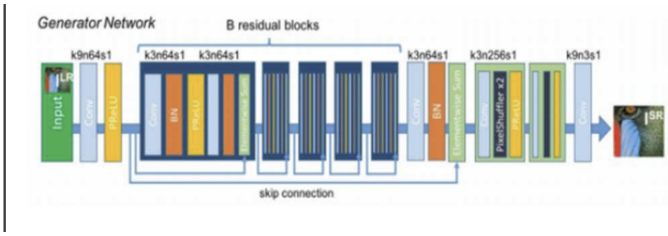

Fig. 1. Generator architecture. [11]

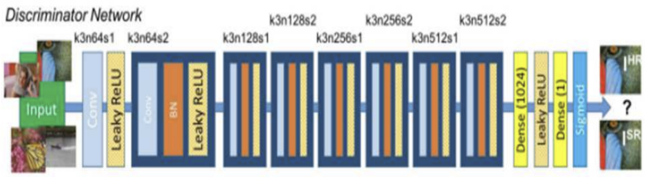

Fig. 2. discriminator architecture. [11]

batch normalization and ParametricReLU activation. These blocks have skip connections, allowing information to flow effectively through the network.(figure 2)

B. The discriminator

The discriminator architecture is a binary classifier that classifies if the input image is real or fake.In this context, the term "real" refers to the ability of the generator to produce an output that is similar to the high-resolution image in the dataset. The figure 3 image describes the architecture of the discriminator network. Distinguishing between genuine HR photographs and artificial SR images is the discriminator's job.

V. Experimnet

A. Dataset

In the context of Super-Resolution Image Generation, providing the model with a suitable dataset can be fulfilled through different approaches. In selecting an appropriate dataset, it's essential to have a set of images with low-quality and high-quality versions. This collection of images, called a dataset, plays a big role in how well your model performs and how useful it is. The Dataset which specifically has been introduced to be used and imported in models for the high-resolution problem is called DIV2K. [15]

DIV2K is a popular single-image super-resolution dataset that contains 1,000 images with different scenes and is split into 800 for training, 100 for validation, and 100 for testing. It was collected for NTIRE2017 and NTIRE2018 Super-Resolution Challenges to encourage research on image super-resolution with more realistic degradation. This dataset contains low-resolution images with different types of degradations. Apart from the standard bicubic downsampling, several types of degradations are considered in synthesizing low-resolution images for different tracks of the challenges.

Fig. 3. DIV2K Dataset sample(HR and LR)

Track 2 of NTIRE 2017 contains low-resolution images with unknown x4 downscaling. Track 2 and track 4 of NTIRE 2018 correspond to realistic mild ×4 and realistic wild ×4 adverse conditions, respectively. Low-resolution images under realistic mild x4 settings suffer from motion blur, Poisson noise, and pixel shifting. Degradations under realistic wild x4 setting are further extended to be of different levels from image to image. In this project, we have utilized the dataset and aimed to enhance the resolution of the image by upsampling. Figure 1 is an example that shows both high and low-resolution images from the dataset.

B. Generator

The process starts with a low-resolution image as input. This could be an image that is, for this project, 128x128 pixels. The input image is first passed through an initial processing step. This step is designed to extract basic features from the low-resolution image. These features might include edges, basic shapes, and general structures present in the image. Next, the image goes through a series of blocks called ResidualBlocks. Each ResidualBlock is like a mini-network within the Generator. These blocks learn and refine detailed features of the image. They capture intricate patterns, textures, and nuances that make up the image content. The information from each ResidualBlock is then combined with the original image. After going through multiple ResidualBlocks, the image features are further refined. Another convolutional block, often called conv, works to enhance the image quality. It sharpens edges, enhances textures, and improves overall visual appearance. This stage is crucial for making the generated image look more realistic and detailed. The refined image features are then passed through an upsampling process. Upsampling increases the size and resolution of the image. This step helps to restore lost details that were downscaled in the low-resolution input. By increasing the image dimensions, it creates a larger canvas for finer details to be added. The upscaled image features are fed into a final convolutional layer. This layer acts as a mapping from the enhanced features to the high-resolution output. It generates the final image that is now much larger and more detailed than the input. The output size is often chosen to match the desired high resolution, our output image is 256x256 pixels. Before the final image is produced, an activation function like sigmoid is applied. This function scales the pixel values to a range between 0 and 1. It ensures that the output image has valid pixel intensities and is ready for display or further processing. The result is a high-resolution image that is detailed, realistic, and suitable for various applications.

C. *Discriminator*

Discriminator acts like a critic in the neural network world, looking at images and deciding if they're real or fake. When an image is given to a Discriminator, it goes through a series of convolutional blocks (ConvBlocks) initially. These blocks help the Discriminator learn about the shapes, edges, and patterns in the image, understanding its basic features. After this analysis, the learned features are then passed through a Multi-Layer Perceptron (MLP). The MLP takes these learned features and makes the final call: Is this image close to the test image, like a high-resolution photo, or not? This process of breaking down the image into features and using the MLP for judgment helps the Discriminator decide the authenticity of the image. Crucially, Discriminator plays a pivotal role in Generative Adversarial Networks (GANs). It aids the Generator in improving its images by providing feedback. By training Generator to fool Discriminator, we get progressively better and more realistic images over time. Essentially, the Discriminator is the discerning eye that helps guide the Generator toward creating images that look nearly indistinguishable from real ones, expanding the realm of artificial creativity within neural networks.

