



University of Brasilia – FGA/UnB
Software Engineering

**CONTEXT-DEPENDENT PROBABILISTIC
PRIOR INFORMATION STRATEGY FOR MRI RECONSTRUCTION**

GABRIEL GOMES ZIEGLER

Advisor: CRISTIANO JACQUES MIOSSO, PhD
Co-advisor: DAVI BENEVIDES GUSMÃO, MSc



UNB – UNIVERSITY OF BRASILIA

FGA – GAMA FACULTY

SOFTWARE ENGINEERING

**CONTEXT-DEPENDENT PROBABILISTIC PRIOR INFORMATION
STRATEGY FOR MRI RECONSTRUCTION**

GABRIEL GOMES ZIEGLER

ADVISOR: CRISTIANO JACQUES MIOSSO, PhD

Co-ADVISOR: DAVI BENEVIDES GUSMÃO, MSc

**BACHELOR THESIS
SOFTWARE ENGINEERING**

BRASILIA/DF, MAY 2021

UNB – UNIVERSITY OF BRASILIA

FGA – GAMA FACULTY

SOFTWARE ENGINEERING

**CONTEXT-DEPENDENT PROBABILITISTIC PRIOR INFORMATION
STRATEGY FOR MRI RECONSTRUCTION**

GABRIEL GOMES ZIEGLER

**SUBMITTED IN PARTIAL FULFILLMENT FOR THE BACHELOR DEGREE IN SOFTWARE
ENGINEERING (HONOR'S DEGREE) TO THE DEPARTMENT OF ENGINEERING OF
UNIVERSITY OF BRASILIA**

APPROVED BY:

Cristiano Jacques Miosso, PhD

(Advisor)

Prof. Adson Rocha, PhD

CATALOGUE SHEET

ZIEGLER, GABRIEL

Context-Dependent Probabilistic Prior Information Strategy for MRI Reconstruction,
[Federal District], 2021.

46p., 210 × 297 mm (FGA/UnB Gama, BEng in Software Engineering, 2020).

Bachelor Thesis, UnB Gama Faculty, Software Engineering

- | | |
|-----------------------------|--|
| 1. Compressed Sensing | 2. Magnetic Resonance Image Reconstruction |
| 3. <i>Prior Information</i> | 4. Deep Learning |
| I. FGA UnB/UnB. | |

REFERENCE

ZIEGLER, GABRIEL (2021). Context-Dependent Probabilistic Prior Information Strategy for MRI Reconstruction. Bachelor Thesis, Software Engineering, UnB Gama Faculty, University of Brasilia, Brasilia, DF, 46p.

ASSIGNMENT OF RIGHTS

AUTHOR: Gabriel Gomes Ziegler

TITLE: Context-Dependent Probabilistic Prior Information Strategy for MRI Reconstruction

DEGREE: Bachelor of Engineering in Software Engineering

YEAR: 2021

The University of Brasilia is granted permission to reproduce copies of this thesis at the end of the course and to lend or sell such copies for academic and scientific purposes only. The author reserves other publishing rights and no part of this thesis may be reproduced without the written permission of the author.

gabrielziegler3@gmail.com

Brasília, DF – Brasil

RESUMO

A obtenção de imagens de ressonância magnética de alta qualidade é uma tarefa árdua pelo funcionamento das máquinas e pela complexidade dos tecidos analisados, tornando praticamente impossível a obtenção de todas as medidas do objeto de estudo em um exame de breve execução. Para lidar com este problema, diversas técnicas foram aplicadas para gerar imagens de alta qualidade com quantidades menores de medidas, entre elas, Compressed Sensing (CS) tem sido a técnica com melhores resultados e mais aprofundada no contexto de reconstrução de imagens de ressonância magnética nas últimas duas décadas. CS é capaz de produzir imagens de altíssima qualidade com proporções muito menores de amostragens, como 10%, 15% do total de medidas disponíveis utilizando técnicas de *undersampling* por sua boa capacidade de lidar com sinais esparsos.

O desempenho do CS evoluiu com as contribuições de uso de filtros esparsificantes e com o uso de informação a priori. Recentemente, o estado-da-arte veio sendo composto pelo uso de redes neurais profundas, muitas vezes em conjunto com o uso de CS e outras de maneira singular. Desta maneira, esta monografia tem como intuito trazer o uso de CS com pré-filtragem e aplicar o uso de redes adversárias generativas para a extração de informação a priori.

A metodologia é composta pela utilização de formas variadas de CS para a reconstrução de imagens comparando métodos, pré-processamentos e parâmetros, a fim de explicitar os diferentes resultados atingíveis com CS. Experimentos com redes adversárias generativas também foram conduzidos com o intuito de demonstrar a capacidade das mesmas para geração de sinais com alta fidelidade.

Os resultados obtidos demonstram que o uso de pré-filtragem é preferível para a reconstrução de imagens, melhorando as métricas de relação sinal-ruído (SNR), relação sinal-ruído de pico (PSNR) e similaridade estrutural (SSIM). Os resultados das GANs mostram que estas são capazes de gerar sinais com alto teor de realismo quando comparadas as imagens criadas com as imagens de treino. Os próximos passos da pesquisa envolvem aplicar o uso de informação a priori de forma manual e posteriormente usando GANs, comparando todas as técnicas e transformações aplicadas.

Contents

1	Introduction	1
1.1	Context	1
1.2	Scientific Problem Definition and Proposal	2
1.3	Objectives	4
1.3.1	General Objective	4
1.3.2	Specific Objectives	4
2	MRI Concepts and Compressed Sensing State-of-Art	5
2.1	Magnetic Resonance Imagery	5
2.1.1	K-space	5
2.1.2	Undersampling	6
2.2	Compressed Sensing	7
2.2.1	Introduction	7
2.2.2	Reconstructing using the LP Regularization	8
2.3	Prior Information	8
2.3.1	Prior Information Retrieval	9
3	Reconstructing Experiments and Future Steps	11
3.1	Metrics and Evaluation Criteria	11

3.2	1-D Direct vs Indirect L1-Minimization	12
3.3	2D Compressed Sensing Reconstruction with Pre Filtered Signal	12
3.3.1	Spiral Undersampling	13
3.3.2	Pre Filtering Sparsifying Transform	13
3.4	Dynamic Sagittal Cardiac Case Study	14
3.4.1	Radial Undersampling	15
3.4.2	Deterministic Prior Information	15
3.4.3	CoDePPI: Context-Dependent Probabilistic Prior Information .	16
3.4.4	<i>L1</i> -minimization Adaptation	20
3.4.5	Parameter Fine-tuning	20
3.5	Future Contribution Literature Review and Experiment	21
3.5.1	Preliminary Tests with Generative Adversarial Networks	21
3.5.2	Data Transformation	22
3.5.3	Generator Network Architecture	22
3.5.4	Discriminator Network Architecture	23
4	CS Reconstruction and Preliminary Deep Learning Results	24
4.1	1-D Compressed Sensing Reconstruction	24
4.2	Compressed Sensing Reconstruction with Pre-Filtered Signal	25
4.2.1	Phantom Reconstruction Evaluation	26
4.2.2	Sagittal Head MRI Reconstruction	26
4.2.3	Sagittal Head Reconstruction Evaluation	27
4.3	Dynamic Sagittal Cardiac Case Study	27
4.3.1	Deterministic Prior Information	27

4.3.2	Context-Dependent Probabilistic Prior Information	28
4.3.3	Dynamic Sagittal Cardiac Reconstruction Evaluation	29
4.4	Future Contributions: Deep Learning Applications	30
4.5	Artificial Neural Networks	30
4.5.1	Biological Inspirations	30
4.5.2	Neuron	31
4.5.3	Multilayer Perceptron	31
4.5.4	Activation Functions	33
4.5.5	Loss Functions	35
4.5.6	Backpropagation	36
4.5.7	Gradient Descent	37
4.6	Generative Adversarial Networks	37
4.6.1	Preliminary Tests with Generative Adversarial Networks	38
5	Conclusion	41

List of Tables

4.1	Phantom reconstruction metrics.	26
4.2	Sagittal reconstruction metrics.	27
4.3	Dynamic sagittal cardiac reconstruction metrics in terms of prior information usage.	30
4.4	Activation functions and respective derivatives.	33
4.5	Loss functions and formulas.	36

List of Figures

2.1	Single-coil k-space points from fully-sampled knee. Source: [1, 2]	6
2.2	Single-coil fully-sampled knee spatial image. Source: [1, 2]	6
2.3	Under-sampling patterns. (a) Cartesian undersampling , (b) radial undersampling , (c) spiral undersampling , (d) isolated samples in the k-space, according to the realisation of a random process [3].	7
2.4	Masks with 15% points of deterministic prior information extraction. . .	10
2.5	Non-deterministic prior information for sagittal head image.	10
3.1	First 10 random 1-D signals	12
3.2	Spiral undersampling method with 30.95% data points.	13
3.3	2-D High pass horizontal filter.	14
3.4	2-D High pass vertical filter.	14
3.5	2-D High pass diagonal filter.	14
3.6	First frame from the dynamic cardiac Magnetic Resonance Imaging (MRI) exam.	14
3.7	Radial undersampling trajectory with 19% of data points.	15
3.8	Example of gaussian distributions with four different levels of variances. .	17
3.9	Fully-sampled cardiac MRI frame.	18
3.10	Manual segmentation: the lighter the colour, the more pixels are dislocated in the region throughout frames.	18

3.11	2D <i>weights</i> matrix.	19
3.12	3D <i>weights</i> matrix visualization.	19
3.13	2D <i>weights</i> matrix visualization.	20
3.14	3D <i>weights</i> matrix visualization.	20
3.15	Sample of digits from MNIST	21
3.16	Generator network architecture of a 3-D image. Source: [4, 5]	23
3.17	Discriminator network architecture of a 3-D image. Source: [4, 5]	23
4.1	SNR x L size for direct and indirect L1-minimization for 200 random 1D signals	25
4.2	Zero-filled	26
4.3	L1-minimization	26
4.4	L1-minimization with pre-filtering	26
4.5	Shepp-Logan phantom reference image	26
4.6	Zero-filled	27
4.7	L1-minimization	27
4.8	L1-minimization with pre-filtering	27
4.9	Reference image	27
4.10	Deterministic prior information with 49.92% data points proportion.	28
4.11	Context-dependent probabilistic prior information with 49.92% data points proportion.	28
4.12	On the left: zero-filled reconstruction. On the right: no prior information reconstruction	29
4.13	From left to right: deterministic prior information; context-dependent probabilistic prior information; reference image	29
4.14	Schematic of an Artificial Neuron. Source: [6]	31

4.15 Artificial Neural Networks (ANN) Architecture Sample.	32
4.16 Rectified linear units (ReLU) activation function	34
4.17 Leaky ReLU activation function	34
4.18 <i>Tanh</i> activation function. Source: [7]	35
4.19 Sigmoid activation function. Source: [7]	35
4.20 Gradient descent example visualization. Source: [8]	37
4.21 Generative Adversarial Network (GAN) training diagram. Source: [9] . .	38
4.22 Discriminator fake loss over epochs	39
4.23 Discriminator real loss over epochs	39
4.24 Generator loss over epochs	39
4.25 GAN generated MNIST digits	40

NOMENCLATURE AND ABBREVIATIONS

MRI Magnetic Resonance Imaging	x
MR Magnetic Resonance	1
ANN Artificial Neural Networks	xii
DL Deep Learning	2
ML Machine Learning	2
GAN Generative Adversarial Network	xii
CNN Convolutional Neural Network	3
GPU Graphics Processing Units	2
CV Computer Vision	2
NLP Natural Language Processing	2
CS Compressed Sensing	1
MLP Multilayer Perceptron	31
DFN Deep Feedforward Networks	31
SNR Signal-to-Noise Ratio	1

SSIM Structural SIMilarity	1
PSNR Peak Signal-to-Noise Ratio	1
DCGAN Deep Convolutional GAN	21
ReLU Rectified linear units	xii
SGD Stochastic Gradient Descent	37
IRLS iteratively reweighted least squares	9
DCT Discrete Cosine Transform	7

1 INTRODUCTION

In this thesis, we propose a novel context-dependent probabilistic prior information extraction algorithm that takes advantage of motion information across frames in a dynamic MRI to weight the confidence that those extracted positions are effectively part of a support structure. In our experiments, our approach provides better reconstruction in terms of the evaluated metrics (Signal-to-Noise Ratio (SNR), Peak Signal-to-Noise Ratio (PSNR), Structural SIMilarity (SSIM), Normalized Mean Squared Error (NMSE), Mean Squared Error (MSE)) and also in computation time when contrasting with classic prior information and with Compressed Sensing (CS) reconstructions that use no prior information. Achieving higher quality with a reduced number of samples allows faster exam procedures, making MRI cheaper, faster and more convenient for both patients and clinics, which is our ultimate goal.

1.1 Context

MRI is a widely used imaging modality in medical practice because of its great tissue contrast capabilities, it has evolved into the richest and most versatile biomedical imaging technique today [10], making MRI the best option for medical imaging whenever it is possible to use.

However, like everything in life, there is a trade-off to consider when using MRI. Typically, reconstructing an MRI is an ill-posed linear inverse task (a problem that has either none or infinite solutions in the desired class). Problems of this nature impose a trade-off between *accuracy* and *speed* [11]. The information obtained from Magnetic Resonance (MR) is commonly represented by individual samples in the k-space, which translates to the Fourier transform of the image to be reconstructed [12]. This MR sampling sparse nature makes CS a liable technique to use when reconstructing MRI, hence we here propose a novel CS prior information approach for better results.

