NOVA SCHOOL OF SCIENCE & TECHNOLOGY

DEPARTMENT OF PHYSICS

ANA MARGARIDA LOPES FERREIRA

Bachelor of Science in Biomedical Engineering

ENHANCING SPINAL MAGNETIC RESONANCE IMAGE RESOLUTION WITH DEEP LEARNING

MASTER IN BIOMEDICAL ENGINEERING

NOVA University Lisbon
September, 2024

NOVA SCHOOL OF SCIENCE & TECHNOLOGY

DEPARTMENT OF PHYSICS

ENHANCING SPINAL MAGNETIC RESONANCE IMAGE RESOLUTION WITH DEEP LEARNING

ANA MARGARIDA LOPES FERREIRA

Bachelor of Science in Biomedical Engineering

Adviser: Prof. Dr. Ricardo Nuno Pereira Verga e Afonso Vigário
Associate Professor, NOVA University Lisbon

Co-adviser: Prof. Dr. Gonçalo Neto D'Almeida
Neurosurgery Coordinator, Hospital Lusíadas Amadora

Examination Committee

Chair: Prof. Dr. Hugo Filipe Silveira Gamboa
Full Professor, NOVA University Lisbon

Rapporteur: Prof. Dr. José Manuel Matos Ribeiro da Fonseca
Associate Professor with Habilitation, NOVA University Lisbon

Advisers: Prof. Dr. Ricardo Nuno Pereira Verga e Afonso Vigário
Associate Professor with Habilitation, NOVA University Lisbon

Prof. Dr. Gonçalo Neto D'Almeida
Neurosurgery Coordinator, Hospital Lusíadas Amadora

MASTER IN BIOMEDICAL ENGINEERING

NOVA University Lisbon
September, 2024

Enhancing Spinal Magnetic Resonance Image Resolution with Deep Learning

This document was created with the (pdf/Xe/Lua)LaTeX processor and the NOVAthesis template (v7.1.18) [1].

Acknowledgements

I dedicate this page to everyone who has contributed directly to this project, as well as to those who have played a pivotal role in my academic journey over the past five years, both in terms of scientific progress and personal development.

First and foremost, I would like to express my deepest gratitude to my advisors, Professor Ricardo Vigário and Dr. Gonçalo Neto D'Almeida, for granting me the invaluable opportunity to work on this thesis, and for their unwavering support, encouragement, and belief in the potential and success of this work. Professor Ricardo has been an essential guide in this project, with his exceptional knowledge and passion for MRI and the transformative power of Artificial Intelligence. His dedication to these fields has greatly influenced the direction of my research. I am equally grateful to Dr. Gonçalo for the opportunity to join his research team and for his constant kindness and support throughout this journey.

Achieving this milestone would not have been possible without the exceptional education I received at NOVA School of Science and Technology. I hold the entire institution in high regard for upholding rigorous standards and fostering an environment that promotes both academic and personal growth. To all my professors, I am profoundly grateful for introducing me to the essence of Biomedical Engineering and for igniting the passion that continues to drive my pursuits.

On a personal note, I extend my deepest thanks to my friends, whose support has been indispensable. Their willingness to listen and provide comfort has been invaluable throughout this journey. To my parents, who have been my steadfast anchors, I owe an immeasurable debt of gratitude. It is with sincere thanks that I acknowledge the opportunity to pursue this education and the values that guide my path. I am truly thankful for all the constant encouragement and support I have received.

I am also incredibly thankful to the rest of my family, whose constant belief in me has been a source of strength, particularly during the most challenging times. My deepest appreciation goes to my grandparents, who hold an irreplaceable place in my heart. Their uplifting and encouraging words have always provided comfort and motivation.

In conclusion, this thesis is a testament to the collective support, guidance, and

encouragement I have received from many exceptional individuals. Their contributions have been invaluable, and I carry the impact of their influence on me as I move forward in my journey. Because of this, and for many other reasons, I extend my sincerest thanks.

”

Just because a machine can do faster calculations, comparisons, and analytical solution creation, that doesn't make it smarter than you. It simply computes faster. In my world, in my belief, smarts have to do with your capacity to love, create, expand, transcend. These are qualities that no machine can ever bear, that are reserved to only humans.

“

— Pinar Seyhan Demirdag

Abstract

Medical imaging plays a crucial role in diagnosing and planning the treatment of various medical conditions. While Magnetic Resonance Imaging (MRI) is highly effective in visualizing soft tissue structures, its spatial resolution is often insufficient for detailed anatomical assessments. This limitation is particularly evident in cases such as lumbar disc herniations, where precise visualization of anatomical features is critical for accurate diagnosis and treatment. As a result, Computed Tomography (CT) scans are frequently employed to complement MRI by offering enhanced information on bone structures. The enhancement of MRI resolution is crucial to reducing dependence on CT, however, achieving this improvement is constrained by hardware-related trade-offs. Recent advances in deep learning-based Super-Resolution (SR) techniques have emerged as a promising approach to overcoming these limitations and enhancing MRI resolution, potentially mitigating the need for supplemental CT imaging.

Three state-of-the-art SR models, Fast Super-Resolution Convolutional Neural Network (FSRCNN), Residual Dense Network (RDN) and Enhanced Super-Resolution Generative Adversarial Network (ESRGAN), were implemented and rigorously tested. The primary objective was to improve the clarity of spinal MRI images, surpassing the traditional Bicubic Interpolation (BCI) method and enabling a clearer view of finer anatomical details. After extensive optimisation, ESRGAN proved to be the most effective model, offering superior perceptual quality and detail recovery. While RDN achieved better results in pixel-to-pixel metrics such as Peak Signal-to-Noise Ratio (PSNR), Spatial Correlation Coefficient (SCC), Structural Similarity Index (SSIM), Gradient Magnitude Similarity Deviation (GMSD), and Visual Information Fidelity (VIF), ESRGAN excelled in free-reference metrics like Signal-to-Noise Ratio (SNR) and the perceptual metric Perceptual Image Quality Evaluator (PIQE), striking a better balance between noise reduction and perceptual detail preservation, making it the preferred model.

When applied to high-quality high-resolution images, ESRGAN showed a tendency to prioritise denoising, focusing more on noise reduction rather than upsampling. However, when trained on images with minimal noise, the model was able to focus on upsampling, and successfully increasing the resolution. Furthermore, ESRGAN output displayed fewer

spatial artefacts, as reflected by its lower PIQE score, although the visual results were comparable to those of BCI.

The study also tested ESRGAN's generalisability across various degradation scenarios and different imaging modalities, including CT scans and grey-scale natural images. ESRGAN demonstrated good adaptability, effectively handling degradation patterns and consistently outperforming BCI in the SSIM, GMSD and VIF metrics.

The increased resolution achieved by ESRGAN paves the way for future advancements in anatomical segmentation and structural reconstruction. This enhancement in resolution holds the potential to enable more precise techniques, particularly in clinical applications. By providing clearer visualisation, especially in cases such as lumbar disc herniations, this improvement could greatly aid in surgical planning and contribute to better patient outcomes.

Keywords: Magnetic Resonance Imaging, Computed Tomography, Natural Grey-Scale Images, Deep Learning, Spatial Resolution, Super-Resolution Reconstruction, Image Degradation, Diagnostic Accuracy, Surgical Planning.

Resumo

A imagiologia médica desempenha um papel indispensável no diagnóstico e planeamento do tratamento de várias condições médicas. Embora a Ressonância Magnética (RM) seja bastante eficaz na visualização de tecidos moles, a sua resolução espacial é frequentemente insuficiente para avaliações anatómicas detalhadas. Esta limitação é particularmente evidente em casos como as hérnias discais lombares, em que a visualização precisa das caraterísticas anatómicas é fundamental para um diagnóstico e tratamento mais correto. Como resultado, os exames de Tomografia Computadorizada (CT) são frequentemente utilizados para complementar a RM oferecendo informações melhoradas sobre as estruturas ósseas. A melhoria da resolução da RM é crucial para reduzir a dependência da CT, no entanto, a obtenção desta melhoria é limitada por compromissos relacionados com o hardware. Avanços recentes em técnicas de *Deep Learning* (DL) baseadas em Super-Resolução (SR) surgiram como uma abordagem promissora para superar essas limitações e melhorar a resolução de RM, potencialmente mitigando a necessidade de imagens suplementares de CT.

Três modelos de SR encontrados na literatura, *Fast Super-Resolution Convolutional Dense Network* (FSRCNN), *Residual Neural Network* (RDN) e *Enhanced Super-Resolution Generative Adversarial Network* (ESRGAN), foram implementados e rigorosamente testados. O principal objetivo foi melhorar a clareza das imagens de RM da coluna vertebral, superando o método tradicional de Interpolação Bicúbica (BCI) e permitindo uma visualização mais nítida dos detalhes anatómicos. Após uma extensa otimização, o ESRGAN provou ser o modelo mais eficaz, oferecendo uma qualidade perceptiva e uma recuperação de detalhes superior. Enquanto RDN obteve melhores resultados em métricas pixel-a-pixel, como *Peak Signal-to-Noise Ratio* (PSNR), *Spatial Correlation Coefficient* (SCC), *Structural Similarity Index* (SSIM), *Gradient Magnitude Similarity Deviation* (GMSD), e *Visual Information Fidelity* (VIF), o ESRGAN destacou-se em métricas de referência livre, como *Signal-to-Noise Ratio* (SNR) e a métrica percetual *Perceptual Image Quality Evaluator* (PIQE), alcançando um melhor equilíbrio entre a redução de ruído e a preservação de detalhes perceptuais, tornando-o o modelo preferido.

Quando aplicado a imagens de alta-resolução e de alta qualidade, o ESRGAN mostrou

uma tendência para priorizar a remoção de ruído, focando-se mais na redução do mesmo do que no aumento da amostragem. No entanto, quando treinado em imagens com ruído mínimo, o modelo conseguiu concentrar-se na sobreamostragem e aumentar a resolução com sucesso. Além disso, os resultados obtidos pelo ESRGAN demonstraram a presença de menos artefactos espaciais, refletido pelo baixo valor de PIQE, embora a qualidade visual fosse comparável à do BCI.

O estudo também testou a capacidade de generalização do ESRGAN em vários cenários de degradação e diferentes modalidades de imagem, incluindo CT e imagens naturais na escala de cinzentos. O ESRGAN demonstrou uma boa adaptabilidade, lidando eficazmente com os padrões de degradação e superando consistentemente o BCI nas métricas de SSIM, GMSD e VIF.

O aumento da resolução obtido pelo ESRGAN abre o caminho a futuros avanços na segmentação anatómica e na reconstrução estrutural. Este aumento de resolução tem o potencial de permitir técnicas mais precisas, nomeadamente em aplicações clínicas. Ao proporcionar uma visualização mais clara, especialmente em casos como as hérnias discais lombares, esta melhoria pode optimizar o planeamento cirúrgico e, em última análise, conduzir a melhores resultados para os doentes.

Palavras-chave: Ressonância Magnética, Tomografia Computorizada, Imagens Naturais em Escala de Cinzentos, Aprendizagem Profunda, Resolução Espacial, Reconstrução por Super-Resolução, Degradação de Imagens, Precisão Diagnóstica, Planeamento Cirúrgico.

Contents

List of Tables

List of Figures

Acronyms

3D Three Dimensional *(p. 4)*

AI Artificial Intelligence *(p. 8)*
ANN Artificial Neural Network *(pp. 2, 8)*

BCI Bicubic Interpolation *(pp. 2, 7, 12, 14, 17, 20, 27–34, 36, 38, 40, 52)*
BLI Bilinear Interpolation *(pp. 7, 12, 40)*

CNN Convolutional Neural Network *(pp. 10, 14, 15)*
CPU Central Processing Unit *(p. 39)*
CT Computed Tomography *(pp. 1, 15, 17, 33, 35–38)*

DL Deep Learning *(pp. 2, 6, 8, 9, 13–18, 20, 28–30, 38, 39, 50, 52)*

ESRGAN Enhanced Super-Resolution Generative Adversarial Network *(pp. 16, 18, 20, 22, 27–32, 34, 36–40, 50, 52)*

FID Free Induction Decay *(pp. 3, 4)*
FOV Field of View *(pp. 4, 5)*
FSRCNN Fast Super-Resolution Convolutional Neural Network *(pp. 14, 16, 18, 20, 21, 27, 29, 30, 38, 50, 52)*
FWHM Full With at Half Maximum *(pp. 4, 5)*

GAN Generative Adversarial Network *(pp. 10, 11, 15)*
GMS Gradient Magnitude Similarity *(p. 25)*
GMSD Gradient Magnitude Similarity Deviation *(pp. 24, 25, 29, 36)*
GPU Graphics Processing Unit *(p. 39)*

HR High Resolution *(pp. 2, 6, 8, 13–16, 18–22, 24, 26–32, 38, 39, 52)*

IBP Iterative Back Projection *(p. 13)*

LR Low Resolution *(pp. 2, 6, 12–16, 18, 19, 27, 30, 33, 39, 40, 52)*

MISR Multi-Image Super-Resolution *(p. 6)*
ML Machine Learning *(pp. 8, 20)*
MRI Magnetic Resonance Imaging *(pp. 1–5, 12, 15–17, 19, 20, 30, 32, 33, 35–40)*
MSE Mean Squared Error *(pp. 15, 20, 21, 23, 24, 27, 38)*

NNI Nearest Neighbour Interpolation *(pp. 7, 12, 15)*

PIQE Perceptual Image Quality Evaluator *(pp. 26, 29, 30, 32, 33, 38, 52)*
PSF Point Spread Function *(pp. 4–6)*
PSNR Peak Signal-to-Noise Ratio *(pp. 12, 15, 24, 28–30, 36)*

RDB Residual Dense Block *(pp. 21, 22, 51)*
RDN Residual Dense Network *(pp. 16, 18, 20–22, 27–30, 38, 50, 52)*
RRDB Residual in Residual Dense Block *(pp. 22, 51)*

SCC Spatial Correlation Coefficient *(pp. 26, 29, 34, 36)*
SISR Single Image Super-Resolution *(p. 6)*
SNR Signal-to-Noise Ratio *(pp. 5, 6, 26, 29, 30, 32, 33)*
SR Super-Resolution *(pp. 2, 5, 6, 12–18, 20, 24, 28, 29, 35, 38, 40)*
SRCNN Super-Resolution Neural Network *(p. 14)*
SSIM Structural Similarity Index *(pp. 24, 29, 36)*

VIF Visual Information Fidelity *(pp. 25, 29, 36)*

1

Introduction

1.1 Context and Motivation

Medical imaging plays a crucial role in enabling healthcare professionals to accurately identify pathological patterns and determine the underlying causes of patients' symptoms [2]. Each imaging modality provides unique and specific information, which is the reason why Computed Tomography (CT) and Magnetic Resonance Imaging (MRI) are often used in a complementary way by doctors when diagnosing certain conditions. CT scans offer high-resolution and detailed images of lesions, but they have limitations when it comes to resolving soft tissue structures. In contrast, MRI excels at providing clear visualisations of soft tissues without the use of ionizing radiation, yet still fails shortly in offering detailed information about tissues such as bones. Consequently, relying on a single imaging modality may not adequately capture the complexity of different tissues, potentially complicating diagnosis and surgical planning [3].

An example of this application is the diagnosis and treatment of lumbar disc herniations. In the majority of cases, non-surgical treatments, including rest, medication and physiotherapy, are an effective course of action. Nevertheless, in the event of persistent or worsening pain, surgical intervention may be recommended. Physical examinations in conjunction with imaging techniques play a pivotal role in this process, enabling a more precise assessment and, subsequently, more efficacious surgical planning [4]. MRI is the primary imaging technique for diagnosing lumbar disc herniations. However, CT scans are still necessary in some cases, as the spatial resolution of MRI is often insufficient due to the large thickness of the slices [5].

Surgical precision requires a level of anatomic accuracy that MRI often struggles to achieve due to its limitations in spatial resolution. Consequently, it is crucial to enhance the spatial resolution of MRI to bring it closer to that of CT, enabling more detailed visualisation of anatomical structures and improving diagnostic accuracy. By doing so, the need for additional CT scans could be minimised or even eliminated. This is particularly advantageous, as CT scans expose patients to ionising radiation, which carries health risks, especially with repeated exposure. Reducing reliance on CT scans offers a safer

diagnostic process, particularly for patients requiring multiple imaging sessions.

Significant efforts have been made to achieve higher resolution MRI images. From a hardware point of view, the acquisition of MRI images with higher spatial resolution is limited by the trade-off between acquisition time, motion artefacts and overall image quality. Super-Resolution (SR) image reconstruction has emerged as a promising approach to address this challenge. This technique involves generating a High Resolution (HR) image from one or more Low Resolution (LR) images. Recently, Deep Learning (DL)-based techniques have gained widespread use in solving SR problems by extracting detailed information from LR images to produce clearer high-quality images. [6].

1.2 Objectives

The primary aim of this study is to enhance the quality and resolution of anatomical MRI images of the spine. To achieve this, the study seeks to implement a DL algorithm that outperforms traditional SR methods. The goal is to improve image sharpness, enabling the clear visualisation of intricate details that may not be fully discernible in the original images. This, in turn, will provide clinicians with a more comprehensive understanding of spinal lesions and assist in more informed surgical planning.

This overarching aim is broken down into specific objectives:

- Compare different Artificial Neural Network (ANN) architectures;
- Evaluate the performance of the DL models against traditional Bicubic Interpolation (BCI);
- Apply the best-performing DL model, trained on LR images, to an already high-quality MRI image, in order to assess the model's maximum potential for enhancing image resolution;
- Test the generalising ability of the best-performing model across different datasets and different types of image degradation schemes.

The research work described in this dissertation was carried out in accordance with the norms established in the ethics code of Universidade Nova de Lisboa. The work described and the material presented in this dissertation, with the exceptions clearly indicated, constitute original work carried out by the author.

2

Theoretical Concepts

2.1 Magnetic Resonance Imaging

MRI is a non-invasive technique that explores the interaction between applied fields and nuclear magnetic moments [7].

