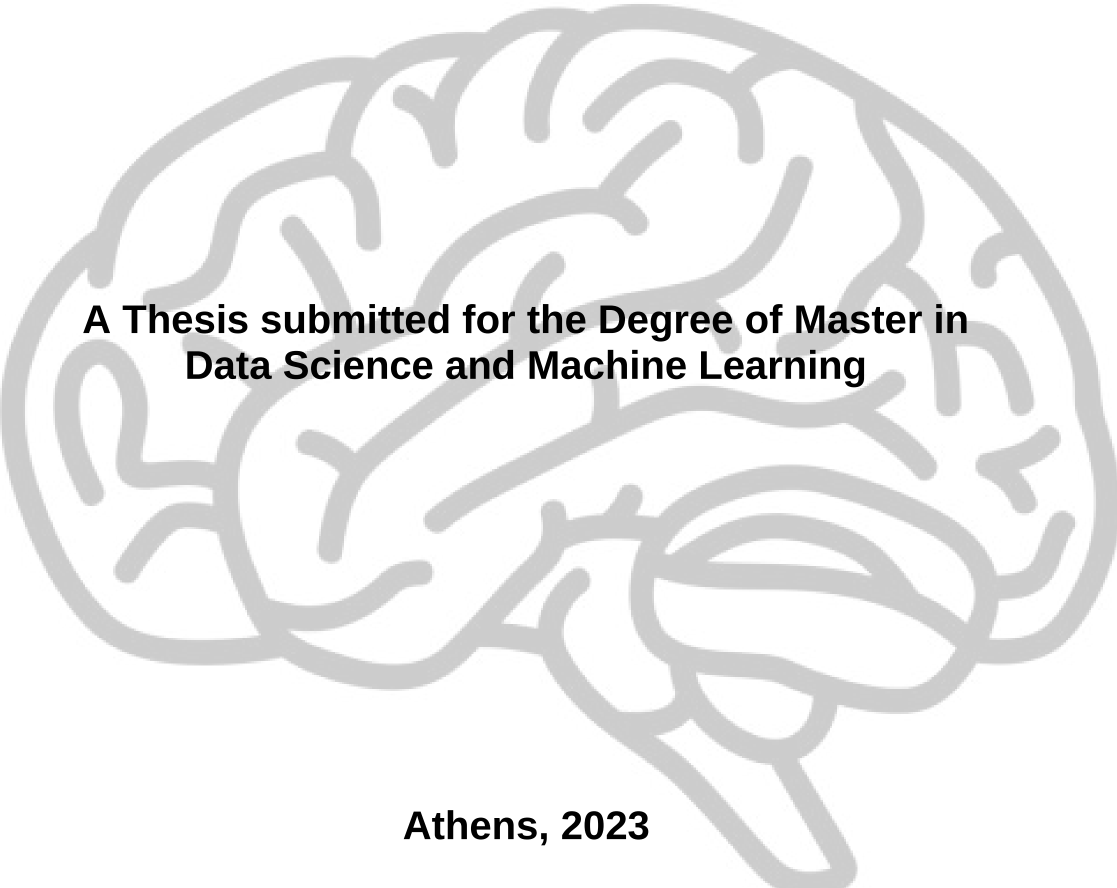
NATIONAL TECHNICAL UNIVERSITY OF ATHENS
SCHOOL OF ELECTRICAL AND ELECTRONIC ENGINEERING

Brain MRI Reconstruction with Cascade of CNNs and a Learnable Regularization

Dimitris Tryfonopoulos



A Thesis submitted for the Degree of Master in Data Science and Machine Learning

Athens, 2023

Master THESIS

ABSTRACT

In the field of Medical Imaging, the recent years special attention has been given to Magnetic Resonance Imaging (MRI). MR imaging is a non-invasive imaging modality capable of producing cross-sectional images with high spatial resolution. Unlike Computed tomography (CT) the nature of the acquisition does not employ ionising radiation. However, MRI acquisitions are characterized by long scan time (acquisition time) which is directly related to the number of samples acquired in the k-space. Being able to reduce the acquisition time will help increase patient satisfaction, reduce several artifacts (mainly motion), and finally reduce in overall the medical costs. The main initiative on reducing the scan time from a software perspective, is Compressed Sensing (CS) optimization [1]-[3] according to which given some constraints that are fulfilled we can reconstruct highly undersampled images. Following the success of deep learning (DL) in a wide range of applications, neural networks based on CS optimization have received significant interest for accelerating MR acquisitions and reconstruction strategies. In CS signal reconstruction, the iterative algorithm unrolled over a deep neural network. The basic strategy is to train the network to learn the weights (CNN kernels) for dealising undersampled MR images from a large dataset containing pairs of aliased and dealiased images. In this work, we are motivated from the 2D Deep Cascaded CNN-Network (DCNN) [4]. DCNN mainly consists of two blocks the CNN and the Data Consistency block, operating on the image and k-space (sampling domain) respectively, linked by a regularization term which adjusts the data fidelity based on the noise level of the acquired measurements. In this work, *we introduce an improved regularization for the data consistency term in two novel settings, a learnable regularization parameter per k-space slice and a spatially learnable regularization parameter. We show that the introduced regularization is independent of the CNN block architecture used and can be incorporated in any DL-CS based optimization network setting. We show that t*he employment of the proposed regularization in any DL-CS based network highly improves the reconstruction performance without adding any computational burden to the network.

ACKNOWLEDGEMENTS

I would like to express my appreciation to my supervisor Prof. Konstantinos Karantzalos for the guidance he provided throughout this thesis. I would also like to thank Mr. Makis Douskos for his valuable insights and time spent reading and providing feedback on my thesis. Finally, I am grateful to my colleague Mr. Marko Panic for his invaluable help and continuous support throughout the entire process.

All the used codes for the current thesis can be found in the following GitHub repository *

*Please ask permission for sharing the latter directory

Cont ents

ACRONYMS

MRI	Magnetic Resonance Imaging
CT	Computed Tomography
RF	Radiofrequency
CS	Compressed Sensing
PI	Parallel Imaging
CNN	Convolutional Neural Networks
DC	Data Consistency
DL	Deep Learning
NMR	Nuclear Magnetic Resonance
PET	Positron Emission Tomography
sMRI	structural MRI
dMRI	diffusion MRI
fMRI	functional MRI
EEG	Electroencephalogram
DNN	Deep Neural Network
RNN	Recurrent Neural Network
GAN	Generative Adversarial Network
AE	Auto Encoders
ReLU	Rectified Linear Unit
SNR	Signal to Noise Ratio
MSE	Mean Squared Error
SSIM	Structural Similarity Index
PSNR	Peak Signal to Noise Ratio
GND	Ground Truth Image
REC	Reconstructed Image
US	Undersampled Image
FF	Fixed λ regularization
TF	Per Slice Trainable λ Regularization
TT	Spatially Trainable λ Regularization

LIST OF FIGURES &TABLES

Chapter 1

Introduction

1.1 Motivation

In the field of Medical Imaging in recent years special attention has been given to Magnetic Resonance Imaging (MRI). MR imaging is a non-invasive imaging modality which unlike Computed tomography (CT) the nature of the acquisition does not employ ionising radiation. By manipulating the main magnetic field (B_0), the gradient coils ($G_x/G_{frequency}$, G_y/G_{phase}, G_x / G_{slice}) and the radiofrequency coils (RF) are capable of producing cross-sectional images with high soft-tissue contrast which again distinguished it from CT and ultrasound. However, MRI acquisitions are characterized by long scan time which is a crucial drawback for the clinical application. The acquisition time which is directly related to the number of samples acquired in the k-space is a lengthy process which not only increases the patient discomfort and the resulted motion artifacts but also increases the scanning cost since fewer patients can be scanned per day. The main approaches on reducing the acquisition time came with the introduction of parallel imaging (PI) [4]-[5] and Compressed Sensing (CS) [1]-[3] from a hardware and software perspective respectively. PI, for 2-3-fold accelerated acquisitions, relies on the information provided by multiple receiver coils that are sensitive to different parts of the object and can operate in both image and k-space.
CS is a collection of algorithms that aim to recover signals from undersampled measurements assuming the following: firstly, the images must have a sparse representation in some transform domain (wavelet, curvelet, dictionary learning, etc.) [7]-[8]. Secondly, the measurements between the sampling and the sparsity domain should be performed in an incoherent manner to guarantee an attainable unique solution. Finally, under the latter assumptions, images can be reconstructed from a broad category of nonlinear optimisation iterative algorithms [9]-[11].
Following the success of deep learning (DL) in a wide range of applications (image classification, image segmentation etc.) [12]-[13], neural networks based on CS optimization have received significant interest for accelerating MR acquisitions and reconstruction strategies. The main power of deep architectures stands in the fact that they can extract features and build representations without the need of hand-crafting algorithms per case as in the CS framework. In CS signal reconstruction,

the iterative algorithm unrolled over a deep neural network. The basic strategy is to train a network to learn the weights (CNN kernels) for dealising undersampled MR images from a large dataset containing pairs of aliased and dealiased images.

In this work, we get motivated from the 2D Deep Cascaded CNN-Network (DCNN) [4]. DCNN mainly consists of two blocks the CNN and the Data Consistency block, operating on the image and k-space (sampling domain) respectively, linked by a regularization term which adjusts the data fidelity based on the noise level of the acquired measurements. It can be viewed as a de-aliasing problem in the image domain under the CNN block guided by the DC block for consistency with the sampled data which imitates the iterative reconstruction through the cascade of the sequential CNNs [4]. In the following chapters, we focus on the interconnection term λ which adjusts the reconstruction based on the two terms. W*e introduce an improved regularization for the data consistency term in two novel settings, a learnable regularization parameter per k-space slice and a spatially learnable regularization parameter. We show that the introduced regularization is independent of the CNN block architecture used and can be incorporated in any DL-CS based optimization network setting. We show that t*he compatibility of the proposed regularization in any DL-CS based network highly improves the reconstruction performance without adding any computational burden to the network.

1.2 Outline

In the following paragraph, the organization of the thesis is presented.

Chapter 2: provides an overview of the background field of MRI on which we are applying the DL reconstruction frameworks on the later chapters.

Chapter 3: focuses on the explanation of the framework we are going to build our reconstruction. The initial ideas on which this sort of reconstructions are based and the mathematical formulation on the derived architectures.

Chapter 4: all the materials and methods used for this work are presented. The used dataset, the novel parts of this work as well as some notations used on the result section.

Chapter 5: our results are presented for all the conducted experiments.

Chapter 6: the findings of our research are discussed in parallel with the origin of this initiative. Remarks on our future directions are discussed as well as the optimal target of our venture.

Chapter 2:

Introduction

2.1 Radiology Evolution

The field of Radiology is based on all the medical imaging modalities that can produce visual representations of the internal structures of the body for both diagnosis and treatment. It is among the most significant medical disciplines as it plays a crucial role in the development of public health. The most critical point for the initiation of the field was the accidental discovery of 'X-rays' by Roentgen in 1895 [15]. During the next year (1986) Becquerel and Curies discovered radium the future of nuclear medicine, and within a couple of years the basic technics of radiography were established. The "nuclear induction" as it was first described during the WWII by Bloch and Purcell, awarded them with the Noble Prize in 1952 for the nuclear magnetic precision measurements. In 1949 Bloch and Bloembergen discovered that a signal can be detected when a sample is placed in a magnetic field, excited it by a radiofrequency pulse (RF) of specific frequency. In 1949 the first ever NMR sequence described by Hahn, stated that you can repeatedly detect an NMR signal at a delayed time by using a second RF pulse (Spin Echo/Hahn sequence). Following the development of the SONAR radar during WWII, the ultrasound came up in 1950, involving non ionising radiation for non-invasive imaging based on the Doppler effect. The next big technological discovery in nuclear medicine came in the 1958 when Anger developed the gamma camera which used in tomographic imagine and more specifically in SPECT single photon emission computed tomography and PET, positron emission tomography. In 1959 Singer from Berkeley proposed the NMR as a non-invasive tool to measure the blood flow. In 1967 Hounsfield invented the first CT-scanner using X-rays in the EMI central research laboratories for which in 1979 awarded with the Nobel prize. In 1971 Damadian discovered that different mouse tumours lead in different relaxation times in comparison with the normal tissue. In 1973 Lauterbur using magnetic field gradients, proposed that using a back-projection reconstruction as in CT, could distinguish between NMR signals originating from different anatomical locations. In 1974 Mansfield invented the selective excitation or the synthesis of tomographic image slices (2003 Nobel Prize) and in 1975 the Ernst's group invented the 2D- Fourier transform imaging (2D FT) (1991 Nobel Prize). The first practical 2D imaging method was invented in 1980 by Edelstein and Hutchison. Finally, the first human head image was published in 1978 by Clow and Young [16]. In 1989 Ramsey a pioneer in spectroscopy developed the theory for the chemical shift on NMR which first described by Rabu in 1944. Lastly, in 2002 Wuthrich established the NMR spectroscopy based on which we can represent 3-D structures of biological macromolecules in solutions [14].

Radiology is a continuously growing field which demands a high dedication and follow up from the specialists in the field to be able to stay in line. With the booming developments and improvements on all the imaging methods several incurable diseases can be better approached, described, and comprehended.

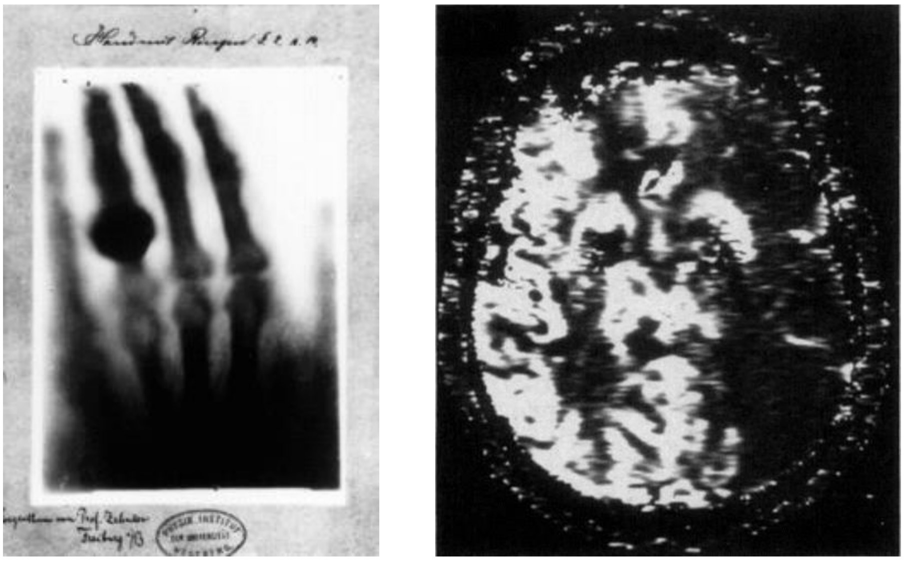


fig.2.1: *LEFT: First medical X-ray by Wilhelm Roentgen of his wife Anna Bertha Ludwig's hand [15]. RIGHT: The 1st MR image of human brain using 0.1T MRI obtained in 1978 by two groups of researchers at EMI Laboratories [16.]*

Fig.2.2: Nobel Prize winners in NMR: a) Purcell – NMR Relaxation b) Bloch – Nuclear Magnetic Precision measurements c) Bloembergen – Electrons have Fixed Energy d) Ernst – Spectroscopy e) Lauterbur & f) Mansfield basis for MRI invention. [14]

2.2 MRI Overview

Magnetic resonance imaging (MRI) is based on the alterations induced when certain atoms are placed in a magnetic field. Atoms with an odd number of protons possess a nuclear spin angular moment and therefore exhibit the MR phenomenon.

In MRI by alternating the magnetic field, the magnetic moments are excited and the induced magnetic field as the atoms relax back to equilibrium is captured. The nature of MR signal is based on the interaction of the spins with three types of magnetic fields: a) the main field B_0, b) the radiofrequency field B_1 and c) the linear gradient fields G [17].

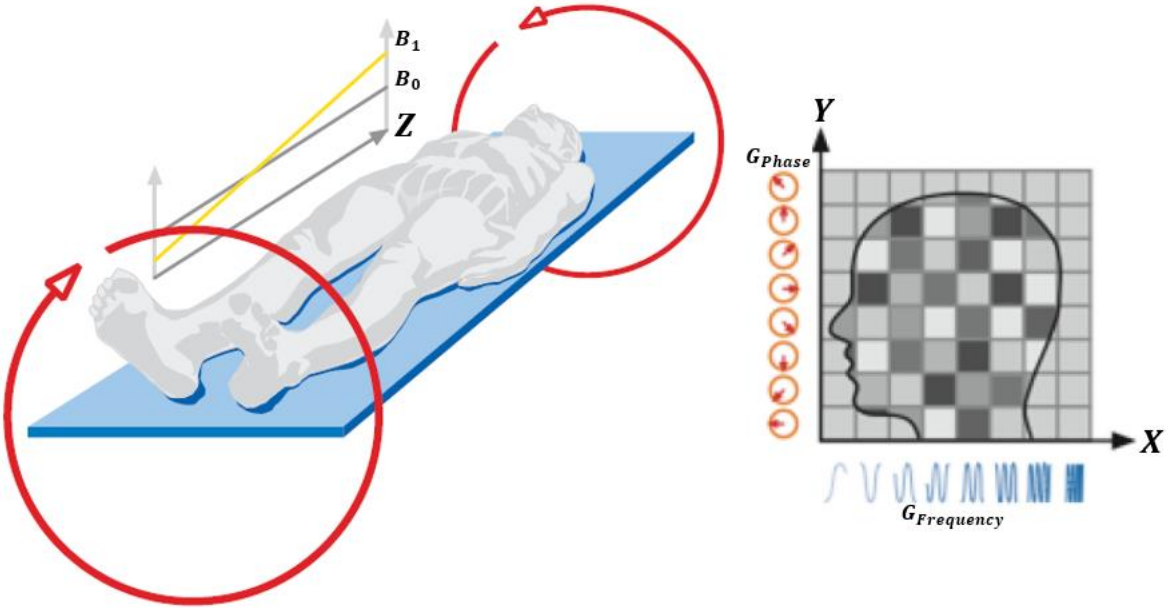


Fig.2.3: All the main magnetic fields are illustrated. The main magnetic field B_0, B_1, the G_y / G_{phase} and the G_x / $G_{frequency}$.

MRI development is based on its predecessor modality, the Nuclear Magnetic Resonance (NMR), mainly used for chemical and physical analysis as it is still today. The nuclear term in the acronym results from the fact that the whole function of the method is based on the magnetic field interactions of the nucleus. The Hydrogen atoms mainly (other atoms can also be measured), under specific circumstances, generate electric signals which can be utilised for several imaging applications. [18]-[19]. MRI is among the most widely used technics to probe the human body. In MRI the subject is not exposed to ionizing radiation as in CT, X-rays or PET and the data acquisitions does not involve any invasive procedure, thus it is considered the safer imaging methodology available in our days [20]-[21].

Since different tissues will react at different rates depending on their chemical composition or molecular structure several imaging applications are available using MRI. In biological matter since most of the body consists of H20 the most sensitive signal is produced by the H atoms. Other atoms such as phosphorus (31P) are of interest for the metabolism but from now on we most commonly assume H proton imaging when we referred to MRI.

MRI has a wide range of applications in medical diagnosis and can show distinct organs or tissue features with high quality and good contrast, under different settings. Using MRI we can obtain information for the Anatomy, the Connectivity, the Activity or even the Chemical structure of soft tissues as well as information about the blood flow in the vessels. More specifically, there are four different submodalities under the umbrella of MRI.

A. sMRI: structural or anatomical MRI. The most recognizable and applied modality, employed to examine the anatomy and pathology of the brain providing images for clinical radiological reporting.

B. dMRI: diffusion MRI. This is an MRI method that measures molecular diffusion in biological tissues. As molecules interact with many different obstacles as they diffuse throughout tissues, dMRI provides insight into the microscopic details of tissue architecture [ref]. One of the earliest applications of dMRI was Diffusion Tensor Imaging (DTI) a 3D visualization of white matter tracts using tractography algorithms [22]-[23].

C. fMRI: functional MRI. In fMRI the brain activity is measured during a cognitive task or at rest by detecting changes associated with blood flow. Cerebral blood flow and neuronal activation are coupled (Oxy/Deoxygenated blood)- BOLD contrast. Through the blood flow (hemodynamic response) related to the energy used by the brain cells, several functions can be captured through the activation areas that the blood flow followed in the brain. [24]

D. Spectroscopy MRI: It is used to measure biochemical changes in the brain, especially in the presence of tumours or other malignancies. The chemical composition between normal and abnormal tissues is compared, capturing several metabolites or products of metabolism, the constitution of the tissue is determined.

