Enhancing X-Ray Images Using ESRGAN: A Deep Learning Approach

Ahmed Mohamed Ali[#1], Osama Mohamed Ali[#2], Ali Sherif Badran[#3] , Muhannad Abdallah[#4]

Systems and Biomedical Engineering Department, Cairo University

[`*] *AI in medical field*

***Abstract*— The analysis of medical imaging is a cornerstone of modern healthcare, enabling early diagnosis, treatment planning, and monitoring of various conditions. However, many older imaging devices, still in use due to resource constraints in low-resource settings or outdated infrastructure, produce low-resolution images that can limit diagnostic accuracy. This issue is particularly significant in chest X-ray imaging, where subtle details such as small nodules, interstitial patterns, or early signs of disease may be missed due to insufficient resolution. The Enhanced Super-Resolution Generative Adversarial Network (ESRGAN) presents a promising solution by offering state-of-the-art super-resolution capabilities that can enhance image resolution by up to four times. This paper investigates the application of ESRGAN to improve the resolution of low-quality chest X-ray images, aiming to bridge the gap between old imaging technology and current diagnostic requirements. By augmenting image quality, this approach has the potential to improve diagnostic precision, reduce misdiagnosis, and optimize healthcare outcomes in resource-limited settings.**

I. Introduction

Medical imaging is a cornerstone of modern diagnostics, enabling the detection, monitoring, and treatment of numerous diseases. Among various imaging modalities, X-ray imaging remains one of the most widely used techniques due to its affordability, accessibility, and ability to quickly deliver diagnostic insights. Chest X-rays, in particular, are invaluable for diagnosing pulmonary diseases, cardiovascular conditions, and thoracic abnormalities. However, the diagnostic utility of these images often hinges on their resolution, as finer anatomical details are critical for accurate interpretation.

In many low-resource healthcare settings, older X-ray devices are still in use, frequently producing low-resolution images that hinder the detection of subtle pathologies such as early-stage lung cancer, interstitial lung disease, or microvascular abnormalities. Enhancing the resolution of these images is crucial to improving diagnostic accuracy, especially in contexts where upgrading equipment is not feasible.

Recent advancements in artificial intelligence (AI), particularly deep learning, have opened new pathways for enhancing medical image quality. AI-driven image processing techniques have shown great promise in improving clarity, contrast, and resolution. Among these, Generative Adversarial Networks (GANs) have emerged as a powerful tool for image synthesis and enhancement. GANs employ a competitive framework between two neural networks—a generator and a discriminator—to produce highly realistic and refined outputs. Super-resolution GANs (SRGANs) and their enhanced variants, such as the Enhanced Super-Resolution GAN (ESRGAN), have proven effective in improving image resolution by capturing high-frequency details and textures. In medical imaging, these techniques offer the potential to transform low-quality scans into diagnostically valuable high-resolution images. Leveraging ESRGAN, the resolution of chest X-rays from older devices can be enhanced significantly, uncovering subtle features crucial for accurate diagnosis.

This paper investigates the application of ESRGAN for enhancing chest X-ray resolution, addressing the challenges of low-resolution imaging, the principles of GAN-based super-resolution, and the implications of this approach for medical diagnostics. By bridging the gap between legacy imaging systems and modern diagnostic requirements, this method offers a scalable and cost-effective solution to enhance healthcare outcomes.

II. Related work:

Enhancing image resolution has long been a challenge in computer vision, with significant progress driven by deep learning. Convolutional neural networks (CNNs) and generative adversarial networks (GANs) have been pivotal, introducing novel approaches for single image super-resolution (SISR).

The introduction of Super-Resolution GAN (SRGAN) by Ledig et al. [1] marked a transformative step, leveraging a generative adversarial framework to produce photo-realistic high-resolution images. Traditional optimization-based methods, though achieving high peak signal-to-noise ratios (PSNR), often failed to capture fine texture details, especially at high upscaling factors. SRGAN addressed this limitation through a perceptual loss function combining adversarial and content losses, enabling high perceptual fidelity. Evaluations, including mean-opinion-score (MOS) tests, demonstrated SRGAN's superior perceptual quality compared to prior methods.

Building on this, Tian et al. [2] proposed the Enhanced Super-Resolution Group CNN (ESRGCNN), which tackled the computational complexity of deep CNNs. Unlike SRGAN's reliance on deep residual networks, ESRGCNN employed a shallow architecture that fused deep and wide channel features, effectively capturing low-frequency information. This balance of performance and efficiency made it ideal for resource-constrained applications. Additionally, its adaptive up-sampling operation allowed ESRGCNN to handle images of varying resolutions, excelling in speed, quality, and visual effect.