Spin, an intrinsic property of particles, varies according the particle type, being ½ for electrons, protons and neutrons. In turn, hydrogen nuclei, with one proton, is one of the most abundant biological elements in the human body [7]. As it has a positive charge, the moving proton generates a small field known as its magnetic moment [8].

MRI phenomena can be described from a quantum perspective, considering each proton's behaviour, as well as from a classical approach, accounting for a group of nuclei. From a classical standpoint, a proton in a uniform external magnetic field B_0 undergoes a force moment, causing it to precess around B_0. The precession frequency, proportional to B_0, is determined by the Larmor Equation (2.1), where γ represents the gyromagnetic constant [8].

$$\omega = \gamma \times B_0 \tag{2.1}$$

The magnetisation M_z, shown in Figure 2.1, represents the sum of all the spins, initially aligned along the z-axis, parallel to B_0. However, to measure the body's magnetisation, detectors are used in the transverse plane. By applying a radiofrequency pulse B_1 at 90°, perpendicular to B_0 and at the Larmor frequency, M_z shifts from the z-axis into the transverse plane. Since the pulse maintains spin coherence, its deactivation leads to spin dispersion, resulting in an exponentially decreasing signal amplitude and generating the Free Induction Decay (FID). The flip angle can be altered depending on the time of application or the intensity of the radiofrequency pulse [8].

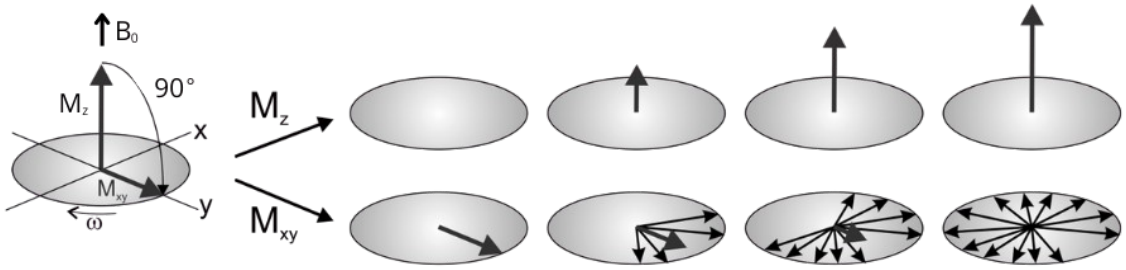

Figure 2.1: Vector representation of the MRI phenomenon. The top row illustrates the longitudinal relaxation, showing the recovery of the magnetisation component along the z-axis (M_z). The bottom row represents the transverse relaxation, showing the dephasing in the xy-plane, leading to the decay of the magnetisation vector M_{xy}. Retrieved from [9].

There are two types of relaxation processes: longitudinal relaxation, characterised by T_1 time corresponding for the recovery of magnetisation along the z-axis, and transverse relaxation, characterised by T_2 time for spin dephasing [8]. Also, T_2^* relaxation is related to T_2 but, in addition to spin-spin interactions, is influenced by field heterogeneity, tissue susceptibility, and spin diffusion, resulting in shorter timescales [10].

MRI doesn't directly measure FID but detects the echoes generated during the imaging process. These echoes arise from the rephasing of spins following the initial radiofrequency pulse, allowing for the distinction and imaging of different tissue characteristics. There are two main ways to achieve those echoes: spin echo, which yields higher-quality image, and gradient echo, known for its shorter acquisition time [8].

2.2 Spatial Resolution

Spatial resolution refers to the sharpness with which an object is represented, determined by the number of pixels in the image: higher resolution means more pixels within the same Field of View (FOV), the observable area that can be captured by the imaging system, leading to a more detailed image. This makes spatial resolution a key measure of imaging system quality [6].

One common way to assess this quality is by measuring the Full With at Half Maximum (FWHM) of the Point Spread Function (PSF), which is the width between points where intensity drops to half its maximum value (see Figure 2.2). Ideally, an object point should be represented by a single point in the image, but due to system imperfections, this point "spreads" into an intensity distribution around it, defined by the PSF. The relationship between the captured image I(x,y,z) and the object O(x,y,z) can be expressed by Equation (2.2), where $*$ represents the convolution operator and h(x,y,z) is the Three Dimensional (3D) PSF [6].

$$I(x, y, z) = O(x, y, z) * h(x, y, z) \tag{2.2}$$

Spatial resolution is defined by the smallest separation between two point sources that allows them to be distinguished by the imaging system. If these sources are too close,

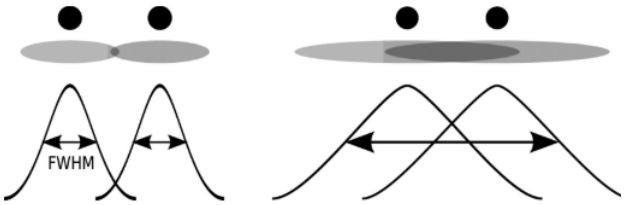

Figure 2.2: Demonstration of spatial resolution with resolvable and non-resolvable point sources based on FWHM. Retrieved from [6].

they may appear as a single blurred source. Two point sources can be resolved if their separation exceeds the FWHM of the PSF. In Figure 2.2, the left side shows two resolvable sources, while on the right, they appear indistinguishable [6].

The PSF is influenced by the number of data points acquired and the FOV. To capture the full frequency content of an object, the sampling rate must satisfy the Nyquist theorem, which requires it to be at least twice the maximum frequency present in the object. This prevents aliasing, where higher frequencies are misinterpreted as lower ones. The width of the PSF is inversely proportional to the number of samples, so the sampling frequency must be high enough to capture fine details and ensure good resolution, avoiding artifacts like data truncation. However, excessively high sampling rates can significantly increase acquisition time allowing motion artifacts [11].

2.3 Enhancing Spatial Resolution in Magnetic Resonance Imaging

In MRI, in addition to factors such as the sampling rate, acquisition time, and moving subjects artefacts already mentioned, the PSF is also influenced by the T_2* time. A longer T_2^* maintains the coherence of the spins for much longer, allowing for a more gradual signal decay during acquisition. In contrast, a low T2* results in a rapidly decaying signal leading to a wider PSF and lower resolution [6].

To further enhance MRI resolution, advancements in hardware are essential, particularly through the increase of magnetic field intensity. Elevating the magnetic field strength B_0 amplifies the magnetization M_0, which in turn generates a stronger signal and improves Signal-to-Noise Ratio (SNR). Another method is increasing the number of receiver coil channels (often referred to as antennas) to cover more volume, enhancing sensitivity and signal quality [6]. Moreover, using smaller coils positioned closer to the area of interest allows the capture of more localised signals while minimizing noise from surrounding tissues, which further boosts the SNR [12].

Increasing resolution through hardware has its advantages, but it also presents challenges. Stronger magnetic fields require more complex infrastructures and higher costs. In addition, the use of dedicated radiofrequency coils can lead to image artefacts and pose a risk of burns due to radiofrequency heating. In contrast, SR methods aim at reconstructing

HR images, with improved SNR from lower quality scans, effectively overcoming these issues [6].

2.4 Super-Resolution Problem

SR is mostly achieved through two methods: Single Image Super-Resolution (SISR) and Multi-Image Super-Resolution (MISR). SISR uses a single LR image to generate a HR image, while MISR combines several LR images of the same object, obtained through different degradation methods and possibly shifted by sub-pixels, to create the HR image. MISR's advantage is that each LR image captures different details, leading to better performance. However, MISR requires more storage, memory, and computational power, making it costly and often impractical. Although SISR is more challenging, it is more feasible in real-world applications due to its simplicity and lower resource demands [13] [14].

2.4.1 Formulation of Single-Image Super-Resolution Problem

The SR problem can be formulated using a linear image acquisition model, represented by Equation (2.3), where the LR image $Y \in \mathbb{R}^n$ is obtained from the HR image $X \in \mathbb{R}^m$ through blurring, downsampling, and the addition of zero-mean Gaussian noise during acquisition. The operator $B \in \mathbb{R}^{m \times m}$ represents the blur matrix modelling the PSF, while $D : \mathbb{R}^m \rightarrow \mathbb{R}^n$ represents downsampling from HR to LR. N denotes the additive noise and $W = D \downarrow B \in \mathbb{R}^{n \times m}$ $(m > n)$ is the global transformation operator encompassing both blur and downsampling [15].

$$Y = D \downarrow BX + N = WX + N \tag{2.3}$$

The aim of SR is to reverse this process and recover the original HR image X free from noise or blur. A common method involves defining the matrix W^{-1} as a combination of a restoration operator $F \in \mathbb{R}^{m \times m}$, which aims to recover lost details, and the magnification operator $S : \mathbb{R}^n \rightarrow \mathbb{R}^m$, which scales up the LR image Y to enhance spatial resolution. This approach is formalised by Equation (2.4), where Z is the high-resolution interpolated image $Z = S \uparrow Y$ and $\hat{X}$ the super-resolved image [16].

$$\hat{X} = W^{-1}(Y) = F(Z) = F(S \uparrow Y) \tag{2.4}$$

Given a training dataset made of pairs of HR/LR images, the DL approach to solving the SR problem involves learning W^{-1} so that the super-resolved images can be predicted. With $\mathcal{L}$ as the loss function used to optimise the network parameters, W^{-1} is learned by solving the expression (2.5), where Θ is the set of possible parameters of the network and $G(\theta)$ is a regularisation function [13]. The aim is to minimise the difference between HR image, denoted as X, and the estimated super-resolved image, denoted as $\hat{X}$ [13].

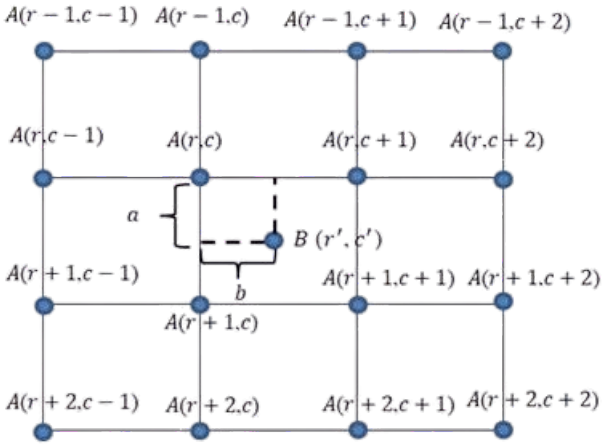

Figure 2.3: Diagram of BCI algorithm. Retrieved from [20].

$$\arg \min_{\theta \in \Theta} [\mathcal{L}(X, \hat{X}) + G(\theta)] \tag{2.5}$$

2.4.2 Bicubic Interpolation

BCI is a reference algorithm used to both downsample and upsample images, reducing and increasing their resolution, respectively. Also known as resampling, this method estimates new pixel intensity values based on the pixels in the original image. The result is a smooth image, although less smooth than the image obtained by Bilinear Interpolation (BLI), and with reduced aliasing effects, compared to Nearest Neighbour Interpolation (NNI) [17].

BCI can be implemented using polynomial, cubic splines or cubic convolution algorithms. During the BCI process (see Figure 2.3) the values of new pixels, B, are calculated using the weighted sum of the 16 neighboring pixels arranged in a 4×4 grid, as described in Equation (2.6). In this equation, $B(r', c')$ represents the interpolated pixel value at the new coordinates r' and c', a_{ij} denotes the intensity value of the pixel located at position (i, j) within the neighboring 4×4 grid, represented in Figure 2.3 by the grid A, and W_{ij} are the corresponding weighting coefficients. The influence of each neighbouring pixel on the interpolated value is determined by its geometric distance, x, from the newly interpolated pixel [18]. This relationship is calculated using bicubic spline interpolation, as described in Equation (2.7), where the constant α is typically set to either -0.5 or -0.75 [19].

$$B(r', c') = \sum_{i=0}^{3} \sum_{j=0}^{3} W_{ij} a_{ij} \tag{2.6}$$

$$W(x) = \begin{cases} (\alpha + 2)|x|^3 - (\alpha + 3)|x|^2 + 1 & \text{for } |x| \leq 1 \\ \alpha|x|^3 - 5\alpha|x|^2 + 8\alpha|x| - 4\alpha & \text{for } 1 < |x| < 2 \\ 0 & \text{otherwise} \end{cases} \tag{2.7}$$

In a matrix-based approach, the Equations 2.8 and 2.9 represent the horizontal, H, and the vertical, V, weight vectors of the interpolation kernel. These vectors are calculated based on the horizontal and vertical distances from the interpolated point $B(r', c')$ to each column and row, respectively, of matrix A. Using these weight vectors, the interpolated value of the subpixel B_{subpixel} is then determined by multiplying the vectors H and V with matrix A, as shown in Equation (2.10) [21].

$$H = (W(1+b), W(b), W(1-b), W(2-b)) \tag{2.8}$$

$$V = (W(1+a), W(a), W(1-a), W(2-a))^T \tag{2.9}$$

$$B_{subpix} = HAV \tag{2.10}$$

2.5 Artificial Intelligence

Artificial Intelligence (AI) is a field within computer science dedicated to developing intelligent machines capable of learning and problem-solving, such as recognizing patterns and relationships within data. These systems are autonomous and adapt dynamically as more data becomes available, similar to human intelligence.

Machine Learning (ML), a subfield of AI, focuses on creating algorithms that enable computers to learn and improve their performance on tasks through experience, without the need for explicit programming [22]. ML systems are generally divided into four categories, based on the nature of supervision provided during training. In supervised learning, the data used for training is labelled, and the algorithm adjusts its predictions to minimise errors against the known outcomes. In contrast, unsupervised learning involves algorithms discovering patterns and structures within unlabelled data. A third approach, known as semi-supervised learning, combines both labelled (in smaller quantities) and unlabelled (in larger quantities) data to enhance learning efficiency. Finally, in reinforcement learning, an agent learns by interacting with its environment, aiming to maximise cumulative rewards, by seeking positive outcomes (rewards) or avoiding negative ones (penalties) [23]. This study will focus on supervised learning, where the objective is to minimise the difference between the super-resolved estimate and the HR ground truth images.

DL is a subfield of ML that leverages ANNs to address more complex problems, often outperforming other ML algorithms. Its efficiency improves with the amount of training data available [23].

2.5.1 Artificial Neural Network

An ANN operates through networks of artificial neurons, which are inspired by the human brain and designed to progressively extract higher-level features from data. These

artificial neurons are computational units that calculate the weighted sum of inputs and apply a non-linear activation function to produce the output [23]. The activation function, denoted as f, processes the weighted inputs and bias, b, and determine whether a neuron should be activated, effectively deciding the relevance of the information. By introducing non-linearity to the network, activation functions enable the model to capture complex patterns in the data. These functions can vary between layers, with common examples including sigmoid, softmax, and rectified linear unit (ReLU). Mathematically, given an i^{th} layer, its output y_i is given by Equation (2.11), where W_i is the weight matrix for the i^{th} layer and, y_{i-1} represents the output of the previous layer [23]:

$$y_i = f(W_i y_{i-1} + b_i) \tag{2.11}$$

A fully connected, or dense layer, is one where every neuron is connected to all neurons in the previous layer [23]. A DL network consists of multiple fully connected layers, with each node containing information from the preceding layers. The data enters through the input layer and passes through hidden layers, which transform the information into different levels of abstraction. The output layer then delivers the final result based on what was learned in the hidden layers. This process, known as feed-forward, directs information from the input to the output layer [24].

A neural network operates in two phases: training and testing. During training, the goal is to minimise or maximise the objective function, which evaluates the network's performance. When the objective function measures the error between the network's predictions and the actual values, it is often called the loss or cost function. The closer this error is to zero, the better the network's performance [23].

The optimal loss value is obtained by finding the best combination of the network's parameters. In a deep neural network, this is done by computing the gradient of the loss function with respect to each parameter, which shows how to adjust them to minimise the loss. This process uses Gradient Descent technique, where parameters are updated in small steps to minimise the loss. The size of these steps is determined by the learning rate, a user-defined hyper-parameter [23].

The learning cycle stops when either the minimum loss value is achieved or the maximum number of epochs, complete iterations over the entire training set, is reached. An appropriate learning rate is crucial for effective training. If the learning rate is too low, the model may get stuck in local minima, slowing down convergence. Conversely, if it's too high, training may become unstable, leading to oscillations or divergence [23].

To properly evaluate a model and avoid overfitting, where the model becomes overly attuned to the training data, including its noise, and thus struggles to generalise to new data, the training data is further divided into training and validation sets. The training data is used to adjust the model's parameters, while validation data offers an unbiased evaluation during training. The test data, unseen during training, is reserved for final assessment once training is complete [23].

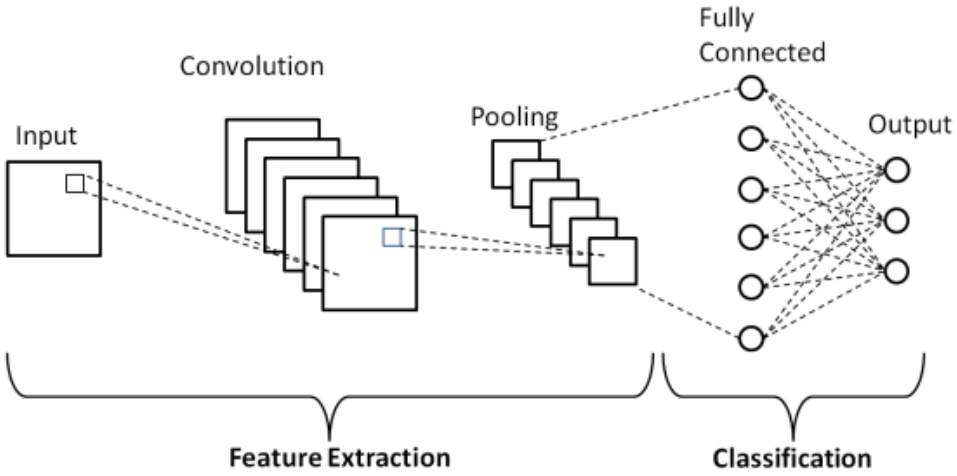

Figure 2.4: Diagram of a CNN. Retrieved from [26].

