FACULDADE DE ENGENHARIA DA UNIVERSIDADE DO PORTO

Deep Learning for Image Super-Resolution: Enhancing Generated CT Scans

Pedro Manuel Santos Pereira

Master in Informatics and Computing Engineering

Supervisor: Hélder Filipe Pinto de Oliveira, PhD

Co-supervisor: Luís Filipe Teixeira, PhD

Tutor: Tânia Maria Pereira Lopes, PhD

Tutor: Tiago Filipe Sousa Gonçalves, MSc

Tutor: Pedro Fernandes Sousa, MSc

July 22, 2024

Deep Learning for Image Super-Resolution: Enhancing Generated CT Scans

Pedro Manuel Santos Pereira

Master in Informatics and Computing Engineering

Approved by . . . :

President: Ana Paula Cunha da Rocha
Referee: António Manuel Trigueiros da Silva Cunha

July 22, 2024

Resumo

Nos dias de hoje, as aplicações da inteligência artificial (IA) estendem-se a diversos domínios e têm cada vez mais impacto no sector médico. A medicina surgiu como uma área promissora, para a exploração do potencial da IA no auxílio ao diagnóstico de doentes através da análise de exames de tomografia computorizada (TC). No entanto, a eficácia dos modelos de IA neste contexto é severamente limitada pela quantidade e qualidade dos dados disponíveis.

Desta forma, a geração de grandes volumes de dados é fundamental para o treino de modelos de IA robustos. Os dados sintéticos de TC são uma estratégia essencial para satisfazer a procura de abundância de dados. No entanto, imagens sintéticas sofrem tipicamente de má qualidade e baixa resolução, comprometendo a eficácia do subsequente treino de modelos.

Este trabalho aborda esta questão através da utilização de técnicas de super-resolução (SR) impulsionadas por um modelo de *deep learning*. Através desta abordagem, o modelo aumenta a resolução dos dados sintéticos de TC, produzindo resultados com uma qualidade significativamente melhorada e detalhes mais pormenorizados, o que resulta em dados sintéticos de TC de maior resolução que servem de input valioso para modelos de diagnóstico de IA de TC. Desta forma, colocamos a hipótese de que a utilização de SR de imagem permite uma deteção e análise mais precisas de doenças em estruturas de imagiologia médica.

Em aplicações anteriores de SR em exames de TC, as redes neuronais convolucionais (CNN) e as redes adversariais generativas (GAN) mostraram sucesso na melhoria da qualidade da imagem. Recentes avanços nos modelos de SR baseados em GANs demonstraram um progresso significativo. Este trabalho tem como objetivo tirar partido do poder dos novos modelos baseados em GAN, oferecendo uma metodologia mais robusta e precisa para melhorar a resolução e a qualidade dos dados de TC.

Um modelo de *deep learning* de SR de imagem focado em TCs de pulmão e mama foi desenvolvido. Este modelo aborda eficazmente os desafios de melhorar e aumentar a resolução de imagens médicas sintéticas, conduzindo a melhorias significativas na resolução e na qualidade das TCs sintéticas. Estes avanços, ao melhorarem os dados de treino para modelos de diagnóstico de IA baseados em TC, contribuem para uma deteção e análise de doenças mais precisas e fiáveis.

Abstract

Nowadays, artificial intelligence (AI) applications extend across diverse fields and are increasingly penetrating the medical sector. Medicine has emerged as a promising area, harnessing AI's potential to aid in patient diagnosis by analysing computed tomography (CT) scans. However, the efficacy of AI models in this context is severely limited by the availability of substantial and high-quality data.

This way, generating large volumes of data is fundamental for training robust AI models. Synthetic CT scan data is an essential strategy to meet the demand for data abundance. Nonetheless, synthetic image data typically suffers from poor quality and low resolution, compromising the effectiveness of subsequent model training.

This work tackles this issue by employing super-resolution (SR) techniques driven by a deep learning model. Through this approach, the model enhances the resolution of synthetic CT scan data, producing outputs with significantly improved quality and finer details, resulting in higher resolution synthetic CT scan data that serves as valuable input for CT scan AI diagnostic models. This way, we hypothesise that employing image SR generates more accurate disease detection and analysis within medical imaging frameworks.

In previous applications of SR in CT scans, convolutional neural networks (CNNs) and generative adversarial networks (GANs) have shown success in enhancing image quality. However, recent advancements in GAN-based SR models have demonstrated significant progress. This work aims to leverage the power of GAN-based models, offering a more robust and accurate methodology to improve CT scan data resolution and quality.

A highly capable image SR deep learning model focused on lung and breast CT scans was developed. This model effectively addresses the challenges of enhancing and upscaling synthetic medical data, leading to significant improvements in the resolution and quality of synthetic CT scans. These advancements, by enhancing the training data for AI diagnostic models based on CT scans, contribute to more precise and reliable disease detection and analysis.

Acknowledgements

I would like to express my gratitude to my main supervisor, Hélder Oliveira, for his guidance and support throughout my research. I am also deeply grateful to my co-supervisors, Tânia Pereira, for all the report reviews and all the other project contributions, and especially Tiago Gonçalves, who has been crucial in guiding me and providing invaluable assistance, helping me with all my questions and doubts I had along the way. I would like to thank Pedro Sousa for his knowledge contributions and presence during this project.

A big thank you to my colleagues, Diogo Campas and João Martins. We went through this journey together, supporting each other along the way, and it made the process much more enjoyable and manageable.

I also want to thank INESC TEC for welcoming me with open arms and providing me with a great experience working on this project.

I would like to thank the co-financing the co-financing provided by Component 5 - Capitalisation and Business Innovation, integrated within the Resilience Dimension of the Recovery and Resilience Plan under the Recovery and Resilience Mechanism (MRR) of the European Union (EU), framed within the Next Generation EU for the period 2021 - 2026, within project HfPT, with reference 41. This support has created a perfect work environment for developing this project.

I am deeply thankful to my family — my mom, dad and brother — for their invaluable support, not only during this phase but since the beginning of my academic journey. Their encouragement and belief in me since the beginning have been a constant source of motivation and a key factor to all my successes. I truly wouldn't have achieved what I have achieved today without their education and guidance, including my brother, who challenges me every day to be the best version of myself and aim high. I also want to thank my grandmother for her enduring support and care since I was born, which significantly eased my life, allowing me to always focus on work and studies. A special mention goes to my grandfather, whose influence was an important factor in my accomplishments and in shaping who I am today.

Lastly, I would like to thank my girlfriend for accompanying my work closely since the beginning. Her essential help, including providing a CT scan, support, and encouragement, helped me get through the toughest times, not only during this project but also since high-school. I hope I can retribute all help next year, when the time to develop your own dissertation comes.

Pedro Manuel Santos Pereira

"A galinha da vizinha é sempre melhor que a minha."

Contents

List of Figures

List of Tables

Abbreviations

AI	Artificial Intelligence
CNN	Convolutional Neural Network
CT	Computed Tomography
DICOM	Digital Imaging and Communications in Medicine
GAN	Generative Adversarial Network
HR	High-Resolution
JPEG	Joint Photographic Experts Group
LIDC	Lung Image Database Consortium
LPIPS	Learned Perceptual Image Patch Similarity
LR	Low-Resolution
MMD	Maximum Mean Discrepancy
MOS	Mean Opinion Score
MRI	Magnetic Resonance Imaging
MS-SSIM	Multi-scale structural similarity
MSE	Mean Squared Error
PIL	Python Imaging Library
PSNR	Peak Signal-to-Noise Ratio
RIDER	Reference Image Database to Evaluate Therapy Response
RNN	Recurrent Neural Network
SDG	Sustainable Development Goals
SR	Super-Resolution
SSI	Structural Similarity Index
SSIM	Structural Similarity Index Measure

Chapter 1

Introduction

1.1 Context

1.1.1 Computed Tomography Scan

Computed tomography (CT) scans stand as a cornerstone of diagnostic imaging, providing medical personnel with intricate cross-sectional views of the human body's internal structures [9]. These scans utilise X-rays to generate detailed images that offer unparalleled insights into anatomical anomalies, tissue abnormalities, and pathologies, as seen in Figure 1.1. By capturing precise anatomical data from various angles, CT scans play a crucial role in identifying conditions that might otherwise remain concealed from conventional imaging techniques [44].

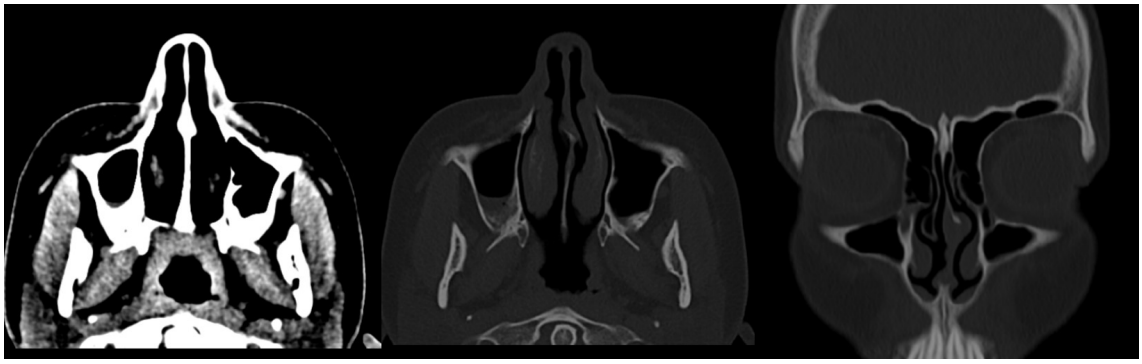


Figure 1.1: Paranasal sinuses CT scan example, courtesy of Mafalda Pereira.

1.1.2 Artificial Intelligence

Artificial Intelligence (AI) [62] has transformed numerous industries, wielding its potential to revolutionise healthcare by delving into medical diagnostics [60]. Among other important applications, AI's integration into the analysis of medical imaging, specifically computed tomography (CT) scans, has become a beacon of hope in improving and increasing the accuracy of diagnoses and treatment plans.

1.2 Motivation

The intersection of AI and medicine has paved the way for groundbreaking advancements. However, the scarcity of substantial and high-quality data is an obstacle to the development of top-notch, reliable and accurate deep learning models [40].

Efforts to bypass this data shortage issue have prompted investigations into synthetic CT scan data as a solution to expand the dataset for model training. However, the adoption of artificially generated data presents a dilemma. Despite the increase in the quantity of available data, these images suffer from bad quality and low-resolution (LR), affecting the resulting model's accuracy and efficacy in diagnosing [71] [32].

Therefore, this thesis aims to equip deep learning models with enhanced images to significantly bolster their accuracy and efficacy in diagnostic tasks. Thus, the aim is to achieve this objective by employing state-of-the-art deep learning-based image super-resolution (SR) models on artificially generated data.

1.3 Goals

The main goals of this project are the following:

1. Development of a deep learning-based image SR model specialised in CT scan data;
2. Deployment of the developed deep learning-based SR model on artificially generated data.

1.4 Contributions

This dissertation introduces a deep learning SR model developed on the basis of the Real-ESRGAN architecture, focused specifically on enhancing breast and lung CT scan data. The efficacy and utility of the developed model have been rigorously demonstrated, showcasing its ability to enhance synthetic medical image data of both lung and breast through superior resolution enhancement and preservation of anatomical details and textures.

By addressing the inherent challenges of LR and poor quality in synthetic images, the model not only improves the visual fidelity of these scans but also enhances their suitability as input for diagnostic AI systems. This advancement is crucial in improving the accuracy and reliability of AI-based diagnostic tools, ultimately contributing to more effective patient care and medical decision-making.

1.5 United Nations Sustainable Development Goals

1.5.1 Goal Identification

This project focuses on developing a model that applies SR to synthetic CT scan images. This work is part of a pipeline aimed at enhancing the development of deep learning diagnostic models

also based on CT scans. Therefore, the primary Sustainable Development Goal (SDG) [64] that this project addresses is goal number 3: *Good Health and Well-Being*.

1.5.2 Specific Targets and Contributions

More specifically, within SDG goal number 3, this project contributes to the following targets:

- Target 3.4 - "*Reduce by one third premature mortality from non-communicable diseases through prevention and treatment and promote mental health and well-being*".

  By contributing to a tool that can assist clinicians in the early detection and diagnosis of non-communicable diseases, this work can facilitate timely and effective treatment, potentially reducing mortality rates associated with these diseases.

- Target 3.8 - "*Achieve universal health coverage, including financial risk protection, access to quality essential health-care services and access to safe, effective, quality and affordable essential medicines and vaccines for all*".

  Also, by enhancing clinical decision-making, improving diagnostic accuracy and speed, the result of this project can also lead to a more efficient use of healthcare resources, potentially reducing the costs associated with prolonged illness, further aligning with Target 3.8.

1.5.3 Performance Indicators

To measure the impact of this project on achieving SDG number 3, the following performance indicators can be assessed:

- Diagnostic Accuracy Rate - The percentage of correct diagnoses made by the deep learning model compared to standard diagnostic methods.
- Patient Outcome Metrics - Improvement in patient outcomes, including survival rates, disease progression, and quality of life, tracked over time for those diagnosed using the model.
- Healthcare Costs - The decrease in healthcare costs achieved by earlier and more accurate diagnoses, measured through cost-benefit analysis.

By systematically tracking these indicators, we can evaluate the success of our project in contributing to SDG 3 and its specific targets, reducing premature mortality from non-communicable diseases and reducing healthcare costs.

1.6 Dissertation Structure

In addition to this Introduction, chapter 1, this dissertation contains 5 more chapters: chapter 2 provides relevant background information and an in-depth analysis of the latest advancements in deep learning models relevant to the subject. Furthermore, chapters 3 and 4 detail the technologies

employed and the methodologies followed in this work. Chapter 5 presents a thorough discussion and analysis of the results obtained. Finally, in chapter 6, a summary of the project's accomplishments is provided, along with a discussion of potential directions for future research. Appendix A contains supplementary statistical tables of the results.

Chapter 2

Literature Review

2.1 Introduction

Developing and training deep learning-based interpretation models for various applications such as image recognition, natural language processing and medical diagnosis using artificially generated data can improve model generalisation, accelerate training, and potentially address data scarcity issues. Taking into account that artificially generated data typically suffers from lower resolution and poor quality, enhancing these images and using them as input fed into the models results in a more accurate and efficient model. This thesis will focus on applying this technology to feed medical CT scans deep learning interpretation models with higher quality artificially generated data. In this chapter, we will explore the background and the latest advancements concerning AI, focusing on artificially generated medical data enhancement techniques.

2.2 Background and Fundamentals

Deep learning [22] is a sub-type of machine learning, a subset of AI that utilises neural networks composed of multiple layers to learn and extract patterns from data. AI was created in an attempt by scientists to make computers imitate human reasoning [62], decision-making, and logic.

These networks [55], heavily inspired by the human brain, consist of interconnected nodes (neurons) organised in layers: an input layer, one or several hidden layers where complex features are extracted from the input, and an output layer. Lower layers recognise simple shapes and patterns in the data, usually also recognisable by humans, while higher layers build upon these to understand more complex and abstract features. This design (see Figure 2.1) allows deep neural networks to automatically learn representations from data without relying on explicit feature engineering, leading to widespread use in diverse fields such as image and speech recognition, natural language processing, and image SR.

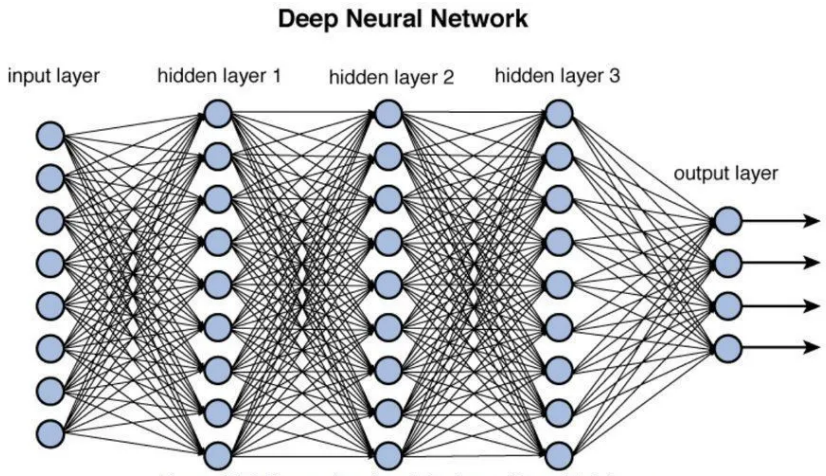


Figure 2.1: Neural network multiple layers, from [51].

Forward propagation and backpropagation are very important concepts. Forward propagation is the initial phase where input data moves through the network, undergoing computations in each layer to produce predictions. This process involves applying weights, biases, and activation functions sequentially. Backpropagation, crucial during training, assesses prediction errors and adjusts the network's internal parameters, such as weights and biases. It evaluates each parameter's contribution to the error and updates them backwards through the network, enabling gradual refinement of predictions. Furthermore, one crucial factor in deep learning algorithms is the chosen loss function [29]. The loss function measures how well an algorithm's prediction matches the ground truth by quantifying the disparity between the predicted output and the actual target values. This function allows the algorithm to iteratively adjust its parameters to minimise this disparity and improve its predictive accuracy.

Activation functions serve as crucial components that introduce non-linearities to the output of neurons. These are essential for the network to learn and represent complex relationships present in data. By applying an activation function to the output of each neuron, the network gains the ability to model and capture intricate patterns, enabling it to handle non-linear data distributions effectively. One of the most commonly utilised and up-to-date activation functions is the ReLU [47] and the more straightforward logistic sigmoid function.

The Perceptron [56], developed in the late 1950s by Frank Rosenblatt, marked the inception of neural networks. This algorithm is a simple neural network aimed at binary classification tasks, learning to distinguish between two classes based on input features by adjusting weights to make predictions. Following this breakthrough, the field of AI experienced what's famously known as the "AI winter", where interest and funding in AI research declined significantly, leading to no significant developments in this area.

The resurgence of interest in neural networks emerged in the late 20th century with the development of the backpropagation algorithm [55], enabling the training of multi-layer neural networks.

Technological advancements played a pivotal role in reshaping the landscape of AI. The invention of better artificial neural network training algorithms, big data availability, and increased computational power led deep learning algorithms to outperform traditional machine learning methods in numerous domains [35].

This era not only reinvigorated AI research but also paved the way for its widespread adoption in diverse fields, ranging from healthcare [18] and finance [26] to autonomous vehicles [6] and natural language processing [77].

2.3 Computer Vision

Computer vision is a field of AI that enables computers to interpret and understand visual information from images [28]. It pursues to mimic human vision to automate tasks that require visual cognition, enhancing capabilities in various sectors. Traditional computer vision involves steps like image acquisition, preprocessing to enhance quality, feature extraction, segmentation, and object recognition. These foundational processes are now complemented by modern techniques such as deep learning, particularly CNNs, which can discern complex visual patterns from vast datasets.

Computer vision requires extensive data, using repeated analyses to recognise images. Nowadays, this is mainly achieved with CNNs. CNNs break images into pixels, label them, and perform convolutions to predict image content. Iterative predictions refine the model's accuracy, simulating human vision. CNNs handle single images, while recurrent neural networks (RNNs) [23] manage video sequences by understanding frame-to-frame relationships.

The main tasks and applications of computer vision are the following:

- Image classification - Classifies images into predefined categories based on their content.
- Object detection - Identifies and localises multiple objects within an image or video.
- Object tracking - Continuously monitors and follows objects of interest across frames in a video or sequence of images.
- Image retrieval - Searches and retrieves images from large datasets based on visual content.

2.4 Image Evaluation Metrics

Image quality evaluation [69] involves assessing visual attributes, focusing on how viewers perceive images. Methods for assessing image quality typically fall into two categories: subjective methods, which rely on human perception and judge how realistic an image appears, and objective computational methods.

While subjective methods align more closely with our needs, they are often time-consuming and costly. Consequently, objective methods have become mainstream. However, these objective approaches may not consistently align with human perception, leading to discrepancies in results.

2.4.1 Objective Methods

2.4.1.1 Peak Signal-to-Noise Ratio

Peak Signal-to-Noise Ratio (PSNR) [27] is one of the most commonly used metrics for image quality assessment of reconstructed or processed images, quantifying image quality degradation. In the context of image SR, the PSNR measurement calculates the ratio between the maximum possible power of a signal and the power of corrupting noise that affects the fidelity of the image, as seen on 2.1. PSNR is typically expressed in decibels (dB). Higher PSNR values imply better quality, indicating lower levels of distortion or noise in the processed image compared to the original.

$$\text{PSNR} = 10 \cdot \log_{10} \left(\frac{\text{MAX}^2}{\frac{1}{n} \sum_{i=1}^{n} (y_i - \hat{y}_i)^2} \right), \tag{2.1}$$

where MAX represents the maximum possible pixel value of the image (usually 255 in 8-bit images).

It's essential to note that while PSNR is a common and straightforward metric, it might not fully align with human perception, especially in scenarios where subjective quality judgements are crucial, performing poorly. Hence, it's often used alongside other metrics or subjective evaluations to provide a more comprehensive understanding of image quality.