The main drawback of MRI is the long imaging time [25] which results in an inherently slow imaging modality. Due to that the subject is required to remain motionless in a tight environment which on one hand reduces subject's compliance, mainly in paediatric populations and on the other hand MRI acquisitions are characterized from many artifacts, especially but not only motion related.
To deal with the inherently slow acquisitions mainly two methods have been introduced for imaging acceleration. The first one is parallel imaging which is a hardware-based method, and it relies on capturing information through multiple coils, sensitive to different part of the object, simultaneously. Depending on the amount of the available coils it can reach up to 2- or 3-fold acceleration [5],[26]. On the other hand, Compressed Sensing (CS) is a software-based technic which aim to recover signals from undersampled measurements given that some constraints are fulfilled based on mathematical formulations using non-linear reconstruction schemes [27]-[28]. Based mainly on the CS methodology and the iterative reconstruction schemes, the recent years the introduction of Deep Learning in the field is seeking to accelerate

further the diagnostic capabilities of MRIs both during acquisition but also in the postprocessing stage.

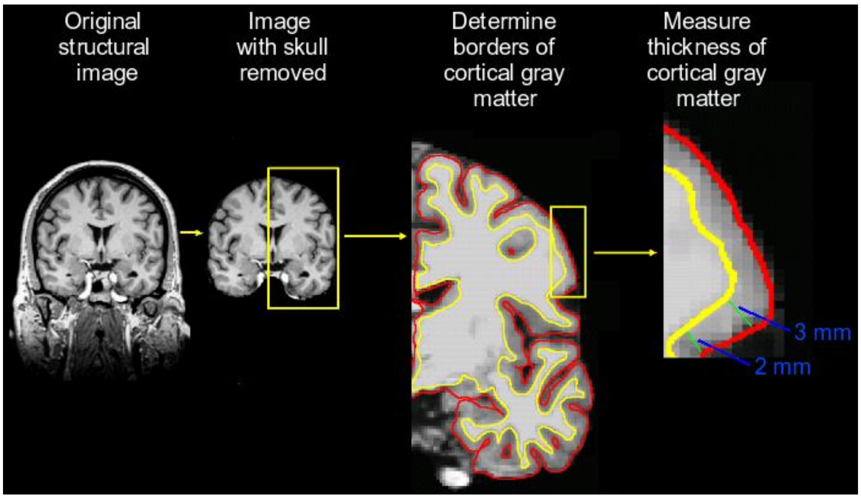


Fig.2.4: *Structural/Anatomical MRI images of brain anatomy. Current application to measure the thickness of the cortex. Cortical thickness changes occur after a pathology (stroke) or after repeated exposure to particular experiences. [28]*

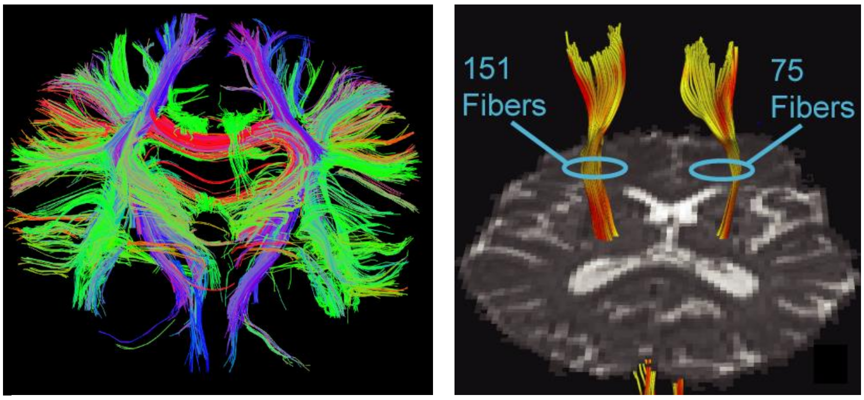


Fig.2.5: DTI to measure microscopic movement of water in the brain. DTI can be used to evaluate the integrity of the white matter in the brain. On the right image the effect of a stroke is shown where fewer fibers are present on the side of the stroke [28].

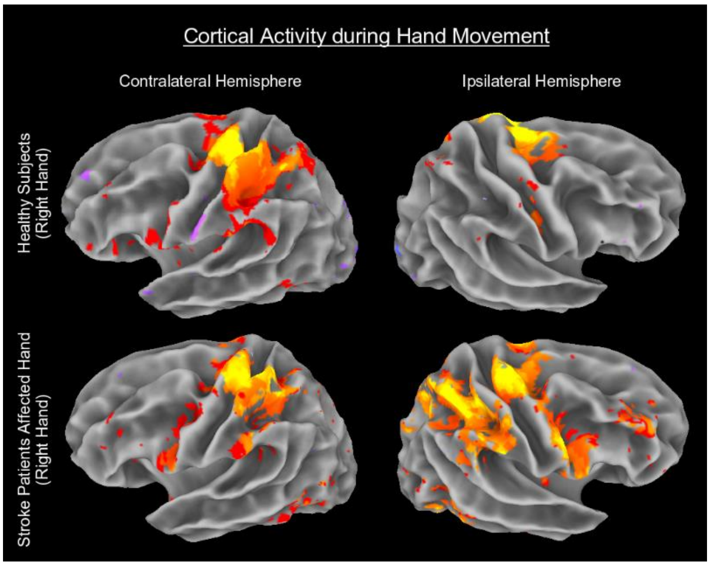


Fig.2.6: fMRI an indirect measurement of brain activity. We can observe the level and pattern of activity alteration between a Healthy subject and a Stroke patient. [28]

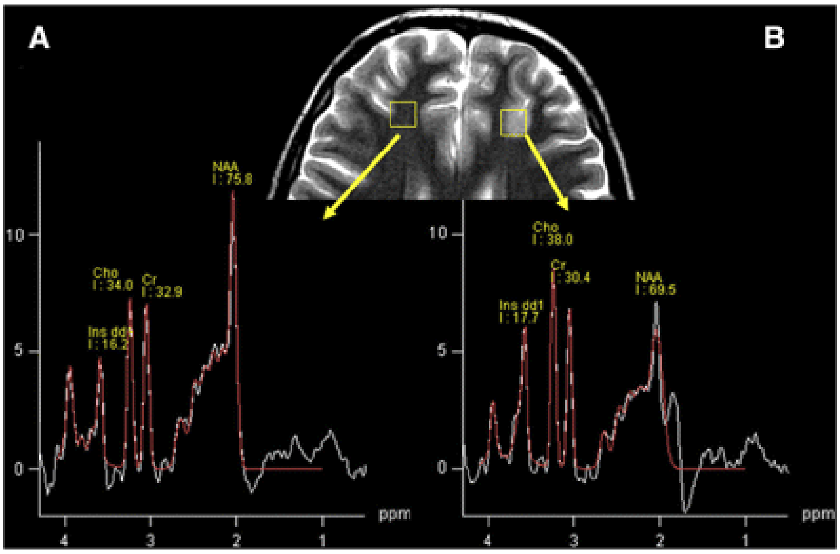


Fig.2.7: Spectroscopy MRI. Normal spectrum and slight Choline elevation with NAA reduction on patients with a seizure [29].

2.3 MRI System

The MRI system mainly consists of the following parts:

a. The magnet
b. Radiofrequency Coils
c. Gradients
d. Computer System
e. Patient handling System

2.3.1 The Magnet

The most important component of the MR system is the magnet. The common field strengths used in clinical scanners may vary between 0.5-1.5 T while in pre-clinical research even 11.7 T magnets (CEA-Paris-Saclay) are used. The highest is the magnetic field the higher is the produced MR-signal, signal-to-noise ratio (SNR). To have an idea of the operational strength of the magnetic field, 1 Tesla equals 10000 Gauss (1G = 0.1 mT) and the Earth's magnetic field is approximately 0.05 mT (0.5G), which means that a magnet of 1T applies magnetic force which is 20K times more than the Earth's magnetic field (gravity). The most modern systems use superconducting magnets in two settings closed or open bore (bore - opening where the subject is placed). The most important characteristic of a magnet is the **homogeneity**, the quality, or the uniformity of the produced magnetic field. [14]. The Magnet produces the main magnetic field which is denoted by Bo.

2.3.2 Radiofrequency Coils

The MR signals are collected from the patient's tissue, produced by the radiofrequency coils (RF pulse). A radiofrequency coil with more than one element (multiple elements) is called *array.* A transmitter coil surround the subject's part of the body under examination and produces the RF pulse. The resulted signal, produced after the initial RF pulse penetrates in the tissue is collected from a receiver coil. Depending on the scanning anatomy several coils exist (body, head, spine, knee, etc.)

2.3.3 Gradients

To localize the MR signals in the body a short-term spatial magnetic field variation needs to be produced. These variations are produced by the gradient coils or gradients. Each system has 3 different gradient coils, in all 3 directions (x, y, z).

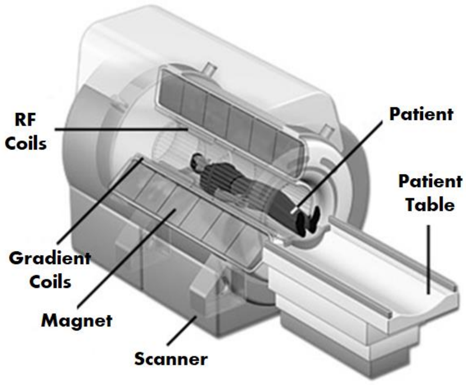
Each scanning anatomy, depending on the clinical protocol under investigation (anatomy, malignancies, neurodegenerative issues) needs to activate all the above coils in a repeatedly manner with specific amplitudes and duration. This series of radiofrequency pulses applied to the sample is called **pulse sequence**.

2.3.4 Computer System

The operation of the whole MR System takes place through a work-station console. Medical details for the subject, as well as commands for the operation of the MR scanner but also for the post processing stage of image reconstruction are communicated through the latter computer system.

2.3.5 Patient Handling System

The last part of the system is the patient's bed, where the subject is placed to be inserted in the magnet bore. During the positioning of the subject, the array coils are placed on the subject, a communication device, and in specific cases further equipment might be connected such as peripheral pulse or EEG.



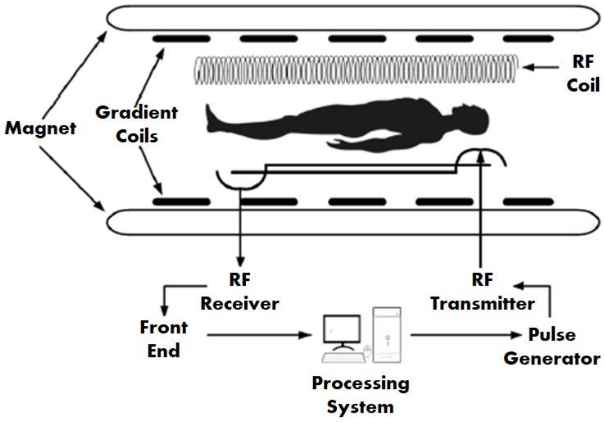


Fig.2.8: MRI System Cutaway.

2.4 MRI Physics for imaging

The MR phenomenon is based on the nuclear spin; atoms with an odd number of protos and/or neutrons possess this property. These nucleons can be visualized as spinning charged spheres and give rise to a small magnetic moment, the spin. Since we humans are 70% water, the ^{1}H is the most prominent element in our body and the most common measured in MR Imaging. The main interactions that can be measured by MRI are the spin-lattice T1 and spin-spin T2 relaxation, through three types of magnetic fields: a) the main field B_0, b) the RF field B_1 and c) the linear Gradient fields G. In the following paragraphs we make a short introduction in each of these fields.

2.4.1 Main Field Bo

The spins, in the absence of an external magnetic field, are randomly oriented and the net macroscopic magnetic moment is zero. However, if we activate an external magnetic field B_0, we can observe two noticeable effects. First, all the magnetic moment vectors tend to align with the direction of the B_0 and a net magnetic moment is produced. The direction of all the aligned magnetic moments is called longitudinal and by convention it is attributed to the z-direction. Second, all the spins exhibit the same resonance at a frequency ω called Larmor. This frequency is proportional to the applied magnetic field B_0:

$$\omega = \gamma B \text{ or } f = \gamma/2\pi B$$

Where γ is the gyromagnetic ratio, unique for each atom type. By placing a sample in a constant magnetic field B_0, all the atoms are polarized inducing a net magnetization parallel with the B_0 field in the z-direction. Also, all the spins exhibit resonance at their Larmor frequency. A radiofrequency magnetic field excites the spins at their specific Larmor frequency, and they produce a signal proportional to this frequency.

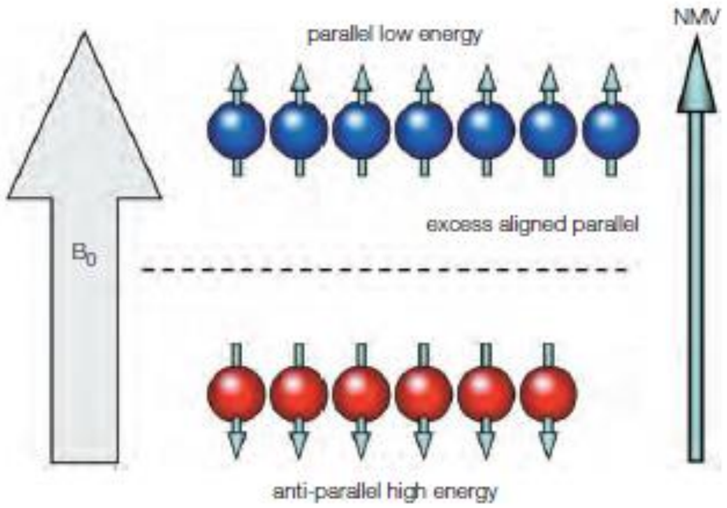


Fig.2.9: Alignment of spins with the main magnetic field B_0. Low energy nuclei align their magnetic moments parallel to the main field. High energy nuclei have enough energy to opposite the main field. In the end all the spins are align to B_0 due to the repulsion between B_0 and the magnetic moments. (NMV- Net Magnetization Vector). [31]

2.4.2 RF Field B1

To obtain the MR signal, a second RF magnetic pulse B_1 is applied to turn the magnetization in the xy (transverse) plane. The B_1 rotates the magnetization vectors by a pre-decided angle (flip angle α), adjusted by the B_1 magnitude (strength) and duration. The activation of the B_1 field lasts only for a few milliseconds and then the tipped vectors are left to process back to equilibrium. The precession takes place initially in the xy plane- transverse magnetization and is characterized by the T2 time decay constant. When the precession of the frequencies occurred along the z-axis the time constant is called T1. These time constants are the most important MR parameters, along with the proton density (PD) values. The generated signal is called free induction decay (FID) and the main objective of MRI is to map the spatial distribution of their amplitudes to reconstruct the scanning organ/tissue.

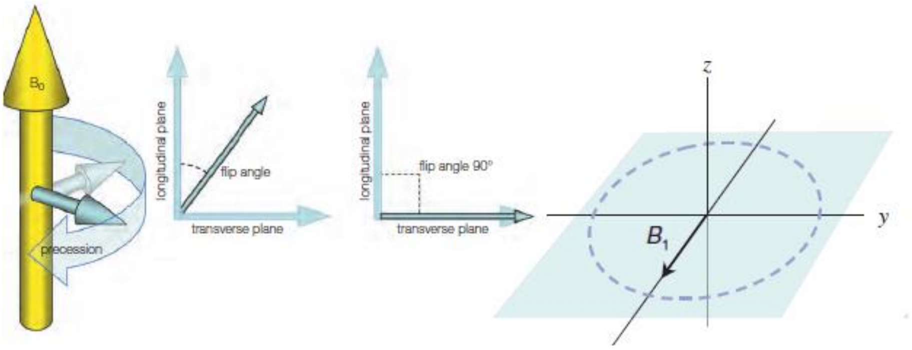


Fig.2.10: B1 rotates the magnetization vectors to the transverse plane (xy)

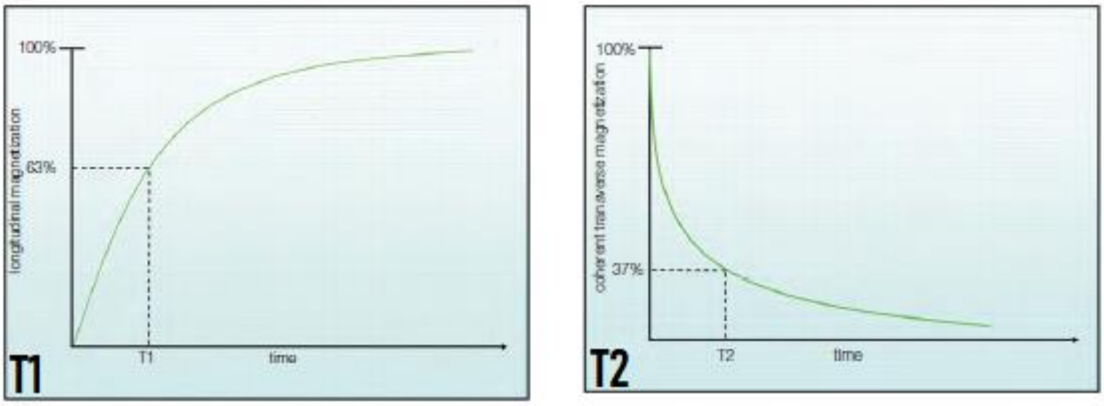


Fig.2.11: Indicative T1(longitudinal magnetization) & T2(transverse magnetization).

2.4.3 Gradient Fields G

With the activation of the main B_0 field, all the spins possess at the same resonance frequency. Thus, we cannot differentiate the signals produced at different spatial locations. In this case we record the signal generated from all the excited regions. In MRI to achieve spatial varying magnetic fields, linear magnetic fields are applied in addition to the Bo through the Gradient fields. And since we want to spatially classify the magnitudes in the 2D plane we vary the magnetic fields in two ways. First, we can introduce a frequency differentiation, in the x-plane denoted by Gx, thus the time signal contains contributions of oscillators emitting signal over a range of frequencies. Second, on the remaining y-axis, we can introduce alteration on the phase of the oscillators through the Gy gradient. Finally, in such a case the resulted MR signal characterized by different frequency and phase for each spatial location can be accurately determined.

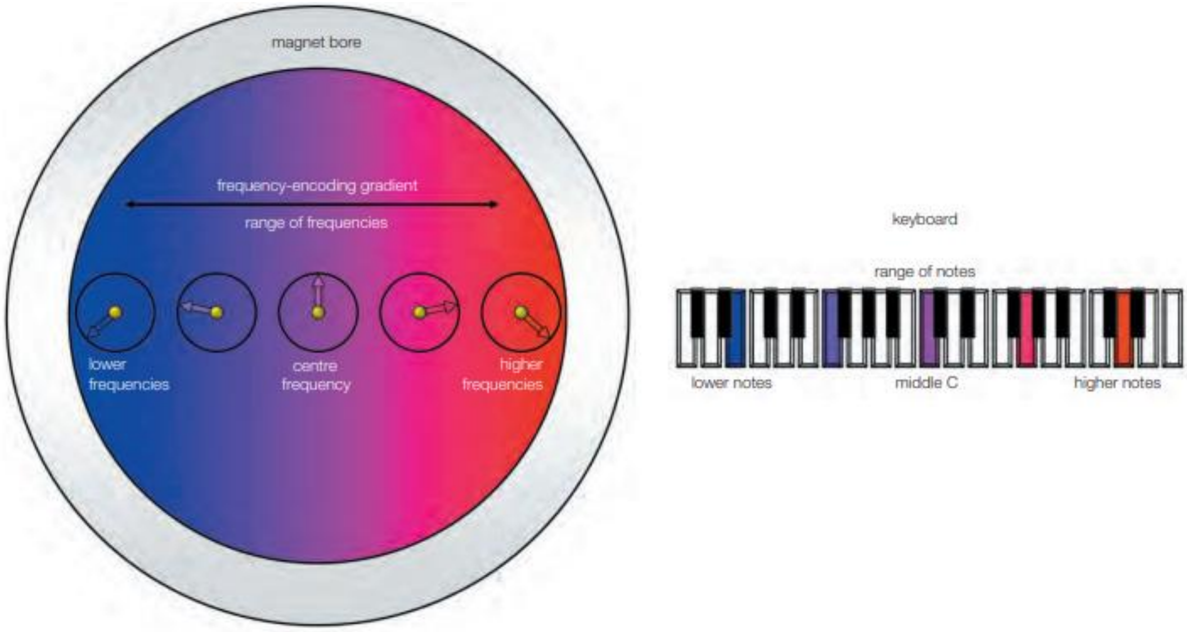


Fig.2.12: Frequency encoding. Analogy with the piano the highest is the frequency the fastest the spins their precessional frequencies increase. [31]

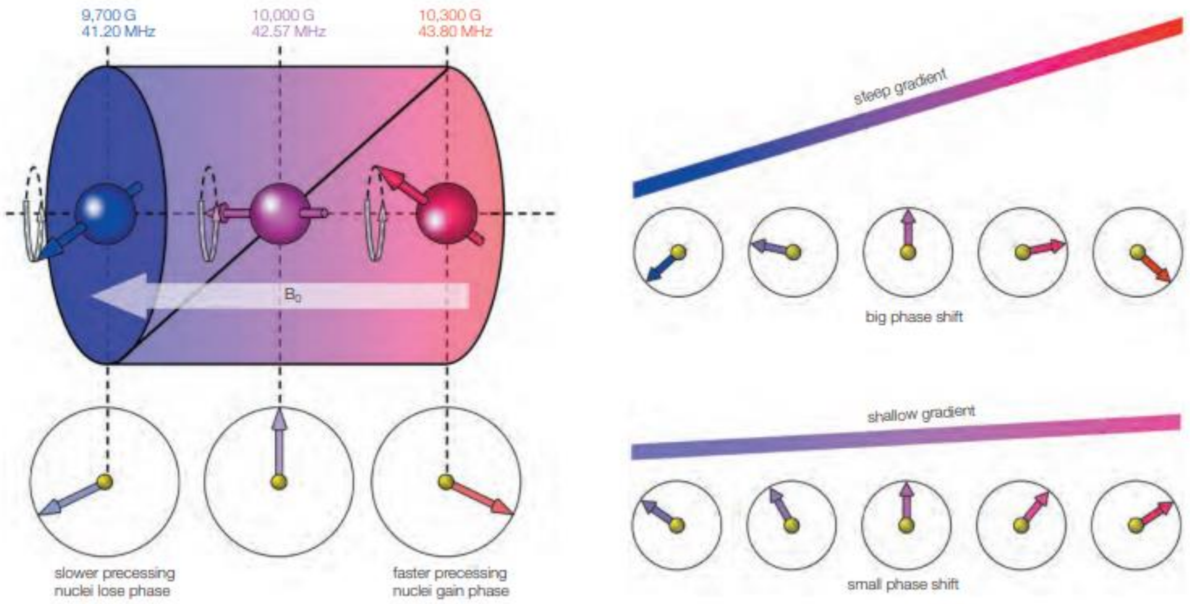


Fig.2.13: Phase Encoding through the Gradients. The highest is the Gradient Amplitude (Stronger) the bigger is the phase shift that is introduced [31].