2.5.2 Convolutional Neural Network

A Convolutional Neural Network (CNN), illustrated in Figure 2.4, is a specialised neural network that efficiently captures spatial correlations between nearby data points through a set of connections between layers. The convolutional layer represents the fundamental building block of a CNN, where it learns feature representations by applying a sliding window, often referred to as a filter or kernel, to the input image. Each convolutional layer uses multiple trainable filters of the same size to convolve the input, producing features maps that highlight specific patterns like edges and shapes. As the number of convolutional layers increases, the network can detect more abstract features, such as textures and complete shapes. During training, those filter coefficients are adjusted, based on the output error, through backpropagation [23] [25].

While deeper networks generally offer better performance, they also have more parameters, leading to increased computational demands. To mitigate this, pooling layers reduce the feature maps dimensionality, while retaining the most significant features, which also helps to prevent overfitting. The output of the final pooling layer is then processed by the fully connected layer, which serves to map the extracted features into the final output space. This layer essentially acts as a classifier or regressor, depending on the task, by combining and weighting the features learned in the previous layers to produce the final prediction or classification [23].

2.5.3 Generative Adversarial Network

Generative Adversarial Networks (GANs), shown in Figure 2.5, consist of two interconnected neural networks: the generator and the discriminator. The generator takes the input from a latent space, which is typically random noise, and generates data that aims to mimic real data as closely as possible. The discriminator, on the other hand, evaluates the generated data to determine whether it is real or synthetic. During training, the

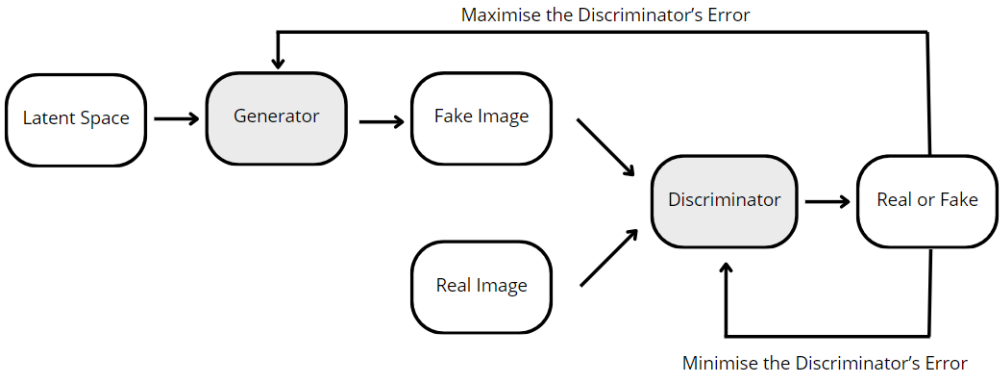

Figure 2.5: Illustration of the GAN training process.

generator and discriminator engage in a competitive process. The generator aims to create increasingly realistic data to fool the discriminator, while the discriminator improves its ability to differentiate between real and generated data. The generator adjusts its weights based on the feedback from the discriminator. If the discriminator identifies the generated data as fake, the generator updates its output in the next iteration to more closely resemble the real data. Meanwhile, the discriminator also updates its weights to enhance its ability to distinguish between real and generated data, aiming to output values near to 1 for real data and close to 0 for generated data [27].

Typically, both generator and discriminator use cross-entropy as their loss function to adjust the weights and optimise performance [27]. Cross-entropy can be calculated using Equation (2.12),

$$H(q, p) = -(q \log(p) + (1 - q) \log(1 - p)) \tag{2.12}$$

where $H(q, p)$ is the value of the cross-entropy loss, q is the true label (0 or 1), and p is the probability that the instance belongs to the positive class, 1 [27].

The GAN loss, $\mathcal{L}_{GAN}$, which combines the generator, $\mathcal{L}_G$, and discriminator, $\mathcal{L}_D$, losses, is often referred to as the min-max loss, as shown in Equation (2.13). This term reflects the adversarial nature of GAN training, where the discriminator aims to maximise its performance by correctly distinguishing between real and fake data, while the generator seeks to reduce the discriminator's effectiveness by producing the most realistic data possible. [27].

$$\mathcal{L}_{GAN} = \min_G \max_D \mathcal{L}_D + \mathcal{L}_G \tag{2.13}$$

3

Literature Review

3.1 Conventional Methods for Image Super-Resolution

Conventional SR algorithms can be classified into three categories: interpolation-based, reconstruction-based and learning-based.

Among interpolation methods, Jiang et al. [28] used the NNI to perform image scaling. However, the authors conclude that this interpolation can generate unwanted artifacts, recommending BLI or BCI for scaling tasks. In turn, Wang et al. [29] [30] showed that BLI and BCI can generate zigzag artefacts. To resolve this, they proposed a parallelogram-shaped interpolation kernel and suggested improvements in the discretisation of the angle of these kernels to improve the model's performance. To reduce staircase artefacts, Suresha et al. [31] applied an adaptive BLI with 7 to 16 neighbours. The generation of the LR image and the denoising performed after upsampling were done in the wavelet domain, and the authors emphasise the importance of this domain in recovering edges. The proposed algorithm proved to be superior to NNI, BLI and BCI methods.

Srilakshmi et al. [18] increased the resolution of MRI images by initially generating LR versions via intensity scaling. Then they applied histogram equalisation, adaptative histogram equalisation CLAHE to improve global and local contrast, and cubic B-spline interpolation. Similarly, Bharati et al. [32] proposed a reconstruction model based on cubic B-spline interpolation to improve image compression and increase information recovery after decompression. The study highlighted that Peak Signal-to-Noise Ratio (PSNR) can be improved with higher order splines and, also, that the SR process can be optimised in the wavelet domain. Lehmann et al. [33], likewise, showed that higher degree B-splines offer greater spatial accuracy and better preserve high Fourier frequencies. However, they warn that large kernels can increase error by preserving the aliasing generated by downsampling. Therefore, to avoid aliasing artefacts, they recommend smoothing the high frequencies before downsampling. Compared to cubic interpolation, Blu et al. [34] demonstrated that shifting the sampling points of linear splines by $1/5$ of the distance between them in the standard uniform spacing results in less smooth images with reduced detail loss.

Most authors conclude that interpolation methods fail to restore high frequencies, making images excessively smooth, as they fail to increase the amount of information. To address this limitation, Irani et al. [35] [36] [37] proposed the Iterative Back Projection (IBP) algorithm, which starts with an estimate of the HR image and applies the image acquisition model to generate LR images that correspond to those observed. In an iterative process, the error between the real LR image and the one generated by the model is minimised by updating the HR image estimate using gradient-based techniques. Although simple to implement, IBP can converge to sub-optimal solutions if the initial estimate is far from the ideal solution. To overcome this limitation, Nayak et al. [38] proposed a meta-heuristic optimisation technique based on cuckoo search, which demonstrated superior results to the gradient method.

Still within reconstruction methods, probabilistic algorithms such as the recursive Baysian algorithm proposed by Alshammri et al. [39] are widely used. This method is based on registration for motion estimation, interpolation for upscaling, restoration to remove noise and blur, and histogram adjustment to improve the distribution of intensities. Although good results were observed, particularly in cases where the levels of noise, blur and motion are unknown, the authors add that the method should be improved using DL techniques.

Machine learning techniques, especially those based on sparse modeling and dictionary learning, have been essential in super-resolving images, significantly improving their quality and resolution.

Sparse modelling is a technique for describing both local smoothness and non-local self-similarities in an image. Dian et al. [40] advanced this concept by introducing a sparse representation algorithm combined with non-local clustering using Kmeans, effectively improving the efficiency of the sparse solution by exploiting similar patterns and structures in LR images. In the same vein, Mikaeli et al. [41] proposed the Adaptive Group-based Domain Selection technique, which refines this approach by grouping similar patches and adaptively selecting sub-dictionaries, considering both local smoothness and self-similarity. Based on these fundamental ideas, Barman et al. [42] developed an algorithm for real-time SR, ideal for medical images, that takes advantage of sparse representations by first calculating the sparse coding with a fixed dictionary, followed by adapting this dictionary to the specific features in the LR image, and incorporating a new feature extraction step to improve accuracy.

Also, Yan et al. [43] developed a SR technique based on dictionary learning that aimed to increase the spectral resolution of multispectral images, bringing them closer to the spatial resolution of hyperspectral images. The proposed algorithm only requires an multispectral image and a spectral library, instead of pre-existing hyperspectral images. After merging the multispectral with the spectral information, dictionaries are learnt to capture the main features, and adaptive sparse coefficients are used to determine the contribution of each dictionary. The authors suggest that the generality of the model can be improved with DL techniques.

Techniques based on dictionary learning improve resolution but face high computational complexity, particularly with large dictionaries or images, limiting their use in real-time medical imaging. DL offers a more efficient alternative by learning hierarchical features and scaling better with large datasets. Despite its computational demands, DL provides faster inference, making it more suitable for real-time applications.

3.2 Deep Learning Methods for Image Super-Resolution

More recently, deep neural networks have significantly boosted the field of image SR, with several innovative approaches being developed to overcome the limitations of traditional methods and improve the quality of the HR images generated.

The first DL method for SR process was proposed by Dong et al. [44] called Super-Resolution Neural Network (SRCNN). This method is an end-to-end CNN with three convolutional layers: the first extracts patches from the LR image and represents them in a high-dimensional vector; the second maps these vectors to another high-dimensional representation; and the last layer reconstructs the super-resolved image. Although SRCNN outperforms traditional methods, it is often criticised for generating blurred edges due to its surface structure [45] [46] [47]. To address these limitations, Kim et al. [48] incorporated a skip connection, which passes the input directly to later layers without modification, helping to preserve information and ensure network stability while enabling global residual learning. Combined with gradient clipping, this strategy enables the use of a higher learning rate, leading to faster convergence and improved deeper networks performance.

Increasing the depth of neural networks improves performance, but can cause overfitting and make storage difficult. The Deeply-Recursive Convolutional Network [49] solves this by using a recursive structure, which increases the depth without adding parameters. To mitigate gradient problems, a recursive supervision and a skip connection are applied, which reduce the need for extra recursive layers. However, this requires more memory to store the intermediate outputs. Building on this approach, Tai et al. [50] proposed a recursive block with residual units sharing weights to further facilitate training and address gradient issues. The suggested framework not only supports local residual learning but also benefits from global residual learning through the skip connection.

There are various approaches to upsampling in neural networks. In that context, Zhang et al. [51] improved the SRCNN [44] by adding an initial layer to perform BCI. Alternatively, Lin et al. [52] used four consecutive deconvolution layers for the same purpose, although this increased the computational cost and reduced the efficiency of the network. Therefore, to improve the efficiency of the SRCNN [44], Dong et al. [53] proposed the Fast Super-Resolution Convolutional Neural Network (FSRCNN), which includes a deconvolution layer at the end of the network for upsampling, a reduction layer to decrease the dimensions of the features before mapping and an expansion layer

to increase the dimensions of the output features. These changes reduced the number of parameters while maintaining performance.

On the other hand, Shi et al. [54] introduced the sub-pixel convolutional layer, which performs convolutions in LR space, activating different parts of the filter according to the location of the sub-pixels. The checkerboard artefacts caused by the shuffle process, in the convolutional layer, were mitigated by an improved NNI approach.

Dong et al. [53] and Shi et al. [54] were pioneers in integrating the upsampling operation into the network, an approach that is now common to improve efficiency in super-resolution tasks.

Haris et al. [55] proposed the idea of alternating up and down sampling, where after each projection the error is calculated by comparing the intermediate result with the map of the features from the previous stage. Li et al. [56] advanced by integrating feedback blocks made up of convolutional and deconvolutional layers with dense connections into a residual network. In addition, the authors employ a curriculum learning strategy to generate LR images from a complex degradation model, improving the recovery of high-frequency details.

Huang et al. [57] introduced a densely connected CNN with skip connections, connecting the output of each block to all previous layers within the block. This approach of concatenating features rather than adding them together optimises the flow of information and gradients, resulting in a more efficient network with improved performance and no overfitting. Building on this concept, Zhang et al. [58] presented the Residual Dense Block, which refines the idea of dense connections by incorporating local residual fusion and global residual learning, better preserving information and resulting in sharper details and reduced blur artefacts.

In addition to these methods, Lai et al. [59] introduced the progressive upsampling idea. The network proprosed features two branches: one for gradual resolution enhancement via convolutional and transposed convolutional layers, and another for combining low and high-frequency information for image reconstruction.

Another type of model widely used in SR are the GANs, which, although they can reduce PSNR, they excel in perceptual quality and high-frequency details preservation.

The Super-Resolution GAN (SRGAN) proposed by Ledig et al. [60] introduces a total loss as the weighted combination of the content loss, composed of Mean Squared Error (MSE) and VGG loss, and the adversarial loss. The authors found an improvement in perceptual quality. Wang et al. [61] improved this approach by incorporating dense blocks with residual connections, without batch normalisation, to facilitate training. They also introduced a relativistic discriminator that compares the probability of a real image being more realistic than a generated image, rather than simply classifying images as real or fake. To overcome situations where it is not possible to obtain LR/HR pairs of images or when the image degradation process is unknown, Yuan et al. [62] proposed the CycleGAN.

Recently, various DL methods have proved effective in super-resolving medical images, such as MRI and CT.

Liu et al. [63] compared the performance of several state-of-the-art image SR models in super-resolving MRI images. Of all the models evaluated, the research article showed that SRGAN [59] is the one that provides the greatest visual richness. Complementing these state-of-the-art architectures, Mithra K K et al. [64] propose an approach based on texture transfer. Feature extraction and comparison is carried out in the wavelet domain, where LR textures are replaced by the corresponding HR textures. A texture loss is also proposed. The results indicate that axial texture transfer is superior to sagittal, suggesting that the HR image's axial textures were richer.

With the same objective of enhancing the resolution of MRI images and also mitigating aliasing, Zhao et al. [65] developed two different algorithms. The 3D algorithm employs zero-fill in k-space and uses the Fermi anti-ringing filter to blur the x-axis, subsequently feeding LR/HR pairs into the network proposed by Lime et al. [66], eschewing the upsampling layer. This was selected due to its capacity for local learning. The 2D algorithm applies a Gaussian filter and introduces aliasing through downsampling and upsampling, and then uses a self-supervised anti-aliasing network. The performance of the algorithms was evaluated in the presence of Rician noise, demonstrating that varying noise levels have a significant impact on their effectiveness.

To mitigate degradation problems and explosive gradients in deep networks, Zhu et al. [67] develop a dense residual network by proposing multi-residual blocks that facilitate information preservation.

Dual-domain methods, such as the one proposed by Li et al. [68], enhance SR by combining learning in the frequency and image domains. Their approach uses cross-attention to transfer high-frequency information from the T_1 modality to T_2, while implicit attention handles upsampling at various scales. Similarly, Xiao et al. [69] applied the Fourier Transform for frequency domain operations and a residual U-net for image domain learning, focusing on adversarial, perceptual, and frequency losses. Eo et al. [70] and Liu et al. [71] further explore both domains to reconstruct fully sampled images from undersampled k-space data, enhancing reconstruction quality through cross-domain interaction.

In essence, traditional SR techniques improve image quality but are often limited by computational inefficiencies, especially with larger datasets or in real-time applications. While DL models also require significant computational resources, they offer better scalability and automatic feature learning, leading to more efficient performance and superior results. This shift has greatly enhanced the precision and quality of image SR, making DL crucial for medical imaging. In this work, FSRCNN [53], Residual Dense Network (RDN) [58], and Enhanced Super-Resolution Generative Adversarial Network (ESRGAN) [61] were selected for their proven ability to efficiently generate HR outputs. FSRCNN's lightweight design is ideal for real-time tasks, RDN balances quality and speed by using dense residual connections, and ESRGAN excels at producing sharp, high-fidelity images, perfect for tasks requiring detailed textures. These models were chosen for their strengths in addressing key SR challenges, combining efficiency with high-quality results.

4

Materials and Methods

The methodology of this study is organised into three stages. The first stage involves selecting the most appropriate datasets for training and testing the DL models. The second focuses on implementing the SR methods. The third stage consists of a quantitative evaluation of the various networks, comparing their performances with one another and against the standard BCI method. This stage also includes the application of the best-performing model to the original MRI image and an analysis of its generalisability across different datasets and image degradation methods.

4.1 Datasets

The MRI data used for training and testing the different models was obtained from the publicly accessible repository *Mendeley Data*, specifically from the set of data entitled *Lumbar Spine MRI* [72]. For the purposes of this study, T_2-weighted scans acquired using the Turbo Spin Echo sequence in the supine position and sagittal views were used, as these are the most commonly used by clinicians for detecting lumbar disc hernias. The dataset comprises anonymised MRI scans from 515 patients with back pain, encompassing a total of 7,994 image slices. The scans predominantly focus on the three lowest vertebrae and intervertebral discs. Table 4.1 provides an overview of the general characteristics of the MRI slices used in this study.

Table 4.1: General information on the magnetic resonance slices used for training and testing the DL models.

Characteristic	Description
Slice thickness	4 mm
Pixel spacing	Uniform
Spatial resolution	320x320 and 310x310 pixels
Intensity resolution	12 bits per pixel

To evaluate the generalisability of the best-performing model, an additional 18 images, from two other datasets were also used. The first dataset comprised CT scans sourced

from the publicly available *Computed Tomography (CT) of Brain - Object Detection* dataset on *Kaggle* [73], which includes scans featuring conditions such as cancer, tumors, and aneurysms. The second dataset consisted of grey-scale natural images, also publicly available, provided by the University of Cape Town [74]. It is worth noting that the images in these two datasets have variable sizes. Examples of images from each modality employed in this study are provided in Appendix A.

4.2 Image Pre-Processing

Image pre-processing is essential to prepare data for efficient integration into a model, thus enhancing training and improving prediction accuracy. For the SR task, a specialised pre-processing method was required: LR images were created from the original HR images through image degradation techniques.