2.4.1.2 Structural Similarity Index Measure

Another commonly used image evaluation tool is the Structural Similarity Index Measure (SSIM) [27]. SSIM measures not only the pixel difference but also the similarity between the structural information (see Equation 2.2), luminance changes (see Equation 2.3) and contrast (see Equation 2.4) of two images. Combining the structural, luminance and contrast equation,

$$\text{Structure} = \frac{\sigma_{XY} + c_3}{\sigma_X \sigma_Y + c_3} \tag{2.2}$$

$$\text{Luminance} = \frac{2\mu_X \mu_Y + c_1}{\mu_X^2 + \mu_Y^2 + c_1} \tag{2.3}$$

$$\text{Contrast} = \frac{2\sigma_X \sigma_Y + c_2}{\sigma_X^2 + \sigma_Y^2 + c_2} \tag{2.4}$$

we get the final SSIM formula (see Equation 2.5):

$$\text{SSIM}(X,Y) = \frac{(2\mu_X \mu_Y + c_1)(2\sigma_{XY} + c_2)}{(\mu_X^2 + \mu_Y^2 + c_1)(\sigma_X^2 + \sigma_Y^2 + c_2)}, \tag{2.5}$$

where μ_X and μ_Y are the means of images X and Y, σ_{XY} is the covariance between μ_X and μ_Y ,σ_X^2 and σ_Y^2 are the variances of images X and Y and c_1 and c_2 are constants to stabilise the division.

2.4.1.3 Multi-scale structural similarity

Multi-scale structural similarity (MS-SSIM) [68] is an extension of the pre-existing SSIM method, also measuring similarity of luminance, contrast and structure between images but also further analyses images at multiple scales by decomposing them into different levels of detail through a process of Gaussian blurring and down-sampling.

2.4.1.4 Maximum Mean Discrepancy

MMD (Maximum Mean Discrepancy) [7] is a kernel based statistical test that measures the distance between two probability distributions. In the context of image SR, MMD compares distributions indirectly (see Equation 2.6) by comparing the distributions of features extracted by images rather than directly treating images as probability distributions.

$$\mathrm{MMD}(P,Q) = \|\mathbb{E}_{X\sim P}[\phi(X)] - \mathbb{E}_{Y\sim Q}[\phi(Y)]\|_{\mathscr{H}} \tag{2.6}$$

where μ_X and μ_Y are the mean embeddings of distributions P and Q in the reproducing kernel Hilbert space $\mathscr{H}$, respectively.

2.4.1.5 Learned Perceptual Image Patch Similarity

LPIPS (Learned Perceptual Image Patch Similarity) [78] is based on a deep neural network model that, unlike other methods, focuses on perceptual differences as perceived by humans instead of pixel-wise differences or structural similarities. This way, LPIPS is useful to evaluate a more humane perception of quality of generated images.

2.4.2 Subjective Methods

2.4.2.1 Mean Opinion Score

The Mean Opinion Score (MOS) [70] is a metric used in quality assessment to measure subjective opinions or perceptions of human observers specialised or familiarised with the context of the project. Human judgement is taken into account by using people to rate perceptual quality scores on a given scale. The final result is obtained by calculating the mean of all the individual ratings.

Although MOS is crucial in quality assessment processes, especially when human perception and preferences play a significant role, it's important to note that MOS has certain limitations. Subjective assessments, while valuable, can be influenced by individual biases, environmental factors during testing, and variations in human perception.

2.5 Image Super-Resolution

Image SR [72] is a technique used in image processing tasks to enhance the quality or resolution of an image beyond its original size or dimensions. The goal is to create a higher-resolution image

or a more detailed version with only a lower-resolution image as input by predicting the missing information using machine learning.

The two main methods to achieve image SR are computer vision-based and deep learning-based. Computer vision-based image SR involves traditional techniques like interpolation [59]. These methods often rely on predefined algorithms, such as bicubic interpolation or edge-preserving filters, to increase the image size and quality without additional training on specific datasets.

On the other hand, deep learning-based image SR utilises neural network architectures, such as convolutional neural networks (CNNs) [15] and generative adversarial networks (GANs) [36], trained on large datasets that, by learning complex patterns and features within images, aim to achieve image SR.

2.5.1 Computer Vision Image Super-Resolution

2.5.1.1 Interpolation

Interpolation is a basic computer vision technique known as one of the simplest and earliest methods. It involves taking into account the average value of the nearest pixels to generate a new pixel and thus increasing the resolution. Interpolation techniques mainly differ by the number of neighbour pixels taken into account to generate a new one: nearest-neighbour interpolation selects only the value of the nearest pixel, bilinear interpolation takes into account the 4 pixels surrounding the newly generated pixel, and bicubic interpolation considers a larger area, usually of 16 pixels, and uses a cubic function to calculate the value of the new pixel (see Figure 2.2).

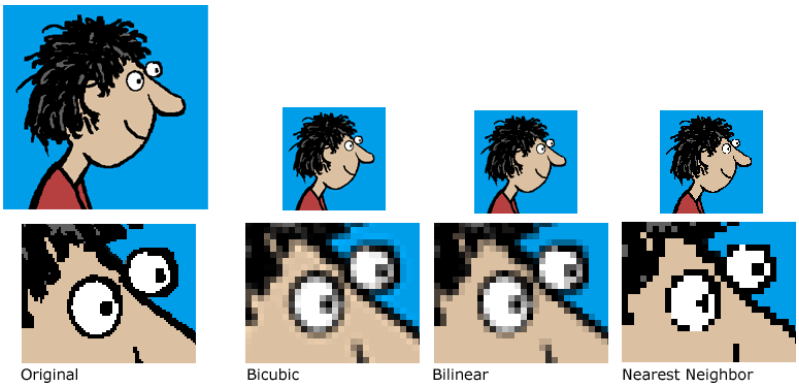


Figure 2.2: Example of interpolation results, from [42].

2.5.2 Deep Learning-Based Super-Resolution

2.5.2.1 Convolutional Neural Networks

Other more complex approaches to image SR include CNN and GANs. A CNN is a specialised artificial neural network designed for processing images. It is structured to learn patterns and features in data by learning hierarchical representations of features from input data, particularly structured grid-like data such as images, detecting various features, like edges, textures, or shapes.

This is achieved through convolutional layers for feature extraction, upscaling layers to increase spatial resolution, and convolutional layers for image reconstruction [14].

2.5.2.2 Generative Adversarial Networks

GANs are a type of neural network architecture comprising two networks that work competitively, designed to create new data resembling the original training examples [13]. The generator tries to create images indistinguishable from real data, while the discriminator tries to distinguish the original image from the one generated by the generator. They improve iteratively: the generator learns to produce more realistic data to fool the discriminator, and the discriminator gets better at distinguishing real from generated data, as seen in Figure 2.3.

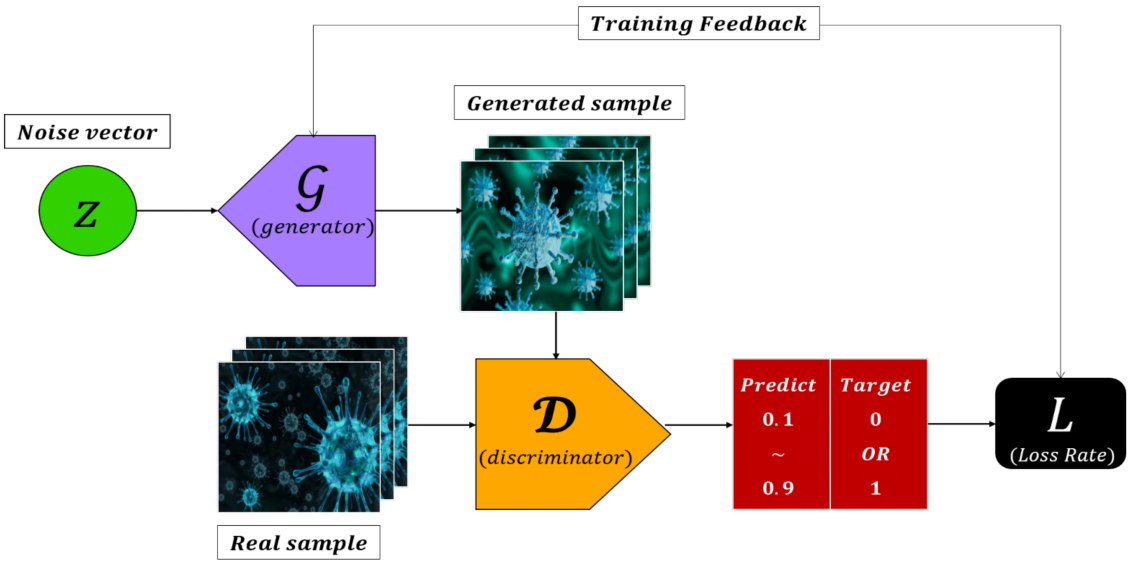


Figure 2.3: GAN's pipeline from [49].

This dynamic between the generator and discriminator in GANs mirrors a non-cooperative, two-player game, each adapting in response to the other, aiming for a strategic equilibrium akin to the Nash equilibrium [46], reflecting the essence of the minimax formula (see Equation 2.7). This equation encapsulates their competitive nature, as the generator **G** minimises its loss while the discriminator **D** maximises its advantage, converging to generate increasingly realistic data:

$$\min_G \max_D V(D,G) = \mathbb{E}_{x\sim p_{\text{data}}(x)}[\log D(x)] + \mathbb{E}_{z\sim p_z(z)}[\log(1-D(G(z)))] \tag{2.7}$$

, where $G(x)$ is the generator function and $D(x)$ is the discriminator function. The expression $\mathbb{E}_{x\sim p_{\text{data}}(x)}[\log D(x)]$ signifies the expected value, computed over real data samples x drawn from the true data distribution $\sim p_{\text{data}}(x)$, of the logarithm of the discriminator's output when processing real data. Maximising this term drives $D(x)$ towards 1, indicating accurate classification of real data. $\mathbb{E}_{z\sim p_z(z)}[\log(1-D(G(z)))]$ represents the expected value, calculated over noise samples z drawn from a prior distribution $\sim p_z(z)$, of the logarithm of the discriminator's outputting when

processing generated data $G(z)$. The generator aims to minimise this term, steering $D(G(z))$ towards 1, implying that generated data is classified as real. The objective involves minimising G's function $\min_G$ to enhance the generator's ability to create realistic outputs while simultaneously maximising D's function $\max_D$ to improve its capability to distinguish between real and generated samples.

As of today, GANs have shown to be one of the most successful in generating realistic images. This way, combining the strengths of deep learning models, adept at learning representations and making predictions, with generative models focused on creating new data from learned patterns seems promising with the use of a SR generative adversarial network algorithm (SRGAN) [37]. Upon reviewing pertinent literature, it appears that there are some underexplored applications of the fusion of deep learning and GANs within the dissertation's focus (CT scans).

2.5.3 Image Degradation Functions

Image degradation functions are models that simulate the degradation of high-resolution (HR) images into low-resolution ones. These functions are crucial for model training, providing low-resolution images that correspond to their HR ground truth counterparts. This way, the process that determines image degradation results is very important as it directly impacts the fidelity and realism of the low-resolution synthetic CT scan data.

By running these low-resolution images through an SR model, we aim to produce output images that are as similar as possible to the original HR ground truth images.

Some of the most common image degradation techniques include downsampling, which reduces image resolution through methods like nearest-neighbour, bilinear, or bicubic interpolation. Blurring functions, using kernels such as Gaussian or motion blur, simulate the loss of sharpness, often combined with downsampling. Real-world images also contain noise due to sensor imperfections or low-light conditions, modelled as Gaussian, Poisson, or salt-and-pepper noise. Compression artefacts from methods like JPEG introduce blocking, ringing, and blurring effects, which are also simulated. Additionally, changes in colour accuracy and saturation and geometric distortions like rotations and translations are considered.

2.5.4 Loss Functions

The loss function applied to image SR guides the training process on supervised deep learning-based SR models by quantifying how well the model's output matches the desired HR image [65]. This way, there's a variety of loss functions that can be employed, each emphasising different aspects of image fidelity and quality.

2.5.4.1 Pixel Loss

The simplest loss function method is pixel loss that includes either mean absolute error (MAE) (see Equation 2.8) and mean squared error (MSE) (see Equation 2.9), which calculate the average of the absolute and squared differences between predicted and actual pixel values.

$$\text{MAE} = \frac{1}{hwc}\sum_{i,j,k}^{n}|\hat{y}_{i,j,k} - y_{i,j,k}| \tag{2.8}$$

$$\text{MSE} = \frac{1}{hwc}\sum_{i=1}^{n}(\hat{y}_{i,j,k} - y_{i,j,k})^2, \tag{2.9}$$

where h, w, and c correspond to the height, width and number of channels of the images.

The Charbonnier loss function [4], a variation of pixel loss, is a differentiable approximation of the L1 (absolute error) and L2 (squared error) loss functions. The Charbonnier loss function is defined in equation 2.10:

$$L_{\text{pixel}}(I;\hat{I}) = \frac{1}{hwc}\sum_{i,j,k}\sqrt{(\hat{I}_{i,j,k} - I_{i,j,k})^2 + \varepsilon^2} \tag{2.10}$$

, where ε is a constant for computation stability.

2.5.4.2 Perceptual Loss

A more complex loss function is perceptual loss, which considers not only the pixel difference but also the higher-level visual content and structure that contribute to and are more related to how humans perceive an image, which results in overall better-looking images as seen on Figure 2.4. This function (see Equation 2.11) leverages neural networks to measure the difference between HR images and their generated counterparts, showing better results than using only pixel loss.

$$\mathscr{L}\text{Perceptual} = \frac{1}{N}\sum i = 1^{N}\left|\Phi(I^i_{\text{HR}}) - \Phi(G(I^i_{\text{LR}}))\right|_2^2 \tag{2.11}$$

Ground Truth Bicubic SRCNN Perceptual loss

Figure 2.4: Example results comparing bicubic interpolation, SRCNN [14], and perceptual loss [30], from [30].

2.5.4.3 Adversarial Loss

Adversarial loss resorts to the use of GANs to introduce a competitive framework in training neural networks [37]. Specifically, in the context of GANs, adversarial loss involves a generator network and a discriminator network engaged in a game-like scenario. As said before, the generator strives

to produce image outputs that resemble samples from a target distribution, aiming to fool the discriminator, while the discriminator aims to distinguish between real and generated samples. This way, adversarial loss, derived from the competitive interplay between the two networks, guides the training process towards generating higher perceptual quality image data, thus classifying it as a loss function.

2.5.4.4 Prior-Based Loss

A prior-based loss function [17] is especially used in the context of generative models, incorporating prior knowledge into the training process. Knowing the content of the generated higher-resolution image in advance, this function encourages the output to align with specific attributes or properties found in the training data. The exact formulation of a prior-based loss varies depending on the specific application, the desired properties of the generated data, and the model architecture being used.

2.6 State-of-the-Art Techniques in Deep Learning-Based Image SR

The state-of-the-art techniques of this field can be divided into two major groups: supervised and unsupervised learning [70]. In supervised learning, models are trained with low-resolution and the corresponding HR images. For the majority of applications, supervised learning serves as a robust and effective approach due to its ability to utilise labelled data for training models, resulting in accurate predictions across diverse domains and tasks. Even so, unsupervised learning shines in scenarios where acquiring matched LR and HR image pairs for supervised learning is challenging. This approach circumvents the need for predefined degradation and instead focuses on unpaired LR-HR images, allowing models to learn the genuine LR-HR mapping without relying on manual priors.

2.6.1 Supervised Image Super-Resolution

2.6.1.1 SRCNN

SRCNN [14], a pioneering deep learning model in image SR, employs a deep CNN architecture to create an end-to-end mapping between low-resolution and HR images. Beginning with the input of a low-resolution image, the SRCNN's initial convolutional layer extracts a series of feature maps as seen on Figure 2.5. These feature maps undergo nonlinear mapping in the subsequent layer to represent HR patches. Finally, the last layer integrates these predictions within a spatial neighbourhood, thus generating a higher-resolution image.

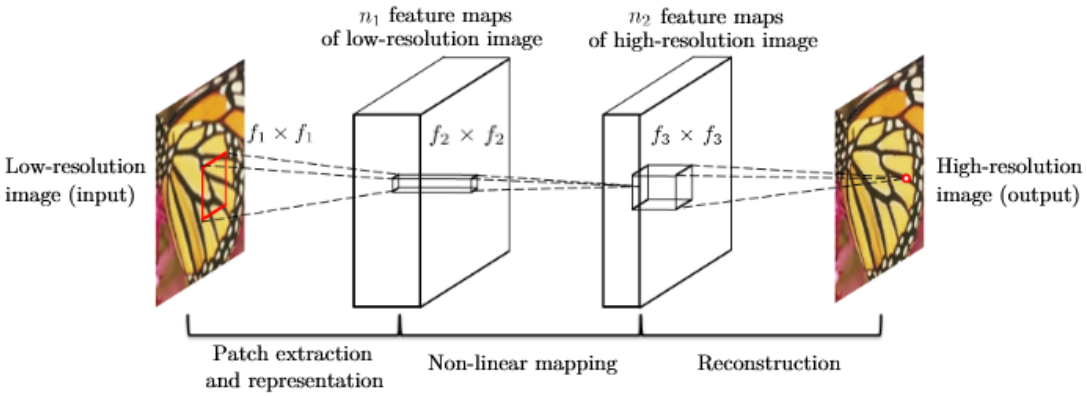


Figure 2.5: SRCNN layers sequence, from [14].

2.6.1.2 SRGAN

GANs offer a potent framework for image SR, and SRGAN [37] is built on top of it. SRGAN, a GAN-based network employing a novel perceptual loss, replaces the conventional MSE-based content loss with perceptual loss calculated on network feature maps, offering greater invariance to pixel space changes.

The fundamental premise behind this formulation involves training a generative model G to deceive a discriminator D, which in turn learns to distinguish upscaled images from real ones. This approach enables our generator to craft solutions remarkably similar to real-life images, challenging D's classification. Such training encourages the generation of perceptually superior and more natural-looking images, differing from SR solutions obtained through pixel-wise error minimisation, like the MSE.

2.6.1.3 ESRGAN

ESRGAN [66], an Enhanced Super-Resolution Generative Adversarial Network, improves upon SRGAN by refining its network architecture, adversarial losses, and perceptual features. It introduces the innovative Residual-in-Residual Dense Block (RRDB), a block that allows better information flow through networks by combining residual connections and dense connections, enabling efficient feature reuse and facilitating gradient propagation while removing batch normalisation layers from the generator, resulting in superior visual quality and more natural textures as seen on Figure 2.6.

Furthermore, ESRGAN employs a network interpolation strategy aimed at striking a balance between perceptual quality and PSNR.

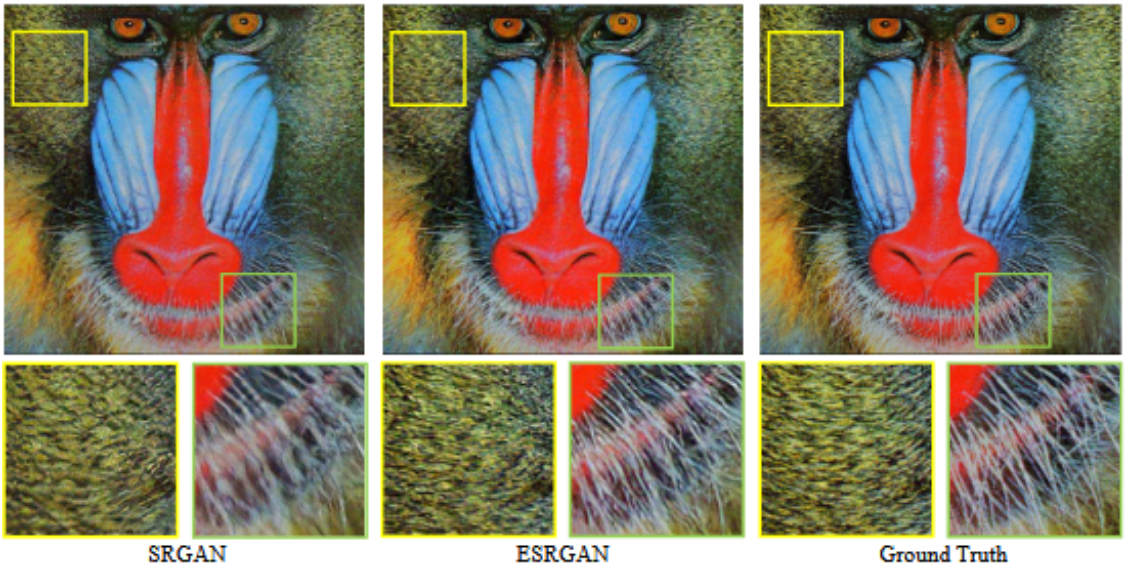


Figure 2.6: SRGAN, ESRGAN and ground truth comparison, taken from [66].

2.6.1.4 Real-ESRGAN

Real-ESRGAN [67] further improves on ESRGAN by introducing a high-order degradation modelling process to better simulate complex real-world degradations. Real-ESRGAN employs a U-Net discriminator with spectral normalisation to enhance discriminator capability and stabilise the training dynamics. This architecture allows the network to capture both local and global features while preserving spatial information, making it especially effective for tasks where spatial details are crucial, such as image synthesis and SR.

All this results in a better image restoration process where, as seen in Figure 2.7, smooth textures remain smooth without neglecting the preservation of sharpness in noisy areas.

Figure 2.7: Comparison between Bicubic interpolation, ESRGAN, RealSR and Real-ESRGAN, taken from [67].

2.6.1.5 RCAN

RCAN [79], Residual Channel Attention Networks, is recognised for its emphasis on channel attention mechanisms and residual learning. RCAN employs residual connections, addressing the problems caused by very deep models by introducing shortcut connections, allowing the network to learn residual functions instead of directly mapping inputs to outputs, bypassing some layers, and directly passing the input from one layer to a deeper layer as seen on Figure 2.8.

The channel attention mechanism's implementation is crucial due to its ability to dynamically emphasise informative features across channels, allowing networks to adaptively assign importance, enhance feature representations, and improve overall learning and performance in various tasks.

