I. PROJECT TITLE

Enhancing Image Super-Resolution with Generative Adversarial Networks (GANs)

II. OBJECTIVE

To upscale low-resolution images using state-of-the-art GAN-based architectures such as SRGAN and ESRGAN. The goal is to improve perceptual quality while comparing the performance with traditional interpolation methods like bilinear and bicubic interpolation.

III. INTRODUCTION

Image super-resolution (SR) is a fundamental problem in computer vision that focuses on reconstructing high-resolution (HR) images from their low-resolution (LR) counterparts. This task has wide-ranging applications in areas such as medical imaging, satellite imaging, security systems, and image enhancement for multimedia. Traditional methods such as bilinear and bicubic interpolation have been widely used due to their computational efficiency; however, these methods often fail to recover fine details and textures, leading to blurred outputs [1] [2].

To address these limitations, deep learning-based methods have been developed, leveraging Convolutional Neural Networks (CNNs) to learn complex mappings from LR to HR images. A breakthrough in this field came with the introduction of Generative Adversarial Networks (GANs)**,** which enable the generation of photorealistic images by combining a generative network and a discriminative network in an adversarial training framework [4][5].

One of the most significant contributions in this area is the Super-Resolution GAN (SRGAN)**,** proposed by Ledig et al. [5]. SRGAN introduced the concept of perceptual loss, which uses high-level features extracted from a pre-trained VGG network to improve perceptual quality. While SRGAN represented a significant improvement over traditional methods, it had limitations in generating sharp textures and maintaining stability during training.

Scientific Foundation

SRGAN (Super-Resolution Generative Adversarial Network)**,** proposed by Ledig et al. [5]**,** introduced adversarial training for super-resolution tasks, combining a generator to create high-resolution images and a discriminator to evaluate perceptual quality. It used perceptual loss based on VGG features to improve image realism, marking a significant shift from traditional pixel-based loss functions. Building on SRGAN, ESRGAN (Enhanced SRGAN)**,** introduced by Wang et al. [6]**,** enhanced performance by replacing Batch Normalization with Residual-in-Residual Dense Blocks (RRDB) to retain more context, and employing a Relativistic GAN (RaGAN) to improve discriminator sensitivity. ESRGAN also refined the perceptual loss function to better align with human visual perception, resulting in more realistic super-resolution images.

IV. METHODS OF TRAINING

Neural Network Architecture

1. Generator (Image Upscaling Network)**:**
   - The generator in this project was inspired by ESRGAN's design. It incorporates:
     - Convolutional Layers**:** For feature extraction and upscaling.
     - Residual Blocks**:** 16 Residual Blocks for feature enrichment, including skip connections to combat vanishing gradients.
     - PreLU Activations**:** ReLU variants that adaptively learn activation slopes.
   - This architecture balances perceptual quality and computational efficiency, as demonstrated in different related works mentioned in the introduction section.

2. Discriminator (Adversarial Training Network)**:**
   - A standard deep convolutional network to classify images as "real" or "fake."
   - LeakyReLU activations with an alpha value of 0.2 improve gradient flow during training.
   - Batch Normalization layers enhance convergence stability.

3. Perceptual Loss Function (VGG Loss)**:**
   - Based on **VGG19** features, this loss measures content similarity in the feature space of a pre-trained VGG19 network.

   - Outputs from the block5_conv4 layer capture high-level details essential for Perceptual realism.

Training Details

1. Data Preprocessing
2. Optimization
3. Training Steps**:**
   - Adversarial Training**:**
     - The discriminator is trained to distinguish real HR images from generator-produced images.
     - The generator aims to produce images indistinguishable from HR images by fooling the discriminator.
   - Content Loss Training**:**
     - Perceptual loss based on VGG19 ensures generated images maintain semantic similarity to HR images.
4. Hyperparameter Optimization**:**
   - **Manual Optimization**: Initial values for learning rates and model architecture were chosen based on insights from ESRGAN and SRGAN literature and then modified for better performance through try and error mechanism.
   - **Iterative Tuning**:
     - Number of residual blocks: Tested with 8, 12, and 16 to balance quality and training time.
     - Batch size: Adjusted based on GPU memory availability (16 for this experiment).

Training Challenges and Adjustments

1. Vanishing Gradients**:**
   - Addressed by adding skip connections in the generator.

2. Stability Issues in GAN Training**:**
   - LeakyReLU and Batch Normalization layers helped stabilize the discriminator.

V. EVALUATION

The performance of ESRGAN and interpolation methods was assessed using:

- **PSNR (Peak Signal-to-Noise Ratio)**: Measures the difference between the generated image and the ground truth. Higher values indicate better fidelity.
- **SSIM (Structural Similarity Index Measure)**: Evaluates perceptual similarity by comparing structural information between images. Higher values indicate closer perceptual quality to the ground truth.

Results

The following results were obtained after running the project:

Method	Average PSNR	Average SSIM
ESRGAN	11.2664	0.1879
Bilinear Interpolation	19.4932	0.7047
Bicubic Interpolation	19.6939	0.7087

Analysis

Comparison of Methods

1. **ESRGAN**:
    - The result indicates that despite lower PSNR and SSIM scores when compared to the traditional methods, ESRGAN can produce images with visually appealing textures and details.
2. **Bilinear Interpolation**:
    - A simple technique that calculates pixel values as a weighted average of surrounding pixels.
    - Achieves good PSNR and SSIM scores.
3. **Bicubic Interpolation**:
    - More advanced than bilinear, using cubic polynomials to interpolate pixel values.

- Outperforms bilinear interpolation slightly in both PSNR and SSIM.

VI. Conclusions

This project demonstrated the potential of ESRGAN for image super-resolution, showing its ability to generate visually appealing results. However, it has low performance in PSNR and SSIM when compared with traditional interpolation methods like bicubic and bilinear.

Future Directions

- Experiment with **fine-tuning ESRGAN** for improved PSNR and SSIM scores.
- Explore **new datasets** to generalize the model's performance.

References

[1]. H. Greenspan, “Super-resolution in medical imaging,” The Com- puter Journal, vol. 52, 2008.

[2]. J. S. Isaac and R. Kulkarni, “Super resolution techniques for medical image processing,” in ICTSD, 2015.

[3]. M. Jahnavi, D. R. Rao, and A. Sujatha, "A Comparative Study Of Super-Resolution Interpolation Techniques: Insights For Selecting The Most Appropriate Method," *Procedia Comput. Sci.*, vol. 233, pp. 504–517, 2024, doi: [10.1016/j.procs.2024.03.240]. Available: https://www.sciencedirect.com/science/article/pii/S1877050924005994.

[4]. N. Sun, H. Li “Super Resolution Reconstruction of Images Based on Interpolation and Full Convolutional Neural Network and Application in Medical Fields” IEEE Access, 7 (2019), pp. 186470-186479, 10.1109/ACCESS.2019.2960828.

[5]. C. Ledig, L. Theis, F. Huszár, J. Caballero, A. Cunningham, A. Acosta, A. Aitken, A. Tejani, J. Totz, Z. Wang, and W. Shi, "Photo-Realistic Single Image Super-Resolution Using a Generative Adversarial Network," arXiv, 2017, [Online]. Available: https://arxiv.org/abs/1609.04802. [Accessed: Dec. 8, 2024].

[6]. X. Wang, K. Yu, C. Dong, and C. C. Loy, "ESRGAN: Enhanced Super-Resolution Generative Adversarial Networks," Proc. Eur. Conf. Comput. Vis. (ECCV), 2018. [Online]. Available: https://arxiv.org/pdf/1809.00219.