CS has been for years the state-of-art technique in MRI reconstruction and has been improved later by the use of sparsifying pre-filtering techniques and prior information [13],

[14]. CS uses the premise that given a signal with a sparse representation in some known domain, it is possible to reconstruct the signal using limited linear measurements taken from a non-sparse representation.

There is still room for improvement in signal processing approaches for MRI reconstruction like the one novel implementation of the prior information algorithm that we present in this thesis, but ultimately, the task of reconstructing a signal from very few extracted measurements can be mostly improved with Machine Learning (ML) techniques. Many times, classical approaches serve as the starting point where ML models should start tackling the problem. In this case, we propose a novel prior information implementation and also investigate possibilities to improve the quality of this prior information even more by utilizing the so-called GANs.

ML methods have been utterly developed and improved recently with the use of higher computing power derived from the invention of Graphics Processing Units (GPU) and other hardware improvements, allowing ANN to come to practicality. These ANN models, often referenced as Deep Learning (DL), have become the state-of-art in various areas, such as Computer Vision (CV), Natural Language Processing (NLP), Recommendation Systems, amongst other fields [15, 16, 17]. These fast-paced developments led to improvements in medical data processing using DL as well. ML techniques can be used in several different manners to improve medical analysis, here we focus on applying GAN in the process of attaining improved prior information to feed the CS algorithm obtaining higher signal-to-noise ratios and faster computation procedures.

1.2 Scientific Problem Definition and Proposal

MRI is great for high-quality tissue images and definitely one of the highest image-quality medical imaging modalities, but there are some drawbacks: MRI exams are often very long and require the patient to be in a static position throughout the whole process, this makes the exam challenging for patients that have difficulties in keeping a still position for several minutes. Another intrinsic complication in MRI procedures is that it is extremely complicated to extract images from moving tissues like a beating heart or flowing blood veins as that would require an enormous amount of samples, which with current technologies used in clinics is not viable. Algorithms that reconstruct MRI try to tackle this sampling issue by producing the best possible quality images from the least amount of samples collected, making the exams faster and less sample-dependent.

CS algorithms have been the state-of-art in MRI reconstruction for the past few years, and now with the advances of DL, new techniques are being produced taking advantages of

how ANNs are powerful in imaging processing, especially Convolutional Neural Network (CNN)s and, more recently, GANs are becoming the new state-of-art techniques in several computer vision areas. Most CS contributions without the use of machine learning cite the usage of sparsifying pre-filtering techniques and prior information that have been proven to improve efficiency and achieve better reconstructions [14, 13]. More recently, the contributions in MRI reconstruction has been more focused in deep learning approaches, with some architectures using CS along ANNs [18, 19, 20, 21, 22, 23].

CS reconstructions often use a preprocessing step called prior information which focuses on the extraction of support positions – regions that normally would not move from a frame to another – and when these positions are fed to the l_p -minimization algorithm (one of the possible ways to solve), the reconstructed image quality is improved significantly [14, 12]. Prior information is normally generated by mathematical approaches like filtering and thresholding on the images and can be leveraged from previous and next frames in the same MRI exam to even medical records from previous scans. These information extraction procedures oftentimes are restricted to few frames and do not take into account the nature of organs and tissues structures. Another observation from this classical prior information extraction is that these support positions are binary, they either are part of a support position or they are not. One could think that using a non-deterministic approach to this technique would likely correct some errors introduced in-between frames and that is exactly what happens when the extracted prior information is treated as a probabilistic support position [24]. Although non-deterministic do perform better than deterministic ones, they still do not consider the different level of confidence in each region of the subject – that is, some regions are different in terms of motion, hence they should be treated as different levels of confidence, which is right what we investigate within this work.

Additionaly, DL models are constantly improving image reconstruction quality like the ones mentioned in [1, 2] and this implies that there is a lot of room for improvement towards MRI and following the same logic, we believe that prior information can be even more improved by the usage of GANs for prior information extraction and so we also documented our research in such contexts aiming towards a more detailed research project in a master’s degree, where more experimentation can be done to validate our hypothesis.

1.3 Objectives

1.3.1 General Objective

Ultimately, the main goal behind this research can be roughly summarized in two intended contributions: faster MRI exams and better MR image quality. To achieve this goal, we have set minor goals that are described more deeply further. In terms of research, our goal is to introduce a novel context-dependent probabilistic prior information algorithm utilizing the different motion across the frames in a dynamic MRI for higher confidence support position extraction. The expected outcome of this algorithm is to improve reconstruction in quality and in runtime directly impacting both our main goals. Additionally, we also present fundamentals and preliminary experiments for a sequel to this study using DL techniques like GANs.

1.3.2 Specific Objectives

In order to achieve the general objective described above, we have set the following specific goals:

- Implement all the CS procedures in Python.
- Implement direct and indirect CS MRI reconstruction algorithm using undersampled *k*-space measurements.
- Demonstrate experiments with CS using pre-filtering and classical prior-information.
- Demonstrate our context-dependent probabilistic prior information algorithm and compare with other prior-information techniques.
- Test our algorithm with fastMRI samples provided by the NYU fastMRI Initiative database (fastmri.med.nyu.edu) [1, 2].
- Present a case study with a dynamic MRI using the context-dependent probabilistic prior information algorithm to show how our technique could eventually impact in real exams.

2 MRI CONCEPTS AND COMPRESSED SENSING STATE-OF-ART

2.1 Magnetic Resonance Imagery

MRI is an indirect process that produces cross-sectional images with high spatial resolution from nuclear magnetic resonances, gradient fields and hydrogen atoms of the subject's anatomy [25]. The acquisition of these measurements is performed by a measuring instrument called *receiver coil* and it can be done by using one receiver coil or in some cases with multiple coils [1, 2]. These receiver coils are placed in proximity to a specific region in the subject to be imaged. During the imaging process, the MRI machine generates a sequence of spatially and temporally-varying magnetic fields which induce the body to emit resonant electromagnetic response fields which are then measured by the receiver coil [1, 2].

2.1.1 K-space

The k-space is the output generated by the MRI machine scan after extracting measurements from a given subject tissue. The k-space is represented in the spatial frequency in two or three dimensions of a subject and may also be referred to as the Fourier space. Its representation contains an implicit sparsity that is exploited when performing undersampling [26] and reinforce the usage of algorithms like CS for MRI reconstruction as CS depends on signals that have a sparse representation in an orthonormal basis [27]. In essence, the k-space is the signal representation in spatial frequency domain.

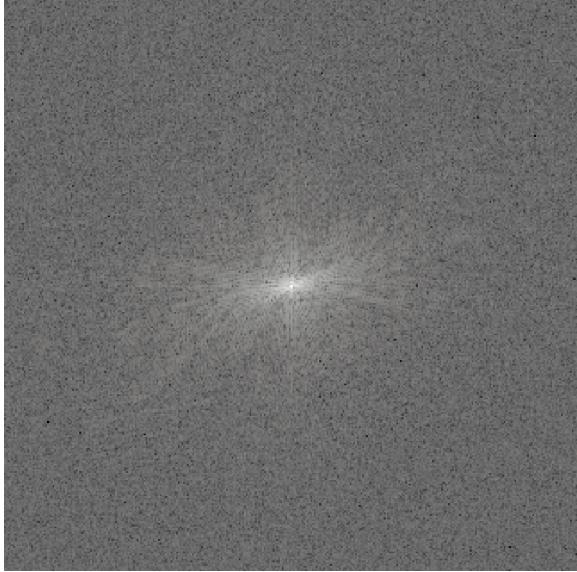


Figure 2.1. Single-coil k-space points from fully-sampled knee.
Source: [1, 2]



Figure 2.2. Single-coil fully-sampled knee spatial image.
Source: [1, 2]

2.1.2 Undersampling

The time required to acquire all the measurements responses from every single atom in a subject would be extremely high and problematic to everyone involved (patients, doctors and clinics). The way machines can do faster MRI is by performing *undersampling*, also referred as subsampling and sampling, when scanning the subject.

Undersampling is performed by giving the machine a known prescribed path in which it will extract measurements from the multidimensional k-space representation. This allows machines to collect only a fraction of data measurements needed for image reconstruction hence speeding up the data acquisition process without critical quality loss.

There are some undersampling patterns to use and each has its benefits depending on several parameters, such as: the subject's region extraction, the algorithm used for reconstruction, acquisition time.

In the figure below we can see some of the most used patterns. In this research, we will focus mostly on the radial, spiral and cartesian undersampling method which we later use in our experiments.

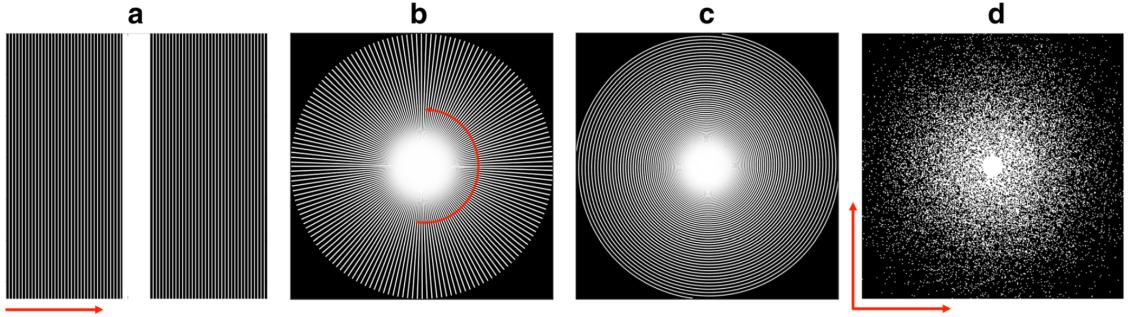


Figure 2.3. Under-sampling patterns. (a) **Cartesian undersampling**, (b) **radial undersampling**, (c) **spiral undersampling**, (d) **isolated samples in the k -space**, according to the realisation of a random process [3].

In MRI, the undersampling techniques are used to acquire η elements where $\eta < 2\eta < N$, violating the Nyquist criterion.

2.2 Compressed Sensing

2.2.1 Introduction

Compressed sensing, often referred to as Compressive Sensing, represented a major breakthrough in signal processing when it was introduced by Donoho, Candès, Romberg, and Tao in 2004 [27, 28, 29] since it allows a sampling at a rate much lower than the Nyquist-Shannon's theorem: the sampling signal rate must be at least twice the maximum frequency present in the signal (Nyquist rate).

The idea behind the CS theory was inspired from questioning the necessity of extracting large portions of samples when much of these very samples are discarded, exposing the inefficiency of trying to gather all signal.

“Why go to so much effort to acquire all the data when most of what we get will be thrown away? Can we not just directly measure the part that will not end up being thrown away?” [27]

CS is a powerful algorithm that implements a novel technique for the acquisition of signals of sparse or compressible nature. CS theory aims to provide accurate reconstruction of unknown sparse signals from underdetermined linear measurements l . CS parts from the principle that if given x , a digital image or signal has a sparse representation in an orthonormal basis (e.g. wavelet, Fourier, Discrete Cosine Transform

(DCT), etc), then the N most important coefficients in that expansion allow reconstruction with l_2 error $O(N^{1/2-1/p})$ [27].

A signal x is known to have a sparse representation if there is a deterministic and invertible matrix $T_{N \times N}$ so that the transformed vector \hat{x} is composed of most of its N components equal to zero [12].

$$\hat{x} = Tx \quad (2.1)$$

MRI l extracted measurements from the MR scanner correspond to some of the coefficients in $\hat{x} = Tx$. In MRI context, the so-called *k-space* is a sparse representation of the image to be reconstructed in the frequency domain and when it is undersampled, the Nyquist criterion is violated [26]. This occurs due to the fact that it is extremely expensive to acquire at $l = 2\eta$, as suggested by the Nyquist criterion [14]. These are favorable conditions for the CS theory, hence all thousands of papers citing the usage of CS for MRI reconstruction and state-of-the-art results involving any application of CS for MRI reconstruction.

2.2.2 Reconstructing using the LP Regularization

Given an N -dimensional complex discrete-time signal x , that is compressible by a linear transformation with a sparse representation tested with 2.1, it is said that this signal is x can be fully reconstructed from an l -dimensional vector of measurements defined as

$$b = Mx \quad (2.2)$$

2.3 Prior Information

The application of prior information in MRI reconstruction was first introduced with the general idea of exploiting the common information shared throughout sequential frames acquired from the MR scan [12, 14]. This would not only improve the reconstruction of MR images but would also make dynamic scans more feasible as large portions of the image could be used as prior information for the reconstruction of future and previous frames. The MRI exam is known for requiring the patient to stand still throughout the exam so that there are not grand differences between different frames.

A dynamic cardiac MRI for instance is naturally a situation that will be impossible for the subject to be still throughout the exam, as the heart will keep beating constantly; hence increasing the level of difficulty for the reconstruction. Although a dynamic cardiac exam has this innate motion characteristic, it also contains crucial portions of *support regions* shared between most of the frames that can be exploited to improve the reconstruction. These support regions are essentially structures that will hold – practically – the same position therefore, these are elements in the image that share the same information and prior information application exploits this nature by slightly reducing the number of variables in our underdetermined system in the CS algorithm.

2.3.1 Prior Information Retrieval

To leverage prior information for a frame – as it was first introduced –, one could apply edge detection filters to first generate support position candidates and then extract the position of these edges and apply a τ factor to increase the value in these specific positions when running the CS algorithm, based on an iteratively reweighted least squares (IRLS) method [14]. In practice, this operation tells the l_p -minimization that these values with a higher magnitude are likely a values in a support position Φ . These procedures are extensively explored in the experiments documented in the experiments chapter.