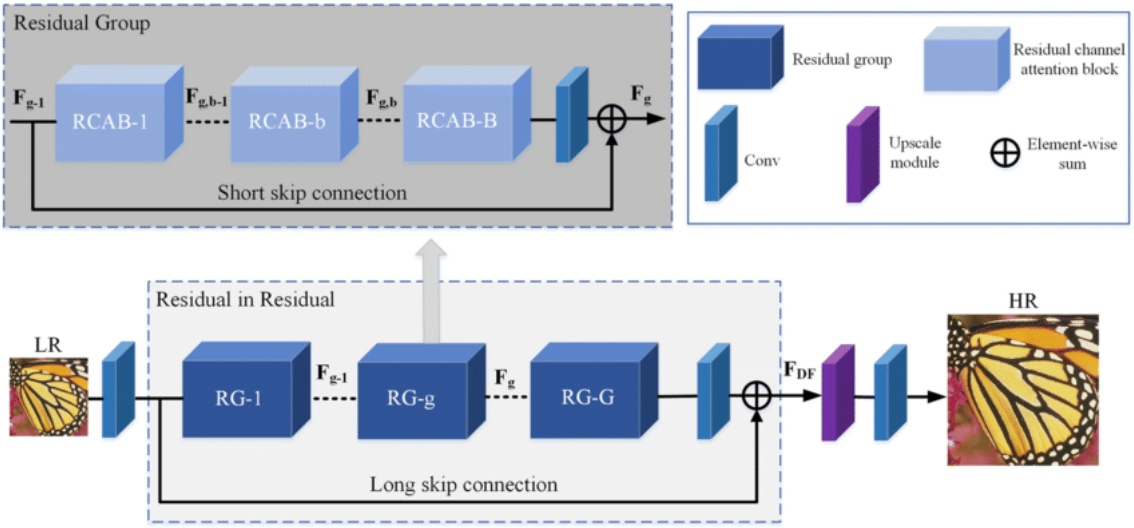


Figure 2.8: RCAN's residual channel diagram, taken from [21].

2.6.1.6 CARN

CARN [2] emphasises efficiency, accuracy, and lightweight architecture. It employs cascading residual blocks to capture fine-grained details. Cascading implies a series or a cascade of these residual blocks stacked one after another within the network architecture. Each block typically consists of multiple layers, such as convolutions and batch normalisation, and skip connections cascaded together to form deeper structures. The cascading arrangement of these blocks (see Figure 2.9) enables the network to learn hierarchical and increasingly abstract representations as information passes through multiple layers.

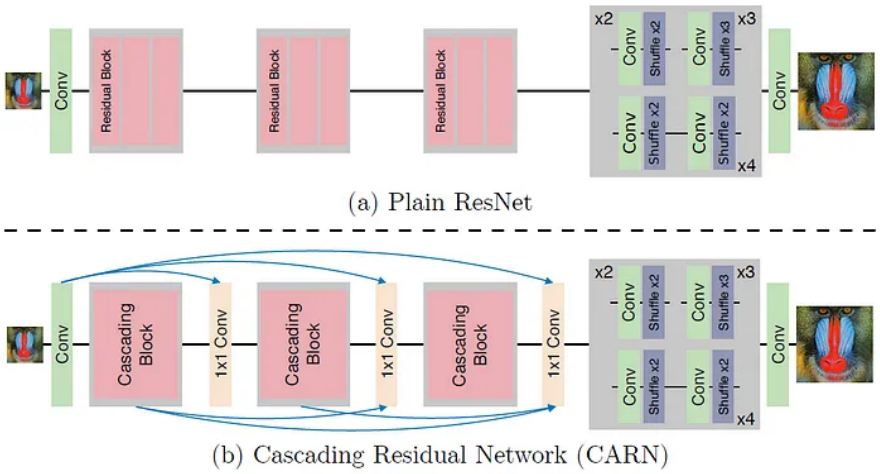


Figure 2.9: CARN's cascading network architecture compared to a standard method, from [61].

2.6.1.7 MS-LapSRN

MS-LapSRN [34] is based on LapSRN [33] that incorporated the concept of laplacian pyramids into its architecture. MS-LapSRN addresses limitations by employing a cascade of CNNs, utilising cascaded convolutional layers to extract feature maps and upsample them with transposed convolutional layers. Then, using convolutional layers, it predicts residuals between image scales, helping reconstruct HR images more effectively. Despite its cascaded structure, LapSRN is trained end-to-end with a Charbonnier [70] loss function, a variation of a pixel loss function.

2.7 Applications of Deep Learning-based Image Super-Resolution

2.7.1 Video

Video SR involves enhancing the resolution and quality of low-resolution video frames. This application has a higher degree of complexity since temporal information such as motion, brightness, and contrast has to be taken into account when applying SR to adjacent frames.

This extra complexity requires extra techniques, such as optical flow-based methods [39] for motion compensation and methods employing recurrent structures to capture spatial-temporal dependencies without relying on direct motion compensation strategies [24].

2.7.2 Depth Map

This application involves applying SR to the quality of depth maps, which represent the distance of objects in a scene, improving the accuracy and detail of the depth perception. As explained on [70], depth map SR is pivotal in tasks like pose estimation, whose purpose is to determine the spatial positioning of objects within an image, and semantic segmentation, which involves assigning specific labels or categories to each pixel in an image.

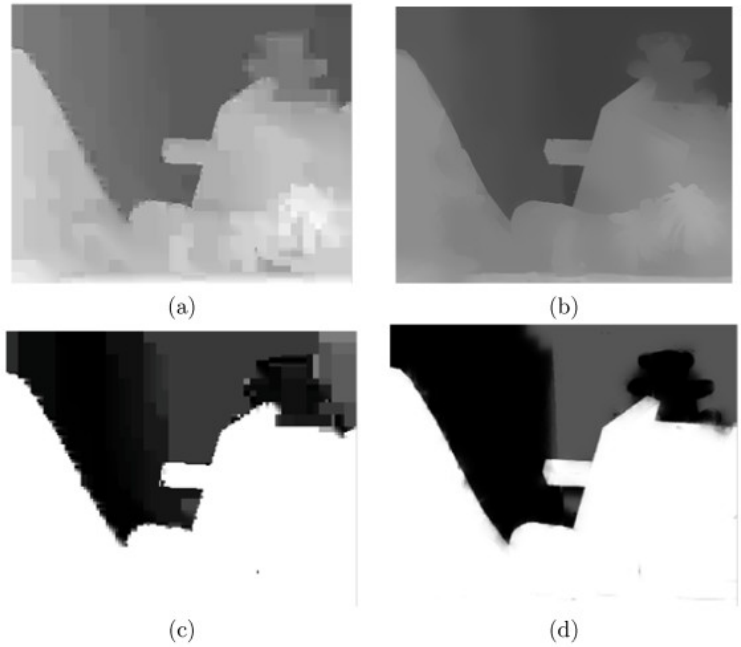


Figure 2.10: SR depth filtering example from [57].

In Figure 2.10, it's visible the effect of applying SR to a depth map. The compressed (a) depth map, after being processed by an SR filter on the image (b), shows better clarity and definition of depth. The same goes for image (d), where bilateral filtering is applied to single cost volume slice (c) constructed from (a).

2.7.3 Face Hallucination

Face hallucination, also known as face SR, is the process of generating HR facial images from low-resolution inputs, assisting in face-related tasks, mainly facial recognition [10] and others such as surveillance and forensic analysis.

Approaches to face hallucination often leverage facial prior knowledge, including landmarks, parsing maps, and identities, to enhance the reconstruction process, as seen on [11].

2.7.4 Medical

Another application of deep learning-based image SR, the focus of this work, is medical imaging. In addition to the various medical imaging modalities interested in image SR techniques, including X-Ray imaging, endoscopy, and others, this work focuses explicitly on CT scans as its primary object of study, with magnetic resonance imaging (MRI) as the second focus.

Both CNN and GAN-based solutions have been investigated as tools for SR on CT scan data. As stated in [38], Umehara K. et al. [63] investigated the use of SRCNN on CT chest images, demonstrating its superiority over traditional linear interpolation methods. Park J. et al. [50] proposed a modified U-Net architecture for 2D brain CT images, integrating SR and denoising

functionalities. This modified U-Net architecture included additional batch normalisation, expediting convergence and preventing improper initialisation.

Mansoor et al. [43] explored SRGAN for CT and MRI imaging, using a network to abstract features and enhance similarity based on image characteristics instead of pixel-level details. They trained the network with 2D slices and generated LR images by down-sampling HR images smoothed with a Gaussian kernel.

A more recent investigation [75] introduced GAN-CIRCLE for CT SR, utilising a circular structure to ensure that the output image aligns with the input image distribution. They incorporated total variation regularisation into the loss function and assumed that feeding a HR image into the generator would yield a similar output, potentially enhancing network robustness, as seen in Figure 2.11.

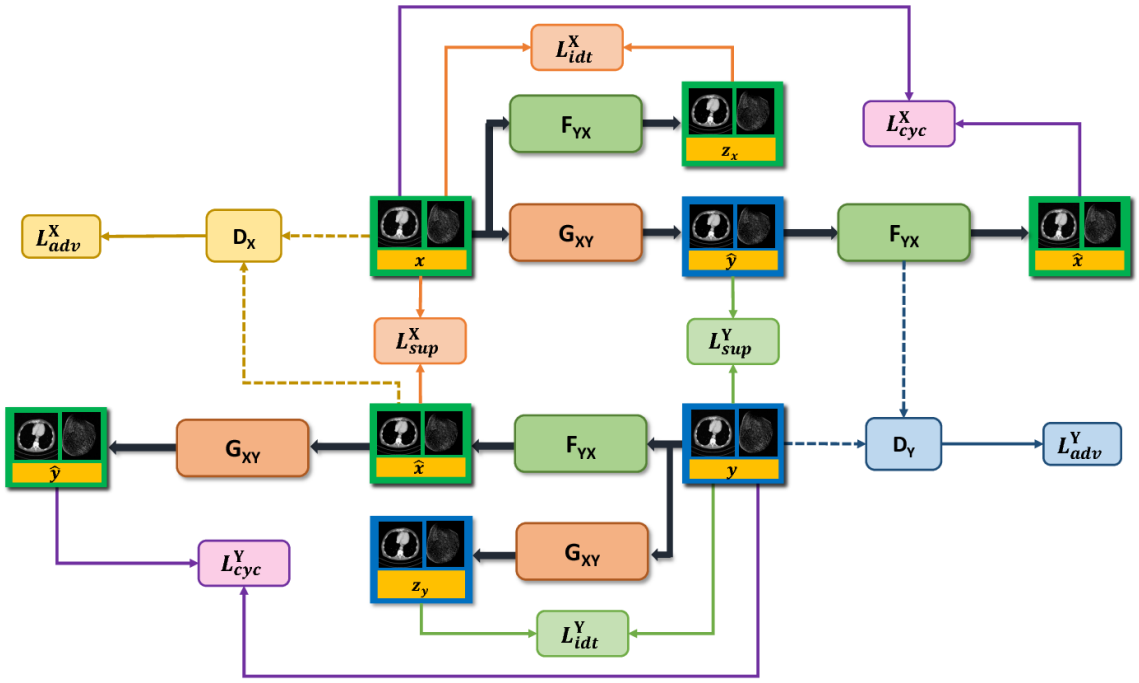


Figure 2.11: GAN cycle framework from [76].

2.8 Summary

In this chapter 2, the key concepts and recent advancements in image SR were reviewed, with a focus on their application in medical imaging. The basics and importance of SR in enhancing image quality were discussed, followed by an exploration of the two main approaches: traditional computer vision methods, such as interpolation, and modern deep learning methods, including CNNs and GANs.

The metrics used to evaluate SR models were also covered, specifically both objective and subjective methods. Then, the latest deep learning-based SR techniques were reviewed and examined.

Furthermore, various applications of SR were reviewed, including its use in videos, depth maps, facial images, and, notably, medical imaging, where it can significantly improve the quality of CT scans for better diagnostics.

In summary, this literature review highlights substantial progress in SR techniques, particularly through deep learning, and lays the groundwork for the development of a specialised SR model for enhancing synthetic CT scan data, which will be discussed in the following chapters.

Chapter 3

Data Preparation

The work started by evaluating the state-of-the-art image SR techniques and deciding on which one to use as a basis for the rest of the project. GAN-based models have recently gained popularity in medical SR research due to their superior ability to recover high-frequency details [73].

After a thorough analysis, which involved examining various GAN architecture types, including ESRGAN [66], it was decided to proceed with Real-ESRGAN [67]. Real-ESRGAN represents an advancement over ESRGAN, incorporating enhanced training strategies, refined network architectures, and more effective perceptual loss functions. The choice was based on the recent advancements of the Real-ESRGAN architecture and comprehensive evaluations of both visual and metric results of the selected model.

3.1 Datasets

Various datasets containing Digital Imaging and Communications in Medicine (DICOM) [31] files were pre-selected and analysed for data collection. DICOM files are commonly used in the medical field to store medical images such as X-rays and CT scans along with the associated patient information and medical data. In the context of this project, only the images within the DICOM files were employed.

3.1.1 RIDER

The project was decided to start solely with the RIDER (Reference Image Database to Evaluate Therapy Response) dataset [53]. This dataset consists of a collection of medical image files (DICOM) intended to evaluate tumours' response to therapy. This specific version contains about 15.000 images of CT scans of 32 different patients focused on the lung. Additionally, breast imagery is included in the dataset.

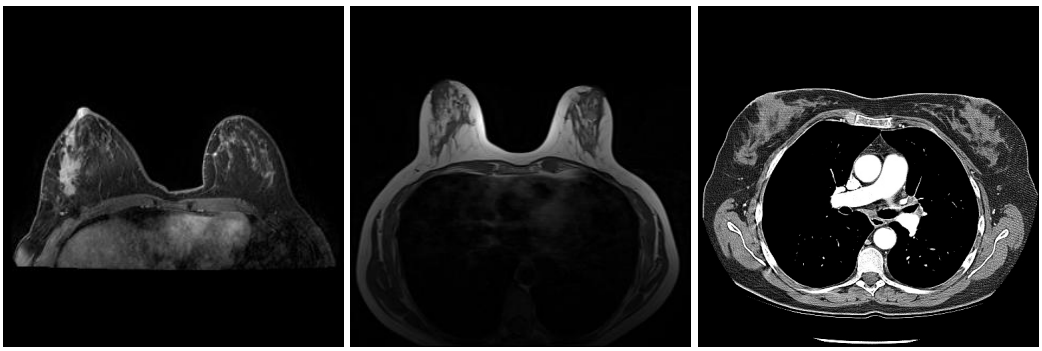
The rationale behind this choice is that the RIDER dataset is relatively small. Therefore, it is a helpful starting point for developing and refining the pre-processing pipeline.

3.1.2 Duke

The utilised Duke dataset on this project consists of a subset of the original Duke Breast MRI dataset [16]. The available subset of this dataset contains approximately 50.000 DICOM MRI files from 992 patients intended to evaluate the extent of disease in breast cancer patients.

3.1.3 LIDC

To ensure a comprehensive amount of thoracic medical imagery, the LIDC (Lung Image Database Consortium) [3] dataset was also incorporated into the project. The LIDC image collection comprises diagnostic and lung cancer screening thoracic CT scans with marked-up annotated lesions. This collection includes imagery from 1010 different subjects and approximately 450.000 DICOM files.



RIDER example Duke example LIDC example

Figure 3.1: Exemplary images retrieved from each one of the datasets.

3.2 Preprocessing pipeline

3.2.1 DICOM to JPEG

The first processing step was to convert all DICOM files into a usable image store format, Joint Photographic Experts Group (JPEG). To do this, a Python script was developed based on the Pydicom library [1].

Given the relatively large amount of data and the inconsistent and irrelevant original directory structure of the datasets in the context of this project, it was also necessary to develop a Python script to organise the datasets in a structured and well-distributed manner. This script ensured that the data was systematically arranged, making the access and analysis of data more coherent and well-structured.

3.2.2 Data correction

After converting the images, an analysis revealed that some images' contrast was sub-optimal. To address this issue, a Python script was employed to balance the contrast and brightness of the medical images. This is achieved by manipulating the window level and window width, which adjust the brightness and contrast of the medical imagery, respectively, enhancing the visibility of specific structures [45] and making the final conversion appear like a normal CT scan or MRI.

Despite these adjustments, it was observed that some images remained utterly black. This is probably due to these images corresponding to the upper or lower slices of the medical scans, where the scanning device had not yet begun capturing the relevant parts of the human body. To resolve this issue, a new Python script was employed that removed all entirely black images from the dataset within a certain degree of sensitivity. This ensured that only images with no discernible anatomical structures were excluded, thereby preserving the integrity of the medical imagery dataset.

In the early stages of planning the work, it was hypothesised that cropping the medical imagery to remove all the black parts, leaving only the sections with actual information, might reduce the dataset size, improve training times, and potentially result in more accurate models. A Python script was created to test this hypothesis and perform this cropping operation on the images. This approach was tested in parallel with the normal procedure to compare the results and determine the effectiveness of the cropping strategy. However, upon implementation, it became evident that this approach did not yield favourable results. Detailed analysis and the outcomes of this technique are discussed in Chapter 3.

3.2.3 Data preparation

To prepare the whole dataset for the training phase, two separate steps were required:

- Image Degradation Process - Some training methods within the Real-ESRGAN development framework automatically generate the low-quality and LR counterparts of the ground truth data needed for the training procedure. Nevertheless, Real-ESRGAN also allows users to create low-quality versions of the ground truth data. Initially, the only transformation applied was downscaling the JPEG files to 1/4th of the original resolution to serve as LR inputs for the image SR model. This process involves reducing the dimensions of each image to 25% of its original width and height, which will create a set of LR images that can be used to train and evaluate the performance of the models by using their upscaled versions to be compared with ground truth data. Eventually, blurring was also added to the image degradation process.

- Text File Generation - Real-ESRGAN's development environment requires a .txt file with the paths to every image in the dataset. This is how the training data is prepared for input into the model. This file should then be split into a training/validation file and a testing file, each with the paths of the images assigned to each function.

3.3 Summary

Chapter 3 began by selecting Real-ESRGAN as the preferred SR technique, following a comprehensive evaluation of various GAN-based SR models.

The data preparation phase was then detailed, focusing on key datasets: RIDER, Duke, and LIDC. These datasets, containing medical images in DICOM format, were converted to JPEG and organised using Python scripts. Data correction steps included adjusting contrast and brightness and removing irrelevant black images. An initial hypothesis to crop black areas from images was tested but did not yield favourable results.

Preparation steps involved generating LR images through downscaling (and blurring eventually) and creating text files to organise image paths for training. These processes established a solid foundation for the training and evaluation of the SR model.

Chapter 4

Implementation

4.1 Real-ESRGAN Network Architecture

4.1.1 Image Degradation

The image degradation process is one of the significant improvements of Real-ESRGAN. The Real-ESRGAN image degradation process creates low-quality, LR versions of the ground truth data that will posteriorly be fed to the model during the training phase.

This degradation process is innovative since it introduces a whole variety of image degradation, including blur, noise (additive Gaussian and Poisson), resize (downsampling and upsampling), and JPEG compression [67]. This technique not only simulates real-world image degradation but is also useful for artefacted and noisy images, similar to the kind of degradation found in synthetic medical imagery, which is the final objective of this project.

Additionally, Real-ESRGAN [67] extends the classical degradation model to a high-order degradation process, incorporating repeated degradation processes with varying hyperparameters to model more complex real-world degradation scenarios, as seen in Figure 4.1. Furthermore, sinc filters are utilised to synthesise ringing and overshoot artefacts commonly observed in images subjected to sharpness enhancement and JPEG compression.

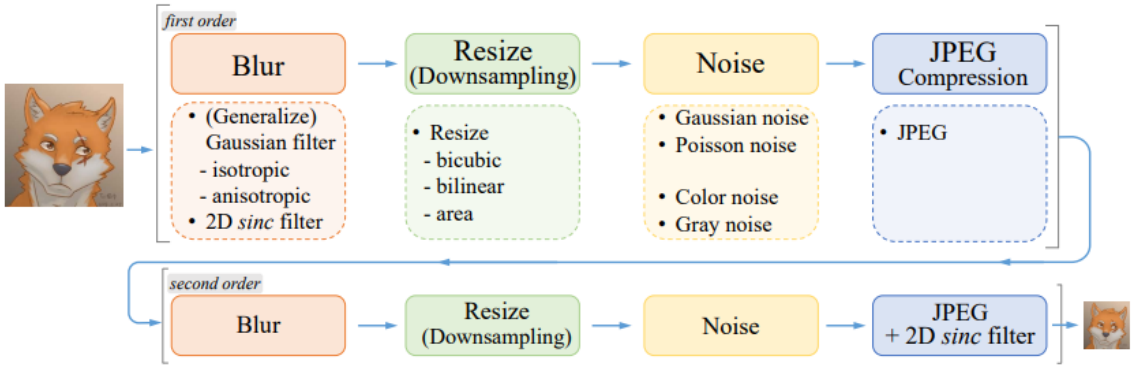


Figure 4.1: Real-ESRGAN's image degradation process steps, from [67].

4.1.2 Generator and Discriminator Network

As per Wang et al. [67], Real-ESRGAN uses the same generator network as in ESRGAN [66], which utilises the RRDB architecture that consists of densely connected residual blocks with residual connections within each block, enabling effective learning of intricate image features. As depicted on Figure 4.2, the generator also employs a pixel-unshuffle operation that rearranges pixel values in an image tensor to convert it from a lower-resolution image with a higher number of channels to a higher-resolution image with fewer channels, aiding in the extraction of spatial information for subsequent processing in SR tasks.