2.4.4 Bloch Equation

The total behaviour of the magnetization vector M for all the x, y, z directions, for the Bo (Mo magnetization) magnetic field and the magnetization arise from the G (Gx, Gy, Gz) gradient fields is given from the Bloch equation:

$$\frac{dM}{dt} = M\ x\ \gamma B - \frac{M_x i + M_y j}{T_2} - \frac{(M_z - M_0)\ k}{T_1}$$

Where i, j, k are the unit vectors in the x, y and z direction respectively.

2.5 Imaging Methods

The main two parts of an MRI pulse sequence are the excitation and the reception. More sophisticated sequences are of course available but for the conventional methods these are the main parts. As we already mentioned in the previous paragraphs the initiation of the sequence starts with an RF pulse to tip the spins in a specified anatomical region. The recordable RF signal occurs when the spins have been tipped to the xy (transversal) plane and the rotation is occurred in this plane. The different excitation pulses are combined to become some function of the MR parameters (T1, T2, PD). At this point, we have excited a specific volume and all the measured signals from the processing magnetic dipoles, with a specific magnetic distribution in the different axis m (x, y, z) need to be reconstructed.

2.5.1 k-space & FOV

The matrix where our data are stored, the energy released from de-excitation of the protons, is called k-space. The stored data are frequencies and kx, ky are used to represent rows and columns respectively on a rectangular grid. The values are complex data represent both frequency and phase information. The central points - low frequencies contain information about the global structure of the subject while the periphery points - high frequencies contain information regarding the edges of the object - resolution. Important is here to clarify that the corresponding points in the k-space are not coincide with the same points in image domain. To reconstruct the image of the subject from the k-space data, the Fourier Transform is used, which will be briefly presented in the next paragraph. [32]-[33].

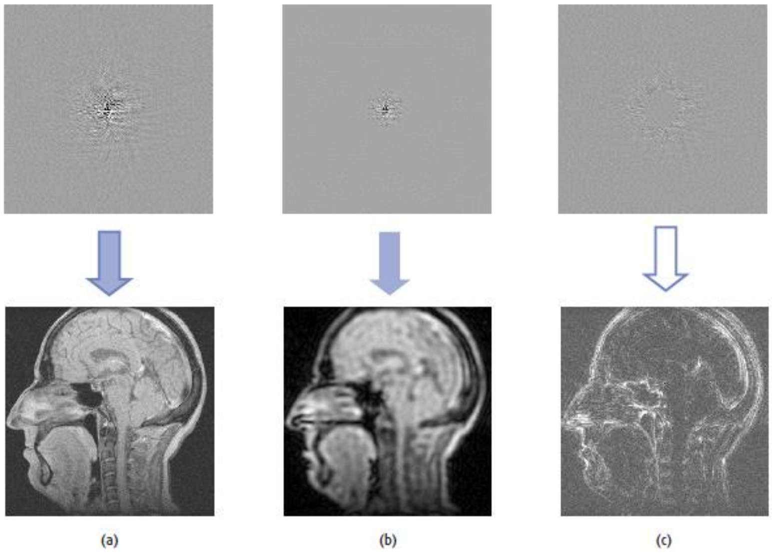


Fig.2.14: k-space sampling and the reconstructed image. A. Fully sampled k-space, B. Only the central part of the k-space – (low frequencies) is sampled – general shape information. C. Only the periphery of the k-space - (high frequencies) is sampled – edges information/resolution. [14]

The size of the area from which the MR signals are sampled is called Field-of-View (FOV). Hence, the received signal ~ pixel/voxel size is proportional to the size of the FOV. An increased FOV results in more sampled points, while the pixel size become bigger, which results in lower spatial resolution. The desired part of the image will appear smaller is the latter case. Also, the size of the covered image depends on the sampling rate, the highest the sampling rate (smaller distance between k-space points) the higher the image size.

2.5.2 Pulse Sequence

In the previous paragraph we tried to present a short description of what is known in the field of MR imaging as a pulse sequence. A pulse sequence, defines and controls the order of appliance, as well as the timings for all the provided RF and gradient pulses [34]. Depending on the pathology under examination, the sequence characteristics for each tissue may vary as a function of the different imaging parameters (duration, amplitude of each pulse). The main timings which characterize the duration of a specific pulse sequence are the TE - echo time and TR - repetition time. TE is the time interval between the initiation of the sequence and the recordable signal, depending on which contrast we are interested in (T1, T2, PD) the TE will vary. TR is the time needed for one repetition of the sequence. The time interval between the termination and the reinitiating of the sequence, is usually used for magnetization recovery. Depending on the size of the scanned volume in association with parameters as the slice thickness for example, each pulse sequence is repeated several times.
A pulse sequence diagram is represented by several axis, stand for the different pulses and gradients applied in a time order manner. The main axis presented are mainly four or five for 2D or a 3D sequence respectively (second G_p). The first axis denoted by RF is the pulse selected for the excitation of a specific volume. The second

axis referred to the z direction gradient, G_z or G_{slice} (G_s). The G_s is occurred simultaneously with the RF pulse to tip the spins to the transversal plane. The remaining two directions G_y, G_x are the G_p / G_{phase} and G_f/ $G_{frequency}$ gradient, encode the spatial frequencies by applying a phase and frequency encoding in the resonant frequencies of the spins respectively. In the following fig.2.15 you can observe the pulse sequence diagram of the Spin Echo sequence. The main steps for the 2D imaging methods are the following:

A selective 90-degree excitation pulse is applied in the presence of G_z such that the nuclear magnetization vectors in a thin plane in parallel with the z-axis are tipped into the xy-plane. The state of the vectors immediately following the excitation - ideally all vectors pointed in the same direction. The resultant signal at this moment equals the 'magnetization volume' of the specific slice. The measured location is the beginning of the k-space since the volume under a 2D function m (x, y) equals the value of its Fourier transform at (k_x=0, k_y=0) at t=0. Position 1 in fig.2.15**.** Next, we need to introduce to the excited spins an amplitude distribution m (x, y) that it will help as distinguish each spatial location. As the excited spins precession and induce an electromagnetic field (EMF) signal in the receiver coil, we encode their spatial location using the other 2 axis gradients G_x (G_f) and G_y (G_p). These steps depend on the used sequence.

Continuing with the sequence, the y-gradient is switched on, which introduces a linear phase difference on the spins along the y-direction. Position 2 in fig.2.15**.** The longer the G_y (G_p) gradient is left on, the larger is the phase variation that is introduced. Position 3 in fig.2.15. The particular spatial frequency of each spin is analogous to the linear phase variation with time. Each x-component is weighted by a cosine function of the vectors at each y-position.

Right after, when the G_y gradient is turned off, the G_x (G_f) gradient is turned on and introduces a frequency difference along the x-axis. Position 4 in fig.2.15**.** The signal at this moment is the sum of all the vectors and equals the k-space value at the appropriate (k_{x_1}, k_{y_2}) position.

Finally, the signal decays and the spin vector return to equilibrium. Several repetitions of the sequence are needed to cover the whole imaging slice/volume.

The pulse sequences of the two most commonly pulse sequences are presented in the following figure. Spin Echo and Gradient Echo.

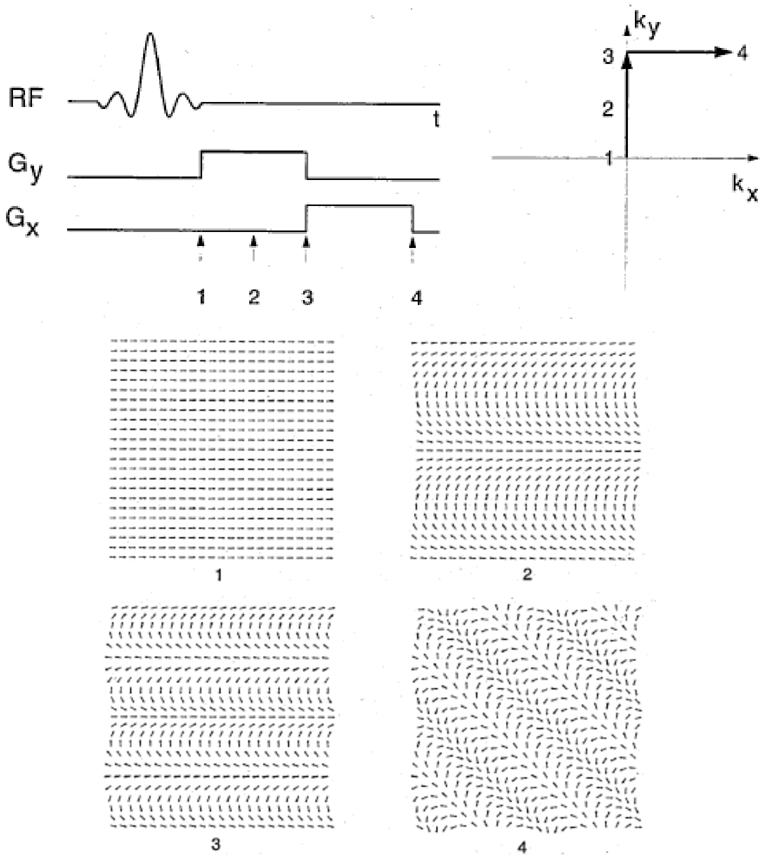


Fig.2.15: A pulse sequence diagram. 1) Initiation of the sequence – (k_x, k_y = 0, 0). 2) Gp=ON a phase variation is introduced. 3) Phase of each nucleus is proportional to the linear phase variation with time. 4) Gx=ON a frequency difference is applied at the nucleus along the x-axis. [35]

The pulse sequences of the two most commonly pulse sequences are presented in the following figure. Spin Echo and Gradient Echo.

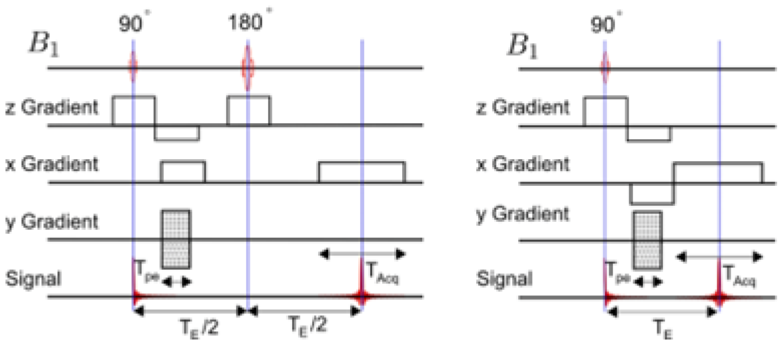


Fig.2.16: Spin Echo & Gradient Echo pulse sequence Diagram [36]

2.5.3 k-space Trajectories

The route that is followed to sample the k-space points is called trajectory. In the more conventional way, trajectories follow a line-by-line path with points lying on a rectangular grid. This sort of trajectories called Cartesian and an inverse Fourier Transform is applied to directly obtain the image of the scanned object. However, other trajectories are available such as the radial or the spiral. These trajectories are non-Cartesian which means that prior to the application of the Fourier transform a regridding strategy is needed to first place the sampling points on a Cartesian grid (interpolation) and then apply the Fourier Transform. Non-Cartesian trajectories result in faster scanning, and they provide an efficient k-space coverage with fewer spokes in Radial case or a smaller number of shots in the spiral trajectory. Apart from acceleration during scanning, these trajectories can be directly undersampled without artifacts that cannot be removed in a postprocessing stage as in the Cartesian case. This is due to the architecture of the trajectory where each spoke or shot contains equal amount of low and high frequency components in comparison with the line-by-line scanning (*either low or high frequencies per line) Cartesian trajectories. The last identity is very important when dealing with acceleration methods using undersampling masks and will further analysed in the next paragraphs. Each of these trajectories is characterized by specific properties but one common important characteristic is the motion tolerance. For example, in radial sampling, due to the nature of sampling, the centre of k-space is densely sampled and thus present motion robustness. [37]

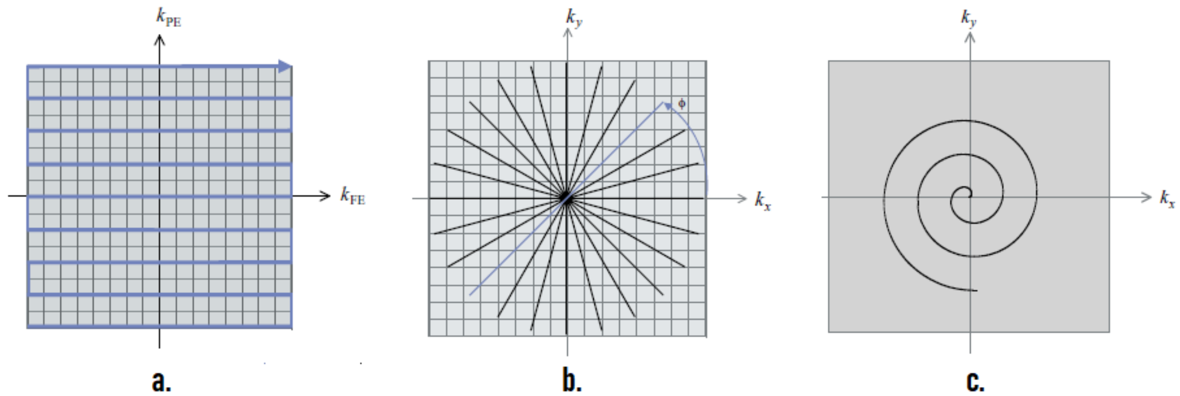


Fig.2.17: k-Space Trajectories. a. Cartesian b. Radial (Non-Cartesian) c. Spiral (Non-Cartesian) [14]

2.5.4 Fourier Transform

The Fourier Transform is a mathematical technique that allows an MR signal to be decomposed into a sum of sine waves of different frequencies, phases, and amplitudes as follows:

$$s(t) = a_0 + a_1 \sin(\omega t + \varphi_1) + \alpha_2 \sin(2\omega t + \varphi_2) + \alpha_3 \sin(3\omega t + \varphi_3) + \cdots$$

where a_i's are amplitudes, φ_i's are phase shifts, and ω is the fundamental frequency. The higher order frequencies $2\omega, 3\omega$, etc. are called harmonics. [35]

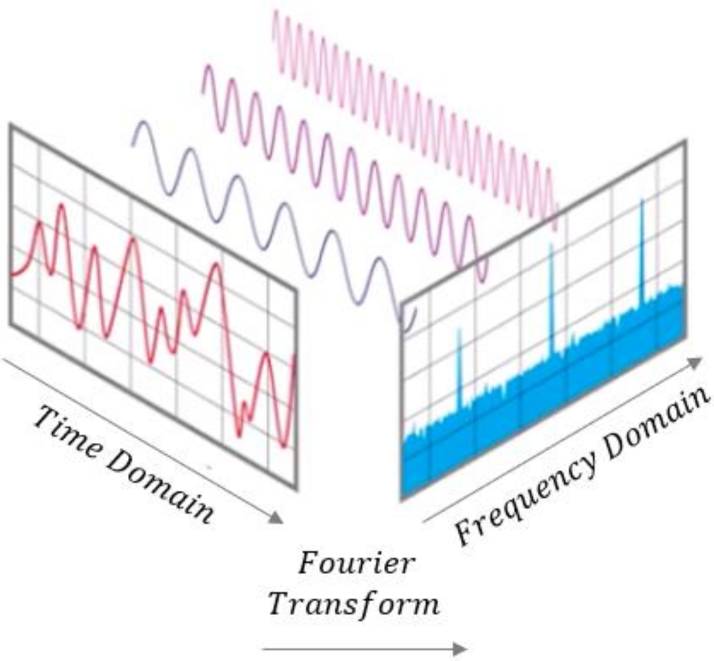


Fig.2.18: Fourier Decomposition of a time signal in a sum of sinusoids and their respective frequencies.

In the above figure we can observe both the Time and the Frequency Domain signals. A brief explanation of Fig.2.18 is that starting from the Time domain signal $s(t)$ we can see the decomposition of it in a sum of sinusoids for each of which we have a specific amplitude $S(\omega)$ in the Frequency domain. The Fourier Transform is a mathematical procedure connecting the $s(t)$ with the $S(\omega)$ as follows:

$$S(\omega) = \int_{-\infty}^{=\infty} s(t)e^{-i\omega t}dt \qquad \boldsymbol{FT}$$

$$s(t) = \int_{-\infty}^{+\infty} S(\omega)e^{i\omega t}d\omega \qquad \boldsymbol{IFT}$$

The 1st equation is the Fourier Transform and the 2nd one the Inverse Fourier Transform, used to convert the k-space to image and vice versa Fig.2.20.

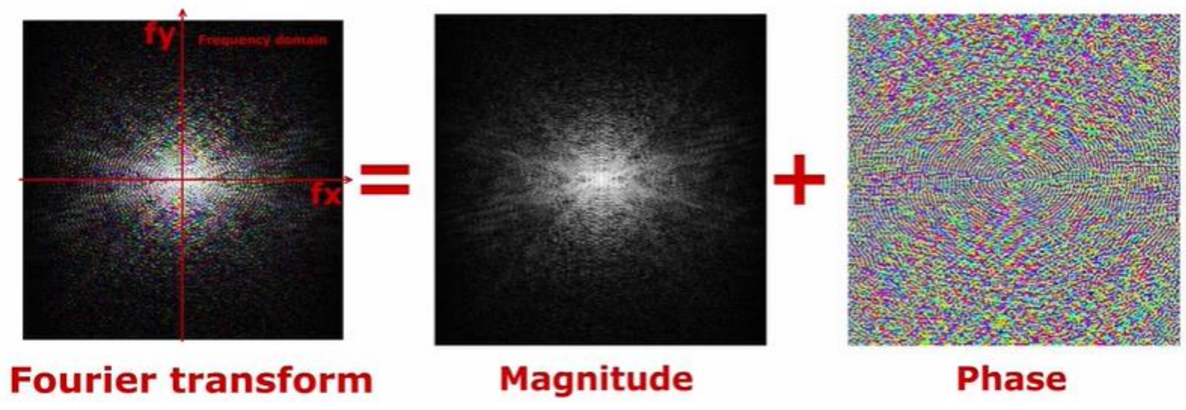


Fig.2.19: The magnitude and frequency of each frequency component along x and y.

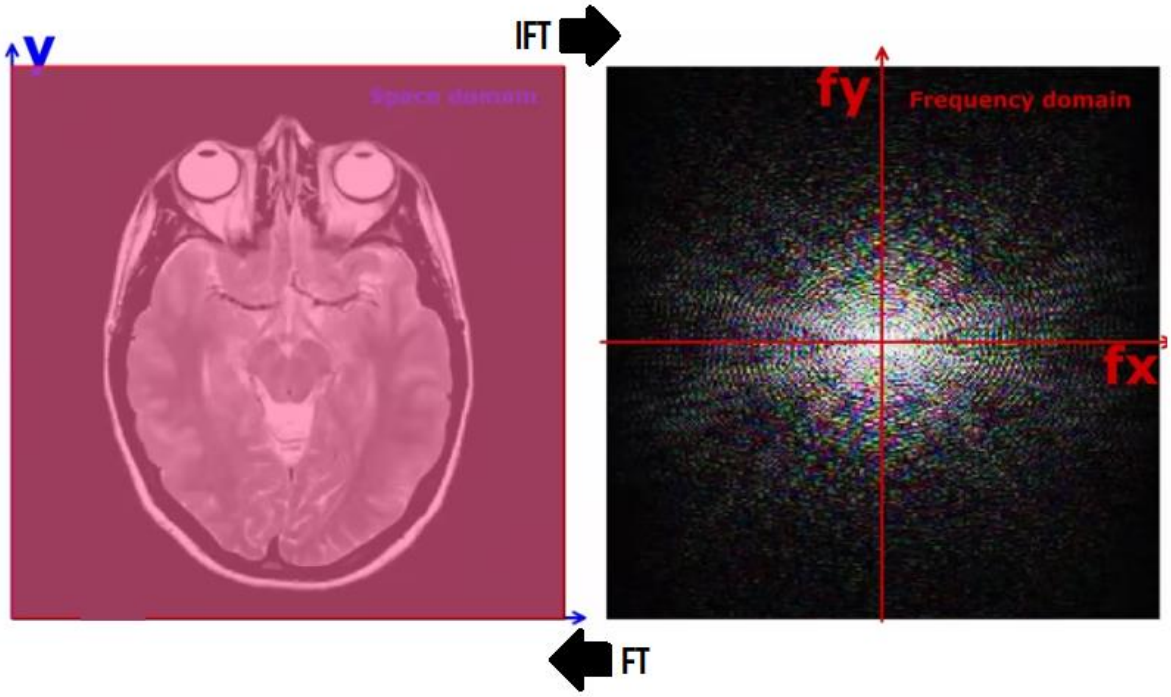


Fig.2.20: The Fourier and the Inverse Fourier transform between the Image and the k-space respectively.