Prior information was then introduced as a deterministic approach – whether the pixel was a support position or it was not. This approach, of course, raises a few questions: what if the subject moved slightly and the support position has been moderately altered? What if the next frame contains support positions that are getting narrower or wider? E.g. Brain MRIs, where the cranial structure’s diameter is increased or decreased frame-by-frame. Prior information theory responds to these question with a high tolerance impact in the reconstruction. It has been documented that using the same amount of mistaken support positions and correct ones will improve SNR in the next frame’s reconstruction [14].

Alternatively, a non-deterministic approach for prior information generation was proposed to achieve more robustness based on more accurate information. This technique uses a Gaussian – or normal – distribution with fixed covariance matrix values for each pixel of the support positions in the space domain instead of a deterministic 0 or 1 value. Using probabilistic support positions improves reconstruction metrics as shown in [24].

The images below are an example to represent a 15% deterministic prior information of a sagittal head MR scan from filters of horizon, vertical and diagonal borders detection.



Figure 2.4. Masks with 15% points of deterministic prior information extraction.

Additionaly, these very same Φ positions have already been represented in terms of probabilities [24] and they look slightly different, where the central positions of the detected edges are greater in magnitude and the edges are smaller, representing different levels of confidence. Below we display what our horizontal low-pass filtered sagittal head image would look like in the non-deterministic approach.

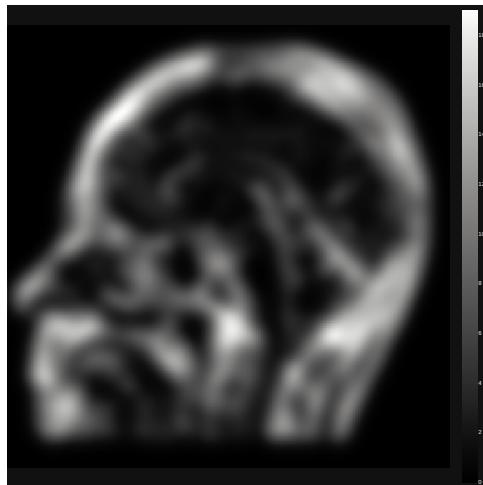


Figure 2.5. Non-deterministic prior information for sagittal head image.

As expected, the borders look very blurry and thicker than the ones presented in the deterministic prior information 2.4. This is useful because it allows for noise/error smoothing between frames and slices.

3 RECONSTRUCTING EXPERIMENTS AND FUTURE STEPS

The following experiments have the purpose to elucidate how the application of pre-filtering and different prior information strategies impact MRI reconstruction in different scenarios. Later, we present a signal generation experiment with GANs aiming to utilize this technology in further research continuation. The generation test is a step before implemeting these neural networks for prior-information leverage on MRI scans.

3.1 Metrics and Evaluation Criteria

All the documented experiments below were evaluated using a set of metrics widely used in the MRI context to analyse image reconstruction quality. The metrics and their respective meanings are:

- PSNR: represents the ratio between the power of the maximum magnitude value in the image and the power of corrupting noise that impacts the fidelity of the reconstruction. PSNR is represented in dB and the greater it is, the greater image quality perceived.
- SSIM: ranges from 0 to 1 and measures the similarity between two images, the closer to 1, the more the two signals are similar.
- SNR: similarly to the PSNR, the SNR measures the level of a desired signal to the level of noise in the reconstructed signal. It is also represented in dB and the greater, the better.
- MSE: measures the mean of the squares of the errors – that is, the difference between the reconstructed signal and the ground truth. The lower the error, the better the reconstruction.
- NMSE: computes the MSE normalized by signal power. NMSE is widely used in

research papers and is the primary measure of choice in fastMRI dataset, but it contains certain bias towards smoothness instead of sharpness [1, 2].

3.2 1-D Direct vs Indirect L1-Minimization

L1-minimization admits both direct and indirect approaches, in which there is the accuracy x resources trade-off. The direct method often produces a higher quality reconstruction but is very memory consuming, whilst the indirect method loses a little bit of quality, but requires much less memory to compute the equations system.

To visualize this trade-off, we have created 200 random 1-D arrays ranging values from the standard normal distribution and have taken 10% of data points randomly to reconstruct the whole signal using different L sizes.

L is the number of linear measurements extracted from the 1-D signal. The figure below displays the first 10 signals created in the standard normal distribution range.

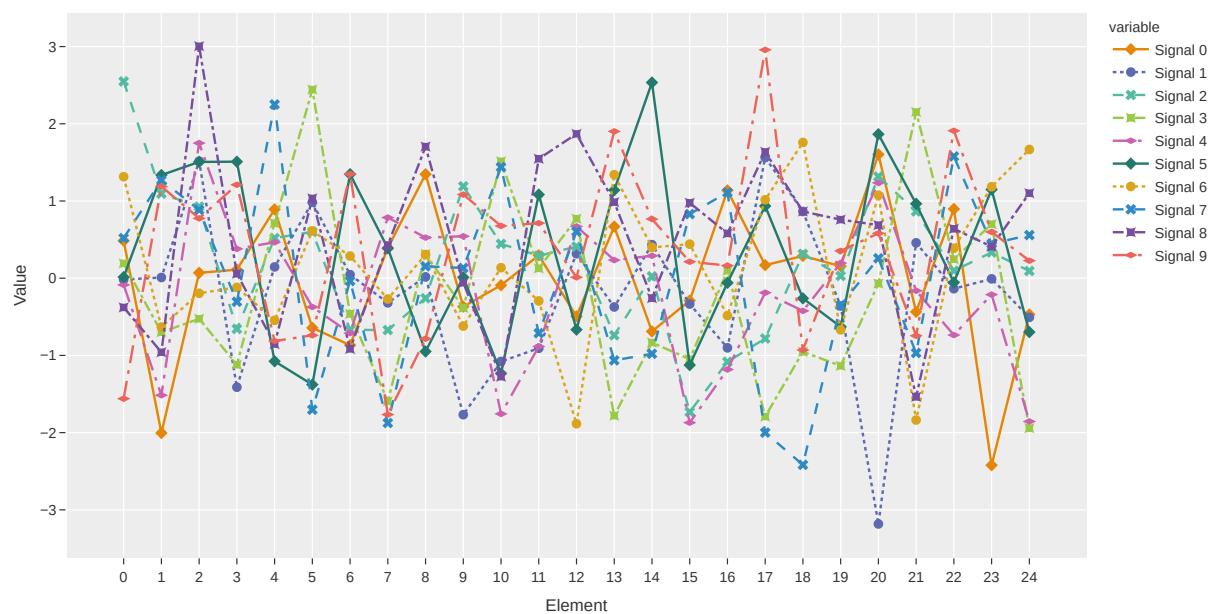


Figure 3.1. First 10 random 1-D signals

3.3 2D Compressed Sensing Reconstruction with Pre Filtered Signal

In order to experiment the sparsifying method of applying pre-filtering processing to the input measurements in the k-space for 2D images, we have conducted experiments

some different images. A simple 2D image of the well-known Shepp-Logan phantom [30] for baseline is used for the first experiment and two MR (sagittal head slice and a singlecoil knee obtained from the NYU fastMRI Initiative database (fastmri.med.nyu.edu) [2, 1]) images that are considerably more complex than the phantom for the reconstruction task.

Within these images, *SNR* improvement gained by applying sparsifying pre-filtering [12, 13] can be best visualized and confirmed for these different scenarios.

3.3.1 Spiral Undersampling

We then created a phantom image with dimension of 256×256 , hence having 65536 data points, using the phantominator python module. Then, we simulated an undersampled phantom image by applying the spiral undersampling pattern achieving approximately 30.95% of data points from the fourier space which resulted in a matrix with 20285 non zero elements. The same procedure was applied to the sagittal head image.

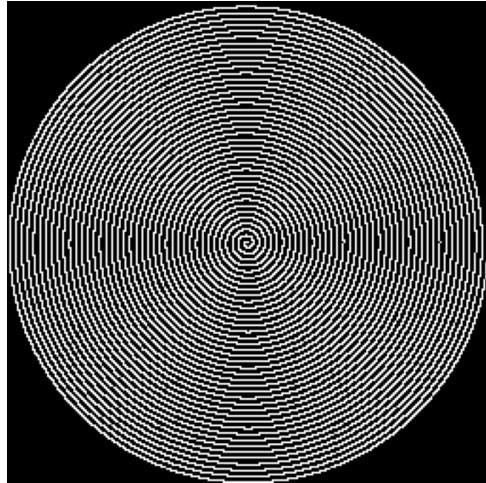


Figure 3.2. Spiral undersampling method with 30.95% data points.

3.3.2 Pre Filtering Sparsifying Transform

For the pre-filtering step, $f = 3$ is used where f is the number of filters applied to increase sparsity in the signal to be reconstructed. The filters are all 2×2 matrices and increase the sparsity in the signal from different perspectives, using more filters leverages the ability to sparsify the signal. The 3 filtered images are then composed into one single image containing the highest gain each filter could provide given a single pixel in the image [13]. The different filters used can be better visualized from the figure below.

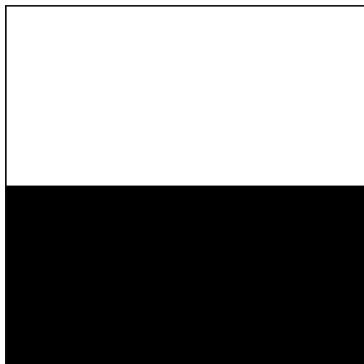


Figure 3.3.
2-D High pass
horizontal filter.

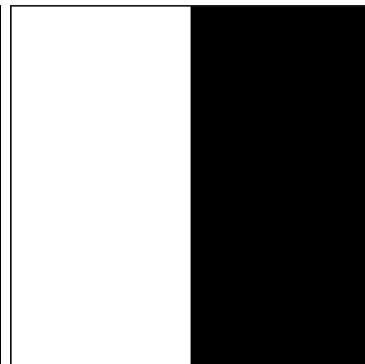


Figure 3.4.
2-D High pass
vertical filter.

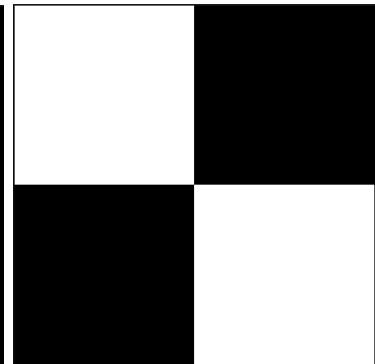


Figure 3.5.
2-D High pass
diagonal filter.

The pre-filtering method is evaluated against the zero-fill reconstruction method (used as a dummy baseline) and an L1-minimization method without pre-filtering with the very same parameters used in the pre-filtering L1-minimization.

3.4 Dynamic Sagittal Cardiac Case Study

The subject is composed of a dynamic sagittal cardiac MRI where the patient's heart is continuously beating. We then extract the first 10 frames from the video. Our reference image looks like this:

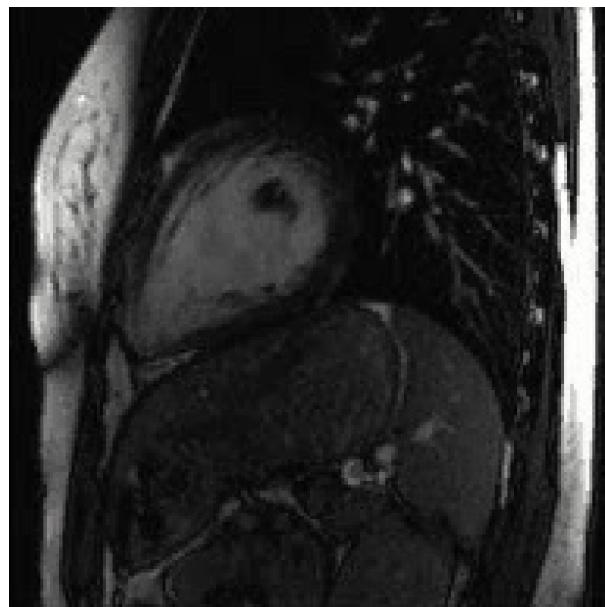


Figure 3.6. First frame from the dynamic cardiac MRI exam.

The first is the one that we do not possess any prior information from previous frames,

hence we perform an l_1 -minimization with pre-filtering step with the same parameters and conditions established in the previous experiments, but without prior information. From the second frame onwards, we will always have the previous frames to look into and extract prior information from support positions. In our experiments, we use the first frame's reconstruction to generate our Φ positions for both the deterministic and context-dependent probabilistic prior information strategies.

3.4.1 Radial Undersampling

For this experiment, we decided to use a radial undersampling mask with 60 lines – which corresponds to 19% of all the data points, concentrating most of the measurements in the low frequency regions.