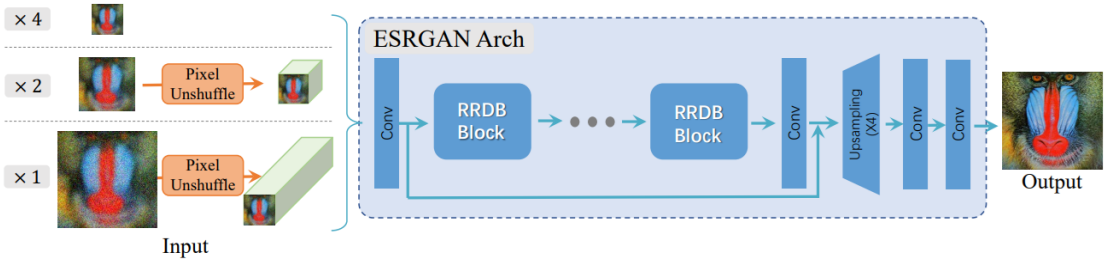


Figure 4.2: Real-ESRGAN's Generator Network Architecture, from [67].

Regarding the discriminator network of Real-ESRGAN, a U-Net [54] design was adopted. A U-Net discriminator (see Figure 4.3) is a neural network architecture inspired by the U-Net model, consisting of contracting and expanding paths with skip connections. It enables per-pixel realness value outputs, providing detailed feedback to the generator.

Spectral normalisation addresses training instability arising from this enhanced structure and complex degradation scenarios. Spectral normalisation constrains the spectral norm of weight matrices in neural networks by dividing each weight matrix by its largest singular value, ensuring smoother optimisation and mitigating training instability. This helps to mitigate oversharp and other undesirable artefacts introduced during GAN training [67].

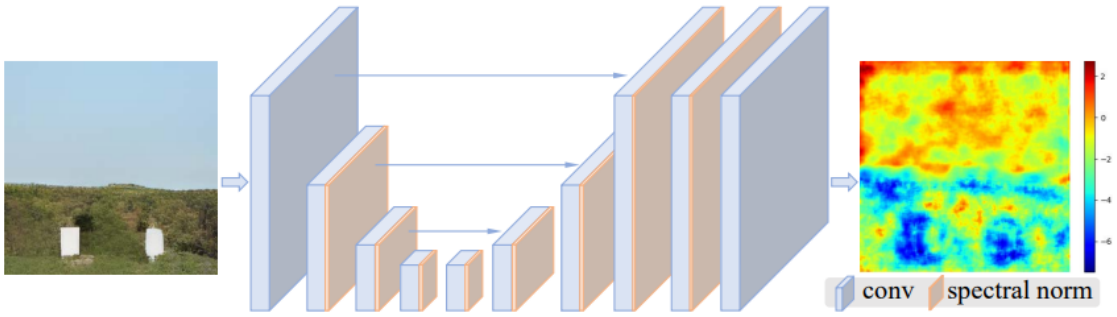


Figure 4.3: Real-ESRGAN's U-Net Discriminator Network Architecture with Spectral normalisation, from [67].

4.1.3 Loss Functions

Real-ESRGAN [67] uses three different techniques as loss function, namely L1 loss, also known as MAE (see Equation 2.8), perceptual loss (see Equation 2.11), and GAN loss 2.5.4.3, with assigned weights of 1, 1, 0.1, respectively.

Regarding the perceptual loss function, specific weights are assigned to feature maps extracted from different layers of the VGG19 network before activation. These weights, namely 0.1, 0.1, 1, 1, 1, determine the relative importance of features from each layer in the perceptual loss calculation, emphasising deeper layers for capturing more abstract image characteristics. By prioritising deeper layers that encode semantic information, Real-ESRGAN ensures that the perceptual loss focuses on preserving overall image structure and content for perceptually faithful results.

4.2 Hardware

All the training and processing required for this project was run on a SLURM [74] platform, an open-source workload manager for efficiently scheduling jobs and managing resources in high-performance computing clusters. Training required at least about 14 GB of available VRAM, so training jobs were run on 24 GB partitions (NVIDIA Quadro RTX 6000), 32GB partitions (NVIDIA Tesla V100), and sometimes on 80 GB partitions (NVIDIA A100).

4.3 Libraries

In addition to the aforementioned libraries, the training process also leveraged PyTorch [52] for its specific functionalities while employing OS [19] and Sys [20] for file management tasks. CV2 (OpenCV) [8] was utilised for handling image files alongside NumPy [25] for efficient numerical operations. Python Imaging Library (PIL) [41] was employed for image manipulation tasks.

4.4 Training Procedures

Real-ESRGAN's [67] development environment provides two main techniques for the training procedure.

4.4.1 Training From Scratch

Training from scratch involves a comprehensive process that does not rely on pre-trained Real-ESRGAN models, ensuring that the model learns almost entirely from the provided dataset. The training process is divided into two main stages. Both stages share the same data synthesis process and training pipeline, differing only in the loss functions used.

Real-ESRNet model is trained in the first stage using the L1 loss function. This phase lays the foundation for the final Real-ESRGAN model.

Moving on to the second stage, the pre-trained Real-ESRNet [67] model serves as the starting point for initialising the generator network of the Real-ESRGAN model. During this phase, the model undergoes finetuning, integrating a blend of loss functions, including L1 loss, perceptual loss, and GAN loss, as mentioned before.

4.4.2 Finetuning

The finetuning training method leverages the use of a pre-trained Real-ESRNet generator and a Real-ESRGAN discriminator, which have been trained on synthetic data, to initialise the generator network.

Finetuning can be performed using Real-ESRGAN's new image degradation process, which simulates low-quality images to train the model. Alternatively, and as said before, it's also possible to use self-sourced low-quality paired data, where ground truth images are paired with user-created low-quality versions.

It's important to note that Real-ESRGAN's development environment for finetuning paired data only supports paired data with 0.25x the resolution of the ground truth image.

4.4.3 Validation

Real-ESRGAN's development environment also supports validation setup. Validation is performed during training, allowing the model to adjust its hyper-parameters and improve performance based on the feedback from validation data. This process helps ensure the model generalises well to unseen data and prevents overfitting.

Choosing an appropriate metric to evaluate something as complex as image SR was challenging. Given the intricacies involved, it was decided to use a straightforward metric that focuses on assessing overall image quality. Consequently, PSNR [27] was selected for validation, providing a clearer and more quantifiable measure of the reconstructed image quality.

4.4.4 Training Settings

Regarding the training settings of all the developed models, 5 workers per GPU and a batch size of 12 were used. Furthermore, training states are regularly saved, making it possible to recover progress in case something goes wrong during the execution of the training process. The training process consisted of 400.000 iterations, which ensured a reasonable training duration without significantly compromising the resulting model's performance.

4.5 Testing Procedure

To conduct the testing procedure, a Python script was developed to upsample the created LR and low-quality versions of ground truth images using the tested developed SR model.

4.5.1 Metrics

As for metrics, a Python script was created to compare the ground truth images with the ones upsampled using the developed model. To cover and assess as many aspects as possible, the following set of metrics was implemented:

- SSIM [27];
- PSNR [27];
- MMD [7];
- LPIPS [78].

In image SR, high quantitative metrics like PSNR and SSIM often don't always align with human-perceived quality [5], as these metrics may not fully capture visual details, textures, and naturalness that humans find essential. This way, it's relevant to include subjective human evaluation of the results to ensure that the enhanced images meet the desired visual standards and preferences.

4.5.2 Statistics

So as to perform a complete analysis, the following statistical data points were extracted:

- Average;
- Quartiles (0.25, 0.5, 0.75);
- Minimum Value;
- Maximum Value;
- Standard Deviation.

To keep the main chapters more concise, only the average values of the metrics are present in the tables. The Appendix A includes all extensive statistical tables.

4.5.3 Pre-Trained Real-ESRGAN Model and Bicubic Interpolation

To establish a baseline for comparison with the models being developed, the outputs of the test image data were also generated using a pre-trained Real-ESRGAN model, which focuses primarily on real-life images, as well as a simpler model based on bicubic interpolation for all variations of models based on different datasets.

4.6 Summary

Chapter 4 covers all the technical parts required to implement and start the model development. It starts by exploring Real-ESRGAN's network architecture, focusing on its image degradation process, the design of generator and discriminator networks and the implemented loss functions.

The chapter also covers the hardware setup used for computational tasks and highlights the software libraries that were utilised. Real-ESRGAN's training procedures are discussed in detail, including starting from scratch, finetuning methods, validation approaches, and specific settings for training. Metrics and statistical methods employed to assess model performance are also covered.

Chapter 5

Experimental Results

5.1 RIDER

In order to finetune the pre-processing pipeline and get accustomed to Real-ESRGAN's [67] development environment, the project started by developing an SR model based on training using the RIDER [53] dataset. This choice was made due to the relatively small size of the RIDER dataset, which facilitated debugging processes, resulted in lower training times and allowed for more effortless method refinement.

5.1.1 Results

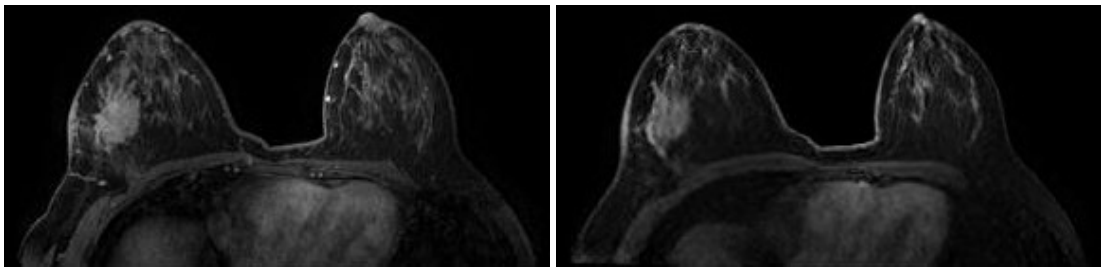
In this initial phase, as the main objective was to recognise and understand the general aspects of the project, without the need for an extensive analysis of the results, only two of the fundamental metrics were implemented: SSIM and PSNR (see Table 5.1), disregarding by now the pre-trained Real-ESRGAN model and bicubic interpolation model output results.

Model/Metrics' Averages	SSIM	PSNR
RIDER Finetuned	0.83	28.62
RIDER Finetuned Paired	**0.88**	**32.12**
RIDER From Scratch	**0.86**	**30.40**

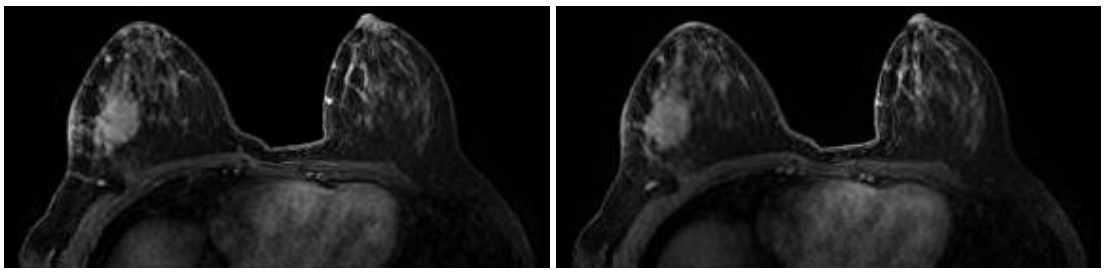
Table 5.1: RIDER metrics' results.

Despite this initial phase not requiring extensive analysis, it is already clear that, given the overall good results (see Figure 5.1), Real-ESRGAN seems to be a good choice for the rest of the project's development.

It's also important to note that the good results obtained with this model could be partly attributed to the relatively LR of the RIDER dataset imagery. The lower resolution might have made it easier for the model to achieve higher similarity and fidelity scores, as measured by SSIM and PSNR.



Ground Truth RIDER Finetuned



RIDER Finetuned Paired RIDER From Scratch

Figure 5.1: Real-ESRGAN's RIDER implementation.

Model/Metrics' Results	SSIM	PSNR
Finetuned	0.84	28.25
Finetuned Paired	**0.89**	**31.01**
From Scratch	**0.88**	**30.26**

Table 5.2: Metric results of the displayed RIDER images in Figure 5.1.

At this starting point, it's already clear that both finetuned paired and from scratch models output the best results (see Table 5.2), maintaining most of the details and features of the ground truth image when compared to the finetuned model.

5.2 Duke

The resulting models from training REAL-ESRGAN with the Duke dataset provided the first insights into the actual use cases and the finality of this project, owing to the dataset's larger size and higher image resolution and quality.

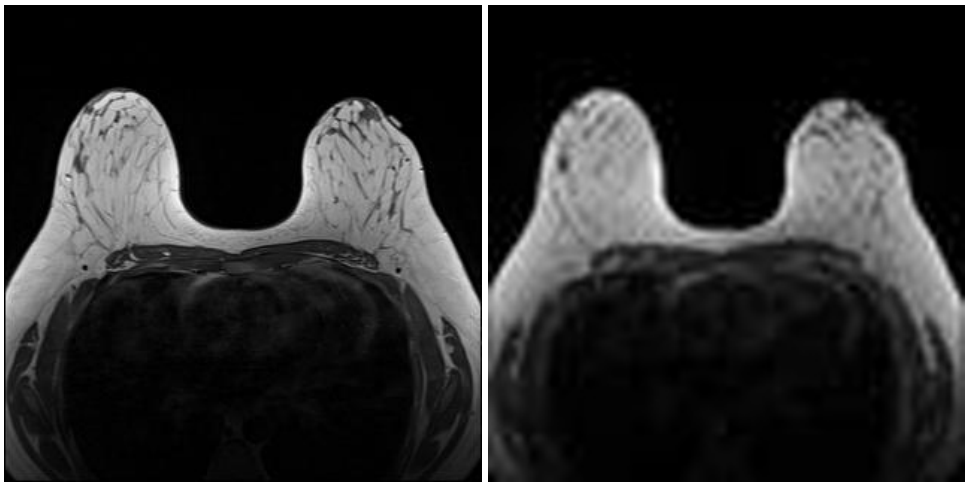
5.2.1 Results

By examining the metric results (see Table 5.3), the standout aspect is the notably high scores achieved by bicubic interpolation upscaling in structural similarity metrics such as SSIM, PSNR, and MMD. These metrics indicate a close match between the interpolated images and the ground

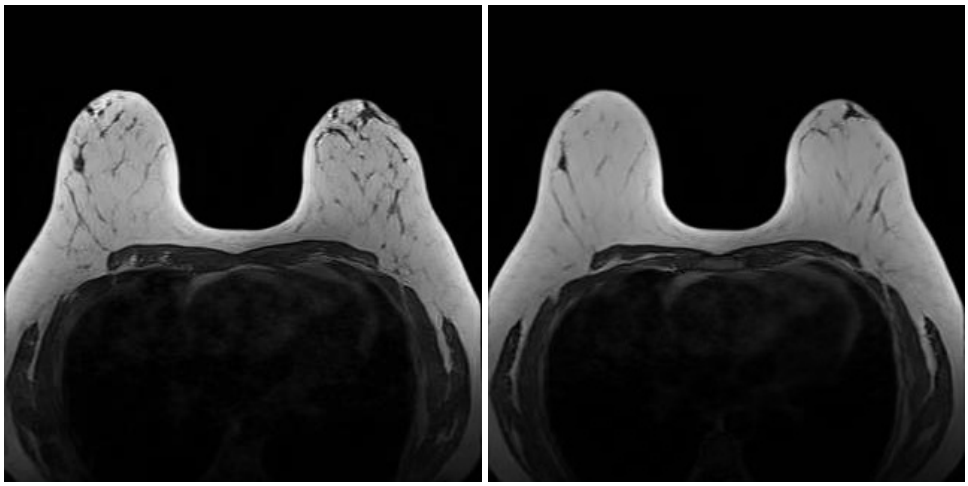
truth regarding structural similarity and noise reduction. However, this correspondence does not align with our humane perception, given the clear pixelation and low detail level (see Figure 5.2).

Model/Metrics' Averages	SSIM	PSNR	MMD	LPIPS
Pre-Trained Real-ESRGAN	0.76	29.62	77.65	0.33
Bicubic Interpolation	**0.86**	**32.21**	**39.40**	**0.27**
Duke From Scratch	**0.81**	**30.70**	**59.69**	**0.23**
Duke Finetuned	0.78	30.05	67.86	0.26
Duke Finetuned Paired	**0.82**	**31.99**	**43.10**	**0.21**

Table 5.3: Duke metrics' results.



Ground Truth Bicubic Interpolation



Duke Finetuned Paired Duke From Scratch

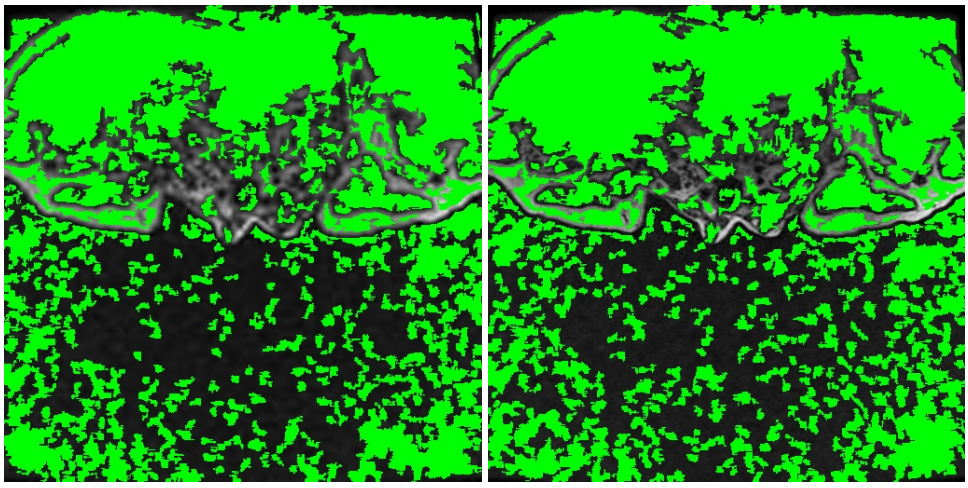
Figure 5.2: Comparison of the output of Duke-based deep learning models with bicubic interpolation and ground truth as a baseline.

This comparison showcases the significant differences in obtained results between the models. Compared to the other results, bicubic interpolation's outputs internal anatomical structures of the breast lack detail and quality. However, observing their presence is still possible, which would explain the result in structure similarity metrics. Despite this, the bicubic interpolation result visually does not meet the desired quality and detail definition, which is best showcased by the LPIPS metric (see Table 5.4). LPIPS, by comparing high-level features extracted from deep neural networks, better evaluates human perception, favouring the models that maintain anatomical structures and details closest to the ground truth.

Model/Metrics' Results	SSIM	PSNR	MMD	LPIPS
Bicubic Interpolation	0.81	27.50	60.77	0.26
Finetuned Paired	0.79	27.53	60.15	0.14
From Scratch	0.76	27.00	68.15	0.15

Table 5.4: Displayed Duke image metrics results.

To better understand the discrepancy between discrete metric results and human perception, a mask based on the SSIM metric was generated (see Figure 5.3). This mask highlights areas of the image where the structural similarity between the image and the ground truth is notable. In this mask, areas with more green indicate more significant structural differences.



Bicubic Interpolation mask Real-ESRGAN model mask

Figure 5.3: Structural similarity mask comparison between a bicubic interpolation model and a Real-ESRGAN based deep learning model.

This analysis was conducted with the expectation that the black areas of the images might have been degrading the structural similarity metric. However, the results indicate otherwise. Both the bicubically interpolated and deep learning-generated images exhibit structural similarity that is dispersed similarly on both masks, with a greater focus on the regions containing the actual medical scan.

Moreover, by analysing the metrics results and evaluating the output of the finetune training method, it's apparent that it consistently proves to be less helpful for this use case, while the finetuned paired approach performs the best. In the following example of an output of the finetuned model (see Figure 5.4), the lack of detail of the internal anatomical structures and overall "over smoothness" is clear compared to the other models' outputs.

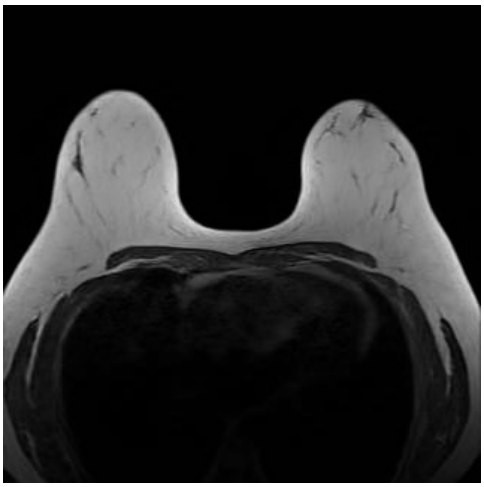


Figure 5.4: Duke-based Real-ESRGAN finetuning training method output.

One possible explanation is that since testing is conducted on downsampled versions of the ground truth, the paired data training method, which is itself only downsampled versions of the ground truth, benefits from this alignment. Nonetheless, these results remain representative because the use case of this project, synthetic generated medical data, does not suffer from some of the data degradation processes applied by the image degradation process of Real-ESRGAN, such as compression and colour deviation. Instead, it primarily suffers from LR and blur, making the finetuned paired approach more suitable for this specific application.

These results also suggest that this specific dataset training benefits the project, given that the pre-trained Real-ESRGAN model scores lower on all metrics overall, which is visually confirmed by the generally lower level of detail of the anatomical structures of the medical imagery.

5.3 RIDER + Duke

In order to evaluate the benefits of using more than one dataset on the same training procedure, the RIDER and Duke datasets were combined to create a dataset with increased image diversity.

5.3.1 Results

The outcomes yielded results (see Table 5.5) that were not vastly different but marginally more positive when incorporating both the Duke and RIDER datasets, compared to utilising only the Duke dataset. This slight improvement stands as a positive indication, as mixing more diverse images could have made the model less accurate, given the risk of it becoming less specialised in

a singular dataset, which could have resulted in worse outcomes. The marginal variance observed in the metrics' results could also be attributed to the relatively limited size of the RIDER dataset. When combined with the DUKE dataset, the RIDER data set holds minimal significance in the training process.