Chapter 3

3. Image Acceleration

MR imaging is a non-invasive, radiation free imaging modality which offers high resolution imaging with various contrast mechanisms to reveal different properties of the underlying anatomy. However, the main drawback of MRI is the inherently slow image acquisition. The sampling process is performed in the k-space (frequency space) in a sequential manner and the speed at which it can be traversed is limited by physiological and hardware constraints [54]. The inherent slow acquisition is associated not only with patient discomfort from the lengthy examinations but also with artifacts mainly motion related. The main two approaches for acceleration are: a) Parallel Imaging (PI) from a hardware perspective and b) Compressed Sensing (CS) from a software perspective. Our main work is based on the CS framework on which acceleration is achieved through k-space undersampling. In the following paragraphs we introduced the CS theory with all the constraints and limitations, as well as the transition to the Deep Learning (DL) methodologies.

3.1 Compressed Sensing (CS)

Compressed Sensing (CS) aims to reconstruct signals and images from significantly fewer measurements than were traditionally thought necessary. The main idea behind this technic originates from transform-based compression a widely used strategy adopted in the JPEG and MPEG standards. The technic is based on a sparsifying transform, mapping image content into a vector of sparse coefficients and then encode the sparse vector by approximating the most significant coefficients and ignoring the smaller ones.

In CS one measures a relatively small number of ‘random’ linear combinations of the signal values - much smaller than the samples nominally defining the signal. Since the signal is compressible the total number of signal samples is a gross overestimate of the ‘degrees of freedom’ of the signal. In MRI the samples are individual Fourier coefficients. The successful application of CS has three requirements:

A. T*ransform Sparsity*: The desired image should have a sparse representation in a known transform domain.
B. *Incoherence* during undersampling: The artifacts produced after the reconstruction should look like noise - be incoherent in the transform domain
C. *Nonlinear reconstruction*: The final image should be iteratively reconstructed by a nonlinear method that enforces both sparsity and consistency of the reconstruction with the acquired samples.

<u>*Sparsity*</u>*: A* predefined mathematical transform such as wavelet, curvelet, discrete cosine transform etc. is used to transform the initial signal in a domain where only a few coefficients remain significant – sparse signal.

<u>*Incoherence:*</u> the way we acquire the samples from the whole signal needs to be random. In case of Cartesian trajectories, mathematical orthogonal basis *[1]-[3]* that produce random measurements need to be used to avoid aliasing artifacts during the reconstruction. For the Non-Cartesian trajectories, the methodology of undersampling is straight-forward without the need of such a base. In radial imaging for example by discard random spokes we produce undersampled measurements.

<u>*Nonlinear Reconstruction:*</u> Let m be a complex vector of measurements, Ψ the sparsifying transform (eg. wavelet), F_s the fourier transform including the undersampled mask (Φ) and y the measured k-space data form the MRI scanner. ε controls the fidelity of the reconstruction. Then our reconstructions are obtained by solving the following constrained optimization problem:

$$\min|\Psi m|_1 \quad s.t. \quad |F_s m - y|_2^2 < \varepsilon$$

$$F_s = F * \Phi \;\; and \;\; \Theta = F_s * \Psi$$

Minimizing the L1-Norm promotes sparsity while constraint by the L2-Norm enforces data consistency. In other words, *among all solutions that are consistent with the acquired data, we want to find a solution that is compressible by the transform Ψ.*

There are many different formulations for the latter definition of the nonlinear reconstruction.

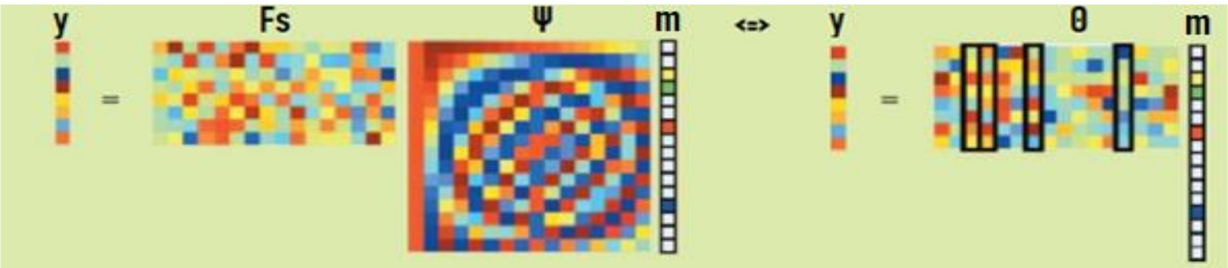


Fig.3.1: Schematic Representation of the CS measurements. Where $Fs = F * \Phi$ the Fourier transform (F) Including the undersampling mask (Φ), Ψ the wavelet transforms and $\Theta = F * \Phi * \Psi$, S is the whole Signal and y and undersampled measurements.

3.2 Machine Learning in Medical Imaging

Machine learning is an active highly growing field of research where algorithms are combined with statistics to extract and discover meaningful patterns based on observations (examples). Beyond human perception, computers by repetitively execute ML algorithms can learn complex patterns. ML algorithms extract image features by eliminating redundant variables or representations to better estimate input data with high accuracy. Next, based on the extracted features learned from the training data, predictions are performed in the 'unseen' data. Deep Learning (DL) is an extension of ML, but the features are identified automatically during the learning process which are useful representations depending on the task.
In Medical Imaging the latter algorithms have been used for segmentation [38]-[39], image reconstruction [40],[4], image registration [41],[42] or classification [43]. The main applications in radiology aim to provide direct interpretation of the results leading in accelerated diagnosis [44]. The state-of-the-art applications on the field target in quantitative learning – results [45].
DL as a branch of ML uses multiple deep layers to model the complex relationships between input and output. The most popular and widely used building pillars of DL are the:

a. Convolutional Neural Networks (CNN)
b. Recurrent Neural Networks (RNN)
c. Generative Adversarial Networks (GAN)
d. Autoencoders (AE)

In our work we are interested in DL reconstruction algorithms for medical imaging applications. In the provided architecture we mainly use CNNs for the current work even thought, the proposed regularization is not limited to these building blocks, but any architecture can be used for the sparsity term.
Since we apply our model in a CNN based architecture, we will perform a brief introduction to these specific models.

CNN

CNN is one of the most used models in computer vision tasks where each neuron receives an input, performs a dot product and a non-linearity function follows to provide the output. This sort of architecture explicitly assumes that the inputs are images and neurons at each layer are connected only to a small region of the previous layer instead of the full image. The main pillars used in a CNN setting are:

A. *Convolutional Layer*: The core element of CNNs, consist of a set of learnable filters - a set of weights and biases for each layer. Each neuron is only connected to a small region of the input. The spatial extend of the total neuron's field of view is a hyperparameter called receptive field. During the forward pass each filter slides across the whole input and for each height x width neighbourhood, the dot product is calculated between the input and the filter weights. This produces an activation map which corresponds to responses of filters at each spatial position. In contrast with the simple neural networks whose neurons at each layer are independent, here the weights are shared for each neuron. In image processing, coloured pictures are represented by 3 channels for Red, Green, and Blue respectively. In MRI reconstruction there are only 2 channels which mainly represent the real and imaginary channel of the complex input.

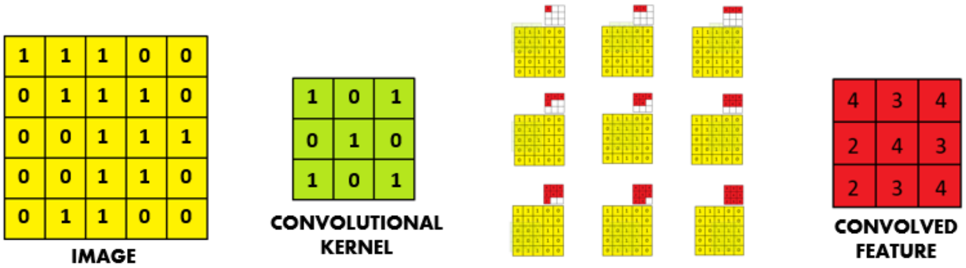


Fig.3.2: Convolution of a 5x5x1 Image with a kernel 3x3x1 to get a convolved feature of 3x3x1.

B. *Pooling Layer*: Mainly common to the CNN architectures this layer is used to reduce the spatial size of the representation - reduce the number of parameters passed in the subsequent layers and thus limit the computational power required to process the data. It is mostly found in two versions: max pooling and average pooling where the max and average of a region is calculated respectively.

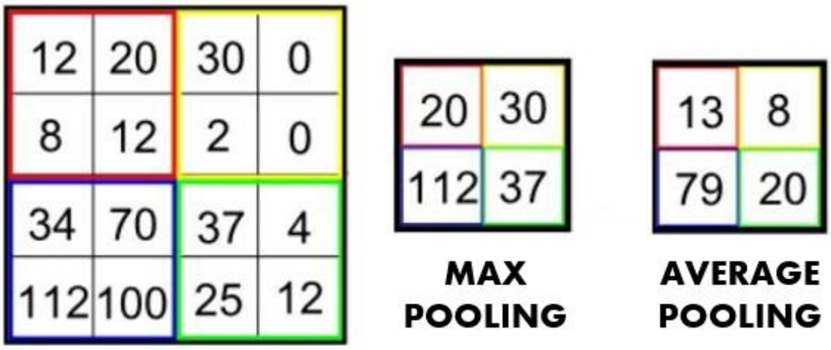


Fig.3.3: Types of Pooling.

c. *Fully Connected Layer*: This layer is mainly the same with the regular neural networks which is connected to all activations from previous layer. The activation is a matrix multiplication with a bias offset.

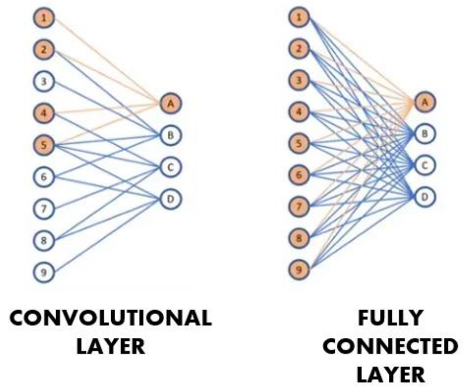


Fig3.4: Convolutional and Fully Connected Layer respectively (FC). In the FC layer all the neurons of a layer are connected to all the neurons of the next layer

A few words for the most recent architectures use CNNs. The ResNet [46] winner of ILSVRC; the main novelty in the ResNet are the skip connections apart from the regular convolutional path. In such a way, the most common issue of deep CNNs, the vanishing gradient, is overcome. The Dense Net, same as the ResNet but with densely connected layers to outperform the performance of ResNet. Finally, U-Net is a recent architecture in which pooling and upsampling layers are used, and the feature maps from the two paths are concatenated. In this setting the gradients from higher layers are propagated to lower layers directly which is highly beneficial for performance gain.

3.2 DL Based Reconstruction

The key strength of DL is its capacity to model complex input-output relationships between large amount of data [47]. From all the different DL techniques for computer vision and image processing we are mainly interested to the nonlinear image reconstruction under the umbrella of CS for accelerated MRI reconstruction.

A DL method learns a non-linear function $f: Y \rightarrow Y$ from a set of all possible mapping functions F. The empirical risk [48] $L(f)$ can be estimated using some loss function as:

$$\hat{\boldsymbol{L}}(\boldsymbol{f}) = \frac{\boldsymbol{1}}{\boldsymbol{2}} \sum_{\boldsymbol{i}=\boldsymbol{1}}^{\boldsymbol{m}} \boldsymbol{l}(\boldsymbol{f}(\boldsymbol{y}_{\boldsymbol{i}}), \boldsymbol{x}_{\boldsymbol{i}})$$

and the generalization error of mapping function $f(.)$ can be measured using some notion of accuracy measurement. In MR image reconstruction using DL we try learning a map from the undersampled k-space measurement $y \in C\ (N1xN2\ or\ N1xN2x2)$ to an unaliased MR image $x \in C\ (N1xN2)$.

The main two broad aspects of DL methods are 1) Generative models, data generation processes capturing the underlying density of data distribution and 2) non-Generative models, that learn complex feature representations of image intending to learn the inverse mapping from k-space measurements to MR images.

3.2.1 CS to DL

In MR imaging the measurement space (k-space) as well the image domain data are complex numbers. To deal with the complex data for the neural network input, we mostly create 2-channel real tensors for the real and imaginary part respectively for both image and k-space data. However, there are methods that deal with complex data at once [49].

Problem Formulation

Let $x \in \mathbb{C}^N$ represent a 2D complex-valued MR image. The reconstruction problem aims to reconstruct x from $y \in \mathbb{C}^M, M \ll N$ undersampled k-space measurements, such that:

$$y = F_u\, x + \varepsilon \qquad (1)$$

Here $F_u \in \mathbb{C}^{MxN}$ is the undersampled Fourier encoding matrix and $\varepsilon \in \mathbb{C}^M$ is acquisition noise. In the case of cartesian acquisition we have:

$$F_u = M\, F$$

Where $F_u \in \mathbb{C}^{NxN}$ applies 2D Discrete Fourier Transform, and $M \in \mathbb{C}^{MxN}$ is the binary sampling mask selecting which lines will be sampled in k-space. The subset of sampled positions in k-space is indicated by Ω. When there is not any undersampled mask $M = N$ and the sequence is reconstructed just by applying the 2D inverse Fourier transform.

The problem formulation in eq.1 is called zero-filled reconstruction and in the current state even in the absence of noise is an ill-posed problem and does not have a unique solution. Consequently, a regularizer $R(x)$ to incorporate prior knowledge is added to reconstruct x and this can be formulated as an unconstrained optimization problem as follows:

$$\min_x R(x) + \lambda \mid y - F_u x|_2^2 \qquad (2)$$

R expresses regularization terms on x and $\lambda \in R$ is a hyperparameter that controls the properties of the reconstructed image x in accordance with the noise level in the acquired measurements y. The regularization term can be optimized using various methods, such as:

1. The Morozov formulation:

$$\min R(x) \;\; s.t. \;\; |F_u x - y| \leq \delta$$

2. The Ivanov formulation:

$$\min |F_u x - y| \;\; s.t. \;\; R(x) \leq \varepsilon$$

3. The Tikonov formulation:

$$\min |F_u x - y| \; + \; \lambda\, R(x)$$

The regularization we will be using in this work is based on the 3rd formulation of Tikonov and can be formulated as:

$$\min_{x} \; |x - f_{CNN}(x_u|\theta)|_2^2 + \lambda|F_u x - y|_2^2 \quad (3)$$

Here the f_{CNN} is the forward mapping of the CNN parametrized by θ, which takes in the $x_u$zero-filled reconstruction and directly produces a reconstruction as an output. The 1st term in eq.3 operates in the image domain and can be seen as a dealising problem. This term is conditioned by the data consistency term (2nd term) using $\Omega \; and \; \lambda$. In this sense the final output is given by:

$$x_{CNN} \; = \; f_{CNN} \; (x_u \; |\theta, \lambda, \Omega)$$

and the training is performed by providing data D containing pairs of undersampled and fully sampled (ground truth) images $D(\, x_u, x_{gnd})$. The reconstruction is obtained by minimizing the following loss:

$$L(\theta) = \sum_{(x_u, x_{gnd})} l(x_{gnd}, x_{cnn}) \quad (4)$$

Where l is the loss function. In this case we use the element-wise squared loss given by:

$$L(x_{gnd}, x_{cnn}) \; = \; \left|x_{gnd} - x_{cnn}\right|_2^2$$

Data Consistency Layer

Denote the reconstructed image by the CNN as: $s_{CNN} = F \, x_{cnn} = F \, f_{cnn}(x_u|\theta)$ where $s_{cnn}(j)$ represents an entry at index j of k-space. The non-acquired indices in k-space are filled with zeros. Therefore, for the reconstructed image regarding the latter 2 cases, sampled and not sampled k-space position respectively, we have the following formulation:

$$s_{rec(j)} = \begin{cases} s_{cnn(j)} & if \;\; j \; \notin \; \Omega \\ \dfrac{s_{cnn(j)} + \lambda \, s_0(j)}{1 + \lambda} & if \; j \; \in \; \Omega \end{cases} \quad (5)$$

And the final reconstruction is then obtained by simply applying the inverse Fourier encoding $x_{rec} = \; F^H \;\; s_{rec}$. Eq.5 states that if there does not exist any k-space value on the specific k-space position, then the only existed value is the one calculated from the CNN. On the other hand, if there is already an existed value on the k-space for the specific j position, then the result is partially accounted depending on λ but also a new value has been calculated from the CNN. In the case where $\lambda \rightarrow \infty$ we simply replace the j-th predicted coefficient in Ω by the original coefficient. That is why this operation is called *data consistency* (DC) step.

During the forward pass the DC term can be decomposed in 3 terms. Receiving the output of the CNN in the image domain a Fourier transform is applied to obtain the k-space coefficient F, then the DC term f_{DC} is applied and finally an inverse $fft\ F^H$ returns the output in the image domain. The latter procedure can be described as:

$$f_{DC}(s, s_0; \lambda) = \Lambda\, s + \frac{\lambda}{1+\lambda}\, s_0 \qquad (6)$$

Where:

$$\Lambda_{kk} = \begin{cases} 1 & if\ j \notin \Omega \\ \dfrac{1}{1+\lambda} & if\ j \in \Omega \end{cases} \qquad (7)$$

In matrix form and the forward pass including all the above operators as:

$$f_L\,(x, y;\ \lambda) = F^H\, \Lambda\, F\, x + \frac{\lambda}{1+\lambda}\, F_u^H\, y \quad (8)$$

Finally, the backward pass is obtained through the Jacobian of the DC layer with respect to the layer input x as follows *[4]*:

$$\frac{\partial f_L}{\partial x^T} = F^H\, \Lambda\, F \qquad (9)$$

3.3 Algorithms for image reconstruction

The image reconstruction on accelerated MRI has been playing the most prominent role on the medical imaging field. The major objective is to acquire high quality images at the minimal cost and patient risk. The patient risk is eliminated by following specific imaging protocols during examinations. On the other hand, the image reconstruction can be formulated in many different optimization settings which is based on mathematical models, but further prior information can be incorporated to easier guide the convergence of the reconstruction. Following the eq.2, for the regularization term different formulations can be used such as 2D wavelet [56], total variation (TV) [57], dictionary [58] etc. Later with the introduction of CS to incorporate the sparsity constrain, pre-constructed basis [14-18], adaptive basis [19-21] or dictionaries [58, 71, 72]. In fig. 3.5 different optimization formulations from several research works are presented:

Smith 2011, Magn. Img Intensity

$$\arg\min_u \|Su\|_1 + \frac{\lambda}{2}\|Fu - d\|_2^2,$$

Madellin 2012. DCT or TV

$$f(x) = \|\mathscr{F}x - y\|_2^2 + \lambda_1\|\Psi x\|_1 + \lambda_2\,\mathrm{TV}(x).$$

Parasoglou 2012, PCA img

$$\min\|F\cdot m - y\|_2^2 + \lambda\|T\cdot m\|_1$$

Rapachi 2014, TV

$$I_{kd} = \mathrm{argmin}\ \ U_{kd}\mathcal{F}(I_{kd}) - (K_2 - K_1)_2^2 + \lambda TV(I_{kd}) + \mu|I_{kd}|_1$$

Geenethan 2012, Daubechies Wavelet

$$\epsilon(m) = \| F_u m - y \|_2 + \lambda_{L1} \| Wm \|_1 + \lambda_{TV}\,\mathrm{TV}\,(m)$$

Ravishankar 2011, Dictionary Learing

$$\min_{x,D,\Gamma} \sum_{i,j} \|R_{ij}x - D\alpha_{ij}\|_2^2 + \nu\|F_u x - y\|_2^2 \ \text{s.t.}\ \|\alpha_{ij}\|_0 \leqslant T_0\ \forall i,j$$

Akcakaya 2011, Voxel Intensity

$$\min\|\Psi \mathbf{m_j}\|_p \ \text{s.t.}\ \|\mathbf{S_j} - F_\Omega(\mathbf{m_j})\|_2 \le \varepsilon,$$

Xiabo 2012, PBDW

$$\hat{x} = \arg\min_x \|A_w x\|_{l_1} + \frac{\lambda}{2}\|y - F_U x\|_2^2$$

Wang 2010, TV or Wavelet

$$\text{minimize}\ \|\Psi f\|_1,\ \text{subject to}\ \|F_u f - g_u\|_2 \le \varepsilon,\ \ g_u \in U,$$

Fig.3.5: Different mathematical formulations from iterative reconstruction algorithms based on equation 1.