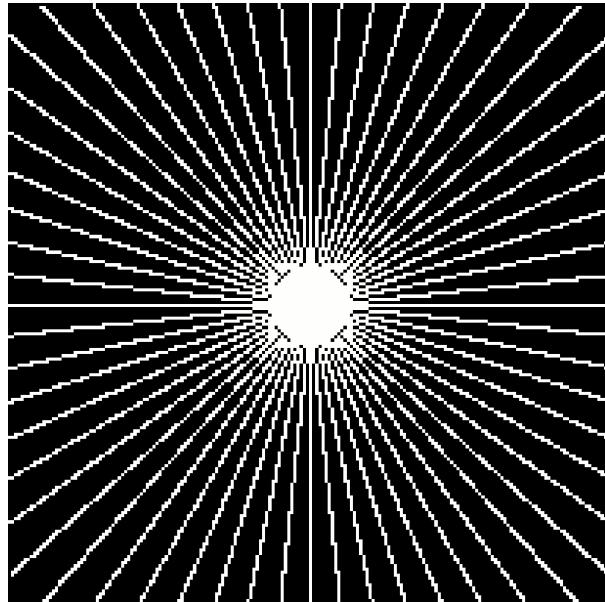


Figure 3.7. Radial undersampling trajectory with 19% of data points.

Following the same procedure as the other experiments, we also perform a zero-filled reconstruction with our undersampled k-space measurements vector b in order to generate a baseline to compare with our latter results.

3.4.2 Deterministic Prior Information

The prior information step consists of creating f 1-D arrays with the Φ positions of non-zero elements from the filtered images generated with the previous frames' reconstructions. These Φ positions are then multiplied by a τ factor of 10 at each iteration

in our l_1 -minimization.

For both deterministic and context-dependent probabilistic prior information we have used the largest proportion of prior information available: 49.92%, which provided the best results in both prior information strategies for this specific scenario.

3.4.3 CoDePPI: Context-Dependent Probabilistic Prior Information

We propose a novel approach to the non-deterministic prior information extraction by using knowledge about regions in the image that are more dislocated than others to apply different variance in the normal distributions procedure. When moving from a frame to another, portions of the image are dislocated in different degrees, that is, some regions will be more likely to contain support positions than others. We call this approach CoDePPI: Context-Dependent Probabilistic Prior Information.

By taking this motion information into account, we segmented the image into four levels of movement – *high*, *medium*, *low* and *close-to-zero* movement. The regions labelled as *high* movement receive a covariance matrix with a higher variance; thus assigning this region as a low-confidence support position, whilst the regions labelled as *low* or *close-to-zero* receive a Gaussian distribution with low variance, representing it as a highly likely support position prior information. Similarly, the *medium* labeled positions are assigned an intermediate variance for an intermediate confidence.

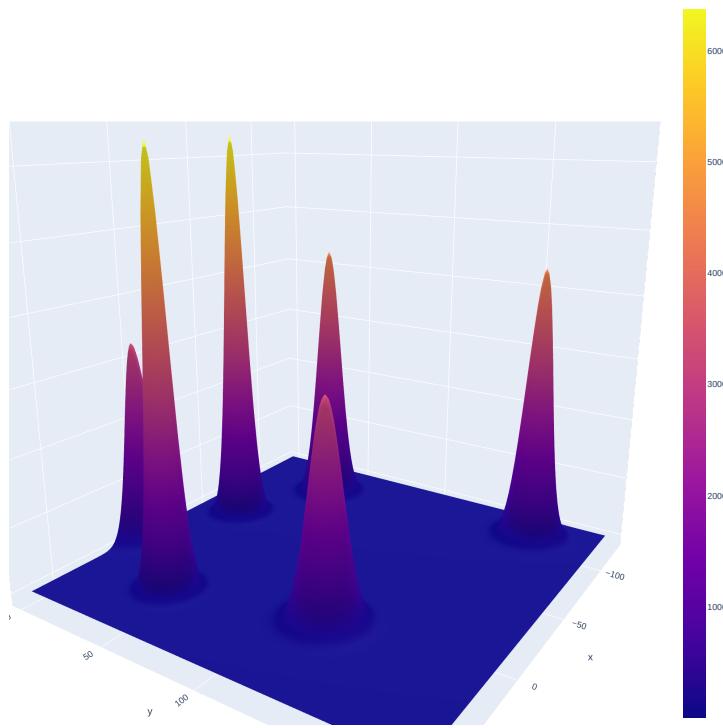


Figure 3.8. Example of gaussian distributions with four different levels of variances.

In this section, we focus on the dynamic sagittal cardiac MR and produce a manual segmentation for this case study specifically, but this could be generalized to other types of dynamic scans. The segmentation step focuses on the different motion contexts related to the exam and subject. In this case, we have a sagittal cardiac MR exam in which the subject's regions with most motion visually detected were the heart region and digestive system, hence these portions of the image are marked as high and medium in terms of motion respectively.

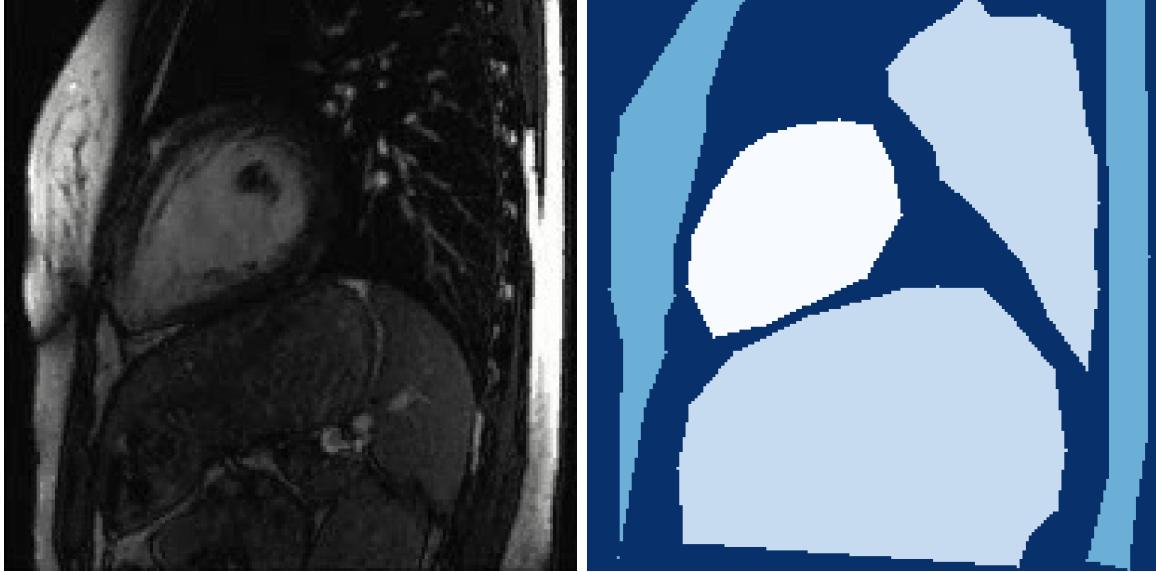


Figure 3.9. Fully-sampled cardiac MRI frame.

Figure 3.10. Manual segmentation: the lighter the colour, the more pixels are dislocated in the region throughout frames.

Identifying these different regions gives us the possibility to apply different weights for each region in the image. The regions labelled as *high* movement receive a covariance matrix with a higher variance; thus assigning this region as a low-confidence support position, whilst the regions labelled as *low* or *close-to-zero* receive a Gaussian distribution with low variance, representing it as a support position prior information. A *weights* matrix is generated by computing an element-wise Gaussian distribution with the according covariance matrix on the segmented mask and then adding these distributions each iteration, forming accumulations on the centre of several support positions. The regions labelled *high* are specifically multiplied by -1 so that the confidence is represented with negative magnitudes. After computing the *weights* matrix, we expect to see distinguishably different magnitude levels, representing different confidences.

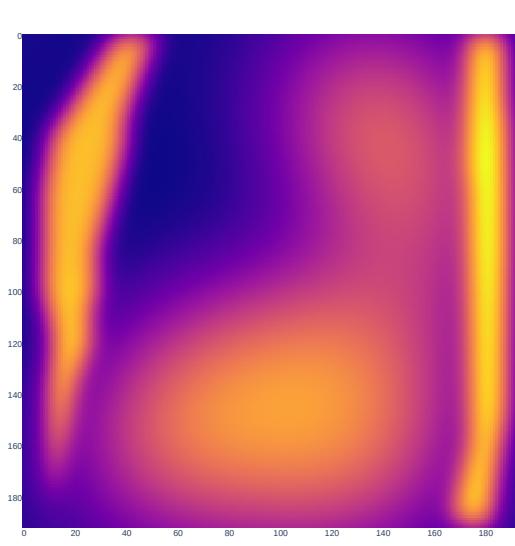


Figure 3.11. 2D *weights* matrix.

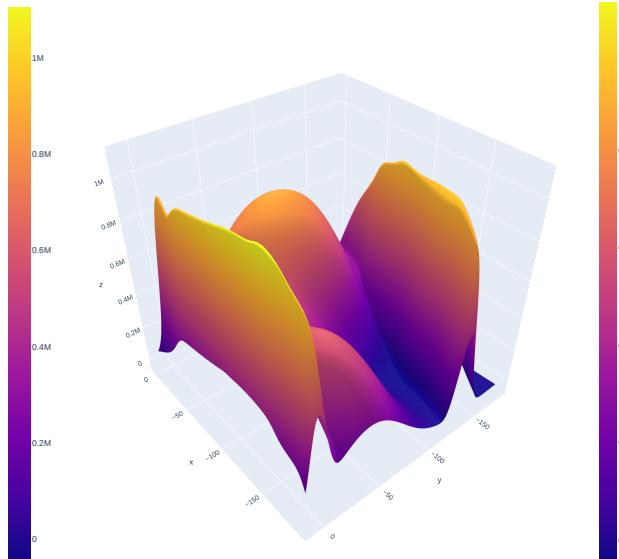


Figure 3.12. 3D *weights* matrix visualization.

Inverting the normal distributions in regions annotated as *high* is key to our implementation due to a *bias* element-wise addition applied after the normal distributions step. The latter step is used to normalize the magnitudes in our *weight* matrix between some constraints. As observed in the previous figures, the generated *weights* matrix contains elements ranging from values below 0 to values over 1×10^6 , we then apply a customized normalization function with a τ factor and a *bias* constant to re-weight the values in this matrix.

$$\omega(x) = (x \oslash \max x_i) \circ \tau + \beta \quad (3.1)$$

Where ω is our re-weighted *weights* matrix, x is the *weights* matrix, τ is the factor by which the values are multiplied and β is the *bias* term that increments all elements in the re-weighted matrix. The bias term defines $\min x_i$ for every x_i in the regions labelled as *medium*, *low* and *close-to-zero* movement.

At this point, the *weights* matrix is computed and re-weighted, so now this matrix is multiplied element-wise with each of the Φ masks created in the same fashion as the deterministic prior information, generating a *weighted phi* mask κ .

$$\kappa = \omega \circ \Phi \quad (3.2)$$

After this operation, our Φ positions no longer holds deterministic values, but rather

dispose of different levels of importance due to the *weights*' values.

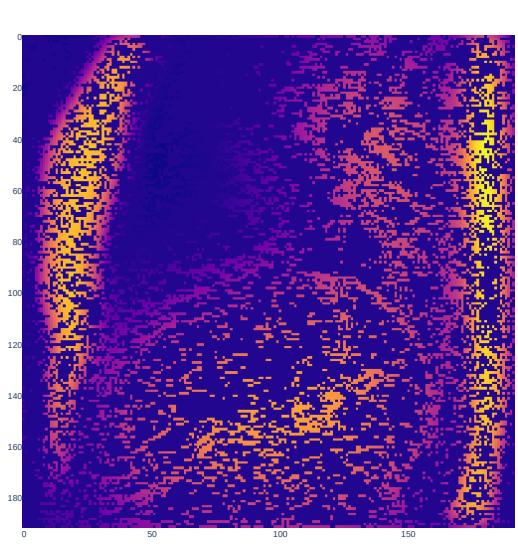


Figure 3.13. 2D *weights* matrix visualization.

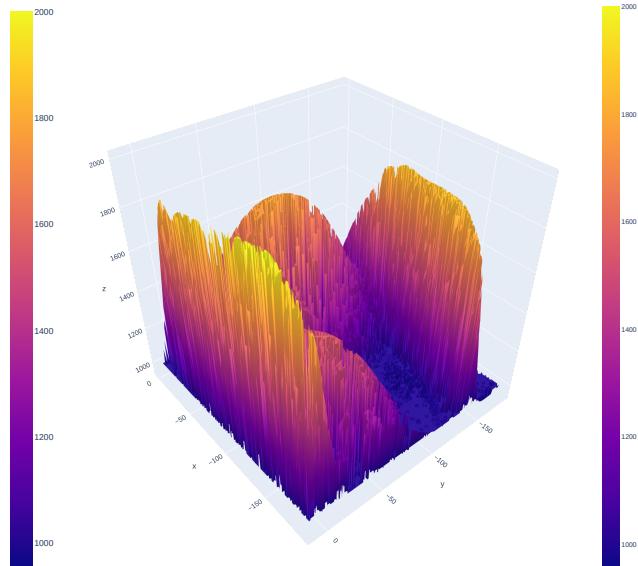


Figure 3.14. 3D *weights* matrix visualization.

3.4.4 L1-minimization Adaptation

After calculating κ , a minor modification is necessary for the $L1$ -minimization algorithm. In essence, we perform an element-wise multiplication between the input and κ every time before an iteration so that the prior information is applied in the computation.

$$x_i = x_i \circ \kappa_k \quad (3.3)$$

Where $x, i, k\dots$

3.4.5 Parameter Fine-tuning

After some fine-tuning experiments in the covariance matrices, biases and τ , we have found that these are the optimal configuration for a cardiac MRI:

3.5 Future Contribution Literature Review and Experiment

3.5.1 Preliminary Tests with Generative Adversarial Networks

In order to test the usage of GANs for data generation and in the future use it along with *prior information* for CS systems, we have developed a GAN capable of generating handwritten digits from 0 to 9 using the notable MNIST dataset. The MNIST dataset contains 60,000 examples for training and 10,000 examples for testing. The digits have been size-normalized and centred in a fixed-size image (28×28 pixels) with values from 0 to 9. For simplicity, each image has been flattened and converted into a 1-dimensional numpy array of 784 features (28×28).

