Improvements and deep learning applications in 19F-NMR

Candidate

Andrea Ciardiello
ID number 1268232

Thesis Advisors

Riccardo Faccini

Stefano Giagu

Co-Advisor

Cecilia Voena

March 2021

"In time, you will know the tragic extent of my failings...."

Wayne June, Darkest Dungeons

Abstract

Improvements and deep learning applications in 19F-NMR

by Andrea CIARDIELLO

A promising new therapeutic approach for pancreatic cancer is Proton Boron Fusion Therapy (PBFT) which produces a highly localized damaging action through nuclear reactions of the incoming proton beam and boron atoms, conveniently administered to the patient before the treatment. Starting from this recent proposal to use boron (and possibly fluorine) as chemical radiosensitizing agents in proton therapy, a new interest has arisen for the study of borate compounds.

To evaluate the effectiveness of these compounds it is necessary to measure the bio-distribution of tracers accumulated in the tissues before the irradiation on a patient by patient basis . It's safe to assume that fluorine/boron mediated sensitization will depend critically from compound concentration that can be achieved in the target nuclei which means that the clinical application of the treatment will need the development of a reliable quantification technique optimized for the tracer of interest. The first Chapter will report our study of the intracellular internalization of fluoro-boron-phenylalanine (F-BPA) ,one of the most promising candidate to be adopted in PBFT as boron carrier, in a cellular model of the pancreatic cancer (PANC-1 cell line) using fluorine magnetic resonance spectroscopy (19F-MRS). The main advantage of F-BPA over the standard molecule adopted in the field, the boron-phenylalanine (BPA), is the addition of the fluorine atom that allows its quantification with magnetic resonance and it may also be used as a tracer for magnetic resonance imaging (MRI).

This is the first step to validate a boron carrier as a proton therapy enhancer since PBFT damage is highly localized and its effects depend on the intracellular concentration of boron.

In the second Chapter, I will discuss the possibility of measuring fluorine accumulation in tissues using 19F-MRS ex vivo in an animal model of pancreatic adenocarcinoma. This experiment will help define the sensitivity that has to be reached to perform an in vivo experiment of localized 19F-MRS and 19F MRI and will provide the data for the validation of these in vivo techniques.

We also believe that this method of quantification ex vivo may be of general interest to screen for fluorine tagged compounds for the utilization in PBFT.

The isotope 19F is characterized by 100% natural abundance, high relative sensitivity, it displays an intense nuclear magnetic resonance signal and it is almost nonexistent in the human body. In contrast to NMR techniques based on proton resonance all the signal detected can be attributed to the tracer introduced and the signal that can be obtained is limited by the tracer concentration in tissues that in turn is constrained by the safety of the dosage administered and the method used for drug delivery.

So, in the context of 19F-MRI, in presence of low signal and no fluorine induced background, it is extremely important to develop tools to remove noise (denoising).

Thus, the third Chapter is a preliminary work on the application of a deep learning convolutional neural network (CNN) to the task of noise reduction in magnetic resonance imaging (MRI). MRI acquisition is performed in the frequency domain, I will show how a newly proposed CNN trained on raw frequency data may outperform a network of the same complexity that is trained in a more conventional way on the reconstructed magnitude images.

The last Chapter consists in an application of this proposed method to a denoising task of a large dataset of parallel imaging to show how the method can be easily transferred to many other acquisition modalities.

Acknowledgements

....

Contents

List of Figures

List of Tables

Chapter 1

Introduction

In this document I will report my activity within the NEPTUNE (Nuclear process driven Enhancement of Proton Therapy UNravEled) collaboration where I conducted my PhD project.

My contribution is divided into two parts. The first part concentrates on the quantification of cellular internalization and biodistribution of compounds tagged with fluorine trough fluorine magnetic resonance (19F-NMR) and the second part on a solution to reduce the effect of noise in Magnetic Resonance Imaging (MRI) in acquisition where the low concentration of resonant molecules produce a low signal as in the case of 19F-MRI.

The final aim of the NEPTUNE collaboration is to study the impact of fluoride and borate (possibly tagged with fluorine) compounds as chemical radio-sensitizing agents in proton therapy.

To evaluate the effectiveness of these compounds it is necessary to measure the bio-distribution of tracers accumulated in the tissues before the irradiation on a patient by patient basis.

It's safe to assume that fluorine/boron mediated sensitization will depend critically from com- pound concentration that can be achieved in the target nuclei which means that the clinical application of the treatment will need the development of a reliable imaging technique optimized for the tracer of interest.

MRI is a non-invasive and non-ionizing radiation diagnostic tool that can be used to directly quantify 19F compounds during the treatment.

The isotope 19F is characterized by 100% natural abundance, high relative sensitivity, it displays an intense nuclear magnetic resonance signal and it is almost non-existent in the human body so its target-non-target signal ratio is intrinsically very large given that endogenous 19F-derived background noise should be very low. Furthermore, fluorine nuclear magnetic resonance differs only by 6% from that of 1H, potentially allowing one to conduct 19F-MRI on existing 1H imaging hardware with little modification.

The validation of a tracer molecule as a PBFT carrier development of a reliable MRI measure to quantify it in in vivo studies has involved the achievement of some important intermediate steps that will be the main topics of this document.

My work concentrated on one particular boron carrier, borono-phenylalanine (BPA), and its fluorine tagged analogue fluoro-boron-phenylalanine (F-BPA) but the methodologies employed are general and, when other tracer candidates will be proposed, the experimental pipeline developed can be easily transferred to their analysis.

The first step in the validation of F-BPA as a proton therapy enhancer is the quantification of boron that is transported inside the cells. The proposed technique, Proton Boron Fusion Therapy (PBFT), produces a highly localized damaging action to

the DNA through nuclear reactions of the incoming proton beam and boron atoms. Due to the short range of the produced particles, to maximize this effect the target boron has to be inside the cell membrane.

In Chapter 4 I will report the results of our studies on cellular accumulation of the F-BPA and I will discuss in more detail the application of PBFT and the motivation for which cellular internalization is critical to its success.

In this section I will also introduce the cancer model that will be the case study of this work, that is the human pancreatic cell line (PANC1), and I will discuss why pancreatic cancer was chosen to be a good candidate to be treated with PBFT.

Although the experiments are conducted on the PANC1 cell line, the methodology developed is general and it could be applied to other cancer models as well.

Also, in this chapter our attempt to propose an alternative boron carrier, that had many promising characteristics but failed in the biological application, will be briefly presented.

The second step of this work is centered on the experiment on the animal model.

The distribution of the drug is not uniform in the body of the patient and for the therapy to be successful, a suitable concentration of boron has to be built in the organ of interest.

Thus, the administration of the drug has to be monitored and the concentration reached in the organ that needs the treatment has to be carefully quantified.

In Chapter 5 our experimental protocol, our somministration system and our measurement strategy will be discussed and analyzed in respect to previous reported attempts.

In particular, the dosage and the administration time of the drug has a great impact on the final concentration in tissues. So we will need a complete pipeline to measure all possible sources of variability in the experiment to find the best approach in order to obtain the best concentration for the therapy to be effective, i.e. the maximum concentration of boron targets in the tumor mass to be treated and the lowest concentration in the surrounding organs to avoid damaging healthy tissues.

The results of this experiment are valuable both because they help improve our experimental protocol and because they help design the image acquisition so that the parameters are tuned on the expected signal and on the desired resolution.

Furthermore, this experiment provides the validation data on the expected concentration in tissues with which we can compare the results in in-vivo quantification.

From general considerations and from the results of our ex-vivo experiment it seems obvious that in a 19F MRI in our experimental conditions the main limitation will be the low signal caused by the low concentration of fluorine inside the tissues and this will result in a noisy image which will compromise the quantification.

This is true in general, 19F-NMR has profoundly different characteristics from 1H-NMR since there are no endogenous sources of 19F. Therefore, while 1H-NMR studies the properties of water within tissues, 19F-NMR looks for signals (even small ones) in a low-background context. It completely changes the focus of research in a framework where noise is the dominant factor.

Thus, in the context of 19F-MRI, in presence of low signal and no fluorine induced background, it is extremely important to develop tools to remove noise (denoising).

In Chapter 6 we implemented two solutions: A novel approach based on the raw frequency data and a more traditional one that works on the already reconstructed

image. In particular, the advantage of working with the frequency based raw-data instead of the images already reconstructed will be discussed and the issues in the application to our use case will be analyzed.

The denoiser will be first tested on simulated data and then in Chapter 7 it will be applied in the task of denoising a large dataset of morphological MRI images. The possibility to rapidly transfer the application of our method to different acquisition modality derives from the fact that it exploits a fundamental property in the noise model common to many different MRI acquisition modality.

The most important results of this section is the possibility to train the denoiser in an unsupervised fashion that means that a ground truth, so the noise free version of the image, is not needed to perform the learning. This perk is directly derived from our strategy of working with raw-data instead of images.

The potential of this technique is not limited to the acquisitions in 19F MRI and it is a topic that should be deepened independently from it since even in traditional images with high signal the noise reduction is important if they are used to extract quantitative information.

In Figure 1.1 is reported the outline of this document. Our prospective work (highlighted in red) is based on the availability of a fluorine tracer with good internalization property and a suitable biodistribution in tissue. Thus our experimental pipeline (on the left) is shaped to select these properties:

- The in-vitro experiment allows us to measure the internalization in cells and to select tracers with good NMR properties that are suitable for imaging.

- The preliminary experiments on the animal model are needed to tune our protocol to reach the best desired concentration in tissues.

These methods are general and can be applied to any proposed tracer. So another molecule that can increase the boron concentration and the fluorine signal since it has multiple atoms of both is tested and then discarded since it has great NMR properties but low biological utility.

On the right, the development of the proposed method is outlined

- The denoising approach that will be applied to the low signal 19F-MRI is first developed on simulated data and then the method proposed is validated on an external dataset composed by examples of another acquisition modality.

- The experience gained working with real data, that is usually more challenging than simulated one, is used to improve our proposed denoiser. In particular, the data required to reach reasonable performance can be evaluated with better accuracy on a real task.

This strategy is applied to improve PBFT, specifically in the treatment of pancreatic cancer, but it is generalizable to any tracer/task selection that involves the pianification of an imaging experiment with low signal where a data driven deep learning denoising can be used.

In the next Chapters, the basic concepts needed to put into context the work are reviewed. In Section 3.1 we will address the topic of NMR spectroscopy while imaging will be treated in Section 3.2. Neural networks, with a particular interest in convolutional neural networks, will be discussed in Chapter 2 .

In the final section instead we will discuss the challenges of the prospective work in the light of the results achieved. In the conclusions, the experiments planned for

the in vivo transfer of the developed techniques and a possible extension of what has been done for the denoising method are reported.
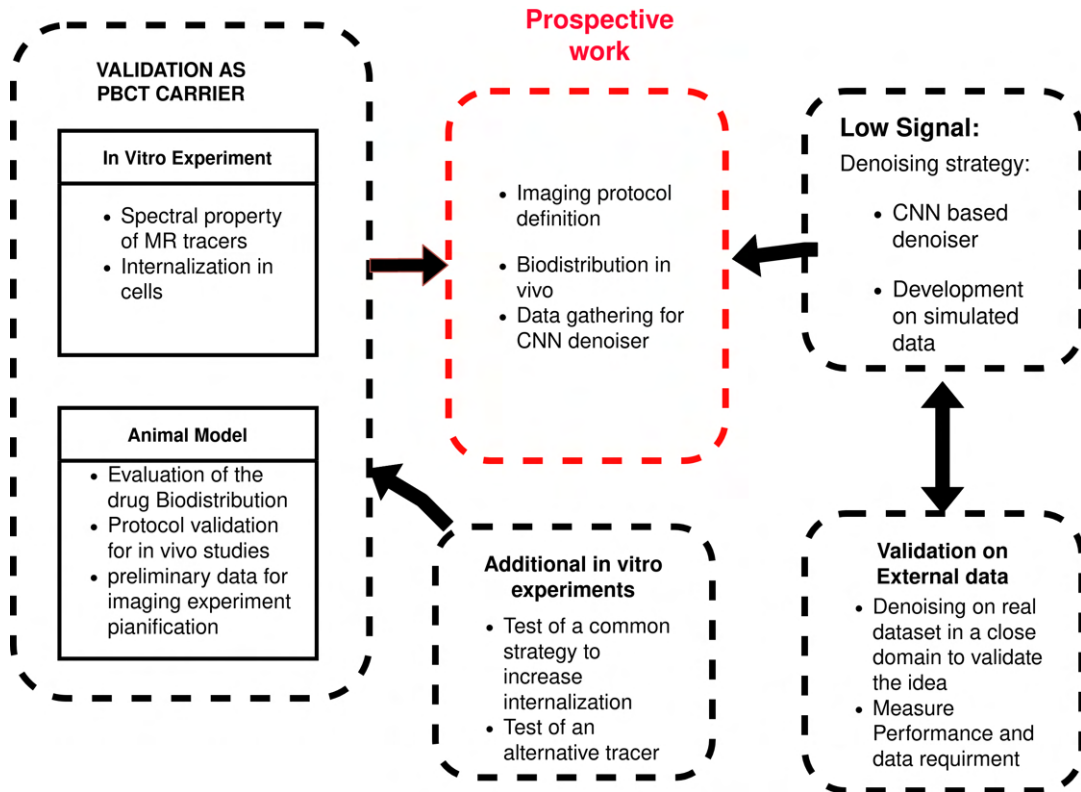
FIGURE 1.1: The structure of the document. The figure outlines the two main parts of my work in relation with the prospective work towards NEPTUNE project aims (center, in Red). On the left, we have the experiments needed for the validation of a proposed molecule as a boron carrier for the PBFT applied to pancreatic cancer. These experiment are needed to select and characterize the tracer to be used in the imaging quantification.
On the right, the proposal of a deep learning based solution first developed on simulated data and then tested in a real task is schematized.
The denoiser will be applied to the intrinsically low signal acquisition of 19-MRI.

Chapter 2

Introduction to Deep Learning

In this Chapter we will review the basic concept of Deep Learning with a particular attention to Convolution Neural Networks (CNN).

CNNs are machine learning models that have immense capacity to create hierarchical representation and have been successfully applied to computer vision problems including object localisation, classification and super-resolution.

Following the breakthrough in image classification with a deep architecture in the ImageNet LSVRC-2010 classification challenge (Krizhevsky, Sutskever, and Hinton, 2012), CNN are now successfully implemented in numerous task in the computer vision domain.

In the medical field, CNNs are nowadays routinely applied to diagnostic imaging (at least at research level) and they are one of the most used model in computer aided diagnosis (CAD) in many application from pathology classification to anatomical segmentation segmentation.

The field is in continuous development as the best model for a given task changes every few months, as can be seen in the leaderboard of any challenge dataset (*Grand Challenge Biomedical challenge database*), that are open competitions to compare solutions on the same ground. For this reason it is not useful to discuss the best approach to every problem so we will concentrate on the fundamentals and on good practices.

A good overview of the widespread application of CNNs in the medical imaging field can be found in this book by Lu et al. (Lu et al., 2017).

The considerable success of CNNs can be attributed to two main reasons that makes them superior to traditional machine learning algorithms and that have driven to a wide adoption of CNNs in the medical image analysis domain:

First, its scalable feature learning architecture that optimize model parameters for a given task and that relies very little on feature-engineering or prior knowledge and secondly, the end-to-end model training strategy which allows to process image data from the raw sources to the task end, even in a complex pipeline, with an automated methodology.

A particular example of CNN models, known as fully convolutional network (FCN), has been shown to offer improved computational efficiency and representation learning capabilities due to simpler model parametrisation and lower computational cost when applied to large images.

This section will introduce the technical details needed to understand the implementation of the CNN based denoiser that we are proposing in Chapter 6.

2.1 Neural networks

The modern standard deep neural network model for machine learning applications is the feedforward network.

This section aims to introduce all the definition needed to explain this sentence. First, these are *models* in the sense that their goals is to approximate a (possibly unknown) function f^*. The simplest example is a classifier that maps an element x to its category y. It can be described as a function $y = f^*(x)$. A feedforward network define a mapping between the set of examples and categories $\{\mathbf{x}, \mathbf{y}\}$ $\mathbf{y} = f(\mathbf{x}; \mathbf{W})$ that models the original function f^* and learns the parameters $\mathbf{W}$ that give the best function approximation.

In these models, the information flows from the input x, to some intermediate computations, to the output y without feedback between steps. For this reason they are called *feedforward*.

Feedforward networks are called *networks* because they are a composition of simple functions built in a chain structure. For example, lets have N function f^n with $n = 1, ..., N$, these functions are subsequently applied to the input x and they are chained together to $y = f^N(f^{N-1}(...f^2(f^1(x)))...))$.

This chain structure is the most used as building block of neural networks and it can be represented as a directed acyclic graph (i.e a graph with no loop, or feedforward).

We will call f^1 the first layer of the network, f^2 the second layer, and so on. The last layer f^N is called the output layer. The length of the chain is called the *depth* of the model. Since a network can be composed of hundreds of layers, modern neural networks are usually referred to as *deep networks*.

During the network training, we aim to match our function f to the original f^*. The training data $\{\mathbf{x}, \mathbf{y}\}$ (i.e pairs of input and output elements) provide a noisy approximation of $\mathbf{y} \sim f^*(\mathbf{x})$ and it is used to force the output of the last layer of the network to reproduce the expected function behaviour (i.e $f(x)$ should be close to y).

Only the last layer is conditioned directly by the training examples so all the other layers can be shaped freely to better approximate f^*: The role of the training algorithm is selecting the best parameter for these layers.

Since these layer are "hidden" during training (i.e their output is not directly influenced from the data), they are called the *hidden layers* of the network. Each hidden layer is composed by elements called hidden units that perform the basic computations in the neural network. The number of the hidden units of a layer is called the *width* of the layer.

One way to build an intuition of how a neural network works is to imagine that the final layer is a simple linear model that operate not on the input itself $\mathbf{x}$ but on a transformation of it $\phi(\mathbf{x})$ that is created trough the other layers. We can describe the action of the hidden layers as the formation of a synthetic description of the input that is shaped by the training algorithm to be useful to correctly represent it. For this reason, these structured intermediate descriptions of the data are called representations or (complex) features.

From a biological point of view, the structure of these networks resembles a biological neural network: The elements of the layers can be seen as neurons and the parameters of the chained functions as the synapses. As a biological neuron, which activity is modulated by a large amount of signals coming from synapses with other neurons, the value of an element of a layer is given by the many inputs that it receives from the previous layer. Also, the modularity and the hierarchical structure is loosely inspired by the brain but the similarity is only superficial since feedforward neural networks do not aim to directly model brain functions.

In Figure 2.1 there is a schematic representation of a network with two hidden layers: An input layer *x* of width 12 is connected to the first hidden layer of width 8 that is connected to the second layer of width 6 that, finally, is connected to the output layer. The direction of the connections is always from input (top) to output (bottom) without loops between elements: Connection came only from the previous layer and elements in the same layer are not directly connected. This architecture is called a multilayer fully connected neural network since it is formed by multiple layers in which every element is connected to all the elements of the previous one so that the connectivity (i.e the ratio between the actual connections and the total connections) between the two layers is 1. This is the most basic example of a neural network but it still finds applications and it will be our case study for the next paragraphs.

Input Layer ^{12}

Hidden Layer

Hidden Layer

Output Layer ^{1}
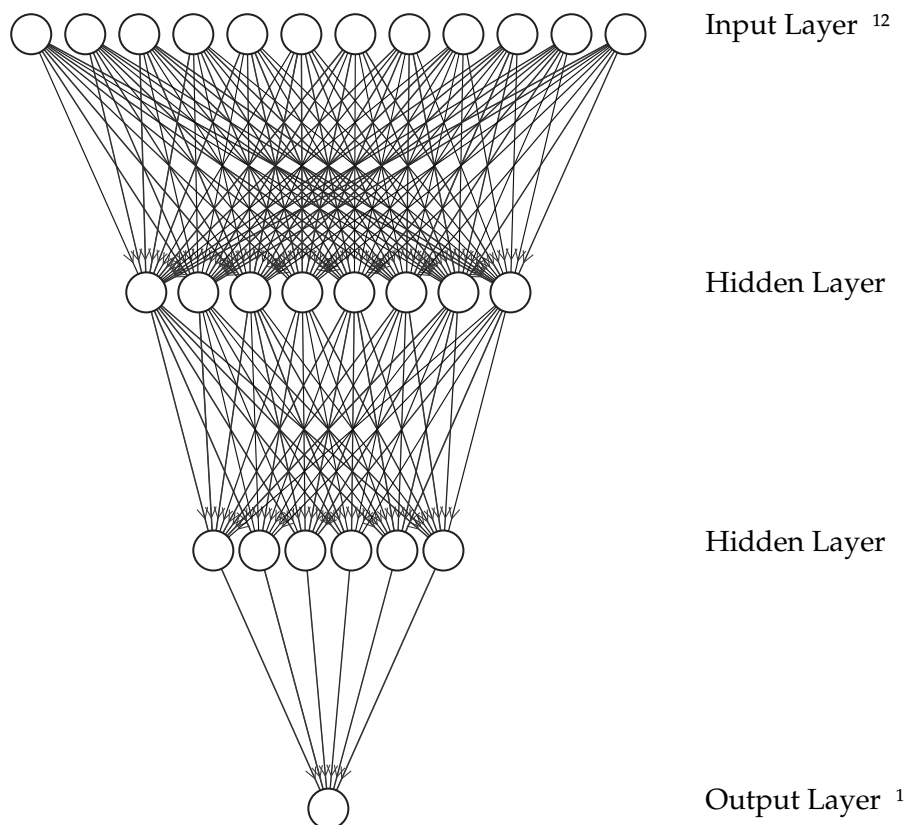
FIGURE 2.1: Schematic representation of a fully connected neural network with two hidden layers. The first layer (top) is the input layer. Arrow represent the weights (parameters) of layer and circles their hidden units. Connections go forward from input to output.

2.1.1 Activation functions and hidden units

To finish our description of the network architecture (i.e the overall structure of the network) in the hidden layers we need to introduce the concept of *activation function* of the hidden unit. Hidden units can be described as the computing elements of an affine transformation of an input vector **x** (that can be the input layer or another hidden layer) to which is subsequently applied a non-liner function *g* that is called the activation function. So, for a vector of hidden units on a generic *n* layer $\mathbf{x^n}$ its

value is computed from the previous $n - 1$ layer with the following relation

$$\mathbf{x}^n = g(\mathbf{W}^{n-1,n}\mathbf{x}^{n-1} + \mathbf{h}^n) \tag{2.1}$$

Where $\mathbf{W}^{n-1,n}$ is the matrix of the weights of the connections between the elements of the layer and the previous layer, $\mathbf{h}^n$ is a bias vector and $\mathbf{x}^{n-1}$ is the vector of the activation of the previous layer. The choice of the activation function g is tied to the task that the network has to perform and the data that is applied to, and there is no definitive answer.

One of the most commonly used activation function is the Rectified Linear Unit (ReLU) that is valued 0 if its input is negative or it is linear if the input is positive.

$$relu(x) = max(x, 0) \tag{2.2}$$

The motivations behind the widespread use of the ReLU activation function are that it is cheap to compute, it is easy to optimize and more importantly it helps reducing the number of active neurons since the activity is zero for negative inputs, but there are many alternative that may perform better or that can have different applications.

An example of activation function that can be used is hidden layers or can have a specialized role as output unit for tasks that require predicting a binary value is the logistic sigmoid $\sigma(x)$ defined as:

$$\sigma(x) = \frac{1}{1 + \exp(-x)} \tag{2.3}$$

There is no straightforward methods to determine which different activation function will be more beneficial to implement for a chosen application so we will use the ReLU function as an example in all the following sections.

The only important comment to make is that two layers with different activation functions will be parametrized differently so the training behaviour and the final function learned may be different if a different activation is used. ReLU has also been implemented in the proposed neural network application in Chapter 6.

The book by Goodfellow et al. (Goodfellow, Bengio, and Courville, 2016) in Chapter 6 gives a more extensive review of possible choices for activation functions that is a topic outside the scope of this introduction. We will discuss briefly the advantages and disadvantages of different activation functions when we will talk about neural networks training is Section 2.3.

2.2   Learning from examples

Machine learning algorithms differ from traditional programming because they consist in instructions on how to learn a task from a set of examples instead of instructions on how to perform the task itself. Thus, in this definition, learning means improving, in respect to some performance metric, the outcome of a task gaining experience from a set of collected examples.

The application of such algorithms may allow computers to tackle problems for which it is difficult to define a sequence of intermediate steps needed to reach the solution.

First, we have to define what learning from experience means: A set of examples is represented as a collection of features that describe quantitatively or qualitatively the objects that we are analyzing; we will refer to this collection of features as the dataset and to the single feature as a data point. For examples, in image recognition

tasks, a set of examples is a collection of images that are represented as a matrix of pixel intensity values so the dataset will be the N images each formed by matrix of data points.

For a ML model, gaining experience from data can be described as modifying their output based on the processing of the dataset: Machine learning algorithms can be divided in two classes, supervised and unsupervised, based on the type of information that can be used in this process.

Unsupervised algorithms learn the structure of the dataset to perform task like clustering (i.e. dividing the data points into groups with similar properties) or density estimations (i.e learning the probability distribution of the features). That can be seen as observing several examples of a vector of features $\mathbf{x}$ extracted from a probability distribution $p(\mathbf{x})$ and learn some interesting properties of that distribution.