To train FSRCNN and RDN, patches, small rectangular regions of size 33×33 and 66×66, were extracted from both LR and HR images, respectively. Each LR patch was paired with its corresponding HR patch. These patches, essentially cropped portions of the original images, were treated as independent training samples, allowing the models to focus on localised features such as textures, edges, and corners. This patch-based approach enhanced the model's ability to learn from fine details. In order to guarantee the randomness of the training process, the patches were extracted in a non-sequential manner. To train the ESRGAN, the images were first cropped to a uniform dimension of 310×310. After generating the LR/HR image pairs with sizes of 155×155 and 310×310, respectively, no patches were created, and these images were fed directly into the model. The decision to avoid using patches was driven by the substantial computational resources and time required for processing, making this approach more efficient and feasible given the constraints.

Finally, the input LR and their corresponding HR images were remodelled so that they could have the correct tensor structure to be fed into the DL model during training and then stored in an HDF5 file.

4.2.1 Image-Domain Degradation Method

The first step in obtaining LR/HR pairs involved cropping the images to ensure that their height and width were integer multiples of the scale factor, 2. The scale factor dictated the resizing proportion, whether the image was scaled down to half size or scaled up to twice the size. This adjustment was fundamental in generating the smaller LR image from the original HR image whose dimensions were not multiples of 2 (such as with some CT images), ensuring that the scaling process remained consistent across the entire image. By maintaining integer multiples of the scale factor, the integrity of the image structure was preserved, preventing any misalignment that could occur during the resizing process by ensuring that the LR and HR pairs corresponded perfectly at every pixel level. These

images were symmetrically cropped around their centres, ensuring that the resulting HR images were evenly balanced along all edges.

In order to create the LR image at half the size of the HR image, the HR image was resized using a bicubic filter from the *OpenCV* library. Subsequently, random Gaussian noise with a mean of 0 and variance of 0.001 was added to the resized image, resulting in the final LR image. This approach follows the strategy commonly adopted by most of the studies reviewed in the literature.

4.2.2 Frequency-Domain Degradation Method

This method is particularly important in MRI since MRI images are acquired in Fourier space and only later converted to image space. In this method, the HR image is generated in the same way as described above regarding the image-domain degradation method in the section 4.2.1. However, the generation of the LR image involves an additional step. Before downsampling and adding noise, the image is converted to the frequency domain using the Fourier Transform.

In the k-space (frequency domain), two different masks were applied to selectively reduce sharpness by eliminating high frequencies, while preserving image contrast. The first mask, known as the central mask, retains the points within a circle whose radius is 25% of the maximum distance from the centre to the corner of the k-space matrix. This means that all points within this circle are preserved (set to 1), while points outside this region are eliminated (set to 0). This retains the low-frequency components, which contribute to global contrast, but discards the high-frequency components responsible for fine details and sharpness. The second mask, referred to as the radial mask, was designed by combining the concept of radial-based matrices, as seen in previous studies such as the one conducted by Yan et al. [75], with the contrast-preserving properties of the central mask. Specifically, the radial mask also retains the points within a circle whose radius is 25% of the maximum distance from the centre to the corner of the k-space matrix, to preserve low-frequency contrast information, but outside this central region, the mask applies a radial sampling pattern to capture the high-frequency components. Points along specific rays extending outward from the central circle are retained (set to 1), while points in between these rays are eliminated (set to 0). This hybrid design effectively preserves both the low-frequency contrast in the central region and selected high-frequency components in specific directions, mimicking more closely the real-world acquisition process in MRI systems.

Once these masks have been applied in the k-space, the image is transformed back to the spatial domain using the Inverse Fourier Transform. This simulates the undersampling and artefacts typically occurred during the frequency-domain acquisition process in MRI, providing a more realistic degradation method than the classical image-domain approach. The LR/HR image pairs generated using this method were then employed to assess the model's ability to generalise to different types of degradation that more closely resemble

real-world conditions during image acquisition.

4.3 Image Super-Resolution Methods

This study used three DL models for SR, alongside the BCI technique implemented with the *OpenCV* Python library. FSRCNN [53] was selected for real-time efficiency; RDN [58] for balancing quality and speed; and ESRGAN [61] for producing high-fidelity, detailed outputs. These choices align with the goal of improving SR while maintaining computational efficiency. The MRI data was split into 70% for training, 20% for validation, and 10% for testing. The models were implemented with the ML frameworks *TensorFlow* and *Keras*, without batch normalisation layers. This choice was made as batch normalisation layers reduces range flexibility and consumes the same memory as convolutional layers. Thus, by excluding them, larger models could be used. Detailed layer parameters for each DL model are provided in Appendix A.

4.3.1 Fast Super-Resolution Convolutional Neural Network

The FSRCNN [53], Figure 4.1, is divided into five key stages: feature extraction, where a convolutional layer with a kernel size of 5×5 is used to output 56 features maps; shrinking, which reduces the dimensionality of the extracted features through a convolutional layer with a kernel size of 1×1; non-linear mapping, consisting of 12 blocks of convolutional layers that map the extracted features to a HR space; expanding, to reverse the shrinking operation and finally, a deconvolutional layer with stride 2 that performs upsampling to generate the HR image. All convolutions are followed by ReLU activation and the loss function used for training was the MSE.

An early stopping strategy with a patience of 3 epochs was implemented, stopping training if validation performance didn't improve over 3 consecutive epochs, to avoid overfitting. Additionally, the learning rate was reduced by 50% after 3 stagnant epochs with a minimum threshold set at 1×10^{-7}, starting at 1×10^{-4}. This helps the model converge more effectively as training progresses, especially when performance plateaus.

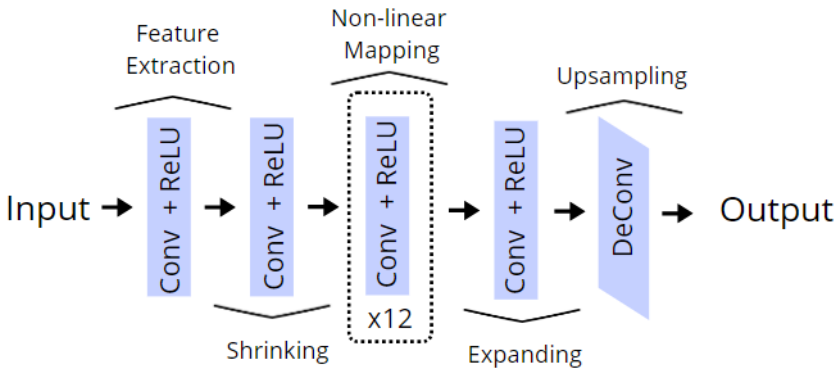

Figure 4.1: Architecture of the FSRCNN.

4.3.2 Residual Dense Network

Residual learning is a technique where the network learns the residual, i.e. the difference between the input and the output, instead of learning their mapping directly. This is done using the residual blocks introduced by He et al. [76], which use skip connections to transfer the information from the input directly to the output, facilitating learning.

The RDN architecture [58], Figure 4.2, combines residual learning with global and local learning strategies (within each patch) to improve feature preservation and performance. The initial layers extract low-level features through convolutions, followed by five Residual Dense Blocks (RDBs), which use dense connections between layers, as shown in Figure 4.3. Local learning occurs within the RDBs, while global learning is ensured by aggregating the outputs of these blocks, which maintains continuous memory. The final upsampling is performed through a transposed convolution with a stride of 2, generating HR outputs.

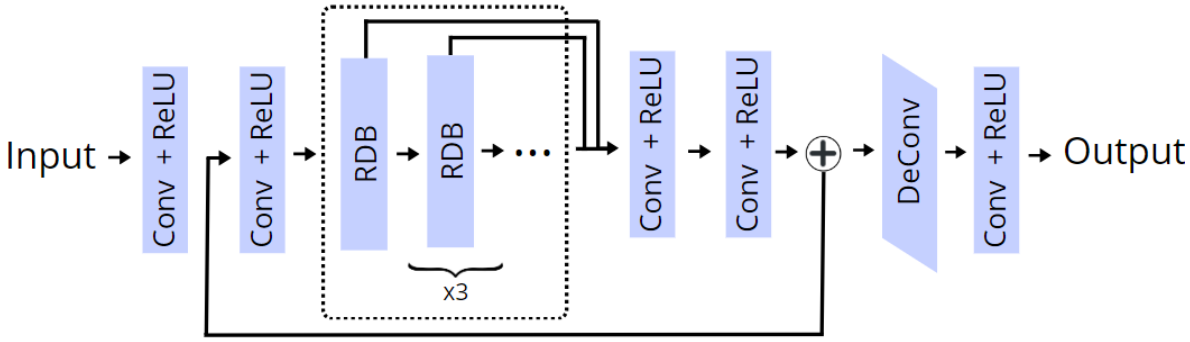

Figure 4.2: Architecture of the RDN.

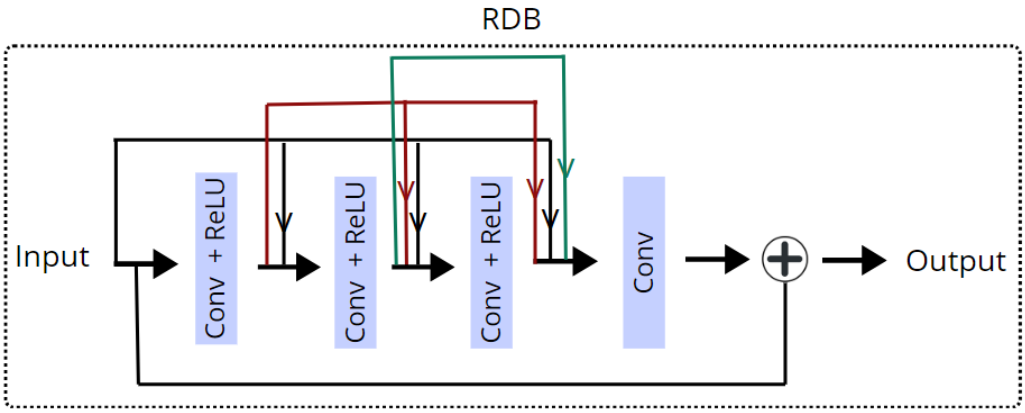

Figure 4.3: Architecture of the RDB of the RDN.

During training, MSE was used as the loss function and the same early stopping and learning rate reduction strategies were used as for the FSRCNN model. However, L2 regularisation was applied, which adds a penalty to large weights, preventing them from dominating predictions. This encourages the network to learn smaller weights, making it more robust and generalisable, while also allowing for faster convergence.

4.3.3 Enhanced Super-Resolution Generative Adversarial Network

The ESRGAN [61] consists of two networks: the generator and the discriminator.

The generator, Figure 4.4, starts by extracting features with a 3×3 convolution. Then the output passes through three Residual in Residual Dense Blocks (RRDBs), which consist of three RDB, Figure 4.5, each with convolutions and internal dense connections. The RDB used in ESRGAN is similar to the RDB in the RDN model described in the section 4.3.2. However, the RDB in ESRGAN includes an additional convolutional layer with ReLU activation. Furthermore, before summing the input of the RDB with the output of the final convolutional layer, a residual scaling factor of 0.2 is applied. This scaling factor diminishes the influence of each RDB, helping to stabilize the training process by preventing large updates and promoting smoother learning. The RRDBs blocks integrate local residual learning, allowing the model to capture fine details. Moreover, the addition of the original input to the final output after the convolution that follows the RRDBs, enables global residual learning, further enhancing feature preservation. A convolution is then applied to increase the number of feature maps. This is essential for the correct operation of the pixel shuffle upsampling technique, which redistributes the channels appropriately, increasing the image's spatial resolution. The upsampling is followed by two more convolutions that refine the final HR image.

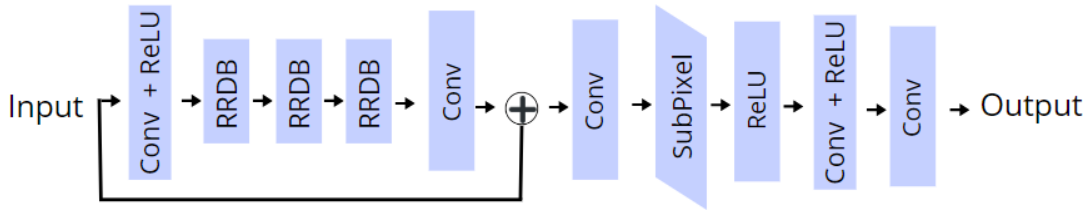

Figure 4.4: Architecture of the generator of the ESRGAN.

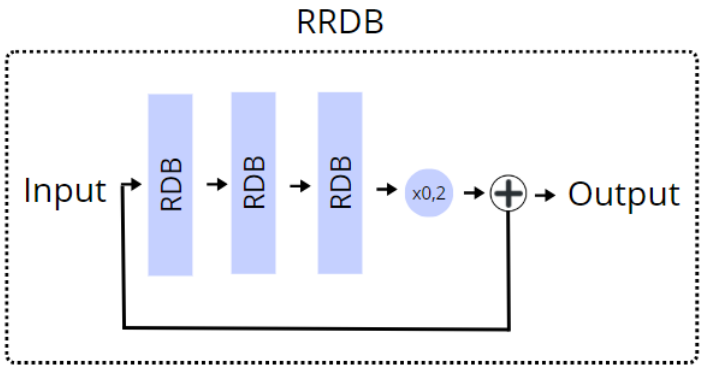

Figure 4.5: Architecture of a RRDB of the ESRGAN.

The discriminator, Figure 4.6, uses multiple convolutional layers with a gradual increase in the number of filters and a reduction in the size of the output by means of strides. After these layers, the network includes a dense layer with 1024 units and a dropout layer to reduce overfitting by randomly deactivating 40% of the neurons during training. Finally,

the output is flattened to a one-dimensional vector and passed through another dense layer that generates a single output.

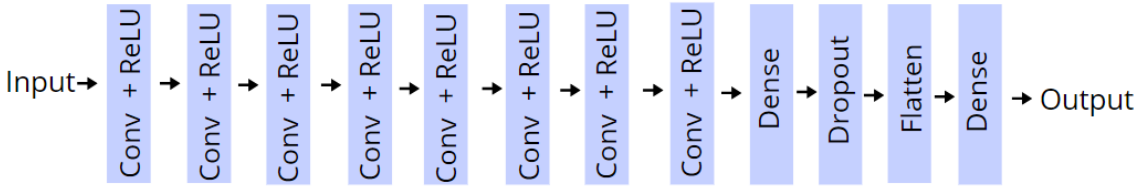

Figure 4.6: Architecture of the discriminator of the ESRGAN.

During training, the discriminator doesn't just assess the probability of an input being real, it determines the probability of a real input being more realistic than a fake input by comparing its output for a real image, $D(\mathbf{x}_{real})$, with the expected value of the outputs for the generated images, $\mathbb{E}_{\mathbf{x}_{fake}}[D(\mathbf{x}_{fake})]$. In other words, the discriminator predicts relative reality rather than an absolute value. Likewise, the output for the generated images, $D(\mathbf{x}_{fake})$, is adjusted by the expected value of the outputs for the real images, $\mathbb{E}_{\mathbf{x}_{real}}[D(\mathbf{x}_{real})]$. The discriminator's total loss $\mathcal{L}_D$, Equation (4.1), is a combination of the losses associated with the real images and the generated images. The discriminator is penalised if it cannot adequately distinguish between real and fake images.

$$\begin{aligned}\mathcal{L}_D = &- \mathbb{E}_{\mathbf{x}_{real}}\left[\log\left(D(\mathbf{x}_{real}) - \mathbb{E}_{\mathbf{x}_{fake}}[D(\mathbf{x}_{fake})]\right)\right] \\ &- \mathbb{E}_{\mathbf{x}_{fake}}\left[\log\left(1 - \left(D(\mathbf{x}_{fake}) - \mathbb{E}_{\mathbf{x}_{real}}[D(\mathbf{x}_{real})]\right)\right)\right]\end{aligned} \tag{4.1}$$

Based on these same principles, the generator's relativistic loss $\mathcal{L}_G$, Equation (4.2), was designed to encourage the generator to produce images that the discriminator is more likely to classify as real. In addition to the relativistic loss, the generator is trained using the MSE loss $\mathcal{L}_{MSE}$, and a perceptual loss $\mathcal{L}_{perc}$, that compares the features extracted from the generated image and the real image using the pre-trained VGG-19, as defined by the Equation (4.3).

$$\begin{aligned}\mathcal{L}_G = &- \mathbb{E}_{\mathbf{x}_{real}}\left[\log\left(1 - \left(D(\mathbf{x}_{real}) - \mathbb{E}_{\mathbf{x}_{fake}}[D(\mathbf{x}_{fake})]\right)\right)\right] \\ &- \mathbb{E}_{\mathbf{x}_{fake}}\left[\log\left(D(\mathbf{x}_{fake}) - \mathbb{E}_{\mathbf{x}_{real}}[D(\mathbf{x}_{real})]\right)\right]\end{aligned} \tag{4.2}$$

$$\mathcal{L}_{G_{total}} = 0.005 \cdot \mathcal{L}_G + 0.01 \cdot \mathcal{L}_{MSE} + \mathcal{L}_{perc} \tag{4.3}$$

4.4 Performance Evaluation

Assessment metrics are vital for evaluating model performance and improvements in image quality. While the human eye can assess quality visually, it lacks objectivity

for comparing methods. Therefore, objective image quality assessment methods are necessary for quantitatively measuring resolution enhancements. The full-reference and free-reference metrics used in this work will be discussed next.

4.4.1 Full-reference Metrics

Full-reference metrics evaluate image quality by directly comparing the super-resolved image to the original HR image, which acts as the ground truth or reference standard.

4.4.1.1 Peak Signal-to-Noise Ratio

The PSNR, expressed in dB, is a widely used metric in SR for assessing the quality of the super-resolved image, $\hat{X}$, relative to the reference HR image, X. It is calculated from the maximum pixel intensity, I (e.g., 255 for 8-bit images), and the MSE between the two images, according to the Equation (4.4) [77]. The open-source image processing library *scikit-image* was used, in order to perform this calculation.

$$\text{PSNR} = 10 \log_{10} \left(\frac{I^2}{\text{MSE}(X,\hat{X})} \right) \tag{4.4}$$

4.4.1.2 Structural Similarity Index

Structural Similarity Index (SSIM) measures the structural similarity between two images and is more aligned with the human visual system, since it incorporates terms that reflect perceptual phenomena such as contrast and luminance [16].

