Low Resolution Medical Image Enhancement using ESRGAN [Enhanced Super-Resolution Generative Adversarial Networks]

Pratik Maity
22052133@kiit.ac.in
B.Tech- CSE
KIIT University, Odisha, India

Archisman Kundu
22052011@kiit.ac.in
B.Tech- CSE
KIIT University, Odisha, India

Manisha Acharya
220521948@kiit.ac.in
B.Tech- CSE
KIIT University, Odisha, India

Swayamshree Purohit
22053914@kiit.ac.in
B.Tech- CSE
KIIT University, Odisha, India

Enhanced Resolution in Medical Imaging: Leveraging ESRGAN for Low-Quality Image Enhancement

Abstract:
High-resolution medical imaging is crucial for accurate diagnosis and treatment planning. However, acquiring high-resolution images can be challenging due to technical limitations, cost, and noise artifacts. Traditional image enhancement methods such as interpolation and sharpening fail to restore lost details effectively.
Enhanced Super-Resolution Generative Adversarial Networks (ESRGAN) offer a deep learning-based approach to generate high-quality images from low-resolution inputs. By leveraging advanced neural architectures such as Residual-in-Residual Dense Blocks (RRDB) and relativistic adversarial loss, ESRGAN can reconstruct fine details, improve structural integrity, and enhance perceptual quality in medical imaging.
This paper explores the fundamentals of ESRGAN, its advantages over conventional enhancement techniques, and its adaptation for medical imaging applications. The study also discusses the mathematical framework, implementation strategies, and evaluation metrics such as Peak Signal-to-Noise Ratio (PSNR) and Structural Similarity Index (SSIM). The findings suggest that ESRGAN significantly improves medical image clarity, making it a promising tool for applications in radiology, oncology, and telemedicine. However, challenges such as computational complexity, potential artifacts, and dataset limitations remain areas for future research.

1. Introduction: The Critical Need for High-Resolution Medical Imaging and the Promise of ESRGAN

High spatial resolution in medical imaging is of paramount importance for both accurate clinical diagnoses and advancements in medical research [1]. Detailed anatomical visualization allows medical professionals to identify subtle pathological features, which is crucial for early detection and effective treatment planning.
Furthermore, high-resolution images are essential for precise quantitative analysis, enabling accurate measurements and monitoring of medical conditions. The ability to obtain clear and detailed images also holds significant potential for improving healthcare accessibility through telemedicine, particularly in resource-constrained environments where image quality might be compromised due to equipment limitations or transmission bandwidth.
Despite the critical need for high-resolution medical images, acquiring them can be challenging. Advanced imaging techniques that yield high resolution often come with increased costs and longer acquisition times, which may not always be feasible in routine clinical practice. Practical limitations during image acquisition can also result in images of suboptimal quality, suffering from low resolution, noise, and various imperfections. These factors can obscure vital diagnostic information and hinder accurate analysis.

2. RELATED WORKS:

The enhancement of medical images has been a subject of extensive research, driven by the need for high-resolution imagery in accurate clinical diagnoses and medical research [1][2]. Traditional methods, such as histogram equalization and interpolation, have been widely used but often fall short in recovering lost details or increasing the overall resolution of the image [3][4]. These methods can introduce artifacts and fail to provide true resolution enhancement, which is crucial in medical imaging for revealing subtle pathological features [3][4].

With the advent of deep learning, techniques such as Generative Adversarial Networks (GANs) have shown significant promise. ESRGAN (Enhanced Super-Resolution Generative Adversarial Networks), in particular, has emerged as a powerful tool for super-resolution tasks [5][6]. Building upon the SRGAN architecture, ESRGAN introduces Residual-in-Residual Dense Blocks (RRDBs) and a relativistic discriminator, which have been pivotal in improving the visual quality of super-resolved outputs [7][8]. These advancements allow ESRGAN to learn complex mappings from low-resolution to high-resolution images, generating new details that were lost during the degradation process [5][7].