Supervised algorithms, on the other hand, process examples $\mathbf{x}$ that are associated with a target or label $\mathbf{y}$ and learn the relations between data points and targets aiming to predict $\mathbf{y}$ from $\mathbf{x}$, usually trough the estimation of the conditional probability $p(\mathbf{y}|\mathbf{x})$. The labels that are provided in supervised learning need to be collected from an "instructor" who is able to perform the task that the algorithm is learning. Thus, one major constrain in this strategy is the scarceness of labelled examples especially in fields where the data labeling is a time consuming process and needs a skilled, human operator to be performed.

Last, we need a metric to evaluate how well the task is performed by our algorithm. This choice is critical since it will drive both the learning phase and measure the ability to generalize the task to other data that will be unlabelled in the supervised learning scenario. So the performance metric has to be general enough to be well defined for all the examples in our data yet precise enough to be a straightforward criteria to check if our objectives are reached: if the performance level that we seek is reached then we expect that the task is solved with a reasonable outcome.

Since we are usually interested in the application of our model to data that are not directly included in our training dataset, performance must also be evaluated on examples that our model do not directly use during learning. Thus, a different data set, the test set, will be used to this purpose and the performance measured on this set will be our estimate on well the model will works on the task when is applied in the real world.

In conclusion, to learn from data we sample the training set to adjust the behavior of the model to maximize our performance metric and then we test the performance on the test set. A machine learning methods will be evaluated by its ability to:

- Reach an high performance level on the training examples.

- Make the gap in performance between the training and test set as small as possible.

These two points are the main goals of machine learning and are related to the concepts of *ovefitting* and *underfittig*.

Underfitting is the inability to reach satisfying performance in the given task and can be caused by the wrong formulation of the performance metric or by the limitation of the training set, both in quantity and in quality, if, for example, noisy or missing labels are present. Also, it can be related to the low *capacity* of the model (i.e the length, or deepness, of the chain that we saw in Section 2.1) that is unable to correctly reproduce the relation between labels and data.

Overfitting instead occurs when the model performs much better on training data than testing data. This means that the model is unable to generalize the learned rule to new examples. Overfitting is linked to the capacity of the model in the opposite way respect to underfitting: the model can learn the noise of the dataset in addition (or in place of) the data structure both because the examples are scarce or because the structure of data is simple.

Machine learning model perform better when their capacity is appropriate for the task they are trained for and for the amount of data they are training with. The capacity of the model is usually controlled by the number of the internal parameters used (i.e the weights of the hidden layers we saw in 2.1) and it is linked to its depth (i.e the number of layers). Deep models have high capacity but are prone to overfit when data is limited, while shallow models may struggle to fit the training set.

We will conclude this section with a few examples of application since the exact definition of task, performance and data depends on the particular domain we are interested in.

One of most common task for machine learning is classification: In this task the program has to link an example to the category the example belongs to. The solution to this problem involves the creation of a function that links a feature vector **x** to a numeric code y that represent an element of a finite set of possibilities. Usually the performance that we are interest in for classification is accuracy (i.e the ratio between correctly classified items and total items). As we will see in Section 2.3, the accuracy metric is ill defined to be optimized during learning since the object association to a class function is difficult to differenziate. In this case the usual metric used is the cross-entropy (but many other choices are available):

$$CE(y, \hat{y}) = -\sum_{c=1}^{M} y_{c,o} \log \hat{y}_{c,o} \tag{2.4}$$

Where M is the number of classes, $\hat{y}_{c,o}$ is the network output in the form of probability for the input to belongs to the class c and $y_{c,o}$ is the label. A low cross entropy score will guarantee an high accuracy.

A few examples of classification task are image recognition, where the picture on an object has to be linked to category of the object it contains, segmentation, where a pixel of an image has to be linked to the object or region it is part of, and topics classification for text documents where a piece of text has to be linked to the topics it talks about.

Another interesting task that can be solved with machine learning is denoising. In this type of problems the algorithm is given as input a corrupted example **x** obtained with an unknown transformation from a clean version **y**. The model will learn to restore the original object **y** from the corrupted one **x**. In this case the performance metric we may use is less straightforward since it may be impossible to perfectly restore the corrupted object. We will need to define a distance metric in the feature space between the two objects: for images it is usually the distance between pair of pixels.

In general, we may be interested to remove the effect of the noise instead of the noise itself, for example by making the accuracy in classification for the denoised objects close to the accuracy of the original objects (if noise affects our hypothetical classification task) and we are not guarantee that trying to minimize the distance between the objects will be an efficient way to reach our goal while, using directly our desired metric in learning may be unpractical.

2.3 loss function and training

We discussed how the process of learning involves the improvement of a performance score. This process can be seen as an optimization (i.e. maximizing or minimizing a function) problem. The function that we want to optimize is called the objective function. In the machine learning community it is preferred to reason in terms of minimization thus the objective function is usually called the Error function, the Energy function or simply the Loss function. The performance score and the loss function are usually directly tied but it is possible that the performance is not directly computable (such as the "visually pleasant" concept we saw in the image denoising example) so it can be optimized only indirectly. thus, we reduce another loss function that, when minimized, we hope will also induce an improvement in performance.

In this introduction, when we discuss to the optimization of an algorithm we will refer to finding the argument $\mathbf{x}$ that minimize the Loss function $L(\mathbf{x})$. To be minimized $L(\mathbf{x})$ needs to be a scalar function while its argument $\mathbf{x}$ in general will be high dimensional: $L : \Re^n \longrightarrow \Re$. As usual, the exact form of $L(\mathbf{x})$ will be problem dependent but in general a low value of $L(\mathbf{x})$ will translate to an high performance score for our algorithm. The argument that minimize a function (i.e the solution of our optimization problem) will be indicated as $\mathbf{x}^*$:

$$\mathbf{x}^* = argmin\, L(\mathbf{x}) \tag{2.5}$$

For neural networks model solving the equation directly $\nabla_{\mathbf{x}} L(\mathbf{x}) = 0$ for $\mathbf{x}$, where $\nabla_{\mathbf{x}} : \{\frac{\partial}{\partial x_1}, \frac{\partial}{\partial x_1}, ..., \frac{\partial}{\partial x_i}, ...\}$ is the gradient in the multidimensional input space, is usually impossible thus the method of gradient descend (or steepest descend) is generally adopted. Gradient methods are treated in details in this recent book by Chong (Chong and Zak, 2001).

The steepest descent is an iterative methods in with at each step a new point is proposed by moving in direction of negative gradient

$$\mathbf{x}' = \mathbf{x} - \epsilon \nabla_{\mathbf{x}} L(\mathbf{x}) \tag{2.6}$$

where ϵ is called the learning rate, a positive constant that determine the size of the step. The method of choosing the best learning rate is an open problem but for now we will assume that ϵ is a small constant.

The steepest descent converges when all the component of the gradient are zeros or reasonably close to zero.

When we train neural networks we are interested in the gradient of the loss function respect to the parameters of the model. Lets write the chain operations defined in 2.1 that act on the input $\mathbf{x}$ as $\hat{y} = NN_W(\mathbf{x})$. Where $\mathbf{W}$ indicates generically the set of parameters of all the hidden layers and $\hat{y}$ is the output of the network. The loss function will depends on the output of the network $\hat{y}$ and, trough it, on the set of parameters $\mathbf{W}$. The gradient that needs to be computed is $\nabla_{\mathbf{W}} L(\hat{y}(\mathbf{W}))$.

As we can see in Eq. 2.2 our neural network may contain instructions that are not differentiable in all points making them not suitable for gradient optimization. In practice, gradient descend can still be used and since, in most application, reducing the loss function to a small value that may not be a minimum is an acceptable results we do not expect to reach a point where the gradient is zero in all direction.

Many elements contribute in the learning phase and in the gradient computation. In particular, ReLU units have a large derivative in every points they are active

making easy for them to contribute to the gradient direction that is most useful for learning. Together with the cheap computation cost of both the their activation and derivative this is the reason for the widespread use of ReLU activation.

On the other hand, activation function as the sigmoid $\sigma(z)$ defined in Eq. 2.3 will saturate across most of their domain (when $|z|$ is very large) making gradients very small and learning difficult thus their use is usually avoided in hidden layers of deep networks.

For most application, computing the analytical expression for the gradient is possible but its numerical evaluation is computationally expensive. To solve this problem, the most common employed solution is the *back propagation algorithm* (Rumelhart, Hinton, and Williams, 1986). The Back Propagation algorithm (BP) computes the chain derivative needed to evaluate the gradient with a specific order of operation that is highly efficient.

2.4   Convolutional Neural Networks

In this section we will examine Convolutional Neural Networks (CNNs) that are network specialized to work with input that are organized in grid-like structures. The typical example of this kind of data is images which can easily be represented as a 2D matrix of pixel intensities.

The name convolutional refer to a particular layer used in this network architecture but it is not the only relevant modification respect to the standard fully connected layer: In this section we will review the principal solutions developed to analyze images with CNNs but the results are easily applied to other matrix like data such as computed tomography volumes or, as we will see in Chapter 6, raw data MRI acquisition in the frequency domain.

In neural networks with fully connected layers like the ones described in Section 2.1 every input is connected to every output unit. So, for each pair of layers the connection matrix $\mathbf{W}^{n,n-1}$ will have $L^n \times L^{n-1}$ parameters where L^n is the number of hidden units in the n layer. For very deep networks with this architecture the total number of parameters will be unfeasibly high even for input of modest dimensions.

The solution of specific problems regarding data with grid structure required the development of networks with specialized architectures where the number of parameters is more limited thanks to the definition of specific prior on the distribution of information in the type of data analyzed.

The main idea in the development of specialized architecture is that, leveraging our additional knowledge about the input and the task, we can find strategies that work better for our specific application domain.

2.4.1   Convolutional operation

Given an input function $x(t)$ and a weighting function $w(t)$, also know as the kernel, the convolution between x and w can be defined as

$$s(t) = \int x(a)w(t-a)da \tag{2.7}$$

Where the output function $s(t)$ is usually called a feature map of x respect to the kernel w. We will denote the convolution operation with an asterisk

$$s(t) = (x * w)(t) \tag{2.8}$$

When the convolution operator is applied to grid like data, where functions are indexed by their position on the grid, we will use the discrete convolution operator

$$s(t) = (x * w)(t) = \sum_a x(a)w(t - a) \tag{2.9}$$

The sum goes over all element for which the kernel differs from zero. Images have usually two axis so we can generalize the discrete convolution operator to more than one direction and for practical purposes we can use the commutative property on the convolution operator to write the general convolution for an image.

$$S(i,j) = (K * I)(i,j) \sum_m \sum_n I(i + m, j + n)K(m,n) \tag{2.10}$$

Where the indexes i, j go over the rows and columns of the image.

The convolutional operator can be seen as a matrix operation between the kernel and a small portion of a larger image. Usually, the kernels adopted for image processing in CNNs are significantly smaller than the image they are applied to because we assume that in images the information is local, so an object (or a part of an object) will be made of spatially close pixels. In a convolutional layer the same kernel will be applied to all elements of the input, meaning that the same operation is repeated in the image space connecting groups of close hidden units.

In Figure 2.2 we can see the effect of a linear kernel applied to an image: A kernel K with 2x2 pixel size is applied to an image I, the kernel slides on the image with a S pixels step. The length of the step is called the stride of the kernel. Modifying the stride we can choose the number of output units of the convolutional operation: a stride equal to 1 will produce a convolution output for each element of the input (How to deal with border pixel will be treated later when we will discuss the network architecture) while a larger stride will subsample the input proportionally to its length. The output of the convolution will usually be a linear combination of input elements, the coefficients of this combination (i.e the kernel operation elements) are learned by the network exactly in the same way the weights of the affine transformation in fully connected layers (see Section 2.1) is learnt during training.

The application of convolutional layers allow us to leverage the structure of information in an image: The convolutional operation with small kernel has four important properties that are extremely efficient when working with grid-like data.

- Parameters sharing.

- Sparse weights.

- Equivariance to translation.

- Input size invariance.

Parameters sharing means that the same operation, thus the same kernel parameters, are applied to every portion of the image. This is useful for manly two reasons: First, we expect that a stimulus (i.e a detail in the image that rise the activity in the convolution) may appear everywhere in the image, so the same kernel can be applied to scan the whole input.

Second, sharing the kernel weights means that only $K \times K$, with K the linear size of the kernel, parameters have to be learned by the network reducing both the capacity and the computational cost. Thus, convolution is an efficient method to apply the same local transformation to each region of an input.

The weights are sparse since the convolution is a local operation that works on spatially close pixels. Connections to the next layer comes from groups of units that are close: if an important detail of a image is composed by close pixels, this information can be captured by a local kernel, this is in direct opposition to fully connected layers where every unit is connected to the input. As we will see, this doesn't mean that convolutional architecture can not use information that comes from distant part of the input together, since usually CNNs are composed by many convolutional operation in sequence. Still, the local sparseness of the weights allows to reduce the computational cost and memory requirement of a CNN.

Since the same kernel is applied on all the image, the effect of the same pattern in different location will rise the same output. Together with the pooling operation (see Section 2.4.2), this property makes CNNs able to correctly consider translation as an identity function to its input. Which is that the same pattern translated a few pixel carries the same information. For images, this definitely is a desired property.

Kernels are learned to respond to to specific pattern of pixels that are useful to perform a task, for example the edges of objects in image segmentation tasks, since edges can appear everywhere in the image they may be recognized with the same operation whenever they are present.

Finally, since the shape of the kernel is fixed and the parameters are shared so it can be applied on inputs of different dimensions (with a bit of attention for border pixels) with the same effect.

After a convolution operation, for each unit associated with a kernel, a non linear activation function (see Section 2.1.1) is applied. This stage is usually called the detector stage. For the ReLU operation (Eq. 2.2) this passage can be interpreted as follows: The convolution kernel recognize a pattern of interest and raises its output, positive output are transmitted to the next layer by the ReLU operation.

Convolution operation is only one of the building blocks of a CNN, we will now review the other basic components and then we will analyze how the complete network architecture is composed.

2.4.2 Pooling operation

The convolution operation identifies a pattern of pixels in a small local region in the input. If we are more interested in the presence in an area of this pattern instead of its precise position (at the scale of the kernel size level, so a few pixels), it may be useful to extract a summary statistic of the input over a region.

Pooling operation reaches this scope by replacing the output of the net in a certain location with aggregate information over all the nearby input units. The most used pooling operation is probably the max pooling (Zhou and Chellappa, 24-2) so we will use it as reference example.

Max pooling extract the maximum activation of a kernel over a neighborhood: it aggregates information on a range equal to its linear size. On top of an convolutional layer with a ReLU activation the pooling operation force an invariance to small translation since the resulting activity is the same regardless of which unit in the neighborhood is actually active.

Since the output of a pooling operation is effectively a compression of its input there is the possibility of subsampling the pooling operation without loosing much information. This reduction in the number of active units will help both the creation on a high level representation of the input , since at the next operation more information will be processed with the same kernel size, and the reduction of computational and memory cost. In Figure 2.3 we can see an example of the action of a max pooling

Image Matrix

$I_{1,1}$	$I_{1,2}$	$I_{1,3}$	$I_{1,4}$
$I_{2,1}$	$I_{2,2}$	$I_{2,3}$	$I_{2,4}$
$I_{3,1}$	$I_{3,2}$	$I_{3,3}$	$I_{3,4}$

Kernel Size 2x2

a	b
b	d

Output Matrix

$a\,I_{1,1} + b\,I_{1,2}$ $+c\,I_{2,1} + d\,I_{2,2}$	$a\,I_{1,2} + b\,I_{1,3}$ $+c\,I_{2,2} + d\,I_{2,3}$	$a\,I_{1,3} + b\,I_{1,4}$ $+c\,I_{2,3} + d\,I_{2,4}$
$a\,I_{2,1} + b\,I_{2,2}$ $+c\,I_{3,1} + d\,I_{3,2}$	$a\,I_{2,2} + b\,I_{2,3}$ $+c\,I_{3,2} + d\,I_{3,3}$	$a\,I_{2,3} + b\,_{2,4}$ $+c\,I_{3,3} + d\,I_{3,4}$

—stride▶
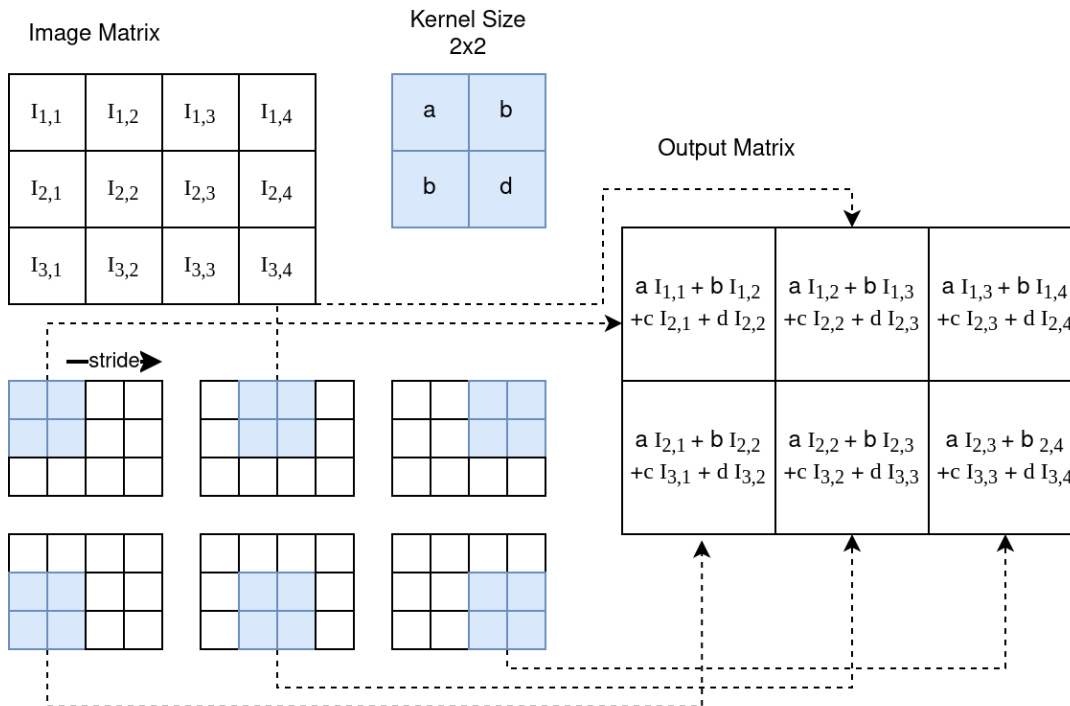
FIGURE 2.2: Example of the application of a convolutional filter that perform an affine transformation. The kernel is applied on the input image and a linear combination of the input elements with the coefficients given by the Kernel parameters is stored in the corresponding output matrix. The kernel is applied at all the position translated by 1 in each direction. This is a schematic illustration of the weights sharing property: the same kernel is applied to all the input image regardless the position

operation on 1D data with and without the subsampling of the output. Max pooling is one of the possible choices of pooling operation, other common alternative are taking the average over a region or the squared sum of the elements of the region. In every case the important concept behind the pooling operation is that in an image (or any grid shaped data) the important information is the presence of one or more patterns and not their position at pixel or unit level accuracy.

Pooling are also useful to create compressed feature representation of the original input by decreasing its dimensionality. In the next section we will see how this behavior is at the base of automated feature extraction of CNNs.

2.4.3 A typical CNN architecture

CNNs have been applied to a multitude of tasks so it is difficult to define a standard architecture that is valid for every implementation. However, when they are used to process images there are some consideration that will be valid most of the time.

As an example we will describe the implementation of the VGG (Simonyan and Zisserman, 2015) architecture that was a popular solution for image classification and many successive approaches to this task took inspiration from it. Besides the particular structure, this network was one of the first to introduce the idea of multiple scalable blocks to develop very deep architecture with many convolutional operations with small kernels that is the base idea of most modern implementations.
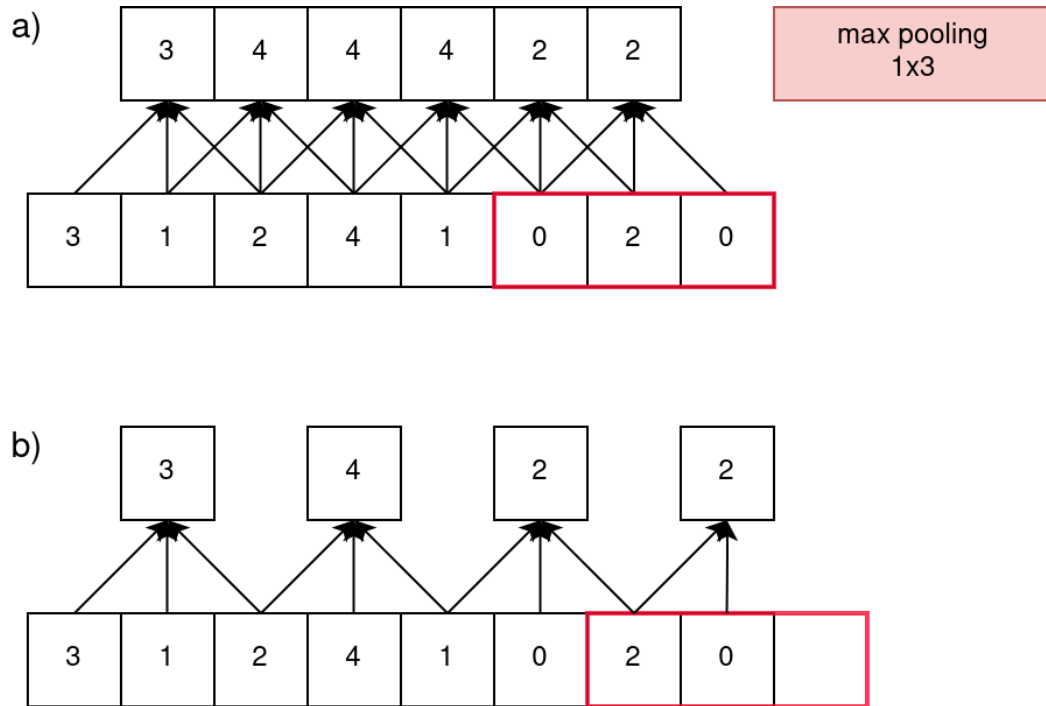
FIGURE 2.3: Example of max pooling operation on a 1D input. The pooling operator select the maximum element in its input. a) Pooling without subsampling: the size of the input is maintained (with the exception of border pixels). b) Pooling with subsampling.

A convolutional layer is usually composed by a set of M convolution operators that perform a linear combination of the KxK (K is the size of the kernel) elements of the input. The number of kernels used controls the number of transformation that a network can learn, the size on the kernel K instead controls how many element are combined together. At this stage the number of parameters P_W of the layer will be (we are not considering the bias.)

$$P_W = N * M * K * K \qquad (2.11)$$

Where N is the size of the input. We can see how the weight sharing dramatically reduce the number of parameters that the network has to learn: In a fully connected architecture the parameters in a layer is the product of the two layers width NxN' that are usually much larger than M.

After the convolutional operations, an activation function is applied to each output element. This is called the detector stage and it is used to introduce the non linearity in the learned transformations.

After the detector stage a pooling operation with downsampling is performed. As we already mentioned, at this stage we reduce the layer size to a factor of the size of the pooling operator to both force invariance to translation and to reduce memory requirements.

This combination of convolutions, activations and poolings is a basic example of a convolutional block.

The CNN is then composed by multiple blocks like the one described above, each of them has a reduced input size due to the pooling stage. This structure is called the contractive or subsampling path. In a subsampling path the input dimension is

continuously reduced allowing move convolutional layers to fit in the same amount of memory. The main reason behind this strategy is to help creating more complex representation using a more compressed version of the input at each step.

At each stage the kernel will have the same size but, since it will process input already processed by convolutional and pooling operation, its output will depend on a large number of pixels. The area of the input that directly influences the output of a unit in a convolutional layer is called its receptive field. The name is inspired from biological networks as we will explain below. The receptive field grows with the depth of the network giving the network the possibility to process more contextual information since the value of a units is directly function of a larger area of the input.

After one or more convolutional blocks the output of the network is then flattened (i.e. all the elements are represented as a vector regardless the original shape) and after one or more fully connected layers this output is passed to a classifier. The flattened vector of the last convolutional block is usually called the feature vector because it will be the high level descriptor (so a vector of feature) of the original input and that will be the input to the classification task. The name came from the standard machine learning algorithm for image processing where the input for the classification are feature of the image extracted by hand crafted filters.

In Figure 2.4 is reported an example of a CNN architecture with the same basic structure described. The network operations are computed from left to right: First, the input a matrix of 128 by 128 units, that can be the data analyzed or the output of the previous convolutional block, is processed by a series of 8 convolutional operators. These convolutional maps are pooled with a subsampling of 2 so their dimensions is reduced to 64 by 64. Then, another convolutional stage with 24 filters is a applied with a kernel with stride 2 and, after the pooling operation, the output is reshaped as a vector (the feature vector) and used as input for a two layer fully connected classifier.

This architecture may not work well in practice but it is a good exemplification on how the goal of classification of images using information from a large patch (or with a large receptive field) is reached by stacking convolutional blocks with small and simple kernels instead of developing large and complex kernels.

Modern versions of the CNN have moved away from this linear form with a single subsampling path and often exhibit more than one convolutional path that process the details of the input with a different level of sampling. These paths are called branches. One of the best examples is the U-net (Ronneberger, Fischer, and Brox, 2015) that combine together multiple branches at different level of details.

Probably the best way to build an intuition an how a CNN works is to make a parallel with the brain visual cortex. This is not a rigorous comparison, but at least in the early development stages, CNNs were inspired by the mammal vision system.

Modern neural networks for computer vision application differ profoundly from the computational model of brain function they derive from, but we can still gain some insights of the their general structure in terms of biological networks.