The idea is to test if the neural network can output liable digits that look both readable (to the extent in which the MNIST dataset is) and also like it has been made by a human, just like the dataset itself.

Each MNIST image contains a 28×28 black and white image, like the following:



Figure 3.15. Sample of digits from MNIST

A Deep Convolutional GAN (DCGAN) was used for the experiment. A DCGAN is an extension of the GAN, except that it explicitly uses convolutional and convolutional-transpose layers in the discriminator and generator networks, respectively [4].

3.5.2 Data Transformation

Each input image used by the *dataloader* went through a computer-vision pre-processing step that includes:

- Grayscale transform: convert the image to greyscale. When loaded, the MNIST digits are in RGB format with three channels. Greyscale reduces these three to one.
- ToTensor: convert the image to a PyTorch Tensor, with dimensions (channels, height, width). This also rescales the pixel values, from integers between 0 and 255 to floats between 0.0 and 1.0.
- Normalize: scale and translate the pixel values from the range 0.0, 1.0 to -1.0, 1.0. The first argument is μ and the second argument is σ , and the function applied to each pixel is:

$$\rho \leftarrow \frac{(\rho - \mu)}{\sigma} \quad (3.4)$$

3.5.3 Generator Network Architecture

The generator network architecture is implemented using PyTorch as:

- A linear *fully-connected* module (or layer) to map the latent space to a $7 \times 7 \times 256 = 12544$ -dimensional space that will later be undersampled several times until we reach $1 \times 28 \times 28$.
- An optional 1-dimensional batch normalization module
- A leaky ReLU module.
- A 2-dimensional convolutional layer with $padding = 2$, $stride = 1$ and 5×5 kernel (or filter).
- Two 2-dimensional transposed convolutional layers with $padding = 1$, $stride = 2$ and 4×4 kernel.
- Two optional 2-dimensional batch normalization modules after each 2-dimensional transposed convolutional layer.
- A *Tanh* activation function, rescaling the images to a $[-1, 1]$ range.

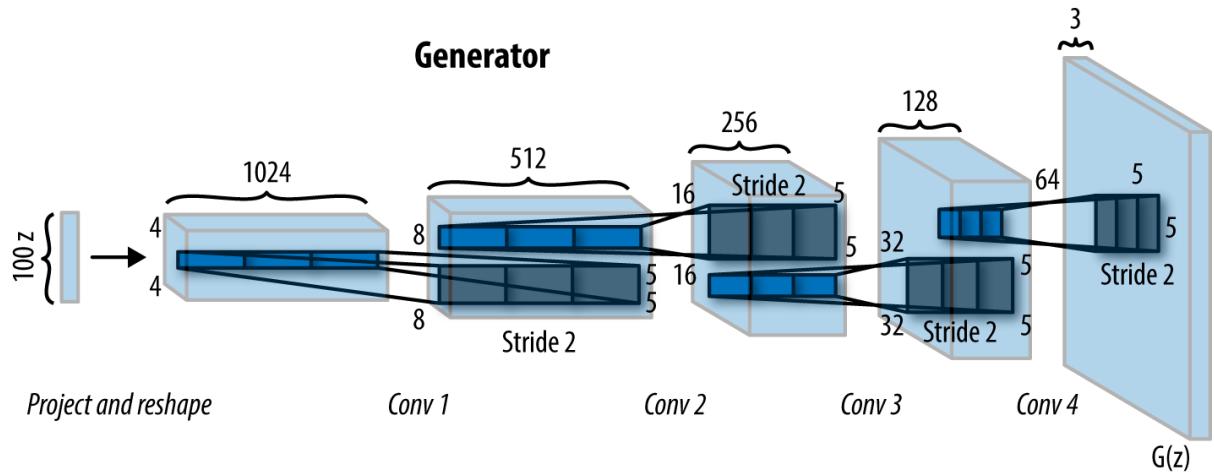


Figure 3.16. Generator network architecture of a 3-D image. Source: [4, 5]

The latent space (random signal) input goes through each layer being upscaled until it reaches the target image dimension 28×28 and then fed into the discriminator network.

3.5.4 Discriminator Network Architecture

The discriminator is a CNN-based image binary classifier network that takes an image as input and outputs a scalar probability that the given image is real or generated. The architecture is quite similar to the Generator network, except backwards. Here, the discriminator takes a $1 \times 28 \times 28$ input image, processes it through a series of convolutions, batch normalizations, and LeakyReLU layers, and outputs the final probability through a Sigmoid activation function.

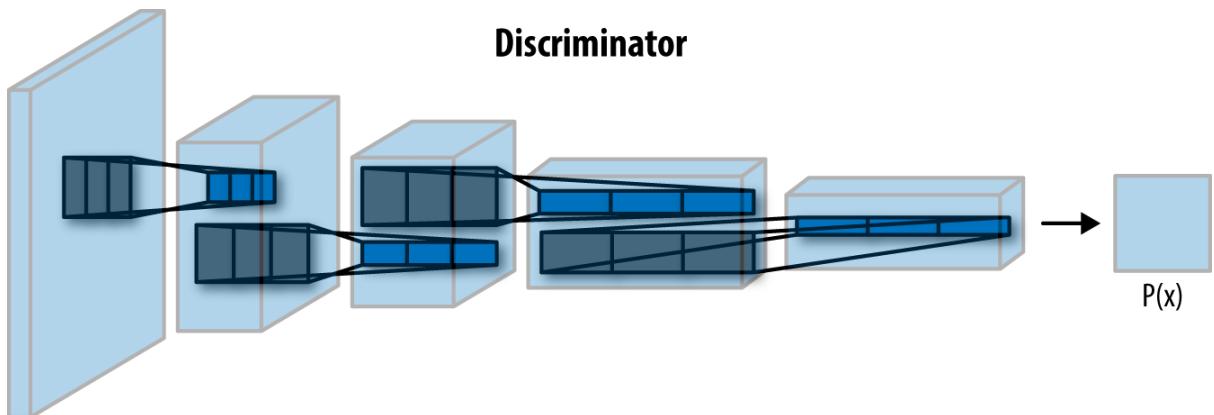


Figure 3.17. Discriminator network architecture of a 3-D image. Source: [4, 5]

4 CS RECONSTRUCTION AND PRELIMINARY DEEP LEARNING RESULTS

In this section we provide experiments in signal reconstruction using the CS theory and contrast reconstruction quality of our proposed method – context-depended probabilistic prior information – with other CS modalities. Additionally, we also present the deep learning foundation for the future work we intend to proceed along with a preliminary generation experiment using GANs.

4.1 1-D Compressed Sensing Reconstruction

The reconstruction quality over computational resources demanded trade-off starts making a difference with L size around 140, which is over 50% of the fully sampled signal, a proportion much higher than the ones normally used in MRI reconstruction. This demonstrates that there is no big prejudice in using indirect reconstruction method for very undersampled signals (close to 10% of the data) such as the ones used in MRI.

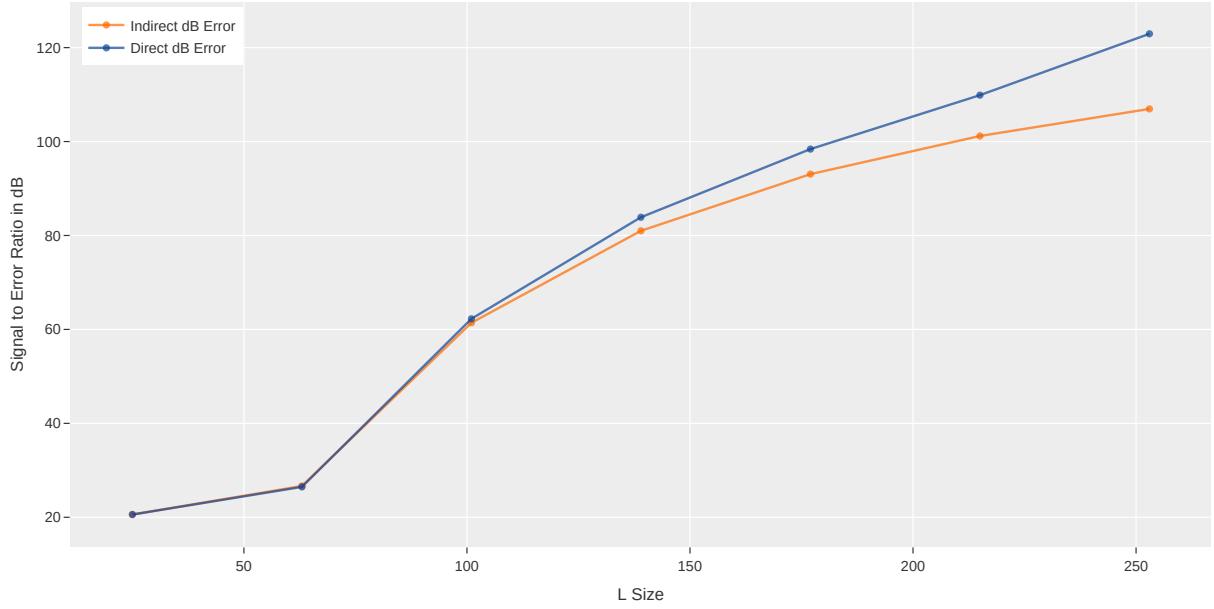


Figure 4.1. SNR x L size for direct and indirect L1-minimization for 200 random 1D signals

4.2 Compressed Sensing Reconstruction with Pre-Filtered Signal

The results in the experiment show a huge gain of resolution in the pre-filtering method reconstructed image compared to using plain l_1 -minimization.

The zero-filled reconstruction (dummy baseline) was unable to reconstruct a high fidelity image and performed very poorly in the PSNR and SNR metrics. The L1-minimization compressed sensing approach reconstructed the image with some noticeable noise artefacts, yet much better than the zero-filling approach. Finally, the L1-minimization along the usage of sparsifying pre-filtering delivered a great looking image without eye-catching artefacts and also increased the metrics largely.

4.2.1 Phantom Reconstruction Evaluation

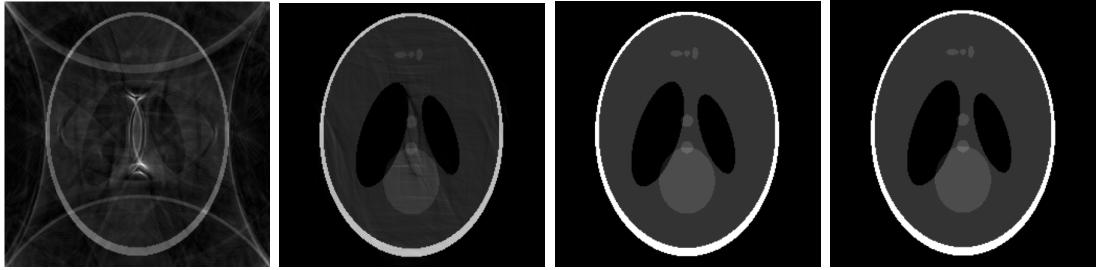


Figure 4.2.
Zero-filled

Figure 4.3.
L1-minimization

Figure 4.4.
L1-minimization
with
pre-
filtering

Figure 4.5.
Shepp-
Logan
phantom
reference
image

It is clear that reconstructing the image using the l_1 -minimization certainly improves the MRI reconstruction and the metrics reinforce that the pre-filtering step to pre-process the image is crucial to achieve higher image quality.

	PSNR	SSIM	SNR	NMSE	MSE
Zero-fill	16.39	0.26	4.22	3.77e-1	2.29e-2
L1-minimization	30.46	0.82	18.28	1.38e-3	8.99e-4
Pre-filtering L1-minimization	76.90	0.99	64.73	3.36e-7	2.04e-8

Table 4.1. Phantom reconstruction metrics.

PSNR and SNR are both represented in the dB scale, thus the results had to have improved by great orders of magnitude in each added step to achieve such results.

4.2.2 Sagittal Head MRI Reconstruction

Another experiment done was using a reference sagittal head image of shape (256×256) . The same spiral undersampling pattern with 30.95% data points used in the phantom experiment was used here for artificial undersampling.

This image poses a harder reconstruction challenge as it is filled with more details and more complex structures than the phantom. That said, it is clear that the undersampling pattern and amount of data points has not been sufficient to reconstruct a high fidelity image in any scenario, but the L1-minimization with pre-filtering reconstruction looks like the winner again, reinforcing the idea that pre-filtering is a good pre-processing strategy.

4.2.3 Sagittal Head Reconstruction Evaluation

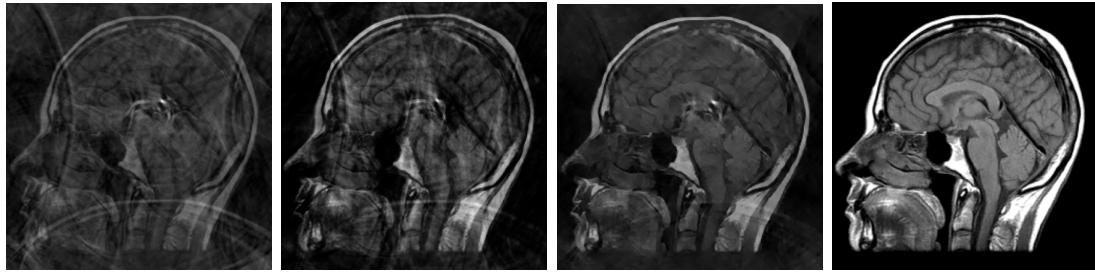


Figure 4.6.
Zero-filled

Figure 4.7.
L1-minimization

Figure 4.8.
L1-minimization with
pre-filtering

Figure 4.9.
Reference image

The metrics evaluated were not affected as much as they were for the phantom experiment, but they were mostly improved by pre-filtering usage.