Practical and effective optimization algorithms are essential, and a large number have been studied to solve various optimization problems. The iterative reconstruction algorithms and the DL networks still have a close connection. The most common algorithms used for solving the latter problems are:

a. Variable splitting with quadratic penalty (VSQP) [66]
b. Proximal gradient descent (PGD) [67]
c. Iterative shrinkage thresholding algorithm (ISTA) [68]
d. Alternate direction method of multipliers (ADMM) [69]

3.4 Cascaded Architecture

For the CS-based methods based on DL the optimisation step in eq.3 is solved using a coordinate-descent type algorithm, iterating between the two terms, the dealising and the data consistency term until convergence. Using just one CNN for dealising and thus one step reconstruction in the absence of vast amounts of training data, is often shows signs of overfitting. The simplest solution is to use a second CNN attached to the output of the first CNN and data consistency step. Following that methodology, very deep architectures can be built to deal with the dealising -reconstruction problem. We term this a cascading network. The main two parameters in a cascading architecture are:

n_d: Number of Consecutive CNN layers followed by a non-linear function – rectified linear unit (ReLU).
n_c: Number of Cascades. Each cascade is characterized by the latter block (number of consecutive CNNs each followed by a ReLU) plus the DC term.

So, for example a Cascaded Architecture of $n_d, n_c = 3, 2$ stands for the following:

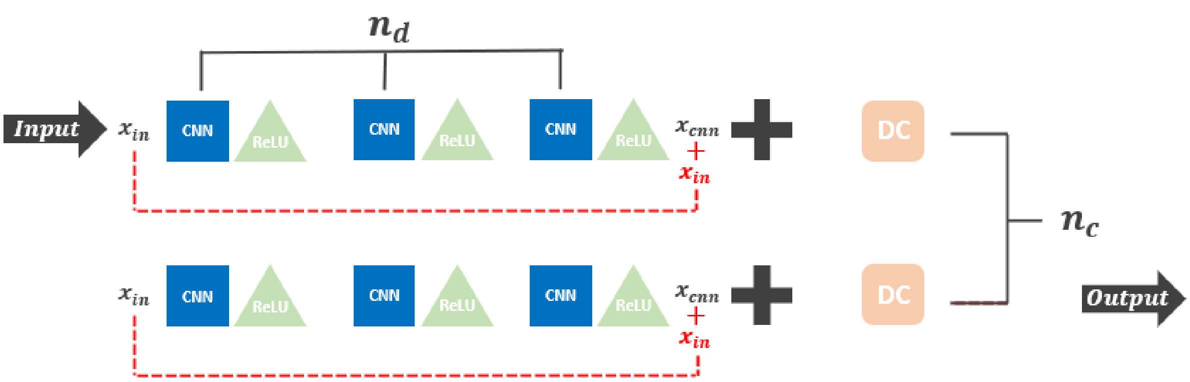


Fig.3.6: Example of a Cascaded Architecture with $n_d, n_c = 3, 2$. The red dashed line represents the residual connection that is used. x_{in}, x_{cnn} the network input and the modified input after the application of the CNNs respectively.

To avoid the main drawback of the CNN networks when the architecture becomes very deep, the vanishing gradient issue, a residual connection* is used.

**Residual Connection:* *The residual connection applies identity mapping of the input just before the output of the series of the applied CNNs. In other words, it performs an element-wise addition of the input with the output after a series of convolutions* $\boldsymbol{x} + \boldsymbol{F_{cnn}}(\boldsymbol{x})$ *prior to the activation function.*

3.5 Regularization Learning

In the previous paragraph we presented the crucial role of the DC layer in the reconstruction performance. In contrast to the convolutional block and the residual connection, which operate in the image domain (mainly) to extract features from the reference image, the DC layer operates in the frequency domain and aims to preserve the accuracy of the learning process through the sampling domain. [4], [50].
After the convolutional layer a Fourier transform is applied to obtain the k-space data, in the frequency domain. As we described in formulation from eq.7; in the case of undersampled data there are two possibilities for the k-space locations. Either the initial data point, s_j, is sampled ($j \in \Omega$) or not ($j \notin \Omega$). In case that a missing data point is reconstructed then the only value existed for the exact location is the output of the CNN-Network. On the other hand, when an existed, initially sampled data point is reconstructed, there arise two values for the same position. The already existed-sampled value, and the one returned from the CNN-Network. However, as we can see in Eq.7 this is not always the case. The latter statement of totally discarding the newly created point is only accurate when we are certain for the total absence of noise or artifacts during sampling. Although, this is an ideal case.
Nevertheless, in real applications and especially in the undersampled cases there is always noise in our measurements. In this case, we will use a sort of convention between the two values and depending on our knowledge about the data (noise existence), we should suitably adjust the weights of them. Here comes the regularizer, λ, which adjusts the interaction of the two terms for the most optimal output. Of course, λ, is highly data- dependent and thus its accurate estimation could formulate an empirical noise model for the current dataset. In such a case, we need to consider the dealising step of the convolutional block in our DC step.

3.5.1 Fixed Regularization Parameter

This hyperparameter λ in the latter formulation (Eq.10) is fixed and predefined in an empirical value, for the whole reconstruction cycle.

$$\left[\frac{\partial f_{dc}(s_{CNN}, s_0; \lambda)}{\partial \lambda}\right]_j = \begin{cases} 0 & if\ j \notin \Omega \\ \dfrac{s_0(j) - s_{cnn}(j)}{(1+\lambda)^2} & if\ j \in \Omega \end{cases} \qquad (10)$$

3.5.2 Learnable Regularization Parameter

At this point we consider the modification of λ, towards a learning scenario. Instead of a predefined λ, we make this parameter trainable. In this way the optimal value is tuned during the reconstruction depending of course on the current dataset. We include λ in the forward and backward pass and in such a way the weights in the Jacobian are adjusted including the derivative of λ.

The eq.10 when used in the network is written in the following format:

$$\boldsymbol{Noiseless\ Case} : s_{rec} = (1 - F_{\Omega}) * s_{CNN} + F_{\Omega} * s_0$$

$$\boldsymbol{Noisy\ Case}: s_{rec} = \frac{(1 - F_{\Omega}) * s_{CNN} + F_{\Omega}\ (s_{CNN} + \lambda\ * s_0)}{1 + \lambda}$$

Slice Learnable Parameter

In this first step of the work, for improved reconstruction, starting from the $j \in \Omega$ case (Noisy Case) of $eq.10$, we convert λ to a learnable network parameter and we calculated the s_{rec} for each k-space frame individually as follows:

$$s_{rec} = \frac{(1 - F_{\Omega}) * s_{CNN} + F_{\Omega}\ (s_{CNN} + \lambda * s_0)}{1 + \lambda}$$

$$s_{rec} = \frac{(1 - F_{\Omega}) * s_{CNN} + F_{\Omega} * s_{CNN} + F_{\Omega} * \lambda * s_0)}{1 + \lambda}$$

$$Instead\ of\ a\ fixed\ \lambda\ value\ we\ use\ a\ learnable\ \lambda: \qquad \boldsymbol{\lambda} \leftrightarrow \boldsymbol{\lambda_{DC}}$$

$$f_{DC}(s_0, s_{CNN}; [\boldsymbol{\lambda_{DC}}]) = s_{rec} = \frac{(1 - F_{\Omega}) * s_{cnn}\ + F_{\Omega} * s_{cnn} + \lambda_{DC} *\ F_{\Omega} * s_0}{1 + \lambda_{DC}} \quad (11)$$

To clarify the conversion to a learnable parameter, we mean that instead of using a fixed integer value for the whole reconstruction, an adjustable through the iterations λ_{DC} is used. The $nn.Parameter$ function from the $Pytorch$ library was used. In that way the λ_{DC} is included in the backpropagation step [73-75].

Spatially Learnable Parameter

Proven the highly optimized results from the learnable λ we moved a step forward in a spatially varying learning formulation. Motivated by the characteristic properties of the frequency space, and more specifically the different role of the low (main object shape) and high frequencies (edges, resolution) in the reconstruction process, as well as the different amount of sampling needed in the two different neighbourhoods respectively, we were interested in a spatially varying learnable parameter set up. Starting again from Eq.10 and separating the λ parameter in 2 terms, for the CNN and the initial k-space output we have the following equation:

$$f_{DC}(s_0, s_{CNN}; [\boldsymbol{\lambda_{CNN}}, \boldsymbol{\lambda_{DC}}]) = s_{rec} = (1 - \lambda_{CNN}\ * F_{\Omega}) * s_{cnn}\ + (F_{\Omega} * \lambda_{DC}) * s_0 \ (12)$$

In the above formulation the distinctive contribution of the different spatial frequencies in k-space was revealed.

The λ_{DC} and λ_{CNN} parameters were again converted through the $nn.Parameter$ function and included in the backpropagation step.

Where:

s_{rec}	Final k-space Output
s_{CNN}	k-space from CNN Output
s_0	Initial k-space
F_{Ω}	Mask & FFT
$1 - F_{\Omega}$	Remaining part without the F_{Ω}
λ_{DC}	Regularization for the DC term
λ_{CNN}	Regularization for the CNN output

3.6 Channel Separated Convolutions

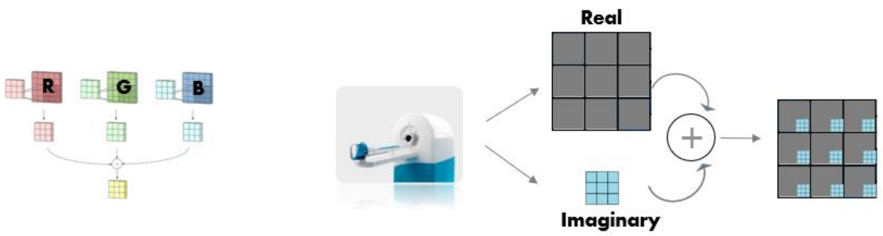


Fig.3.7: A. CNN operation on 3 Channel colour input (RGB). B. Complex MRI analogy – The two parts are intentionally presented with such a big size difference to denote the order of magnitude difference in real sampled data.

Regarding the CNNs', several settings can differentiate each CNN unit, such as the kernel size, the number of applied filters, the padding, or the stride etc. Another important aspect of the CNNs, is the number of channels, especially in the MRI case where the input is complex. Of course, all the data inputs are provided using real numbers for both channels (real and imaginary) for both image and k-space entries.

However, the way the data are provided to the network leads in different results for mainly two reasons. First, during the CNN operation, the different channels interact to provide a final single channel output (fig.3.6). This is important here because the 2 channels in our case contain frequency and phase information (complex data) rather than just different channels of colour.
Second, when dealing with real and not synthesized MRI data, the Real and the Imaginary channel of our measurements have several orders of magnitude difference. In such a case when we provide the data as 2-channel input, with the imaginary part being several orders of magnitude smaller than the Real channel, the imaginary part is finally neglected. The different channels in MRI data contain useful information

about the nature of the acquisition and both are of vital importance for a successful reconstruction. Thus, in the current work presented in the next chapter, all the convolutions were performed separately per channel and the convolved outputs were concatenated prior to the DC block.

Chapter 4

4. Materials & Methods

In the current chapter all the used pillars for the current reconstruction are presented in a Materials & Methods way.

4.1 Dataset

The Dataset used for the reconstruction algorithms presented in this work is the Calgary Campinas Public Dataset [53]. It is composed of healthy T1-weighted MR brain images acquired with a General Electric – Discovery MR 750 scanner. The dataset consists of both Single and Multi-Coil Data but for the current work only the Single-Coil Data were used. 45 fully sampled acquisitions are provided. For train, validation and test we divided the data in 25, 10 and 10 subjects respectively. However, in the total amount of 4.524, 1.700 and 1.700 images for train, validation, and test respectively, the SNR was measured, and part of each dataset was discarded

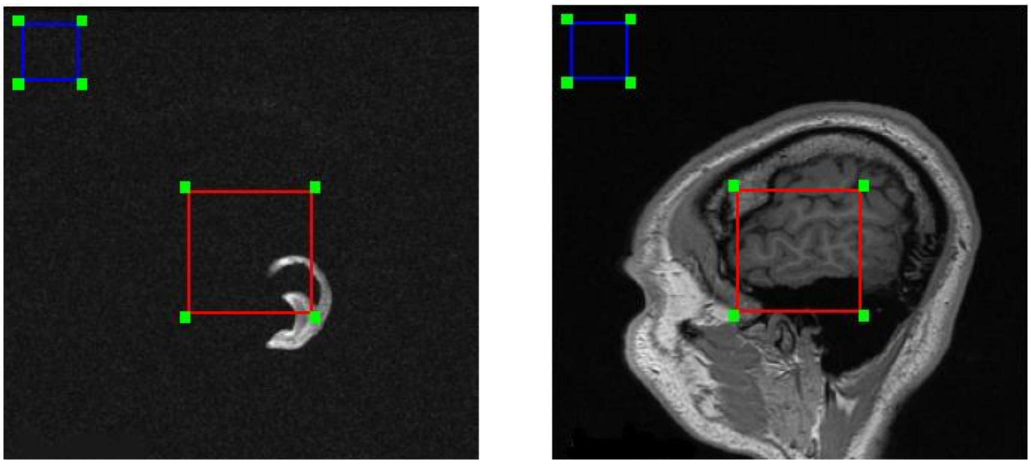


Fig.4.1: ROIs for SNR measurements. The red window measures the Signal while the blue one the Noise. The image on the left- hand side is an example of a discarded image from the initial dataset which has an SNR<20.

when the SNR was below 20 for better reconstruction results. This method led to a total amount of 3724, 1527 and 1553 for train, validation, and test respectively. The SNR was measured using two Regions of Interest (ROI) as it is shown in the following figure. The blue one represents the noisy counterpart while the red one the area with the signal.

The final SNR was calculated using the following formula:

$$SNR = \frac{mean(Signal)}{std(Noise)}$$

4.2 Convolutional Schemes

For the current work several architectures for the convolutional block were used. The main idea was to build and test all the most common network topologies used. A quick overview of all the created architectures will follow.

General Network Characteristics

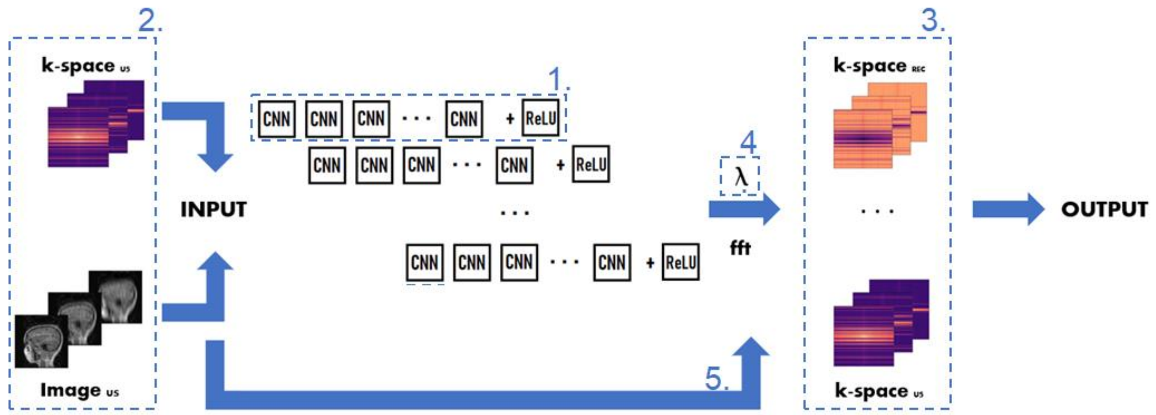


Fig.4.2: Main building blocks of the built Networks. 1.Cascade of CNN's, 2. Dual Domain Setting, 3. Data Consistency, 4. λ Regularization Learning, 5. Residual Connection.

In Fig. 4.2 we can observe the main pillars existing in all the following used architectures.

1. CNN Block (Cascade of CNN's): A chain of CNNs followed of a non-Linear function is called a cascade of CNNs. In each cascade we can select the number of the latter scheme (CNN + Non-Linearity)

2. *Dual Domain Setting*: All the used architectures have been built to be able to be used for both image and k-space data as an input to the CNN-Block.

3. *Data Consistency Block:* The CNN block is always followed by a DC Block which compares the converted to k-space reconstructed output of the CNN Block with the initial existed-sampled values.

4. *λ Regularization Learning*: All the networks consisted of an adjustable λ regularization parameter in three different settings (FF-fixed value, TF-learnable, TT-spatially learnable).
5. Residual Connection: To avoid the vanishing gradient problem of this kind of architectures (deep sequential layers of CNNs) a residual connection is used. The initial input to the CNN is added to the convolved output prior to the DC-block.

In the following paragraphs the main architectures used are briefly presented:

Cascades of CNNs

Our initial architecture borrowed from Schlempler et al. [4] with several sequential CNNs fig4.2, a residual connection prior to the ReLU function and a DC term. The kernel size can be adjusted but is conserved through the different layers. The number of CNNs as well as the number of cascades is also an option. The undersampled image is provided as a Single-Two Channel Input (separate convolution for each channel and concatenation prior to the DC term).

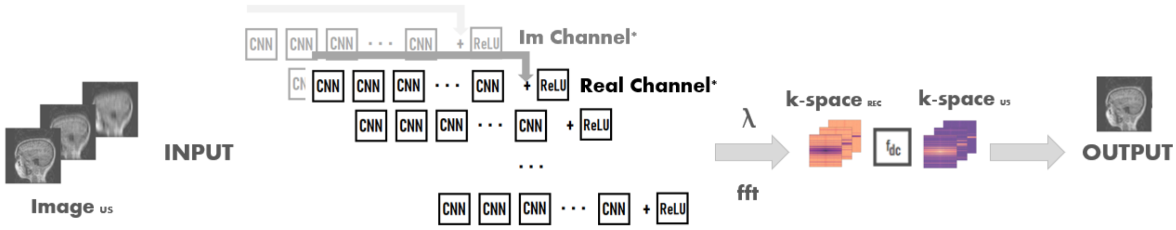


Fig.4.3: Cascade of CNNs

Cascades of Increasing Filters & Concatenation Block

The 2nd architecture we implemented consists of increasing filter sizes per layer, 3x3, 5x5, and 7x7, followed by a concatenation block of all the previous outputs. The concatenate block at the end prior to the DC term is mimicking a fully connected layer. In this format we moderate the computational burden of the model while increasing the field of view using this technique. The undersampled image is provided as a Single-Two Channel Input (separate convolution for each channel and concatenation prior to the DC term).

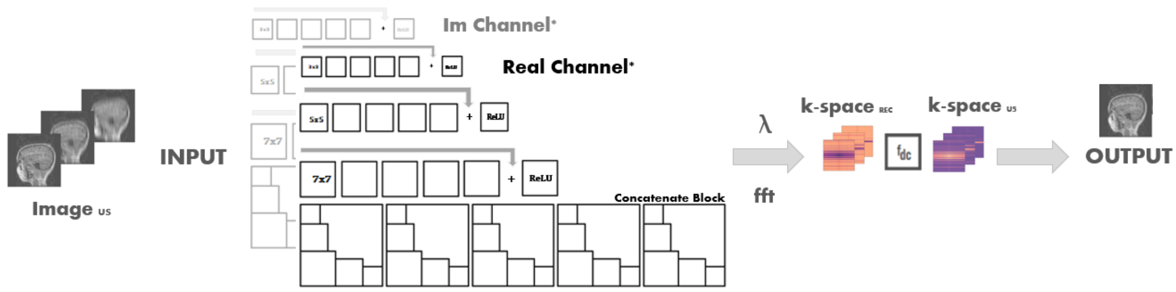


Fig.4.4: Cascade of Increasing Filters & Concatenation Block

Cascades of U-Net like shape

The last architecture, which was used for all the results presented in the next chapters is mimicking in a way the U-Net architecture just from the point of increasing-decreasing number of filters but with the same kernel size in all layers as well as the size of the outputs per layer. In all the experiments the set of applied filters was 8-16-32-32-16-8 followed by a ReLU and a DC term respectively. The option of this architecture is related with the memory and can be adjusted to 16-32…. Or 32-64… etc. The undersampled image is again provided as a Single-Two Channel Input (separate convolution for each channel and concatenation prior to the DC term).

