# Enhancing chest CT Image Resolution using UNet-based models

Sanho Lee Jiyeong Ha Euijae Kim
New York University
sh18607@nyu.edu jh7685@nyu.edu ek3955@nyu.edu

## A. Introduction

Acquiring high-resolution medical images is crucial for accurate disease diagnosis. Nonetheless, many medical imaging techniques, including computed tomography (CT), frequently encounter challenges from low-resolution images, due to various factors such as patient motion, efforts to reduce radiation dose, and limitations in reconstruction algorithms. Thus, it is important to address these challenges through advancement in imaging-processing techniques is essential to enhance the effectiveness of medical imaging modalities.

The technique of single image super-resolution (SISR), which involves restoring a high-resolution image from its degraded counterpart using deep-neural networks (DNN), is being considered as a promising approach to achieve high-resolution medical images. Previous studies have shown that the deep-learning based super-resolution methods can be successfully applied to medical images. For example, studies have shown that Super-Resolution Convolutional Neural Network (SRCNN, [2]) and many variants of Generative Adversarial Networks (GAN) based methods have demonstrated significant improvements in chest CT [6], brain tumor, skin cancer datasets [1]. However, these methods have limitations in that SRCNN may be too shallow to extract fine features and GAN methods often suffer from unstable training.

Motivated by this, We propose to use a modern state-of-the-art SISR method based on UNet architecture to improve chest CT dataset. A robust UNet (RUNet) architecture proposed by Hu et al (2019) takes a unique approach in that it contains 'residual' blocks in the down path to allow the model to learn more complex features. Moreover, the authors used perceptual loss functions [5] that measure the distance between the predicted image and the target image in feature space ("feature-distance").

Furthermore, we explored how different image interpolation methods impact the performance of RUNet. Specifically, we compared RUNet results obtained from resized low-resolution images using bi-linear versus nearest-neighbor interpolation methods. Our findings indicate that the RUNet architecture significantly enhances the visual quality of medical images.

## B. Methods

### B.1. Experimental Design

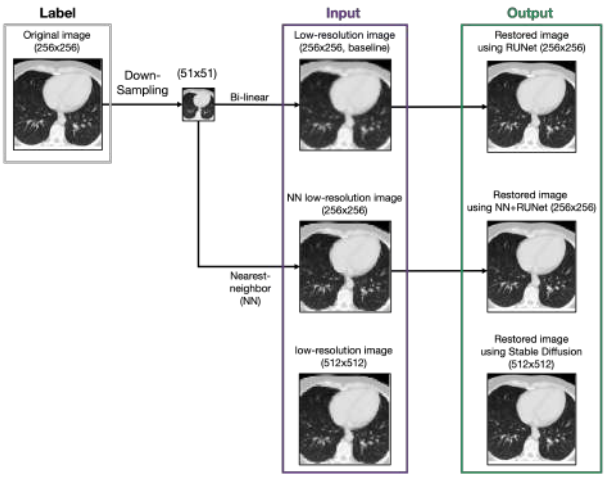


Figure 1. Overview of experimental design for the testing phase.

Figure 1 shows the experimental design with a publicly available chest CT scan dataset (see B.2.2 Dataset Preparation for more details). First, low-resolution images were generated by scaling the original images by 0.2 (image size: 51x51). Then the images were again up-sampled in order to match the size with the labels. In this process, we used two different interpolation methods, 1) bi-linear or 2) nearest-neighbor. This was to explore whether different image interpolation methods have any impact on the model's outcome. Thus, the size of the resulting high-resolution image was the same size as that of the original test image (256x256). This allowed us to assess whether the resulting high-resolution images were correctly restored or not, in comparison with the original images.

We also used another UNet-based stable diffusion model to restore the medical images to explore which model performs better at restoring medical image details.

Finally, the image restoration quality of the predicted

high-resolution images was quantified by measuring two image quality metrics with the original images: peak signal-to-noise ratio (PSNR) and structural similarity (SSIM).

The main methodologies employed in this study are Robust UNet and SR3 (Super-Resolution via Repeated Refinement). Each methodology utilizes a variety of different techniques such as nearest neighbor interpolation, image down-sampling, and diffusion process to produce an experimental dataset. The detailed implementation of the methods employed in this study can be found in our group GitHub repository (https://github.com/euijae/nyu_computer_vision_project).

## B.2. Robust UNet

### B.2.1 Model Architecture

Recent research introduced Robust UNet (RUNet) architecture which is a UNet-based super-resolution method [4]. One of the most significant features that differentiates between RUNet and UNet is a residual-blocks which allows the model to learn more complex features.