The first layers of a CNN can be thought to play the role of the primary visual cortex (V1) that is the first step of the visual input processing in the brain. The image is first collected on the retina and then transported trough the optical nerve and the Lateral geniculate nucleus to the visual cortex. Here the V1 process the signal and passes it to the other areas of the visual system. V1 is a two dimensional structure that map the activation of the retina (i.e visual stimuli captured in the same area of

the retina will be collected in the same area of V1). This is our first similarity with the convolutional operation: the activation of the convolutional operation, that produce the image feature, captures local properties of the input.

Then in V1, specialized cells called simple cells, that are neurons that respond to stimuli with particular orientations to objects within their receptive field, process the signal. Their action is similar to the detector stage in the CNN.

In V1 are also present complex cells that will respond primarily to oriented edges and gratings, similarly to simple cells but their function has a degree of spatial invariance like the pooling stage that we defined for the artificial network.

The visual cortex presents a hierarchical structure, the V1 area is strongly linked trough feedforward connection to a second area, the secondary visual cortex (V2), also called prestriate cortex, where cells are tuned to simple properties such as orientation, spatial frequency, and color. The function of V2 is close to V1 but the neuron response in V2 modulate more complex properties of the stimulus such as contours, shapes and and whether the stimulus is part of the figure or the background. We can think of the connection between V1 and V2 as the successive stacking of convolutional blocks where at each step the input is processed with larger receptive field that produce more complex representations.

V2 sends strong connections to the successive steps of the visual processing pathway that we will simply indicate as V3, V4, and V5 but it also sends strong feedback connections to V1.

As we already discussed, in feedforward artificial neural network, the feedback connection are not implemented. This is the most crucial difference in our parallel that invalidates any serious relationship between the two processes.

In any case, we will go a bit further in the analogy to compare the feature vector of the last stage of a CNN to a brain area called the inferotemporal cortex (IT). The IT, that can be taken as one of the last steps of the visual processing, is associated with the representation of objects, places, faces, and colors. It may also be involved in face perception and in the recognition of numbers and symbols. We can say that the activation of its neurons respond to complex information and not just geometrical or statistical ones.

The features extracted by deep CNNs may have the same role: They encodes the content of an image instead of just its description in terms of simple properties.

Even if, as we have seen, the analogy between brain and artificial networks is at least naive, it is still useful to report it both because it helps in the understanding of artificial neural networks and because the study of biological networks can lead to the introduction of useful solutions in the development of computer vision.

The CNN described here is one of the most basic implementation of the concept and it is probably not used anymore in real applications. This is not a problem since our objective is not making a survey of CNN architectures but to highlight the motivation behind their success.

CNNs architecture perform extremely well in their field of application because they are specialized architectures that exploit our prior knowledge on the data we are applying them to. They are useful to recognize spatial patters in the input regardless of their exact position and to integrate information from a local area while retaining the ability to process together the information extracted in different parts of their inputs.

We can apply the same architecture where the same properties of the input are valid regardless of the nature of the input. For this reason, convolutional network are also employed in time series analysis, in sound processing or in video processing

that are data that share the possibility to be well represented by locally extracted high level features.
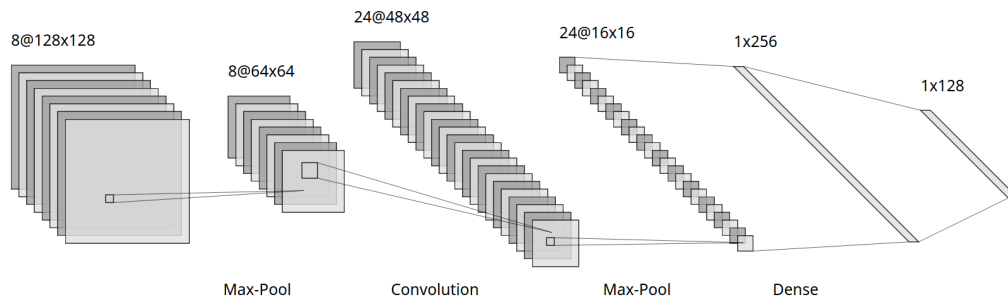
FIGURE 2.4: Schematic description of a convolutional network with convolutional and pooling operations. After the second pooling operation the output is flatten to a an array that is passed to fully connected network.

2.4.4 The fully convolutional architecture

We will now discuss a particular class of CNNs where the output is structured as a map that has the same shape as the input. This architecture is called Fully Convolutional Neural Network (FCNN) and it was first introduced for the task of image segmentation (Long, Shelhamer, and Darrell, 2015). The segmentation task can be described as the assignment of each pixel of the input to the class of the object it is a part of. So we can say that he FCNN uses convolutional operations to transform image pixels to pixel categories.

A simple example can be the separation of white matter zone from gray matter in a brain image.

Unlike the convolutional neural networks previously introduced, a FCNN transforms the height and width of the intermediate layer feature map in the subsamplig path back to the size of input image through the transposed convolution operation (i.e when a unit is mapped to a larger patch trough a kernel operation), so that the predictions have a one-to-one correspondence in shape with input image. Thus each pixel in the original input will be associated to one output unit.

As opposed to CNN with a full contractive path, in FCNN is possible that each output unit depends only on a part of the input. This area is called, as is the case of intermediate layers of a CNN, the receptive field or the effective patch size of the network. Every pixel outside this area will not contribute to determine the unit output.

The receptive field width is particularly interesting in case there are no pooling operations. In this case the ability to process larger area depends only by the deepness of the network. The receptive field, from the initial size of the kernel K grows by the stride S of the operator in each direction for each consecutive layer. Thus after D layer the effective patch size is

$$P_s = K + 2S(D-1) \tag{2.12}$$

In Figure 2.6 we can see an example for a network with a kernel of size 3 by 3 with a stride 1 and depth d that will have an effective patch width of $2d + 1$.

This kind of architecture has a relative small receptive field but it retains the spatial information at the same level of detail that is present in the input. This is possible because of the lack of pooling operation which helps to improve translation invariance but at the same time makes the output less dependent on the exact spatial position. For task in which the definition and sharpness of the output is an important quality, this network will perform extremely well even with small details at the cost of loosing the general context of the image.

Thus, when a greater area is needed two or more branches with different level of pooling are combined to increase the effective patch size of the output layer without disregarding the spatial information of the input. The more used solution of this type is the U-net (Ronneberger, Fischer, and Brox, 2015) that find numerous application in segmentation tasks where small details in the border between two areas need to be maintained.
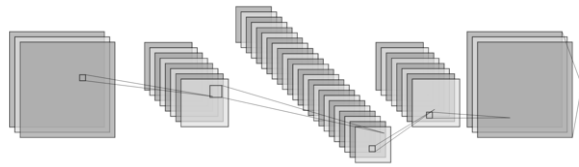
FIGURE 2.5: Example of a fully convolutional architecture with a symmetric downsampling and upsamplig path.
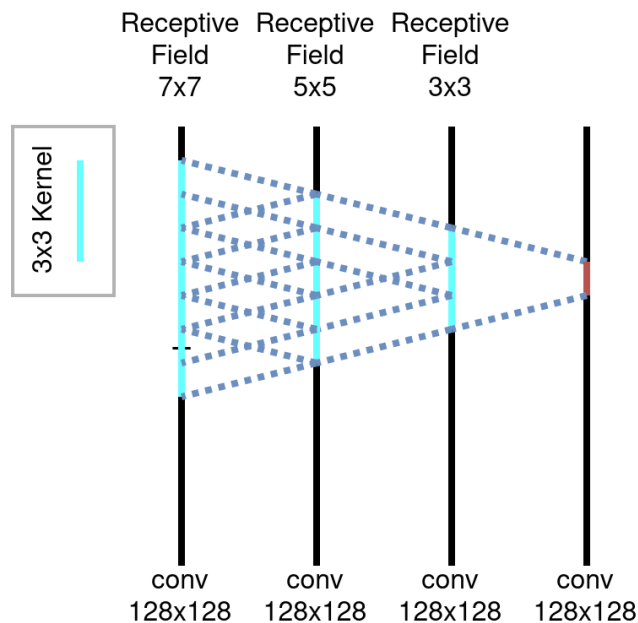
FIGURE 2.6: Receptive field in a CNN without pooling operation with a kernel size of 3 and stride 1. The layer operations are performed left to rigth. A unit in the output layer (Right) will depend on 7 by 7 pixel in the first layer. The size of the receptive field is given by $2D + 1$ where D is the deepness of the network.

2.4.5 Batch Normalization

The efficient use of parameters in CNNs allows the development of very deep architecture with a relatively low number of parameters respect to other solutions.

While this strategy helps deriving high level representations it also introduce problems in the application of gradient learning techniques. As we saw in Section 2.3 the gradient descend algorithm compute the forward pass sequentially, the output of each layer, given an input, is computed in series feeding to each layer the input on the precedent one. The learning step, instead, is computed and the gradient applied to all the parameters at the same time (Eq. 2.5) so that all the layers are updated concurrently. This situation may generate a strange behaviour since we update layers under the assumption that all the other layer in the chain remain constant and often it is not the case.

To study this effect we can build a toy network of the type examined in Section 2.1 but with only linear operation, with one unit per layer (so one parameter) and no bias. The chain function, of length L, that this toy network computes can be written explicitly as:

$$NN_{\mathbf{W}}(x) = \hat{y} = x * w_1 * w_2 * w_3 * w_4 * ... * w_L \tag{2.13}$$

Where $\hat{y}$ is the network output and $*$ denote a multiplication between scalar since layers, in this example, have only one value. When the network compute the forward pass, each layer hidden unit activation h_i is computed in sequence: In this example we have that $h_i = h_{i-1}w_i$.

In the toy example layer operations commute but this is not true in general. The network function is linear $NN_{\mathbf{W}}(x)$ in the output x but it is not linear in the elements of the weight vector $\mathbf{W}$ (if we change more than one element). After we compute our Loss function (an identity function for our toy model) we can use the BP algorithm to compute the gradient and then apply it. As usual ϵ denote the learning rate.

$$\mathbf{g} = \nabla_W \hat{y} \tag{2.14}$$

$$\hat{y}' = \hat{y} - \epsilon \mathbf{g} \tag{2.15}$$

$$\hat{y}' = x(w_1 - \epsilon g_1)(w_2 - \epsilon g_2)...(w_L - \epsilon g_L) \tag{2.16}$$

In this simple example we can see how a learning step can be influenced by higher order terms such as $\epsilon^2 g_1 g_2 \prod_{l=3}^{L} w_l$ that can be extremely large, depending on how the weights are distributed, thus making difficult to move in small steps.

In this simple model we can explicitly compute the second and higher order effects in the step computation and modify the learning accordingly but in general is more straightforward to add an explicit normalization for the activity in the network operation so when we backpropagate trough them they will compensate the increasing in mean and variance of the hidden layers given by the gradient.

This is the basic idea of Batch Normalization (BN) (Ioffe and Szegedy, 2015) that reparametrize the activity of each unit in a layer with its z-scored value computed over a batch of examples. The reparametrization is performed by computing the average activation and its standard deviation of a hidden layer $\mathbf{H}$ over a batch of b examples

$$\mu = \frac{1}{b} \sum_i \mathbf{H}_i \tag{2.17}$$

$$\sigma = \sqrt{\frac{1}{b}\sum_i (\mathbf{H}_i - \mu)} \qquad (2.18)$$

And then the normalization is applied at units level

$$\mathbf{H}' = \frac{\mathbf{H} - \mu}{\sigma} \qquad (2.19)$$

In this way BN assure that units are always standardized by definition reducing the effect of large, unbalanced weighing in the gradient step. Usually BN is applied after a series of convolutional block to keep the block activation under control. In Chapter 6 we will implement batch normalization to stabilize the learning in a deep fully convolutional architecture to help convergence during training.

2.5   Practical remarks and training strategy

We have reviewed the main ideas behind deep learning as a powerful framework for supervised learning. In contrast to traditional machine learning in which simple parametric models are implemented, we saw how deep learning is used to approximate functions of increasing complexity by adding more units and layers to networks.

This strategy allows the construction of machines that can learn very complex tasks from example with data that is extremely complex to represent in terms of simple features.

Most tasks that we overview consisted in mapping an input vector or multidimensional matrix to another one and this is where these methods produce the best results, given that sufficiently large capacity and a sufficiently large dataset of labeled examples are provided.

This introduction had focused on the approaches that are already applied at industrial level and are essentially working technologies, even if there is yet open research in the field, that are now being implemented outside their usual scope on problems that can be addressed with the same strategies such as automated analysis on clinical images in the medical field.

Still, the difficulty of the application of a deep learning model should not be underestimated since a lot of work has to be devoted to the preparation of the dataset and to studying the performance that the method obtains, in particular compared to the alternatives that are normally used and are not based on machine learning.

Even if the main perk of deep learning is the automated extraction of features that allow the processing of raw data in many applications we must not think that this means that a profound knowledge of the data is not necessary to build a reliable machine learning application.

The success of convolutional networks is a clear example of how building data driven models with the correct prior will produce better architecture that needs less data to be trained and can better generalize the learned task.

Also, even if we have available a clever strategy to train a model with a large number of parameters with a few standard approaches that may work in most of the cases, there is still a large number of hyperparameters, that are every parameters that is not in the weight matrix such as the learning rate, and many arbitrary choices that have to be made , and that need extensive testing and fine tuning.

Finally, reaching a good level of performance is critically dependent on the amount and quality of the data available. A deep learning solution can be easily applied

when a large dataset of correctly labeled examples is already organized but when data is lacking or worse it is conflicting, the implementation of DL may be an impossible task.

Often the large part of the effort in the development phase is absorbed in curating the dataset and this aspect must be considered when weighing the benefits of applying a machine learning method in substitution of a different approach.

After the development phase we usually aim to deploy our solution to the task, thus we want our deep learning methods to be general enough to be applied outside the research facility. This needs lots of fine tuning to assure it will maintain stable performance in situations where the input may deviate a bit from the curated data used for training and testing.

This is particularly true in tasks where critical decisions are made based on a machine learning model prediction or when the operator is not able to check and eventually refuse the predictor output. Unfortunately, this is exactly the case of the medical field so particular attention at the output stability should be made.

All these considerations must be made before deciding to solve a problem with deep learning. In particular proposing a machine learning solution for a specific problem where data are scarce or they are not easily obtained, can deeply impact the chances of success and can ruin the credibility of the proposer and the mental well being of the developer.

For these reasons, we will first test the performance of our new proposed denoiser methods on simulated data that is easily accessible and then we will try to validate it on already collected data from an open dataset.

If these tests will be considered passed we may obtain an estimate of the data needed to train our methods in real conditions and we will decide if the deep learning approach is reasonable in denoising magnetic resonance imaging of eterenuclei acquisitions or not.

Chapter 3

Introduction to Magnetic resonance

3.1   NRM Spectroscopy for biological samples

The use of NMR to analyze the content of a biological sample was first used in 1977 by Brown et al. (Brown, Campbell, and Kuchel, 1977) that shown that 1 H signals could be observed from a range of compounds in a suspension of red blood cells. A great deal of metabolic information can be derived from such metabolic studies and it was soon recognized that 1H-MRS has a considerable role to play in areas of pharmacology, toxicology and drug discovery (Ross et al., 2007). More recently, the presence of fluorine in a significant part of newly approved drugs made 19F-MRS an important analysis and quantification tool for both in vitro and in vivo application to biological samples (Mattes et al., 2016).

3.1.1   NMR principles

We will present a phenomenological description of magnetic resonance needed to understand the MRS quantification in cells and tissues that are reported in Chapter 4 and 5.

   This section is meant to be an overview on the application of NMR to metabolic profiling of biological samples. It is not meant to be a theoretical treatment of the matter so not only a semi-classical description on NRM will provided but also a few detail on data acquisition and processing in an experimental setting.

   To this end, the physical phenomena, the hardware needed, the acquisition protocol and the preprocessing of the signals will be briefly treated.

   While in this section the discussion of the phenomena will be general and we will often refer to proton spectroscopy (1H), our main interest is in fluorine (19F) spectroscopy for which both the challenges and the application are different. Thus, in section 3.1.2 we will analyze in more details those differences.

   A first-principle formulation based on quantum mechanics is available but it is beyond the scopes of this introduction and can be found, for example, in (Goldman and Webb, 1989; Ziessow, 1988).

Semi-classical NMR

When atomic nuclei are immersed in an external magnetic field and when an electromagnetic radiation with specific frequency is applied, a resonance transition between magnetic energy levels can be observed. NMR is based on this physical phenomenon. By detecting the absorption signals, one can acquire an NMR spectrum. According to the positions, intensities and fine structure of resonance peaks, people can study the structures of molecules quantitatively. In NMR spectroscopy (MRS) molecules under investigation are exposed to an external magnetic field B_0 that is

also called the *longitudinal* field. We will assume that the direction of this field is along the z axis. At equilibrium, the magnetic moments of the molecules in the sample align along B_0 accordingly to a Boltzmann distribution.

To produce the NMR signal, the total magnetization is flipped orthogonal to B_0 using an high frequency magnetic field B_1 orthogonal to B_0 applied for a fixed time interval. This is called the radio-frequency pulse (RF-pulse) or the $B1$ pulse. The total magnetization $\vec{M}$ now is orthogonal to B_0 and will precess with a frequency f_0 know as its resonance given by:

$$f_0 = \frac{\gamma_X}{2\pi} B_0 \tag{3.1}$$

It is important to notice that the resonance frequency depends on the gyromagnetic ratio, that is a property of the molecule, and on the external field, that is under experimental control.

Thus, NMR spectrometers are usually indicated almost interchangeably with their magnetic field strength in Tesla (T) or with the resonance frequency of the hydrogen atom (1H) at the provided field strength. For example, given the gyromagmetic ratio of the hydrogen:

$$\gamma_{1H} = 42.58 \frac{MHz}{T} \tag{3.2}$$

A spectrometer operating at 9.4 Tesla will be called a 400 MHz and one operating at 14.1 T will be called a 600 MHz. Other nuclei will rotate in the same field at other frequency depending on their gyromagnetic ratio. The fluorine atom gyromagmetic ratio is $\gamma_{19F} = 40.05 \frac{MHz}{T}$ so it will resonate at 376 MHz in the 400 MHz (or 9.4 T) spectrometer.

Detection of signal

Following the schematic representation in Figure 3.1 a sample (i.e a glass vials with a resonant molecule in solution) is inserted in a detection coil and it is immersed in the static magnetic field B_0. The processing magnetization produced by the B_1 field (the RF pulse) will induce a voltage U_{ind} modulated with the molecule resonance frequency f_0. The amplitude of U_{ind} is directly proportional to the magnetic moment $\vec{M}$ and thus with the number of spins rotating with the same resonance frequency f_0 located inside the coil volume. This signal is called "Free Induction Decay" (FID).

In a molecule, the local magnetic field, that influences a spin, deviate slightly from the external field B_0 due to the effects of its electron clouds. Thus the differences in frequency measured between the spin f_i and the reference spin f_0 is an indication of the electronic and chemical neighborhood of the nucleolus observed. This relative difference measured in parts per million (ppm) respect to B_0 is defined as the *chemical shift* of the spin. Spectral distances (i.e differences in frequency) measured in ppm are independent from B_0 and can be compared between different acquisitions at different field strength while spectral distances measured in Hz will depends on B_0. Different spins species of the same atom have different ranges of chemical shift: 1H spins resonate within 15 ppm while 19F has a larger chemical shift range of 700 ppm. Usually the reference frequency f_0 is derived from the frequency of a reference compound that contain the atom of interest. The choice of this reference compound is linked to experimental necessity and may vary but for each application of MRS it is limited to a small number of standard molecules (Rosenau et al., 2018).
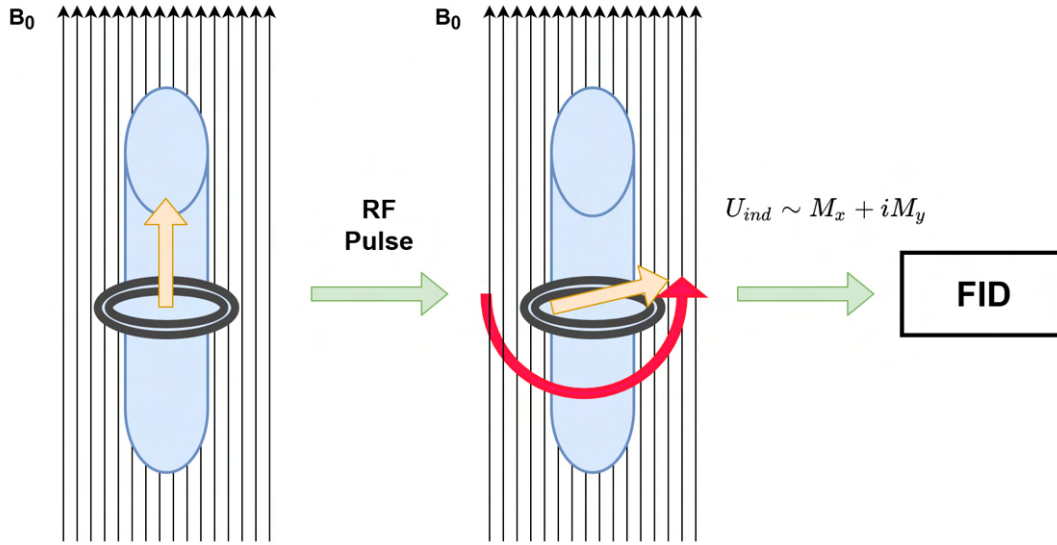
FIGURE 3.1: Detection of NMR signal: (left) a sample with many spins is immersed in an external field B_0 oriented along the z-axis. At equilibrium the total magnetization (Orange arrow) is parallel to this static field. The sample is enclosed in a detection coil (Black). After the application of a RF pulse (center), the total magnetization is now parallel to the x,y-plane and it starts to precess around B_0. The NMR signal is the voltage induced U_{ind} in the coil by the varying magnetization.

Relaxation

At the thermodynamic equilibrium, the the total magnetization is parallel to the external field so that $\vec{M} = M_0\hat{z}$. The effect of a resonance RF pulse (or B_1 pulse) is to disturb the spin system out of its equilibrium. When the perturbation stops, in time, the equilibrium will be restored by a process known as the spin-lattice relaxation. This process involves an energy exchange between the spin system and the surrounding thermal reservoir (the lattice). Since this process restore the state of the magnetization that is parallel to the longitudinal field, it is also called longitudinal relaxation. The equation that describes this longitudinal relaxation is:

$$\frac{dM_z}{dt} = -\frac{M_z - M_0}{T_1} \tag{3.3}$$

With solution:

$$M_z(t) = M_z(0)e^{-t/T_1} + M_0(1 - e^{-t/T_1}) \tag{3.4}$$

The time constant of this process T_1 is known as the spin-lattice or longitudinal relaxation time. At room temperature for a biological sample it is in the range of 0.1 to 10 seconds.

Spins also interact between themselves giving rise another relaxation process that controls the lifetime of the transverse magnetization. This process is known as the spin-spin or transverse relaxation. Phenomenologically the equation for transverse relaxation is written as:

$$\frac{dM_{x,y}}{dt} = -\frac{M_{x,y}}{T_2} \tag{3.5}$$

With solution:

$$M_{x,y}(t) = M_{x,y}(0)e^{-t/T_2} \tag{3.6}$$

The time constant of this process T_2 is known as the spin-spin or transverse relaxation time. In biological samples usually $T_2 << T_1$ and it is in the range 10 μs to 1 s. This exponential description applies only when the interaction terms responsible for transverse relaxation are weak. This regime is called Bloembergen, Purcell, and Pound (BPP) and it is valid for spins residing in liquid state molecules (Callaghan, 1991). For our NMR applications we will always assume to observe slowly relaxing spins for which the this phenomenological approach is appropriate.

Bloch equation

If we combine Eq. 3.3 and Eq.3.5 in the rotating frame we obtain a set of relationships known as the Bloch equations.

These equations provide a valuable reference for the phenomenological description of NMR.

$$\frac{\partial}{\partial t} \begin{pmatrix} M_x \\ M_y \\ M_z \end{pmatrix} = \gamma_X \begin{pmatrix} M_x \\ M_y \\ M_z \end{pmatrix} \times \begin{pmatrix} B_x \\ B_y \\ B_z \end{pmatrix} - \frac{1}{T_2} \begin{pmatrix} M_x \\ M_y \\ 0 \end{pmatrix} - \frac{1}{T_1} \begin{pmatrix} 0 \\ 0 \\ M_z - M_0 \end{pmatrix} \qquad (3.7)$$

Suppose that a coil is placed around the sample as in figure 3.1 transverse to the static field B_0. In the laboratory frame any transverse magnetization precessing at the larmor frequency will produce an induced voltage oscillating at ω_0 frequency.

The amplitude of the signal is in the order of μV and it will be proportional to the transverse magnetization vector present inside the coil.

The RF receiver works by mixing the induced signal with the output from a reference radio frequency oscillator. This process is called heterodyning and it is used to obtain a phase sensitive measure with a single receiver.

Thus, by separately mixing the signal with two heterodyne oscillator each out of phase $\pi/2$ we obtain two output signals that are proportional to the orthogonal phases of the magnetization so in practice we are measuring M_x and M_y.

When the mixing reference is oscillating at ω_r any the output signal will oscillate at the offset frequency $\Delta\omega = \omega_0 - \omega_r$.

Conventionally the output signal is represented as a complex number in which the real part is M_x and the imaginary part is M_y.