Model/Metrics' Averages	SSIM	PSNR	MMD	LPIPS
Pre-Trained Real-ESRGAN	0.77	29.56	77.15	0.33
Bicubic Interpolation	0.86	32.23	38.67	0.27
Duke + RIDER From Scratch	**0.81**	**30.75**	**57.20**	**0.23**
Duke + RIDER Finetuned	0.77	29.73	72.75	0.27
Duke + RIDER Finetuned Paired	**0.81**	**31.81**	**44.53**	**0.21**

Table 5.5: Duke + RIDER metrics' results

When comparing the results of the different models, the bicubic interpolation model still yielded the best overall results. By looking at the statistics in the appendix A.7, specifically at standard deviation and quartiles, the bicubic interpolation model still reveals a consistent performance across images, particularly evident in SSIM and MMD metrics.

By looking at the Real-ESRGAN's models results, it's notable a commendable result of the PSNR metric and still good results with LPIPS, specifically with the finetuned paired and from scratch models, which reaffirms the previous presumption that LPIPS is the metric in this set that better resonates with human perception, and the one that better maintains anatomical and textural detail. As we can see in the following subset of images (see Figure 5.5), bicubic interpolation visual results lack clarity and detail despite the high metric scores. Also, the textures are very smooth, indicating a high blur level and loss of fine details.

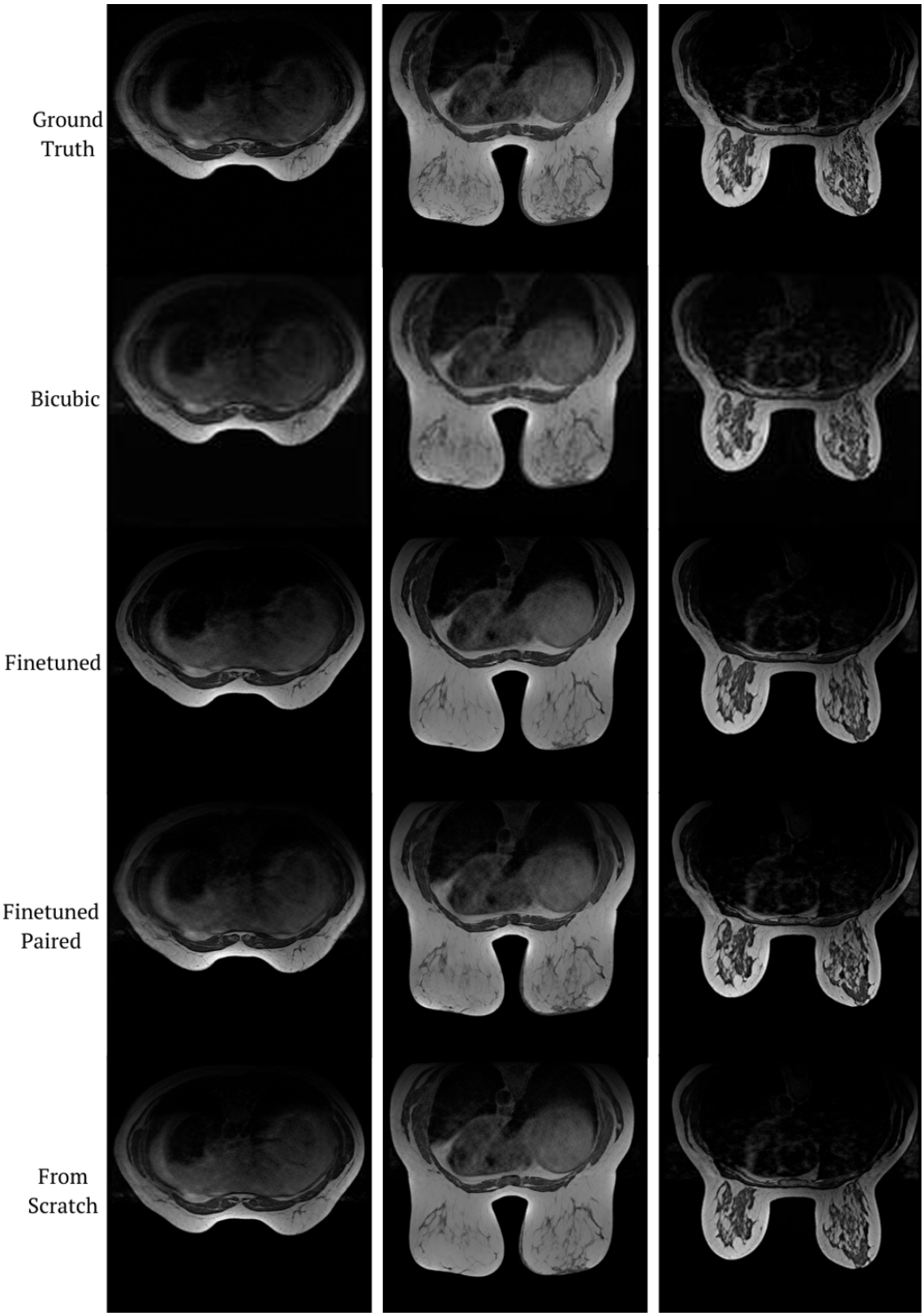


Figure 5.5: Duke and RIDER-based Real-ESRGAN models' results.

Analysing the rest of the images, it's still clear that anatomical details appear more present and refined in the Real-ESRGAN's models, especially in the finetuning paired but also in the from scratch variants. Despite being clearly more detailed than the bicubically interpolated image, the finetuned output still lacks crispness and clarity. Furthermore, the images generated through both finetuned paired and from scratch models exhibit enhanced visibility and detail compared to the output from the finetuned model alone. However, the advantage of the finetuned paired model output is not entirely clear, as it may prioritise slight increases in detail at the expense of texture, resulting in a slightly noisier output, as seen on Figure 5.6.

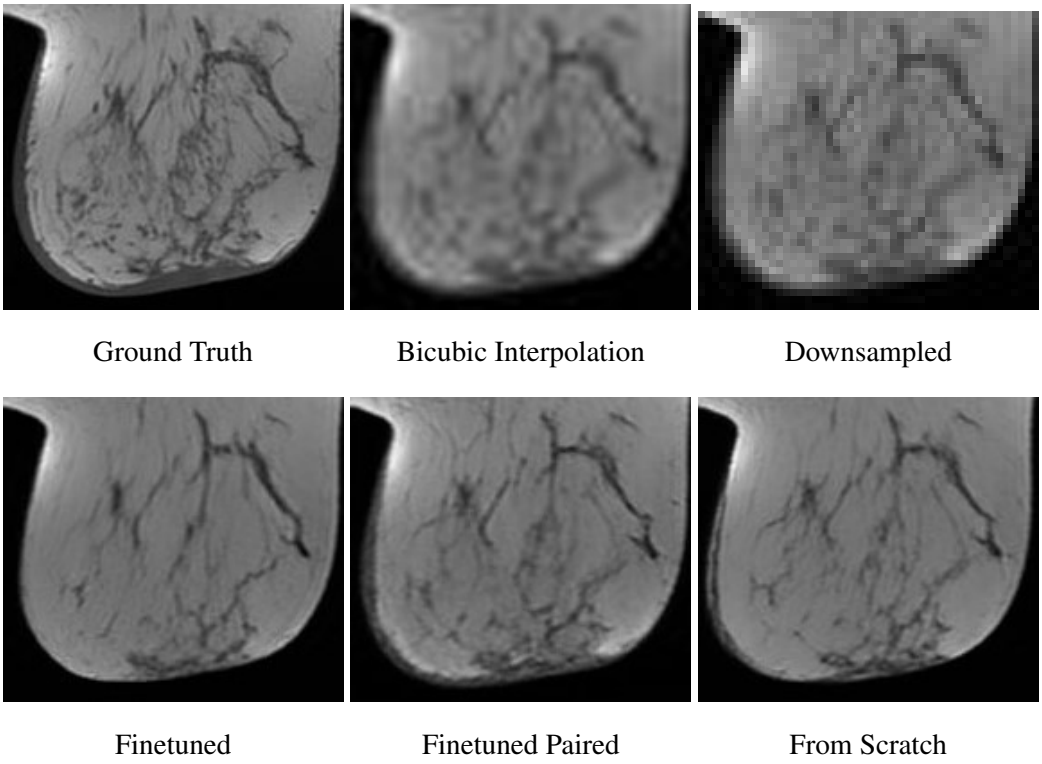


Figure 5.6: Comparison of the output of Duke and RIDER-based deep learning models with 4x downsampled version of the ground truth, bicubic interpolation and ground truth as a baseline.

Based on all the conducted analyses, both the finetuned paired and from scratch models emerge as highly useful in the context of this project. The finetuned paired models are the most competent, likely due to their use of downsampling as the image degradation method. This method not only better simulates the low quality and resolution of synthetic medical imagery, the project's primary focus, but also aligns more closely with the image degradation found in the testing data. On the other hand, from-scratch models also prove beneficial, likely owing to their requirement for tailored training of both discriminator and generator, without entirely relying on pre-trained models derived from real-life imagery. This aspect is particularly pertinent, as utilising real-life images could confuse the model, considering that the project's context revolves solely around medical imagery.

Moreover, it's noteworthy to mention that the finetuned paired models excel in preserving details, whereas from scratch models demonstrate superiority in capturing texture and contrast.

5.3.2 Cropped Images

As mentioned before, it was initially hypothesised that cropping the image and removing the black parts of the medical imagery could result in a smaller image size, improving training times, given the relatively smaller size of every image. It could also potentially improve the developed models' accuracy by focusing the training on the relevant parts of the medical scan where the actual information is located.

After passing the Duke dataset through the cropping pipeline, the resulting dataset size did not decrease significantly, going from about 900 MB to 650 MB. This is likely because the cropped-out portions were predominantly black and did not contain much information, making their exclusion insignificant regarding overall data volume.

Despite this, training was conducted using only cropped images of the Duke dataset to evaluate this change's effect on the final models' metric results. In order to have a baseline for comparison, the same crop was applied to ground truth images, and a testing subset of Duke and RIDER models trained with uncropped data, and the metrics results were also generated (see Table 5.6).

5.3.2.1 Results

The comparative analysis of the results (see Table 5.6) reveals that the metrics for both the original and cropped image datasets remained virtually unchanged. This strongly suggests that cropping the images used in the training process adds no value to the final result and accuracy of the developed model.

Model/Metrics' Results	SSIM	PSNR	MMD	LPIPS
Duke Finetuned	0.71	28.07	72.66	0.16
Duke Finetuned Cropped	**0.71**	**28.07**	**72.66**	**0.16**
Duke Finetuned Paired	0.77	30.18	44.31	0.11
Duke Finetuned Paired Cropped	**0.77**	**30.17**	**44.44**	**0.11**
Duke From Scratch	0.76	29.01	57.13	0.12
Duke From Scratch Cropped	**0.76**	**29.01**	**57.13**	**0.12**

Table 5.6: Duke + RIDER cropped metrics' results

Considering these results, it was concluded that the cropping technique was not beneficial, so it was discarded.

5.3.3 Combined Training Methods

The following solution emerged to benefit from each training method's strengths: initiating with the from scratch approach and refining the resultant generator and discriminator through finetuning

using the paired approach with the same dataset. This method benefits from the diversity of image degradation processes and the extensive training without pre-trained data.

5.3.3.1 Results

Based on the provided results (see Table 5.7), it's evident that this model outperforms finetuned paired and from scratch models in several key metrics. Firstly, the SSIM metric average results are the highest observed amongst the deep learning models, and the standard deviation, as seen on A.11, is also one of the lowest, suggesting that the model's performance is not only superior but also highly consistent across different samples.

Model/Metrics' Averages	SSIM	PSNR	MMD	LPIPS
Duke + RIDER Combined	**0.82**	**32.00**	**42.21**	**0.20**
Pre-Trained Real-ESRGAN	0.77	29.56	77.15	0.33
Bicubic Interpolation	0.86	32.23	38.67	0.27
Duke + RIDER From Scratch	**0.81**	**30.75**	**57.20**	**0.23**
Duke + RIDER Finetuned	0.77	29.73	72.75	0.27
Duke + RIDER Finetuned Paired	**0.81**	**31.81**	**44.53**	**0.21**

Table 5.7: Duke + RIDER metrics' results

In addition to SSIM, the models also excel in LPIPS, the only metric that has shown to be the most trustworthy at evaluating the results, by achieving the best average score in this metric until this point. Besides, the standard deviation for LPIPS is respectable, ranking amongst one of the lowest. PSNR and MMD also have the best average scores when comparing the models, though bicubic interpolation scores better in these metrics.

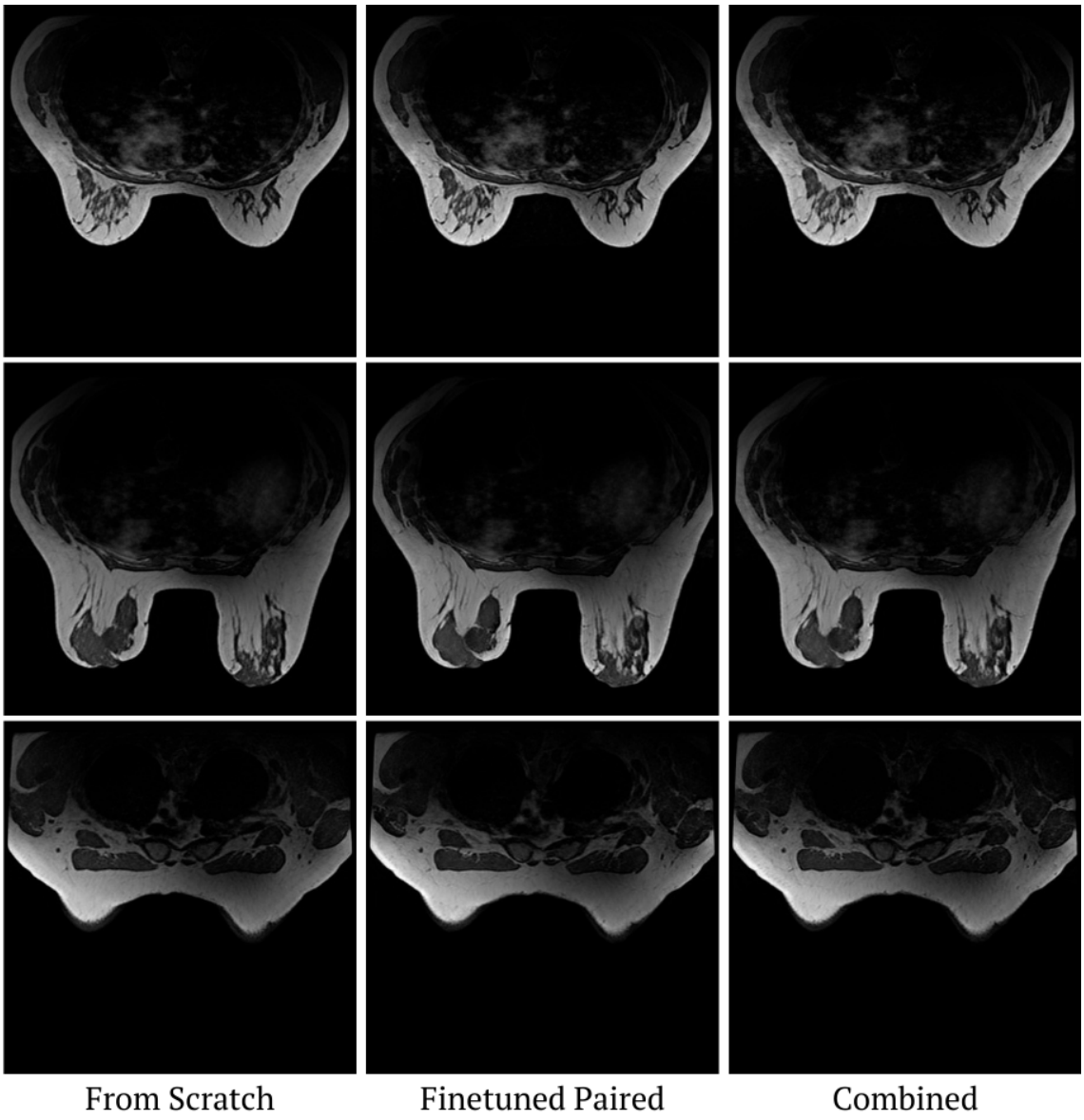
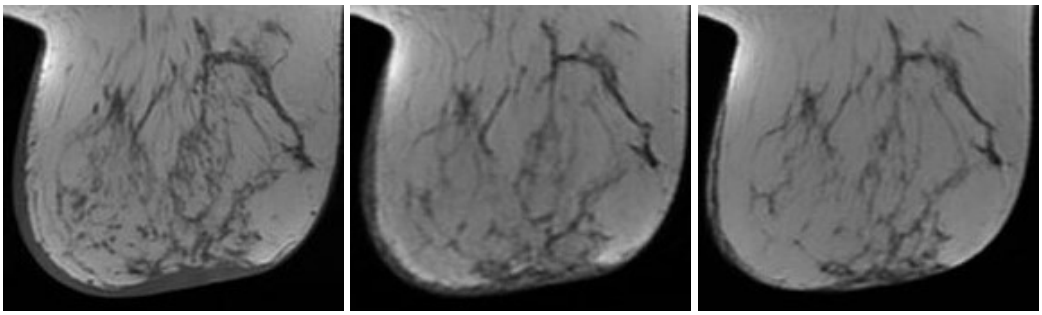
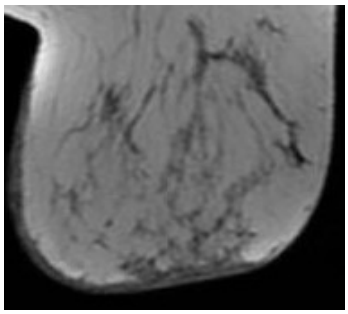


Figure 5.7: Duke and RIDER based Real-ESRGAN models' results comparison.

At this level of similarity, it becomes quite challenging to visually distinguish the results produced by each model from one another and determine which one looks better (see Figure 5.7). The differences between the images are subtle and not easily noticeable without close inspection. Despite this, by comparing specific parts of the resulting images to the ground truth images, we can still identify minor differences that set each model apart (see Figure 5.8).



Ground Truth Finetuned Paired From Scratch



Combined

Figure 5.8: Comparison of the output of Duke and RIDER-based deep learning models with ground truth as a baseline.

As mentioned earlier, the differences between these results are not very clear to the human eye. The most noticeable distinction is that the model trained from scratch produces images with slightly less detail than the other two models.

Among the other models, the combined approach adds more anatomical detail to the image than the finetuned paired model. It is important to note that this is an extreme case where many fine details are concentrated in a small part of the image, making it particularly hard for the models to generate these details accurately.

After considering all these observations and considering the improvement in the metric scores, it becomes clear that the models benefit from the introduction of new datasets. These datasets not only create more diversity but also increase the volume of images available for training.

Moreover, the combined training method shows substantial advantages, evident from the metrics results. SSIM, PSNR and MMD, despite giving high scores even for basic methods like bicubic interpolation, have shown to be effective at the performance of deep learning models, and, as said before, the combined training method obtained the best results. Furthermore, the LPIPS score, which is particularly sensitive to perceptual differences, supports this conclusion.

These scores, combined with the visual analysis of the resulting images, underscore that the models benefit not only from the increased diversity and volume of training data but also from the combined training method itself.

5.4 Rider + Duke + LIDC

So as to increase the volume of images for training while also incorporating greater image diversity, the LIDC dataset was introduced. This dataset significantly balanced the overall dataset by adding a substantial amount of lung medical data, complementing the already existing large volume of breast data. The model gains exposure to a broader variety of anatomical structures by including both lung and breast images. Only the combined version of paired finetuning and from-scratch training was developed, given that it has already been concluded that this method is the best approach to achieve a more accurate and efficient model.

5.4.1 Results

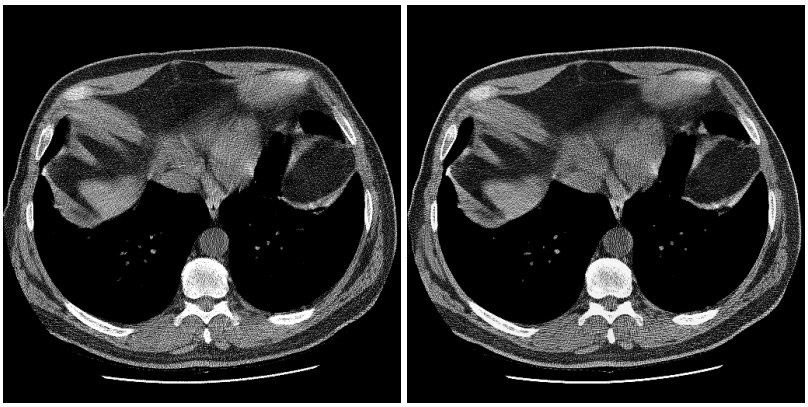
The obtained results (see Table 5.8) are strange and unexpected, considering the efforts and methodologies applied so far. The SSIM, PSNR, and MMD scores are surprisingly poor, while the LPIPS score is outstandingly good.

Model/Metrics' Averages	SSIM	PSNR	MMD	LPIPS
Duke + RIDER + LIDC Combined	0.71	24.19	1112.60	0.09

Table 5.8: Duke + RIDER + LIDC metrics' results

Upon examining the metric scores for each tested image individually, it became evident that the LIDC test images were responsible for poor SSIM, PSNR, and MMD performance.