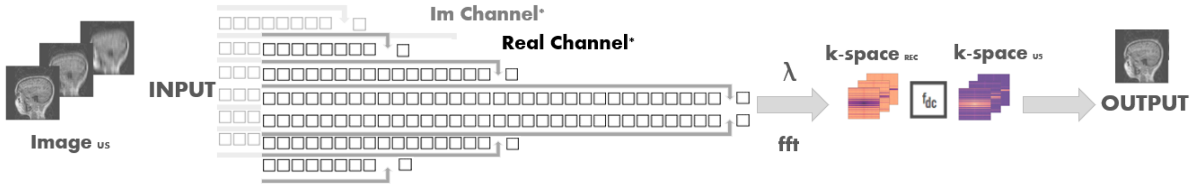


Fig.4.5: Cascades of U-Net like shape

Important is here to mention that all the DC blocks were applied after each non-linearity function (per-layer). For the sake of illustration, the DC block is presented just once per Network.

4.3 Evaluation Metrics

For the Loss function the mean square error (MSE) was used in the image domain, comparing the reconstructed image with the reference image.

$$MSE = |x_{GND} - x_{REC}|^2$$

The quantitative metrics used for validation of the reconstruction quality was:

a. Peak Signal to Noise Ratio:

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

Where MAX_I is the maximum possible pixel value of the image

b. Structural similarity index (SSIM):

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

μ_x: *the pixel sample mean of* x
μ_y: *the pixel sample mean of* y
σ_x^2: *the variance of* x
σ_y^2: *the variance of* y

σ_{xy}: $the\ covariance\ of\ x, y$
$c_1 = (k_1 L)^2, c_2 = (k_2 L)^2$ $two\ variables\ to\ stabilize\ the\ division$
L: $the\ dynamic\ range\ of\ the\ pixel - values$
$k_1 = 0.01\ \&\ k_2 = 0.03$

The SSIM function that used was from the [52].

4.4 The undersampling schemes

For all the presented work data from Cartesian acquisitions were used. Thus, the undersampling schemes are based on randomly excluding lines from the k-space based on a Poisson distribution as in [53]. The PSNR/SNR from undersampled data is highly dependent on the imaging data and the undersampling mask. For a fair comparison we aligned an arbitrary but fixed undersampling mask for all the used datasets (train-validation-test).
The acceleration rates applied was 4.0, 8.0 and 20.0 corresponds to undersampling of 0.25, 0.125 and 0.05 (of the initial data) respectively ($acc = \frac{1}{undersampling}$). The undersampling masks are presented in the following figure.

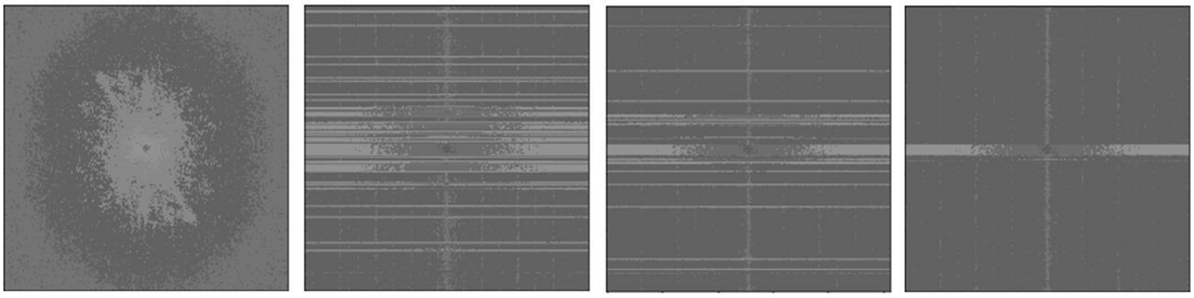


Fig.4.6: From right to left hand side: Fully Sampled k-space, 4 acc., 8 acc., and 20 acc.

4.5 Network Notation

The main architecture used in all the performed reconstructions is based on CS-DL reconstruction architecture. The built of the architecture is based on the two main block formulation of iterative optimization under the CS framework consisted of a convolutional block (CNN_{bl}) followed by the data consistency (DC) block. The notation for this type of networks is mainly regards the formation of the CNN_{bl}; $\boldsymbol{n_d}$: the number of sequential CNNs' followed by a non-linear function (ReLU) and $\boldsymbol{n_c}$: the number of cascades- the repetitions of the latter block.
For the simple examples in the simplest architecture of the "Cascade of CNN'', the notation D5C2 is a network of 5 consecutive CNNs& ReLU per Cascade and in total of two such cascades.
In the rest of the results where the "Cascades of U-Net" architecture is used the 8-16-32 notation is used to denote the 8-16-32-32-16-8 number of consecutive CNNs& ReLU. The meaning of Cascade in this case is more abstract.

Chapter 5

5. Results & Future Research

5.1 2 vs 1 Channel Convolution Results

In this first section we would like to present the results regarding the different weight matrices produced when the convolution is performed using a 2 single-channel input vs a two-channel input data. For these experiments a D4C2 architecture of the simple "Cascade of CNNs" was used with 16 filters per layer. In the following figure the characteristics of the network as well as the dimensions in each case are presented.

	Module List					
		INPUT	OUTPUT	KERNEL	STRIDE	PADDING
0	Conv2D	2	16	(3,3)	(1,1)	(1,1)
1	Leaky Relu					
2	Conv2D	2	16	(3,3)	(1,1)	(1,1)
3	Leaky Relu					
4	Conv2D	2	16	(3,3)	(1,1)	(1,1)
5	Leaky Relu					
6	Conv2D	2	16	(3,3)	(1,1)	(1,1)
7	Data Consistency					
8	Conv2D	2	16	(3,3)	(1,1)	(1,1)
9	Leaky Relu					
10	Conv2D	2	16	(3,3)	(1,1)	(1,1)
11	Leaky Relu					
12	Conv2D	2	16	(3,3)	(1,1)	(1,1)
13	Leaky Relu					
14	Conv2D	2	16	(3,3)	(1,1)	(1,1)
15	Data Consistency					

	2 Channels
	CNN Block Input:
0	[2,16,256,256]
1	[2,16,256,256]
2	[2,16,256,256]
3	[2,16,256,256]
4	[2,16,256,256]
5	[2,16,256,256]
6	[2,2,256,256]
7	[2,2,256,256]
8	[2,16,256,256]
9	[2,16,256,256]
10	[2,16,256,256]
11	[2,16,256,256]
12	[2,16,256,256]
13	[2,16,256,256]
14	[2,2,256,256]
15	[2,2,256,256]

	1 Channels	
	CNN Block Input:	
0	[2,16,256,256]	[2,16,256,256]
1	[2,16,256,256]	[2,16,256,256]
2	[2,16,256,256]	[2,16,256,256]
3	[2,16,256,256]	[2,16,256,256]
4	[2,16,256,256]	[2,16,256,256]
5	[2,16,256,256]	[2,16,256,256]
6	[2,1,256,256]	[2,1,256,256]
7	[2,2,256,256]	
8	[2,16,256,256]	[2,16,256,256]
9	[2,16,256,256]	[2,16,256,256]
10	[2,16,256,256]	[2,16,256,256]
11	[2,16,256,256]	[2,16,256,256]
12	[2,16,256,256]	[2,16,256,256]
13	[2,16,256,256]	[2,16,256,256]
14	[2,1,256,256]	[2,1,256,256]
15	[2,2,256,256]	

Fig.5.1: The Network Characteristics and the dimensions of the layers in each case. The 2-Single Channel CNN block can be seen as a twice version of the 1-Dual Channel CNN block.

In the following figure a schematic representation of the filter weights is shown for the above case. Only the Convolutional layers are presented for all the layers (not the biased or the Leaky ReLU layer)

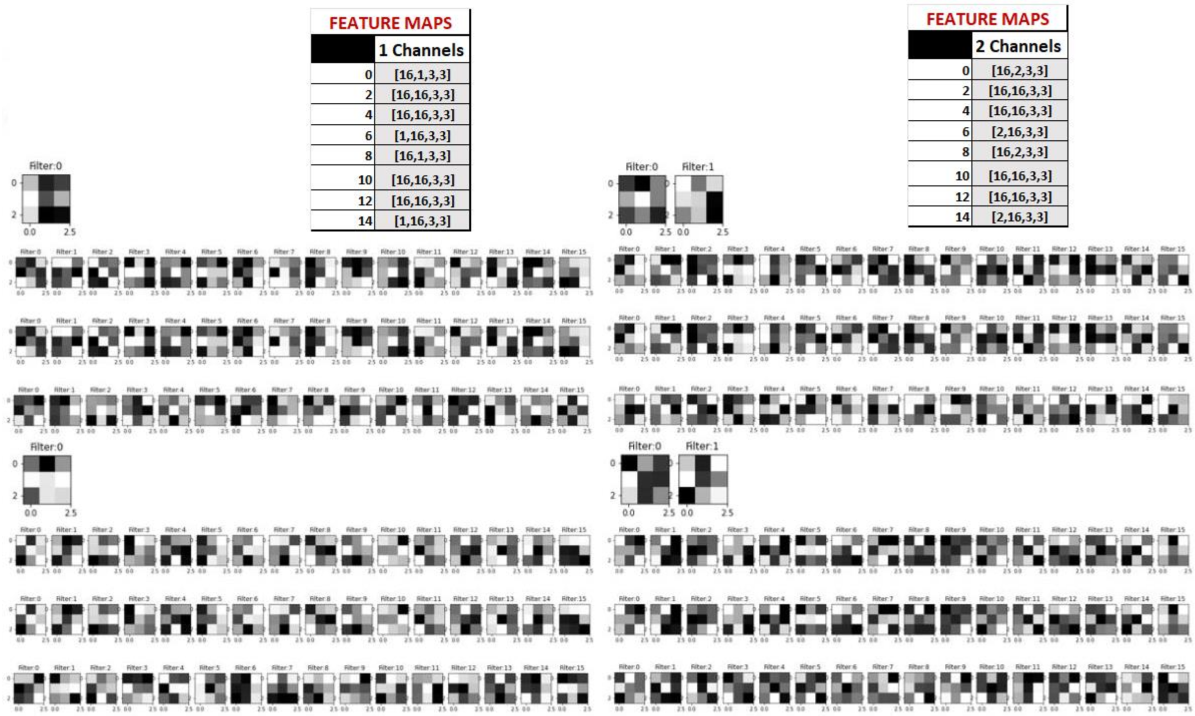


Fig.5.2: The two convolution cases 2 Single-Channel and 1 Dual-Channel respectively. Only the CNN layer weights are shown in the above figure.

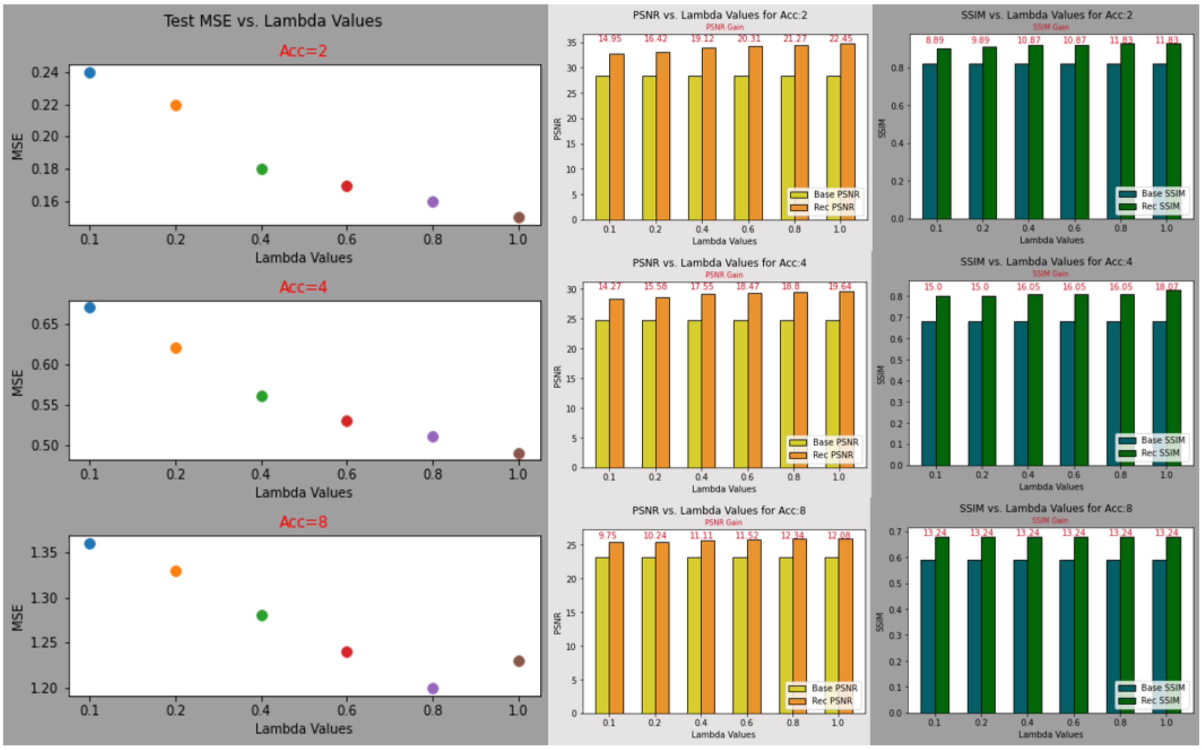
The difference is highly profound in the above results. Here we would like to pinpoint the importance of using the 2 Single-Channel convolutions. The most important argument for using this format is mainly due to the nature of the involved data. In MRI acquisitions both the Real and the Imaginary part are of high importance in real sampled data. As we have noticed in several acquisitions we have performed, the real and the imaginary absolute values are characterized by several orders of magnitude difference. In the way the CNNs are operating in the Single Dual-Channel input leads in almost absolute ignorance of the imaginary values in the resulted convolved features. The higher is the absolute value the more important is a feature for the reconstruction proceeding. In the following presented results both cases are used. However, in the last part of experiments the 2 Single-Channel input is used for optimal reconstruction. To the best of our knowledge this is the first time such an argument is introduced at least in the specific domain.

5.2 Learnable Regularization Parameter Results

Fixed Regularization Parameter

To prove the need of learnable regularization parameters, we first show that different fixed $\lambda \in [0,1]$ values lead in different reconstruction results. Thus, there will be a specific λ value that leads in the most improved reconstruction results.

Using the simplest 'Cascade of CNNs'-D2C5 architecture***** with a fixed λ regularization parameter, we performed reconstructions for three different acceleration rates (4.0,



8.0, 20.0), using five different λ values (0.1, 0.2, 0.4, 0.6, 0.8, 1.0). The MSE and image evaluation metrics (PSNR and SSIM) were measured. For ease of comparison the performance gain between the undersampled, and the reconstructed images was included.

***D2C5**:
Network Params.: Epochs:30-60-100, Batch Size:16, Learning Rate:0.001, Optimizer: Adam
CNN: Channels:2, Filters:32, Kernel Size:3x3, Activation: Leaky ReLU

From the above results there are several important findings to be discussed.

MSE

As we can observe from the 1st column of fig.3 the MSE decreased inversely proportional to the λ value almost for all the accelerations except for acc.=8. For acc.=8 we noticed that the λ = 0.8 leads in a smaller MSE than the λ=1.0. This outcome suggests that there is not always a specific (expected) trend as the previous accelerations suggest and that the prominent λ value is highly case specific.

PSNR

The PSNR values follow the same trend as the MSE results even for the acc.=20 for which the value λ=0.8 leads in more optimised PSNR.

SSIM

For the SSIM the results for acc.=2, seems to follow the inversed proportional trend between the λ values and the performance gain. For acc.=4, there is some similar behaviour but only for the edge values. The λ=0.4, 0.6, 0.8 value lead in the same performance. Finally, for acc.=8 for all the λ values the performance gain remain constant. The last results suggest that for such high accelerations the number of

Fig. 5.3: *Columns A: MSE, B: PSNR (& PSNR Gain), C: SSIM (& SSIM gain) vs. Lambda Values for 3 different acceleration rates (2,4,8). The MSE error is decreased inversely proportional to Lambda value.*

epochs used is very limited to lead in more differentiable results just for a slightly change in λ.

The latter results underline the importance of approaching the optimal λ value for each reconstruction case. At this point, we would like to pinpoint that the role of finding the optimal λ value is of secondary importance for the final reconstruction results by itself. Having said that and to avoid any misconception, the right architecture – network depth as well as all the tuning parameters (number of epochs, batch size, etc.) is of course of primary importance and the most crucial part of the reconstruction algorithm. However, our proposed results suggest that the optimal λ estimation highly improves the reconstruction results for a given architecture and thus a more solid strategy for approaching the optimal λ value is needed.

Learnable Regularization Parameter (Slice)

The next step was to convert the regularization parameter λ into a trainable network parameter as follows:

$$\boldsymbol{s_{rec}} = \frac{(\mathbf{1} - \boldsymbol{F_\Omega}) * \boldsymbol{s_{CNN}} + \boldsymbol{F_\Omega} * \boldsymbol{s_{CNN}} + \boldsymbol{\lambda} * \boldsymbol{F_\Omega} * \boldsymbol{s_0}}{\mathbf{1} + \boldsymbol{\lambda}}$$

Where $\boldsymbol{s_{CNN}(j)}$ represents an entry at index j in k-space and $\boldsymbol{s_0}$ the initial acquired k-space data using zero filling. In this formulation the λ parameter was learned-estimated for each k-space frame.
We used a fixed λ value $\lambda = 10$ and we performed training for 3 different accelerations (acc.=2.0, 4.0, 8.0) for a fixed number of epochs. The architecture used was the $\boldsymbol{D2C5}$. In the next table we can observe the MSE and image evaluation metrics (PSNR, SSIM) for all the above combination of experiments. The performance gain was used again for the image evaluation metrics.

Lambda 10	FIXED									TRAINABLE								
Acceleration	2			4			8			2			4			8		
40 EPOCHS																		
MSE	0.20			0.43			1.03			0.23			0.37			0.95		
	BASE	REC.	GAIN%	BASE	REC.	GAIN%	BASE	REC.	GAIN%	BASE	REC.	GAIN%	BASE	REC.	GAIN%	BASE	REC.	GAIN%
PSNR	28.45	36.41	27.99	24.8	30.31	22.22	23.13	26.69	15.39	28.38	32.95	16.11	24.83	31.06	25.07	23.08	26.70	15.70
SSIM	0.82	0.94	12.77	0.69	0.84	17.86	0.64	0.76	15.79	0.82	0.90	8.89	0.69	0.86	19.77	0.59	0.71	16.90
60 EPOCHS																		
MSE	0.10			0.43			1.43			0.15			0.39			1.00		
	BASE	REC.	GAIN%	BASE	REC.	GAIN%	BASE	REC.	GAIN%	BASE	REC.	GAIN%	BASE	REC.	GAIN%	BASE	REC.	GAIN%
PSNR	28.45	36.36	27.83	24.83	30.34	22.22	24.83	30.34	22.22	28.36	32.73	15.42	24.84	30.77	23.90	23.12	32.90	29.72
SSIM	0.84	0.95	11.58	0.68	0.84	19.05	0.68	0.84	19.05	0.82	0.93	11.83	0.68	0.88	22.72	0.62	0.78	20.51
100 EPOCHS																		
MSE	0.09			0.41			1.22			0.17			0.39			1.13		
	BASE	REC.	GAIN%	BASE	REC.	GAIN%	BASE	REC.	GAIN%	BASE	REC.	GAIN%	BASE	REC.	GAIN%	BASE	REC.	GAIN%
PSNR	28.42	36.77	29.41	24.81	30.55	23.13	23.12	26.71	15.53	28.41	32.42	14.08	24.86	30.73	23.63	23.11	29.27	21.04
SSIM	0.83	0.95	12.63	0.71	0.88	19.32	0.64	0.76	15.79	0.82	0.91	9.89	0.66	0.84	21.43	0.62	0.78	20.51

Table 5.1: *All the Results (MSE, PSNR, SSIM) for the Fixed and Trainable lambda, for the 3 different Acceleration Rates (2,4,8) are presented briefly. With Green colour are highlighted the better performed lambda (fixed or trainable) respectively.*

As the results suggest for small acceleration, acc.=2 the fixed λ seems to perform slightly better, but not for higher acceleration rates (4, 8). Furthermore, as the number of epochs increases the trainable λ led to better performance which was expected (longer training). Important is here to mention that our results are of the same order of magnitude with similar state of the art architectures [4], [53].

Spatially Learnable Regularization Parameter

The last step in our research was to divide the regularization parameter in two parts and learn it independently for the two building blocks of our CNN architecture. The $\boldsymbol{\lambda_{DC}}$ is the regularization parameter regarding the initial k-space sampled locations on which the DC block operates, and the $\boldsymbol{\lambda_{CNN}}$ which is the regularization parameter for the reconstructed locations of the k-space from the CNN block fig.4.