Consider an experiment in which a RF pulse that flips of $\pi/2$ the spin magnetization is applied to a system in equilibrium with the field in which the magnetization **M$_0$** is along the z-axis. In this case the total magnetization in function of time can be written following Eq. 3.7 as

$$\mathbf{M}(t) = [M_0 \cos{(\omega_0 t \mathbf{i})} + M_0 \sin{(\omega_0 t \mathbf{j})}] \exp\{(-t/T_2)\} \qquad (3.8)$$

In the complex notation it becomes

$$\mathbf{M}(t) = M_0 \exp\{(i\omega_0 t)\} \exp\{(-t/T_2)\} \qquad (3.9)$$

So the output signal at the offset frequency simply is

$$S(t) = S_0 exp(i\phi) \exp\{(i\Delta\omega_0 t)\} \exp\{(-t/T_2)\} \qquad (3.10)$$

Where ϕ is the absolute phase of the detector and S_0 is the initial amplitude which is proportional to M_0.

This time domain oscillating signal induced by a free precession is the FID.

Limit of detection and noise

The detected current in the coil is proportional to the total magnetic moment of the sample.

Thus, the signal depends linearly on the number of atoms N in the sample and the difference in population between spins aligned parallel with the B_0 field and those aligned anti-parallel. According to the high-temperature approximation of Boltzmann's law this difference increases linearly with B0.

In the signal term, there is an additional B0 dependence in the detection process itself (the energy provided by any spin flip, so the difference in energy).

Other important factors influencing the sensitivity of the measurement are the gyromagnetic ratio γ_x, the sensitivity of the detector and the noise created by the sample and the electric resistance , the latter summarized in S_D.

The sensitivity defined as the signal-to noise ratio (SNR) of a single repetition of the acquisition is given by:

$$Sensitivity_{Scan} \sim \gamma_x^3 \cdot N \cdot B_0^2 \cdot S_D \tag{3.11}$$

There exist two approaches to increase SD: First, we can reduce the contribution of the circuit and losses. Noise in electric circuits is caused by thermal motion of electrons in wires (Edelstein et al., 1986).

Consequently it is proportional to the temperature and it can be reduced by cooling the wires of the detection coil and the preamplifier down to cryogenic temperatures of a few K. Such probeheads are called cryogenic or cold-probes. They are unfortunately expensive and specialized hardware.

In addition S_D can be increased by reducing the size of the detection coil since Jonson noise is also linearly dependent on the resistance it is clear that shorter wires and smaller coils create less noise.

In biological samples usually the concentration is given by the biology of the system and it can be extremely low so using a small volume may be disadvantageous since less resonant atoms will contribute to the magnetization.

In this case usually the strategy is to concentrate the metabolites in smaller volumes during sample preparation for example by the lyophilization and reconstitution of the sample. A specialized guide for biological sample preparation can be found in (Kostidis et al., 2017).

Also, A sample of high conductivity in which many solutes are present reduces the sensitivity of the detection by induced eddy-currents. This situation is often encountered for biological samples having high ionic strength like blood or urine.

From Equation 3.11 it is also clear that a higher static magnetic field does increase the sensitivity of the NMR detection. This is one reason why magnets with higher B_0 fields are attractive.

If the same experiment is repeated by averaging the detected signal, it is important to realize that the signal sums coherently (proportional to the number of repetitions called scans (NSA)). While statistic averaging leads to a decrease of noise linear in the root of NSA.

Thus increasing the SNR by two enforces a fourfold increase in experimental time. Another important factor influencing the outcome is the rate of repetition of scans (TR).

A slow rate, that is long inter-scan delays, allows for full relaxation of the magnetization. This ensures that the area of the resonance peaks is perfectly linear respect to the concentration of the analytes. The trade-off in this regime is a long experimental time for a given number of repetitions.

This aspect is important when we are interested in the quantification of a molecule in the sample and will be discussed in Section 3.1.2.

Since biological concentration of many analytes of interest may be low, easily in the order of 10^{-8} Molar, many strategies to improve the SNR and detection for metabolomic analysis have been developed in NMR. This topic is discussed in depth in chapter 3 of the book by Ross et al. (Ross et al., 2007).

Elements of signal processing

The spectrometer coil detects magnetization along the x-direction and the y-direction over time.

The detection is done by a measurement of induced voltages with a single coil using two-channel phase-shifted high-frequency mixing.

For any time-point, two values M_x and M_y are recorded, this technique is called phase-sensitive detection.

The NMR signal is a complex valued time domain signal constructed as $M_x(t) + iM_y(t)$ that is transformed to the frequency domain through a fourier transform (FT).

The fourier transform of the FID signal will show individual lines with positions given by the chemical shifts. The width of the lines is characterized by $1/T_2$ and shows a so-called absorptive lorentzian line-shape.

This relation holds if all magnetization flipped in the x-y plane is perfectly aligned along the x-axis at the beginning of the acquisition. Because of the imperfect strength calibration of the RF pulse and due to limitations of spectrometer electronics this is not true in practice and the spins are not perfectly flipped so resonances are not perfect lorentzian.

Initial magnetization is oriented under a small initial and frequency-dependent phase with respect to the x-axis.

As a consequence, the real part of the FT will also contain a frequency-dependent contribution of the imaginary part of a perfect signal. This imperfection can be removed after FT by a process called phasing.

Thus, the phase shift also varies with frequency and, in FT NMR spectroscopy, the real output of the FT is taken as the frequency domain spectrum so the NMR spectra require both constant and linear corrections to the phasing of the Fourier transformed signal.

$$\phi' = m \cdot \phi + b \qquad (3.12)$$

Constant phase corrections, b, arise from the inability of the spectrometer to detect the exact M_x and M_y. Linear phase corrections, m, arise from the inability of the spectrometer to detect transverse magnetization starting immediately after the RF pulse.

A magnitude spectrum might occasionally be used in some applications. The magnitude signal is equal to the square root of the sum of the squares of M_y and M_x. Magnitude spectra can be useful when phasing correction is difficult since it discards the phase completely.

Another problem that can appear in NMR spectra is a distortion of the baseline due to imperfections and non-linearities of the electronic detection process.

This non flat baseline can be corrected by the subtraction of a polynomial function from the raw spectrum obtained. This correction can be obtained automatically by a fitting procedure of the spectra after the removal of the points relative to the resonances.

Phasing and baseline correction can be performed automatically but in some occasions a manual procedure can be advantageous. Thus, Visual inspection of processed spectra for artifacts is important before the analysis of the data is performed.

The signal to noise and line-shape of spectra can be adjusted by an apodization (i.e. the time-domain signal is multiplied with a window function) prior to FT.

In metabolomics, the window function used for this purpose is typically the so-called exponential-window given by $\exp\{-LB \cdot t\}$ employed with a line broadening factor LB of 1 - 3 Hz. Since the first point of the FID is not modified the integral areas of the peaks of the spectra do not change after this filtering is applied.

The LB weighs the points in the FID prior to the FT. An exponential weight makes more important the first points recorded, limiting the effect of the noise at the expense of spectral resolution.

An LB of 1 - 3 Hz offers an acceptable compromise between signal-to-noise and spectral resolution.

Since the integral of an NMR signal in the spectrum (not its amplitude) is linear in the number of NMR active nuclei present in the sensible volume of the coil, the concentration of this molecule can be determined from this integral.

Comparison of integral values with a reference compound allows an absolute value determination of concentration of the selected molecules in the sample.

Quantification will be discussed more in detail in the next Section.

3.1.2   19F MRS quantification

Magnetic resonance spectroscopy (MRS) can be used for quantification purposes such as the determination of the content of a drug and its impurities when. This technique has many applications in pharmaceutical R&D and it is also used for metabolites quantification and profiling in plasma and excreta (Holzgrabe, 2010). In particular 19F-MRS can be used to track fluorine labeled compound and it has been proposed as an alternative to radiolabeling in mass balance studies (James et al., 2017; Athersuch et al., 2010). Also, it has been used to quantify fluorine labeled T-cells in tissues samples in ex vivo experiments (Chapelin et al., 2017). Those application are possible because, with the correct experimental setup, the area of a resonance peak in the 1-dimensional spectrum is proportional to the number of nuclei giving rise to it .

To be used in a quantification experiment, the acquisition and processing parameters have to be set carefully. The first concern is to obtain complete longitudinal ($T1$) relaxation: The repetition time (TR, the time interval at which the pulse is repeated) has to be set taking into account the longitudinal relaxation time $T1$ of the signal used for quantification. Relaxation in $T1$ can be described as : $M_z = M_0(1 - \exp\{-TR/T1\})$ with M_0 and M_z are respectively the magnetization along the z-axis at equilibrium and after the repetition time. To obtain a signal that is proportional to the equilibrium magnetization, one should set $TR >> T1$. When in the tested sample there are multiple sources of signal, in theory one for each atoms in each molecule, T1 should be measured for all of them and TR should be set accordingly.

Usually to determine T1 an inversion recovery pulse sequence is used (see Section 5.2.5) and when the tested samples are in solutions with different solvents, a measure should be repeated for each solvent since T1 can be different in different solvents.

Quantification can be achieved by means of relative or absolute methods, using either internal or external standards: a reference signal given by a know source can be used to normalize all other unknown signals.

For the purpose of this experiment the choice of this standard reference should follows a few considerations(Giraudeau et al., 2014; Cullen, Ray, and Szabo, 2013) : First, the reference peak should not overlap any those of the injected fluoridated compound and any of its metabolites. Second, the signal of the reference should not be exceedingly larger than the expected signal of the unknown molecules to avoid the modification of the receiver gain for lower and higher concentration samples that might influence the stability of the measure . Third, the reference should be in the same chemical shift region of the tested molecules to avoid off resonance effect and the pulse excitation should be uniform in the spectral width of interest. This is particularly important for the fluorine spectra because of its wide chemical shift range. Finally, the reference should be chemically stable for the duration of the experiment and should tested for impurities to check its stability.

3.2   NMR imaging

We saw how the NMR signal is detected by the receiver, and it is simply a voltage induced in the coil. In this Section we will discuss how spatial information needed to form an image is encoded in this signal.

The NMR signal is centred at a resonance frequency that is given by the strength of the static field , but it contains a range of different frequencies that encode information about the location in the scanned volume.

This centre RF frequency is removed digitally from the signal leaving data organized in a bandwidth, which is typically several kilo hertz (kHz) wide.

This signal is acquired in frequency so most of the signal-to-noise and contrast information is in the low frequencies, while the higher frequencies contain information about resolution in the image.

Electronic noise can be thought as distributed evenly across the whole bandwidth so High receive bandwidths have a worse signal-to-noise ratio than low receive bandwidths simply because they include more noise.

This data is usually called the k-space acquisition of the MRI. From this matrix the final image is reconstructed through a fourier transform.

3.2.1   Gradients and sequences

In MRI the static magnetic field B_0 is constantly present. The gradients are applied in a controlled fashion to form a pulse sequence.

In MRI the term gradient means a spatially dependent additional component to the magnetic field in the z direction, i.e. along B_0.

For example an x gradient (G_x) will add to the static field another contribution to the total field intensity at different points along the x axis.

We saw how the resonance frequency depends on the magnetic field so when all the protons (spins) experience the same field and have the same frequency but when a a gradient is added to the magnetic field produced by the gradient then spin resonance frequency became spatially dependent

Faster or slower precession is detected as higher or lower frequencies in the NMR signal, and so frequency measurements may be used to distinguish between NMR signals at different positions in space.
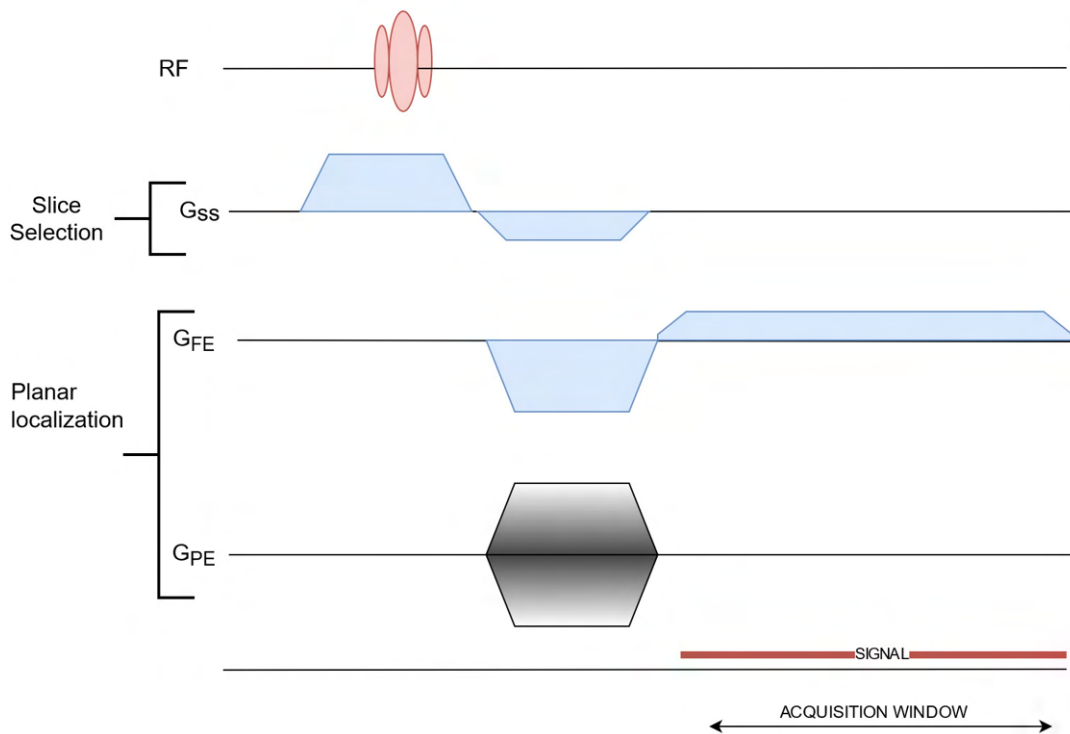
FIGURE 3.2: Basic gradient-echo MR imaging sequence. Amplitude is shown vertically, time horizontally. G_{SS} is the slice-selective gradient, G_{PE} the phase-encoding gradient and G_{FE} the frequency-encoding gradient

Gradients can be applied in any direction or orientation. Usually in a NMR system there are three main direction axes G_x, G_y and G_z and they are combined to produce gradients in any desired direction. Gradients are normally applied only for a short time so they are called pulses as pulses and they are used to produce sequences.

An NMR pulse sequence diagram is a simple means of showing how the RF and gradients are applied.

The basic gradient-echo MR imaging sequence that will be used to illustrate the image-formation process is shown in Figure 3.2 . The vertical axis represents pulse amplitude and the horizontal axis is time.

First, an RF pulse is applied simultaneously with a slice-selective gradient G_{SS} . The RF pulse stimulates the NMR interactions in tissue which lead to the NMR signal.

By combining the RF excitation with a gradient the MR interactions are restricted to a two-dimensional plane, the slice. Any physical gradient G_x, G_y or G_z or combinations of these can be used for this purpose, allowing us to produce axial, sagittal or coronal slices.

Then, gradients are applied in the directions orthogonal to the slice selection to produce the localization in the plane with phase and frequency encoding.

This encodes the NMR signal in the phase-encode direction and the frequency-encode direction. The signal is acquired after the application of the gradient. Note that this is during the frequency-encode gradient but after the phase encoding.

This sequence has to be repeated for every line of data, corresponding to a different value of phase-encoded gradient thus a time period, TR, needed to let the system relax, occurs between the application of one RF excitation and the next.

Once all the data are acquired, a two dimensional Fourier transform is applied. This converts the data, already encoded as spatial frequencies, into an image.

3.2.2 Pixels in images and k-space

The pixels in NMR images are organized into rows and columns in a matrix. Each pixel in the reconstructed image can be thought of as a number which represents the signal Intensity in a given small volume that is indicated by its spatial coordinates and its elements are real values or integers.

The raw data (k-space) instead is conventionally organized in the following order: frequency-encode dimension, phase-encode dimension, number of slices. As opposed to images, k-space elements are complex valued.

Each time the sequence is repeated a full line of data in the frequency-encoded direction is acquired. The phase-encode gradient is changed for each repetition and each line has a different position in the phase-encoded direction. Thus, as the sequence is acquired, k-space is filled row by row in the raw data matrix.

When the number of phase-encoded elements (the rows in k-space) is defined it sets the number of times the acquisition sequence must be repeated. Each time a TR (see 3.1.1) has to be waited to allow the system relaxation so this setting strongly influences how long the scan will take.

On the other hand, the frequency-encode matrix has only a negligible effect on the scan time in the order of ms, which is why we often have a larger matrix in the FE direction.

Conventionally we show the FE direction as columns in k-space, and the PE direction as rows. So if you choose 256 for frequency encoding, each MR echo will have 256 sample points, thus requiring 256 columns in the k-space matrix.

When the PE matrix is set, it is defined how many echoes have to be acquired (the rows in k-space), so every digitized sample point of the echo is stored in the k-space.

Another important difference is in how information from the volume is stored in the data. Pixels in images are close if they are geometrically close in the volume while in k-space pixels are close if they are close in frequency or in phase.

Thus while pixels in the images encode local information in a portion of the scanned volume, in the k-space instead all points encode information about all the volume but at a different frequency. Thus, data in the middle of k-space contain all the signal-to-noise and contrast information for the image, and data around the outside contain all the information about the image resolution (edges and boundaries).

Pixel (and voxel) size is an important parameter in the acquisition since it controls the resolution of an image and its SNR.

We can define the voxel size in all three dimensions, usually one dimension is the slice thickness and the other two spatial dimensions of the field of view (FOV) are relative to the single slice.

We have to calculate the size in the frequency - and phase-encode directions separately, because they can have different matrix sizes. For example

$$FE\ pixel\ size = \frac{FE\ FOV}{NFE} \tag{3.13}$$

Where NFE is the number of elements in the k-space in the frequency encoded dimension. Images and k-space data will be the again discussed in Chapter 6 when a neural network will be applied to both representations of the signal.

3.2.3   Resolution and signal

Generally MRI resolution is pixel limited. That means that the smallest object or detail you can visualize in the image has the dimensions of a single pixel.

The dimension of a pixel is set by the length of the FOV and the number of pixels acquired. So for a 256 matrix and a 25 cm FOV details of the order of 1 mm should be visible.

Contrast, resolution and signal to noise ratio (SNR, see Section 6.3.1) are the three most important factors that determine if a detail is visible in an MRI acquisition.

First, The contrast is the intensity difference with the surrounding tissues. If the contrast is too low the area of the object can be confused with adjacent areas. In general the contrast is modulated by leveraging differences in T1 and T2 relaxation times modulating the time parameters of the acquisition sequence.

Second, the resolution should be sufficient to clearly identify the object. If resolution is too low, the signal of the object will be mixed in the same voxels with the signal on the surrounding tissues. Third, if the SNR is too low, the details of the borders may be confused by image noise.

3.2.4   Resolution and scanning time

Increasing spatial resolution will both increase the scan time and reduce the SNR.

Spatial resolution in the frequency-encoding (FE) direction will contribute only marginally on the scan time (but not in terms of SNR) if the matrix is increased while keeping the FOV constant but changing the phase-encoding (PE) matrix, means acquiring more lines of data, which takes time.

$$Scan\ time = NSA \times TR \times NPE \tag{3.14}$$

Where NSA is the number of averages of the signal and NPE is the number of lines acquired.

The signal on the other hand is proportional to the number of atoms in a region of space so it will increase if slice thickness and pixel size are increased at the expense of resolution if the FOV is kept constant.

Another important acquisition parameter is the signal averaging (NSA) that does not affect resolution and reduces the noise by averaging the signal over multiple acquisitions. Acquiring more scans increases linearly the acquisition times but only increases SNR as the square root of the number of signal averages. Also, in the case of living subjects there may be other motivations to keep the scan time low, for example the discomfort of the patient or to limit the effect of movement when it can not be controlled (i.e fetal MRI).

Also changing NFE can affect the noise, but the effect depends on what happens to the bandwidth and FOV. If FOV is kept constant, the bandwidth increases and noise is proportional to the square root of total bandwidth.

In general, the best set of acquisition parameters and sequence is defined by the imaging task. Each application has its own constrains and its impossible to give indication to which is the best imaging configuration.

In Chapter 7 a particular acquisition modality where multiple coils are used at the same time. At first approximation we can think that all the topics discussed remain valid at single coil level.

Actually, the reason the data used in that chapter are acquired in that way, accelerated imaging, should be discussed separately and a nice introduction can be found in the book by McRobbie et al. (McRobbie et al., 2017).

Chapter 4

Evaluation of F-BPA internalization for Proton Boron Fusion Therapy applications in pancreatic cancer cells

Chapter Abstract

A promising new therapeutic approach for pancreatic cancer is Proton Boron Fusion Therapy (PBFT) which produces a highly localized damaging action through nuclear reactions of the incoming proton beam and 11B, conveniently administered to the patient before the treatment. In this Chapter we report our evaluation of the internalization of fluorinated-boronphenilalanina (F-BPA) in the PANC-1 cellular line. The internalization of the boron is vital for this technique to be adopted for the treatment of a particular tumor since its effect depends on the concentration of target boron inside the cell membrane.

Compared to the most used boronophenylalanine (BPA), F-BPA has the advantage that its bio-distribution may be in principle monitored in vivo using 19F- Magnetic Resonance (MR) so its adoption for PBFT may be advantageous since the possibility to track the drug biodistribution in vivo can greatly help the application in clinical practice of this therapy.

Thus, to further strengthen our results The F-BPA internalization in PANC-1 cells was evaluated with three independent techniques: neutro autoradiography, which is sensitive to boron, liquid chromatography, and 19F-MR Spectroscopy, which is sensitive to fluorine atoms.

Since the internalization is confirmed, we propose to further investigate F-BPA as a boron carrier for BPFT for the therapy of pancreatic cancer and we will adopt it in our experiment on the animal model.

Together with these positive results, we will also present the measure performed on a proposed alternative tracer that was promising because it has a large number of fluorine and boron atoms but it is not internalized so its application is not advantageous in this particular case.

4.1   Introduction

In this Chapter we will introduce our case study that is developed within the NEP-TUNE collaboration that is the treatment of pancreatic cancer. The newly proposed technique, the Proton Boron Fusion Therapy (PBFT), will be briefly described and we will discuss why it is important to research new therapeutic approaches for this tumoral pathology and how the PBFT can be a valuable tool in its treatment.

The Chapter is centered on the importance to measure the internalization of the boron inside cells since it is a necessary condition for applicability of PBFT.

In particular, we will draw a parallel with a related technique that involves a similar nuclear reaction, the Neutron Boron Capture Therapy (BNCT), that has the same constraint to reach a high concentration of boron inside the tumoral area but to which we believe that PBFT may be superior in this specific use case.

Then we will discuss how to measure the intracellular internalization of the boron carrier and we will apply the technique to the study of F-BPA that we would like to adopt as a dual purpose boron carrier and magnetic resonance tracer.

The techniques presented are general and can be applied to any problem that need the quantification of cellular uptake of molecules tagged with fluorine and is not limited to the application in PBFT.

In addition, we will also present the tentative proposal of a new boron carrier that has great magnetic resonance properties but unfortunately can not be applied to pancreatic cancer with this treatment since it does not reliably internalize into cells.

4.1.1   Application to pancreatic cancer

Pancreatic cancer is an aggressive and highly lethal disease. Optimal treatment is a multimodal approach that combines surgical resection, chemotherapy and/or radiation therapy (Sereti et al., 2018) but, despite the improvements in the therapeutic strategies over the last 30 years, the 5-year survival rate remains at 9% (Rawla, Sunkara, and Gaduputi, 2019), indicating that the overall management of the disease is still insufficient. For this type of cancer, chemotherapeutic approaches are limited by the reduced stroma permeability and heterogeneous blood supply, while the effectiveness of radiation therapy is reduced by hypoxia (Seshacharyulu et al., 2017) . As the relative biological effectiveness (RBE) of proton therapy (PT) is only 10% higher than photon radiation therapy (XRT), PT alone is not expected to guarantee a better outcome than XRT for this type of cancer. Carbon-ion therapy, which utilizes heavier particles inducing more severe DNA damages compared to PT and XRT, is providing hopeful results, while alternative pathways to fight radiation resistance are also being investigated, including the introduction of hypoxia activated drugs (Dell'Oro et al., 2020).

Given the characteristic of pancreatic tumour, a possible therapeutic approach for pancreatic tumour could be the Boron Neutron Capture Therapy (BNCT), a binary therapy approach (Barth et al., 2014; Moss, 2014) which exploits thermal capture of 10B, suitably accumulated inside the tumor before thermal or epithermal neutron (n) irradiation. The n+10B reaction is able to generate short range high-LET particles (alpha particles), causing a highly localized damaging action. The drawback of BNCT for pancreatic cancer is that it may be difficult to obtain a sufficient thermal neutron fluence at the tumour site to ensure a therapeutic effect without exceeding the tolerance dose of surrounding tissues/organs.

Another crucial event for the success of BNCT concerns the 10B tumor / healthy(normal) tissue concentration ratio (T/N) which has a minimum threshold of 2 in current clinical applications (Barth, Mi, and Yang, 2018). The most commonly used boron carriers used in BNCT are 10B enriched sodium borocaptate (BSH) and borono-phenylalanine (BPA, currently the gold standard) (Barth, Mi, and Yang, 2018); In this

contest, tumour to normal tissue (T/N) boron concentration ratio may not be optimal for pancreatic this cancer as BPA is taken up in healthy pancreas with higher concentration than other normal tissues: See Section 5.1.1 for a short review on the subject.

The idea to exploit nuclear reactions to produce short-range high LET particles has been revived in a recently proposed approach, the Proton Boron Fusion Therapy (PBFT), which goal is to enhance the biological effectiveness of proton (p) therapy inducing p-11B nuclear reactions (p-11B -> 3 alpha).