A visual analysis of these images revealed the probable cause of the low scores: the LIDC dataset images come with an intended preset window width and window level values that sometimes result in noisier images where smooth textures are typically expected (see Figure 5.9).



Ground Truth Combined

Figure 5.9: Comparison between ground truth image and combined method Real-ESRGAN model image upscaled.

The dashed texture details present in the images contribute to the low scores in the referenced metrics, as it is exceedingly difficult for the model to replicate the exact noisy or dashed patterns.

To more accurately assess the model's performance beyond visual inspection, the metrics were recalculated using only the Duke and RIDER datasets for testing (see Table 5.9), excluding the LIDC dataset. By doing so, we can more reliably compare the results of the new model with those obtained previously, offering a clearer understanding of the model's true capabilities without the confounding influence of the LIDC dataset's preset window width and level values.

Model/Metrics' Averages	SSIM	PSNR	MMD	LPIPS
Duke + RIDER + LIDC Combined	0.84	31.82	43.91	0.18

Table 5.9: Duke + RIDER + LIDC metrics' results (Duke and RIDER testing subset)

These results are the best observed in most metrics overall, confirming that the training method and dataset expansion contributed significantly to developing a more capable and accurate model. As evidenced by the exemplary images (see Figure 5.10), the generated results are perceptually very similar to the ground truth images. This enhanced performance demonstrates that incorporating diverse datasets and employing the combined training method has effectively refined the model's ability to produce high-quality, detailed outputs that closely resemble the original images.

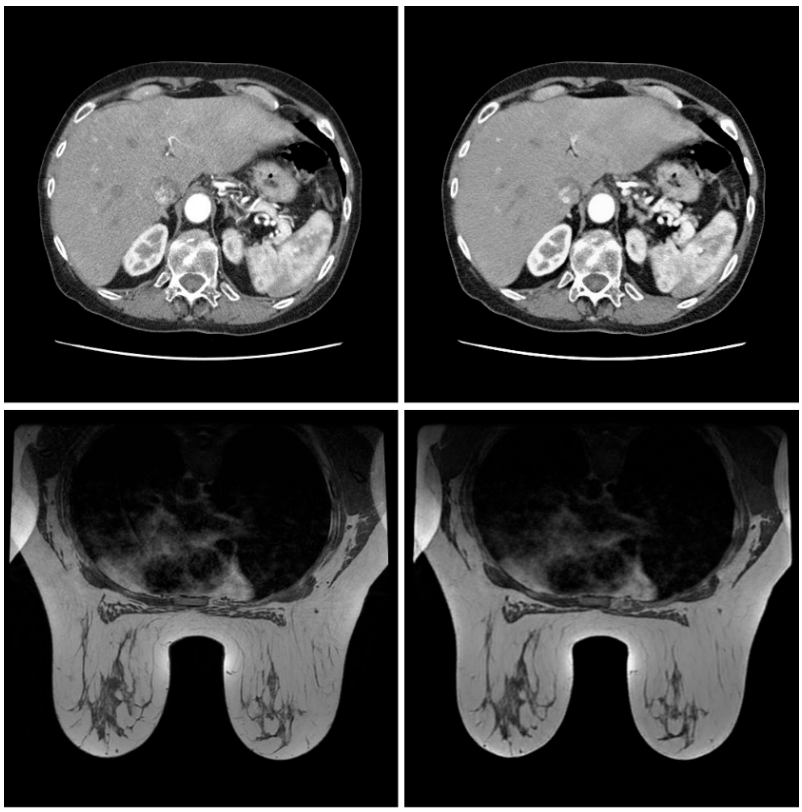


Figure 5.10: Duke, RIDER and LIDC based Real-ESRGAN model with combined training method results comparison with ground truth.

5.4.2 New Image Degradation Process

At this point in the project's development timeline, access was obtained for the first time to the actual synthetic images generated by the final models of the projects being developed in parallel. After upscaling some samples of the synthetic images through the latest developed model of this project, it was evident that, despite the added detail and improved quality, the images suffered from significant blurring as seen on (see Figure 5.11). This blurring was identified as also being present in the synthetic images.

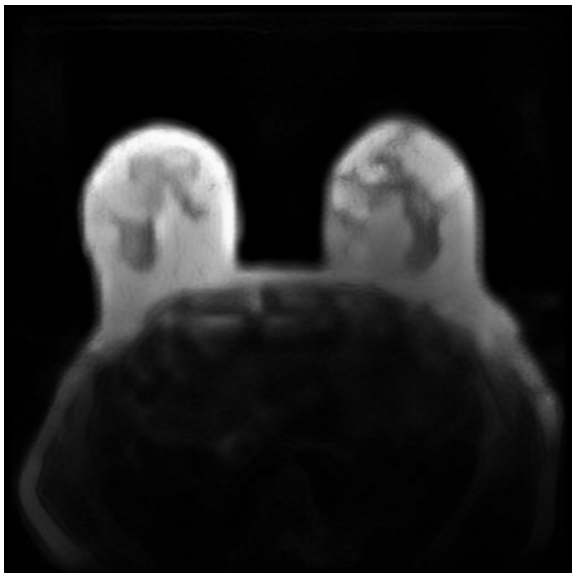


Figure 5.11: RIDER + Duke + LIDC based combined training method model upscale of a synthetic image.

Although the developed model introduced blur in the image degradation process of the from-scratch training method, this was insufficient as it is not fully capable of addressing the degradation found in synthetic imaging.

In order to solve this issue, heavy blurring was introduced into the image degradation process during the finetuning paired part of the combined training procedure in the hopes of better matching the images of the LR version of the ground truth images of the datasets with the characteristics of synthetic imaging. Using the same dataset, despite the persistence of some blurring, resulted in a much more capable model of handling the characteristics of synthetic imaging, as seen on Figure 5.12.

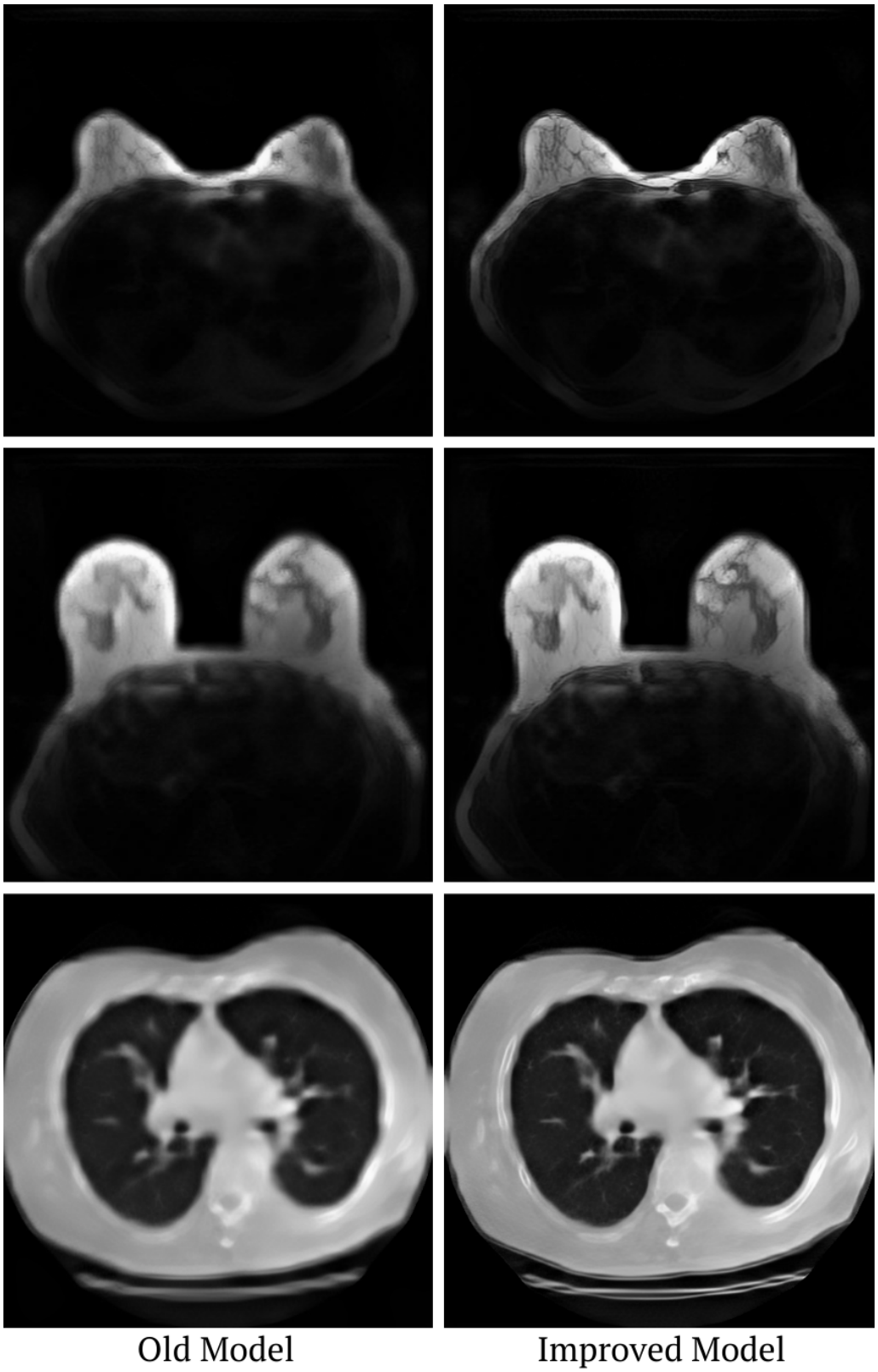


Figure 5.12: Comparison of upscaled synthetic images between the final trained model and its improved version using blurring on the paired finetuning phase of the combined training method.

5.4.2.1 Results

Metric results are not particularly meaningful in the context of the improved final model due to the tested subset being different from the one used during training. As said before, the improved model introduced blurriness in the image process degradation. Hence, the testing subset of this model differs from all the other, not allowing for direct comparisons with other models' results. Despite this, it might still be useful to analyse the chosen metrics evaluation (see Table 5.10) of the upscales performed by the improved model, as well as to visually examine some upscaled images of the testing subset (see Figure 5.13). In order to maintain the metrics results as comparable as possible with other developed models, these were also calculated based solely on the Duke and RIDER datasets, which exclusively constituted the testing subset of this evaluation.

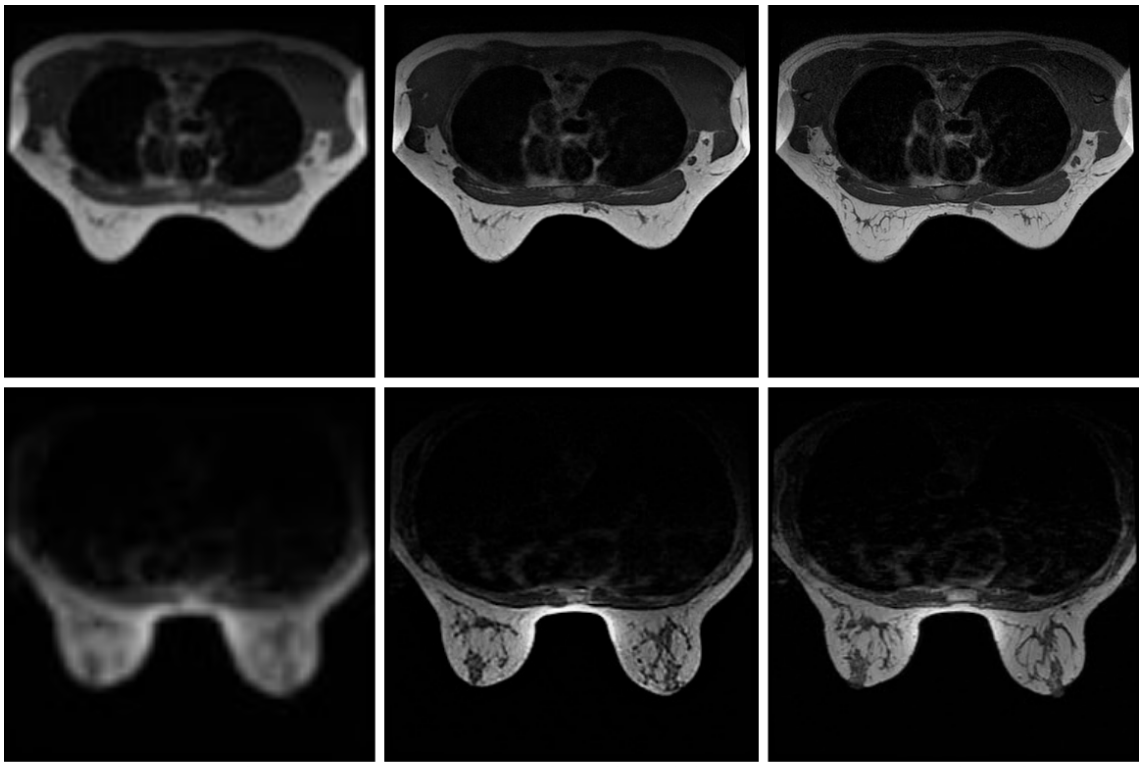


Figure 5.13: Comparison of upscaled images from the testing subset between the bicubic interpolation model and the final model with specialised training directed at synthetic imaging.

Model/Metrics' Results	SSIM	PSNR	MMD	LPIPS
Improved Final Model (Full testing subset)	0.70	23.72	1143.55	0.10
Improved Final Model (Duke and RIDER testing subset)	0.82	31.01	52.51	0.18

Table 5.10: Metrics' results of the final model with improved training methods to better handle the characteristics of synthetic images.

By visually analysing the images, the significant advantage of the developed model over bicubic interpolation is still evident. Despite this, and in comparison with previous results, an expected increase in noise and a decrease in overall anatomical detail is noticeable, considering the over-degraded testing subset used for these upscales.

The metrics results do not differ too much from the previously obtained results of the final model without the new degradation process. Despite this, these reflect the slightly diminished structural similarity, as per SSIM metric, given the decreased anatomic detail in comparison with other models. Furthermore, the increased noise is also reflected in these metrics, specifically by the PSNR score. The LPIPS score is one the best achieved so far, and while it's challenging to provide a definitive explanation due to the metrics' inherent abstraction, this result is clearly positive. LPIPS has proven its accuracy in capturing human perception and overall model quality, making this score a robust indicator of the improved model's performance despite the more challenging LR images with increased blurriness used for upscaling in the testing procedure.

5.5 Final Analysis

The final model, which was trained with both a complex image degradation process and one focused on synthetic data characteristics, has shown to be more versatile and capable of handling different types of image degradation present in synthetic images. It's also notorious the improvement in the metric results caused by the tuning of training methods, but also by introducing new datasets to the training procedure (see Table 5.11).

Model/Metrics' Results	SSIM	PSNR	MMD	LPIPS
Duke Finetuned Paired	0.82	31.99	43.10	0.21
Duke From Scratch	0.81	30.70	59.69	0.23
Duke + RIDER Finetuned Paired	0.81	31.81	44.53	0.21
Duke + RIDER From Scratch	0.81	30.75	57.20	0.23
Duke + RIDER Combined	0.82	32.00	42.20	0.20
Duke + RIDER + LIDC Combined	**0.84**	**31.82**	**43.91**	**0.18**
Duke + RIDER + LIDC Combined Improved	**0.82**	**31.01**	**52.51**	**0.18**

Table 5.11: Best performing models metrics' results (LIDC dataset not used in testing)

The evolution of the LPIPS and SSIM metrics aligns with the visual improvements observed in the model's results. These metrics specifically reflect the enhanced quality and amount of detail recovered, as well as the improved textures. While other metrics, such as PSNR and MMD, do not show clear improvement or deterioration, they remain constant and stable, still achieving good results. An exception to this is the combined improved model, which, as explained before, the results can't be directly comparable due to the use of different testing subset images.

When visually comparing the final model's results with those from a pre-trained Real-ESRGAN model and bicubic interpolation, it's clear that the final developed model has a significant advantage over traditional methods like bicubic interpolation. Additionally, when compared to a standard pre-trained Real-ESRGAN model designed for real-life synthetic images, the final model shows noticeable improvements, especially in terms of texture and anatomical details, with the generated images now showing textures more similar to the ground truth and recovering more anatomical details (see Figure 5.14).

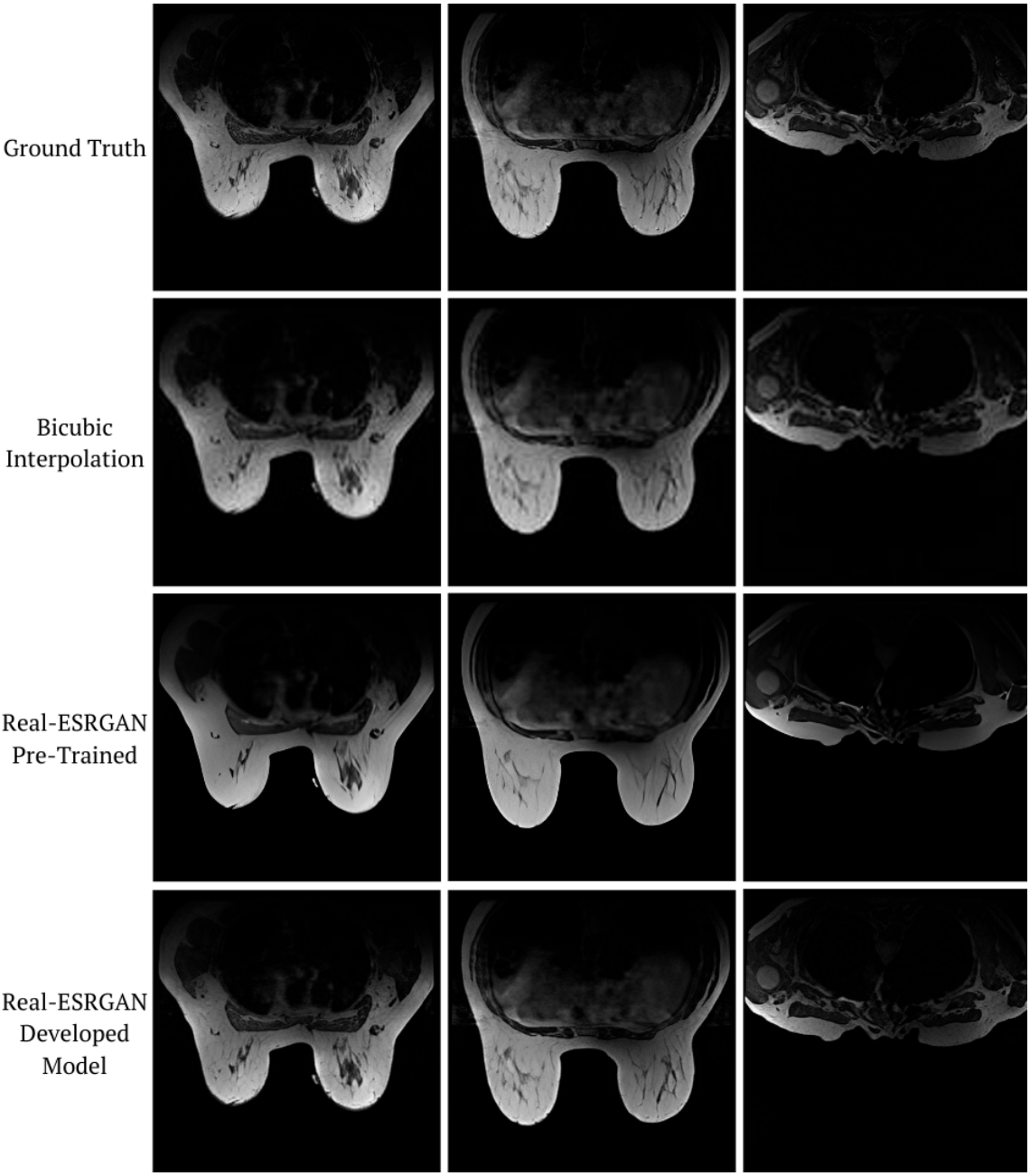


Figure 5.14: Comparison between 3 sets of examples of ground truth, bicubic interpolation, Real-ESRGAN pre-trained and Real-ESRGAN final developed model images.

This way, the resulting developed image SR model has shown significant improvement over other standard models in handling medical imaging, proving that the conducted training with the datasets and the training method used were essential and well-targeted.

5.5.1 ChatGPT Image Quality Assessment

According to Chen et al. [12], chatGPT 4 demonstrates remarkable proficiency in image quality assessment. OpenAI recently unveiled the chatGPT 4o model, which, as outlined in the launch notes [48], offers improved speed and capabilities across text, voice, and vision. Specifically, the vision capabilities of GPT 4o are a key area of interest for assessing image quality.

To gain insights from one of the most advanced LLM models, GPT 4o was tasked with evaluating six images. The set included the ground truth image (the reference image), one bicubic interpolation upscale of the LR version of the ground truth image, and five upscales of the LR version produced by various developed models. Each image was linked with a unique identifier to ensure anonymisation of the attribution of image generation to specific models. This procedural step enabled chatGPT to evaluate images independently, devoid of any association with their originating model. Such anonymisation safeguards against biases in the evaluation process, thereby maintaining the integrity and objectivity of performance metrics analysis. chatGPT 4o was asked to assess the images on the following criteria:

- Contrast;
- Structural similarity and detail;
- Texture;
- Overall Image Quality.

The results A.3.3 obtained from this prompt were very positive. In the case of bicubic interpolation, chatGPT noted that the resulting images tended to have lower contrast and higher brightness levels compared to the ground truth. This method also struggled to reproduce fine details accurately, resulting in blurred finer structures and smoother textures than what was present in the original images. As a result, there was a noticeable loss of important textural information from the images.

The finetuned training method model produced outputs with more precise details compared to bicubic interpolation. However, these details were not as sharp as expected compared to the ground truth. The textures in the outputs appeared relatively smooth, which led to a loss of delicate textural information.