A 7×7 sliding window traverses the images, calculating the SSIM index at each window using Equation (4.5). Here, X and $\hat{X}$ denote the HR reference and super-resolved images, respectively, μ and, σ^2 represent the mean and variance of each image and $\sigma_{X\hat{X}}$ indicates their covariance. The constants C_1 and C_2 stabilise the division. The final SSIM value is the average of the SSIM values from all windows [16]. For this calculation, the *scikit-image* library was also used.

$$\text{SSIM}(X, \hat{X}) = \frac{(2\mu_X\mu_{\hat{X}} + C_1)(2\sigma_{X\hat{X}} + C_2)}{(\mu_X^2 + \mu_{\hat{X}}^2 + C_1)(\sigma_X^2 + \sigma_{\hat{X}}^2 + C_2)} \tag{4.5}$$

4.4.1.3 Gradient Magnitude Similarity Deviation

Gradient Magnitude Similarity Deviation (GMSD) was created to overcome limitations of SSIM in evaluating textures and fine edges, offering greater precision in critical scenarios such as medical images, and standing out for its optimisation and real-time performance [78].

To compute the GMSD, first it is necessary to obtain the gradient magnitudes of the reference image X, denoted by m_X, and that of the super-resolved image $\hat{X}$, denoted by $m_{\hat{X}}$. This is achieved by Equations (4.6) and (4.7), respectively. These gradient magnitudes

are determined by taking the square root of the sum of the squared directional gradients, obtained by applying a 3×3 linear filter, such as the Sobel, along the orthogonal directions, h_x and h_y. In the equations below $*$ denote the convolution operator [78].

$$m_X(i) = \sqrt{(X * h_x)^2(i) + (X * h_y)^2(i)} \tag{4.6}$$

$$m_{\hat{X}}(i) = \sqrt{(\hat{X} * h_x)^2(i) + (\hat{X} * h_y)^2(i)} \tag{4.7}$$

The Gradient Magnitude Similarity (GMS) map assesses the similarity in gradient magnitude between two images, calculated using the Equation (4.8) , where c is a positive constant that stabilises the computation. The maximum value of the GMS is 1, indicating a perfect match between the gradients m_X(i) and $m_{\hat{X}}$(i) [78].

$$\text{GMS}(i) = \frac{2m_X(i)m_{\hat{X}}(i) + c}{m_X(i)^2 + m_{\hat{X}}(i)^2 + c} \tag{4.8}$$

Different structures in an image can undergo varying levels of distortion. For example, flat structures tend to be less impacted by blur compared to textured ones. The GMSD takes these local differences into account by computing the standard deviation of the GMS values across the image. This is represented by Equation (4.9), where N denotes the total number of pixels and $\bar{\text{GMS}}$ denotes the average of the GMS across all pixels [78]. The GMSD calculation was conducted using the Python *sporco* package.

$$\text{GMSD} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(GMS(i) - \bar{\text{GMS}}\right)^2} \tag{4.9}$$

4.4.1.4 Visual Information Fidelity

Visual Information Fidelity (VIF) assesses the perceptual quality of the super-resolved image $\hat{X}$ by comparing the amount of information preserved relative to the reference image X. In the calculation of VIF, the focus is on how much of the original and true information from the reference image is retained in the super-resolved image. Although VIF involves complex Information Theory calculations, it can be simplified as Equation (4.10) [79].

$$\text{VIF} = \frac{\text{Information preserved in } \hat{X}}{\text{Original information in } X} \tag{4.10}$$

VIF considers human visual perception, especially in terms of contrast and noise sensitivity. Both reference and super-resolved images are decomposed into sub-bands of various scales and orientations, to model how the human eye perceives visual information at different levels of detail. The information preserved from the reference image is calculated for these sub-bands and aggregated to determine the overall visual fidelity of the super-resolved image [79]. For this calculation it was used the Python's open source library *sewear*.

4.4.1.5 Spatial Correlation Coefficient

The Spatial Correlation Coefficient (SCC) [80] measures texture and detail preservation by correlating variance between a reference image, X, and a super-resolved image, $\hat{X}$. SCC is computed from local variance, σ^2, and covariance, $\sigma_{X\hat{X}}$, maps derived from high-frequency components of both images, extracted using a 3×3 high-pass filter like the Laplacian. The variance maps are then produced after passing these components through a smoothing filter of size 8×8, such as a moving average, in order to make it possible to measure the degree of dispersion of pixel values around the mean. The SCC is defined through Equation (4.11). For this calculation, the *sewar* library was also used.

$$\text{SCC} = \frac{\sigma_{X\hat{X}}}{\sqrt{\sigma_X^2} \times \sqrt{\sigma_{\hat{X}}^2}} \tag{4.11}$$

4.4.2 Free-reference Metrics

Free-reference metrics assess image quality without the need for a reference HR image, only relying on the analysis of the super-resolved image itself.

4.4.2.1 Perceptual Image Quality Evaluator

Perceptual Image Quality Evaluator (PIQE) was calculated using the *Python Image Processing and Quantitative Evaluation* (*pypiqe*) library. The algorithm begins by computing the Mean Subtracted Contrast Normalised coefficient for each pixel, which measures how pixel values deviate from the local mean, normalised by the local variation (contrast), following the method proposed by Venkatanath et al. [81]. The image is then split into non-overlapping 16×16 pixel blocks. In each block, it assesses distortions, such as blocking artefacts and noise, through the aforementioned coefficients, identifying blocks with high variance. Then, the algorithm aggregates this information to determine the overall image quality, generating the final PIQE score (the average of the block scores). Additionally, masks are created to highlight regions with potential distortions, noticeable artefacts, and Gaussian noise. Areas of the Activity mask that are white, where pixels have a value of 1, correspond to regions with higher variability due to the presence of artefacts and noise. Conversely, black areas, with pixels set to 0, signify regions of low activity, where little to no distortions are present [82].

4.4.2.2 Signal-to-Noise Ratio

The SNR measures image quality by comparing the strength of the signal with the variation in noise. It is calculated as the ratio between the average pixel intensity, μ_{signal}, and the standard deviation of pixel intensities, σ_{noise}, as shown by Equation (4.12) [83].

$$\text{SNR} = \frac{\mu_{\text{signal}}}{\sigma_{\text{noise}}} \tag{4.12}$$

5

Results and Discussion

5.1 Comparative Analysis of DL Models and BCI

Figure 5.1 shows a region of the input LR image, generated by the image-domain degradation method, alongside its ground truth HR image. The LR image is pixelated with less sharpness, especially around the vertebrae, while the HR image has much higher resolution, providing clearer details. The goal was to estimate a super-resolved image that closely matched the HR ground truth, using only the LR image.

Figure 5.1: Comparison of input LR image (left) and HR ground truth (right), illustrating the difference in detail and clarity between the initial input and the reference standard.

Figure 5.2 shows a region of the super-resolved images generated by FSRCNN, RDN, ESRGAN and BCI, highlighting noticeable differences in quality. The FSRCNN output appears more blurred, with less defined edges, likely due to the simplicity and limited depth of its architecture, which hinders its ability to capture finer details, particularly around the vertebrae. In contrast, RDN produces a sharper image with better-defined edges, thanks to the introduction of dense feature fusion and residual learning. However, some smoothing remains, indicating that, despite its complexity, it does not fully capture all high-frequency details. ESRGAN excels in texture preservation and retains more detail than the other models, with well-defined edges. This superior detail preservation is due to the combination of perceptual loss, MSE, and adversarial learning, which enhance

perceptually pleasing details. Additionally, ESRGAN's training on full images, rather than patches like the other models, allows it to capture both local and global contextual information more consistently across the entire image.

In the super-resolved image produced by BCI, there is noticeable noise and reduced sharpness. BCI computes the value of each new pixel as a weighted combination of neighbouring pixels but cannot learn or adapt to specific visual features like edges or textures. It applies the same method uniformly across all pixels, without distinguishing between noise and actual image details. Consequently, because the neighbouring pixels contain noise, this noise is also incorporated into the interpolated value. As a result, noise in the input image is preserved and may even become more prominent. In contrast, DL models are trained to distinguish between noise and true features, preserving textures and structural details, resulting in more accurate reconstructions with fewer distortions.

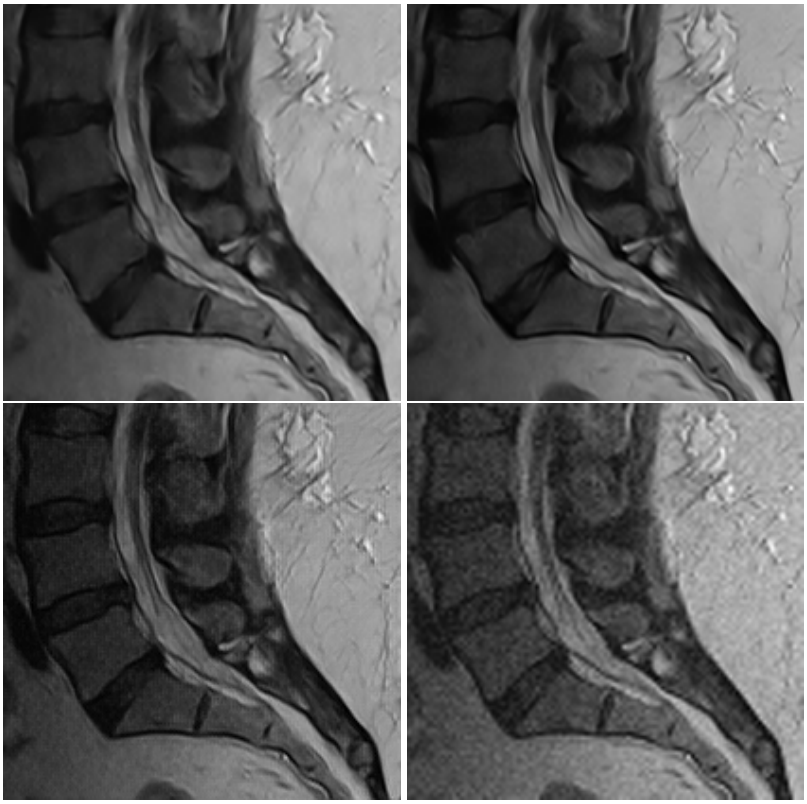

Figure 5.2: Top row: super-resolved outputs generated by FSRCNN (left) and RDN (right). Bottom row: super-resolved outputs from ESRGAN (left) and BCI (right).

To provide an objective analysis of the super-resolved quality of the images, full-reference metrics were used to compare each super-resolved image with the HR ground truth. The average assessment metric values for each SR method, and their standard deviations, are shown in Table 5.1.

Among the SR methods, RDN achieved the highest PSNR, showing it can more faithfully reconstruct the super-resolved image with fewer differences from the HR reference, and better noise reduction. Additionally, RDN most accurately reproduced the spatial

Table 5.1: Full-reference metrics results (mean ± standard deviation) for each SR method.

SR Methods	**Full-Reference Assessment Metrics**				
	PSNR(dB)	**SCC**	**SSIM**	**GMSD**	**VIF**
FSRCNN	25 ± 2	**0,30 ± 0,03**	0,69 ± 0,05	**0,049 ± 0,005**	**0,43 ± 0,03**
RDN	**30 ± 3**	**0,32 ± 0,03**	**0,89 ± 0,02**	**0,047 ± 0,005**	**0,45 ± 0,03**
ESRGAN	**29 ± 1**	0,25 ± 0,03	0,84 ± 0,02	**0,048 ± 0,004**	0,40 ± 0,03
BCI	21 ± 1	0,27 ± 0,02	0,53 ± 0,04	0,084 ± 0,006	0,36 ± 0,02

patterns of the original HR image, as indicated by its highest SCC value.

However, the PSNR and SCC values don't always reflect the human visual perception. To provide a more perceptually accurate evaluation, additional metrics like SSIM, GMSD, and VIF were computed. SSIM was used to evaluate luminance, contrast, and structure by comparing the structural similarity between the original and super-resolved images. RDN performed best, with a SSIM value closest to 1, indicating better preservation of structural integrity. In contrast, FSRCNN showed a low SSIM, indicating a weaker perception of structural integrity. This is particularly noticeable around the vertebrae, where the texture appears less detailed, largely due to the blur effect, which reduces contrast and sharpness.

In terms of GMSD, RDN and ESRGAN showed similar values, indicating both models effectively preserve edge transitions and consequently minimizing distortion. However, the slightly higher GMSD for ESRGAN suggests that while it retains more perceptual details, it may introduce subtle distortions in finer areas.

Finally, RDN outperformed with a higher VIF value, indicating better preservation of visual details. However, despite having a low VIF value, ESRGAN appears perceptually detailed. This suggests that in its effort to enhance sharpness, ESRGAN may introduce subtle distortions or even noise to create the appearance of a sharper image, which reduces the amount of information actually preserved, explaining its lower VIF score.

The results for BCI are consistently inferior across most metrics compared to the DL methods, aligning with the visual analysis. However, BCI demonstrated a slightly higher SCC than ESRGAN, indicating better reproduction of spatial correlations in some cases. The lower PSNR confirms greater noise, indicating a larger discrepancy from the original HR image. The SSIM score shows poorer structural preservation, while the higher GMSD highlights more edge distortions. Additionally, the reduced VIF suggests that BCI retains less visual information overall.

The PIQE metric was used to better assess image quality from a human perceptual perspective. A free-reference metric was chosen because it doesn't rely on a reference image, which may contain noise or distortions. Full-reference metrics, which assume the HR image is flawless, can be misleading if the reference image itself has imperfections. By using free-reference metrics like PIQE, the focus is on practical visual quality as perceived by users, which full-reference methods might overlook. Additionally, the SNR metric was computed to assess the balance between signal preservation and noise reduction. Results

for the three DL methods are shown in Table 5.2. ESRGAN achieved the lowest PIQE score, indicating superior preservation of perceptual details like textures and patterns, which are crucial for visual quality and fidelity. ESRGAN also recorded the highest SNR, highlighting its ability to manage noise effectively and produce visually pleasing images. Although RDN showed a higher PSNR, reflecting better pixel-to-pixel accuracy, ESRGAN's higher SNR combined with its lower PIQE, suggests that it balances noise reduction and perceptual detail preservation more effectively. Despite its slightly lower pixel-to-pixel accuracy compared to the reference HR image (yet still comparable to that of RDN), the markedly better PIQE value of ESRGAN combined with its visibly superior detail recovery, establishes it as the preferred model for further analysis. The activity masks showing PIQE-detected distortions are presented in Appendix B.

Table 5.2: Free-reference assessment metrics (mean ± standard deviation) for each DL model.

DL Methods	**Free-Reference Assessment Metrics**	
	SNR	**PIQE**
FSRCNN	**1,5 ± 0,4**	56 ± 6
RDN	1,3 ± 0,2	52 ± 6
ESRGAN	**2,0 ± 0,6**	**25 ± 7**

5.2 Application of the ESRGAN to the Original MRI Image

Figure 5.3 compares a region of the super-resolved MRI images generated by ESRGAN and BCI, both obtained from the original HR image. The ESRGAN image appears overly smoothed, indicating that the model prioritises denoising at the expense of preserving fine details and sharpness. This is due to its training on noisy images, making it more effective at suppressing noise. Thus, when applied to the original HR image, which is of good quality with minimal noise, ESRGAN tends to over-smooth removing both noise and details, particularly in the tissues surrounding the vertebrae. This highlights its difficulty in balancing noise reduction with detail preservation, leading to difficulties in distinguishing between areas that need denoising and those requiring sharpness. In contrast, the BCI image is sharper, as it doesn't focus on noise removal. Although BCI retains more noise, it preserves greater detail and sharpness.

The ESRGAN was then re-trained using images with low-level noise to prevent it from prioritising denoising. These training images were created using the image-domain degradation method, omitting the addition of Gaussian noise.

Figures 5.4 and 5.5 compare patches from the re-trained ESRGAN and BCI methods, applied to both the LR image with low-level noise and the original HR image, to evaluate the maximum resolution enhancement achieved by each. Comparing the super-resolved images generated from the input LR image with low-level noise, the ESRGAN produces noticeably sharper and more detailed results than BCI. This superior performance is

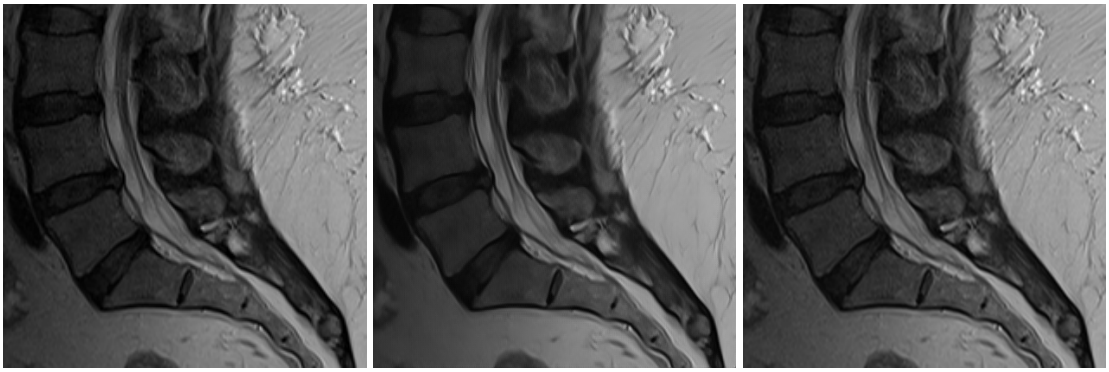

Figure 5.3: The original HR image (left) and its corresponding ESRGAN (middle) and BCI (right) enhanced versions.