Promising results for BPFT have been obtained in (Cirrone et al., 2018), where, using BSH, an increase in DU-145 cell killing and DNA damage was measured at the mid-SOBP (Spread Out Bragg Peak) of a clinical proton beam.

It has been suggested that this increase is due to the high-LET alpha particles produced in the p-11B reaction, whose maximum cross section is at low proton energy, corresponding to the Bragg peak localized in the tumor region. The constraint of a differential uptake of the carrier between normal and cancer cells as needed in the BNCT is thus eliminated and BPA remains a suitable carrier.

Given that PBFT, as BNCT, needs carriers to be administered to the patients before irradiation and internalized by cancer cells, it is necessary to check that suitable concentrations are achieved inside the tumors. The use of F-BPA has been suggested in BNCT to estimate the boron concentration in the tumor before the neutron irradiation via PET scan or Magnetic Resonance Imaging (19F-MRI) (Ishiwata, 2019a). PET is used to evaluate the eligibility of a patient for BNCT, by measuring with a scan taken days before the treatment the T/N ratio. This is because the tracer used for PET, 18F-BPA for example, cannot be used as boron carrier during the therapy, given that the quantity needed for the treatment would expose the patient to a very high dose.

Moreover, using different molecules (18F-BPA vs BPA) at different times, can introduce biases in the measurement of the bio-distribution of 10B inside the patient, due to effects like different pharmacokinetics, non linear uptakes etc.).

19F-MR Imaging or Spectroscopy, on the other hand, exploits a non radioactive fluorine isotope (19F), thus a molecule like 19F-BPA can be in principle used simultaneously as tracer for T/N quantification imaging and as boron carrier in BNCT.

In BPFT, as recently pointed out, fluorine can also contribute to the enhancement of the radiobiological effectiveness of the therapy since the nuclear reaction p+19F has similar properties, producing a single alpha of $\mathcal{O}(10)$ MeV energy, with a cross section peak in the tumor region. The use of 19F-MRI is currently limited by a low signal to noise ratio (SNR) (Tirotta et al., 2015) but some promising results have been obtained in in-vivo on animal model imaging with 19F-BPA (Porcari et al., 2009; Porcari et al., 2008).

Our main goal is to to investigate the feasibility of the usage of 19F-BPA as a carrier and as tracer in BPFT for the treatment of pancreatic cancer. To this purpose, as a first step, we present an in-vitro measurement of the internalization of 19F-BPA in the PANC-1 cellular line with a multimodal approach. We concentrated on 11B because BPFT seems more promising as a therapy for pancreatic cancer. We employed neutron autoradiography, nowadays one of the entrusted techniques suitable for the quantification of boron in cellular and tissue systems (Wittig et al., 2008), to quantify boron taken up by PANC-1 cells and also to obtain imaging of boron distribution in

the cell samples, allowing to verify if 19F-BPA administration protocol had guaranteed a uniform availability of boron to all cells. The needing to develop protocols, to detect and quantify boron-based carriers at cellular level, based on experimental approaches and advanced analytical techniques (e.g. UHPLC-HR-MS; UHPLC-UV-DAD), allowed us to hypothesize the use of cellular metabolomics as a useful tool. In fact, metabolomics offers the advantage of exploring the biological system's response to an external chemical insult and how this, eventually, changes from a structural point of view. For this purpose, ultra-high-pressure liquid chromatography electrospray ionization quadrupole time-of-flight mass spectrometry (UHPLC-ESI-QqTOF-MS) , through the measurement of accurate mass and isotope patterns, provides a reliable contribution following a protocol of quenching/extraction that takes into account the chemical-physical properties of the carrier.

The presence of the fluorine atoms allows us to considered another technique, Magnetic Resonance Spectroscopy (19F-MRS) to qualitatively evaluate the internalization with a completely independent approach (measuring fluorine instead of boron) with the perspective to set-up an imaging protocol (19F-MRI).

In particular we will be able to validate the 19F-MRS results against the two other techniques that are more established in the field. Thus, we performed a multimodal approach using three different techniques to evaluate the in-vitro internalization of 19F-BPA in PANC-1 cells: neutron autoradiography, liquid chromatography and magnetic resonance spectroscopy.

In this Chapter I will report our experiment aimed to verify the internalization of fluorine tagged version of borono-phenylalanine (F-BPA) in a human pancreatic (PANC-1) cell line. Together with the experiment that I performed with fluorine magnetic resonance spectroscopy, I will report also results from experiments performed by our collegues and members of the Neptune collaboration, Silva Bortolussi and Ian Postuma from the BNCT lab in Pavia and Severina Pacifico and collaborators at University of Campania "Luigi Van-vitelli". Tables and Figures that reports results that I did not personally obtain are credited in the caption.

4.2 Methods and Materials

4.2.1 Cell line and protocol (19F-MRS)

Human pancreatic cancer cell line (PANC-1) was purchased from ATCC (American Type Culture Collection). Cells were grown in DMEM medium supplemented with 10% Fetal Bovine Serum, 50.0 U/mL penicillin, and 100.0 mg/mL streptomycin, at 37 °C in a humidified atmosphere containing 5% CO2. 2-Fluoro-4-boronophenylalanine (19F-BPA) has been prepared following a previously reported procedure (Kabalka et al., 2000a).

Cells were grown in 10 ml DMEM medium supplemented with 10% FBS, 1% penicillin/streptomycin and 2mM glutammina. Cells were seeded at a density of $3.0 \cdot 10^6$ in 2 T75 flasks with 48 hours incubation time. Cells were treated with 120 ppm (13.6 mM) 19F-BPA with 4 hours exposure time, dissolved directly in cell medium. At the end of the exposure period the cell medium was removed and 3 washes with PBS were performed (2 ml/each). The liquid of the 3rd wash is preserved and used as "negative control" sample for the analysis. Adherent cells were detached using Trypsin-EDTA solution (trypsin, 0,25%, EDTA 0,02%), collected and centrifuged to obtain the "pellet" sample. Another sample ("lysate" sample) is obtained with the same procedure but cells were lysed in a 400 μL lysis buffer (TRIS 50mM, Triton 1%, SDS 0,1%, NaCl 150 mM, EDTA 5 mM). The "lysate" sample is

created because the MRS measure is simpler on an homogeneous sample whose volume is comparable with the active volume of the spectrometer and because we aim to avoid short T2 relaxation time of intracellular 19F-BPA that may damp the NMR signal. In order to evaluate the fraction of 19F-BPA internalized by the cells (internalization fraction), a "positive control" sample is built with the same number of fluorine atoms that is expected if the concentration inside the cells were the same of the external medium (i.e the internalized fraction is equal to 1). The same number of cells as in the "pellet" (and thus the "lysate") sample is considered and added to a 400 μL lysis buffer. The 19F-BPA is then added in a quantity obtained by the initial medium concentration (13.6 mM) and the total cellular volume, computed as the measured number of cells times the cellular volume estimated from microscopy imaging ($2.25 \, 10^{-5} L$). The corresponding 19F-BPA is 3.8 μ moles.

Neutron autoradiography and UHPLC-ESI-QqTOF-MS analysis

The same cell line and cellular medium used in Section 4.2.1. For neutron autoradiography cells were seeded at a density of $2 \cdot 10^6$ in 100 mm Petri dishes and treated with 19F-BPA (120 ppm; 4 h), after 24. At the end of the exposure period, cell pellets were obtained in the following way: the borated medium was removed, cells were washed three times with PBS, harvested by trypsinization, counted, centrifuged and deposited on Mylar support. Cell pellets thus obtained were let drying. Dry pellets were then irradiated in contact with the CR-39 track detectors for boron quantification and imaging (see below). For liquid chromatography, cells were seeded at a density of $2 \cdot 10^6$ in 100 mm Petri dishes and treated with 19F-BPA (80, 100, and 120 ppm; 4 h), after 24. The cell culture medium was instead collected in Falcon and lyophilized using the FTS-System Flex-dry instrument (SP Scientific, Stone Ridge, NY, USA). Cells were first quenched by addition of 1 mL of MeOH:H2O (4:1, v:v), and, after collection by physical scraping, extracted using a solution MeOH:H2O (1:1, v:v). The extraction was carried out dipping the tubes with the samples into liquid N2 for 30 s to snap-freeze the cells and then allow them to thaw on dry ice (Bi et al., 2013; Luo, Gu, and Li, 2018). Sample tubes underwent sonication for 5 min in an ultrasound bath. After centrifugation at 3500 rpm and 0 °C for 10 minutes, the supernatant was recovered and lyophilized. The lyophilized supernatants were reconstituted in methanol and subjected to analysis.

4.2.2 19F-MRS

19F-MRNMR spectra were collected on each sample in 5mm glass capillary using a 9.4T vertical bore high field spectrometer (Bruker Avance-400). We performed two independent experiments (experiment n.1 and experiment n.2), to check for experiment repeatability. For the first experiment, the 19F-MRS protocol was organized with a first acquisition on the "pellet" sample followed by a second acquisition on the "lysate" sample, the "positive control" sample and the "negative control" sample. The 19F-MR spectra were acquired using 18 μs pulse for the 90 flip-angle radiofrequency pulse, 8192 free induction decay (FID) points, 180 ppm spectral width, repetition time (TR)=3s and number of signal averages (NSA) = 4096 with the exception of pellet sample in experiment 2 where NSA was 1024. Experiment n.1 and experiment n.2 were conducted with $(12 \pm 2) \cdot 10^6$ and $(10 \pm 2) \cdot 10^6$ cells, respectively and "positive control" sample was prepared with $(10 \pm 2)10^6$. Results were consequently normalized respect to the cells number. Cells count was performed with Burker's chamber.

After phase and baseline corrections spectral quantification was achieved by peak-resonance integration. Raw data was imported in MATLAB (The MathWorks Inc,2018b) baseline noise level was subtracted and spectral data was filtered with a mean filter of (10 point) Hz length. Numeric integration of peak areas was performed. The area of the 19F-BPA resonances in the "lysate" sample is proportional to the quantity of 19F-BPA internalized by the cells. This has to be divided by the area of the peak of the "positive control" sample which is proportional to the signal expected if the 19F-BPA concentration inside the cells were the same as the external medium concentration (internalization fraction equal to one). This ratio is an estimate of the internalized fraction. The "pellet" sample is used as a cross-check while the "negative" control sample is used to check that non internalized 19F-BPA has been correctly washed out.

Neutron autoradiography

The neutron autoradiography technique used in this work is based on the irradiation of samples enriched with 10B in direct contact with a passive detector that is only sensitive to charged particles. The charged particles coming from neutron capture in 10B, create latent tracks in the detector which can be made visible by chemical etching. Depending on the calibration of the technique (neutron fluence, etching parameters), it can be employed to obtain quantitative information on boron concentration or qualitative imaging of boron spatial distribution. For boron quantification, samples in contact with the detectors are irradiated with a neutron fluence of $(1.97 \pm 0.01)10^{10} cm^{-2}$ and etched for 10 minutes in PEW solution at 70°C. The track density is measured in 40 sequential pictures taken with a light microscope. Each image has an area of about 0.3 mm^2 (0.632 mm x 0.474 mm): the total measured area is about 12 mm^2. The 10B concentration is obtained by counting the tracks over the entire measured area, the associated error is the square root of the counted tracks.

This value is then converted in boron concentration through an appropriate calibration curve as reported in (Postuma et al., 2016). 11B concentration is inferred by multiplying isotopic natural abundance (1:4). With this technique another variable can be measured, by computing the standard deviation of the track density over all the images obtained in one sample, we are able to distinguish homogeneous from heterogeneous boron uptake. Dosimetry due to proton irradiation is calculated with an average boron concentration in cells, thus the standard deviation of the concentration is a representative uncertainty to propagate in dose calculation if boron concentration is heterogeneous. Qualitative analysis is obtained by irradiating samples with higher neutron fluence and by etching for longer time, thus causing the tracks to overlap forming a map of gray levels, representing areas of the sample taking up higher boron concentration (Bortolussi et al., 2011). This technique and its variants have been set-up and validated for BNCT, where the relevant isotope is 10B (Portu et al., 2015). 11B uptake experiments have been carried out with 19F-BPA.

Liquid Chromatography

UHPLC-ESI-QqTOF-HR MS/MS analysis

The chromatographic analyses were carried out using the Shimadzu NEXERA UH-PLC system and the Omega Luna C18 column (1.6 μm, 50 × 2.1 mm i.d.; Phenomenex, Torrace, CA, USA). The mobile phase consisted of a binary solution A: 0.1% HCOOH in H2O, B: 0.1% HCOOH in CH3CN. A linear gradient was started at 2% B, held for 1 min, and linearly ramping to 98% B in 2.50 min. The mobile phase

composition was maintained at 98% B for further 50 seconds, then returned to the starting conditions and allowed to re-equilibrate for 2 min. The total analysis time was 5.0 min, the flow rate was 0.5 mL/min, and the injection volume was 2.0 μL.

MS analysis was performed using a hybrid Q-TOF MS instrument, the AB SCIEX Triple TOF® 4600 (AB Sciex, Concord, ON, Canada), equipped with a DuoSprayTM ion source (consisting of both electrospray ionization (ESI) and atmospheric pressure chemical ionization (APCI) probes), which was operated in the positive ESI mode. The APCI probe was used for automated mass calibration using the Calibrant Delivery System (CDS). The CDS injects a calibration solution matching the polarity of ionization and calibrates the mass axis of the TripleTOF® system in all scan functions used (MS and/or MS/MS). The Q-TOF HRMS method combined TOF-MS and MS/MS with Information Dependent Acquisition (IDA) and consisted of a full scan TOF survey and a maximum number of eight IDA MS/MS scans. The MS parameters were as follows: curtain gas (CUR) 35 psi, declustering potential (DP) 70 V, collision Energy (CE) applied was 30 V with a collision energy spread (CES) of 20 V, ion-spray voltage, 5500 V; ion source heater, 500 °C. The instrument was controlled by Analyst® TF 1.7 software, while data processing was carried out using PeakView® software version 2.2.

HPLC-UV-DAD analysis

To quantify the phenylalanine-derived carrier, HPLC-UV-DAD analyses were performed using the HPLC 1260 INFINITY II system (Agilent, Santa Clara, CA, USA), equipped with an Agilent G7129A autosampler, an Agilent GY115A DAD-UV-visible detector, and a Quaternary pump Agilent G711A. The separation was accomplished on Luna® Phenyl-Hexyl column (150 × 2 mm i.d.; 3.0 μm). Mobile phases were water (A) and acetonitrile (B), both with 0.1% formic acid and the gradient condition was: 0–3.50 min, 2 → 98% B; 3.50–4.0 min, 98% B. Finally, starting conditions were restored and the system re-equilibrated for other 1 min. The total analysis time was 5.0 min, the flow rate was 0.5 mL/min Injection volume was 5.0 μL. The diode array UV-Vis detector was set at 230 and 270 nm.

4.2.3 Statistical analysis

The internalization fractions are expressed as averages over different experiments/samples ± standard error. In the case of neutron autoradiography errors are propagated from the square root of the counted tracks. Errors for the MRS signal area are derived from the noise level of the acquisition and the peak width, errors reported in the internalized fraction are propagated from them.

4.3 Results

4.3.1 19F-MRS

Figure 4.1, panel a) shows the obtained 19F MR spectra for the "negative control" (dottet black line), the "positive control" (dotted red line) and the "lysate" (yellow line, purple line) samples while panel b) shows a comparison between the spectra of the "pellet" and "lysate" samples. In cell cultures 19F-BPA exhibits two resonances. The result suggests that 19F BPA is found in two main compartments characterized by a different chemical environment.The absence of signal in the negative control confirms the absence of fluorine that was not internalized by the cells. Table 4.1

TABLE 4.1: Position, area of the two resonances peaks of the 19F-MRS spectra, total area in the various samples. Error on the resonance peak position is half of the size of the smoothing filter. Errors on the resonance peak area are estimated from the noise level of the acquisition and the resonance peak width.

	Peak 1 (ppm)	Area 1 (σ ppm)	Peak 2 (ppm)	Area 2 (σ ppm)	Total Area (σ ppm)
"Lysate" sample exp. 1	107 ± 0.1	10 ± 1	105.7 ± 0.1	12 ± 1	21 ± 2
"Lysate" sample exp. 2	106.8 ± 0.1	5 ± 1	105.7 ± 0.1	9 ± 1	14 ± 2
positive control	106.9 ± 0.1	24 ± 1	105.5 ± 0.1	16 ± 1	40 ± 2

shows the position and the area of the two peak-resonances, and the total area for the various samples.

For each experiment the measured internalization fraction is obtained as the ratio of the total area in the positive control sample normalized in respect to the total number of cells in the sample. We obtained 0.5 ± 0.1 and 0.4 ± 0.1 for the two experiments. Given the biological variability of the samples we believe the two measures are compatible.
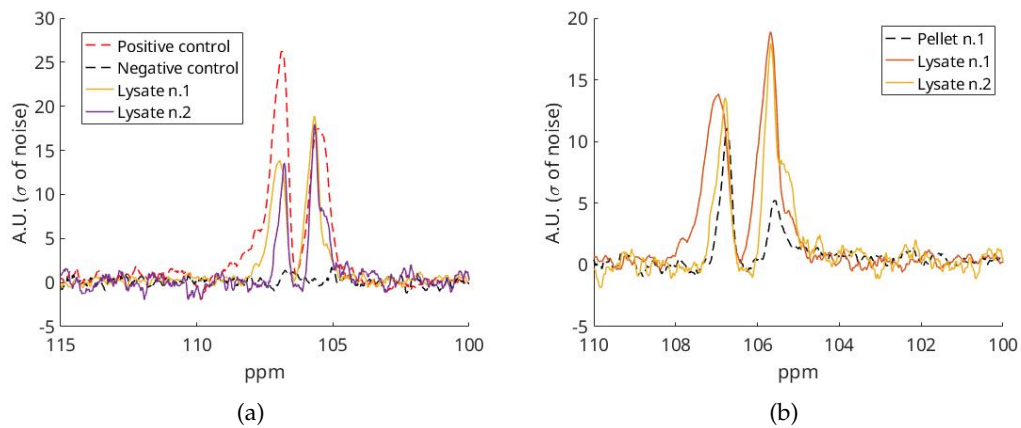
FIGURE 4.1: (a) NMR spectra for the "negative control", the "positive control" and the "lysate" samples in experiment n.1. and n.2 (b) "pellet sample" and "lisate sample" for experiment n.1 and n.2. Spectra are normalized so that noise, signal far from the peak, has mean 0 and std 1. Signal is smoothed with a linear filter of 0.2 ppm.

Neutron autoradiography

A first set of measurements highlighted a poor solubility of the fluorinated BPA: Figure 4.2 (a) Left panel shows qualitative neutron autoradiography where the white

TABLE 4.2: 11B concentration and internalization fraction obtained by quantitative neutron autoradiography in samples treated with 19F-BPA (120 ppm, 4 hours). The results have been obtained by measuring track density in a central area of the sample as explained in the text. Boron uptake was homogeneous in all samples. The error reported is statistical only.
Courtesy of Silva Bortolussi and Ian Postuma from the BNCT lab in Pavia.

Experiment	Sample	Inside 11B ppm from F-BPA	Internalized fraction
1	1	63 ± 2	0.52 ± 0.03
1	2	66 ± 2	0.55 ± 0.03
1	3	68 ± 2	0.56 ± 0.03
2	1	56 ± 2	0.47 ± 0.04
2	2	54 ± 2	0.45 ± 0.04
2	3	55 ± 2	0.46 ± 0.04

spots (corresponding to the yellow areas on the right) indicate a high boron concentration. A quantitative map of 10B distribution in a portion of the sample, showing spots where concentration is up to 5 times higher, is presented in panel (a) on the right. In this case, 19F-BPA was dissolved in culture medium as normal BPA, requiring the use of fructose at Ph 9.5-10 and then the use of HCL to readjust Ph to 7.4 (Coderre et al., 1997). This prompted a revision of the protocol of 19F-BPA solution in the culture medium, which consists in the simpler procedure described in without the formation of the complex with fructose.

The following experiments have been performed with an improved protocol to dissolve the formulation in the medium. PANC-1 cells have been cultured and exposed to 120 ppm of 11B for 4 hours. This time neutron autoradiography proved a very uniform boron uptake (Figure 4.2, left panel) as also demonstrated by the values of boron concentration, showing a very low variability in the track density over the area sampled in the image (panel d). The results are shown in Table1. Cells were treated with BPA in natural isotopic composition, thus images of 10B were less contrasted because the overall concentration is only $\frac{1}{5}$ of the total. 11B concentration was inferred by quantitative neutron autoradiography pointing out 10B, which concentration was multiplied by 4. It should be noted that a significant difference in the internalization fraction between the two experiments is observed, to be attributed to the known biological variability in cell growing process.

Liquid Chromatography

Behaviour of 19F-BPA on the mass spectrometer

19F-BPA was dissolved in water and analysed by tandem mass spectrometry in positive ion mode (Figure 4.3, panels A and B). The proposed carrier's fragmentation pathway is reported in Figure 4.3, panel C. The ion $[M + H]+$ at m/z 228.0838 provided, by loss of a residue of HCOOH or CO+H2O, the ion TOF-MS2 at m/z 182.0783, which can undergo a loss of metaboric acid, providing the ion at m/z 138.0716, or to dehydration to give the less abundant ion at m/z 164.0679. The ion
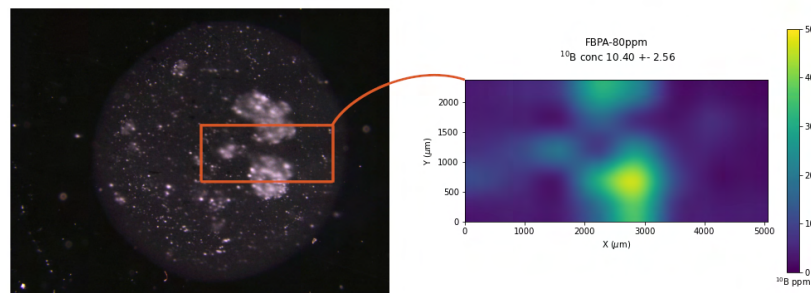
(a)
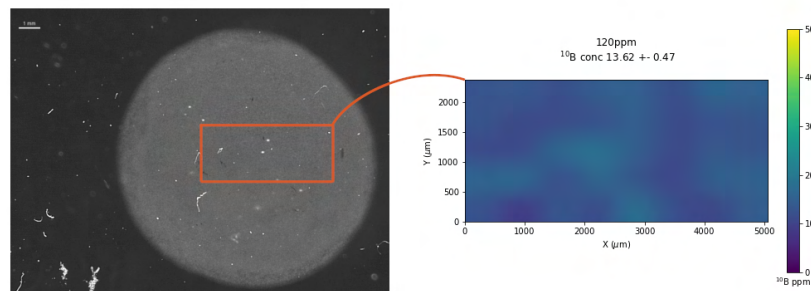
(b)

FIGURE 4.2: (a) The figure on the left shows the qualitative neutron autoradiography of cells treated with 19F-BPA, showing a map of boron distribution in the cell pellet. White areas represent part of the sample taking up high boron concentration. This imaging shows poor uniformity and different availability of the compound to different cell populations. On the right, quantitative map of 10B distribution in a portion of the sample shows spots where concentration is up to 5 times higher; (b) "The figure on the left shows a qualitative neutron autoradiography of cells treated with 19F-BPA, showing a map of boron distribution in the cell pellet. White areas represent parts of the sample taking up high boron concentration. In this case the white area covers the entire sample region. Consequently the B distribution is well homogeneous. The image on the right shows the B quantitative map on a representative portion of the sample. Also this analysis shows an homogeneous B uptake.
Courtesy of Silva Bortolussi and Ian Postuma from the BNCT lab in Pavia.

TABLE 4.3: Concentration of 11B in PANC-1 cells after treatment with 19F-BPA. The values are reported as the mean ± SD of the measurements carried out on 3 samples, analysed 2 times.
Results from Severina Pacifico, University of Campania "Luigi Vanvitelli"

Treatment dose (11B ppm)	Inside 11B ppm from F-BPA	Internalized fraction
80	40 ± 2.9	0.50 ± 0.04
100	54 ± 2.3	0.54 ± 0.02
120	63 ± 1.0	0.52 ± 0.01

at m/z 182.0788, the base peak of the spectrum, could also be generated as a result of fluorine protonation, with following formation of the enaminium residue and loss, favoured by the intramolecular hydrogen bond, of the CO+H2O units (Figure 4.3, panel C). The ion at m/z 138.0716 by deamination provided the ion with m/z 121.0456 or more favourably by loss of HF the fragment ion at m/z 118.0654, from which the formation of the tropilium ion is favoured (m/z 91.0549).

Following 19F-BPA treatment, PANC-1 cells appeared to assume a mesenchymal phenotype (See Figure 4.4 for a discussion)

4.4 Discussion

Pancreatic cancer, one of the most lethal types of cancer, still lacks an effective therapeutic approach. Among the possible therapies currently under study, PBFT is a promising method which exploits nuclear reactions that produce high-LET alpha particles, triggered by a proton beam on boron atoms, conveniently introduced in the tumor by borated carriers. However, for the therapy to be successful is necessary to accumulate a suitable quantity of 11B and 19F carrier in the tumors and to be able to reliability quantify it before the irradiation. As opposed to BNCT, a high T/N ratio is not necessary, given that PBFT concentrates the dose release in the Bragg peak, where the proton energy is suitable to trigger the nuclear reaction with the 11B. Standard carriers used in BNCT are BSH and BPA that have been studied both in vitro and in vivo on various cellular lines.

In this Chapter, we characterized the internalization of F-BPA in PANC-1 cellular line with a multimodal approach.