ChatGPT found that both models generally improved contrast compared to bicubic interpolation and the finetuned training method model when comparing models trained from scratch with finetuned paired trained models. The finetuned paired trained models tended to exhibit more prominent anatomical details. However, they did not always preserve textures as effectively as

models trained from scratch, which was classified as being better than the finetuned model in maintaining finer textural details.

Furthermore, models trained using combined training methods were considered the best-performing by chatGPT. These models showed good contrast and brightness levels, resembling the ground truth images. They were both evaluated as effectively preserving details and textures, maintaining fine textural details better than other methods. ChatGPT closed the answer by concluding that the combined model with the three datasets used in training exhibits the highest overall image quality among the choices presented.

Overall, the results of the image quality assessment performed by the chatGPT 4o model are positive, given that they are aligned with the LPIPS metric, which has been this project's most trustworthy guiding metric.

5.5.2 Enhancing Synthetic CT Scans

The development of the generator models was not within the scope of this project, although its development coincided with this project's timeline. As a result, the model's progress, training, and tuning could not rely on examples of the synthetic data throughout the entire process. Despite this limitation, it was still possible to adjust the final model training procedure to better cope with synthetic imaging limitations.

The generator models of the synthetic data are based on the HA-GAN (Hierarchical Amortised GAN) [58] architecture, which was trained in two ways: first, a LR image is generated for the model to get accustomed to the 3D structure of the image, then a HR image is generated so as to learn the intrinsic details of each 2D slice. The breast synthetic image model was trained using the Duke dataset [16] and a private dataset, while the lung model was trained with the LIDC dataset [3].

The goal of incorporating various image degradation processes was to create a highly effective model that could enhance synthetic image data. This method aimed to ensure the model's robustness and utility for the project's objectives, which focus on improving the quality of artificial images.

Contrary to the primary testing data used during the development of the model, the synthetic images exhibit significant blurriness and artefacts such as noise and low contrast (see Figure 5.15).

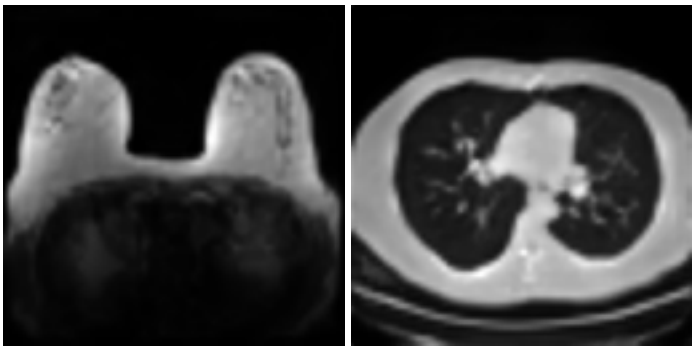


Figure 5.15: Example of a synthetic slice of a breast and lung CT Scan, respectively.

Despite the inclusion of blur in the image degradation processes during training, specifically in the from scratch training procedure where the built-in Real-ESRGAN's image degradation process is used, the resulting output of the upscale of the synthetic images using the deep learning model is not as good as we have seen previously in testing, which is expected given the aspects and characteristics of synthetic imaging. To obtain the positive results shown in Figure 5.16, it was necessary to add blurring to the image degradation process, as mentioned before. Compared to the results of bicubic interpolation and previous models without blurring in the image degradation process, the upscaled image with the deep learning model demonstrates increased clarity and detail, providing a much more refined and valuable representation of the synthetic data.

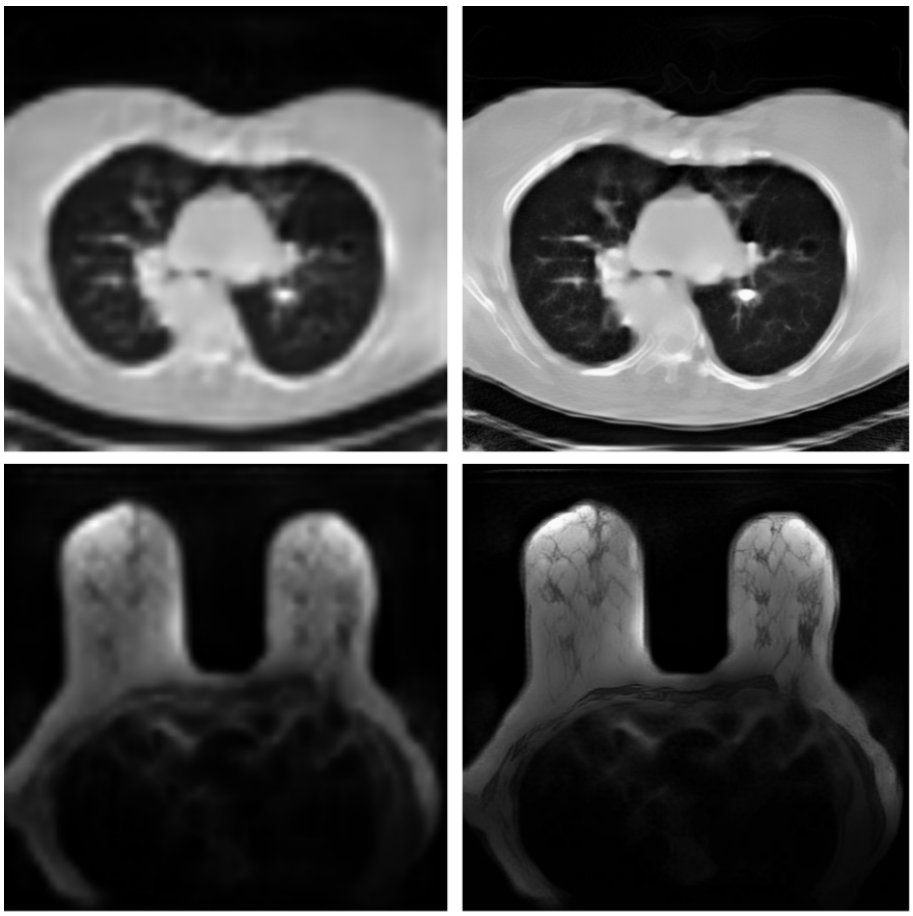


Figure 5.16: Comparison between bicubic interpolation and the developed deep learning model upscaling synthetic images.

5.6 Summary

Chapter 5 presents the results of applying Real-ESRGAN using diverse training methods, various datasets for training/validation/testing, and different image degradation processes. The chapter begins by detailing the sequential inclusion of datasets in the training process, demonstrating how

additional data from the Duke and LIDC datasets impacted the model's performance. Alongside the inclusion of datasets, the efficacy of different training methods was evaluated to determine their impact on the final model performance.

Furthermore, it was described how an image degradation process was introduced to replicate synthetic image data, incorporating blur effects. This new approach aimed to enhance the model's ability to handle synthetic CT scans effectively.

The chapter concludes with an analysis of the overall effectiveness of these techniques. It also includes a qualitative assessment using chatGPT for evaluating image quality.

Chapter 6

Conclusion and Future Work

6.1 Overview

In this thesis, a specialised image SR model based on Real-ESRGAN [67] was developed, specifically targeted at CT scans medical imagery. The primary objective was to enhance the resolution and quality of synthetic medical images, thereby improving their usability for training further deep learning models for medical purposes such as diagnostics.

The work undertaken involved several key steps to achieve the defined goal. Initially, appropriate and relevant datasets were selected, ensuring that the model would be trained on data that accurately represents the type of imagery encountered in medical settings.

Subsequently, various training and parameter tuning methods were evaluated, with a strong emphasis on optimising the employed metric results while constantly visually analysing the results to ensure the metrics were well employed and aligned with human perception.

Finally, the optimised model was applied to synthetic medical imagery. This application demonstrated the model's capability to improve image quality and anatomic detail, bringing LR synthetic images up to a more useful standard for medical diagnostics. The enhanced images exhibited improved anatomical detail, overall image quality, and texture similarity, aligning with the goals initially defined in chapter 1.

In conclusion, the development of the Real-ESRGAN-based image SR model for medical imagery was successful, given that the developed model demonstrated the ability to enhance medical imagery. Due to the concurrent development of GAN models that generate synthetic image data, it was not possible to design this project from the beginning with the specific characteristics of the synthetic images in mind. However, the simulated synthetic images used for training demonstrated excellent results, and the final model fine-tuned to synthetic data presented significant improvements over other standard methods in upscaling synthetic medical imagery.

6.2 Future Work

6.2.1 Image Degradation Process

One of the primary areas for future work involves finetuning even more the image degradation process to more accurately resemble the synthetic data used in this project. By refining this process, the model would be better equipped to handle real-world synthetic data, ultimately producing higher-quality upscaled images.

6.2.2 Metrics

In this project, LPIPS was the only metric that consistently provided a useful, precise classification and evaluation of generated images across all tested models and upscaling techniques. Future work should focus on investigating and validating other metrics that could offer additional insights into the quality and performance of SR models.

6.2.3 Further Data Acquisition

The results obtained in this project indicated that adding more datasets into the training procedure improved the models' metric results. Therefore, future work should investigate the impact of integrating further datasets into the training process. This could involve collecting a more diverse dataset, including medical imagery focused on different organs and zones of the human body, in order to enable the model to handle a wider diversity of anatomical structures and imaging conditions.

6.2.4 Alternative SR Technologies

While Real-ESRGAN has proven to be one of the state-of-the-art models for image SR, it might be useful to further investigate other technologies in this field more deeply. Exploring alternative models could reveal approaches more appropriate for this specific application of image SR.

Appendix A

Supplementary Statistical Results

A.1 Duke

A.1.1 Pre-Trained Real-ESRGAN

Duke P-T Real-ESRGAN	SSIM	PSNR	MMD	LPIPS
Average	0.77	29.62	77.65	0.33
Quartiles (0.25)	0.72	28.26	40.62	0.24
Quartiles (0.50)	0.77	29.68	67.14	0.32
Quartiles (0.75)	0.82	31.02	101.32	0.41
Min, Max	0.42, 0.95	20.66, 38.93	5.45, 420.69	0.10, 0.76
Standard Deviation	0.07	2.11	49.75	0.11

Table A.1: Pre-trained Real-ESRGAN model full results with Duke dataset.

A.1.2 Bicubic Interpolation

Duke Bicubic Interp.	SSIM	PSNR	MMD	LPIPS
Average	0.86	32.22	39.40	0.27
Quartiles (0.25)	0.84	30.96	24.73	0.23
Quartiles (0.50)	0.86	32.18	36.23	0.27
Quartiles (0.75)	0.88	33.50	51.11	0.31
Min, Max	0.58, 0.96	25.64, 42.21	4.10, 134.22	0.10, 0.56
Standard Deviation	0.03	1.93	18.84	0.06

Table A.2: Bicubic interpolation full results with Duke dataset.

A.1.3 Finetuning

Duke Finetuned	SSIM	PSNR	MMD	LPIPS
Average	0.77	30.05	67.86	0.26
Quartiles (0.25)	0.75	28.88	37.32	0.19
Quartiles (0.50)	0.79	30.09	62.94	0.25
Quartiles (0.75)	0.82	31.21	89.47	0.32
Min, Max	0.31, 0.92	22.60, 37.75	5.73, 479.82	0.10, 0.62
Standard Deviation	0.07	1.70	37.61	0.08

Table A.3: Finetuned trained Real-ESRGAN model full results with Duke dataset.

A.1.4 Finetuning Paired

Duke Finetuned Paired	SSIM	PSNR	MMD	LPIPS
Average	0.82	31.99	43.10	0.21
Quartiles (0.25)	0.80	30.77	24.69	0.15
Quartiles (0.50)	0.83	31.99	40.09	0.20
Quartiles (0.75)	0.85	33.20	57.59	0.26
Min, Max	0.54, 0.93	25.58, 39.89	3.50, 172.35	0.07, 0.56
Standard Deviation	0.04	1.82	22.83	0.08

Table A.4: Finetuned paired trained Real-ESRGAN model full results with Duke dataset.

A.1.5 From Scratch

Duke From Scratch	SSIM	PSNR	MMD	LPIPS
Average	0.81	30.70	59.69	0.23
Quartiles (0.25)	0.78	29.40	31.09	0.16
Quartiles (0.50)	0.82	30.65	55.14	0.22
Quartiles (0.75)	0.85	31.93	82.29	0.30
Min, Max	0.36, 0.94	24.36, 39.45	3.88, 320.51	0.07, 0.59
Standard Deviation	0.06	1.82	33.98	0.09

Table A.5: Trained from scratch Real-ESRGAN model full results with Duke dataset.

A.2 Duke + RIDER

A.2.1 Pre-Trained Real-ESRGAN

Duke + RIDER P-T Real-ESRGAN	SSIM	PSNR	MMD	LPIPS
Average	0.77	29.56	77.15	0.33
Quartiles (0.25)	0.72	28.18	39.38	0.24
Quartiles (0.50)	0.77	29.66	66.43	0.32
Quartiles (0.75)	0.82	31.03	101.08	0.41
Min, Max	0.42, 0.99	19.53, 42.11	1.70, 420.80	0.27, 0.75
Standard Deviation	0.07	2.29	50.66	0.11

Table A.6: Pre-trained Real-ESRGAN model full results with Duke and RIDER datasets.

A.2.2 Bicubic Interpolation

Duke + RIDER Bicubic Interp.	SSIM	PSNR	MMD	LPIPS
Average	0.86	32.23	38.67	0.27
Quartiles (0.25)	0.84	30.93	23.91	0.23
Quartiles (0.50)	0.87	32.19	35.70	0.27
Quartiles (0.75)	0.88	33.55	50.72	0.31
Min, Max	0.58, 0.99	24.75, 45.04	0.87, 134.22	0.08, 0.56
Standard Deviation	0.03	2.06	19.23	0.06

Table A.7: Bicubic interpolation full results with Duke and RIDER datasets.

A.2.3 Finetuning

Duke + RIDER Finetuned	SSIM	PSNR	MMD	LPIPS
Average	0.77	29.73	72.75	0.27
Quartiles (0.25)	0.74	28.46	37.53	0.20
Quartiles (0.50)	0.79	29.81	65.50	0.26
Quartiles (0.75)	0.82	31.03	96.67	0.34
Min, Max	0.26, 0.92	21.08, 40.10	2.70, 573.78	0.10, 0.70
Standard Deviation	0.08	1.94	43.95	0.09

Table A.8: Finetuned trained Real-ESRGAN model full results with Duke and RIDER datasets.

A.2.4 Finetuning Paired

Duke + RIDER Finetuned Paired	SSIM	PSNR	MMD	LPIPS
Average	0.81	31.81	44.53	0.21
Quartiles (0.25)	0.78	30.58	24.89	0.15
Quartiles (0.50)	0.81	31.86	41.01	0.19
Quartiles (0.75)	0.84	33.13	59.42	0.26
Min, Max	0.53, 0.92	22.46, 45.15	1.68, 214.35	0.06, 0.57
Standard Deviation	0.04	2.00	24.76	0.07

Table A.9: Finetuned paired trained Real-ESRGAN model full results with Duke and RIDER datasets.

A.2.5 From Scratch

Duke + RIDER From Scratch	SSIM	PSNR	MMD	LPIPS
Average	0.81	30.75	57.20	0.23
Quartiles (0.25)	0.78	29.52	29.48	0.14
Quartiles (0.50)	0.82	30.76	52.96	0.21
Quartiles (0.75)	0.85	32.00	78.88	0.30
Min, Max	0.26, 0.99	21.88, 41.66	1.88, 332.71	0.01, 0.69
Standard Deviation	0.06	1.92	32.68	0.10

Table A.10: Trained from scratch Real-ESRGAN model full results with Duke and RIDER datasets.

A.2.6 From Scratch + Finetuning Paired

Duke + RIDER Combined	SSIM	PSNR	MMD	LPIPS
Average	0.82	32.00	42.20	0.20
Quartiles (0.25)	0.80	30.84	24.13	0.14
Quartiles (0.50)	0.83	32.05	39.59	0.19
Quartiles (0.75)	0.85	33.27	56.18	0.26
Min, Max	0.52, 0.93	22.94, 42.77	1.46, 206.66	0.06, 0.60
Standard Deviation	0.04	1.92	22.61	0.08

Table A.11: Finetuned paired trained Real-ESRGAN model of a trained from scratch Real-ESRGAN model full results with Duke and RIDER datasets.

A.3 LIDC + Duke + RIDER

A.3.1 From Scratch + Finetuning Paired

LIDC + Duke + RIDER Combined	SSIM	PSNR	MMD	LPIPS
Average	0.71	24.19	1112.60	0.09
Quartiles (0.25)	0.62	20.04	113.15	0.05
Quartiles (0.50)	0.75	25.64	237.18	0.08
Quartiles (0.75)	0.84	28.67	865.35	0.12
Min, Max	0.09, 0.99	6.17, 43.43	1.26, 21315.99	0.01, 0.53
Standard Deviation	0.15	6.39	2246.97	0.06

Table A.12: Finetuned paired trained Real-ESRGAN model of a trained from scratch Real-ESRGAN model full results with Duke, RIDER and LIDC datasets.

LIDC + Duke + RIDER Combined	SSIM	PSNR	MMD	LPIPS
Average	0.84	31.82	43.91	0.18
Quartiles (0.25)	0.82	30.54	25.57	0.12
Quartiles (0.50)	0.84	31.88	40.54	0.16
Quartiles (0.75)	0.86	33.20	58.22	0.23
Min, Max	0.52, 0.99	22.28, 43.43	1.26, 195.94	0.01, 0.53
Standard Deviation	0.04	2.12	26.62	0.07

Table A.13: Finetuned paired trained Real-ESRGAN model of a trained from scratch Real-ESRGAN model full results with Duke, RIDER and LIDC datasets. Testing performed only with Duke and RIDER datasets.

A.3.2 From Scratch + Finetuning Paired with new Image Degradation Process

LIDC + Duke + RIDER Combined	SSIM	PSNR	MMD	LPIPS
Average	0.70	23.72	1143.55	0.10
Quartiles (0.25)	0.60	19.70	133.77	0.06
Quartiles (0.50)	0.73	25.07	270.76	0.09
Quartiles (0.75)	0.82	27.95	935.77	0.13
Min, Max	0.08, 0.99	6.05, 42.40	1.59, 21682.07	0.01, 0.55
Standard Deviation	0.16	6.16	2264.60	0.06

Table A.14: Finetuned paired trained Real-ESRGAN model of a trained from scratch Real-ESRGAN model full results with Duke, RIDER and LIDC datasets with new image degradation process.

LIDC + Duke + RIDER Combined	SSIM	PSNR	MMD	LPIPS
Average	0.82	31.10	52.51	0.18
Quartiles (0.25)	0.80	29.79	29.38	0.12
Quartiles (0.50)	0.83	31.12	48.09	0.17
Quartiles (0.75)	0.85	32.49	70.59	0.23
Min, Max	0.50, 0.99	21.34, 45.39	1.59, 216.20	0.01, 0.55
Standard Deviation	0.04	2.13	29.27	0.07

Table A.15: Finetuned paired trained Real-ESRGAN model of a trained from scratch Real-ESRGAN model full results with Duke, RIDER and LIDC datasets with new image degradation process. Testing performed only with Duke and RIDER datasets.

A.3.3 ChatGPT 4o Image Quality Assessment Response

A.3.3.1 Combination

Image-Model Combination:

- Image A - Bicubic Interpolation (Duke + RIDER datasets);
- Image B - Finetuned Model (Duke + RIDER datasets);
- Image C - Finetuned Paired Model (Duke + RIDER datasets);
- Image D - From Scratch Model (Duke + RIDER datasets);
- Image E - Combined Training Method Model (Duke + RIDER datasets);
- Image F - Combined Training Method Model (Duke + RIDER + LIDC datasets).

A.3.3.2 Image Quality Assessment

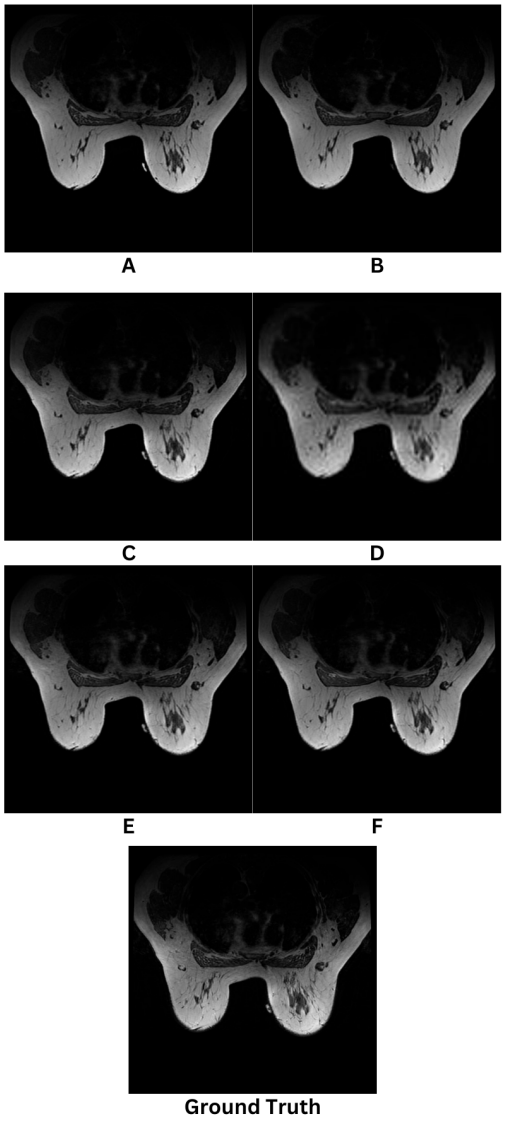


Figure A.1: Imaged used in the prompt for the chatGPT 4o image quality assessment analysis.

