Image Super Resolution using GAN framework

Abstract— **Image Super Resolution is a rising field of computer vision. It aims to enhance the resolution of an image from low to high. Through the development of technology and deep learning, the application of Image Super Resolution ranges from simple resolution enhancement in phones to analysis of medical data such as MRI. There are various ways to perform Image Super Resolution. This paper focuses on SRGAN and ESRGAN applications.**

I. Topic Introduction

Image Super Resolution is a significant system in computer vison and image processing to improve the visual perception of the poor-quality images. The primary goal is to convert blurred, unclear, low-resolution images into sharp, high-resolution images. This process is often referred to by terms such as interpolation, image scaling, enlargement, and upscaling [1]. There are two main approaches to super-resolution (SR): multi-frame and single-frame based on the input low-resolution (LR) data. Multi-frame SR utilizes multiple images of the same scene with slight sub-pixel shifts to leverage and reconstruct a higher-resolution image or sequence. However, in cases where multiple LR images are unavailable, limited LR data is used to generate the high-resolution (HR) image, an approach known as single-frame SR [2].

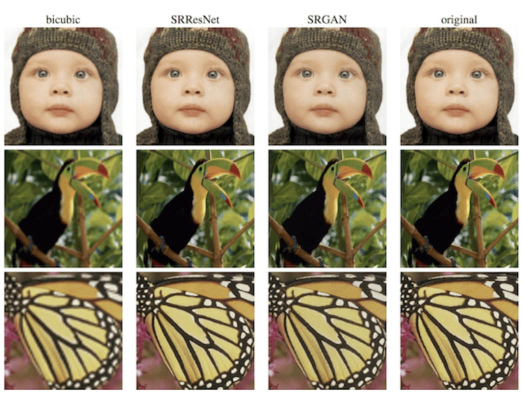

Fig. 1. Examples of Super Image Resolution [3].

II. Existing methods

There are various methods to perform Image Super Resolution as listed in Figure 1. Among all methods, this paper focuses on GAN-based Connections, specifically on SRGAN.

GAN or generative adversarial network is a deep learning model that uses two competing neural networks (generator and discriminator) to generate realistic new data based on a given dataset. The generator creates new data samples,

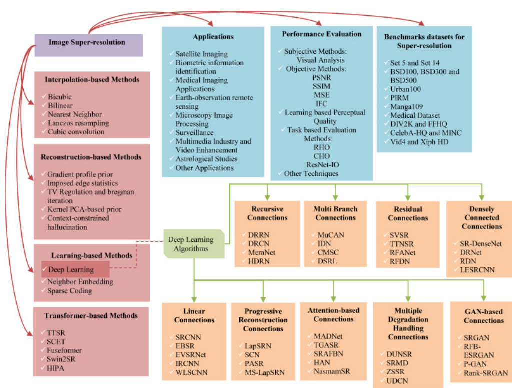

Fig. 2. Taxonomy of the existing state-of-the-art super resolution techniques [1].

while the discriminator evaluates if these samples are real or fake compared to the original data. Over time, the generator improves its output until the discriminator can no longer tell the difference between generated and real data [4].

SRGAN (Super-Resolution Generative Adversarial Network) applies a GAN framework to single image super-resolution (SISR). The generator creates high-resolution (HR) images from low-resolution (LR) inputs and a discriminator distinguishes between real and generated HR images. SRGAN enhances the image quality by using a perceptual loss that includes adversarial loss and content loss based on high-level features from a pre-trained VGG network, which is a approach that helps recover realistic image details [5].

ESRGAN (Enhanced SRGAN) improves on SRGAN by enhancing the generator and discriminator. The generator includes Residual-in-Residual Dense Block (RRDB) to increase stability and performance, while the Relativistic GAN discriminator compares how realistic generated images are against real ones, producing sharper and more detailed images. ESRGAN also improves the perceptual loss by using features in the VGG network before activation, improving brightness consistency and texture detail [6].

The evaluation of super-resolution methods is carried out using both objective and subjective assessments to measure image quality. Objective metrics include Peak Signal-to-Noise Ratio (PSNR), which calculates pixel-wise accuracy between super-resolved images and ground truth, and Structural Similarity Index (SSIM), which evaluates structural and perceptual similarities. Subjective evaluation involves techniques like Mean Opinion Score (MOS), where human raters evaluate the visual appeal of images, and qualitative assessments, which examine the sharpness, texture, and absence of artifacts. These methods collectively provide a comprehensive understanding of the effectiveness of super-resolution approaches in generating high-quality, realistic images [5], [6].