Hetherington et al. [3] further expanded the field by integrating modal decomposition algorithms with deep learning to address super-resolution and repair noisy datasets. Their approach utilized singular value decomposition (SVD) and high-order SVD with neural networks, achieving robust reconstruction for complex and incomplete data. This work highlighted data-driven methods for uncovering the underlying physics of datasets, enhancing precision and reducing noise.

These advancements demonstrate diverse strategies in image super-resolution. SRGAN prioritized perceptual quality via adversarial learning, ESRGCNN focused on computational efficiency, and modal decomposition combined with deep learning showcased innovative data reconstruction techniques. Building on this foundation, this paper explores the use of Enhanced Super-Resolution GAN (ESRGAN) to address challenges in low-resolution chest X-ray imaging, leveraging GAN-based perceptual optimization and efficient architectures.

III. Dataset

We used the NIH Chest X-ray dataset, a large-scale collection of 112,120 frontal-view chest X-ray images from 30,805 patients, developed by the NIH Clinical Center to advance computer-aided detection and diagnosis of thoracic diseases. The dataset includes 14 pathology labels, such as Atelectasis, Pneumothorax, and Cardiomegaly, mined with an NLP pipeline at ~90% accuracy, extending previous datasets with six additional disease categories.

Images are provided at a 1024 × 1024 resolution, along with metadata (e.g., patient demographics, view position, pixel spacing) and bounding box annotations for ~1,000 images. Data is split at the patient level into training, validation, and testing subsets to prevent data leakage. Its scale (~60% of NIH Clinical Center's PACS database) and diversity make it highly representative of clinical settings, enabling robust evaluation of deep learning models.

While limitations include potential labeling inaccuracies and limited bounding box annotations, the dataset's size and comprehensive scope support the development of multi-label models for medical imaging. For this study, we used 10,000 images, standardizing intensity values and normalizing pixel spacing. Enhanced resolution images generated by the ESRGANs model were compared to the original dataset to evaluate super-resolution performance.

IV. Methods

We employed the Enhanced Super-Resolution Generative Adversarial Network (ESRGAN) for the super-resolution task, adapting its architecture to upscale medical images with high fidelity. The generator network was designed to reconstruct high-resolution images from low-resolution inputs by leveraging a series of advanced feature extraction modules. The network incorporated Residual Dense Blocks (RDBs) and Residual-in-Residual Dense Blocks (RRDBs) to enable efficient feature learning and hierarchical feature reuse. The generator begins with an initial convolutional layer to extract low-level features, followed by a series of eight RDBs. These blocks operate through local feature fusion and concatenation mechanisms, ensuring effective learning of fine details. The features are subsequently globally fused, upsampled through pixel-shuffling operations, and passed through additional non-linear activations to produce the final high-resolution images.

The discriminator network, on the other hand, was implemented as a progressively deeper convolutional network designed to distinguish between real high-resolution images and those generated by the model. It consists of alternating convolutional and downsampling layers, complemented by batch normalization and LeakyReLU activations, culminating in a dense output layer that produces a scalar probability score. This adversarial feedback encourages the generator to produce images that closely mimic the ground truth. Additionally, we incorporated a pre-trained VGG19 network as a feature extractor to compute the perceptual loss. The features were extracted from the block5_conv4 layer of the network, allowing the model to measure high-level perceptual similarity between the generated and ground truth images. To maintain the integrity of the extracted features, the VGG19 network was frozen during training.

The dataset used in this study consisted of a curated subset of 10,000 chest X-ray images, selected from the larger publicly dataset. Each image was preprocessed to produce high-resolution (HR) and low-resolution (LR) counterparts. High-resolution images were resized to 512 \times 512 pixels using bicubic interpolation, while low-resolution images were downsampled to 128 \times 128 pixels(with other variations discussed in the experiments section). To enhance model robustness, random Gaussian noise with a standard deviation of 0.01 was added to the low-resolution images. All images were normalized to the range [-1, 1] to ensure compatibility with the ESRGAN architecture. The dataset was divided into training and testing subsets, following an 80/20 split, with no overlap of images from the same patient to avoid data leakage. The data was processed into mini-batches of size four and preloaded using shuffling and prefetching techniques to optimize training efficiency.