We propose F-BPA since it presents some advantages with respect to BPA. First, F-BPA could be directly imaged via 19F-MRI, during the treatment. Second, in PBFT, fluorine can also trigger nuclear reactions with high-LET particles in the final state. Other advantages are related to the higher solubility of F-BPA that makes the administration easier. Due to the possible translation of 19F MR in in vivo studies, we used 19F MRS to investigate F-BPA accumulation in PANC-1 cells validated by multimodal evaluation with well established techniques such as autoradiography and liquid chromatography.

Autoradiography exploits the same nuclear reaction used in BNCT showing that at 120 ppm in the medium F-BPA is significantly internalized in PANC-1 cells, with
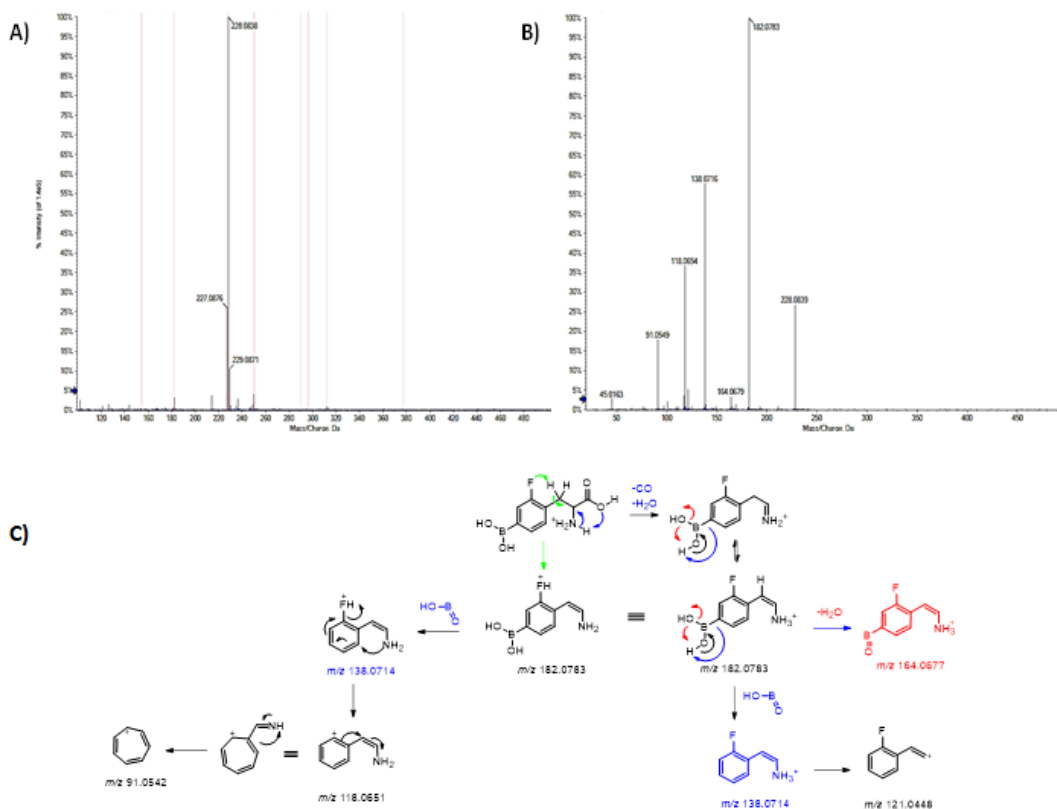
FIGURE 4.3: a) TOF/MS; b) TOF-MS/MS spectra of 19F-BPA in positive ion mode; c) proposed fragmentation pathway of 2F-BPA. Measured exact mass of each fragment ion was at m/z to within 5 ppm vs. its relative theoretical m/z value. This latter is reported below each structure.
Results from Severina Pacifico, University of Campania "Luigi Vanvitelli"

a fraction which is experiment dependent. In fact, treating cells in different experiments, may result in different boron uptake depending on the proliferation state of the cells, on environmental conditions and on confluence. In this sense, the recommendation for radiobiological experiments, which aims to measure the biological damage on cells derived from radiation exposure (Cirrone et al., 2018), is to measure boron uptake each time it is necessary to evaluate the effective radiation dose difference between boron enriched and control samples due actual boron concentration internalized and to the nuclear interaction with protons. Experiments with F-BPA with neutron autoradiography allowed individuating the best protocol of compound administration, because this technique points out the uniformity of boron distribution in cells, which may depend on the level of solubility of the borated formulation in the culture medium. After the optimization of the administration protocol, we found that the increase in the solubility of F-BPA gives a is better internalizationed in PANC-1 cells, and that a uniform distribution can be obtained as shown in Figure 5.4b.

We applied, as an alternative technique, also cellular metabolomics techniques that allows us to determine, together with the internalization fraction, also possible cellular changes from a structural point of view. During this experiment we found that treated cells appeared to assume a mesenchymal phenotype. This finding is in line with pleomorphic pancreatic cells with epithelial-mesenchimal transition

(EMT) phenotype characterized by the functional loss of E-cadherin (Gradiz et al., 2016). However, optical microscopy analysis made during the 19F-MRS experiment in Rome seems to exclude this phenomenon in the sample analyzed with 19F-MRS (see Figure 4.4). We are investigating if this difference had an impact on our results.

PANC-1 cells show increased levels of cholesterol, polyunsaturated fatty acids and octanoic acids with following increased membrane fluidity (Watanabe et al., 2012) which favoured F-BPA intracellular accumulation. Comparably to natural amino acids and borophenylalanine, commonly used in BNCT, F-BPA can diffuse into cells.

We found a dose dependent internalization fraction in agreement, at 120 ppm, with the fraction obtained with the neutron autoradiography technique. We can conclude that we intervalidated the two methods: this allows using one or the other depending on the needs.

In agreement with the autoradiography and liquid cromatography, 19F-MRS showed that F-BPA is internalized at 120 ppm of medium concentration. 19F-MRS can be used to quantitatively measure fluorine concentration in samples through the introduction of a reference sample and is infact often adopted in pharmaceutical application (Holzgrabe, 2010; Holzgrabe, 2015) however its limited sensitivity compared to the other proposed techniques convinced us to test only a qualitative agreement.

The results obtained in the different experiments and measured with different techniques, conclude that the boron internalization fraction in PANC-1 is about 50%, which allows sufficient intracellular boron concentration for radiobiological measurements.

In BNCT experiments, this means that it is possible to deliver representative radiation dose by neutron capture reactions in short irradiation times provided a thermal neutron beam of at least 10^9 neutrons per square centimeter per second. For PBCT, the discussion on the boron concentration needed to obtain a therapeutic change in tumour control is still underway.

We envisage to move to ex-vivo cell on mice models, to check that enough F-BPA can reach the pancreas, once administered in-vivo. To this end, neutron autoradiography will be also useful to obtain maps of boron distribution in thin tissue sections, to be compared with contiguous histological preparation. Results will be again intervalidated among different and complementary techniques.

In Chapter 5 we will provide a first estimation of boron concentration in an orthotopic model of pancreatic adenocarcinoma after the administration of F-BPA. In particular, 19F-MRS will be used to quantitatively estimate fluorine concentrations in tissues.

4.5 Additional internalization experiments

The adoption of F-BPA as a boron carrier has great benefits: It is a well known and highly tested drug that is well tolerated both at cellular level and at systemic level in the animal model. In particular, its high tolerance allows the administration of a high dosage in the animal model as we will see in Chapter 5. One of its weaknesses is having a single atom of fluorine and a single atom of boron that makes it harder to reach higher concentration of both the tracer atom for NMR and the target atom for the therapy. If a molecule has a greater abundance of fluorine we can reach a N
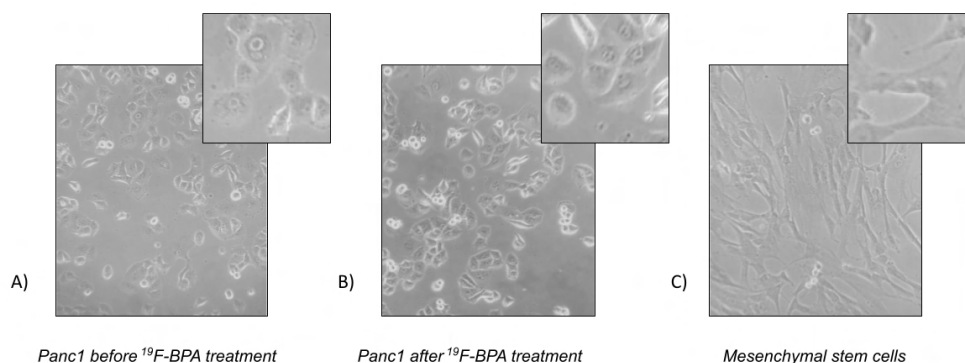
FIGURE 4.4: (A) Cells before and (B) after the treatment with F-BPA used for the 19F-MRS experiment. (C)Example of mesenchymal phenotype
After the administration of F-BPA S. Pacifico from University of Campania "Luigi Vanvitelli" reports that cells seem to assume a mesenchymal phenotype. We do not find this happening in our sample and we are verifying any differences in cellular sample preparation.

(number of identical fluorine atoms) fold increase in signal and concentration at the same drug concentration.

To overcome this weakness we started testing alternative boron carriers that also contain fluorine: one promising candidate was a compound of the family of the dodecaborate that exhibits great NMR properties and it may have application also in NMR imaging. Unfortunately, we found that it is not suitable for in vivo application in PANC1 cells but we report our initial tests since it is of interest to discuss the generalizability of this method of analysis.

Dodecaborate dianios was recently proposed by Warneke et al. (Warneke et al., 2016) as a possible carrier for BNCT. We performed some preliminary experiments to assess their uptake in PANC cells. In particular we used dodecafluoro dodecaborate (B12F12) a molecule with 12 identical fluorine atoms to allow the tracking through NMR spectroscopy.

These molecules can form a complex with beta cyclodextrins (β CD) that can improve their affinity for the tumor cells (Uekama, Hirayama, and Irie, 1998; Warneke et al., 2016).

We first confirm the formation of the complex in our cellular growth medium through a measure of chemical shift with 19F-MRS and we found that this value is comparable to what is reported in the original paper. Figure 4.6 shows the spectra and the chemical shift for the B12F12 in the cellular growing medium describet in Section 4.2.1 .

Then we qualitatively assess the internalization in PANC1 cells employing the 19F-MRS methodology that was presented in this Chapter.

The cell culture for the MRS was prepared as reported in Section 4.2.1 with the following modifications:

The incubation medium was prepared with 1.2 mM di F12B12 + 1.4 nM of betaCD, since B12F12 has 12 fluorine atoms this concentration gives the same fluorine concentration of F-BPA. To favor the formation of the complex the medium was heated to 45 degrees and sonicated for 10 minutes.

The incubation time was 22 hours instead of 4 hours for the F-BPA as reported in the original paper.

To have a benchmark we first test the internalization in fibroblast L929 cell line and we qualitatively confirm the results of Werneke: a rough estimate of the internalization fraction can be made from the SNR ratio between the positive control, defined as the quantity of B12F12 expected if the internalization fraction was 1 (See the F-BPA experiment for the details), thus we can indicate a 5% internalized in L929. In PANC1 cells we found a much lower signal and compatible with a crude estimate of an internalization factor < 1%. Table 4.4 summarizes these results while the spectra are reported in Figure 4.5.

These were only a preliminary measure but we decided to put aside for the moment any additional test in vivo with this molecule:

We verified the internalization but the results were not encouraging since with this experimental setup there may not be an advantage in the use of B12F12 over F-BPA.

First, the low solubility didn't allow us to reach higher fluorine concentration with respect to F-BPA in the cellular growing medium. Also, our colleagues report that B12F12 used at the same concentration of F-BPA (13.6 mM) impairs cellular growth and cells appear to be different in form so a toxicity test should be performed. This quite invalidates the advantage of B12F12 that is the abundance of fluorine atoms since the actual fluorine concentration is the same reached with F-BPA

Second, the internalization fraction was at least an order of magnitude lower than the one of F-BPA. The problem is that for the PBFT the only relevant concentration is the boron concentration inside the cell. Also the internalization was at least 5 times lower in PANC1 cells in respect to fibroblast and this may mean that there may not be an affinity for PANC1 cells.

One possible explanation is the difference in the mechanism of internalization between the two molecules B12F12 enters the cells through diffusion through the cell walls while F-BPA is transported by a membrane protein:

Amino acid (AA) transporters are proteins in the cellular membrane that actively carry AAs in and out the cell. The amino acid transporter family has many transporters and they are assembled based on the specificity of the amino acid molecule being transported as well as dependence on sodium ion for transport activity: They accept groups of AAs rather than individual AAs, such as small neutral AAs, large neutral AAs, anionic AAs, and cationic AAs. Some transporters are Na+ dependent while others are Na+ independent.

AA transporters can be classified as "systems". The system that is responsible for the transport of BPA and F-BPA is the L system:

This transporter is responsible for the inward flux of several essential amino acids such as phenylalanine, leucine, isoleu-cine, tryptophan, histidine, tyrosine in antiport with histidine,tyrosine and a non-essential amino acid glutamine (Puris et al., 2020).

Thus, if PANC1 cells express this transporter more than healthy cells this may be the motivation for the high internalization fraction observed.

TABLE 4.4: Internalization B12F12 in PANC1 cells and in fibroblast L929 cells. Cells were incubated for 22 hours with 1.2 mM and b12F12 + 1.4 mM β CD. In the positive control sample 0.17 μg of B12F12 and 0.36 μg of β CD was added to an untagged cellular sample to have the resonance area of a sample with internalization factor equal to 1. The L929 measure is performed with 1024 NSA so a factor 2 should be used to compare its SNR to the other measures.
TR is the repetition time, NSA is the number of signal averages and SNR is the signal to noise ration.

Experiment	Peak (ppm)	TR	NSA	SNR
PANC1	-41.76	3 s	4096	4.8
L929	-41.53	3 s	1024	11.1
Positive control	-41.88	3 s	4096	448.2

Lysate sample F12B12 1.2 mM , beta CD 1.4 mM

—41.761

(a)

FIBROBLAST – Lysate sample 1.2 mM F12B12

—41.533

(b)

Positive control: lysate 1.2 mM in cellular volume, B12F12 = 0.17 ug

—41.881

(c)
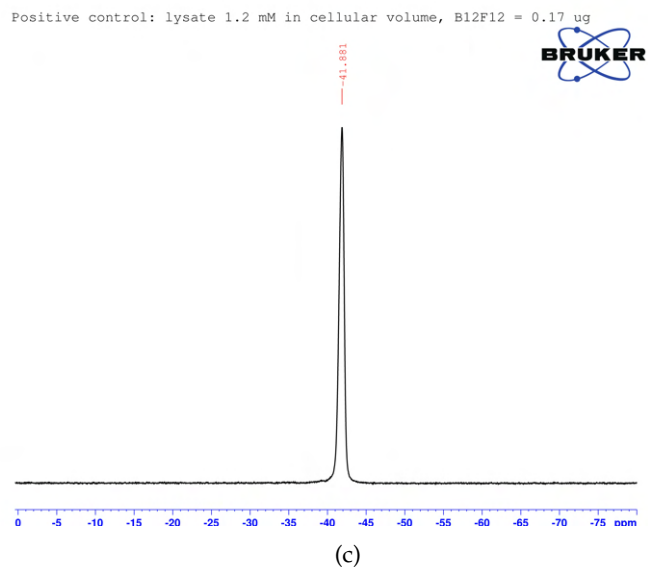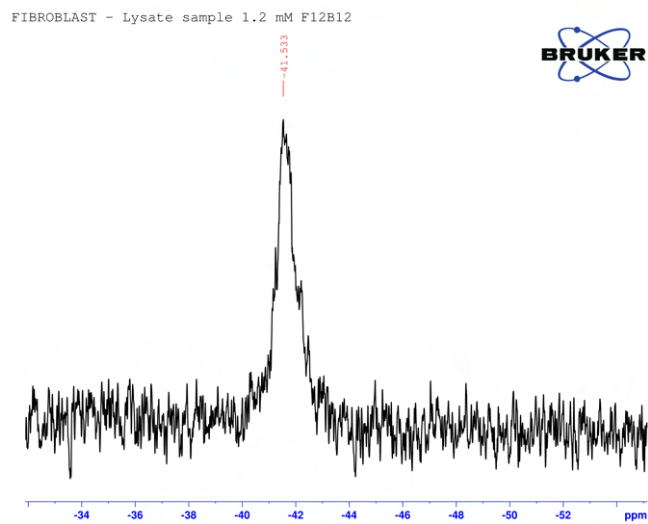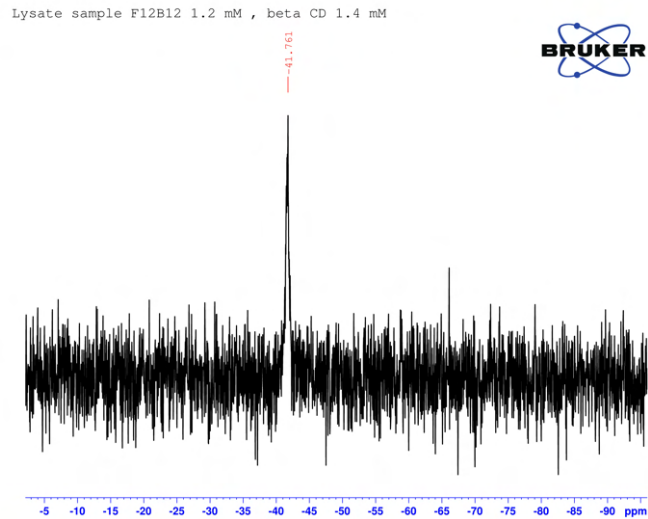
FIGURE 4.5: 19F-MRS spectrum of a) lysate sample of PANC1 cells incubated with B12f12 in complex with *beta* CD. b) lysate sample of L929 cells incubated with B12f12 in complex with *beta* CD. c) Positive control: lysate sample of L929 cells incubated with normal medium to which is added a quantity of B12F12 in complex compatible with an internalization factor equal to one.

(a)

(b)
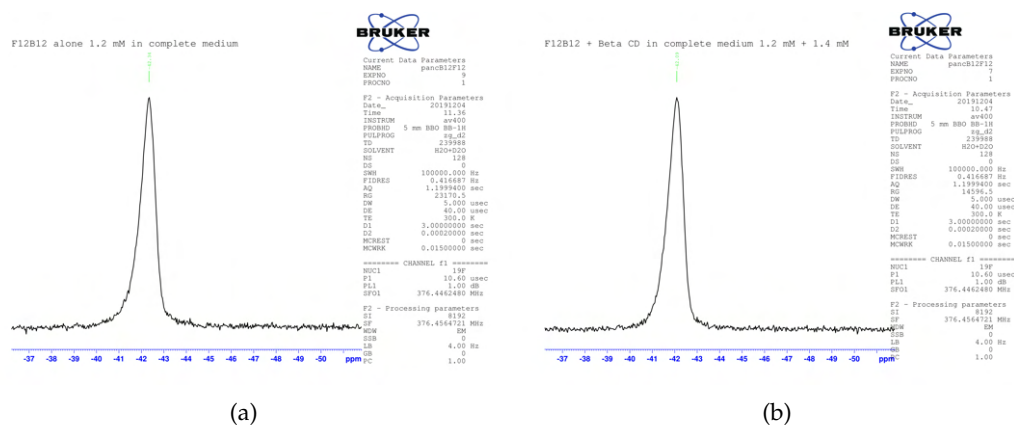
FIGURE 4.6: 19F spectra of a) B12F12 1.2 mM in complete medium and b) B12F12 1.2 mM + 1.4 mM *βCD* after heating to 45 and sonication. Resonance peak of B12F12 alone is - 42.34 ppm and -41.88 for the complex.

Chapter 5

The feasibility of F-BPA quantification in tissue trough 19F-MRS

Chapter Abstract

Boron quantification in tissue is essential for BNCT/PNCT treatments. Direct quantification of boron is usually made with mass spectroscopy or boron neutron autoradiography. BPA is one of the gold standard drugs in BNCT application and its biodistribution is usually indirectly monitored in vivo with a PET scan using 18F tagged F-BPA. In this Chapter we aim to develop a methodology to directly monitor the non radioactive isotope 19F-BPA trough 19F-MRS.

We tested the technique on an animal model ex-vivo on tissue and blood samples and we performed the fluorine quantification with 19F-MRS and the results are compared with the quantification obtained with neutron autoradiography.

19F-MRS may have the sensitivity to detect fluorine when F-BPA is used at BNCT treatment concentration. Unfortunately, tissue samples showed less fluorine accumulation than expected but our result may be affected by human error.

Still, 19F-MRS can be useful in Fluorine quantification in tissues when FBPA is administered at high dosage. Furthermore this technique is interesting since it can be used to screen for molecules with in-vivo imaging properties and can be transposed to in-vivo localized spectroscopy.

5.1   Introduction

In this Chapter we will report our first test to develop a 19F-NMR based methodology to asses the uptake of F-BPA in a orthotopic model of pancreatic cancer.

An accurate calculation of boron concentration (BC) in tumoral and normal tissues is necessary to determine the feasibility of a BNCT or PNCT treatment. One of the main limitation on BNCT is the risk of radiation damage in normal tissues localized near the treated tumoral site. To be more specific, BNCT is conditional to a suitable BC reached in tumoral tissue while a much lower concentration is present in surrounding healthy tissues. This quantity is referred as T/N (the ratio of BC in Tumor and healthy (Normal) tissue) and its determination prior to treatment is one of the main concern in the research aimed to improve this treatment. Direct quantification of BC is difficult in-vivo in a clinical setting so its often derived from boron concentration in blood or trough a PET scan of radio marked molecule.

The newly proposed PNCT, described briefly in Chapter 4, should mitigate this limitation.

In the context of this new treatment possibility we are investigating into the benefits of adopting 19F-BPA (instead of BPA) as a boron source in the treatment of pancreatic adenocarcinoma. Boronophenylalanine (BPA) is one of the gold standards in BNCT application : it is proposed as the boron carrier of choice in multiple tumoral site and it is used as a reference when other borated compound is proposed. Additionally the fluorine labelled analogous 18F-BPA is commonly used as an image-probe for PET applications and many studies support the possibility to derive BC after BPA administration trough 18F-BPA PET imaging. PET imaging can be used to estimate the BC of tissues before irradiation to calculate the radiation dose due to the nuclear reaction with boron and to confirm the safety of healthy tissues and the efficacy of the treatment.

While the usefulness of 18F-BPA PET for BNCT therapy with BPA is a well researched topic and its potential as tumor-imaging tool was demonstrated in various application (Ishiwata, 2019b), chemical differences between the two molecules and the difference in the administration protocol of the two compounds suggest that the reliability of BC estimate in tumor from PET scan should be examined on a case by case basis. The main differences between BPA and 18F-BPA can be summarized as follows (Watanabe et al., 2016): From a chemical prospective the different chemical structure caused by the addition of the fluorine atom gives rise to a change in the chemical properties: F-BPA is more soluble in water and in lipids than BPA and this difference in solubility has an impact in both the formulation adopted for the clinical use and more importantly the lipophilicity of F-BPA that may affect, in theory, the bio-distribution in tissues.

From an operative point of view, in BNCT BPA is administered at high dosage to help increase the BC in tissues since, to obtain the desired radiobiological effect, BC in the target tumor must be over a minimal threshold. Therapeutic dosage of BPA in BNCT is 250-900 mg/kg of body weight and, to obtain a suitable pharmaceutical preparation to reach this dosage, BPA is usually formulated as the fructose complex to increase solubility. This is also true when 19F-BPA is used instead as reported in (Porcari et al., 2006; Capuani et al., 2008). On the other hand, the radio-pharmaceutical analogue 18F-BPA is used at tracer dose without the formulation of the fructose complex. Moreover, especially in human clinical application, BPA is administered with a continuous infusion due to the large volume of the preparation while 18F-BPA is administered in a single shot. Significant variation in boron bio-distribution in tissues, when the same dose is administered in different ways, is reported for both BPA (Garabalino et al., 2011) and F-BPA (Watanabe et al., 2016).

The last note is on the detection modalities of the two compounds: boron accumulated in tissues after BPA administration is usually measured trough inductively coupled plasma atomic emission (ICP-AES) or boron neutron autoradiography while boron accumulation reached with 18F-BPA at tracer dose can't be reliability detected by those technique so 18F is manly measured by PET imaging instead.

These differences suggest caution into the assessment of the predictive value of F-BPA PET for BPA accumulation in the tumor and normal tissues: The research on the spread between the distributions of 18F-BPA and BPA is still an open problem and often consist in a difficult calibration between the two techniques and it may introduce constrains on the administration and treatment protocols for the BNCT (Lin et al., 2020).

As already discussed in Section 4.1 our case study is the pancreatic adenocarcinoma for which we are evaluating the applicability of PNCT. Unfortunately the pancreas is a well-known site of boron accumulation when BPA and F-BPA are used: the risk of damaging healthy tissue with neutron irradiation during the BNCT excludes this technique for pancreas cancer treatment and in any tumor near the pancreas (Hanaoka et al., 2014; Chou et al., 2009).

In particular, the need to safeguard healthy tissues has led to the proposal of invasive treatments such as liver autotrasplant following an extracorporeal irradiation (Pinelli, 1996) that may not be indicated for other organs that can be transplanted but for which this procedure is even riskier than for the liver.

We believe that PNCT may ease those constrain and its application may be promising: It has been suggested that the increase in radiobiological effectiveness due to the high-LET alpha particles produced in the p-11B reaction, whose maximum cross section is at low proton energy, corresponding to the Bragg peak can be localized in the tumor region with a precision of few milliliters.