The similarities and differences of SRGAN and ESRGAN are summarized in Table 1.

TABLE I
COMPARISON BETWEEN SRGAN AND ESRGAN BASED ON ARCHITECTURAL AND PERFORMANCE IMPROVEMENTS.

Feature	SRGAN	ESRGAN
Network Architecture	Uses standard residual blocks with batch normalization [5].	Introduces Residual-in-Residual Dense Block (RRDB) and removes batch normalization, enhancing stability and feature extraction [6].
Adversarial Loss	Standard GAN discriminator that predicts if an image is real or fake [5].	Uses Relativistic GAN (RaGAN), evaluating the relative realism of generated images against real ones to enhance texture fidelity [6].
Perceptual Loss	Combines adversarial loss with content loss based on VGG network feature maps after activation, which can result in brightness inconsistency [5].	Uses VGG features before activation in perceptual loss, leading to improved brightness consistency and texture detail [6].
Detail and Texture Quality	Improves perceptual quality with photorealistic detail recovery but may still produce artifacts and lack fine detail [5].	Produces sharper and more natural textures, particularly in complex areas such as edges, fur, and grass [6].
Applications	Effective for general single-image super-resolution tasks where moderate visual quality is acceptable [5].	Ideal for applications needing high perceptual quality, such as medical imaging and satellite image enhancement, due to sharper detail recovery [6].

III. OPEN CHALLENGES

A. Current Limitations

Even though GAN model is widely used and well accepted in deep learning field, especially in super image resolution, there are several limitations. Common failures include mode collapse, where the generator maps different inputs to a single output class, resulting in low diversity among samples. Non-convergence can occur when training becomes unstable due to oscillations or divergence between the generator and discriminator. Additionally, vanishing or exploding gradients can slow down or completely halt the learning process entirely [7].

SRGAN challenges occur mostly due to network depth issue because SRGAN are harder to train and they often produce high-frequency artifacts, resulting in repetitive or noisy texture instead of realistic details. Thus, as network depth increases, it becomes harder to balance the adversarial and perceptual loss functions effectively. The deeper layers can exaggerate minor inaccuracies, resulting in repetitive or noisy textures instead of realistic image details. Also, The deeper layers of SRGAN can amplify small inaccuracies, leading to repetitive or noisy textures instead of realistic image details. Additionally, its deep architecture, while improving perceptual quality, requires significant computational resources [5].

ESRGAN often struggles to recover fine local details, leading to blurry or unnatural artifacts. This may be due to the limitations of the pre-trained VGG network, which may not fully capture the potential features in images, thus restricting the generator's capacity to produce highly accurate outputs [8].

Training both SRGAN and ESRGAN is resource-intensive and requires considerable GPU power, making it less practical for real-time applications or deployment on limited-resource devices [5], [6].

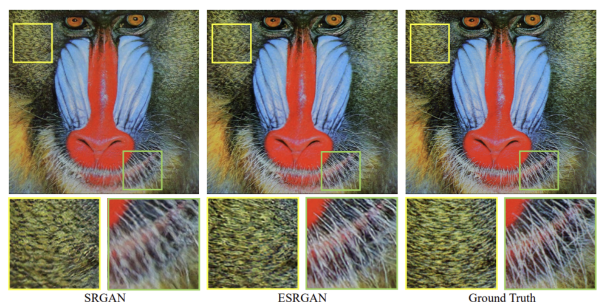

Fig. 3. Image Resolution Comparison between SRGAN, ESRGAN, and ground truth [6]. Ground truth typically represents the original high-resolution images that a model aims to replicate or approximate from low-resolution inputs.