The training process optimized the ESRGAN model using a combination of loss functions. Content loss was computed as the mean absolute error (MAE) between the pixel intensities of the high-resolution ground truth and the super-resolved images, ensuring accurate reconstruction. Perceptual loss was calculated as the MAE between feature representations of the generated and ground truth images in the VGG19 feature space, promoting the preservation of perceptual details. Adversarial loss was derived from the discriminator's predictions, guiding the generator to produce realistic outputs. The total generator loss was a weighted sum of these components, with equal weights assigned to content and perceptual loss, and a lower weight assigned to adversarial loss as follows:

$$L_{\text{total}} = 1.0 \cdot L_{\text{content}} + 1.0 \cdot L_{\text{perceptual}} + 0.1 \cdot L_{\text{adversarial}}$$

The discriminator loss was computed using binary cross-entropy to distinguish between real and generated images.

The ESRGAN model was trained for 10 epochs using the Adam optimizer with a learning rate of 1x10^{-4} for both the generator and discriminator. Each training step involved generating super-resolved images from low-resolution inputs,

evaluating them using the discriminator, and updating the model parameters based on the computed gradients. Loss metrics, including content, perceptual, generator, and discriminator losses, were recorded at each epoch to monitor the training process. Periodic checkpoints were saved to ensure reproducibility and enable subsequent fine-tuning.

To evaluate the performance of the model, we employed Peak Signal-to-Noise Ratio (PSNR) and Structural Similarity Index Measure (SSIM) metrics, which quantify image reconstruction quality and structural similarity, respectively. Qualitative results were visualized by displaying side-by-side comparisons of low-resolution, high-resolution, and super-resolved images. Training dynamics were analyzed through visualizations of loss functions over epochs, while PSNR and SSIM distributions provided additional insights into model performance. The implementation was carried out using TensorFlow 2.15.1 on a GPU-enabled system, with GPU memory growth enabled to handle large data batches efficiently. The training pipeline incorporated robust error-handling mechanisms to manage resource constraints and ensure uninterrupted execution.

This comprehensive methodology allowed us to successfully train and evaluate the ESRGAN model for the super-resolution task, achieving high-quality reconstructions of medical images while maintaining computational efficiency.

V. Experiments

5.1 Model Architecture and Training Setup

The Enhanced Super-Resolution Generative Adversarial Network (ESRGAN) was employed to upscale low-resolution chest X-ray images, addressing the diagnostic limitations posed by older imaging equipment. The generator consists of eight Residual-in-Residual Dense Blocks (RRDB), designed to capture fine details and high-frequency information, which is essential for medical imaging. A discriminator network was trained concurrently to distinguish between the super-resolved and ground truth high-resolution images, reinforcing adversarial learning.

The model leverages a pre-trained VGG19 network to extract perceptual features, contributing to the content and perceptual loss calculations. This VGG19 component guides the generator toward producing images that are not only visually convincing but also diagnostically relevant.

5.2 Hyperparameter Selection and Resource Constraints

The choice of hyperparameters was influenced by extensive manual tuning and limited access to advanced GPU resources. Larger batch sizes or higher image resolutions led to GPU memory exhaustion, necessitating a more conservative selection of parameters. Consequently, the batch size was fixed at 4, and the low-resolution images were consistently set to one-quarter the size of their high-resolution counterparts to mitigate memory constraints.

The adversarial weight was adjusted across experiments to evaluate the trade-off between perceptual sharpness and adversarial stability. A higher adversarial weight promotes sharper images but may introduce artifacts, while lower weights emphasize content and perceptual loss.

Hyperparameter	Experiment 1	Experiment 2	Experiment 3
Learning Rate (Generator)	1e-4	1e-4	1e-4
Learning Rate (Discriminator)	1e-4	1e-4	1e-4
Content Weight	1.0	1.0	1.0
Perceptual Weight	1.0	1.0	1.0
Adversarial Weight	0.1	0.9	0.1
Batch Size	4	4	4
HR Image Size	256	512	128
LR Image Size	64	128	32
Epochs	10	15	50

VI. Results

6.1 Quantitative Evaluation

Three key metrics were used to evaluate the performance of the super-resolution models:

- **Peak Signal-to-Noise Ratio (PSNR)** – Measures the ratio between the maximum possible power of a signal and the power of corrupting noise.
- **Structural Similarity Index (SSIM)** – Evaluates the perceived quality of the images by comparing luminance, contrast, and structure.
- **Mean Squared Error (MSE)** – Measures the average squared difference between the original high-resolution image and the super-resolved image.