Figure 5.4: Top row: the input LR image patch (left) and the corresponding ESRGAN-predicted patch (right). Bottom row: the original HR patch (left), and its corresponding ESRGAN-enhanced version (right).

attributed to ESRGAN's specific training on this type of degraded images, where it learned to reconstruct the missing details in such inputs effectively. In contrast, BCI did not undergo any training and, as a result, only relies on the existing information within the input, without learning to recover lost details, which limits its performance in this context. When applied to the original HR image, ESRGAN also provided a improvement in sharpness, with a slight better edge definition and enhanced fine textures. In this case, where the original image contains minimal noise, BCI achieved a resolution increase comparable to that of ESRGAN. This is because ESRGAN, having been trained on images

Figure 5.5: Top row: the input LR patch (left) and the corresponding BCI-predicted patch (right). Bottom row: the original HR patch (left) and its corresponding BCI-enhanced version (right).

with low-level noise, emphasised upsampling rather than prioritising denoising, much like BCI. The full images resulting from the application of both BCI and ESRGAN to the original HR image are presented in Appendix B.

The free-reference metrics for the super-resolved images produced by both methods, applied to the HR image, are shown in Table 5.3. The SNR values were similar, but since noise in the original HR image is minimal, SNR reflects more signal variance rather than noise. The slightly lower SNR for ESRGAN could be due to variations introduced when generating new details. However, ESRGAN's better PIQE score suggests fewer perceptual artefacts. The PIQE masks, shown in Figure 5.6, indicate areas of distortion and noise, with white regions signifying high activity and black regions representing minimal artefacts. The BCI mask has more white regions, indicating more blocking artefacts than ESRGAN. In the ESRGAN mask, artefacts are mainly concentrated along the vertebral edges and textured areas. In contrast, the BCI mask reveals artefacts not only along the edges but also in more homogeneous regions, suggesting lower image quality overall.

5.3 Analysis of the ESRGAN's ability to Generalise

This section evaluates ESRGAN's ability, trained on MRI images, to generalise across different degradation scenarios and modalities. The model was first tested on data from

Table 5.3: Free-reference assessment metrics (mean ± standard deviation) for the super-resolved images produced by ESRGAN and BCI, from the original HR input image.

Methods	**Free-Reference Assessment Metrics**	
	SNR	**PIQE**
ESRGAN	**0,9 ± 0,1**	**23 ± 3**
BCI	**1,0 ± 0,1**	56 ± 3

Figure 5.6: Activity masks for the super-resolved images generated by ESRGAN (left) and BCI (right) from the original HR image.

the frequency-domain degradation method, where the LR image undergoes a degradation step in k-space simulating MRI acquisition limitations. Its performance was then assessed on other modalities, such as CT and grey-scale natural images, to examine its versatility beyond the clinical context it was trained for.

5.3.1 Analysis of the ESRGAN Performance in Frequency-Domain Degradation Scenarios

Figure 5.7 illustrates the two masks used to generate the LR image along with a region of the resulting images after their application. Both masks operate in k-space, however, the central mask retains only low-frequency information related to structure and contrast, while the radial mask additionally captures some high-frequency information in specific directions, introducing anisotropy by preserving details in certain orientations. This mask simulates data acquisition in k-space using a series of radial trajectories, where the centre is densely sampled, while the outer regions are sampled less frequently and in a more targeted manner. The central mask, circular and filled, maintains sharpness in the central area but introduces ringing artefacts in the regions corresponding to the abrupt transition at the edge of the circle in the mask. In contrast, the radial mask results in inconsistent sharpness, with some regions showing better detail preservation along the radial lines and others displaying slight blurring. This directional bias leads to variations in the resolution depending on the orientation of the sampled lines. Despite its anisotropy, the radial mask retains more high-frequency information than the central mask, which filters out

more high-frequency details. As a result, the radial mask offered better overall resolution, capturing finer details in some orientations, while the central mask more aggressively reduces high-frequency content.

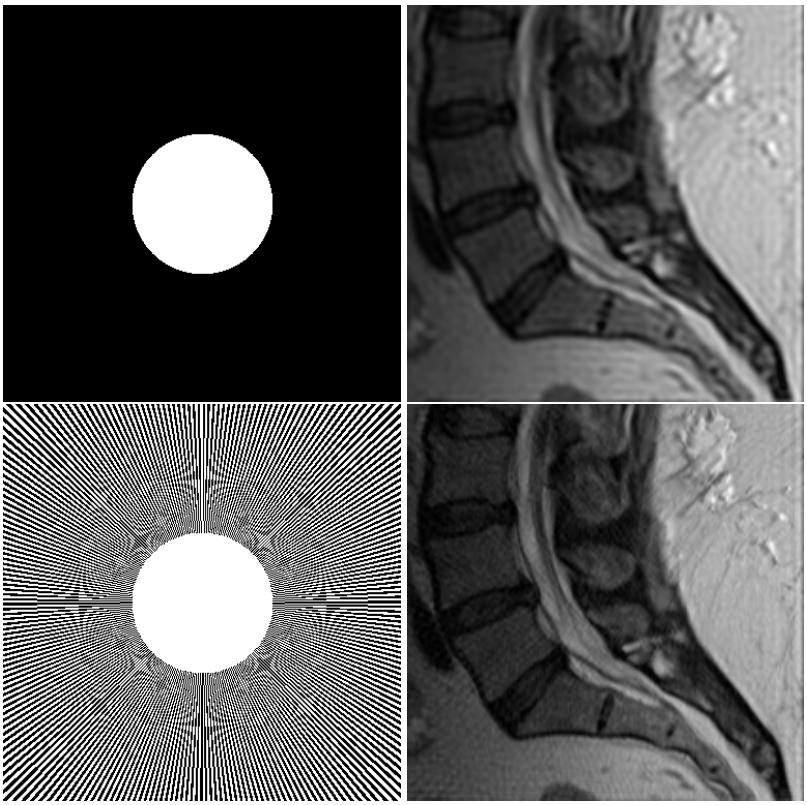

Figure 5.7: Top row: central mask (left) and image resulting from its application (right). Bottom row: radial mask (left) and image resulting from its application (right).

Observing Figure 5.8, the super-resolved images from ESRGAN show more detail and less noise than those from BCI, highlighting ESRGAN's superior denoising ability. Although both masks produce images with similar detail and sharpness, the radial mask results in a slight loss of detail in the tissue around the vertebrae compared to the central mask. This can be attributed to the anisotropic sampling of the radial mask that preserves details more effectively along certain radial lines while losing clarity in other directions. As a result, both BCI and ESRGAN face minor challenges in consistently recovering details. The central mask, on the other hand, more evenly cuts high-frequency information, leading to fewer issues with uneven sharpness. Despite these differences, the metrics in Tables 5.4 and 5.5 show no significant variations between the two masks, indicating comparable image fidelity and detail preservation. While BCI performed similarly across both masks, it consistently underperformed compared to ESRGAN, except in the SCC metric, where it achieved a higher score, reflecting better reconstruction of spatial correlation. In conclusion, ESRGAN performed better overall, demonstrating greater robustness in handling different degradation patterns and reconstructing finer details than BCI, despite minor challenges with the radial mask.

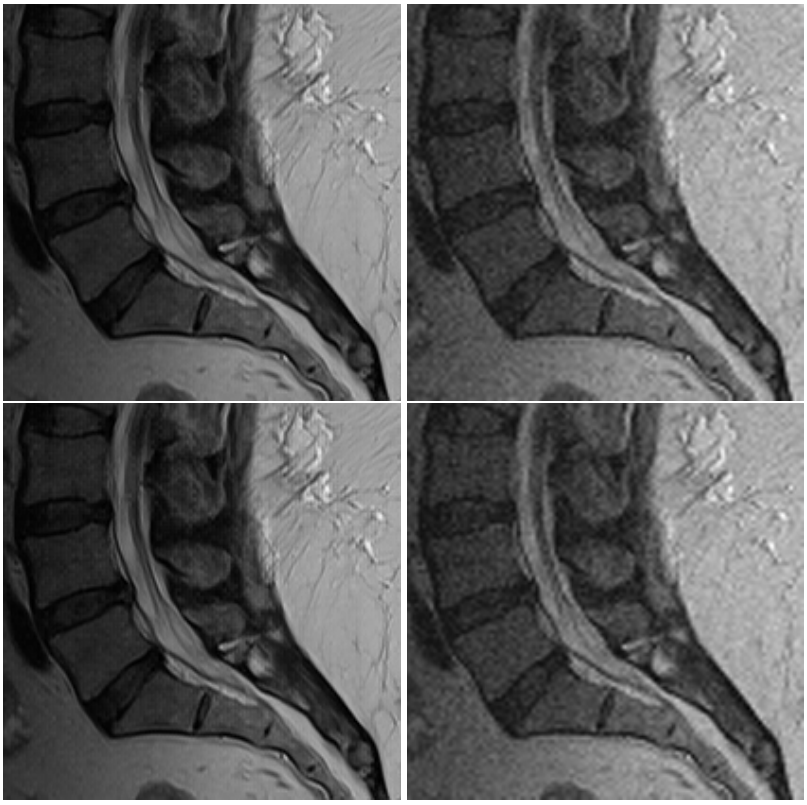

Figure 5.8: Top row: super-resolved images generated by ESRGAN (left) and BCI (right) from LR image generated with central mask. Bottom row: super-resolved images generated by ESRGAN (left) and BCI (right) from LR image generated with radial mask.

Table 5.4: Full-reference assessment metrics results (mean ± standard deviation) for each SR method, for the central mask.

Method	Full-Reference Assessment Metrics				
	PSNR (dB)	SCC	SSIM	GMSD	VIF
ESRGAN	**30 ± 2**	0,25 ± 0,03	**0,83 ± 0,02**	**0,051 ± 0,006**	**0,39 ± 0,03**
BCI	23 ± 2	**0,31 ± 0,03**	0,61 ± 0,04	0,088 ± 0,006	**0,37 ± 0,02**

Table 5.5: Full-reference assessment metrics results (mean ± standard deviation) for each SR method, for the radial mask.

Method	Full-Reference Assessment Metrics				
	PSNR (dB)	SCC	SSIM	GMSD	VIF
ESRGAN	**30 ± 2**	0,25 ± 0,03	**0,83 ± 0,02**	**0,051 ± 0,007**	**0,39 ± 0,03**
BCI	23 ± 2	**0,31 ± 0,03**	0,61 ± 0,04	0,089 ± 0,006	**0,37 ± 0,02**

5.3.2 Analysis of the ESRGAN Performance Across Different Imaging Modalities

To assess the model's ability to perform SR on different types of images and thus be used outside the clinical MRI context in which it was trained, the model was also tested on CT

data and grey-scale natural images.

Observing the Figures 5.9 and 5.10, a clear increase in sharpness is evident for both CT and natural images. In both cases, ESRGAN consistently outperformed BCI in terms of visual sharpness and detail recovery by enhancing fine textures and sharp edges, particularly in the soft tissue of CT image and in finer details like the textures of the clothing and camera equipment in the natural image. In contrast, BCI produced less sharp results, with more blurring and noise. However, when visually comparing the super-resolved images generated by ESRGAN to the ground truth, while ESRGAN improved resolution and outperformed BCI, it did not fully recover all perceptible details. In contrast, for MRI images, where ESRGAN was specifically trained, the super-resolved images achieved a greater recovery of fine details compared to the ground truth.

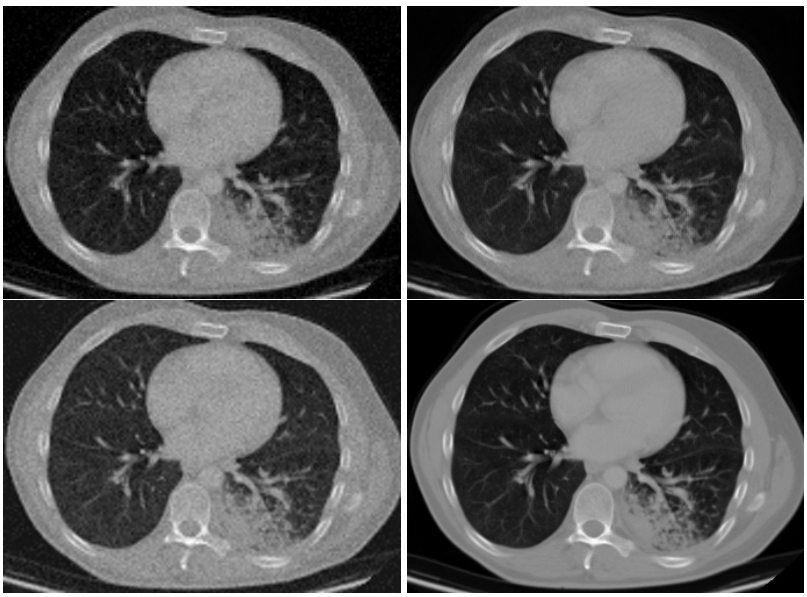

Figure 5.9: Resolution enhancement for CT images. Top row: the LR input (left) and the ESRGAN's output (right). Bottom row: BCI's output (left) and HR ground truth image (right).

The visual observations are consistent with the evaluation metrics (see Table 5.6 and 5.7). For the CT images, ESRGAN outperforms BCI across all key metrics: PSNR, SSIM, GMSD, and VIF, confirming its ability to generate sharper images with fewer distortions. The higher PSNR for ESRGAN reflects superior signal preservation, while the higher SSIM indicates better structural integrity. The lower GMSD demonstrates improved edge detail retention, and the higher VIF score highlights enhanced preservation of visual information, ultimately improving the perceptual quality of the reconstructed images. BCI shows a slightly higher SCC, suggesting better spatial correlation, but this doesn't improve visual clarity. For grey-scale natural images, ESRGAN again achieves better SSIM, lower GMSD and higher VIF, producing sharper results, while BCI scores higher in PSNR and SCC, indicating better pixel-level accuracy. Overall, ESRGAN enhances resolution for both CT and natural images, outperforming BCI in the majority of the metrics. However, the super-resolved MRI image visually resembles the ground-truth image the most, with

Figure 5.10: Resolution enhancement for natural grey-scale images. Top row: the LR input (left) and the ESRGAN's output (right). Bottom row: BCI's output (left) and HR ground truth image (right).

better recovery of perceptual details. This suggests that while ESRGAN generalises well to CT and natural images, its performance is slightly better in MRI, where it has been specifically trained.

Table 5.6: Full-reference assessment metrics results (mean ± standard deviation) for each SR method, for CT images.

SR Method	Full-Reference Assessment Metrics				
	PSNR (dB)	SCC	SSIM	GMSD	VIF
ESRGAN	**29 ± 2**	0,25 ± 0,04	**0,79 ± 0,02**	**0,076 ± 0,017**	**0,43 ± 0,03**
BCI	24 ± 2	**0,32 ± 0,04**	0,63 ± 0,06	0,094 ± 0,010	**0,42 ± 0,03**

Table 5.7: Full-reference assessment metrics results (mean ± standard deviation) for each SR method, for natural grey-scale images.

SR Method	Full-Reference Assessment Metrics				
	PSNR (dB)	SCC	SSIM	GMSD	VIF
ESRGAN	**24 ± 4**	**0,23 ± 3,56**	**0,77 ± 0,06**	**0,079 ± 0,018**	**0,36 ± 0,07**
BCI	**25 ± 2**	**0,30 ± 0,07**	**0,72 ± 0,05**	0,096 ± 0,011	**0,35 ± 0,06**

6

Conclusion

6.1 Main Conclusions

This study aimed to implement a DL method to acchieve SR of spinal MRI images, seeking to surpass the current reference BCI technique. Three state-of-the-art models (FSRCNN, RDN, and ESRGAN) were implemented and optimised to achieve their best performance.

Among all three models, ESRGAN demonstrated to be the most effective. Its generator's residual architecture, with dense connections, enabled learning at both local and global levels across the image, resulting in sharper detail recovery. The use of perceptual loss, alongside MSE improved the learning of textures and patterns, leading to enhanced visual quality. While RDN achieved better full-reference metric scores, ESRGAN had the lower PIQE value and preserved more visual details, indicating superior perceptual quality. This highlights the limitation of relying only on full-reference metrics, which may miss visible improvements, especially when the reference image is suboptimal.

When applied to the original HR image, ESRGAN prioritised denoising, sometimes at the cost of fine detail. However, when re-trained on images with low-level noise, ESRGAN successfully improved the definition of anatomical structures, although the results remained visually similar to those of BCI, likely due to the minimal noise present in the original image. With little noise to remove, ESRGAN focused more on upsampling, similar to BCI, resulting in comparable outcomes without amplifying or retaining noise.

Another key aspect of this study explored the ESRGAN's generalisation ability across different image degradation scenarios and different modalities, such as CT scans and grey-scale natural images. Despite minor challenges like preserving anisotropic details, the results suggest that ESRGAN has versatility, consistently outperforming BCI. For both CT and grey-scale natural images, ESRGAN also surpassed BCI in the majority of the assessment metrics, offering a better visual sharpness and detail recovery. However, as expected, the model did not recover lost visual details as effectively as it did in MRI, where it had been specifically trained and optimised.

This study successfully demonstrated the efficacy of DL methods, particularly ESRGAN, in the task of SR, outperforming the traditional BCI technique and enhancing image

resolution. The findings lay a strong foundation for future advancements in medical image processing. The ability of these models to improve resolution and recover perceptual details, especially in MRI, highlights the potential of DL to enhance diagnostic accuracy and treatment planning. The enhanced images produced by these techniques can be integrated into clinical workflows, enabling more precise segmentation and reconstruction of anatomical structures, thus providing healthcare professionals with more effective tools for diagnosis, surgical planning, and treatment.

6.2 Limitations and Future Work

Several limitations were found in this study, potentially impacting the results and offering opportunities for future improvements.

A key limitation is the lack of detailed information about the specific origin of the MRI images, including whether they were acquired using a single machine or multiple machines. This uncertainty affects data reliability, as noise and resolution can vary between different equipment, impacting the consistency and generalisability of the results. Additionally, using a Central Processing Unit (CPU) for model training and testing proved inefficient, as CPUs are not well-suited for the parallel computations required by DL models. This led to longer processing times and increased memory usage. Transitioning to a Graphics Processing Unit (GPU), designed for parallel tasks, such as the NVIDIA range, would significantly improve training speed and model performance.

Another limitation is ESRGAN's sensitivity to noise levels during training. When trained on noisy images, the model prioritised denoising over preserving detail, leading to excessive smoothing. This highlights the need to balance denoising and upsampling. Future improvements could involve separating these tasks into specialised models or exploring architectures that integrate these processes more effectively.