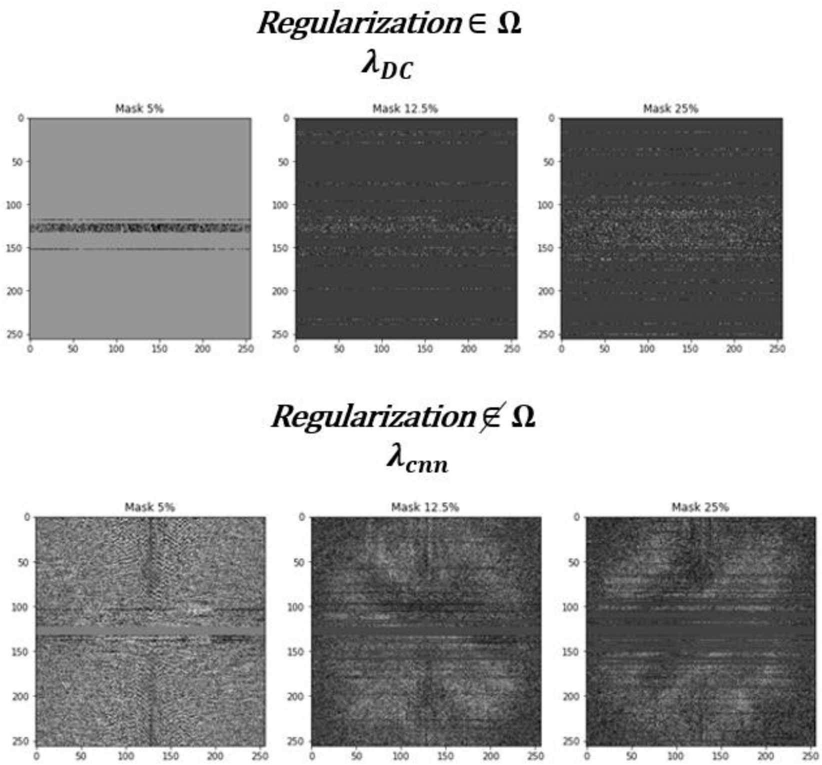


Fig. 5.4: The different regularization parameters for the initially sampled k-space points $\boldsymbol{\lambda_{DC}}$ and the reconstructed k-space points from the CNN block $\boldsymbol{\lambda_{cnn}}$

The new formulation for the spatially learnable regularization parameter for the DC block is stated as follows:

$$\boldsymbol{s_{rec}} = (\mathbf{1} - \boldsymbol{\lambda_{cnn}} * \boldsymbol{F_{\Omega}}) * \boldsymbol{s_{CNN}} + (\boldsymbol{F_{\Omega}} * \boldsymbol{s_0}) * \boldsymbol{\lambda_{DC}}$$

On the following figures the reconstruction results are presented for all the three different regularization cases: a) The fixed λ, b) The trainable λ (per slice), and c) The spatially trainable λ denoted by FF, TF, and TT respectively.

'Cascades of U-Net like shape'

In the following experiments we used the "Cascade of U-Net like shape" architecture with all the regularization settings to test the performance of the regularization settings in another architecture for the CNN block.

The above network architecture regarding the CNN block followed the setting for the low-memory with 8-16-32-32-16-8 sequential CNNs respectively. For the DC block all the different settings were used, and the results are presented in the following figures:

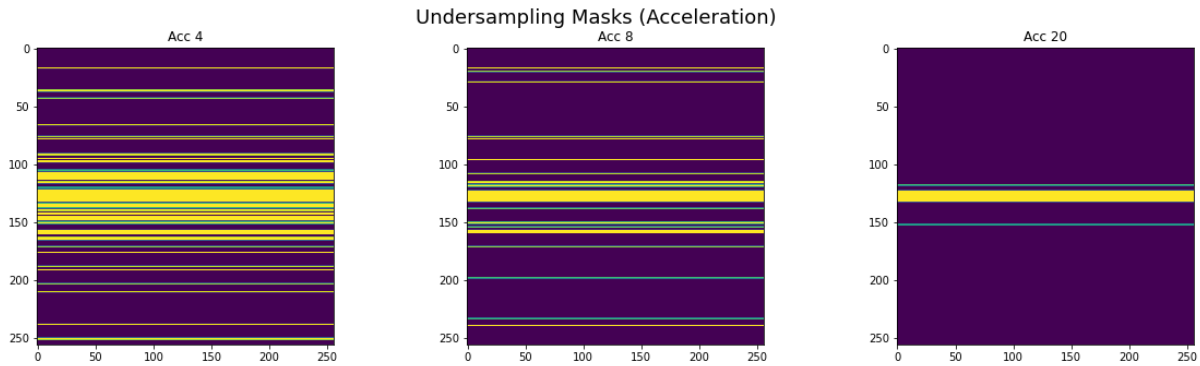


Fig.5.5: The undersampling masks for acc.=4, 8 and 20 respectively.

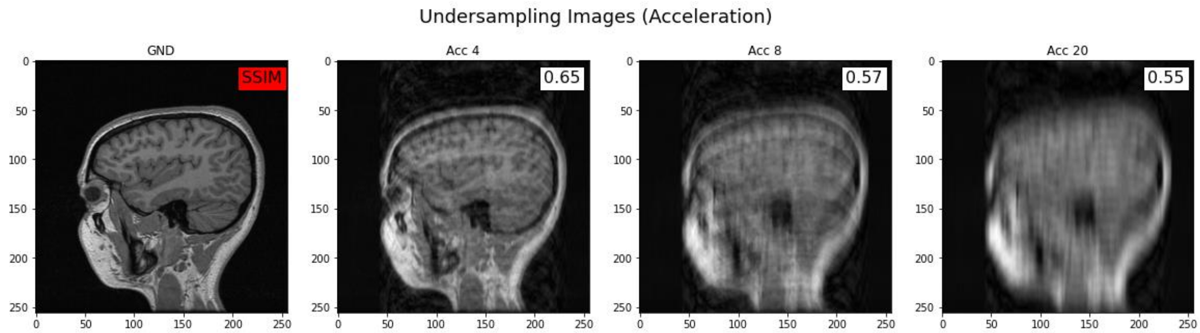


Fig. 5.6: The Ground Truth and the undersampled images for acc.=4, 8 and 20 respectively including the SSIM index in the upper right corner.

In fig. 5.6 the k-space undersampling masks that used for the following experiments are depicted as well as the undersampled images for the latter accelerations can be seen in fig. 5.7. For all the undersampled images the SSIM index between the undersampled and the ground truth image is denoted at the upper right corner. As we can observe even for acc. 8.0 the reconstruction is really challenging. In the following experiments we used the acc. 20.0 just to prove that the current network setting is able to learn even in such extreme cases.

In fig. 5.8 the MSE vs the number of epochs is summarized for all the different λ regularization settings and for all the accelerations. Green, blue, and red were used for the FF, TF, and TT setting respectively, while the solid and dashed lines used to differentiate between the training and validation results. Finally, the diamond sign was used for the mean test error in each case.

Independently of the regularization setting all methods showed that the current architecture can learn the hidden features and reduce the reconstruction error (MSE). Also, all the different methods produced different results which highlights the adding value of each λ regularization setting.

As we can observe the TT setting outperforms all the other settings in all the above cases. Again, in this network setting for the lowest acceleration (acc. 4.0) the FF case outperforms the TF as in the previous results. For the medium acceleration (acc. 8.0) the FF and TF showed significant results but a mild better performance on the test case for the TF setting. In the highest acceleration (acc. 8.0) the TF performance was better than the FF.

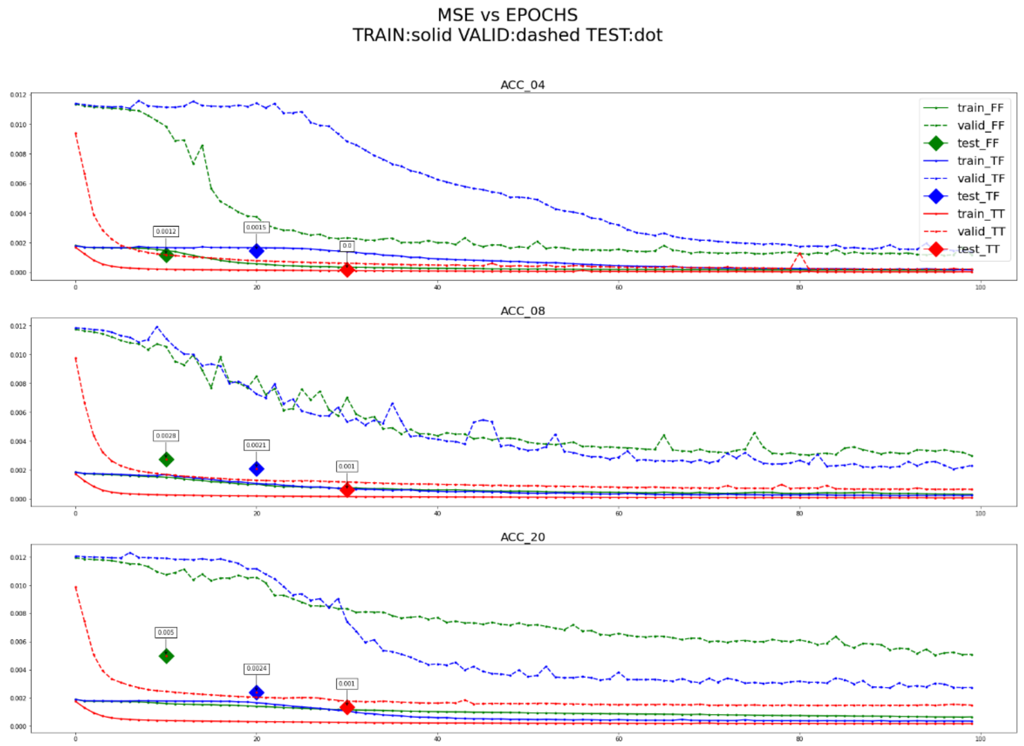


Fig. 5.7: MSE vs number of Epochs for all the different λ settings and for all the accelerations for the Training and the Validation test. The Test Error for each setting is annotated above the different markers per λ setting.

Regarding the PSNR measurements, highly similar performance was observed as we can see in fig. 5.9. At this point we would like to underpin the role of these two-validation metrics MSE and PSNR. Both measurements are closely related with the degree of undersampling, and the opposite behaviour is expected (MSE-decrease, PSNR-increase) proportional with the number of training epochs.

In fig. 5.10 the mean SSIM per batch per epoch is shown for all the pre-mentioned cases. Also, in fig. 5.11 the same results are presented in bar-plot format. What is important here to mention is that of course the TT setting was always outperformed all the rest but also for the extreme case of acc. 20.0 the TT setting was the only one capable of even slightly overcome the SSIM index of the undersampled image (thick-dashed- black line). This finding highlights the significance of the proposed setting on upgrading the performance of such CS-DL based architectures.

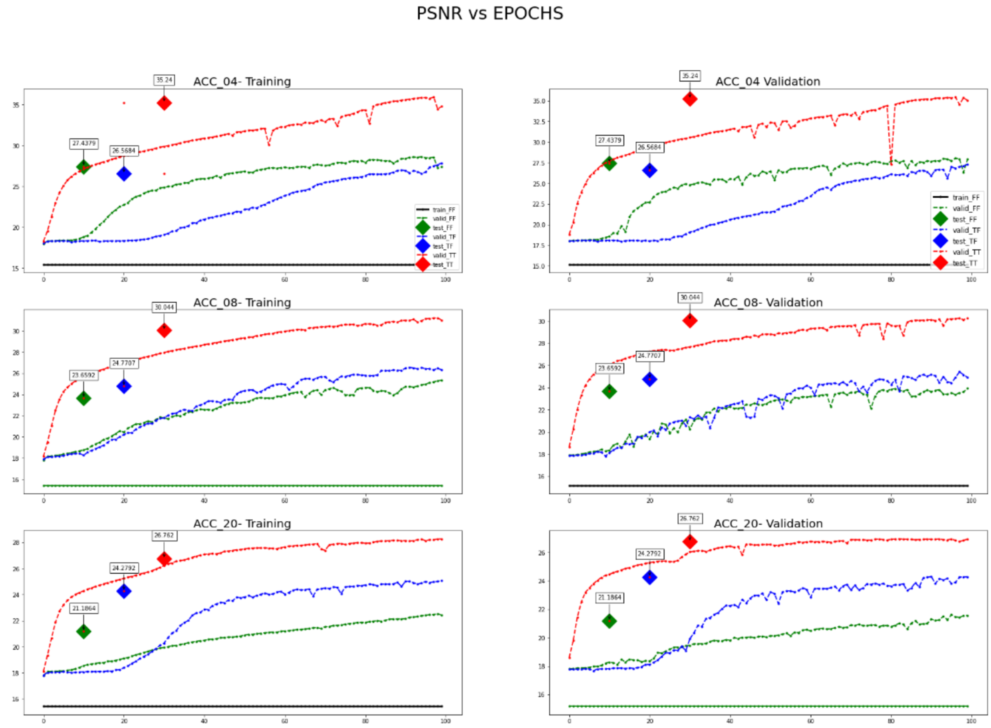


Fig.5.8: PSNR vs number of Epochs for all the different λ settings and for all the accelerations for the Training and the Validation test. The Test Error for each setting is annotated above the markers per λ setting.

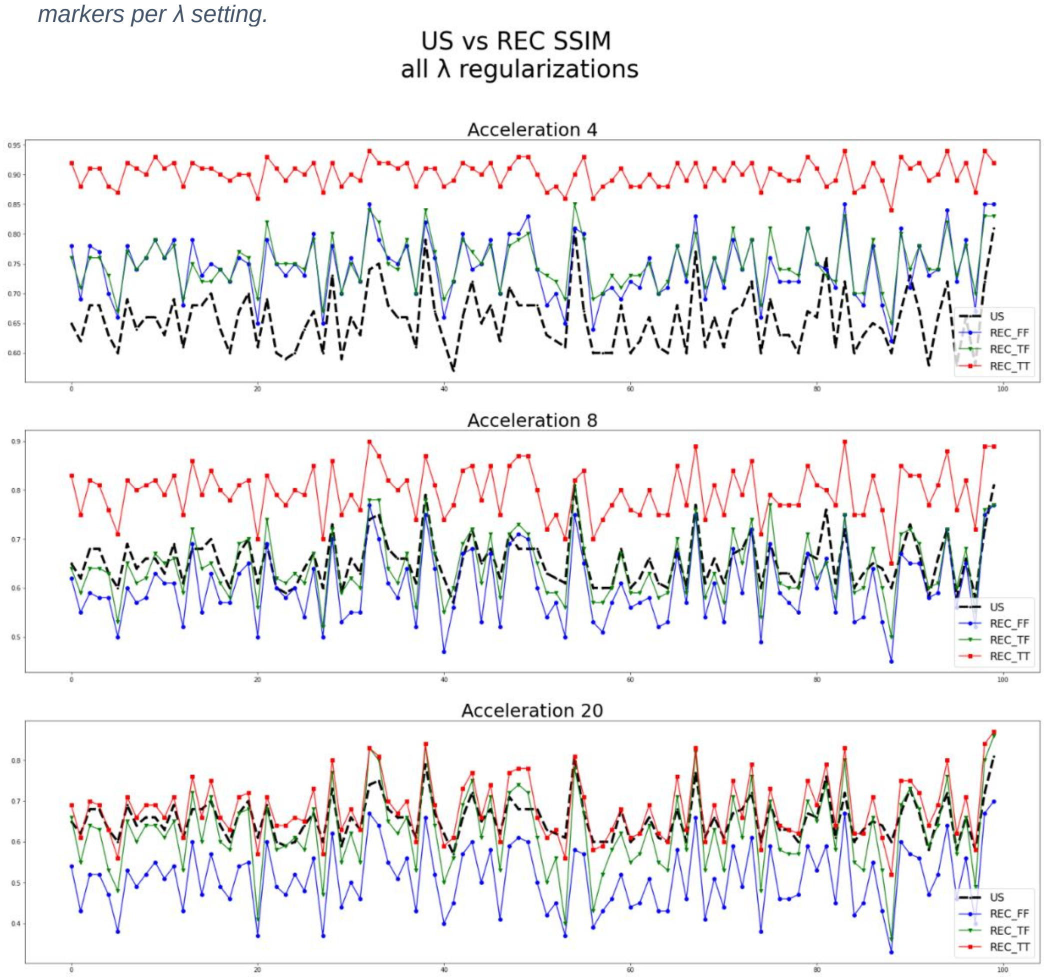


Fig. 5.9: SSIM for all saved Images for all the λ regularization parameters and all accelerations. The black dashed line is the base SSIM of the undersampled image.

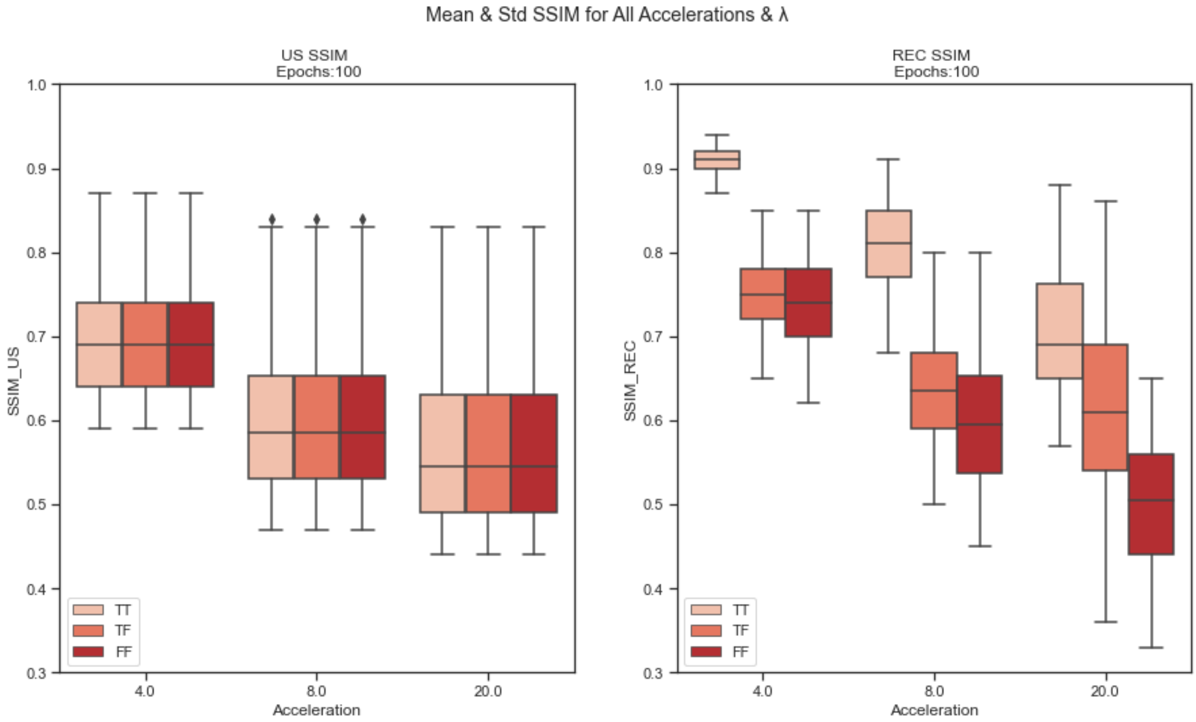


Fig. 5.10: Barplot showing the mean and std SSIM for all the λ regularization settings and all the accelerations.

Finally in fig. 5.12 the reconstruction results for the same brain slice are shown for all the different λ regularization settings and accelerations. On the upper right corner, the SSIM is mentioned. The different colours yellow-green-red denote the SSIM in increasing performance order. For all the above experiments the same order was conserved on all the data (train- validation- test) for the sake of comparison. (code)

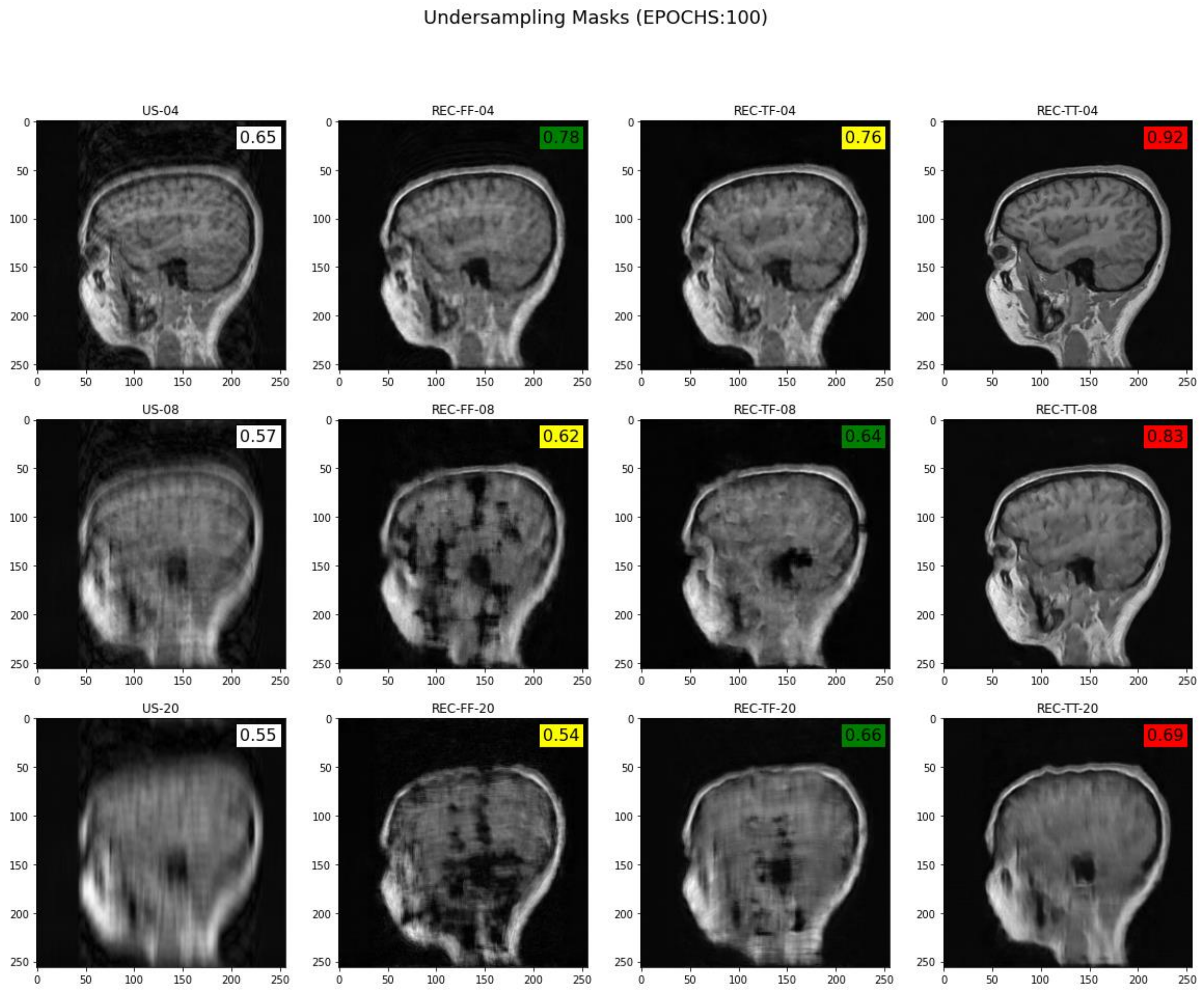


Fig.5.11: An indicative reconstruction for all the acceleration rates and all the λ regularizations. In the upper right corner, the SSIM is denoted between the illustrated image and the ground truth. The best performance is indicated with by colours with the following increasing order yellow<green<red.