### B.2.2 Dataset Preparation

In this study, we used Chest CT-Scan images Dataset that were publicly available on Kaggle (https://www.kaggle.com/datasets/mohamedhanyyy/chest-ctscan-images).The dataset contained 586 and 315 images for training and test respectively that are categorized into one of four types: three different types of cancer (Adenocarcinoma, Large cell carcinoma, and Squamous cell carcinoma) or normal.

- **Baseline:** The dimension of images in the dataset is inconsistent. We reshaped all images to 256 by 256 by cropping at the center.
- **Adjusted Baseline:** This is an augmented dataset. We first apply the same effect as the baseline step and then rotate all images by 20 degrees.
- **Low-Resolution:** It's obtained by reshaping baseline images to 128 by 128 first and then expanding reshaped dimensions with a scale factor of 2. It ends up with a size of 256 by 256.
- **Low-Resolution with Nearest Neighbor Interpolation:** The Low-Resolution step produces an image of size 128 by 128. Then we expand low-resolution images to 256 by 256 using Nearest Neighbor interpolation.

## B.3. Stable Diffusion Model

Our approach, SR3 (Super-Resolution via Repeated Refinement), is a new method for conditional image generation. Inspired by Denoising Diffusion Probabilistic Models (DDPM) [3], SR3 transforms a standard normal distribution into an empirical data distribution through a sequence of refinement steps, resembling Langevin dynamics.

### B.3.1 Model Architecture

The SR3 model is built with a U-Net architecture, adapted to image super-resolution. The key component is a denoising objective which iteratively removes various levels of noise from an image. The model utilizes a U-Net-based architecture with self-attention mechanisms. This architecture has been modified to handle the conditional generation tasks effectively

### B.3.2 Dataset Preparation

- **Low-Resolution Images (LR):** Denoted as `lr_64`, these images form the initial input for the SR3 model.
- **High-Resolution Images (HR):** Referred to as `hr_512`, these images serve as the target for super-resolution.
- **Preprocessing Steps:** Images are resized and converted to fit the model's input and output requirements.

### B.3.3 Process Steps

**Denoising Process**: The forward Markovian diffusion process is at the core of the model, gradually corrupting the data distribution, starting from the original data distribution $q(y_0)$ to a known noise distribution over a predefined number of steps $T$. The process is defined as:

$$q(y_1 : T|y_0) = \prod_{t=1}^{T} q(y_t|y_{t-1}), \tag{1}$$

where each transition probability $q(y_t|y_{t-1})$ is modeled as a Gaussian distribution:

$$q(y_t|y_{t-1}) = \mathcal{N}(y_t; \sqrt{1-\beta_t}y_{t-1}, \beta_t\mathbf{I}), \tag{2}$$

and $\{\beta_t\}_{t=1}^{T}$ represents a variance schedule of the added noise.

**Reverse Process**: The reverse process aims to learn a parameterized model $p_\theta(y_{t-1}|y_t)$ to reverse the diffusion process. It's modeled as a denoising step where the noise is predicted and then subtracted from the noisy image. The objective function for the model parameter $\theta$ is the expected L2 norm between the real and predicted noise:

$$L(\theta) = \mathbb{E}_{t,y_0,\epsilon}[\|\epsilon - \epsilon_\theta(y_t, t)\|_2^2], \tag{3}$$

where $\epsilon \sim \mathcal{N}(0, \mathbf{I})$ and $y_t$ is the noisy image at time step $t$.

**Training and Inference**: During training, the model learns the denoising function over various noise levels. This is achieved by randomly sampling different time steps and applying the corresponding noise level. Inference involves iteratively applying the reverse process starting from a noise sample, gradually denoising it to obtain the super-resolved image.

**Super-Resolution Process**: Super-resolution is treated as a conditional generation task, where the low-resolution image is used as a condition. The process involves upscaling the low-resolution image using bicubic interpolation, which is then fed into the model along with noise to generate a high-resolution output. The model refines details iteratively to achieve super-resolution:

$$\hat{y}_0 = \sqrt{\alpha_t} y_t + \sqrt{1-\alpha_t}\epsilon_\theta(y_t, t), \tag{4}$$

where $\alpha_t = \prod_{s=1}^{t}(1-\beta_s)$.

**Cascading for High-Resolution**: For very high-resolution image generation, SR3 models can be cascaded. This involves using multiple SR3 models in sequence, each progressively enhancing the resolution. This cascading approach enables efficient, scalable super-resolution while maintaining image fidelity.

These methods collectively represent a novel approach in the field of super-resolution, effectively leveraging denoising diffusion processes to produce high-quality images from low-resolution inputs.