Equations:

$$PSNR = 10 \cdot \log_{10} \left(\frac{MAX^2}{MSE} \right)$$

Where MAX is the maximum pixel intensity (1.0 for normalized images).

$$SSIM(x,y) = \frac{(2\mu_x\mu_y + c_1)(2\sigma_{xy} + c_2)}{(\mu_x^2 + \mu_y^2 + c_1)(\sigma_x^2 + \sigma_y^2 + c_2)}$$

Where μx and μy are the mean pixel intensities, σx^2 and σy^2 are the variances, and σxy is the covariance.

$$MSE = \frac{1}{N}\sum(I_{original} - I_{sr})^2$$

Experiment Results (Completed Runs):

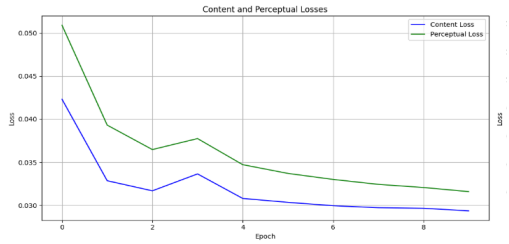
Experiment	Mean PSNR	Mean SSIM	Mean MSE
1	30.0764	0.89	0.0012
2	32.9906	0.9257	0.0006
3	27.7965	0.8675	0.0020

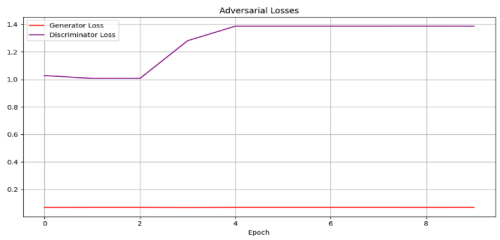
6.2 Loss Visualization

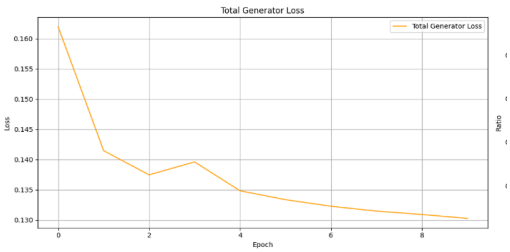
The training process was monitored through three primary loss plots per experiment:

- Content and Perceptual Loss
- Adversarial Loss (Generator and Discriminator Losses)
- Total Generator Loss

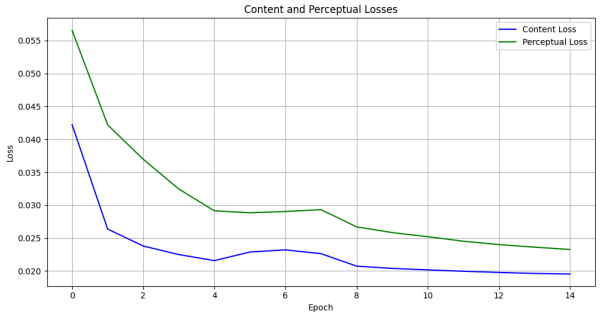
Experiment 1:

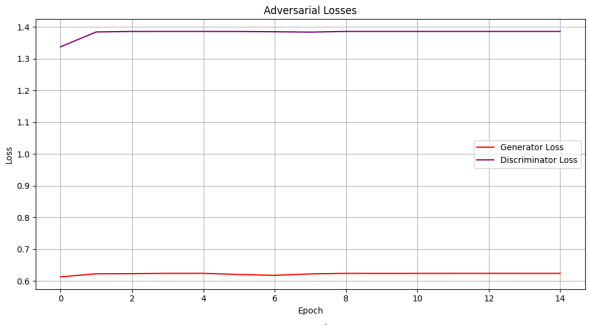


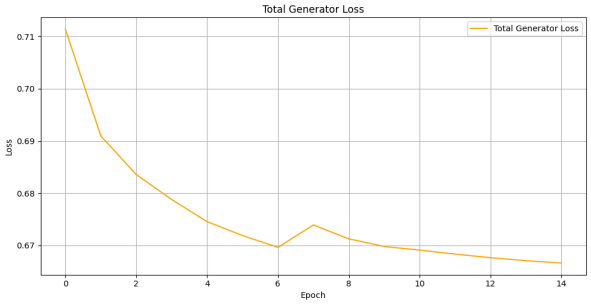




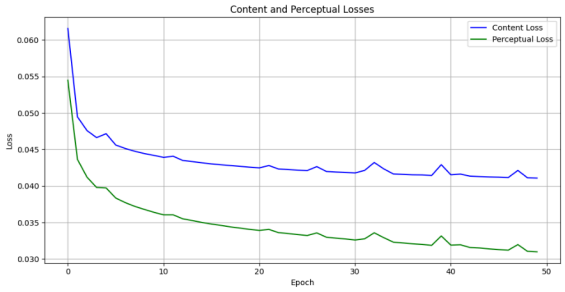
Experiment 2:

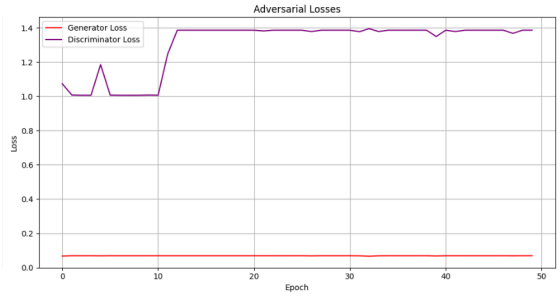


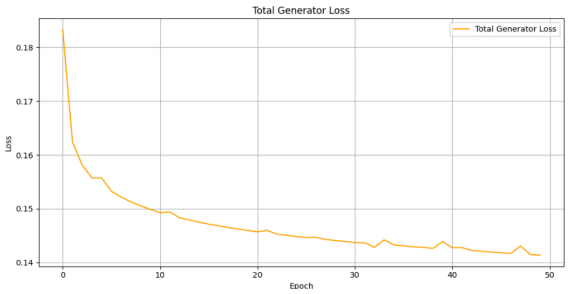




Experiment 3:







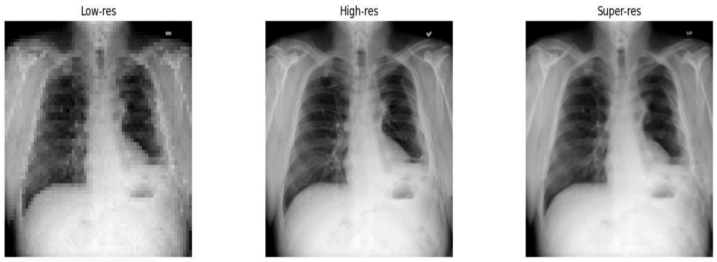
6.3 Qualitative Results

Visual comparisons between low-resolution, high-resolution, and ESRGAN-generated super-resolution images highlight the model's capacity to reconstruct fine details critical for medical diagnosis.

Sample Visualization:

- Column 1: Low-resolution images
- Column 2: Ground truth high-resolution images
- Column 3: ESRGAN super-resolved images

Experiment 1 : (64 -> 256)



Experiment 2 : (128 -> 512)

Experiment 3 : (32 -> 128)

VII. Discussion

The results indicate that ESRGAN is effective at enhancing chest X-ray resolution, with PSNR and SSIM values suggesting high fidelity to ground truth images. However, the following observations emerged:

- Experiment 2 (higher adversarial weight) produced sharper images but exhibited occasional artifacts. This highlights the trade-off between sharpening and introducing noise.
- Experiments 1 and 3 (lower adversarial weights) generated smoother images, with fewer artifacts but slightly reduced sharpness.

Challenges:

- GPU memory limitations constrained model size and batch size, preventing the exploration of higher-resolution inputs.
- Experiment 3 (ongoing) investigates whether the model can effectively enhance extremely low-resolution images (32 × 32).

VIII. Conclusion and Future Work

This study demonstrated the effectiveness of ESRGAN in improving the resolution of chest X-ray images, with promising results across various configurations.

Future Work:

- Larger models and datasets: With access to more powerful GPUs, future experiments could involve higher-resolution inputs (e.g., 1024 × 1024).
- Clinical evaluation: Collaborating with radiologists to assess the clinical relevance of enhanced images.
- Advanced architectures: Exploring transformer-based models (e.g., SwinIR) for super-resolution.

References

[1] Ledig, C., et al. *Photo-Realistic Single Image Super-Resolution Using a Generative Adversarial Network*. arXiv preprint arXiv:1609.04802, 2017. [Online]. Available: https://arxiv.org/abs/1609.04802

[2] Tian, C., Yuan, Y., Zhang, S., Lin, C.-W., Zuo, W., & Zhang, D. *Image Super-resolution with An Enhanced Group Convolutional Neural Network*. arXiv preprint arXiv:2205.14548, 2022. [Online]. Available: https://arxiv.org/abs/2205.14548

[3] Hetherington, A., Serfaty, D., Corrochano, A., Soria, J., & Le Clainche, S. *Data repairing and resolution enhancement using data-driven modal decomposition and deep learning*. arXiv preprint arXiv:2401.11286, 2024. [Online]. Available: https://arxiv.org/abs/24286

01.11