D. *Implemenation*

The model was built by including all the specified architectural details mentioned earlier. The dataset was used to train the model, with the following hyperparameters set: epoch = 30, batch size = 6.

In Figure 4-6, we can see the differences between the low-resolution image and the predicted output from the model, as well as the original high-resolution image in the dataset. This comparison highlights the model's progress in identifying edges and finer details, thus improving the resolution of the input. The learning process occurs through adjustments made by the discriminator, demonstrating how the model gradually improves its ability to discern intricate features over time.
Discriminator and generation loss have been extracted from the training for each epoch so it could be an element of discrimination in terms of model evaluation.[Figure 7] In a Generative Adversarial Network (GAN), the Generator Loss measures how effectively the generator can deceive the discriminator by encouraging the production of samples that closely resemble real data, often computed using binary cross-entropy or mean squared error loss. Conversely, the Discriminator Loss assesses the discriminator's ability to differentiate between real and fake samples, penalizing misclassifications with loss functions like binary cross-entropy or hinge loss. Through iterative training, the generator minimizes its loss while maximizing the discriminator's loss, fostering an adversarial dynamic

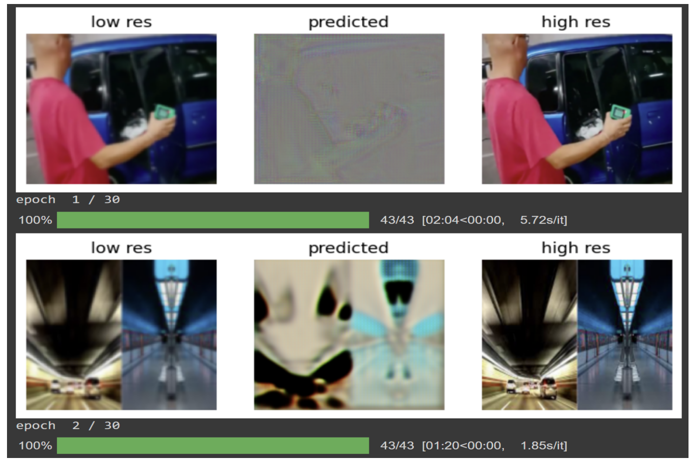

Fig. 4. Epoch one and two (Input of model, Output of generator and the Target(HL))

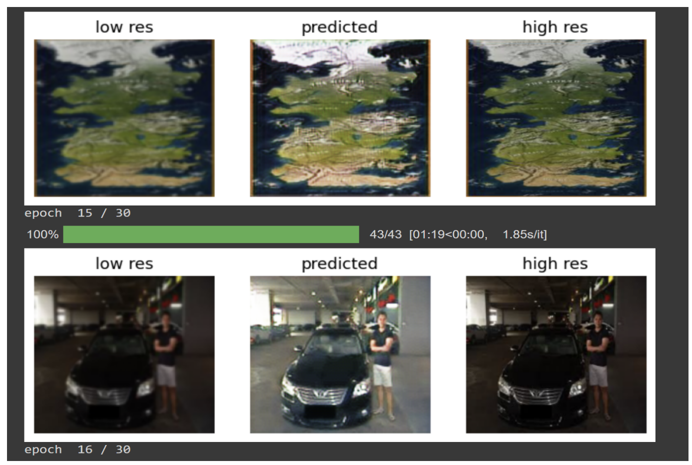

Fig. 5. Epoch fifteen and sixteen (Input of model, Output of generator and the Target(HL))

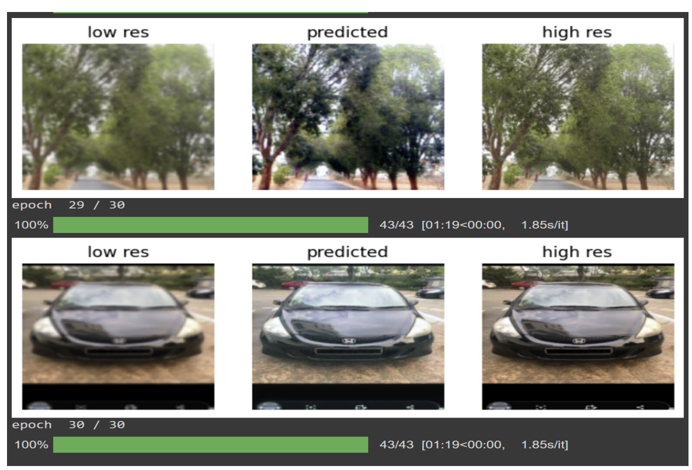

Fig. 6. Last two epochs (Input of model, Output of generator and the Target(HL))