The fixed 2× upsampling factor is another constraint. In clinical settings, the required upsampling factor may vary, but training separate models for each factor is impractical due to the significant time, computational resources, and storage requirements. Developing a dynamic model capable of adjusting to different scaling factors would enhance flexibility, potentially using multi-branch networks or adaptive upsampling layers.

Moreover, the image degradation method used to artificially generate the training data may not accurately reflect real-world conditions, where factors such as equipment quality, patient movement, and variations in imaging protocols introduce specific artefacts. Future improvements could involve incorporating real-world data, with HR images captured using advanced imaging antennas that provide finer detail and reduce noise, while LR images would be generated from systems lacking these antennas. This approach would create more realistic training conditions, better representing the degradation typically encountered in clinical environments.

The model also demonstrated its versatility by enhancing the resolution of images beyond the scope of MRI. However, even though it visually produced superior detail

recovery when applied to MRI, further specialising in the unique characteristics of MRI, such as incorporating T_1-weighted information to improve T_2 resolution or using k-space data, could improve its accuracy even more, particularly in addressing aliasing artefacts.

Furthermore, operating exclusively in the image domain may limit the model's ability to handle non-structural and non-local artefacts commonly found in real-world LR images. Future work could explore frequency domain operations to address these issues better.

Additionally, this study only compared ESRGAN with BCI, and future evaluations could include other conventional SR methods, such as BLI or dictionary learning techniques, for a more comprehensive assessment.

Finally, the lack of qualitative assessment to validate the practical applicability of the model is a significant limitation. Without this evaluation, it is challenging to confirm that the observed improvements in resolution genuinely contribute to better diagnostic accuracy.

Bibliography

[1] J. M. Lourenço, *The NOVAthesis LaTeX Template User's Manual*, NOVA University Lisbon, 2021. [Online]. Available: https://github.com/joaomlourenco/novathesis/raw/main/template.pdf (cit. on p. i).

[2] C. J. Liew, P. Krishnaswamy, L.-T. Cheng, C. H. Tan, A. C. Poh, and T. C. Lim, "Artificial intelligence and radiology in singapore: Championing a new age of augmented imaging for unsurpassed patient care," *Ann Acad Med Singapore*, vol. 48, no. 1, pp. 16–24, 2019. DOI: 10.47102/annals-acadmedsg.V48N1p16 (cit. on p. 1).

[3] T. Zhou, Q. Cheng, H. Lu, Q. Li, X. Zhang, and S. Qiu, "Deep learning methods for medical image fusion: A review," *Computers in Biology and Medicine*, vol. 160, p. 106 959, 2023. DOI: https://doi.org/10.1016/j.compbiomed.2023.106959 (cit. on p. 1).

[4] A. M. Awadalla, A. S. Aljulayfi, A. R. Alrowaili, *et al.*, "Management of lumbar disc herniation: A systematic review," *Cureus*, vol. 15, no. 10, 2023. DOI: 10.7759/cureus.47908 (cit. on p. 1).

[5] M. T. Modic, T. Masaryk, F. Boumphrey, M. Goormastic, and G. Bell, "Lumbar herniated disk disease and canal stenosis: Prospective evaluation by surface coil mr, ct, and myelography," *American journal of neuroradiology*, vol. 7, no. 4, pp. 709–717, 1986. DOI: 10.2214/ajr.147.4.757 (cit. on p. 1).

[6] E. Van Reeth, I. W. Tham, C. H. Tan, and C. L. Poh, "Super-resolution in magnetic resonance imaging: A review," *Concepts in Magnetic Resonance Part A*, vol. 40, no. 6, pp. 306–325, 2012. DOI: 10.1002/cmr.a.21249 (cit. on pp. 2, 4–6).

[7] C. Guy and D. Ffytche, *An Introduction to the Principles of Medical Imaging*. Londres, Reino Unido: Academic Press, 2005, accessed: 1 August 2024, ISBN: 1-86094-502-3. [Online]. Available: https://www.academia.edu/9798762/An_Introduction_to_the_Principles_of_Medical_Imaging (cit. on p. 3).

[8] D. W. McRobbie, E. A. Moore, M. J. Graves, and M. R. Prince, *From Picture to Proton*, 2nd. Cambridge, UK: Cambridge University Press, 2017, Accessed: 1 August 2024, ISBN: 9781107706958. [Online]. Available: https://www.academia.edu/37299892/From_Picture_to_Proton (cit. on pp. 3, 4).

[9] B. A. Jung and M. Weigel, "Spin echo magnetic resonance imaging," *Journal of Magnetic Resonance Imaging*, vol. 37, no. 4, pp. 805–817, 2013. DOI: https://doi.org/10.1002/jmri.24068 (cit. on p. 4).

[10] M. Dominietto and M. Rudin, "Could magnetic resonance provide in vivo histology?" *Frontiers in genetics*, vol. 4, p. 298, 2014. DOI: 10.3389/fgene.2013.00298 (cit. on p. 4).

[11] H. Greenspan, "Super-resolution in medical imaging," *The computer journal*, vol. 52, no. 1, pp. 43–63, 2009. DOI: 10.1093/comjnl/bxm075 (cit. on p. 5).

[12] H. Benveniste and S. Blackband, "Mr microscopy and high resolution small animal mri: Applications in neuroscience research," *Progress in neurobiology*, vol. 67, no. 5, pp. 393–420, 2002. DOI: 10.1016/s0301-0082(02)00020-5 (cit. on p. 5).

[13] Y. Li, B. Sixou, and F. Peyrin, "A review of the deep learning methods for medical images super resolution problems," *Irbm*, vol. 42, no. 2, pp. 120–133, 2021, Accessed: 26 May 2024. [Online]. Available: https://doi.org/10.1016/j.irbm.2020.08.004 (cit. on p. 6).

[14] V. Labs. "Image super resolution: The ultimate guide." Accessed: 25 September 2024. (2023), [Online]. Available: https://www.v7labs.com/blog/image-super-resolution-guide#single-image-vs-multi-image-super-resolution (cit. on p. 6).

[15] C.-H. Pham, "Deep learning for medical image super resolution and segmentation," NNT: 2018IMTA0124, tel-02124889, PhD thesis, École nationale supérieure Mines-Télécom Atlantique, Nantes, França, 2018 (cit. on p. 6).

[16] H. Yang, Z. Wang, X. Liu, C. Li, J. Xin, and Z. Wang, "Deep learning in medical image super resolution: A review," *Applied Intelligence*, vol. 53, no. 18, pp. 20 891–20 916, 2023. DOI: https://doi.org/10.1007/s10489-023-04566-9 (cit. on pp. 6, 24).

[17] W. Dourado, "Avaliação de técnicas de interpolação de imagens digitais.," M.S. thesis, Universidade Estadual Paulista, 2014. [Online]. Available: http://www2.fct.unesp.br/pos/mac/dissertacoes/2014/wesley_dourado.pdf (cit. on p. 7).

[18] P. Srilakshmi, C. Mounika, S. G. PG, K. Shijith, S. S. Chinchole, and S. A. Selvaraj, "Resolution enhancement of mr brain images using interpolation method," in *2022 International Conference on Smart and Sustainable Technologies in Energy and Power Sectors (SSTEPS)*, IEEE, 2022, pp. 136–140. DOI: 10.1109/SSTEPS57475.2022.00043 (cit. on pp. 7, 12).

[19] D. Liu, G. Zhou, X. Zhou, C. Li, and F. Wang, "Fpga-based on-board cubic convolution interpolation for spaceborne georeferencing," *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, vol. 42, pp. 349–356, 2020. DOI: https://doi.org/10.5194/isprs-archives-XLII-3-W10-349-2020 (cit. on p. 7).

[20] F. A. Raheem and A. Abdulwahhab, "Deep learning convolution neural networks analysis and comparative study for static alphabet asl hand gesture recognition," *Journal of Xidian University*, vol. 14, no. 4, pp. 1871–1881, 2020. DOI: 10.37896/jxu14.4/212 (cit. on p. 7).

[21] J.-L. Li and G.-F. Zhao, "Further development of digital image correlation (dic) to measure the discontinuous deformation of rock on small-scale tests," *Rock Mechanics and Rock Engineering*, vol. 56, no. 2, pp. 1163–1183, 2023. DOI: https://doi.org/10.1007/s00603-022-03120-2 (cit. on p. 8).

[22] J. Bajwa, U. Munir, A. Nori, and B. Williams, "Artificial intelligence in healthcare: Transforming the practice of medicine," *Future healthcare journal*, vol. 8, no. 2, e188–e194, 2021. DOI: 10.7861/fhj.2021-0095 (cit. on p. 8).

[23] A. Géron, *Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow*, 2nd Edition. Sebastopol, CA, USA: O'Reilly Media, Inc., 2019, ISBN: 9781492032649 (cit. on pp. 8–10).

[24] Y. He, "Deep-learning methods for mri super-resolution," M.S. thesis, The University of Queensland, 2022. DOI: https://doi.org/10.14264/38367c2 (cit. on p. 9).

[25] DataCamp, *Introduction to convolutional neural networks (cnns)*, Accessed: 1 May 2024, n.d. [Online]. Available: https://www.datacamp.com/tutorial/introduction-to-convolutional-neural-networks-cnns (cit. on p. 10).

[26] S. Shekhawat. "Types of neural networks - convolutional neural networks." Accessed: 24-Aug-2024. (2024), [Online]. Available: https://medium.com/@shekhawatsamvardhan/types-of-neural-networks-convolutional-neural-networks-bd973e4fe78c (cit. on p. 10).

[27] M. D. Pra. "Generative adversarial networks." Accessed: 2024-07-14. (2023), [Online]. Available: https://medium.com/@marcodelpra/generative-adversarial-networks-dba10e1b4424 (cit. on p. 11).

[28] N. Jiang and L. Wang, "Quantum image scaling using nearest neighbor interpolation," *Quantum Information Processing*, vol. 14, pp. 1559–1571, 2015. DOI: https://doi.org/10.1007/s11128-014-0841-8 (cit. on p. 12).

[29] Q. Wang and R. K. Ward, "A new orientation-adaptive interpolation method," *IEEE transactions on image processing*, vol. 16, no. 4, pp. 889–900, 2007. DOI: 10.1109/TIP.2007.891794 (cit. on p. 12).

[30] Q. Wang, R. Ward, and H. Shi, "Isophote estimation by cubic-spline interpolation," in *Proceedings. International Conference on Image Processing*, IEEE, vol. 3, 2002, pp. III–III. DOI: 10.1109/ICIP.2002.1038990 (cit. on p. 12).

[31] D. Suresha and H. Prakash, "Single picture super resolution of natural images using n-neighbor adaptive bilinear interpolation and absolute asymmetry based wavelet hard thresholding," in *2016 2nd International Conference on Applied and Theoretical Computing and Communication Technology (iCATccT)*, IEEE, 2016, pp. 387–393. DOI: 10.1109/ICATCCT.2016.7912029 (cit. on p. 12).

[32] N. Bharati, A. Khosla, and N. Sood, "Image reconstruction using cubic b-spline interpolation," in *2011 Annual IEEE India Conference*, IEEE, 2011, pp. 1–5. DOI: 10.1109/INDCON.2011.6139378 (cit. on p. 12).

[33] T. M. Lehmann, C. Gonner, and K. Spitzer, "Addendum: B-spline interpolation in medical image processing," *IEEE transactions on medical imaging*, vol. 20, no. 7, pp. 660–665, 2001. DOI: 10.1109/42.932749 (cit. on p. 12).

[34] T. Blu, P. Thévenaz, and M. Unser, "Linear interpolation revitalized," *IEEE Transactions on Image Processing*, vol. 13, no. 5, pp. 710–719, 2004. DOI: 10.1109/TIP.2004.826093 (cit. on p. 12).

[35] M. Irani and S. Peleg, "Super resolution from image sequences," in *[1990] Proceedings. 10th International Conference on Pattern Recognition*, IEEE, vol. 2, 1990, pp. 115–120. DOI: 10.1109/ICPR.1990.119340 (cit. on p. 13).

[36] M. Irani and S. Peleg, "Improving resolution by image registration," *CVGIP: Graphical models and image processing*, vol. 53, no. 3, pp. 231–239, 1991. DOI: https://doi.org/10.1016/1049-9652(91)90045-L (cit. on p. 13).

[37] M. Irani and S. Peleg, "Motion analysis for image enhancement: Resolution, occlusion, and transparency," *Journal of visual communication and image representation*, vol. 4, no. 4, pp. 324–335, 1993. DOI: https://doi.org/10.1006/jvci.1993.1030 (cit. on p. 13).

[38] R. Nayak, S. Monalisa, and D. Patra, "Spatial super resolution based image reconstruction using hibp," in *2013 Annual IEEE India Conference (INDICON)*, IEEE, 2013, pp. 1–6. DOI: 10.1109/INDCON.2013.6726146 (cit. on p. 13).

[39] G. H. Alshammri, A. K. Samha, W. El-Shafai, *et al.*, "Three-dimensional video super-resolution reconstruction scheme based on histogram matching and recursive bayesian algorithms," *IEEE Access*, vol. 10, pp. 41 935–41 951, 2022. DOI: 10.1109/ACCESS.2022.3153409 (cit. on p. 13).

[40] R. Dian, S. Li, and L. Fang, "Non-local sparse representation for hyperspectral image super-resolution," in *2016 IEEE International Conference on Image Processing (ICIP)*, IEEE, 2016, pp. 2832–2835. DOI: 10.1109/ICIP.2016.7532876 (cit. on p. 13).

[41] E. Mikaeli, A. Aghagolzadeh, and M. Azghani, "Single image super resolution via adaptive group-based sparse domain selection," in *Electrical Engineering (ICEE), Iranian Conference on*, IEEE, 2018, pp. 731–736. DOI: 10.1109/ICEE.2018.8472489 (cit. on p. 13).

[42] T. Barman, B. Deka, and A. Prasad, "Gpu-accelerated adaptive dictionary learning and sparse representations for multispectral image super-resolution," in *2021 IEEE 18th India Council International Conference (INDICON)*, IEEE, 2021, pp. 1–7. DOI: 10.1109/INDICON52576.2021.9691521 (cit. on p. 13).

[43] H.-F. Yan, Y.-Q. Zhao, J. C.-W. Chan, and S. G. Kong, "Spectral super-resolution based on dictionary optimization learning via spectral library," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 61, pp. 1–16, 2022. DOI: 10.1109/TGRS.2022.3229439 (cit. on p. 13).

[44] C. Dong, C. C. Loy, K. He, and X. Tang, "Learning a deep convolutional network for image super-resolution," in *Computer Vision–ECCV 2014: 13th European Conference, Zurich, Switzerland, September 6-12, 2014, Proceedings, Part IV 13*, Springer, 2014, pp. 184–199. DOI: https://doi.org/10.1007/978-3-319-10593-2_13 (cit. on p. 14).

[45] D. Mahapatra, B. Bozorgtabar, and R. Garnavi, "Image super-resolution using progressive generative adversarial networks for medical image analysis," *Computerized Medical Imaging and Graphics*, vol. 71, pp. 30–39, 2019. DOI: https://doi.org/10.1016/j.compmedimag.2018.10.005 (cit. on p. 14).

[46] F. A. Dharejo, M. Zawish, F. Deeba, *et al.*, "Multimodal-boost: Multimodal medical image super-resolution using multi-attention network with wavelet transform," *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, vol. 20, no. 4, pp. 2420–2433, 2022. DOI: 10.1109/TCBB.2022.3191387 (cit. on p. 14).

[47] H. Liu, J. Xu, Y. Wu, Q. Guo, B. Ibragimov, and L. Xing, "Learning deconvolutional deep neural network for high resolution medical image reconstruction," *Information Sciences*, vol. 468, pp. 142–154, 2018. DOI: https://doi.org/10.1016/j.ins.2018.08.022 (cit. on p. 14).

[48] J. Kim, J. K. Lee, and K. M. Lee, "Accurate image super-resolution using very deep convolutional networks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 1646–1654. DOI: https://doi.org/10.48550/arXiv.1511.04587 (cit. on p. 14).

[49] J. Kim, J. K. Lee, and K. M. Lee, "Deeply-recursive convolutional network for image super-resolution," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 1637–1645. DOI: https://doi.org/10.48550/arXiv.1511.04491 (cit. on p. 14).

[50] Y. Tai, J. Yang, and X. Liu, "Image super-resolution via deep recursive residual network," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 3147–3155. DOI: 10.1109/CVPR.2017.298 (cit. on p. 14).

[51] Y. Zhang and M. An, "Deep learning-and transfer learning-based super resolution reconstruction from single medical image," *Journal of healthcare engineering*, vol. 2017, no. 1, p. 5 859 727, 2017. DOI: https://doi.org/10.1155/2017/5859727 (cit. on p. 14).

[52] J. Lin, N. T. Clancy, Y. Hu, *et al.*, "Endoscopic depth measurement and super-spectral-resolution imaging," in *Medical Image Computing and Computer-Assisted Intervention-MICCAI 2017: 20th International Conference, Quebec City, QC, Canada, September 11-13, 2017, Proceedings, Part II 20*, Springer, 2017, pp. 39–47. DOI: https://doi.org/10.48550/arXiv.1706.06081 (cit. on p. 14).

[53] C. Dong, C. C. Loy, and X. Tang, "Accelerating the super-resolution convolutional neural network," in *Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11-14, 2016, Proceedings, Part II 14*, Springer, 2016, pp. 391–407. DOI: https://doi.org/10.48550/arXiv.1608.00367 (cit. on pp. 14–16, 20).

[54] A. Aitken, C. Ledig, L. Theis, J. Caballero, Z. Wang, and W. Shi, *Checkerboard artifact free sub-pixel convolution*, US Patent 11,308,361, 2022-4 19. DOI: https://doi.org/10.48550/arXiv.1707.02937 (cit. on p. 15).

[55] M. Haris, G. Shakhnarovich, and N. Ukita, "Deep back-projection networks for super-resolution," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2018, pp. 1664–1673. DOI: https://doi.org/10.48550/arXiv.1803.02735 (cit. on p. 15).