To further strengthen this claim, we will review the past literature in Section 5.1.1 and argue that is reasonable to assume that the pancreas will have a selective uptake of F-BPA. The boron concentration will be high immediately after the administration of the drug and after there will be a rapid washout but also a long retention of boron. Our working assumption is that, as suggested in our in-vitro experiment described in Chapter 4, also the pancreatic adenocarcinoma will significantly internalize the F-BPA but the washout will be slower since tumoral tissues often exhibit an higher retention level. In addition, the administration of a large dose of F-BPA instead of BPA may open the possibility of a direct quantification of the target compound trough 19F-NMR. Our hope is that the difference in concentration combined with the precision of proton therapy and the possibility to have a direct method to quantify the BC in the target will be sufficient to lay the foundation of a boron enhanced radio therapy for the pancreatic adenocarcinoma.

5.1.1 Quantification of Boron for BNCT/PNCT for FBPA in pancreas

The quantification of BC in the pancreas is essential to test the feasibility of BNCT in the neighbor organs. For this reason quantification of BPA is available from the literature even if this molecule is not usually proposed in the BNCT treatment of pancreatic cancer. Often, also data regarding F-BPA used for PET application, so when it is administered in tracer dose, and the possibility of using PET signal to correctly estimate pancreas BC, are reported. On the contrary, we didn't find any previous work that quantify directly F-BPA administered at high dosage in the Pancreas to be used for treatment purposes.

Nevertheless it seems that both BPA at therapeutic dosage and F-BPA at tracer dosage are selectively accumulated in the pancreas. Also, it seems that a consensus has been reached on the relation between BC level in pancreas and in blood. The BC in pancreas is higher than the BC in blood and other tissues, with the probable exception of the kidneys but it may depend on the administration protocol, up to 3 h after BPA administration.

The first concern in the application of BNCT is radiation dose damage to healthy tissues. In this context Chou and collaborators (Chou et al., 2009) point out that

TABLE 5.1: Summary of the literature where BPA accumulation in the Pancreas is reported when BPA dosage is at BNCT therapeutic level. The table reports: publication year, animal model used, BPA dose administered, administration modality (Adminis.), time of sacrifice (tos) after the drug administration , Boron Content (BC) in pancreas expressed in *ppm* at each recorded tos, Boron Content in Blood expressed in *ppm* at each recorded tos, the ration between BC of Pancreas and BC of Blood (P/B), technique used in BC measure, if a quantitative 18F-BPA PET scan was also performed (PET 18F-BPA) and if a significant correlation between PET signal and BC measure was found for the Pancreas. Since uncertainties on quantities are not consistent between each other and data are often extracted from figures, only mean values are reported.

(*) means data was extrapolated and not directly reported by the authors

-	(Chou et al., 2009)	(LIN et al., 2012)	(Hanaoka et al., 2014)	(Grunewald et al., 2017)
Year	2009	2012	2014	2017
Animal	rats	mice	rats	mice
BPA dose	150 (*mg/kg*)	400 (*mg/kg*)	167 (*mg/kg*)	200 (*mg/kg*)
Adminis.	i.p.	tail i.v.	tail i.v.	i.v. (infusion)
Tos	3 h	3, 15, 30 / 45, 60 min	1 h	2 h
BC pancreas	17.4 *ppm*	-, 170, 170 / 118, 80 *ppm*	52.3 *ppm*	72 *ppm*
BC blood	4 *ppm*	-, 12, 11 / 10, 9 *ppm*	12.3 *ppm*	6.3 *ppm*
P/B	4.16	-, 14.2, 15.5 * / 11.8, 8.9 *ppm*	4.3*	11*
Technique	ICP-AES	ICP-AES	ICP-AES	PGAA
PET 18F-BPA	no	yes only tumor	yes	yes
Corr. with BC	-	-	yes	yes

in rats the pancreas has higher BC than other organs and blood: "The pancreas had a higher boron concentration than blood; the accumulation ratios of boron in liver,pancreas and kidney to boron in blood were 0.83, 4.16 and 2.47, respectively, in BPA-treated rats" and they clearly warn that: " BPA may not be a suitable agent for the BNCT of hepatoma because of the high uptake and retention of BPA in the pancreas." The reason is easily understood since any treatment attempt with BNCT may result in a higher radiation dose inside the healthy pancreas. Also in mice, pancreas uptake of BPA is higher than other tissues. In a related work, Lin and collaborators (LIN et al., 2012) measure pancreas boron uptake at 15, 30, 45 and 60 minutes after BPA injection and say that "the BC was mainly accumulated in the pancreas, and secondly in the kidney" at all measured times. He also shows that BC in pancreas increases rapidly after the administration of the drug reaching a maximum at 15 minutes and then it decreases rapidly in the following hour but it remains always higher than the BC in blood.

The same warning on the dangerous level of BPA accumulation in the pancreas is issued by Hanaoka and collaborators (Hanaoka et al., 2014) that specify how: "The highest concentration was found in the kidney, followed by that in the pancreas and the glioma[1] tumor tissue." and argue that: "The accumulation levels of both BPA and FBPA in the pancreas were remarkably high, ..., This finding suggests that radiation injury of the pancreas should be anticipated when BNCT is employed for abdominal cancers." In this study they also found a significant correlation between BC in healthy organs and the accumulation of F-BPA used for PET scans despite the difference in dose administered and in the chemical structure of the two drugs.

At the same time they also suggest that it is important to check whenever the administration protocol can affect the correlation between BC and PET signals:

"Further studies are required to estimate the 10B concentrations in tissues following BPA-fr[BPA in complex with fructose to help solubility -Ed.] administration according to a clinical protocol, such as slow infusion and drip infusion of BPA during neutron irradiation".

Most importantly, in order to motivate our suggestion to directly use F-BPA as a boron carrier, they noted how in their results the correlation between BPA and F-BPA concentrations and the subsequent opportunity to estimate BC from it, may not hold on human since the level of the enzymes responsible for the active cellular uptake in the pancreas are not the same in the two species: "Especially in the pancreas, the expression levels of LAT1 between rats and humans are different. The accumulation levels of F-BPA and BPA determined in this research may therefore not be applicable to clinical cases. Further clinical F-BPA PET study is required to determine BPA-fr distribution in the cancer patients.".

In a recent article, Grunewald and collaborators (Grunewald et al., 2017) confirm the higher BC of pancreas after BPA administration:

"At 110 min p.i.[past injection -Ed.] the order of accumulation starting with the highest was as follows for all study groups: pancreas, kidneys, bone, tumor, liver, heart, stomach and brain."

They also explore the possibility of determine BC from a PET exam and they are confident that this estimation method may hold:

"The major finding of our study is a significant correlation between F-BPA and BPA uptake in tumors and various organs. For both compounds high accumulation levels in pancreas and kidneys.".

---

[1]In their works,the authors studied a xenograft model of glioma implanted subcutaneously in the back of the rats

In table 5.1 we summarize some recent works that directly measure BC of pancreas when BPA is administered in small rodents detailing the dosage used, the administration method and how the measurement was performed.

Recent works that directly study the accumulation of boron in tissues from the administration of F-BPA at BNCT therapeutic dosage don't report the boron concentration reached in the pancreas (Capuani et al., 2008; Watanabe et al., 2017) but at least in the case of the work of Watanabe and collaborators (Watanabe et al., 2017) they found that, for the examined tissues, there is no significant difference between BPA and F-BPA pharmacokinetics when they are administered with the same dose, modality and in the same formulation. They found instead some differences in BC of tissues between different administration protocols that should be taken into account in dose calculation from 18F-BPA PET for BNCT.

While not extensive or exhaustive, previous literature clearly state that the pancreas selectively accumulate F-BPA, that the main limitation of 18F-BPA-PET is that it can not be performed at the same time of irradiation and that the possibility to estimate BC from PET scans have to be tested in each application. In any case, the possibility to detect F-BPA directly at therapeutic concentration may overcome those problems. This solution, obtained trough 19F-NMR imaging, has already been proposed for the glioma in rats (Capuani et al., 2008) but there is no developments of this modality for the pancreas.

This experiment may also contribute to transfer 19F-NMR imaging in this organ by giving the exact sensitivity constrain that have to be reached to quantitatively image properly in this context.

Thus, Magnetic resonance of fluorine has been already used to visualize F-BPA biodistribution in vivo (Porcari et al., 2006; Capuani et al., 2008) in tumoral tissue but there is no other attempt to direct quantification on fluorine in tissues.

The main goal of this experiment is to quantify the fluorine content in tissue (and particularly in the pancreas) after the intravenous administration of F-BPA. We aim to give a lower bound to the fluorine that can be reliably detected and find the optimal measurement modality to screen fluorinated compounds accumulation in tissue with this technique. The results of our measures of fluorine content in tissue samples will be cross-validated by a neutron autoradiography measure: The simultaneous use of those two techniques is significant since they are based of different physical properties and the quantification of boron and fluorine at the same time are useful to check for the stability of the tagged boron carrier that is important if an indirect estimate of BC is used since a modification in the molecular composition will make invalid any indirect quantification. This especially need attention if the nuclear reaction that exploits the presence of fluorine analog to the reaction with protons (described in Chapter 4) is used.

5.2   Materials and methods

5.2.1   chemicals and standards used

2-Fluoro-4-boronophenylalanine (19F-BPA) was synthesized as reported in (Kabalka et al., 2000b). PFTP-DOPA is an analogue of of the non-proteinogenic aminoacid 3,4-dihydroxy-L-phenylalanine (L-DOPA) tagged with perfluoro-tert-butoxy (Orlandi et al., 2020) and it was kindly supplied by the authors of the original paper. It is

used as the external standard in the quantification experiment. To shorten its $T1$ relaxation time, it was prepared in a solution of 0.026 mM $CuSO_4$. Isoflurane was in the wrong place at the wrong time. The structure of chemicals used is reported in Table 5.2.
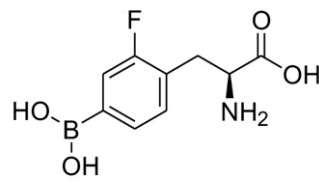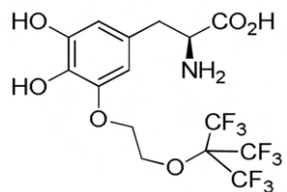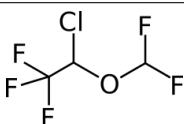
Name	Abbreviation	Structure
2-Fluoro-4-boronophenylalanine	F-BPA	
perfluoro-tert-butoxy 3,4-dihydroxy-L-phenylalanine	PFTP-DOPA	
Isoflurane	Isoflurane	

TABLE 5.2: Details of the chemical and the internal standards used for 19F-MRS quantification.

5.2.2 Animal model

Three male NOD Scid rats (4 weeks; 27.8 g, 22.0 g, 27.1 g) obtained from Charles River Laboratories, were used for this study. Mice were keep in a cage for a week before the experiment and had access to unlimited food and water. Human pancreatic cancer cell line (PANC-1) was purchased from ATCC (American Type Culture Collection) and grown as described in Chapter 4.

PANC-1 cells ($1.5 \cdot 10^6$) were implanted orthotopically as a cell/Matrigel mixture into the pancreas of two of the mice (mouse 1 and 2). The other mouse was left as control (mouse CTR). After three weeks 200 μL of F-BPA-fructose solution (see Section 5.2.3) was administered intravenously (i.v.) trough the tail vein. After 45 -50 minutes the mice were sacrificed and blood and tissue sample collected. Details can be found in Table 5.3.

This study was autorized from "Ministero della salute" (italian Ministery of Health) n°732/2019-PR released in 4 November 2019 ai sensi dell'art.31 del D.lgs. 26/2014.

5.2.3 Preparation of F-BPA-fructose solution

Solubility in water of F-BPA is low (2.6 mg/ml) so it is prepared as a complex with fructose to increase solubility.
There is a constrain (in terms of animal welfare) on the volume that can be administered in a single shot in mice and it is recommended to . Thus, the solution was prepared to deliver a dose of at least 200 mg/kg of F-BPA, that is in line with reported treatment dosage in BNCT (see Table 5.1), while keeping the volume under 10 $\mu L/g$

TABLE 5.3: Mice body weights (bw), F-BPA-f(ructose) solution dose at concentration 30 mg/ml administered in mice and time of sacrifice (Tos) after injection.

	Mouse CRT	Mouse 1	Mouse 2
bw (g)	27.8	22.0	27.1
F-BPA-f (vol.)(μL)	200	200	200
F-BPA (m.)(mg)	6	6	6
dose/bw (mg/kg)	216	272	221
Tos (min.)	50	45	45

of body weight. F-BPA-fructose solution has been prepared following a previously reported procedure in (Watanabe et al., 2017):

19F-BPA and fructose were dissolved in distilled water at a molar ratio of 1:1.5, and then a 1.15 M ratio of 1 N NaOH was added. The mixture was stirred until F-BPA had completely dissolved, and the pH value was titrated to 7.6 with 1 N HCl. Before the administration the solution was filtered through a 0.22-μm pore syringe filter for sterilization. The final concentration of F-BPA was set to 30 mg/ml. (For example, a mouse of 25 g of body weight that is administered with 200 μL of preparation with a concentration of 30 mg/ml, receives 6 mg of F-BPA that correspond to 240 mg/kg) Delivered dose is reported in Table 5.3.

5.2.4 Sample preparation

Blood samples were collected trough cardiac puncture in anesthetized (with $2 - 3\%$ isofluorane[2]) mice and refrigerated with heparin to avoid coagulation. Animals were euthanized via cervical dislocation under deep anesthesia, and intact tissues of interest (liver, spleen, kidney, pancreas, fat, skin, lungs, hearth, stomach, genitourinary system) were harvested. Blood and organs were weighted and tissue samples were stored in nunc vials and frozen in liquid nitrogen.

For 19F-MRS analysis each tissue sample was thawed at room temperature and afterwards they were mechanically homogenized.

Tissue and Blood samples were prepared for 19F-MRS in 10 mm vials with a sealed glass vial containing $300 \mu L$ of 0.29 mM solution of PFTP-DOPA, prepared with $CuSO_4$ to shorten its 19F T1 (see Section 5.2.1), was inserted inside the samples to be used as reference peak (rif.). The sample size/volume fit the whole receptive field of the NMR detector coil to assure that all 19F nuclei can be detected. Also, blood samples were diluted with PBS to mach the reference level to avoid discontinuity between the level of the reference and the level of the sample if needed, while lysis buffer (TRIS 50mM, Triton 1%, SDS 0,1%, NaCl 150mM, EDTA 5mM) was added in tissue samples.

Volume of blood collected and tissues weight is reported in table 5.4.

Tissue samples that were not analysed with MRS will be measured trough neutron autoradiography at Pavia.

---

[2]This detail was discovered at a later time

TABLE 5.4: Weight of the tissue samples collected after F-BPA injection and sacrifice. Blood samples were taken pre-mortem in anaesthetized mice. In red: Samples that were analyzed with 19F-MRS. All other sample are being analyzed with neutron autoradigraphy @Pavia

Sample	Mouse CRT	Mouse 1	Mouse 2
Blood (μL)	650	550	530
Liver (g)	1.45	1.06	1.36
Spleen (mg)	52	18	38
Kidney (mg)	393	364	464
Pancreas (mg)	130	88	149
Fat (mg)	162	104	187
Skin (mg)	78	128	186
Lungs (mg)	180	132	185
Heart (mg)	156	143	132
Stomach (mg)	290	248	365
Genitourinary sys. (mg)	219	266	323

5.2.5 19F measure

One-dimensional 19F spectra were acquired at a temperature of 296 K using a Bruker 400 MHz Avance NMR spectrometer equipped with a 10 mm 1H/19F-Probe with a (x,y,z)-gradient system to optimize the magnetic field homogeneity.

Acquisition parameters for each sample are reported in Table 5.6 and in Table 5.7. Data were zero-filled to 65,536 complex points prior to Fourier transformation, and an exponential window function was applied with a line-broadening factor of 5.0 Hz. The spectra were manually phase and baseline corrected and referenced to the PFTP-DOPA standard. The pulse repetition time was (TR=9s); Based on the parameters in Table 5.6 and in Table 5.7 the total acquisition time was approximately 10 h per sample.

The 19F signal of each tissue sample was determined by calculating the ratio of the F-BPA resonance peak (1.5 ppm) integrated area to the PFTP-DOPA reference (50 ppm) integral and confronted against the calibration line. The signal-to-noise ratio (SNR) and integral ratio of the signal were calculated using TopSpin 3.6 software (Bruker). The limit of detection (Lod) was defined as the 19F signal where the $SNR <= 2$. Errors on the integral areas ratio are derived from the noise level of the spectrum and from peak width but they are also operator dependent (as explained in (Holzgrabe, 2015)) since phasing and baseline correction are performed manually. This user-dependent variability in area estimation is reported to be up to 2% of the reference peak area. Taking into account this variability may result in an overestimation of the error but it should not impact the results when a signal is clearly recognisable (i.e $SNR > 3$).

For a quantification experiment total longitudinal relaxation should be achieved. To quantify T1, an inversion recovery (IR) acquisition sequence with 8 IR times, t_{IR} from 0.1 ms to 10 s were used, number of averaged scans N=512, repetition time TR=20 s. Data were fitted to the mono-exponential function:

$$S(t_{IR}) = S_0(1 - 2\exp\left\{-\frac{t_{IR}}{T1}\right\})$$

(5.1)

Blood samples, F-BPA in lysis buffer and the reference peak was tested. We found that T1 for F-BPA in blood sample (1.5 ppm), the isoflurane in blood (40 ppm), the reference peak and the F-BPA in lysis buffer were respectively (1140 ± 70) *ms*, (860 ± 90) *ms*, (1190 ± 70) *ms*, (750 ± 70) *ms*. Therefore we set $TR = 9s$ that should always satisfy the criteria $TR > 5 \cdot T1$ that guarantee the proportionality between peak area and the number of fluorine atoms (Holzgrabe, 2015). Spectra and inversion recovery experiment data can be found in Appendix **??**. Relaxation times of F-BPA in tissues were difficult to measure since the signal was low but it is reasonable to estimate them to be $< 2s$.

5.2.6   19F measure calibration

To test the detection performance and sensitivity of our measure we performed a series of acquisitions on F-BPA samples with a known FBPA molecules content. F-BPA content were prepared from serial dilutions as reported in Table 5.5.

Samples were diluted with lysis buffer to match the solvent used for tissue samples (see Section 5.2.4) up to a volume of 1.5 *mL* in a 10 mm NMR vial.

The 19F NMR spectra were acquired using: 17 1*mus* pulse, 74k points acquired for free induction decay (FID), 140 ppm spectral width, 4096 averages, 10 h acquisition time, and a recycle delay of 9 s. The results were plotted as the signal in reference peak area units versus 19F-BPA content measured for each sample as shown in Figure 5.1. A linear fit was performed on the experimental point $(y = p1 * x + p2)$;

In theory, the calibration if affected both by errors in the determination of the dilution (so the reference concentration) and the error over the integral area calculated. We can say that the latter is dominant since the error in signal determination when SNR is low can be difficult due to imperfect phasing and baseline correction while volume determination with air displacement pipettes in this volume range should be affected by a relative error less than 0.1 % of the nominal value (Ferroni and Letang, 1989). For this reason we are not considering concentration error in the fitting procedures.

Where y is the Peak Area Ratio and x is the F-BPA concentration measured in μ moles. Results from the fit are: $p1 = 0.635 - 95\%CI(0.590, 0.681)$, $p2 = 0.017 - 95\%CI(-0.018, 0.050)$, $R^2 = 0.9994$. From them is possible to estimate the boron concentration in ppm in the sample as:

$$BC = B_f \left(\frac{AR - p2}{p1} * M_w \right) * \frac{1}{m_{sample}} \ ppm \tag{5.2}$$

Where B_f is the percentage in weight of boron of the molecule, AR is the ratio between the area of the unknown signal and the area of the reference, M_w is the molecular weight expressed in $g/mole$ and m_{sample} is the mass of the sample in grams. This relation hold only if there is a relation between the number of boron a fluorine atoms in the molecule. In the case of metabolic defluorination where the fluorine atoms are stripped from the F-BPA the relation between boron and Fluorine concentration is invalid. This is also true for PET scans and it is one main bias sources in BC estimation from fluorine concentration measure (Romanov et al., 2020).

TABLE 5.5: Samples used for the calibration line and results from 19F spectra. Each sample consist in a known quantity of F-BPA-fructose in a 10 mm NMR vial. A sealed glass vial containing $300\mu L$ of 0.29 mM solution of PFTP-DOPA was inserted inside the samples to be used as reference peak (rif.). For each data point the following columns are reported: Moles of F-BPA in the calibration samples, Integral Area Ratio between the reference sample peak and the F-BPA total peaks area, SNR of the reference peak, SNR of the F-BPA peak. SNRs are calculated from magnitude spectra.

	μmol	**Area Ratio**	**SNR rif.**	**SNR F-BPA peak**
S1	1.30	0.84 ± 0.03	115.2 ± 0.5	47.2 ± 0.5
S2	0.65	0.43 ± 0.03	123.1 ± 0.5	17.2 ± 0.5
S3	0.13	0.09 ± 0.03	138.4 ± 0.5	4.0 ± 0.5
S4	0.06	0.05 ± 0.03	109.3 ± 0.5	3.4 ± 0.5

5.3 results

5.3.1 In blood and tissue boron quantification

19F spectra analysis performed on blood samples shows an unexpected signal at 41 ppm and 35 ppm respect to the reference 0 (F-BPA in water). These peaks are shown in Figures 5.2 and the signal relative to reference peaks is reported in Table 5.6. A large chemical shift can be the result of drugs metabolization, that is effectively the breakdown of the molecules in different compound, but it is difficult to attribute these signals to F-BPA injection due to their large signal. We feared that there may have been an external source of fluorine and after we realized that anesthesia was performed with isofluorane and this is the most likely source of that line, albeit still under investigation. From Table 5.6 appears that this area peak also diminish in time. The blood was stored in a sealed but not airtight container so this compound may still be volatile.

Only Mouse 2 had a detectable signal raised from F-BPA in blood. From the calibration (eq. 5.2) we can estimate the fluorine content and the boron concentration in blood. Its peak area is consistent with (0.26 ± 0.05) $\mu moles$ of F-BPA. The error associated is propagated from error in Area Ratio estimate and errors in the calibration line coefficients. Considering the molecular weight of F-BPA ($Mw_{F-BPA} = 227g/mole$) and its boron content (4.8% of natural boron in weight) that correspond to (2.8 ± 0.5) μg of natural boron. Given the blood volume (see Table 5.4) the boron concentration estimate is (5 ± 1) ppm.

Tissue samples shows a weak signal from F-BPA in mouse 2 genitourinary system and mouse CTR kidney. Signal in Mouse 2 sample can be quantified (0.08 ± 0.05) $\mu moles$ that correspond to (0.9 ± 0.5) μg of natural boron. Given the organ mass (see Table 5.4) the boron concentration estimate is (4 ± 2) ppm. Signal in Mouse CTR kidney is too low to be reliably quantified. All measures in the pancreas fail to shown any meaningful source of signal outside the one related to isofluorane. Details can be found in Table 5.7.

TABLE 5.6: Quantification for the blood samples from 19F spectra. Acquisition parameters are reported: Number of averages (N), Number of FID points (FID pts.), Sweep Width (SW). A sealed glass vial containing $300 \mu L$ of $0.29 \, mM$ solution of F-L-Dopa was inserted inside the samples to be used as reference peak (rif.). For each sample the following columns are reported: Integral Area Ratio between the reference sample peak and the F-BPA total peaks area (A.R. F-BPA), SNR of the reference peak (SNR rif.), SNR of the F-BPA peak (SNR F-BPA). SNRs are calculated from magnitude spectra. Integral Area Ratio between the reference sample peak and the Isoflurane total peaks area (A.R. ISOF.). In notes column is indicated if the acquisition was performed in the same day of the sacrifice (fresh) or in the following weeks (w)

sample	N	FID pts.	SW	SNR rif. (mag)	SNR F-BPA (mag)	A.R. F-BPA	A.R. ISOF.	notes
Mouse CTR	4096	74k	140	173 ± 1	< 2	< lod	1.30 ± 0.03	fresh
Mouse CTR	256	56k	135	80 ± 1	< 2	< lod	1.18 ± 0.05	1w
Mouse 2	4096	74k	140	211 ± 1	11 ± 1	0.19 ± 0.03	0.60 ± 0.03	fresh
Mouse 2	4096	19k	123	235 ± 1	11 ± 1	0.18 ± 0.03	0.28 ± 0.03	2w
Mouse 1	4096	19k	123	208 ± 1	< 2	< lod	1.40 ± 0.03	fresh

TABLE 5.7: Quantification for the tissue samples from 19F spectra. Acquisition parameters are reported: Number of averages (N), Number of FID points (FID pts.), Sweep Width (SW). A sealed glass vial containing $300\mu L$ of $0.29\ mM$ solution of F-L-Dopa was inserted inside the samples to be used as reference peak (rif.). For each sample the following columns are reported: Integral Area Ratio between the reference sample peak and the F-BPA total peaks area (A.R. F-BPA), SNR of the reference peak (SNR rif.), SNR of the F-BPA peak (SNR F-BPA). SNRs are calculated from magnitude spectra. Integral Area Ratio between the reference sample peak and the Isoflurane total peaks area (A.R. ISOF.) and the mass of the sample before homogenization. In notes column is indicated if the acquisition was performed on the homogenized sample and if the reference was present.

| sample	N	FID pts.	SW	SNR rif. (mag)	SNR F-BPA (mag)	A.R. F-BPA	A.R. ISOF.	notes	mass (mg)
pancreas 2	4096	67k	140	164 ± 1	< 2	< lod	0.18 ± 0.03	-	149
pancreas 2	4096	74k	140	160 ± 1	< 2	< lod	< lod	homo.	149
pancreas 2	4096	22k	280	none	< 2	-	-	homo.; no rif.	149
pancreas CTR	4096	19k	123	211 ± 1	< 2	< lod	< lod	homo.	130
pancreas CTR	4096	19k	123	209 ± 1	< 2	< lod	< lod	homo.	130
kidney CTR	4096	74k	140	270 ± 1	5 ± 1	0.07 ± 0.03	0.12 ± 0.03	homo.	293
genitour.sy.2	4096	19k	123	218 ± 1	5 ± 1	0.05 ± 0.03	trace	homo.	232
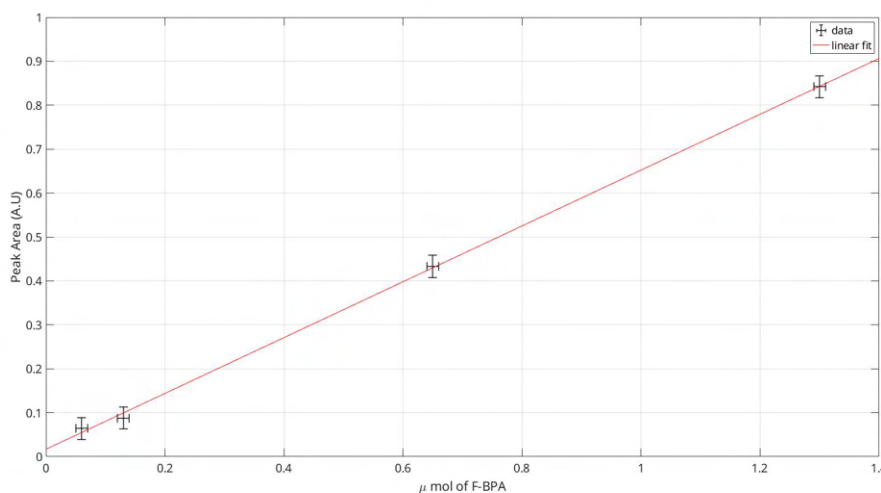
FIGURE 5.1: Calibration samples signal vs. samples F-BPA content. Signal (Black crosses) is expressed in reference peak area units. Red line: linear fit ($y = p1 * x + p2$); $p1 = 0.635$ - $95\%CI(0.590, 0.681)$, $p2 = 0.017$ - $95\%CI(-0.018, 0.050)$, $R^2 = 0.9994$.