# C. Results

## C.1. Robust UNet Model

### C.1.1 Visual Examples

Figure 2 illustrates an example of the resulting high-resolution images of the chest CT dataset obtained using RUNet. Visual assessment confirmed that the RUNet restored some of the fine details that were missing in the low-resolution images. In particular, it seems that RUNet show a better performance for some parts of the images, especially around the edges of the chest consisting of bones, skins, and fat. Moreover, RUNet with bi-linear seems to reconstruct images more similar to the original image compared to the model with nearest-neighbor condition. Compared to the original labels, specifically, the images from the nearest-neighbor condition seem to have lower contrast and less details in the central part of the chest.

### C.1.2 Comparison of image quality

We Also quantified the model performances using two different image similarity metrics: peak signal-to-noise ratio (PSNR) and structural similarity (SSIM).

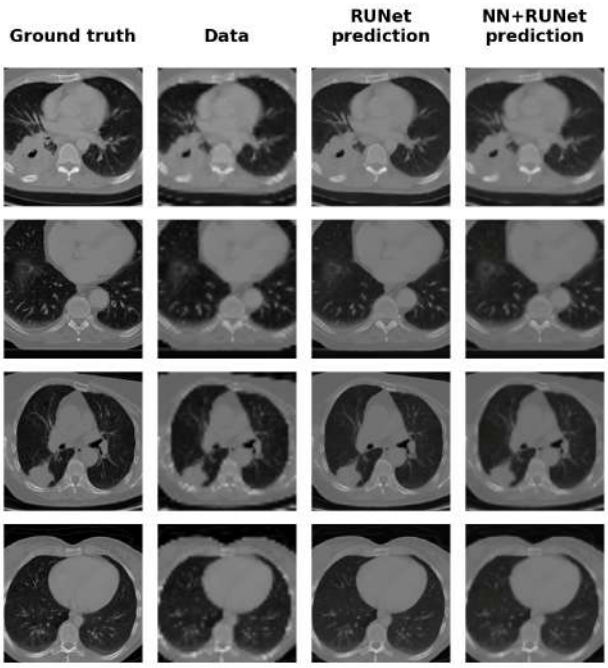


Figure 2. Example data of the original image, low-resolution image, and model outputs. Each row indicates different chest images. The model outputs are reconstructed high-resolution images using RUNet and RUNet and nearest-neighbor interpolation, respectively.

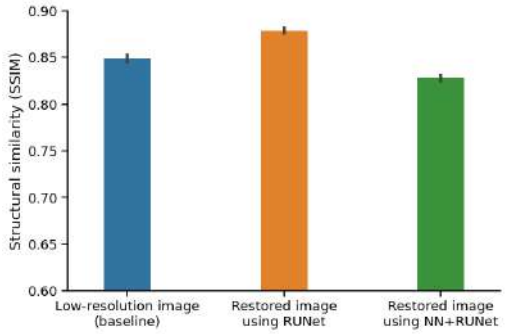


Figure 3. Comparison of the SSIM between baseline, RUNet condition, and NN+RUNet condition. The PSNR was calculated using all the images in the test dataset and averaged for each condition. Error bar indicates 68% confidential interval calculated using the bootstraps.

Figure 3 shows SSIM comparisons whereas figure 4 shows PSNR values. We first noticed that even for the baseline condition, which represents similarity between the low-resolution images and original images, are comparatively high, given that the low-resolution images were intentionally degraded using size rescaling. This may be due to the particular image characteristics of anatomical structure. For example, the center and the surround of the

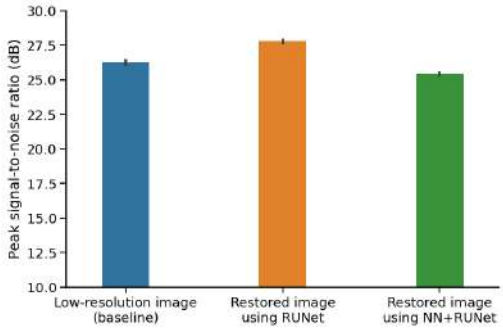


Figure 4. Comparison of the PSNR between baseline, RUNet condition, and NN+RUNet condition.

chest CT images will have the same contrast across all images, which will be likely to be remained after size rescaling. However, the RUNet model still performed superior to the baseline, showing a slight increase in all of the image metrics. This indicates that the model was able to restore some of the fine details that were lost in the low-resolution images. Both SSIM and PSNR metrics from the condition with nearest-neighbor interpolation method showed decreased values compared to the baseline condition. This may be due to the lower contrast of the images that was observed in the visual examples.