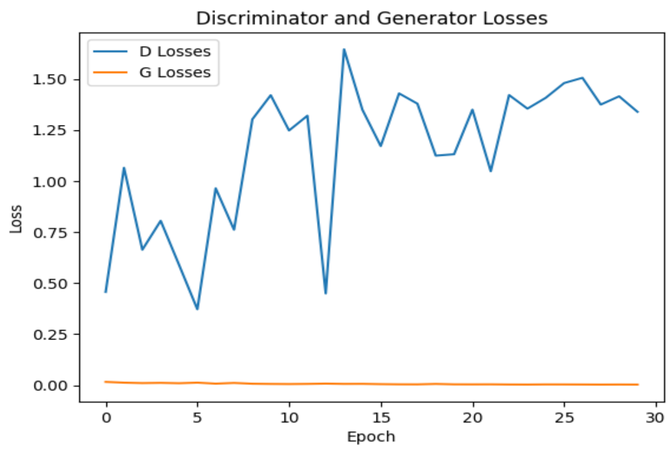

Fig. 7. Evaluation of the model in 30 epochs

Fig. 8. visual testing of the model by self-captured picture

where the generator generates increasingly realistic samples, deceiving the discriminator, and the discriminator improves its discrimination skills. The overarching objective is for the generator to create samples indistinguishable from real data. In the end, I have taken a photo of myself to be given to the model as a low-resolution picture to be tested by an image out of the dataset, figure 8 illustrates both the input and output of the model.

VI. Conclusion

References

[1] P. Milanfar, Ed., Super-Resolution Imaging. CRC Press, 2017

[2] Wang, Z.; Chen, J.; Hoi, S.C. Deep learning for image super-resolution: A survey. IEEE Trans. Pattern Anal. Mach. Intell. 2020, 43, 3365–3387.

[3] Liebel, L.; Körner, M. Single-Image Super Resolution For Multispectral Remote Sensing Data Using Convolutional Neural Networks. Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci. 2016, 41, 883–890.

[4] Goodfellow, I.; Pouget-Abadie, J.; Mirza, M.; Xu, B.; Warde-Farley, D.; Ozair, S.; Courville, A.; Bengio, Y. Generative adversarial networks. Commun. ACM 2020, 63, 139–144.

[5] H. Greenspan, "Super-Resolution in Medical Imaging," in The Computer Journal, vol. 52, no. 1, pp. 43-63, Jan. 2009, doi: 10.1093/comjnl/bxm075.

[6] H. Wang, W. Wu, Y. Su, Y. Duan and P. Wang, "Image Super-Resolution using an Improved Generative Adversarial Network," 2019 IEEE 9th International Conference on Electronics Information and Emergency Communication (ICEIEC), Beijing, China, 2019, pp. 312-315.

[7] R. Sustika, A. B. Suksmono, D. Danudirdjo and K. Wikantika, "Generative Adversarial Network with Residual Dense Generator for Remote Sensing Image Super Resolution," 2020 International Conference on Radar, Antenna, Microwave, Electronics, and Telecommunications (ICRAMET), Tangerang, Indonesia, 2020, pp. 34-39.

[8] Tian, Chunwei et al. "Generative Adversarial Networks for Image Super-Resolution: A Survey." ArXiv abs/2204.13620 (2022).

[9] P. Cobelli, S. Nesmachnow and J. Toutouh, "A comparison of Generative Adversarial Networks for image super-resolution," 2022 IEEE Latin American Conference on Computational Intelligence (LA-CCI), Montevideo, Uruguay, 2022, pp. 1-6.

[10] D. Kim and R. Kyung, "Improving Image Quality Using Deep Learning Based Super Resolution," 2022 IEEE 13th Annual Information Technology, Electronics and Mobile Communication Conference (IEMCON), Vancouver, BC, Canada, 2022, pp. 0275-0279.

[11] D. P. B, A. Muthukumar and A. Lakshmi, "Super-resolution of Retinal Fundus Images Using Generative Adversarial Networks," 2022 Second International Conference on Next Generation Intelligent Systems (ICNGIS), Kottayam, India, 2022, pp. 1-4, doi: 10.1109/ICNGIS54955.2022.10079882.

[12] Wang, Xuan et al. "A Review of GAN-Based Super-Resolution Reconstruction for Optical Remote Sensing Images." Remote. Sens. 15 (2023): 5062.

[13] N. Kawabata and T. Nakaguchi, "Optimal Design of Color Laparoscopic Super-Resolution Image Quality Based on Generative Adversarial Networks," 2023 International Conference on Computer Graphics and Image Processing (CGIP), Tokyo, Japan, 2023, pp. 1-7.

[14] W. Shi, F. Tao and Y. Wen, "Structure-Aware Deep Networks and Pixel-Level Generative Adversarial Training for Single Image Super-Resolution," in IEEE Transactions on Instrumentation and Measurement, vol. 72, pp. 1-14, 2023, Art no. 5007614

[15] E. Agustsson and R. Timofte, "NTIRE 2017 Challenge on Single Image Super-Resolution: Dataset and Study," 2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), Honolulu, HI, USA, 2017, pp. 1122-1131, doi: 10.1109/CVPRW.2017.150.