	PSNR	SSIM	SNR	NMSE	MSE
Zero-fill	14.34	0.35	5.52	0.28	2392.57
L1-minimization	17.54	0.48	8.73	0.13	1144.12
Pre-filtering L1-minimization	17.79	0.49	8.97	0.12	1081.42

Table 4.2. Sagittal reconstruction metrics.

4.3 Dynamic Sagittal Cardiac Case Study

In order to test if the extracted prior information from one frame is relevant and how it impacts the reconstruction of the next frame, we have tested reconstructing frames from a dynamic cardiac MRI exam. The extracted frames are $(n \times n)$ in shape

4.3.1 Deterministic Prior Information

The three prior information Φ obtained from the extraction of non-zero elements in the filtered images transformed to 2-D masks for visualization. For this dynamic cardiac images, we used the maximum amount of prior information available: 49.92%.



Figure 4.10. Deterministic prior information with 49.92% data points proportion.

These images look very noisy due to the proportion of prior information extracted and that these values are all represented by the same magnitude. It becomes clearer when the masks above are visually observed that by using deterministic prior information, where all positions have the same likelihood to belong to a support structure, some error will be induced. Also, we can note some artifacts generated in the above images that do not necessarily are part of support positions.

4.3.2 Context-Dependent Probabilistic Prior Information



Figure 4.11. Context-dependent probabilistic prior information with 49.92% data points proportion.

Although we are still using a great amount of prior information proportion and with many noisy points, the regions that should have a high confidence of belonging to support structures in our subject are much better defined when the Φ values are multiplied by our weights matrix x 3.1. This definition observed in the visualizations above is also reflected in the l_1 -minimization as it can be noted in the results 4.3.

4.3.3 Dynamic Sagittal Cardiac Reconstruction Evaluation

All experiments conducted for the dynamic sagittal cardiac were done in order to compare the different applications of prior information, thus all of the reconstructions were done with the usage of pre-filtering within the l_1 -minimization context.



Figure 4.12. On the left: zero-filled reconstruction. On the right: no prior information reconstruction

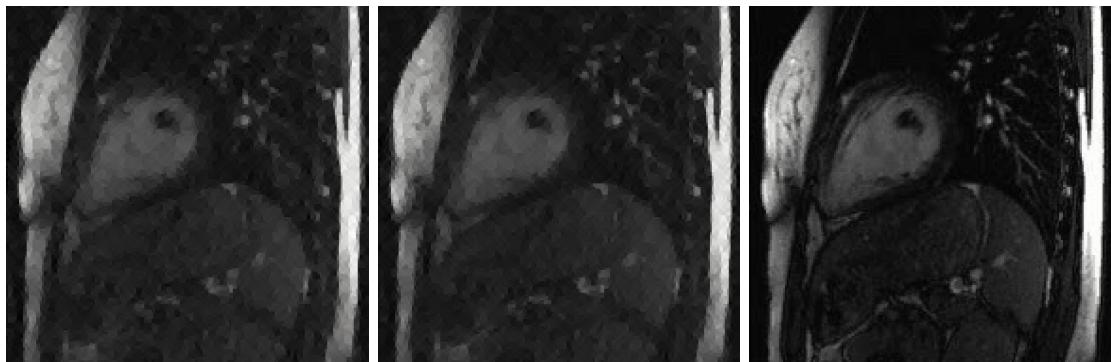


Figure 4.13. From left to right: deterministic prior information; context-dependent probabilistic prior information; reference image

The images presented above are displayed from left to right in terms of best SNR. These images represent the different strategies we took for MRI reconstructing and are hard to distinguish and perceive actual gain in terms of image quality, thus we should also analyse the metrics from our dynamic MR reconstruction.

	PSNR	SSIM	SNR	NMSE	MSE
Zero-fill	25.41	0.68	15.42	2.86e-2	186.70
No Prior Information	27.75	0.74	17.76	1.67e-2	109.08
Deterministic	27.75	0.74	17.76	1.67e-2	109.04
CoDePPI (ours)	28.40	0.76	18.41	1.4e-2	93.88

Table 4.3. Dynamic sagittal cardiac reconstruction metrics in terms of prior information usage.

4.4 Future Contributions: Deep Learning Applications

Our implementation counts on the usage of signal processing techniques without any machine learning usage, but that is only because this work is intended to be continued in a graduate degree where we would like to investigate more on two aspects: the generation of prior information support positions using GANs and a computer-vision system capable of segmenting the image based on levels of motion throughout the frames in a dynamic exam.

In this work, we have initiated the documentation of the DL necessary for our next steps as well as have we also document an experiment for the generation of handwritten MNIST numbers using a GAN in order to mitigate viability of our hypothesis to use generative algorithms in the near future.

4.5 Artificial Neural Networks

4.5.1 Biological Inspirations

ANNs, as the name suggests, have been (loosely) inspired by biological neural networks (brains) from animals. The concept of using many layers of vector-valued representation is drawn from neuroscience. The choice of the functions $f^{(i)}(x)$ used to compute these representations is also loosely guided by neuroscientific observations about the functions that biological neurons compute [31]. Another trait they share is that just like the human brain can be trained to pass forward only meaningful signals to achieve larger goals of the brain, the neurons on a neural network can be trained to pass along only useful signal [7].

4.5.2 Neuron

The most basic unit in ANNs is the *artificial neuron*. Neurons act as feature detectors and this is one of the advantages of deep learning techniques in contrast to classical machine learning as the ANN is responsible for doing feature engineering and selection, and often outperform humans in this task.

These artificial neurons that are modelled mirroring the behaviour of the biological neuron as both of them are stimulated by inputs and carry some information they receive to other neurons. Artificial neurons take in inputs x_1, x_2, \dots, x_n , each and multiply them by their respective weights w_1, w_2, \dots, w_n . Then these weighted inputs are summed together producing the *logit* of the artificial neuron, $z = \sum_{i=0}^n w_i x_i + b$, with b being a constant number added called *bias*. After this, the logit is passed to a function f in order to generate the value $y = f(z)$.

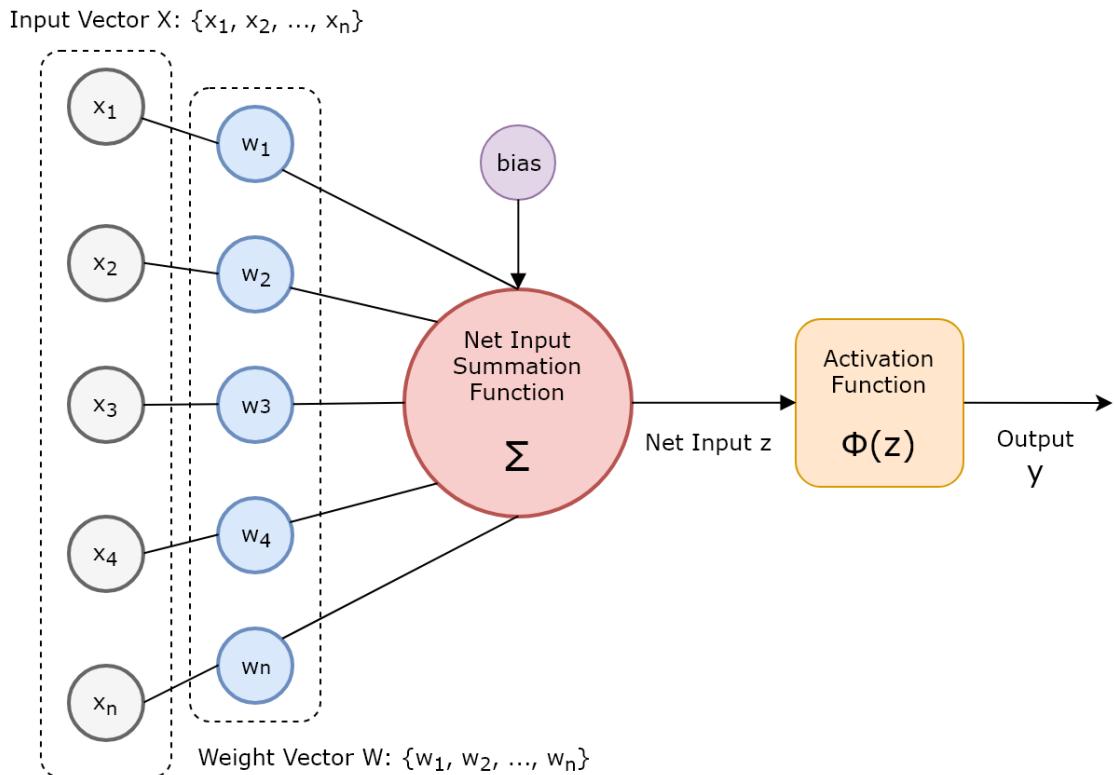


Figure 4.14. Schematic of an Artificial Neuron. Source: [6]

4.5.3 Multilayer Perceptron

Deep Feedforward Networks (DFN) or Multilayer Perceptron (MLP)s are a type of ANN very commonly used. It is the foundation of many famous architectures like CNNs.

DFNs have an input layer followed by one or many hidden layers and a single output layer. Each layer is fully connected to the adjacent layer.

MLPs are computational models that flow information through the function f that evaluates x . The goal is to approximate some function f^* . For instance, a classifier $y = f^*(x)$ maps an input x to a category y . The feedforward defines a mapping $y = f(x; \theta)$ and learns the value of the parameters θ that result in the best function approximation [31].

The behaviour of an ANN is shaped by its architecture, which describes the number of units it should have and how these units connect to each other and how complex the model is. Often adding too much complexity to the network will lead to overfitting the training set, which occurs when the model shapes the training data too precisely and cannot generalise new data fed.

Most ANNs are organized into rows of neurons called layers. These layers are arranged in a chain-like structure, with each layer being a function of the layer before it. These layers' goal is to extract *representations* out of the data fed and generalize what is meaningful towards minimizing the error rate. This architecture scheme is represented by the following equation, where i is the layer index:

$$h^{(i)} = g^{(i)}(W^{(i)T}x + b^{(i)})$$

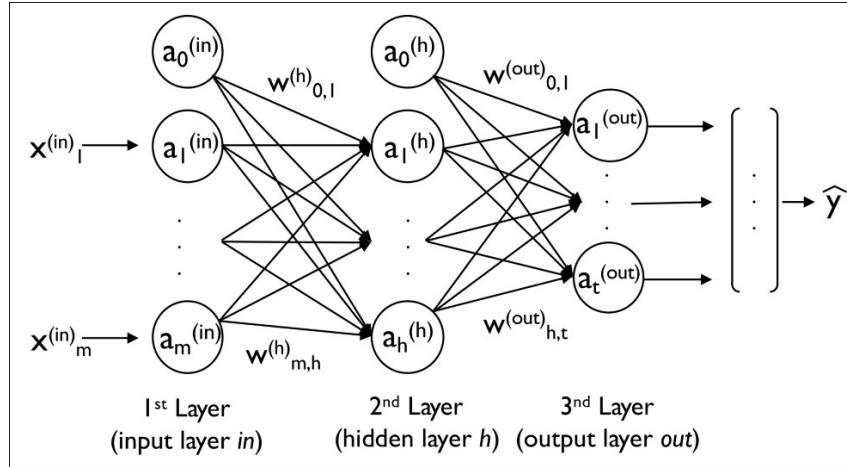


Figure 4.15. ANN Architecture Sample.

4.5.4 Activation Functions

Activation functions are a scalar-to-scalar function used to propagate the output of one layer's neurons forward to the next layer. There are several types of activation functions for different purposes and network architectures.

The DCGAN architecture used in my MNIST experiment is built using ReLU activations in the generator network and a *tanh* in the output layer. The discriminator network uses LeakyReLU activations for all layers and a *sigmoid* function for the output layer. LeakyReLU activation functions [32, 33] have been proven to work well for higher resolution modelling [4] in contrast to the usage of maxout activation functions that were first proposed in the original GAN paper [34].

Name	Function	Derivative
Sigmoid	$\phi(x) = \frac{1}{1 + e^{-x}}$	$\phi'(x) = \phi(x)(1 - \phi(x))$
TanH	$\phi(x) = \frac{2}{1 + e^{-2x}} - 1$	$\phi'(x) = 1 - \phi(x)^2$
ReLU	$\phi(x) = \begin{cases} 0 & x \leq 0 \\ x & x > 0 \end{cases}$	$\phi'(x) = \begin{cases} 0 & x \leq 0 \\ 1 & x > 0 \end{cases}$
Leaky ReLU	$\phi(x) = \begin{cases} \alpha x & x \leq 0 \\ x & x > 0 \end{cases}$	$\phi'(x) = \begin{cases} \alpha & x \leq 0 \\ 1 & x > 0 \end{cases}$

Table 4.4. Activation functions and respective derivatives.

Some of the most used and also required activation functions in the use of GANs and other widely used neural networks are described below.

4.5.4.1 ReLU

The ReLU [35] transform activates a node only if the input is above a certain threshold having a linear relationship with the dependent variable and outputs zero for every input below zero.