## C.2. Stable Diffusion Model

Our primary dataset comprised Chest CT-Scan images, delineating various types of chest cancer, including Adenocarcinoma, Large Cell Carcinoma, Squamous Cell Carcinoma, and normal cells. The dataset was curated with an aim to assist in chest cancer detection using machine learning and deep learning techniques. The images were sourced from multiple datasets to create a comprehensive collection conducive for training a CNN model.

**Preprocessing:** The CT-Scan images, originally in diverse formats, were transformed into a uniform jpg/png format suitable for model processing. We divided the dataset into training (80%), and validation (20%) sets.

To fit the model requirements and facilitate efficient data handling, we applied a script for resizing images into three distinct sets:

- Low Resolution (LR) - Downsampled images simulating lower quality inputs.
- High Resolution (HR) - Original image quality maintained for reference.
- Super Resolution (SR) - Upscaled from LR, aiming to reconstruct HR-like quality.

This resizing process was crucial in preparing the data for subsequent super-resolution tasks.

### C.2.1 Training Configuration

The model's training was configured as follows:

- **Dataset:** The training utilized the CT-Scan dataset for high-resolution images and the processed CT dataset for LR images.
- **Model Architecture:** A UNet architecture with multiple channel multipliers and attention mechanisms.
- **Optimizer:** Adam optimizer with a learning rate of $3 \times 10^{-6}$.
- **Iterations:** Training for 840,000 iterations with checkpoints and validation frequencies set at every 10,000 iterations.

### C.2.2 Model Evaluation

Model evaluation was performed on the preprocessed CT-scan dataset. The performance metrics included:

- **PSNR:** Peak Signal-to-Noise Ratio, evaluating the reconstruction quality.
- **SSIM:** Structural Similarity Index, assessing perceived image quality.

The final average PSNR and SSIM values for the SR3 model were recorded as 15.234 and 0.17228, respectively.

### C.2.3 Comparative Analysis

A comparative analysis of the SR3 model and Runet was conducted based on the metrics PSNR and SSIM. The results are presented in the following table:

| Model | Average PSNR | Average SSIM |
|---|---|---|
| SR3 Model | 15.23 | 0.17 |
| Runet | 27.79 | 0.87 |

Table 1. Comparative performance of SR3 Model and Runet

### C.2.4 Result Visualization

Figure 5 presents a visual comparison of HR, inferred LR, and generated SR images for selected samples, showcasing the model's enhancement capabilities.

## D. Discussion

We applied two state-of-the-art UNet-based models to reconstruct high-resolution chest CT scan images. The

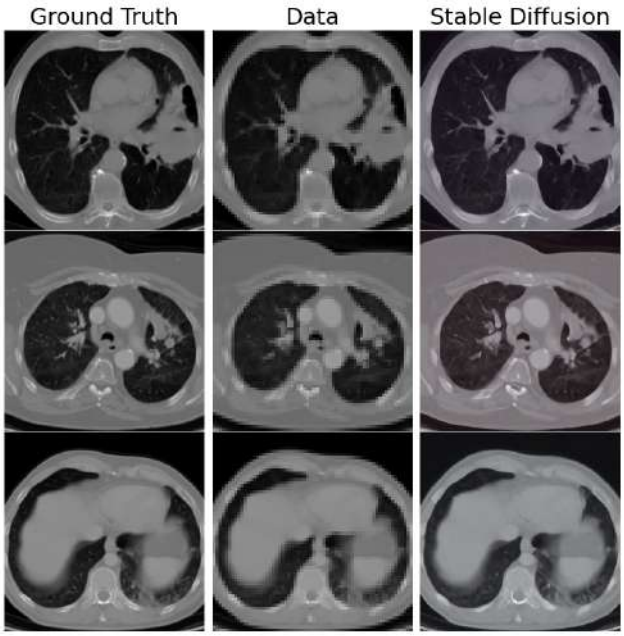


Figure 5. Comparison of HR, LR, and SR images from the dataset, illustrating the enhancement performance of the SR3 model.

RUNet and stable diffusion model successfully reconstructed some of the fine details from the low-resolution inputs. The experiment indicates that the UNet-based super-resolution models can be generalized to medical image synthesis with promising results, implying that the model can have practical applications in more accurate disease diagnosis.

### D.1. Model comparison: RUNet vs. SR3

Our SR3 model achieved an average SSIM of 0.177 and PSNR of 15.2334, which are lower compared to the Runet model's SSIM of 0.87 and PSNR of 27.5. Several factors contribute to these differences:

- **Nature of the SR3 Model:** The SR3 model, based on a diffusion framework and iterative refinement, focuses on generating high-quality, perceptually convincing images rather than achieving pixel-perfect accuracy. This approach often results in lower PSNR and SSIM scores, as these metrics favor models like Runet that prioritize exact pixel alignment with the target image.