Image A:

- Quality: Image A has a lower contrast compared to the ground truth image. The overall brightness seems higher, which results in less distinction between different tissue types.
- Detail: The details, especially in the bone structures and soft tissues, are less pronounced. The finer structures appear slightly blurred.
- Texture: The texture is smoother, leading to a loss of important textural information that is visible in the ground truth image.

Image B:

- Quality: Image B also has lower contrast, similar to image A. The overall brightness is quite high, causing the image to appear overexposed.
- Detail: The details in the bones and surrounding tissues are not as sharp. There is a noticeable loss of detail compared to the ground truth image.
- Texture: The texture is quite smooth, leading to a washed-out appearance and loss of fine textural information.

Image C:

- Quality: Image C shows better contrast than images A and B, closer to the ground truth image. However, it still appears slightly overexposed.
- Detail: The details are better preserved in image C. The bone structures and soft tissues are more distinguishable.
- The texture is improved compared to A and B, but it still lacks the fine granularity seen in the ground truth image.

Image D:

- Quality: Image D has a similar quality to image C but appears slightly darker, which helps in preserving more details.
- Detail: The details in image D are quite good, with visible bone structures and finer soft tissue distinctions.
- Texture: The texture is better preserved than in previous images, showing a closer resemblance to the ground truth image.

Image E:

- Quality: Image E has a good balance of contrast and brightness, closer to the ground truth image.
- Detail: The details in image E are well-preserved, with clear distinctions in bone and soft tissues.
- Texture: The texture in image E is quite close to the ground truth image, maintaining fine granularity and textural details.

Image F:

- Quality: Image F also has good contrast and brightness, similar to image E, and is close to the ground truth image.
- Detail: The details are well-preserved, with clear visibility of bone structures and soft tissues.
- Texture: The texture in image F is very similar to the ground truth image, maintaining the fine textural details effectively.

Summary:

- Images A and B are of lower quality with reduced detail and texture.
- Images C and D show improvements in contrast and detail, with D being slightly better.
- Images E and F are the closest to the ground truth image in terms of quality, detail, and texture, with F possibly being the best match overall.

References

[1] PydiCom Library. URL https://pydicom.github.io/.

[2] N. Ahn, B. Kang, and K.-A. Sohn. Fast, accurate, and lightweight super-resolution with cascading residual network. *Proc. Eur. Conf. Comput. Vis.*, pages 256–272, 2018.

[3] S. G. Armato III, G. McLennan, L. Bidaut, M. F. McNitt-Gray, C. R. Meyer, A. P. Reeves, B. Zhao, D. R. Aberle, C. I. Henschke, E. A. Hoffman, E. A. Kazerooni, H. MacMahon, E. J. R. Van Beek, D. Yankelevitz, A. M. Biancardi, P. H. Bland, M. S. Brown, R. M. Engelmann, G. E. Laderach, D. Max, R. C. Pais, D. P. Y. Qing, R. Y. Roberts, A. R. Smith, A. Starkey, P. Batra, P. Caligiuri, A. Farooqi, G. W. Gladish, C. M. Jude, R. F. Munden, I. Petkovska, L. E. Quint, L. H. Schwartz, B. Sundaram, L. E. Dodd, C. Fenimore, D. Gur, N. Petrick, J. Freymann, J. Kirby, B. Hughes, A. V. Casteele, S. Gupte, M. Sallam, M. D. Heath, M. H. Kuhn, E. Dharaiya, R. Burns, D. S. Fryd, M. Salganicoff, V. Anand, U. Shreter, S. Vastagh, B. Y. Croft, and L. P. Clarke. Data from lidc-idri, 2015. URL https://doi.org/10.7937/K9/TCIA.2015.LO9QL9SX.

[4] Jonathan T. Barron. A general and adaptive robust loss function. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2019.

[5] Yochai Blau and Tomer Michaeli. The perception-distortion tradeoff. In *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 6228–6237, 2018. doi: 10.1109/CVPR.2018.00652.

[6] Mariusz Bojarski, Davide Del Testa, Daniel Dworakowski, Bernhard Firner, Beat Flepp, Prasoon Goyal, Lawrence D Jackel, Mathew Monfort, Urs Muller, Jiakai Zhang, et al. End to end learning for self-driving cars. In *arXiv preprint arXiv:1604.07316*, 2016.

[7] Karsten M. Borgwardt, Arthur Gretton, Malte J. Rasch, Hans-Peter Kriegel, Bernhard Schölkopf, and Alex J. Smola. Integrating structured biological data by Kernel Maximum Mean Discrepancy. *Bioinformatics*, 22(14):e49–e57, 7 2006. doi: 10.1093/bioinformatics/btl242. URL https://doi.org/10.1093/bioinformatics/btl242.

[8] G. Bradski. The opencv library, 2000. URL https://opencv.org/. Accessed: 2024-06-20.

[9] Jerrold T. Bushberg, J. Anthony Seibert, Edwin M. Leidholdt, and John M. Boone. *The Essential Physics of Medical Imaging*. Lippincott Williams & Wilkins, 3rd edition, 2011.

[10] Rama Chellappa, Carrie Wilson, and Sastry Sirohey. Human and machine recognition of faces: A survey. *Proceedings of the IEEE*, 83(5):705–740, 1995.

[11] Yudong Chen, Ying Tai, Xiaoming Liu, Chunhua Shen, and Jian Yang. Fsrnet: End-to-end learning face super-resolution with facial priors. *Proc. IEEE*, 2018.

[12] Zhihao Chen, Bin Hu, Chuang Niu, Tao Chen, Yuxin Li, Hongming Shan, and Ge Wang. Iqagpt: Image quality assessment with vision-language and chatgpt models, 2023.

[13] Antonia Creswell, Tom White, Vincent Dumoulin, Kai Arulkumaran, Biswa Sengupta, and Anil A. Bharath. Generative adversarial networks: An overview. *IEEE Signal Processing Magazine*, 35(1):53–65, 2018. doi: 10.1109/MSP.2017.2765202.

[14] Chao Dong, Chen Change Loy, Kaiming He, and Xiaoou Tang. Image super-resolution using deep convolutional networks, Jun 2015. URL https://ieeexplore.ieee.org/abstract/document/7115171.

[15] Chao Dong, Chen Change Loy, Kaiming He, and Xiaoou Tang. Image super-resolution using deep convolutional networks. In *European Conference on Computer Vision (ECCV)*, pages 184–199. Springer, 2016.

[16] Duke Breast Cancer MRI Dataset. URL https://sites.duke.edu/mazurowski/resources/breast-cancer-mri-dataset/.

[17] Rosana El Jurdi, Caroline Petitjean, Paul Honeine, Veronika Cheplygina, and Fahed Abdallah. High-level prior-based loss functions for medical image segmentation: A survey. *Computer Vision and Image Understanding*, 210:103248, 2021. ISSN 1077-3142. doi: https://doi.org/10.1016/j.cviu.2021.103248. URL https://www.sciencedirect.com/science/article/pii/S1077314221000928.

[18] Andre Esteva, Alexandre Robicquet, Bharath Ramsundar, Volodymyr Kuleshov, Mark DePristo, Katherine Chou, Cheryl Cui, Greg Corrado, Sebastian Thrun, and Jeff Dean. A guide to deep learning in healthcare. *Nature Medicine*, 25(1):24–29, 2019.

[19] Python Software Foundation. Python os module, 2024. URL https://docs.python.org/3/library/os.html. Accessed: 2024-06-20.

[20] Python Software Foundation. Python sys module, 2024. URL https://docs.python.org/3/library/sys.html. Accessed: 2024-06-20.

[21] Bo Fu, Liyan Wang, Yuechu Wu, Yufeng Wu, Shilin Fu, and Yonggong Ren. Weak texture information map guided image super-resolution with deep residual networks. *Multimedia Tools and Applications*, 81, 06 2021. doi: 10.1007/s11042-021-11085-7.

[22] Ian Goodfellow, Yoshua Bengio, and Aaron Courville. *Deep Learning*. MIT Press, 2016. http://www.deeplearningbook.org.

[23] S. Grossberg. Recurrent neural networks. *Scholarpedia*, 8(2):1888, 2013. doi: 10.4249/scholarpedia.1888. revision #138057.

[24] J. Guo and H. Chao. Building an end-to-end spatial-temporal convolutional network for video super-resolution. *Proc. 31st AAAI Conf. Artif. Intell.*, pages 4053–4060, 2017.

[25] Charles R. Harris, K. Jarrod Millman, Stéfan J. van der Walt, Ralf Gommers, Pauli Virtanen, David Cournapeau, Eric Wieser, Julian Taylor, Sebastian Berg, Nathaniel J. Smith, Robert Kern, Matti Picus, Stephan Hoyer, Marten H. van Kerkwijk, Matthew Brett, Allan Haldane, Jaime F. Del Río, Mark Wiebe, Pearu Peterson, Pierre Gérard-Marchant, Kevin Sheppard, Tyler Reddy, Warren Weckesser, Hameer Abbasi, Christoph Gohlke, and Travis E. Oliphant. Array programming with numpy, 2020. URL https://numpy.org/. Accessed: 2024-06-20.

[26] John Heaton, Nick Polson, and Craig Witte. Deep learning for finance: Deep portfolios. *Applied Stochastic Models in Business and Industry*, 33(1):3–12, 2017.

[27] Alain Horé and Djemel Ziou. Image quality metrics: Psnr vs. ssim. *2010 20th International Conference on Pattern Recognition*, pages 2366–2369, 2010. doi: 10.1109/ICPR.2010.579.

[28] IBM. What is computer vision?, 2024. URL https://www.ibm.com/topics/computer-vision. Accessed: 2024-06-28.

[29] Shruti Jadon. A survey of loss functions for semantic segmentation. *arXiv.org*, Sep 2020. URL https://arxiv.org/abs/2006.14822.

[30] Justin Johnson, Alexandre Alahi, and Li Fei-Fei. Perceptual losses for real-time style transfer and super-resolution. *Springer International Publishing*, pages 694–711, march 2016.

[31] Charles E. Kahn, John A. Carrino, Michael J. Flynn, Donald J. Peck, and Steven C. Horii. DICOM and Radiology: past, present, and future. *Journal of the American College of Radiology*, 4(9):652–657, 9 2007. doi: 10.1016/j.jacr.2007.06.004. URL https://doi.org/10.1016/j.jacr.2007.06.004.

[32] Eyal Klang, Lior Deutsch, Itai Dgani, and Eli Zemel. Deep learning and medical imaging: A state-of-the-art review. *Radiology*, 290(3):650–669, 2019.

[33] W.-S. Lai, J.-B. Huang, N. Ahuja, and M.-H. Yang. Deep laplacian pyramid networks for fast and accurate super-resolution. *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, pages 5835–5843, 2017.

[34] W.-S. Lai, J.-B. Huang, N. Ahuja, and M.-H. Yang. Fast and accurate image super-resolution with deep laplacian pyramid networks. *IEEE Trans. Pattern Anal. Mach. Intell.*, 41(11): 2599–2613, Nov 2018.

[35] Yann LeCun, Yoshua Bengio, and Geoffrey Hinton. Deep learning. *Nature*, 521(7553): 436–444, 2015.

[36] Christian Ledig, Lucas Theis, Ferenc Huszár, Jose Caballero, Andrew Cunningham, Alejandro Acosta, ..., and Wenzhe Shi. Photo-realistic single image super-resolution using a generative adversarial network. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 4681–4690, 2017.

[37] Christian Ledig, Lucas Theis, Ferenc Huszar, Jose Caballero, Andrew Cunningham, Alejandro Acosta, Andrew Aitken, Alykhan Tejani, Johannes Totz, Zehan Wang, and Wenzhe Shi. Photo-realistic single image super-resolution using a generative adversarial network. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, July 2017.

[38] Y. Li, B. Sixou, and F. Peyrin. A review of the deep learning methods for medical images super resolution problems. *IRBM*, 42(2):120–133, 2021. ISSN 1959-0318. doi: https://doi.org/10.1016/j.irbm.2020.08.004. URL https://www.sciencedirect.com/science/article/pii/S1959031820301408.

[39] Renjie Liao, Xin Tao, Ruiyu Li, Zehuan Ma, and Jiaya Jia. Video super-resolution via deep draft-ensemble learning. *Proc. IEEE Int. Conf. Comput. Vis.*, pages 531–539, 2015.

[40] Geert Litjens, Thijs Kooi, Babak E. Bejnordi, Arnaud Arindra Adiyoso Setio, Francesco Ciompi, Mohsen Ghafoorian, and Clara I. Sánchez. A survey on deep learning in medical image analysis. *Medical Image Analysis*, 42:60–88, 2017.

[41] Fredrik Lundh. Python imaging library (pil), 2009. URL https://python-pillow.org/. Accessed: 2024-06-20.

[42] Chris Madden. URL chrismadden.co.uk.

[43] Awais Mansoor, Tan Vongkovit, and Marius George Linguraru. Adversarial approach to diagnostic quality volumetric image enhancement. *Biomedical Imaging (ISBI 2018), 2018 IEEE 15th International Symposium on*, pages 353–356, 2018.

[44] Cynthia H. McCollough and Richard L. Morin. The technical design and performance of whole-body ct scanners. *Radiology*, 259(3):660–679, 2011.

[45] Craig Morris and Paul Bell. Abdominal ct windows (advanced), 2022. URL https://litfl.com/abdominal-ct-windows-advanced/. Accessed: 2024-06-30.

[46] John F Nash. Non-cooperative games. *Annals of Mathematics*, 54(2):286–295, 1951.

[47] Chaity Banerjee University of Central Florida, Chaity Banerjee, University of Central Florida, Tathagata Mukherjee University of Alabama in Huntsville, Tathagata Mukherjee, University of Alabama in Huntsville, Eduardo Pasiliao Air Force Research Labs, Eduardo Pasiliao, Air Force Research Labs, Kennesaw State University, and et al. An empirical study on generalizations of the relu activation function: Proceedings of the 2019 acm southeast conference, Apr 2019. URL https://dl.acm.org/doi/abs/10.1145/3299815.3314450?casa_token=25xUt8UzDvUAAAAA%3ATwOe4Z3w-1mdYEonA4S8Y7Yv3HKxE-yZITKfLT3ve3_7iiiVVsquKBSqq6GYc61CWVDHXke7G3K2rg.

[48] OpenAI. Gpt-4 turbo and more tools to chatgpt (free). https://openai.com/index/gpt-4o-and-more-tools-to-chatgpt-free/, 2024. Accessed: 2024-06-16.

[49] Sung-Wook Park, Jae-Sub Ko, Jun-Ho Huh, and Jong-Chan Kim. Review on generative adversarial networks: Focusing on computer vision and its applications, May 2021. URL https://www.mdpi.com/2079-9292/10/10/1216.

[50] Kim K. Y. Kang S. K. Kim Y. K. Park J., Hwang D. and J. S. Lee. Computed tomography super-resolution using deep convolutional neural network. *Physics in Medicine & Biology*, 63, 2018.

[51] Ravindra Parmar. Training deep neural networks, Sep 2018. URL https://towardsdatascience.com/training-deep-neural-networks-9fdb1964b964.

[52] Adam Paszke, Sam Gross, Soumith Chintala, Gregory Chanan, Edward Yang, Zachary DeVito, Zeming Lin, Alban Desmaison, Luca Antiga, and Adam Lerer. Pytorch: An imperative style, high-performance deep learning library, 2019. URL https://pytorch.org/. Accessed: 2024-06-20.

[53] RIDER Collections - The Cancer Imaging Archive (TCIA) Public Access - Cancer Imaging Archive Wiki. URL https://wiki.cancerimagingarchive.net/display/Public/RIDER+Collections.

[54] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In Nassir Navab, Joachim Hornegger, William M. Wells, and Alejandro F. Frangi, editors, *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*, pages 234–241, Cham, 2015. Springer International Publishing. ISBN 978-3-319-24574-4.

[55] David E Rumelhart, Geoffrey E Hinton, and Ronald J Williams. Learning representations by back-propagating errors. *Nature*, 1986. URL https://www.nature.com/articles/323533a0.

[56] Terence Sanger and P. Baljekar. The perceptron: A probabilistic model for information storage and ... *Semantic Scholar*, Nov 1958. URL https://www.semanticscholar.org/paper/The-perceptron%3A-a-probabilistic-model-for-storage-Sanger-Baljekar/5d11aad09f65431b5d3cb1d85328743c9e53ba96.

[57] Sergey Smirnov, Atanas Gotchev, and Karen Egiazarian. Methods for depth-map filtering in view-plus-depth 3d video representation. *EURASIP Journal on Advances in Signal Processing*, 2012, 12 2012. doi: 10.1186/1687-6180-2012-25.

[58] Li Sun, Junxiang Chen, Yanwu Xu, Mingming Gong, Ke Yu, and Kayhan Batmanghelich. Hierarchical amortized gan for 3d high resolution medical image synthesis. *IEEE Journal of Biomedical and Health Informatics*, 26(8):3966–3975, 2022. doi: 10.1109/JBHI.2022.3172976.

[59] Marshall F Tappen. Comparison of image interpolation methods. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 1–8, 2007.

[60] Eric J. Topol. High-performance medicine: the convergence of human and artificial intelligence. *Nature Medicine*, 25(1):44–56, 2019.

[61] Sik-Ho Tsang. Reading: Carn-cascading residual network (super resolution), Jul 2020. URL https://sh-tsang.medium.com/reading-carn-cascading-residual-network-super-resolution-cb86a34d8d1b.

[62] A. M. TURING. I.—COMPUTING MACHINERY AND INTELLIGENCE. *Mind*, LIX (236):433–460, 10 1950. ISSN 0026-4423. doi: 10.1093/mind/LIX.236.433. URL https://doi.org/10.1093/mind/LIX.236.433.

[63] Keigo Umehara, Jun Ota, and Takayuki Ishida. Application of super-resolution convolutional neural network for enhancing image resolution in chest ct. *Journal of digital imaging*, 31(4):441–450, 2018. doi: 10.1007/s10278-017-0033-z.

[64] United Nations. Sustainable development goals, n.d. URL https://sdgs.un.org/goals. Accessed: 2024-06-26.

[65] Qiang Wang, Yun Ma, Kai Zhao, and et al. A comprehensive survey of loss functions in machine learning. *Annals of Data Science*, 9:187–212, 2022. doi: 10.1007/s40745-020-00253-5.

[66] X. Wang et al. Esrgan: Enhanced super-resolution generative adversarial networks. *Proc. Eur. Conf. Comput. Vis. Workshop*, 2018.

[67] Xintao Wang, Liangbin Xie, Chao Dong, and Ying Shan. Real-esrgan: Training real-world blind super-resolution with pure synthetic data. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV) Workshops*, pages 1905–1914, October 2021.

[68] Z. Wang, E.P. Simoncelli, and A.C. Bovik. Multiscale structural similarity for image quality assessment. In *The Thrity-Seventh Asilomar Conference on Signals, Systems Computers, 2003*, volume 2, pages 1398–1402 Vol.2, 2003. doi: 10.1109/ACSSC.2003.1292216.

[69] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli. Image quality assessment: From error visibility to structural similarity. *IEEE Trans. Image Process.*, 13(4):600–612, Apr 2004.

[70] Zhihao Wang, Jian Chen, and Steven C. H. Hoi. Deep learning for image super-resolution: A survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 43(10):3365–3387, 2021. doi: 10.1109/TPAMI.2020.2982166.

[71] Jelmer M. Wolterink, Tim Leiner, Max A. Viergever, and Ivana Isgum. Generative adversarial networks for noise reduction in low-dose ct. *IEEE Transactions on Medical Imaging*, 36 (12):2536–2545, 2017.

[72] Chao Yang, Jinliang Ma, Xiaokang Yang, and Ming-Hsuan Yang. Deep learning for single image super-resolution: A brief review. *IEEE Transactions on Multimedia*, 22(12):3034–3057, 2020.

[73] Hujun Yang, Zhongyang Wang, Xinyao Liu, Chuangang Li, Junchang Xin, and Zhiqiong Wang. Deep learning in medical image super resolution: A review, Apr 2023. URL https://link.springer.com/article/10.1007/s10489-023-04566-9#citeas.

[74] Andy B. Yoo, Morris A. Jette, and Mark Grondona. *SLURM: Simple Linux Utility for Resource Management*. Lawrence Livermore National Laboratory, 2003. URL https://slurm.schedmd.com/.

[75] Chenyang You, Yu Zhang, Xuan Zhang, Guanbin Li, Shan Ju, and et al. Zhao, Zhizhong. Ct super-resolution gan constrained by the identical, residual, and cycle learning ensemble (gan-circle). *Preprint arXiv:1808.04256*, 2018.

[76] Chenyu You, Yi Zhang, Xiaoliu Zhang, Guang Li, Shenghong Ju, Zhen Zhao, Zhuiyang Zhang, Wenxiang Cong, Punam K. Saha, and Ge Wang. Ct super-resolution gan constrained by the identical, residual, and cycle learning ensemble (gan-circle). *IEEE Transactions on Medical Imaging*, 39:188–203, 2018. URL https://api.semanticscholar.org/CorpusID:51969563.

[77] Tom Young, Devamanyu Hazarika, Soujanya Poria, and Erik Cambria. Recent trends in deep learning based natural language processing. *IEEE Computational Intelligence Magazine*, 13 (3):55–75, 2018.

[78] Richard Zhang, Phillip Isola, Alexei A. Efros, Eli Shechtman, and Oliver Wang. The unreasonable effectiveness of deep features as a perceptual metric. 2018.

[79] Y. Zhang, K. Li, K. Li, L. Wang, B. Zhong, and Y. Fu. Image super-resolution using very deep residual channel attention networks. *Proc. Eur. Conf. Comput. Vis.*, pages 294–310, 2018.