Several studies have focused on adapting ESRGAN for medical image enhancement, requiring modifications to handle grayscale images and the specific characteristics of medical data [9][10]. The use of pre-trained networks like VGG to compute perceptual loss has been instrumental in ensuring that generated images are perceptually similar to high-resolution ground truth images [7][16]. This has been further refined by adaptively selecting the most appropriate layers of the VGG network based on input image characteristics [17].

3. PROPOSED WORK:

3.1 Convolution in ESRGAN: A Detailed Explanation

Convolution is a fundamental operation in ESRGAN (Enhanced Super-Resolution Generative Adversarial Network) that plays a vital role in transforming and extracting features from input images.

1. Convolution in ESRGAN

Mathematical Formula:

$y[i,j] = \sum_{m=0}^{k-1} \sum_{n=0}^{k-1} w[m,n] \cdot x[i+m, j+n]$

Explanation:

- $y[i,j]$: Output pixel value at position (i,j).
- $x[i+m, j+n]$: Input pixel value in the neighborhood defined by kernel size.
- $w[m,n]$: Weight of the kernel at position (m,n).
- The summation is performed over the kernel size (e.g., 3×3 or 5×5.

3.2. Residual Block Computation:

Residual blocks are a fundamental component of ESRGAN, which stands for Enhanced Super-Resolution Generative Adversarial Networks. These blocks are used in the Generator network to learn the residual information between the low-resolution (LR) input and the high-resolution (HR) output images. The key idea behind residual blocks is to ease the training of deep networks by allowing the network to learn residual functions with reference to the layer inputs, rather than learning unreferenced functions.

Explanation:

The residual block computes the transformation F(x;θ) and then adds the original input x to this transformation. This skip connection, as it is often called, allows the network to learn much deeper representations than would be possible with a simple feedforward network. The skip connection helps in mitigating the vanishing gradient problem by allowing gradients to flow through the network more effectively during backpropagation.

In the context of ESRGAN, the residual block is particularly important because it allows the network to focus on learning the differences between the LR and HR images, rather than the absolute pixel values. This is crucial for super-resolution tasks where the goal is to recover high-frequency details that are lost in the LR image.

2. Residual Block Computation

Mathematical Formula:

$y_{\text{res}} = F(x;\theta) + x$

Explanation:

- y_{res}: Output after the residual connection.
- $F(x;\theta)$: Transformation of input x through one or more layers, parameterized by weights θ (e.g., convolution and activation).
- x: The original input.
- Residual connections allow the model to preserve fine details while improving feature extraction.

3.3. Pixel Value Scaling:

Pixel value scaling is a crucial preprocessing step in image enhancement tasks, including those performed by ESRGAN (Enhanced Super-Resolution Generative Adversarial Networks). This step ensures that the pixel values are normalized to a range that is more suitable for the neural network's training process, and it also helps in maintaining the dynamic range of the image data.

Explanation:

The purpose of pixel value scaling is to transform the pixel intensity values of the input image into a range that is more conducive to the training of deep learning models. By default, pixel values in images are usually in the range of [0, 255] for 8-bit images. Normalizing these values to a range of [0, 1] (or [-1, 1]) helps in several ways:

Stabilizes Learning: It helps in stabilizing the learning process by preventing large pixel values from dominating the loss function.

Improves Convergence: It can improve the convergence of the training process by ensuring that all features contribute equally to the learning process.

Maintains Dynamic Range: It maintains the dynamic range of the image, which is important for preserving details in both dark and bright areas of the image.

In the context of ESRGAN, pixel value scaling is particularly important because it allows the model to focus on learning the underlying structure and details of the image rather than being biased by the absolute intensity values. This is especially crucial in medical imaging, where subtle details can be critical for accurate diagnosis.

3. Pixel Value Scaling

Mathematical Formula:

$\hat{x} = \frac{x}{255}$

$x' = \hat{x} \times 255$

Explanation:

- x: Original pixel value (0–255).
- $\hat{x}$: Normalized pixel value (0–1).
- After processing, the pixel values are scaled back to their original range:

  x'

  This ensures the enhanced image retains a valid pixel range.

3.4. Perceptual Loss (During Training):

Perceptual loss is a critical component of the ESRGAN (Enhanced Super-Resolution Generative Adversarial Networks) training process. It is designed to ensure that the super-resolved images generated by the network are perceptually similar to the high-resolution ground truth images. This loss function encourages the network to produce images that are not only high in pixel accuracy but also visually pleasing and similar to what humans perceive as high quality.