- **Trade-off between Perceptual Quality and Metric Accuracy:** As observed in previous studies, conventional metrics like PSNR and SSIM do not always correlate well with human perception, especially for models generating high-frequency details. SR3, with its iterative refinement process, introduces such details but may not align perfectly with the target image, leading to lower metric scores.

- **Consistency with Low-Resolution Inputs:** While SR3 shows strong performance in maintaining consistency with low-resolution inputs, this consistency does not necessarily translate to higher PSNR and SSIM values. SR3's methodology, which does not rely on MSE regression, tends to generate outputs that are more diverse and visually appealing but may diverge slightly from the original high-resolution target, affecting these scores.

- **Model Capacity and Architecture:** The architecture and capacity of SR3, designed for balancing detail generation and consistency, might inherently lead to lower PSNR and SSIM scores compared to Runet, which may have an architecture optimized for these metrics.

- **Training Data and Objectives:** The training objectives and data used for SR3 emphasize perceptual quality over metric optimization. In contrast, models like Runet, which are trained with a direct focus on reducing pixel-level errors, achieve higher PSNR and SSIM values.

### D.2. Implications

These results were particularly evident in images where the model was able to maintain high levels of detail and texture fidelity, contributing to higher PSNR and SSIM scores. This observation indicates that while the average performance across a broad dataset may show lower metric scores for SR3, there are instances where the model's performance is highly competitive. Our findings reaffirm the limitations of conventional reference-based metrics in super-resolution. High PSNR and SSIM scores, as seen with the Runet model, do not always imply superior perceptual quality. The SR3 model's lower scores on these metrics highlight the trade-off between generating perceptually realistic images and achieving metric-based accuracy. This observation underscores the need for more comprehensive evaluation metrics that can better capture human perception of image quality in super-resolution tasks.

### D.3. Notable Performance on Selected Images

Despite the overall lower PSNR and SSIM scores of SR3, it is important to highlight that the model achieved a fairly good performance on certain images. Notably, the top-ranking images in terms of PSNR and SSIM demonstrated a closer parity with the performance of the RUnet model. These results suggest that in specific cases, the SR3 model has potentials comparable to traditionally metric-optimized models like RUnet.

- **Top-Performing Images:**

    1. Rank 1: PSNR: 24.705, SSIM: 0.62192; suggesting near parity with Runet in certain scenarios.
    2. Rank 2: PSNR: 24.138, SSIM: 0.68346; indicating the model's capability to achieve high-quality super-resolution.
    3. Rank 3: PSNR: 24.020, SSIM: 0.52956; showcasing the model's strength in certain challenging conditions.

These results were particularly evident in images where the model was able to maintain high levels of detail and texture fidelity, contributing to higher PSNR and SSIM scores. This observation indicates that while the average performance across a broad dataset may show lower metric scores for SR3, there are instances where the model's performance is highly competitive.

The performance of SR3 on these top-ranking images underscores the complexity of image super-resolution as a task. While average metric scores provide a useful overview, they may not fully capture the nuanced performance of models across different images and conditions. This further emphasizes the need for a more nuanced evaluation approach, taking into account both average performance and the ability to excel in specific scenarios.

## References


[1] Waqar Ahmad, Hazrat Ali, Zubair Shah, and Shoaib Azmat. A new generative adversarial network for medical images super resolution. *Scientific Reports*, 12(1):9533, 2022. 1

[2] Chao Dong, Chen Change Loy, Kaiming He, and Xiaoou Tang. Learning a deep convolutional network for image super-resolution. In *Computer Vision–ECCV 2014: 13th European Conference, Zurich, Switzerland, September 6-12, 2014, Proceedings, Part IV 13*, pages 184–199. Springer, 2014. 1

[3] Jonathan Ho, Ajay Jain, and Pieter Abbeel. Denoising diffusion probabilistic models. *Advances in neural information processing systems*, 33:6840–6851, 2020. 2

[4] Xiaodan Hu, Mohamed A Naiel, Alexander Wong, Mark Lamm, and Paul Fieguth. Runet: A robust unet architecture for image super-resolution. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, pages 0–0, 2019. 2

[5] Justin Johnson, Alexandre Alahi, and Li Fei-Fei. Perceptual losses for real-time style transfer and super-resolution. pages 694–711, 2016. 1

[6] Kensuke Umehara, Junko Ota, and Takayuki Ishida. Application of super-resolution convolutional neural network for enhancing image resolution in chest ct. *Journal of digital imaging*, 31:441–450, 2018. 1