DISCLOSURE Content

Chapter 6

Discussion & Future Research

In the current work, we used the CS-DL based architectures which composed of the two blocks, the CNN, and the DC block with an intermediate regularization term λ which controls the properties of the reconstructed image in accordance with the noise level in the acquired measurements.
The main novelty proposed on this work focused on converting the regularization term in a learnable setting and as the results suggest the performance is highly improved. We used three different network architectures for the CNN term and as our findings suggest in all the different architectures the spatially learnable λ regularization highly improves the reconstruction results.
All the used architectures and the presented results could be further improved through fine tuning, but this was out of the scope of the current work. The main aim was to underpin the importance of incorporate a learnable regularization and that the current setting can be easily adapted in this kind on networks. The most important aspect of the proposed setting, apart from the improved reconstructions is that it only needs a very limited additive computational burden.

Future Research

The optimal target of the proposed setting would be to derive a model for the prominent noise presence through the learning of the spatially learnable regularization parameter. Since our idea on working on this topic was conceived while working on the MRI pulse sequences optimization for accelerated reconstructions our target aims on that direction. In the optimal case that we will be able to derive a noise model for this source of acquisitions many pulse sequence related artifacts will further be understood. From the imaging acceleration perspective, when the undersampling of measurements should be performed on a real time setting, the prior knowledge about the sources of the noisy artifacts of a specific anatomy will highly increase the efficiency of the examinations.

On the same path our future research would like to focus on the CNN block and more specifically on the dual domain networks [55]. Under this kind of networks, the CNNs are tuned to operate not only on the image domain but also on the frequency domain (k-space). The interaction of the two domains allows learning the mapping between the sampled data (frequency) and the reconstructed image. Again, the more macroscopic outcome of this venture is again to better understand the sources of noise associated with the hardware components of the MRI during sampling.
This is our current research interest on CNNs operate on the frequency domain and more specifically we would like to focus on non-cartesian trajectories such as radial and spiral. Due to the optimised motion resistance and direct ability of undersampling the recent years this kind of trajectories are broadly used.
To conclude we showed how we can further optimise the reconstructions resulted from the very flexible CNN based on the CS-DL architectures, with a very simple and highly adaptive setting on the regularization term. Still CS remains challenging to implement since not all reduced-rate sampling is equal. Also, compression rates interplay with the properties of the input, the anatomy in our case, to impact the ability of the reconstruction. The main pros of the CS-DL is that the CNNs' surpass conventional formulations and sidestep the sparsity utilized from predefined prior knowledge assumptions. CNNs' capture and exploit relevant features from the data. However, depending on the current reconstruction anatomy further investigation should be performed on the formulation of the optimization problem. Finally, we are confident that by analysing the optimised performance offered by the regularization term we will be able to derive a noise model for the associated artifacts especially for the real sampled and not on synthetic data.

REFERENCES

1. E. Candès, J. Romberg, and T. Tao, “Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information,” *IEEE Trans. Inform. Theory*, vol. 52, no. 2, pp. 489–509, 2006.

2. D. Donoho, “Compressed sensing,” *IEEE Trans. Inform. Theory*, vol. 52, no. 4, pp. 1289–1306, 2006.

3. E. Candès and T. Tao, “Near optimal signal recovery from random projections: Universal encoding strategies?” *IEEE Trans. Inform. Theory,* vol. 52, no. 12, pp. 5406–5425, 2006.

4. J. Schlemper, J. Caballero, J. Hajnal, A. Price, and D. Rueckert, “A deep cascade of convolutional neural networks for dynamic MR image reconstruction,” IEEE Transactions on Medical Imaging, vol. 37, no. 2, pp. 491–503, 2018.

5. Pruessmann KP, Weiger M, Scheidegger MB, Boesiger P. SENSE: sensitivity encoding for fast MRI. Magnetic resonance in medicine: official journal of the Society of Magnetic Resonance in Medicine / Society of Magnetic Resonance in Medicine 1999;42(5):952-962.

6. Griswold MA, Jakob PM, Heidemann RM, Nittka M, Jellus V, Wang J, Kiefer B, Haase A. Generalized autocalibrating partially parallel acquisitions (GRAPPA). Magnetic resonance in medicine: official journal of the Society of Magnetic Resonance in Medicine / Society of Magnetic Resonance in Medicine 2002;47(6):1202 1210.

7. Ansari, R.A., Budhhiraju, K.M. A Comparative Evaluation of Denoising of Remotely Sensed Images Using Wavelet, Curvelet and Contourlet Transforms. *J Indian Soc Remote Sens* 44, 843–853 (2016). https://doi.org/10.1007/s12524-016-0552-y

8. S. Ravishankar and Y. Bresler, “MR image reconstruction from highly undersampled k-space data by dictionary learning,” IEEE Trans. Med. Imag., vol. 30, no. 5, pp. 1028–1041, May 2011.

9. Z. Zhang, Y. Xu, J. Yang, X. Li and D. Zhang, "A Survey of Sparse Representation: Algorithms and Applications," in *IEEE Access*, vol. 3, pp. 490-530, 2015, doi: 10.1109/ACCESS.2015.2430359.

10. Yang AC, Kretzler M, Sudarski S, Gulani V, Seiberlich N. Sparse Reconstruction Techniques in Magnetic Resonance Imaging: Methods, Applications, and Challenges to Clinical Adoption. Invest Radiol. 2016 Jun;51(6):349-64. doi: 10.1097/RLI.0000000000000274. PMID: 27003227; PMCID: PMC4948115.

11. Pilastri, A.L., & Tavares, J.M. (2016). Reconstruction Algorithms in Compressive Sensing: An Overview.
12. K. He, X. Zhang, S. Ren, and J. Sun, “Deep residual learning for image recognition,” arXiv preprint arXiv:1512.03385, 2015.

13. O. Ronneberger, P. Fischer, and T. Brox, “U-Net: convolutional networks for biomedical image segmentation,” in Proc. 18th Int. Conf. MICCAI, Munich, 2015, pp. 234–241.

14. McRobbie, D., Moore, E., Graves, M., & Prince, M. (2006). MRI from Picture to Proton (2nd ed.). Cambridge: Cambridge University Press. doi:10.1017/CBO9780511545405

15. B. Kevles (1996). ‘Naked to the Bone Medical Imaging in the Twentieth Century’. Camden, NJ: Rutgers University Press. pp. 19–22. ISBN 978-0-8135-2358-3

16. Edelman RR. The history of MR imaging as seen through the pages of radiology.2014 Nov; 273(2 Suppl.): S181-200. doi:10.1148/radiol.14140706. PMID: 25340436.

17. Berger A. Magnetic resonance imaging. BMJ. 2002 Jan 5;324(7328):35. doi: 10.1136/bmj.324.7328.35. PMID: 11777806; PMCID: PMC1121941.

18. R. A. Pooley, “Fundamental physics of MR imaging,” Radiographics, vol. 25, no. 4, pp. 1087–1099, 2005.

19. H. Hashemi, W. G. Bradley, and C. J. Lisanti, MRI: the basics. Lippincott Williams & Wilkins, 2012.

20. J. F. Schenck, “Health and Physiological Effects of Human Exposure to Whole-Body Four-Tesla Magnetic Fields during MRI,” Annals of the New York Academy of sciences, vol. 649, no. 1, pp. 285–301, 1992.

21. S.-B. Park, S.-Y. Kim, J.-H. Hyeong, and K.-R. Chung, “A study on the development of image analysis instrument and estimation of mass, volume, and centre of gravity using CT image in Korean,” Journal of Mechanical Science and Technology, vol. 28, no. 3, p. 971, 2014.

22. Conturo TE, Lori NF, Cull TS, Akbudak E, Snyder AZ, Shimony JS, McKinstry RC, Burton H, Raichle ME. Tracking neuronal fiber pathways in the living human brain. Proc Natl Acad Sci U S A. 1999 Aug 31;96(18):10422-7. doi:10.1073/pnas.96.18.10422. PMID: 10468624; PMCID: PMC17904.

23. Jones DK, Horsfield MA, Simmons A. Optimal strategies for measuring diffusion in anisotropic systems by magnetic resonance imaging. Magn Reson Med. 1999 Sep;42(3):515-25. PMID: 10467296.

24. Huettel, S. A.; Song, A. W.; McCarthy, G. (2009), Functional Magnetic Resonance Imaging (2 ed.), Massachusetts: Sinauer, ISBN 978-0-87893-286

25. Griswold MA, Jakob PM, Heidemann RM, Nittka M, Jellus V, Wang J, Kiefer B, Haase A. Generalized autocalibrating partially parallel acquisitions (GRAPPA). Magnetic resonance in medicine : official journal of the Society of Magnetic Resonance in Medicine / Society of Magnetic Resonance in Medicine 2002;47(6):1202-1210.

26. Gorodnitsky IF, Rao BD. Sparse signal reconstruction from limited data using FOCUSS: A re-weighted minimum norm algorithm. IEEE Signal Processing 1997;45(3):600-616.

27. Tibshirani R. Regression shrinkage and selection via the Lasso. J Roy Stat Soc B Met 1996;58(1):267-288.

28. "Center Leadership | Martinos Center". 2019-04-01. Retrieved 2021-10-30.

29. W.D. Heiss, P. Raab,H. Lanfermann, (2011) Journal of Nuclear Medicine , 52 (10) 1585-1600; DOI: https://doi.org/10.2967/jnumed.110.084210

30. C. Westbrook (2016) MRI at a Glance, (3rd ed.) Wiley. 978-1-119-05355-2

31. N. D. Prionas, Quantitative Targeting and Radiation Therapy of Breast Cancer using a Dedicated Breast Computed Tomography Platform. University of California, Davis, 2012.

32. M. Lustig, D. L. Donoho, J. M. Santos, and J. M. Pauly, “Compressed sensing MRI,” IEEE signal processing magazine, vol. 25, no. 2, pp. 72–82, 2008.

33. R. Bitar, G. Leung, R. Perng, S. Tadros, A. R. Moody, J. Sarrazin, C. McGregor, M. Christakis, S. Symons, A. Nelson, et al., “MR pulse sequences: what every radiologist wants to know but is afraid to ask,” Radiographics, vol. 26, no. 2, pp. 513–537, 2006.

34. D.G. Nishimura, (1996) Principles of magnetic resonance imaging (1st ed.) Selfpublished, URL: http://www-ee.stanford.edu/$~$dwight/book.html.

35. Edler, Karl. (2010). Spherical Harmonic Detection Coils and Dynamic Preemphasis. 10.13140/RG.2.1.4390.1046.

36. D. J. Larkman and R. G. Nunes, “Parallel magnetic resonance imaging,” Physics
in medicine and biology, vol. 52, no. 7, p. R15, 2007.

37. A. Karpathy and F. F. Li. Deep visual-semantic alignments for generating image descriptions. In Proceedings of the IEEE conference on computer vision and pattern recognition, pages 3128–3137, 2015.

38. D. Zikic, Y. Ioannou, M. Brown, A. Criminisi, and A. Criminisi. Segmentation of brain tumor tissues with convolutional neural networks. International Conference on Medical Image Computing and Computer-Assisted Intervention-BRATS, pages 36–39, 2014.

39. K. Hammernik, T. Klatzer, E. Kobler, M. P. Recht, D. K. Sodickson, T. Pock, and F. Knoll. Learning a variational network for reconstruction of accelerated MRI data. Magnetic resonance in medicine, 79:3055–3071, 2018.

40. J. Krebs, T. Mansi, B. Mailh´e, N. Ayache, and H. Delingette. Unsupervised probabilistic deformation modeling for robust di↵eomorphic registration. In Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support, pages 101–109. Springer, 2018.

41. B. D. de Vos, F. F. Berendsen, M. A. Viergever, H. Sokooti, M. Staring, and I. Isgum. A deep learning framework for unsupervised affine and deformable image registration. Medical image analysis, 52:128–143, 2019.

42. W. Shen, M. Zhou, F. Yang, C. Yang, and J. Tian. Multi-scale convolutional neural networks for lung nodule classification. In International Conference on Information Processing in Medical Imaging, pages 588–599. Springer, 2015.

43. B. J. Erickson, P. Korfiatis, Z. Akkus, and T. L. Kline. Machine learning for medical imaging. Radiographics, 37(2):505–515, 2017.

44. Tofts, P., Ed. (2003) Quantitative MRI of the Brain: Measuring Changes Caused by Disease. Wiley, Chichester, Hoboken. http://dx.doi.org/10.1002/0470869526

45. K. He, X. Zhang, S. Ren, and J. Sun. Deep residual learning for image recognition. In The IEEE Conference on Computer Vision and Pattern Recognition, June 2016.

46. Tezcan, K. C., Baumgartner, C. F., Luechinger, R., Pruessmann, K. P. & Konukoglu, E. MR Image Reconstruction Using Deep Density Priors. *IEEE Transactions on Medical Imaging* 38, 1633–1642, DOI: 10.1109/TMI.2018.2887072 (2019). Publisher: ieeexplore.ieee.org.

47. Vapnik V. Principles of risk minimization for learning theory. Advances in Neural Information Processing Systems, 4, 1991.

48. Dedmari, M.A.; Conjeti, S.; Estrada, S.; Ehses, P.; Stöcker, T.; Reuter, M. Complex fully convolutional neural networks for mr image reconstruction. In International Workshop on Machine Learning for Medical Image

Reconstruction; Springer: Berlin/Heidelberg, Germany, 2018; pp. 30–38.

49. T. Dieleman, " Lasagne: First Release," 8 2015. [Online]. Available: https://lasagne.readthedocs.io/en/latest/modules/nonlinearities.html#lasagne.nonlinearities.softmax.

50. R. Souza, O. Lucena, J. Garrafa, D. Gobbi, M. Saluzzi, S. Appenzeller, L. Rittner, R. Frayne, R. Lotufo. (2017) Neuroimage, "An open, multi-vendor, multi-field-strength brain MR dataset and analysis of publicly available skull stripping methods agreement." pp. 482-494.

51. Marcel, S., & Rodriguez, Y. (2010, October). Torchvision the machine-vision package of torch. In Proceedings of the 18th ACM international conference on Multimedia (pp. 1485-1488)

52. H. Jung, J. C. Ye, and E. Y. Kim. Improved k–t BLAST and k–t SENSE using FOCUSS. Physics in medicine and biology, 52(11):3201, 2007.

53. Souza R, Bento M, Nogovitsyn N, Chung KJ, Loos W, Lebel RM, Frayne R. Dual-domain cascade of U-nets for multi-channel magnetic resonance image reconstruction. Magn Reson Imaging. 2020 Sep;71:140-153. doi: 10.1016/j.mri.2020.06.002. Epub 2020 Jun 17. PMID: 32562744.

54. M. Lustig and D. Donoho, “Compressed sensing MRI,” IEEE Signal Process. Mag., vol. 25, no. 2, pp. 72–82, Mar. 2008.

55. Zhang Y, Yap PT, Qu L, Cheng JZ, Shen D. Dual-domain convolutional neural networks for improving structural information in 3 T MRI. Magn Reson Imaging. 2019 Dec;64:90-100. doi: 10.1016/j.mri.2019.05.023. Epub 2019 Jun 5. PMID: 31175927; PMCID: PMC6874896.

56. Qu X, Zhang W, Guo D, Cai C, Cai S, Chen Z. Iterative thresholding compressed sensing MRI based on contourlet transform. Inverse Probl Sci Eng. 2010;18(6):737–58.

57. Knoll F, Bredies K, Pock T, Stollberger R. Second order total generalized variation (TGV) for MRI. Magn Reson Med. 2011;65(2):480–91.

58. Zhan Z, Cai JF, Guo D, Liu Y, Chen Z, Qu X. Fast multiclass dictionaries learning with geometrical directions in MRI reconstruction. IEEE TransBiomed Eng. 2016;63(9):1850–61.

59. Lustig M, Donoho D, Pauly JM. Sparse MRI: The application of compressed sensing for rapid MR imaging. Magn Reson Med. 2007;58(6):1182–95.

60. Block KT, Uecker M, Frahm J. Undersampled radial MRI with multiple coils. Iterative image reconstruction using a total variation constraint. Magn Reson Med. 2007;57(6):1086–98.

61. Liu Y, Zhan Z, Cai JF, Guo D, Chen Z, Qu X. Projected iterative soft thresholding algorithm for tight frames in compressed sensing magnetic resonance imaging. IEEE Trans Med Imaging. 2016;35(9):2130–40

62. Eslahi SV, Dhulipala PV, Shi C, Xie G, Ji JX. Parallel compressive sensing in a hybrid space: application in interventional MRI. In: Annual international conference of the IEEE engineering in medicine and biology society (EMBC); 2017. p. 3260–63

63. Lai Z, Qu X, Liu Y, Guo D, Ye J, Zhan Z, Chen Z. Image reconstruction of compressed sensing MRI using graph-based redundant wavelet transform. Med Image Anal. 2016;27:93–104.

64. Qu X, Guo D, Ning B, Hou Y, Lin Y, Cai S, Chen Z. Undersampled MRI reconstruction with patch-based directional wavelets. Magn Reson Imaging. 2012;30(7):964–77.

65. Ravishankar S, Bresler Y. Data-driven learning of a union of sparsifying transforms model for blind compressed sensing. IEEE Trans Comput. Imaging. 2016;2(3):294–309.

66. Yaman B, Hosseini SAH, Moeller S, Ellermann J, Uğurbil K, Akçakaya M. Self-supervised learning of physics-guided reconstruction neural networks without fully sampled reference data. Magn Reson Med. 2020;84(6):3172–91.

67. Combettes PL, Pesquet J-C. Proximal splitting methods in signal processing. In: Fixed-point algorithms for inverse problems in science and engineering. Springer; 2011. p. 185–212.

68. Daubechies I, Defrise M, De Mol C. An iterative thresholding algorithm for linear inverse problems with a sparsity constraint. Commun Pure Appl Math. 2004;57(11):1413–57.

69. Boyd S, Parikh N, Chu E. Distributed optimization and statistical learning via the alternating direction method of multipliers. Found Trends Mach Learn. 2011;3(1):1–122.

70. Ravishankar S, Bresler Y. MR image reconstruction from highly undersampled k-space data by dictionary learning. IEEE Trans Med Imaging. 2010;30(5):1028–41.

71. Wang Y, Ying L. Compressed sensing dynamic cardiac cine MRI using learned spatiotemporal dictionary. IEEE Trans Biomed Eng. 2013;61(4):1109–20.

72. Zhan Z, Cai JF, Guo D, Liu Y, Chen Z, Qu X. Fast multiclass dictionaries learning with geometrical directions in MRI reconstruction. IEEE Trans Biomed Eng. 2016;63(9):1850–61.

73. Strong, D., & Chan, T.Edge-preserving and scale-dependent properties of total variation regularization. Inverse problems, 2003;19(6), S165.

74. Calvetti, D., & Somersalo, E. Hypermodels in the Bayesian imaging framework. Inverse Problems, 2008;24(3), 034013.

75. Ntouskos, V., & Pirri, F. Confidence driven tgv fusion. 2016;arXiv preprint arXiv:1603.09302.