B. Research Idea

This research aims to further develop and apply Image Super Resolution methodologies for enhancing microscopic brain imaging through GAN-Based Super-Resolution. High-resolution imaging of brain structures is crucial for neuroscience research, particularly in mapping complex neural networks and monitoring cellular changes. However, current imaging technologies face physical limitations that hinder the ability to capture ultra-fine details quickly or with minimal photo toxicity. This research seeks to develop a super-resolution model tailored for neuroscience applications, enabling clearer visualization of fine neural structures while reducing the need for ultra-high-magnification imaging. This might be achieved through utilizing techniques that adapt GAN-based models specifically for biological and microscopic imaging contexts such as model training with pre-learned features from high-resolution natural images and gradually fine-tuning it on neuroscience imaging data and incorporating noise reduction techniques or artifact suppression layers into the GAN.

IV. Concept of Code

This code implements the Enhanced Super-Resolution Generative Adversarial Network (ESRGAN) for super-resolving low-resolution images. The objective is to apply ESRGAN to enhance the resolution of specific images (e.g., MRI brain scans), enabling clearer visualization of fine details.

Dataset link: https://www.kaggle.com/api/v1/datasets/download/masoudnickparvar/brain-tumor-mri-dataset

Code link: https://www.kaggle.com/code/nnsssadithya/esrgan-enhanced-super-resolution-gan/comments

1) **Environment Setup and Import Libraries**
2) **Download and Set Paths for Dataset and Model**
3) **Define Constants and Paths**
4) **Image Preprocessing Function**
   - Preprocesses the high-resolution image to make it compatible with the ESRGAN model.
5) **Save and Plot Image**
6) **Load the ESRGAN Model and Generate Super-Resolution Image**
7) **Downscale Image Function (for comparison)**
8) **Display Low-Resolution Image and Calculate PSNR**
   - Calculates the PSNR between the super-resolved and original high-resolution images, providing a metric of quality.
9) **Final Visualization of All Images**
10) **Next Step**
    - As it can be observed, super resolution image is rather lower resolution due to artifact issues as mentioned in open challenges section.
    - The next implementation should be focused on how to improve the resolution when utilizing ESRGAN for medical MRI images.

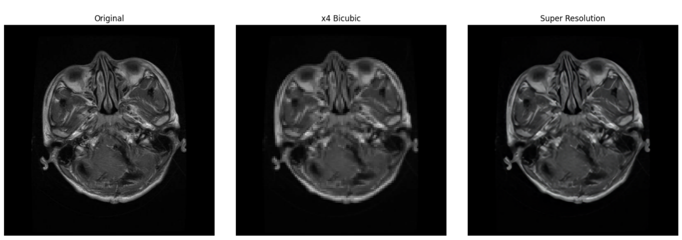

Fig. 4. Output images from sample code: Original, Low resolution, Super resolution.

References

[1] A. D. V. G. Dawa Chyophel Lepcha, Bhawna Goyal, "Image super-resolution: A comprehensive review, recent trends, challenges and applications," *Infomation Fusion*, vol. 91, pp. 230–260, 2023.

[2] J. L. Q. Y. H. Z. L. Z. Linwei Yue, Huanfeng Shen, "Image super-resolution: The techniques, applications, and future," *Signla processing*, vol. 128, pp. 389–408, 2016.

[3] J. Brownlee. (2019) 18 impressive applications of generative adversarial networks (gans). Accessed: 2024-11-11. [Online]. Available: https://machinelearningmastery.com/impressive-applications-of-generative-adversarial-networks/

[4] AWS. What is a gan? Accessed: 2024-11-12. [Online]. Available: https://aws.amazon.com/what-is/gan/?nc1=h_ls

[5] F. H. J. C. A. C. A. A. A. A. T. J. T. Z. W. W. S. Christian Ledig, Lucas Theis, "Photo-realistic single image super-resolution using a generative adversarial network," *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 105–114, 2017.

[6] S. W. J. G. Y. L. C. D. C. C. L. Y. Q. X. T. Xintao Wang, Ke Yu, "Esrgan: Enhanced super-resolution generative adversarial networks," *Computer Vision – ECCV 2018 Workshops*, vol. 11133, p. 63–79, 2019.

[7] M. Wiatrak and S. V. Albrecht, "Stabilizing generative adversarial network training: A survey," *ArXiv*, vol. abs/1910.00927, 2019. [Online]. Available: https://api.semanticscholar.org/CorpusID:203626752

[8] Y. Choi and H. Park, "Improving esrgan with an additional image quality loss," *Multimed Tools Appl*, vol. 82, p. 3123–3137, 2023.