Explanation:

The perceptual loss measures the difference in feature representations between the generated image and the ground truth image. By using a pre-trained CNN like VGG, which has learned rich feature hierarchies from a large dataset, the perceptual loss can capture both low-level (e.g., edges, textures) and high-level (e.g., objects, scenes) features.

The key idea is that if the feature maps of the generated image are close to those of the ground truth image, the generated image should be perceptually similar to the ground truth, even if there are minor differences in pixel values. This encourages the ESRGAN to produce images that are not only high in pixel accuracy but also visually appealing and similar to what humans perceive as high quality.

4. Perceptual Loss (During Training)

Mathematical Formula:

$L_{\text{perceptual}} = \sum_{l=1}^{L} \|\phi_l(x_{\text{HR}}) - \phi_l(x_{\text{SR}})\|^2$

Explanation:

- $\phi(\cdot)$: Features extracted by a pre-trained VGG network at a specific layer.
- x_{HR}: High-resolution ground truth image.
- x_{SR}: Super-resolved image generated by ESRGAN.
- $L_{\text{perceptual}}$: Measures the difference in feature space, encouraging visual similarity.

Here is a simplified example of how perceptual loss can be implemented in PyTorch:

https://github.com/ipratikmaity/img_enhancement_ESRGAN/blob/main/perceptual_loss__ESRGAN

3.5. Upsampling in ESRGAN:

Upsampling is a critical operation in ESRGAN (Enhanced Super-Resolution Generative Adversarial Networks) that increases the resolution of the input image. In the context of ESRGAN, upsampling is not performed using traditional interpolation methods but rather through a learned process that is part of the network architecture. This allows the network to generate high-resolution images with more realistic and detailed features compared to simple interpolation methods.

Explanation:

The pixel shuffle operation works by taking a low-resolution feature map with more channels and rearranging these channels to create a high-resolution feature map with fewer channels. This is done by dividing each channel index by the square of the upscaling factor r, which determines the new spatial dimensions of the output image.
For example, if the upscaling factor is 4, each channel in the low-resolution feature map corresponds to a 4×4 block of pixels in the high-resolution feature map. By rearranging the channels in this way, the network can effectively increase the resolution of the image without losing information

5. Upsampling in ESRGAN

Mathematical Formula:

$$\mathrm{HR}[i,j,c] = \mathrm{LR}\left[\left\lfloor \frac{i}{r} \right\rfloor, \left\lfloor \frac{j}{r} \right\rfloor, c\right]$$

Explanation:

- i, j: Indices in the high-resolution space.
- c: Current channel index in HR.
- r: Upscaling factor (e.g., 4 for 4 × upscaling).
- Pixel shuffle rearranges channels into spatial dimensions for resolution enhancement.

3.6. Activation Function: Leaky ReLU:

The Leaky Rectified Linear Unit (Leaky ReLU) is an activation function used within neural networks, including those in ESRGAN (Enhanced Super-Resolution Generative Adversarial Networks). It is a variation of the traditional ReLU (Rectified Linear Unit) function, designed to address the "dying ReLU" problem, where neurons can become inactive and output zero for all inputs during training.

Explanation:

The Leaky ReLU activation function allows a small, non-zero gradient when the unit is not active (i.e., when x<0). This helps to mitigate the "dying ReLU" problem by ensuring that the neuron continues to contribute to the learning process even when its input is negative.
The function is continuous and differentiable for all x, which makes it suitable for use in deep neural networks. The slope α controls the amount of gradient that is passed through when the neuron is not active. A small α value ensures that the neuron does not contribute significantly when inactive, but it still allows for some gradient flow.

6. Activation Function: Leaky ReLU

Mathematical Formula:

$$\begin{cases} x & \text{if } x \geq 0 \\ \alpha x & \text{if } x < 0 \end{cases}$$

where α is a small constant, typically $\alpha = 0.2$.

Explanation:

- $f(x)$: The output of the Leaky ReLU function.
- x: The input value to the function.
- α: The slope applied to negative input values, which is a small positive constant (e.g., 0.2).