[56] Z. Li, J. Yang, Z. Liu, X. Yang, G. Jeon, and W. Wu, "Feedback network for image super-resolution," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2019, pp. 3867–3876. DOI: https://doi.org/10.48550/arXiv.1903.09814 (cit. on p. 15).

[57] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 4700–4708. DOI: https://doi.org/10.48550/arXiv.1608.06993 (cit. on p. 15).

[58] Y. Zhang, Y. Tian, Y. Kong, B. Zhong, and Y. Fu, "Residual dense network for image super-resolution," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2018, pp. 2472–2481. DOI: https://doi.org/10.48550/arXiv.1802.08797 (cit. on pp. 15, 16, 20, 21).

[59] W.-S. Lai, J.-B. Huang, N. Ahuja, and M.-H. Yang, "Deep laplacian pyramid networks for fast and accurate super-resolution," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 624–632. DOI: 10.1109/CVPR.2017.618 (cit. on pp. 15, 16).

[60] C. Ledig, L. Theis, F. Huszár, *et al.*, "Photo-realistic single image super-resolution using a generative adversarial network," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 4681–4690 (cit. on p. 15).

[61] X. Wang, K. Yu, S. Wu, *et al.*, "Esrgan: Enhanced super-resolution generative adversarial networks," in *Proceedings of the European conference on computer vision (ECCV) workshops*, 2018, pp. 0–0. DOI: https://doi.org/10.48550/arXiv.1809.00219 (cit. on pp. 15, 16, 20, 22).

[62] Y. Yuan, S. Liu, J. Zhang, Y. Zhang, C. Dong, and L. Lin, "Unsupervised image super-resolution using cycle-in-cycle generative adversarial networks," in *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*, 2018, pp. 701–710. DOI: https://doi.org/10.48550/arXiv.1809.00437 (cit. on p. 15).

[63] H. Liu, J. Liu, J. Li, J.-S. Pan, and X. Yu, "Dl-mri: A unified framework of deep learning-based mri super resolution," *Journal of Healthcare Engineering*, vol. 2021, no. 1, p. 5 594 649, 2021. DOI: https://doi.org/10.1155/2021/5594649 (cit. on p. 16).

[64] K. M. Mithra, S. Ramanarayanan, K. Ram, and M. Sivaprakasam, "Reference-based texture transfer for single image super-resolution of magnetic resonance images," in *2021 IEEE 18th International Symposium on Biomedical Imaging (ISBI)*, IEEE, 2021, pp. 579–583. DOI: 10.1109/ISBI48211.2021.9433961 (cit. on p. 16).

[65] C. Zhao, B. E. Dewey, D. L. Pham, P. A. Calabresi, D. S. Reich, and J. L. Prince, "Smore: A self-supervised anti-aliasing and super-resolution algorithm for mri using deep learning," *IEEE transactions on medical imaging*, vol. 40, no. 3, pp. 805–817, 2020. DOI: 10.1109/TMI.2020.3037187 (cit. on p. 16).

[66] B. Lim, S. Son, H. Kim, S. Nah, and K. Mu Lee, "Enhanced deep residual networks for single image super-resolution," in *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*, 2017, pp. 136–144. DOI: 10.1109/CVPRW.2017.151 (cit. on p. 16).

[67] D. Zhu and D. Qiu, "Residual dense network for medical magnetic resonance images super-resolution," *Computer Methods and Programs in Biomedicine*, vol. 209, p. 106 330, 2021. DOI: https://doi.org/10.1016/j.cmpb.2021.106330 (cit. on p. 16).

[68] G. Li, W. Xing, L. Zhao, *et al.*, "Dudoinet: Dual-domain implicit network for multi-modality mr image arbitrary-scale super-resolution," in *Proceedings of the 31st ACM International Conference on Multimedia*, 2023, pp. 7335–7344. DOI: https://doi.org/10.1145/3581783.3612230 (cit. on p. 16).

[69] Y. Xiao, C. Chen, L. Wang, *et al.*, "A novel hybrid generative adversarial network for ct and mri super-resolution reconstruction," *Physics in Medicine & Biology*, vol. 68, no. 13, p. 135 007, 2023. DOI: 10.1088/1361-6560/acdc7e (cit. on p. 16).

[70] T. Eo, Y. Jun, T. Kim, J. Jang, H.-J. Lee, and D. Hwang, "Kiki-net: Cross-domain convolutional neural networks for reconstructing undersampled magnetic resonance images," *Magnetic resonance in medicine*, vol. 80, no. 5, pp. 2188–2201, 2018. DOI: https://doi.org/10.1002/mrm.27201 (cit. on p. 16).

[71] Y. Liu, Y. Pang, X. Liu, Y. Liu, and J. Nie, "Diik-net: A full-resolution cross-domain deep interaction convolutional neural network for mr image reconstruction," *Neurocomputing*, vol. 517, pp. 213–222, 2023. DOI: https://doi.org/10.1016/j.neucom.2022.09.048 (cit. on p. 16).

[72] S. Sudirman, A. Al Kafri, F. Natalia, *et al.*, *Lumbar spine mri dataset*, version 2, Published: 3 Apr 2019, Accessed: 2 March 2024, 2019. DOI: 10.17632/k57fr854j2.2. [Online]. Available: https://doi.org/10.17632/k57fr854j2.2 (cit. on p. 17).

[73] TrainingDataPro, *Computed tomography (ct) of the brain - object detection*, Accessed: 26-Aug-2024, 2024. [Online]. Available: https://www.kaggle.com/datasets/trainingdatapro/computed-tomography-ct-of-the-brain (cit. on p. 18).

[74] U. of Cape Town, *Standard greyscale images for image processing*, Accessed: 26-Aug-2024, 2024. [Online]. Available: http://www.dip.ee.uct.ac.za/imageproc/stdimages/greyscale/ (cit. on p. 18).

[75] C. Yan, G. Shi, and Z. Wu, "Smir: A transformer-based model for mri super-resolution reconstruction," in *2021 IEEE International Conference on Medical Imaging Physics and Engineering (ICMIPE)*, IEEE, 2021, pp. 1–6. DOI: 10.1109/ICMIPE53131.2021.9698880 (cit. on p. 19).

[76] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 770–778. DOI: https://doi.org/10.48550/arXiv.1512.03385 (cit. on p. 21).

[77] Z. Zhang, G. Dai, X. Liang, S. Yu, L. Li, and Y. Xie, "Can signal-to-noise ratio perform as a baseline indicator for medical image quality assessment," *IEEE Access*, vol. 6, pp. 11 534–11 543, 2018. DOI: 10.1109/ACCESS.2018.2796632 (cit. on p. 24).

[78] W. Xue, L. Zhang, X. Mou, and A. C. Bovik, "Gradient magnitude similarity deviation: A highly efficient perceptual image quality index," *IEEE transactions on image processing*, vol. 23, no. 2, pp. 684–695, 2013. DOI: 10.1109/TIP.2013.2293423 (cit. on pp. 24, 25).

[79] H. R. Sheikh and A. C. Bovik, "Image information and visual quality," *IEEE Transactions on image processing*, vol. 15, no. 2, pp. 430–444, 2006. DOI: 10.1109/TIP.2005.859378 (cit. on p. 25).

[80] J. Kochhar, *How to evaluate image quality in python: A comprehensive guide*, Accessed: 2024-08-18, 2020. [Online]. Available: https://medium.com/@jaikochhar06/how-to-evaluate-image-quality-in-python-a-comprehensive-guide-e486a0aa1f60 (cit. on p. 26).

[81] V. N, P. D, M. C. Bh, S. S. Channappayya, and S. S. Medasani, "Blind image quality evaluation using perception based features," in *2015 Twenty First National Conference on Communications (NCC)*, 2015, pp. 1–6. DOI: 10.1109/NCC.2015.7084843 (cit. on p. 26).

[82] N. Venkatanath, D. Praneeth, M. C. Bh, S. S. Channappayya, and S. S. Medasani, "Blind image quality evaluation using perception based features," in *2015 twenty first national conference on communications (NCC)*, IEEE, 2015, pp. 1–6. DOI: 10.1109/NCC.2015.7084843 (cit. on p. 26).

[83] P. E. Azun, *Analyzing mri image parameters based on signal-to-noise ratio*, Accessed: 5-Jul-2024, 2023. [Online]. Available: https://medium.com/@preciousebiteazun/analyzing-mri-image-parameters-based-on-signal-to-noise-ratio-577c685c5705 (cit. on p. 26).

A

Image Modalities and Layer Parameters for DL Models

Figure A.1 presents examples of images from the three different datasets used in this study, representing distinct image modalities.

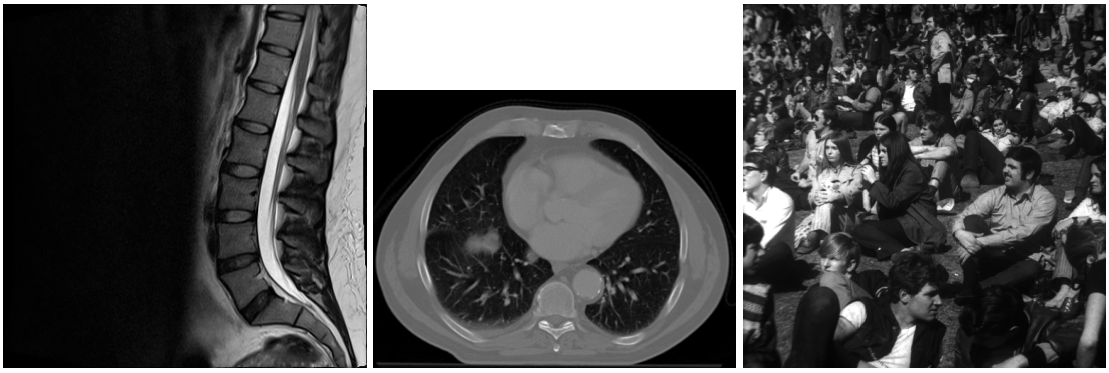

Figure A.1: Representation of different image modalities used in the study. From left to right: MRI scan of the spine, CT scan of the chest, and a natural greys-scale image of a crowd.

The tables below outline the layer parameters used to construct each DL model, providing insight into their respective architectures. Table A.1 provides details of the configuration for the FSRCNN model, while Tables A.2 and A.3 relate to the RDN model. Finally, Tables A.4 through A.6 outline the parameters for the ESRGAN. These tables include key information such as the number of filters, kernel sizes, strides, and activation functions used in each layer, illustrating the specific design choices made for each network.

Table A.1: Parameters of the FSRCNN.

Layer Type	Number of Filters	Kernel Size	Stride	Padding	Activation Function
Conv2D (1st)	56	5×5	1	same	ReLU
Conv2D (2nd)	12	1×1	1	same	ReLU
Conv2D (Repeated 12×)	12	3×3	1	same	ReLU
Conv2D (Pre-DeConv)	56	1×1	1	same	ReLU
Conv2DTranspose (Upsample)	1	9×9	2	same	-

Table A.2: Parameters of the RDN.

Layer	Number of Filters	Kernel Size	Stride	Padding	Activation Function	Operation
Conv2D (1st)	32	9×9	1	same	ReLU	-
Conv2D (2nd)	64	9×9	1	same	ReLU	-
RDB (Repeated 5×)	64	3×3	1	same	ReLU	-
Concatenate	-	-	-	-	-	All RDB Outputs
Conv2D (3rd)	64	3×3	1	same	ReLU	-
Conv2D (4th)	32	3×3	1	same	ReLU	-
Add	-	-	-	-	-	Input + Conv2D (4th)
Conv2DTranspose (Upscaling, 2×)	1	2×2	2	same	-	-
Conv2D (5th)	1	9×9	1	same	ReLU	-

Table A.3: Parameters of the RDB of the RDN.

Layer Type	Number of Filters	Kernel Size	Stride	Padding	Activation Function	Operation
Conv2D (1st)	64	3×3	1	same	ReLU	-
Concatenate	-	-	-	-	-	Input, Conv2D (1st)
Conv2D (2nd)	64	3×3	1	same	ReLU	-
Concatenate	-	-	-	-	-	Input, Conv2D (1st + 2nd)
Conv2D (3rd)	64	3×3	1	same	ReLU	-
Concatenate	-	-	-	-	-	Input, Conv2D (1st + 2nd + 3rd)
Conv2D (4th)	64	3×3	1	same	-	-
Add (Residual)	-	-	-	-	-	Input + Conv2D (4th)

Table A.4: Parameters of the generator of the ESRGAN.

Layer	Number of Filters	Kernel Size	Stride	Padding	Activation Function	Operation
Conv2D (1st)	64	3×3	1	same	ReLU	-
RRDB Block (Repeated 3×)	64	3×3	1	same	ReLU	-
Conv2D (2nd)	64	3×3	1	same	-	-
Add (Residual)	-	-	-	-	-	Input + Conv2D (2nd)
Conv2D (3rd)	256	3×3	1	same	-	
SubPixel (Upscaling, 2×)	-	-	-	-	-	-
ReLU	-	-	-	-	ReLU	-
Conv2D (4td)	64	3×3	1	same	ReLU	-
Conv2D (5th)	1	3×3	1	same	-	-

Table A.5: Parameters of the RDB of the generator of the ESRGAN.

Layer Type	Number of Filters	Kernel Size	Stride	Padding	Activation Function	Operation
Conv2D (1st)	64	3×3	1	same	ReLU	-
Concatenate	-	-	-	-	-	Input, Conv2D (1st)
Conv2D (2nd)	64	3×3	1	same	ReLU	-
Concatenate	-	-	-	-	-	Input, Conv2D (1st + 2nd)
Conv2D (3rd)	64	3×3	1	same	ReLU	-
Concatenate	-	-	-	-	-	Input, Conv2D (1st + 2nd + 3rd)
Conv2D (4th)	64	3×3	1	same	ReLU	-
Concatenate	-	-	-	-	-	Input, Conv2D (1st + 2nd + 3rd + 4th)
Conv2D (5th)	64	3×3	1	same	-	-
Multiply	-	-	-	-	-	0.2 × Conv2D (5th)
Add (Residual)	-	-	-	-	-	Input + (0.2 × Conv2D (5th))

Table A.6: Parameters of the discriminator of the ESRGAN.

Layer Type	Number of Filters	Units	Kernel Size	Stride	Padding	Activation Function	Dropout Rate
Conv2D (1st)	64	-	3×3	1	same	ReLU	-
Conv2D (2nd)	64	-	3×3	2	same	ReLU	-
Conv2D (3rd)	128	-	3×3	1	same	ReLU	-
Conv2D (4th)	128	-	3×3	2	same	ReLU	-
Conv2D (5th)	256	-	3×3	1	same	ReLU	-
Conv2D (6th)	256	-	3×3	2	same	ReLU	-
Conv2D (7th)	512	-	3×3	1	same	ReLU	-
Conv2D (8th)	512	-	3×3	2	same	ReLU	-
Dense	-	1024	-	-	-	-	-
Dropout	-	-	-	-	-	-	0.4
Flatten	-	-	-	-	-	-	-
Dense	-	1	-	-	-	-	-

B

Additional Results

Figure B.1 presents the activity masks generated by the PIQE metric for the super-resolved images produced by the three DL-models, using the LR image created through the image-domain degradation method. More regions with a value of 1, representing high activity, indicate that FSRCNN and RDN have more distortions than ESRGAN.

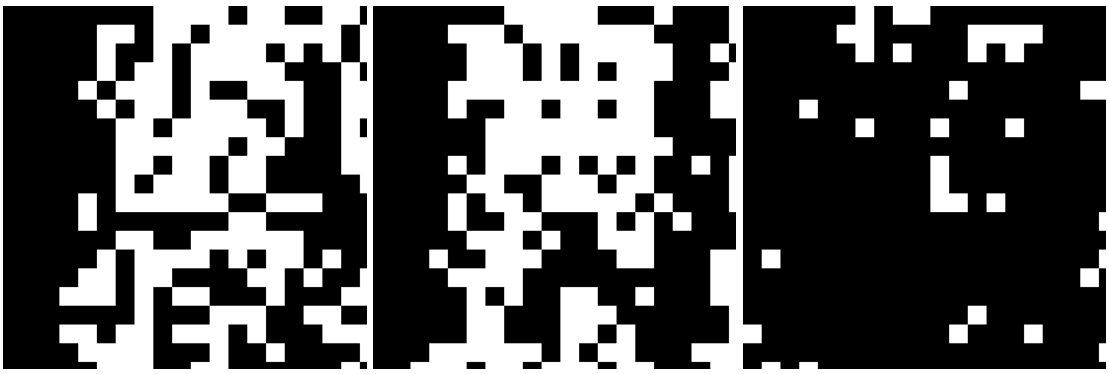

Figure B.1: Activity masks for the super-resolved images generated by FSRCNN (left), RDN (middle) and ESRGAN (right).

Figure B.2 shows the full images produced by applying the re-trained ESRGAN and BCI to the original HR image. Both methods display a similar level of detail and sharpness, with comparable performance in preserving textures and edge clarity.

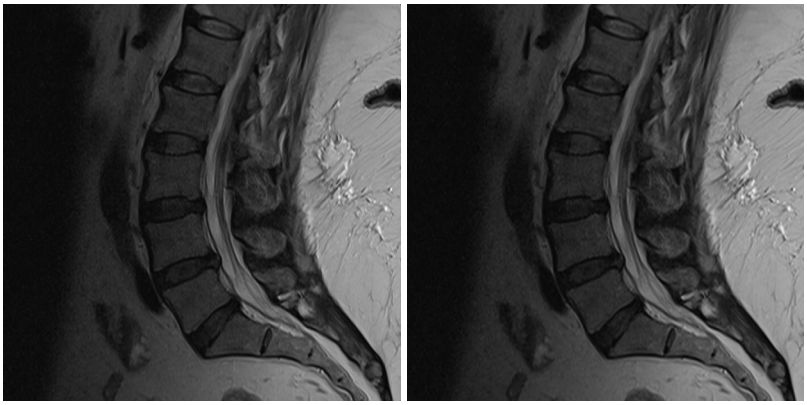

Figure B.2: The ESRGAN (left) and BCI (right) enhanced versions from the original HR image.