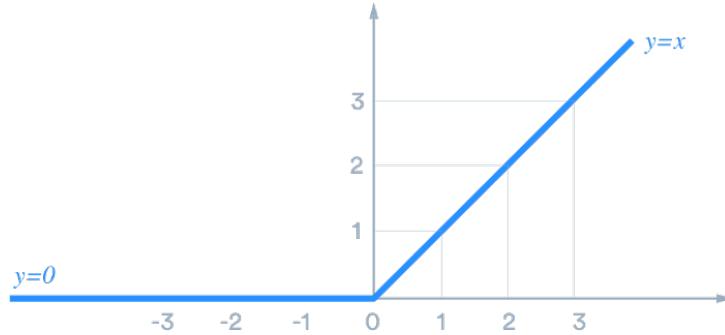


Figure 4.16. ReLU activation function

4.5.4.2 Leaky ReLU

ReLU activation functions have the “dying ReLU” problem, where a ReLU neuron is stuck in the negative side and always outputs 0 [36, 37]. This happens when the slope of ReLU in the negative range is also 0, once a neuron gets negative, it is unlikely for it to recover. These “dead” neurons are not playing any role in discriminating the input and are essentially useless.

To mitigate this issue within ReLUs, LeakyReLUs are a strategy that opposed to having the function being zero when $x < 0$, it has instead a small negative slope (most times with $\alpha = 0.01$).

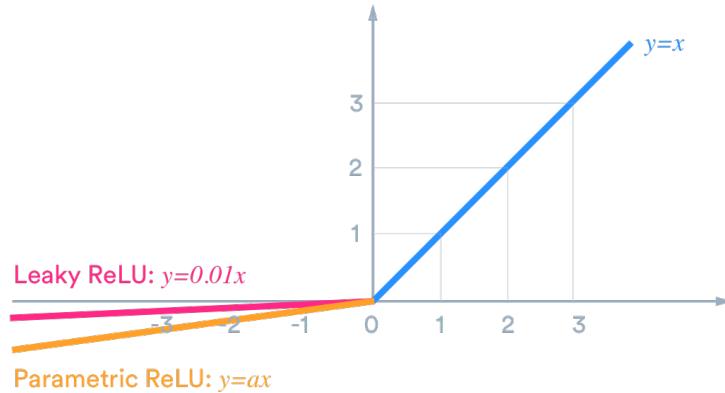


Figure 4.17. Leaky ReLU activation function

4.5.4.3 Tanh

Tanh is a hyperbolic trigonometric function that deals more easily with negative numbers [7]. Unlike the Sigmoid function, tanh ranges from -1 to 1.

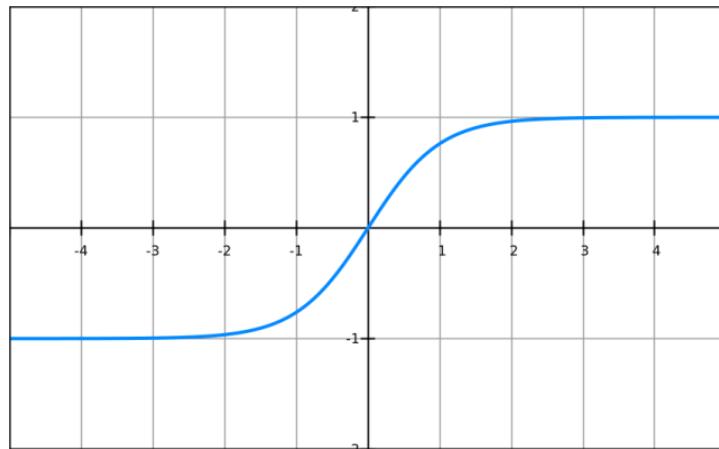


Figure 4.18. *Tanh* activation function. Source: [7]

4.5.4.4 Sigmoid

Sigmoids can reduce extreme values or outliers in data without removing them, framing the input from 0 to 1 and most outputs will be close to either 0 or 1.

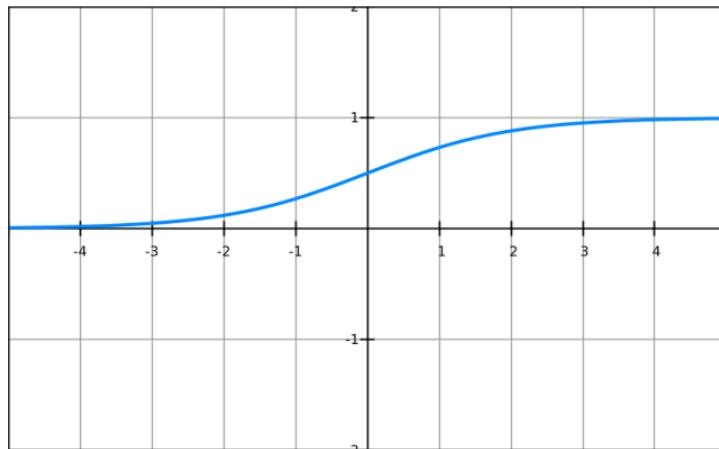


Figure 4.19. Sigmoid activation function. Source: [7]

4.5.5 Loss Functions

Loss functions are used to determine how a neural network is performing on the given data. A metric is calculated based on the error observed in the network's predictions and the model then tries to minimize this error in an optimization problem fashion.

Some of the most commonly used functions are described in the table below.

Mean squared error	MSE	$= \frac{1}{n} \sum_{t=1}^n e_t^2$
Root mean squared error	RMSE	$= \sqrt{\frac{1}{n} \sum_{t=1}^n e_t^2}$
Mean absolute error	MAE	$= \frac{1}{n} \sum_{t=1}^n e_t $
Mean absolute percentage error	MAPE	$= \frac{100\%}{n} \sum_{t=1}^n \left \frac{e_t}{y_t} \right $

Table 4.5. Loss functions and formulas.

In the case of generative networks, the original GAN paper presents a loss function called *minmax* [34], that is described as.

$$\min_G \max_D V(D, G) = E_{x \sim P_{data}(x)}[\log D(x)] + E_{z \sim p_z(z)}[\log(1 - D(G(z)))] \quad (4.1)$$

Where x is the input data representing an image, $D(x)$ is the discriminator network, $G(x)$ is the generator function and z represents the latent vector that is mapped to data-space by G . Hence, the scalar probability that the output of the generator G is a real image is given by $D(G(z))$ [34].

4.5.6 Backpropagation

Backpropagation is a technique used to implement gradient descent in weight space for an MLP [38, 39]. In essence, backpropagation computes the error partial derivatives of an approximating function $F(w, x)$ computed by the ANN with respect to the weight and input vector for each training example [40].

The development of the backpropagation algorithm is a milestone in neural networks development and research as it made computationally efficient to train MLPs, thus confirming that ANNs research field was filled with potential in the mid-1980s.

In order to evaluate the derivatives of the function $F(w, x)$ with respect to all the elements in the weight vector w for an input vector $x = [x_1, x_2, \dots, x_m]^T$ for an MLP with layer $l = 2$, we have the following equation where φ is the activation function, w is the ordered weight vector and x is the input vector fed into the MLP [40]:

$$F(\mathbf{w}, \mathbf{x}) = \sum_{j=0}^{m_1} w_{oj} \varphi \left(\sum_{i=0}^{m_0} w_{ji} x_i \right) \quad (4.2)$$

4.5.7 Gradient Descent

Gradient descent is an optimization algorithm frequently used in ANNs to find the values of coefficients of a function that minimizes a cost function. Gradient descent can be very time consuming on large datasets due to the necessity of having a prediction for each instance in the training set. In scenarios where there is a large number of data instances, a variation of gradient descent called Stochastic Gradient Descent (SGD) can be used. SGD updates the coefficients for each training instance or batch instead of at the end after running through all the training set instances.

Most deep learning models are powered by the SGD and it can be visualized as the figure below demonstrates: the function starts in a random point in the loss function and after each iteration, the SGD calculates how it should adjust parameters in order to reach the minimal point in the loss function, hence moving towards the valley as illustrated.

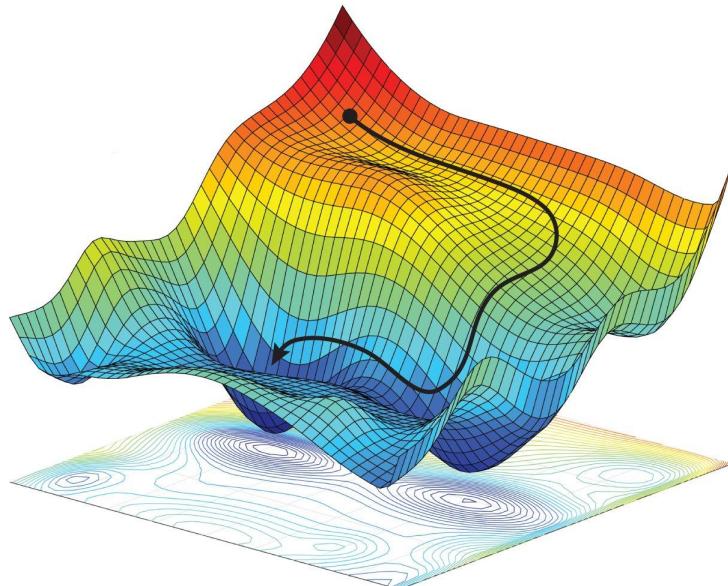


Figure 4.20. Gradient descent example visualization. Source: [8]

4.6 Generative Adversarial Networks

GANs are a machine learning strategy proposed in 2014 by Ian Goodfellow [34] that consists of two simultaneously trained models: the *Generator* $G(x)$ and the *Discriminator*

$D(x)$. The generator has the role to generate fake data whilst the discriminator is trained to discern whether the given input is real or fake.

In essence, the generator takes a vector of random numbers (z) as input and outputs a fake example that strives to look as close as possible to the training data pattern. The discriminator takes an image (x) as input from two sources: real examples from the training set and fake examples generated by the generator network, then the discriminator outputs a scalar probability that the image is real [9].

GANs play a minimax two-player game in which D tries to maximize the probability to correctly classify real and fake samples ($\log D(x)$), whilst G tries to minimize the chance that D will correctly predict its generated outputs are fake ($\log(1 - D(G(x)))$).

Ideally, this minimax game would resolve to a solution with $pg = p_{data}$, where the discriminator is incapable of distinguishing real from fake inputs. However, GANs are still new neural network techniques and its convergence theory is still being highly researched and hardly reaching this point in reality [34].

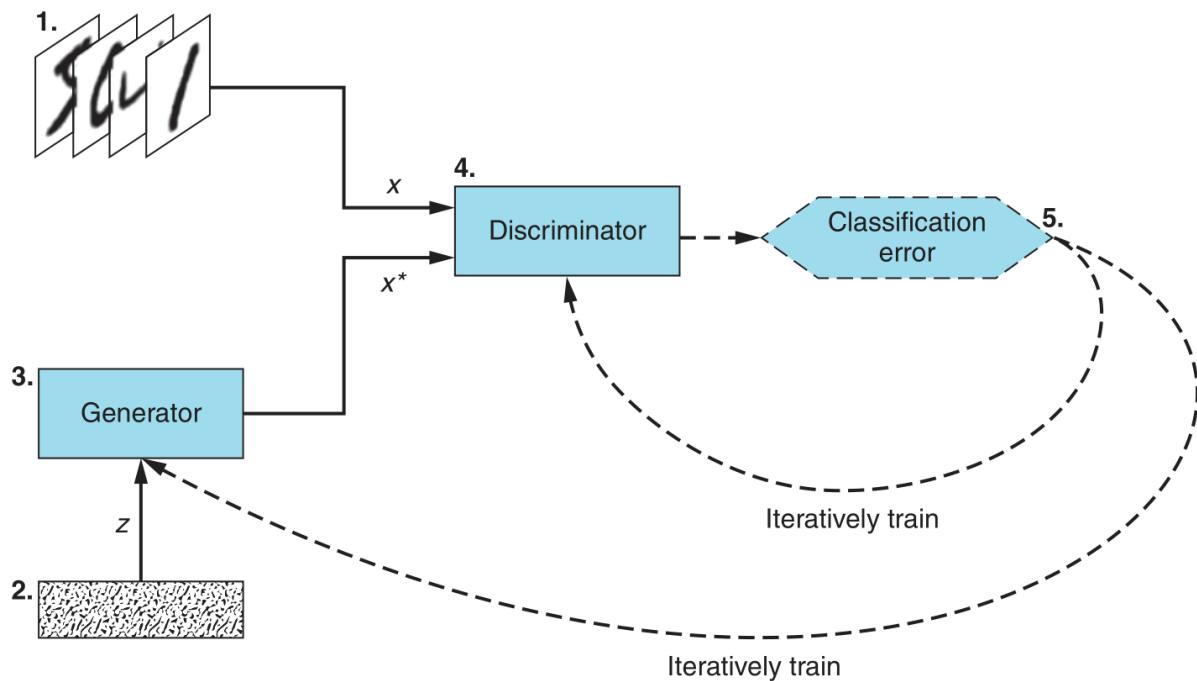


Figure 4.21. GAN training diagram. Source: [9]

4.6.1 Preliminary Tests with Generative Adversarial Networks

Both discriminator losses (fake and real) start very high and quickly decreases as the generator loss curve goes up in an invertedly correlated manner. This happens especially

because the generator starts by tricking the discriminator network very easily as it is naïve to determine if an image is real or generated. Quickly the discriminator starts to detect how the data is disposed and manages to interpret the generated images are different from the training examples it is seeing.