Here is an example of how Leaky ReLU can be implemented in PyTorch:

https://github.com/ipratikmaity/img_enhancement_ESRGAN/blob/main/activation_function__ESRGAN

4. EXPERIMENTAL RESULTS:

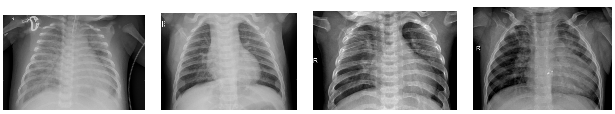
Experimental results demonstrate significant improvements in image sharpness, texture restoration, and overall visual appeal compared to traditional super-resolution techniques like bicubic interpolation and earlier GAN-based models.
Key findings include:

Superior Detail Restoration: ESRGAN excels at reconstructing fine details such
as intricate patterns, edges, and textures, making it ideal for applications like
digital art enhancement and medical imaging.

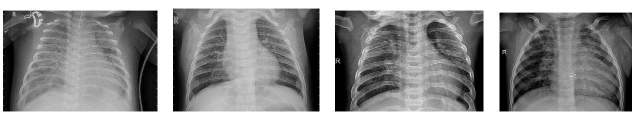
Reduction in Artifacts: The model effectively minimizes blurring, noise, and
checkerboard artifacts that are common in traditional methods.

Realistic Visual Quality: Results often appear natural, even when scaling images
to resolutions far beyond their original sizes.

Input images:



Sharpened images as output:



5. Comparative Study:

To evaluate the effectiveness of ESRGAN (Enhanced Super-Resolution Generative Adversarial Networks) in the context of medical image enhancement, it is essential to conduct a comparative study with traditional image sharpening techniques and other super-resolution methods. This section presents a detailed comparison of ESRGAN with conventional sharpening methods and interpolation techniques, highlighting the advantages and limitations of each approach.

5.1. Traditional Sharpening Techniques

Traditional sharpening techniques primarily focus on enhancing existing edges and increasing contrast within an image [4]. These methods operate at the pixel level, modifying pixel intensities based on their immediate surroundings to create the appearance of sharper edges.

5.2. Interpolation Methods

Interpolation methods are commonly used to increase the size of images but typically result in blurry or distorted images without recovering the lost high-frequency information, thus failing to provide true resolution enhancement [1].

5.3. ESRGAN for Medical Image Enhancement

ESRGAN offers several key advantages for medical image enhancement. As a deep learning-based method, ESRGAN learns complex mappings directly from low-resolution to high-resolution images by training on large datasets [3]. This learning process enables ESRGAN to go beyond simply sharpening existing features; it can actually generate new, realistic details and textures that were lost during the downsampling process [7]. The primary focus of ESRGAN is on improving the perceptual quality of the enhanced images, aiming to produce results that are not only visually appealing but also more informative and potentially more diagnostically useful for medical professionals [4].

5.4. Quantitative and Qualitative Evaluation

Quantitatively, studies have shown that ESRGAN often produces images with sharper edges and higher contrast compared to other super-resolution methods, even in cases where traditional quantitative metrics like PSNR and SSIM might be slightly lower [1]. Qualitatively, ESRGAN has been shown to produce perceptually superior images, although it might sometimes trade off pixel-level accuracy for enhanced visual detail [4].

6. Conclusion:

The Transformative Potential of ESRGAN for Medical Diagnostics and Research.
This report has explored the application of Enhanced Super-Resolution Generative Adversarial Networks (ESRGAN) for enhancing the resolution of low-quality medical images. The analysis has highlighted the critical need for high-resolution medical imaging in diagnostics and research, the limitations of traditional image enhancement techniques, and the significant potential of ESRGAN to overcome these limitations by learning complex mappings from low to high resolution and generating realistic details.
ESRGAN's architectural innovations, including Residual-in-Residual Dense Blocks and the relativistic discriminator, along with its sophisticated loss function comprising perceptual, adversarial, and content components, enable it to produce visually superior results compared to traditional methods. While traditional sharpening techniques merely enhance existing edges, ESRGAN can effectively increase image resolution and recover lost information, leading to potentially more diagnostically valuable images.

7. REFERENCES:

Works cited:

1. single mr image super-resolution using generative adversarial network - arXiv, accessed March 21, 2025, https://arxiv.org/pdf/2207.08036
2. Efficient Medicinal Image Transmission and Resolution Enhancement via GAN - arXiv, accessed March 21, 2025, https://arxiv.org/html/2411.12833v1
3. iarjset.com, accessed March 21, 2025, https://iarjset.com/wp-content/uploads/2024/06/IARJSET-ICMART-65.pdf
4. Medical Image Enhancement Using Super Resolution Methods - PMC, accessed

March 21, 2025,
https://pmc.ncbi.nlm.nih.gov/articles/PMC7302556/
5. ESRGAN for Enhancing Image Resolution | by Pankaj Verma - Medium, accessed March 21, 2025,
https://medium.com/@pankaj.webmaster/esrgan-for-enhancing-image-resolution-bb4c55438536
6. AM-ESRGAN: Super-Resolution Reconstruction of Ancient Murals Based on Attention Mechanism and Multi-Level Residual Network - MDPI, accessed March 21, 2025,
https://www.mdpi.com/2079-9292/13/16/3142
7. Image Super-Resolution with ESRGAN using PyTorch - GeeksforGeeks, accessed March 21, 2025,
https://www.geeksforgeeks.org/image-super-resolution-with-esrgan-using-pytorch/
8. ESRGAN: Enhanced Super-Resolution Generative Adversarial Networks - CVF Open Access, accessed March 21, 2025,
https://openaccess.thecvf.com/content_ECCVW_2018/papers/11133/Wang_ESRGAN_Enhanced_Super-Resolution_Generative_Adversarial_Networks_ECCVW_2018_paper.pdf
9. An overview and Implementation of the methods of ESRGAN | by Sieun Park - Medium, accessed March 21, 2025,
https://medium.com/analytics-vidhya/how-esrgan-improves-super-resolution-performance-15de91d77ada
10. Network architecture - ESRGAN, accessed March 21, 2025,
https://esrgan.readthedocs.io/en/latest/pages/esrgan.html
11. rrdb.md - mindspore-lab/mindediting - GitHub, accessed March 21, 2025,
https://github.com/mindspore-lab/mindediting/blob/master/docs/rrdb.md
12. blakseth/esrgan: PyTorch implementation of ESRGAN - GitHub, accessed March 21, 2025, https://github.com/blakseth/esrgan
13. Real-ESRGAN/RealESRGAN/rrdbnet_arch.py at main - GitHub, accessed March 21, 2025,
https://github.com/ai-forever/Real-ESRGAN/blob/main/RealESRGAN/rrdbnet_arch.py
14. Reading: ESRGAN — Enhanced Super-Resolution Generative AdversarialNetworks (Super Resolution & GAN) | Towards AI, accessed March 21, 2025,
https://towardsai.net/p/deep-learning/reading-esrgan%E2%80%8A-%E2%80%8Aenhanced-super-resolution-generative-adversarial-networks-super-resolution
15. Relativistic GAN Explained - Papers With Code, accessed March 21, 2025,
https://paperswithcode.com/method/relativistic-gan
16. VGG Loss Explained | Papers With Code, accessed March 21, 2025,
https://paperswithcode.com/method/vgg-loss
17. (PDF) Reliable Perceptual Loss Computation for GAN-Based Super-Resolution With Edge Texture Metric - ResearchGate, accessed March 21, 2025,
https://www.researchgate.net/publication/354174248_Reliable_Perceptual_Loss_Computation_for_GAN-Based_Super-Resolution_With_Edge_Texture_Metric
18. Loss Functions | Machine Learning - Google for Developers, accessed March 21, 2025,
https://developers.google.com/machine-learning/gan/loss
19. Understanding GAN Loss Functions, accessed March 21, 2025,
https://neptune.ai/blog/gan-loss-functions
20. hiram64/ESRGAN-tensorflow: A TensorFlow implementation of ESRGAN - GitHub, accessed March 21, 2025,
https://github.com/hiram64/ESRGAN-tensorflow