This phenomenon exposes how bad the generator is in the first epochs and how easily the discriminator can distinguish between created and real. Then as the epochs go by and both networks get more sophisticated, the generator starts to get better at creating the desired signal style and makes the discriminator's loss get higher again as it is observed in the loss curves below.

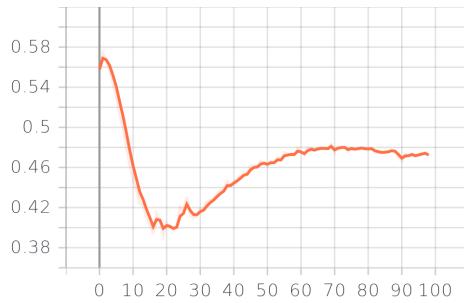


Figure 4.22.
Discriminator
loss over epochs

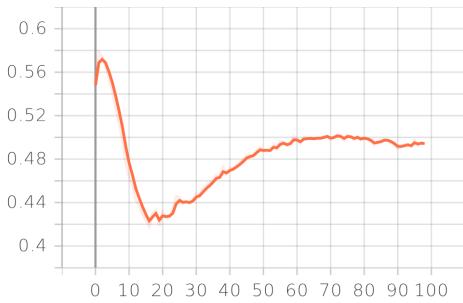


Figure 4.23.
Discriminator real loss
over epochs

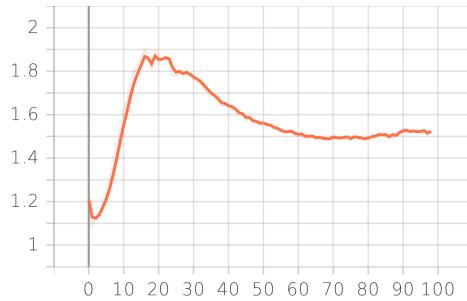


Figure 4.24. Generator
loss over epochs

After 100 epochs, the DCGAN for MNIST number generation had an exceptionally good performance when the generated images are displayed. It is hard to tell if these are generated images or if they are part of the training set. The generated images sometimes have a bit more blur to them, but certainly with more epochs and more training samples this could be minimized. The following results were not cherry-picked in any manner.

Generated Images								
6	5	6	1	1	7	7	1	
9	0	4	8	7	9	4	0	
9	7	8	1	2	5	0	9	
7	3	6	3	1	9	7	1	

Figure 4.25. GAN generated MNIST digits

5 CONCLUSION

We have successfully proposed an alternate approach for prior information extraction that uses motion context from segmented images and this method proves efficient and produce better quality for the experiments we have conducted. Further investigation would be more appropriate with more adequate statistics validations in order to effectively assert that CoDePPI is always a better solution rather than the classical prior information algorithm presente in [14]. Additionaly, many modules for CS reconstruction have been developed in Python featuring the usage of undersampling techniques, lp -minimization function, pre-filtering pre-processing, classical and CoDePPI implementation of the prior information and modules to convert and utilize fastMRI images. All this research and development led to results that led to further hypothesis and investigations that we have initiated in this very document regarding the usage of DL models for prior information generation and computer-vision approaches for automatic image segmentation for an automatized context detection for CoDePPI. These ideas and hypothesis are to be continued and mitigated in a graduate degree after the completion of this period.

Furthermore, we believe that the proposal of an alternate prior information algorithm can lead other scientists to new possibilities and contributions to the process of prior information retrieval, after all, our goal was to not only create better results with a novel algorithm, but also contribute to the field of MRI that is undoubtly very relevant to medicine and our society.

List of References

- [1] Jure Zbontar, Florian Knoll, Anuroop Sriram, Tullie Murrell, Zhengnan Huang, Matthew J. Muckley, Aaron Defazio, Ruben Stern, Patricia Johnson, Mary Bruno, Marc Parente, Krzysztof J. Geras, Joe Katsnelson, Hersh Chandarana, Zizhao Zhang, Michal Drozdal, Adriana Romero, Michael Rabbat, Pascal Vincent, Nafissa Yakubova, James Pinkerton, Duo Wang, Erich Owens, C. Lawrence Zitnick, Michael P. Recht, Daniel K. Sodickson, e Yvonne W. Lui. fastMRI: An Open Dataset and Benchmarks for Accelerated MRI. *arXiv:1811.08839 [physics, stat]*, December 2019. arXiv: 1811.08839.
- [2] Florian Knoll, Jure Zbontar, Anuroop Sriram, Matthew J. Muckley, Mary Bruno, Aaron Defazio, Marc Parente, Krzysztof J. Geras, Joe Katsnelson, Hersh Chandarana, Zizhao Zhang, Michal Drozdal, Adriana Romero, Michael Rabbat, Pascal Vincent, James Pinkerton, Duo Wang, Nafissa Yakubova, Erich Owens, C. Lawrence Zitnick, Michael P. Recht, Daniel K. Sodickson, e Yvonne W. Lui. fastMRI: A Publicly Available Raw k-Space and DICOM Dataset of Knee Images for Accelerated MR Image Reconstruction Using Machine Learning. *Radiology: Artificial Intelligence*, 2(1):e190007, January 2020.
- [3] Jong Chul Ye. Compressed sensing MRI: a review from signal processing perspective. *BMC Biomedical Engineering*, 1(1):8, December 2019.
- [4] Alec Radford, Luke Metz, e Soumith Chintala. Unsupervised Representation Learning with Deep Convolutional Generative Adversarial Networks. *arXiv:1511.06434 [cs]*, January 2016. arXiv: 1511.06434.
- [5] Dominic Monn. Deep convolutional generative adversarial networks with tensorflow, 2017. Available at: <https://www.oreilly.com/content/deep-convolutional-generative-adversarial-networks-with-tensorflow/>. Last access on November 18th, 2020.
- [6] J. Quddus. *Machine Learning with Apache Spark Quick Start Guide: Uncover patterns, derive actionable insights, and learn from big data using MLlib*. Packt Publishing, 2018.

- [7] Josh Patterson e Adam Gibson. *Deep Learning: A Practitioner's Approach*. O'Reilly, Beijing, 2017.
- [8] Daniela Rus. Alexander Amini. Gradient descent relies on trial and error to optimize an algorithm, aiming for minima in a 3d landscape. adapted by m. atarod/science, 2020. Available at: <https://www.sciencemag.org/news/2018/05/ai-researchers-allege-machine-learning-alchemy>. Last access on November 30th, 2020.
- [9] Jakub Langr e Vladimir Bok. *GANs in action: deep learning with generative adversarial networks*. Manning Publications, Shelter Island, New York, 2019. OCLC: on1050335878.
- [10] R Nick Bryan. *Introduction to the Science of Medical Imaging*. Cambridge University Press, Cambridge, 2009.
- [11] S. I. Kabanikhin. Definitions and examples of inverse and ill-posed problems. *Journal of Inverse and Ill-posed Problems*, 16(4), January 2008.
- [12] Cristiano Jacques Miosso. *Compressive Sensing with Prior Information Applied to Magnetic Resonance Imaging*. PhD Thesis, Department of Electrical and Computer Engineering, University of Texas at El Paso (UTEP).
- [13] C. J. Miosso, R. von Borries, e J. H. Pierluissi. Compressive sensing method for improved reconstruction of gradient-sparse magnetic resonance images. In *2009 Conference Record of the Forty-Third Asilomar Conference on Signals, Systems and Computers*, pages 799–806, Pacific Grove, CA, USA, 2009. IEEE.
- [14] C.J. Miosso, R. von Borries, M. Argaez, L. Velazquez, C. Quintero, e C.M. Potes. Compressive Sensing Reconstruction With Prior Information by Iteratively Reweighted Least-Squares. *IEEE Transactions on Signal Processing*, 57(6):2424–2431, June 2009.
- [15] Li Wan, Matthew Zeiler, Sixin Zhang, e Yann LeCun. Regularization of Neural Networks using DropConnect. page 12.
- [16] Jacob Devlin, Ming-Wei Chang, Kenton Lee, e Kristina Toutanova. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, pages 4171–4186, Minneapolis, Minnesota, June 2019. Association for Computational Linguistics.

- [17] Daeryong Kim e Bongwon Suh. Enhancing VAEs for collaborative filtering: flexible priors & gating mechanisms. In *Proceedings of the 13th ACM Conference on Recommender Systems*, pages 403–407, Copenhagen Denmark, September 2019. ACM.
- [18] yan yang, Jian Sun, Huibin Li, e Zongben Xu. Deep ADMM-Net for Compressive Sensing MRI. In D. D. Lee, M. Sugiyama, U. V. Luxburg, I. Guyon, e R. Garnett, editors, *Advances in Neural Information Processing Systems 29*, pages 10–18. Curran Associates, Inc., 2016.
- [19] Guang Yang, Simiao Yu, Hao Dong, Greg Slabaugh, Pier Luigi Dragotti, Xujiong Ye, Fangde Liu, Simon Arridge, Jennifer Keegan, Yike Guo, e David Firmin. DAGAN: Deep De-Aliasing Generative Adversarial Networks for Fast Compressed Sensing MRI Reconstruction. *IEEE Transactions on Medical Imaging*, 37(6):1310–1321, June 2018.
- [20] Morteza Mardani, Enhao Gong, Joseph Y. Cheng, Shreyas S. Vasanawala, Greg Zaharchuk, Lei Xing, e John M. Pauly. Deep Generative Adversarial Neural Networks for Compressive Sensing MRI. *IEEE Transactions on Medical Imaging*, 38(1):167–179, January 2019.
- [21] Dong Liang, Jing Cheng, Ziwen Ke, e Leslie Ying. Deep MRI Reconstruction: Unrolled Optimization Algorithms Meet Neural Networks. *arXiv:1907.11711 [physics, stat]*, July 2019.
- [22] Elizabeth K Cole, John M Pauly, Shreyas S Vasanawala, e Frank Ong. Unsupervised MRI Reconstruction with Generative Adversarial Networks. page 8, 2020.
- [23] D. Liang, J. Cheng, Z. Ke, e L. Ying. Deep magnetic resonance image reconstruction: Inverse problems meet neural networks. *IEEE Signal Processing Magazine*, 37(1):141–151, 2020.
- [24] Daniel Lucas Ferreira e Almeida. ReconstruÃ§Ã£o de imagens de ressonÃ¢ncia magnÃ©tica com base em compressive sensing usando informaÃ§Ã£o a priori estrutural em abordagem estocÃ¡stica. Msc thesis, Universidade de BrasÃlia.
- [25] P. C. Lauterbur. Image Formation by Induced Local Interactions: Examples Employing Nuclear Magnetic Resonance. *Nature*, 242(5394):190–191, March 1973.
- [26] Michael Lustig, David Donoho, e John M. Pauly. Sparse MRI: The application of compressed sensing for rapid MR imaging. *Magnetic Resonance in Medicine*, 58(6):1182–1195, December 2007.

- [27] D.L. Donoho. Compressed sensing. *IEEE Transactions on Information Theory*, 52(4):1289–1306, April 2006.
- [28] E.J. Candes, J. Romberg, e T. Tao. Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information. *IEEE Transactions on Information Theory*, 52(2):489–509, February 2006.
- [29] Emmanuel J. Candes e Terence Tao. Near-Optimal Signal Recovery From Random Projections: Universal Encoding Strategies? *IEEE Transactions on Information Theory*, 52(12):5406–5425, December 2006.
- [30] L. A. Shepp e B. F. Logan. The Fourier reconstruction of a head section. *IEEE Transactions on Nuclear Science*, 21(3):21–43, 1974.
- [31] Ian Goodfellow, Yoshua Bengio, e Aaron Courville. *Deep Learning*. MIT Press, 2016.
- [32] Andrew L Maas, Awni Y Hannun, e Andrew Y Ng. Rectifier Nonlinearities Improve Neural Network Acoustic Models. page 6.
- [33] Bing Xu, Naiyan Wang, Tianqi Chen, e Mu Li. Empirical evaluation of rectified activations in convolutional network. *CoRR*, abs/1505.00853, 2015.
- [34] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, e Yoshua Bengio. Generative Adversarial Nets. In Z. Ghahramani, M. Welling, C. Cortes, N. D. Lawrence, e K. Q. Weinberger, editors, *Advances in Neural Information Processing Systems 27*, pages 2672–2680. Curran Associates, Inc., 2014.
- [35] Richard H. R. Hahnloser, Rahul Sarpeshkar, Misha A. Mahowald, Rodney J. Douglas, e H. Sebastian Seung. Digital selection and analogue amplification coexist in a cortex-inspired silicon circuit. *Nature*, 405(6789):947–951, June 2000.
- [36] L. Trottier, P. Giguere, e B. Chaib-draa. Parametric exponential linear unit for deep convolutional neural networks. In *2017 16th IEEE International Conference on Machine Learning and Applications (ICMLA)*, pages 207–214, 2017.
- [37] Abien Fred Agarap. Deep learning using rectified linear units (relu). *CoRR*, abs/1803.08375, 2018.
- [38] David E. Rumelhart, Geoffrey E. Hinton, e Ronald J. Williams. Learning Representations by Back-propagating Errors. *Nature*, 323(6088):533–536, 1986.
- [39] Paul John Werbos. *The Roots of Backpropagation: From Ordered Derivatives to Neural Networks and Political Forecasting*. Wiley-Interscience, USA, 1994.

- [40] Simon S. Haykin e Simon S. Haykin. *Neural networks and learning machines*. Prentice Hall, New York, 3rd ed edition, 2009. OCLC: ocn237325326.