5.3.2 Isofluorane quantification in blood

We can quantify isofluorane concentration in blood since for both resonance peaks we have full relaxation. From the ratio of resonance peak area in Table 5.6 we derive, knowing that isofluorane has 5 fluorine atoms, that: The concentration in blood on Mouse CTR was (0.65 ± 0.03) mM for the first measure and (0.56 ± 0.04) mM after 1 week. Mouse 2 had an isofluorane concentration of (0.33 ± 0.02) mM for the first measure and (0.17 ± 0.03) mM after 2 weeks. Mouse 1 had an isofluorane concentration of (0.83 ± 0.04) mM. Both Mouse CTR and Mouse 2 exhibit a lower concentration in the second measure: Isofluorane may remain volatile in blood and can be released in air during the sample conservation since the NMR vials were closed but not airtight. Unfortunately, the original gas mixture concentration and the exposition time were not recorded and in blood concentration of anesthetic gasses depends strongly on both.

5.4 Discussion

During the analysis of data acquired in the experiment we become convinced that an anomalous amount of fluorine was present in the blood samples. We later discover that isofluorane instead of ketamine and xylazine was used to anesthetize the mice before the cardiac puncture. As shown in Table 5.2, isofluorane contains 5 atoms of fluorine and, from data available in literature (Mandal and Pettegrew, 2008; Chen et al., 1992), we can reasonably attribute those signals to it. The main peak is raised by (at 40.8 ppm) the CF_3 group and the secondary peak (at 35.4 ppm) by the CHF_2 group. The distance between the two peaks is consistent with the cited literature. Currently an experiment with a sample of blood of an anaesthetized mouse is been performed to dispel any doubt.

Clearly human error had an impact on our experiment but, once we had confirmed the presence of other fluorine sources, our results should not be negatively impacted since the isofluorane resonance do not overlap our signal of interest.

Mouse 1 Blood Sample

(a)
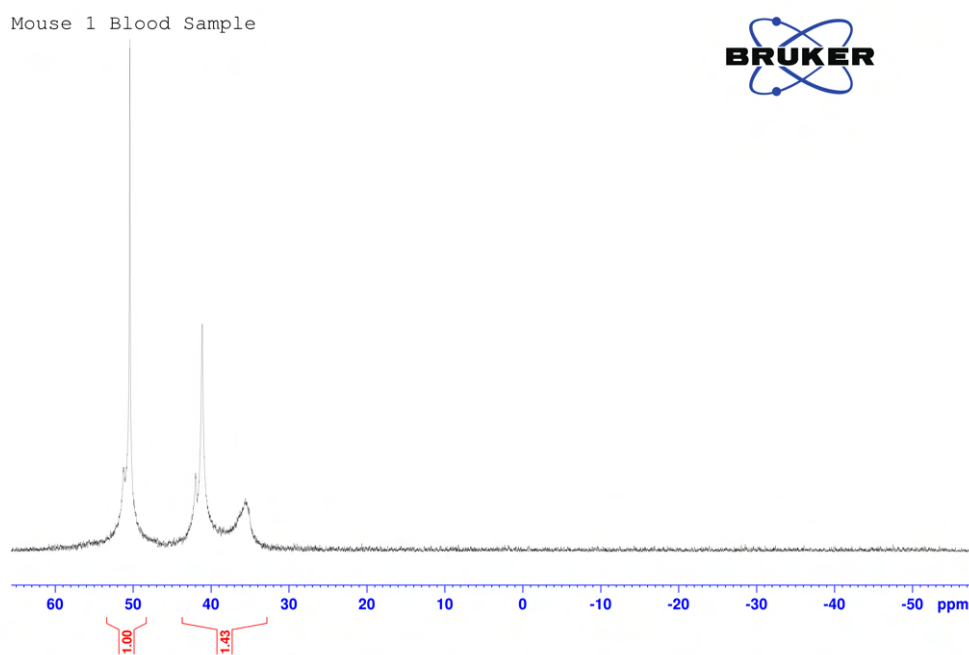
Mouse 2 Blood Sample
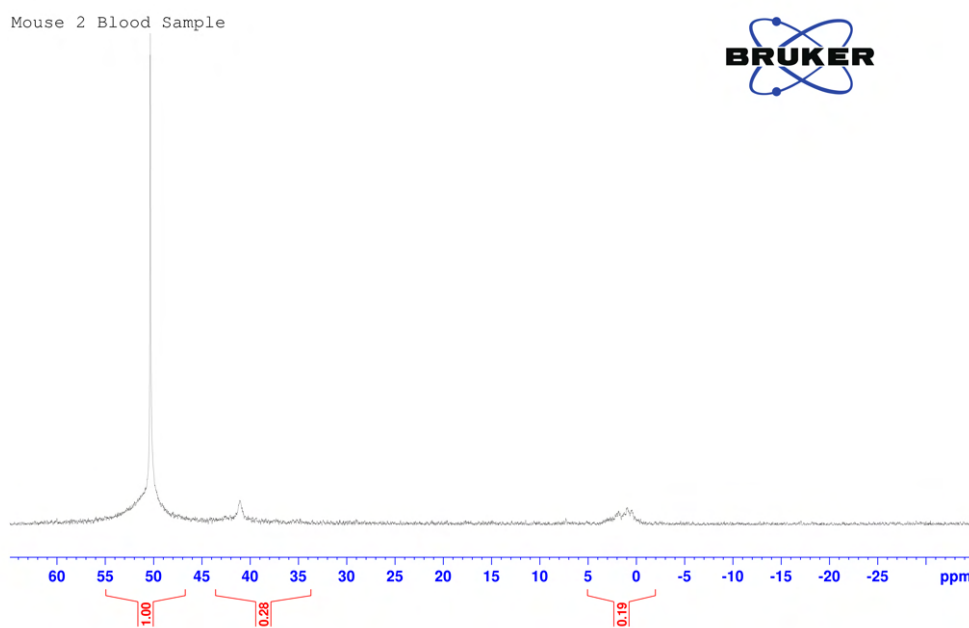
(b)

FIGURE 5.2: Spectra on blood samples referenced to F-BPA resonance in water. Red notes are the area integral normalized to the area of the reference peak. (a) Magnitude Spectrum of Mouse 1 blood sample (b) Magnitude Spectrum of Mouse 2 blood sample (c) Magnitude Spectrum of Mouse CRT blood sample.
Reference peak (FTP-DOPA) in at 50 ppm, two other peaks are visible at 41 and 35 ppm and we attributed them to the anesthesia with isoflurane. Only in Mouse 2 sample we were able to detect in blood a signal related to F-BPA.

(a)

(b)

FIGURE 5.3: Spectra on blood samples referenced to F-BPA resonance in water. Red notes are the area integral normalized to the area of the reference peak. (a) Magnitude Spectrum of Mouse CTR kidney (b) Magnitude Spectrum of Mouse 2 prostate.
Reference peak (FTP-DOPA) in at 50 ppm, two other peaks are visible at 41 and 35 ppm and we attributed them to the anesthesia with isoflurane.

Our main problem is the absence of signal from F-BPA in the pancreas samples. It is clear from the works reviewed in Section 5.1.1 that pancreas sho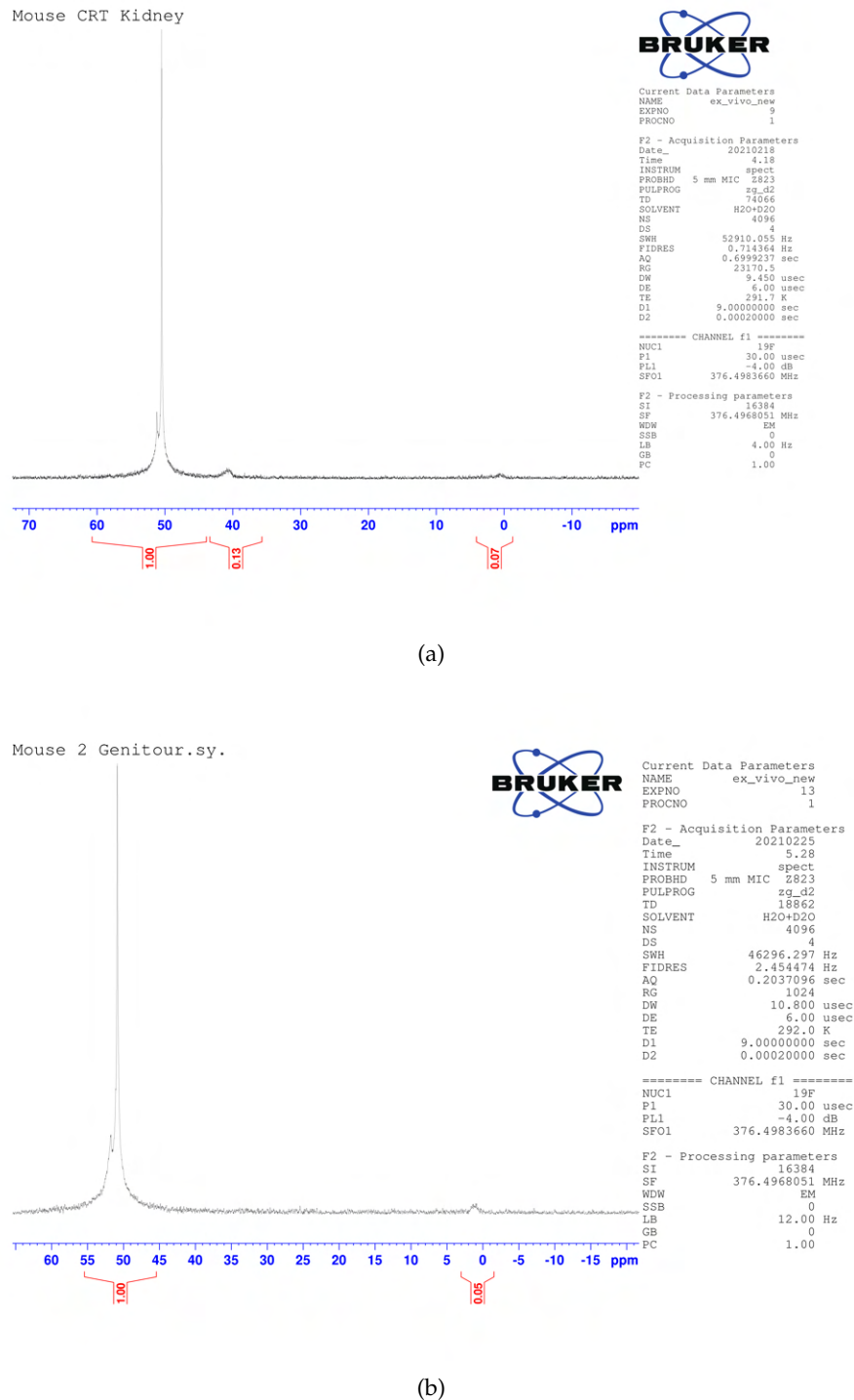uld have a large boron uptake and at least 18F-BPA used in PET accumulate in it. It is also fair to assume that the ratio between BC in pancreas and BC in blood is always larger than 1 for intravenous administration of BPA. However we fail to detect any sign of it.

Samples of Mouse 1 was sent to the the BNCT lab in Pavia to be analyzed trough neutron autoradiography. Results for the pancreas sample can be found in Figure 5.4. From preliminary results, we understand that the signal measured is close to their detection limit so it may be around 1 ppm of 10B that correspond to 5 ppm of natural boron. Since pancreas are small, so total fluorine concentration in the sample is low compared to larger samples at the same concentration, their measure is consistent with our results that put under our limit of detection the fluorine content of pancreas of Mouse 2 and Mouse CRT. It also worth noting that we did not find F-BPA signal in blood of Mouse 1 so our observation may be consistent with the measure of the pancreas on Mouse 1.

There are multiple explanations that are currently under evaluation: F-BPA preparation may have been not pure or at a much lower concentration than reported. The same batch was used for the calibration line and purity was tested by an independent 1H-MRS exam but we do not test directly the sample injected. For the next experiment we will save a fraction of the administered preparation for later analysis. Another explanation can be human error in the injection trough the tail vein, although operators tend to exclude it. This procedure may be tricky and an error can result in the dispersion of the drug outside the mouse body or a subcutaneous injection of the drug. In the latter case the pharmacokinetics of F-BPA may be different and probably much slower. In the next tests we will take a sample of the mouse tail to evaluate this possibility.

In this experiment we aimed to test 19F-MRS as a tool for fluorine quantification in tissues to evaluate BC for BNCT. We met this results in half: this technique can detect low amount of F-BPA in blood and tissues but we fail to reproduce previous results with our experimental model. This is probably cause by our inexperience with animal models.

In general, the main limitation of 19F-MRS are sensitivity and that it is only applicable to fluorinated drugs. The first limitation is partially removed by our primary task that is to find high accumulation in tissue prior to a BNCT experiment. In this case a low limit of detection compared to other techniques is less important. The second limitation is shared with FBPA-PET that is the currently preferred method for BC estimation (Ishiwata, 2019b) and may become a strength if the nuclear reaction of proton and fluorine will be proved as radiobilogical relevance. Another weakness is the duration of our experiments. Ten hours is unreasonable for most application. We choose this duration principally because we had to confront our results with neutron autoradiography, a technique that has a greater sensitivity.

Nevertheless, with the correct setup we believe we can reach the accuracy in detection needed to screen compound to be tested as PNCT/BNCT drugs and we believe that localized 19F-MRS can be eventually used in vivo and our results may provide a calibration benchmark for future experiments.
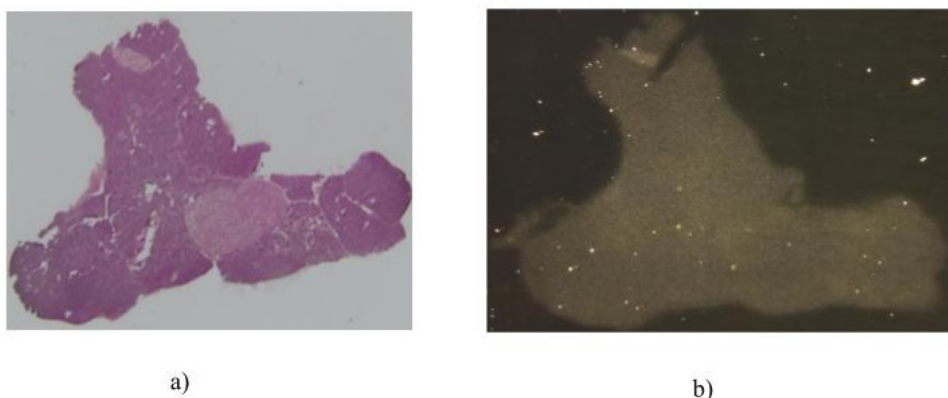
FIGURE 5.4: (a) Histological analisis of a pancreas sample stained with hematoxylin-eosin. (b) qualitative autoradiograf of a pancreas section of Mouse 1. Images courtesy of Silva Bortolussi, unpublished results.

5.5 Remarks

5.5.1 Tumor growth monitoring

This experiment has highlighted two different questions on our experimental protocol. First, our result are not consistent with what is reported in literature. Second, we still are not sure how reliably we are able to grow a tumor in our orthotopic cancer model.

The first concern raises from the fact that in this single preliminary measure we are not finding FBPA accumulation in the pancreas when it seems to be a consensus that the pancreas is the first or second most active organ for BPA/F-BPA uptake. In this case I think that human error is the most probable cause since our two orthogonal techniques didn't find any relevant amount of boron/fluorine in any on the tested organs.

Regarding the second concern, one of the main setback in our experimental design was the difficulty in monitoring the tumor growth. A critical part of our experiment is a way to assure that the tumor reach a suitable size before F-BPA administration. I performed CT scans two weeks after the inoculation of the tumor (See Appendix **??** for an example). The diagnostic image was negative and the veterinarian was unable to detect any tumoral mass in both the tested mice. We still decided to perform the treatment with F-BPA the following week and we confirmed that no tumoral growth visible to the naked eye was present. The absence of tumoral growth was later proved by histological analysis.

We need to device a strategy to monitor the growth of the tumor mass. A fast methodology for this screening can the monitor of the level of blood sugar level: Pancreas adenocarcinoma are know to disrupt organ functionality and to unbalance the insulin production. Blood sugar level imbalance may be one of the first symptoms of tumoral growth so we may implement hyperglycemia monitoring. Ecography can also be useful for this task but this exam need a skilled ecographist that was not available in our facility and its sensibility is more operator dependant than CT-scans.

Even considering the difficulty we faced in this first experiment, I still believe that

our choice of an ortothopic pancreas cancer model should be preferred to a xenograft model since the increase in technical difficulty and the greater probability of failure are completely justified by the scientific relevance in respect to our hypothesis.

Even if our tests did not reach the original goal, data collected has been extremely useful to make our protocol more robust and to develop techinal steps is some delicate process. Furthermore this first experimet let us test our collaboration between the laboratory at ISS, the NMR facility at CNR and the autoradiography lab in Pavia.

5.5.2 Sample preparation

In this experiment sample masses ranged from 50 mg to 1.5 g and volumes ranged from few 50 μL to 1.5 mL. This made sample preparation non trivial. In Figure 5.5 there is an example on the problem of working with large volume samples and with an external reference. This problem can be easily overcome by using different probe sizes and a different reference sample for lower volume samples but for technical reasons we decided to dilute the samples when needed instead. The next experiment will be tried with an internal standard to be added to the solvent in our samples. Especially for tissue samples we are evaluating the possibility to increase the concentration to keep the fixed volume to less than 500 μL by drying the homogenized tissues and reconstituting them in methanol, as described in (James et al., 2017) for excreta analyses, to maximize the concentration of the sample and to minimize the solvent/analyte ratio. Another important steps in sample preparation is to exploit the possibility to measure the same sample with both our techniques. Neutron autoradiography needs a few slices of the samples hundreds of micron thick while 19F-NMR can work on the whole organs. They can be used together to have both the global and local F-BPA concentration in the sample.
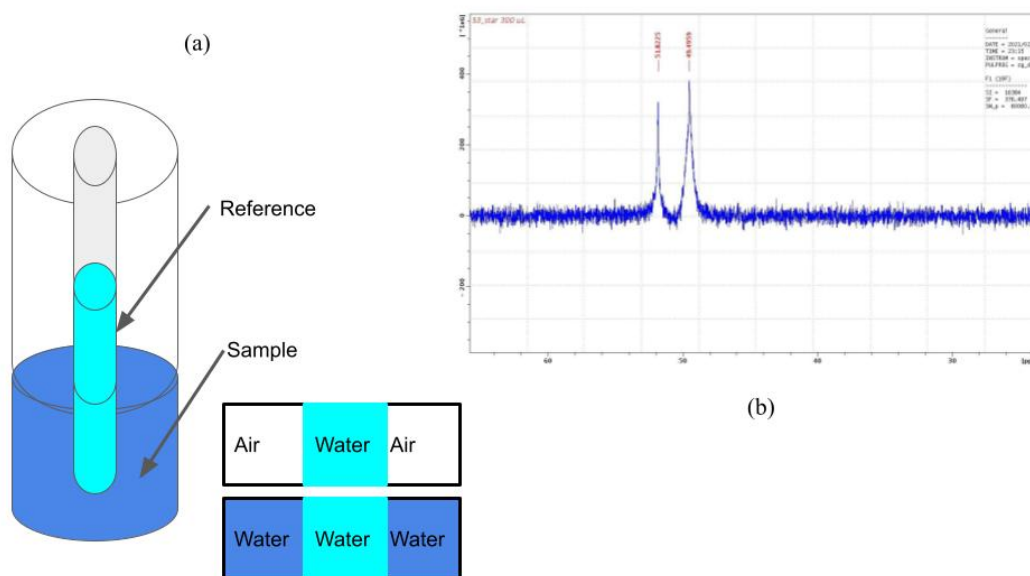

FIGURE 5.5: (a) Schematic configuration of our setup with two coaxial cylindrical vials. (b) Example of the effect of the discontinuity on the homogeneity of the field that results in the reference peak to be split. The effect has been increased to be seen clearly. In the case of this kind of dishomogeneity a "shimming" correction usually fail

Chapter 6

MRI denoising in K-space

Chapter Abstract

Magnetic Resonance Imaging (MRI) raw data is acquired in the frequency domain (k-space) and then transformed to magnitude image that are commonly used in many clinical applications. MRI can be afflicted by noise that can diminish the image quality and clinical information available. In this Chapter we will propose a new deep learning method to mitigate the effect of the noise directly on the data acquired in the k-space.

We simulated with the MRiLab software a dataset of MRI acquisitions in the k-space and their respective magnitude images. We tested the performance of a denoising convolutional neural network (DnCNN)in the denoising task for both the images in k-space and in magnitude images.

When the denoising task is performed in the k-space both the Peak-Signal-to-Noise-Ratio and visual quality is improved respect to the denoising performed directly on the magnitude images. Also, when used before the measure of relaxation times from noisy images, denoising in the k-space improves the estimation significantly.

We found that on our simulated dataset a DnCNN applied on frequency data is superior to a network with the same architecture applied on magnitude images. Thus, if a neural network approach has to be implemented in a MRI acquisition pipeline it should be applied to the data in k-space.

Our next step is to perform additional test on real data for a complete assessment of the proposed model.

6.1   Introduction

Magnetic Resonance Imaging (MRI) is a diagnostic image modality that has found a large number of applications in clinical practice. MRI can be greatly impacted from noise, thus noise removal (denoising) is a key preprocessing steps in many image analysis tasks. In this Chapter we will test a deep-learning neural network based approach for noise reduction in MRI. Our aim is to evaluate when, during the acquisition pipeline of a MRI, a simple convolutional neural network (CNN) can bring the greatest benefits.

MRI is based on a phenomenon known as Nuclear Magnetic Resonance (NMR). When certain atomic nuclei are subjected to a Magnetic field, they rotate with their own characteristic frequency that depends on the strength of the magnetic field (and the nature of the atoms). Atoms are able to absorb energy at this frequency if they are excited with a radio frequency (RF) pulse, and once the pulse is removed, they emit this electromagnetic energy at the resonance frequency. This principle can be used to obtain images: Since the resonance frequency depends on the strength of the magnetic field applied, if a spatial gradient of the magnetic field is present (so different locations are associated with different magnetic field strengths) atoms at

different locations will resonate at different frequencies. Thus, listening to different frequencies is the same as studying different locations. In medical imaging this effect is used to study the properties of tissues in the imaging field. The important message for the following sections is that the acquisition is made in function of frequency and not of space.

We do not need the details of the physical process involved (our CNN models aims to be only data driven) but one of our performance metric, which we will use to test our denoising algorithm, is based on correct estimation of atoms relaxation times from MRI acquisitions so we need to define them. A more rigorous definition can be found in Section 3.1.1.

At equilibrium the spins of the atoms are aligned with the external magnetic field (B_0), the cumulative effect of all the spins of the atoms aligned is called the net magnetization vector that, in this situation, is parallel to the external field. When a specific RF pulse is emitted, it causes the net magnetization to flip to a certain angle so two component of the magnetization can be defined the longitudinal magnetization (the projection on the field axis, M_z) and the transverse magnetization (the projection on the plane orthogonal to the field, M_{xy}).

After the excitation each atom returns to its equilibrium state and the net magnetization is again parallel with the external field: This phenomenon is called relaxation and it is associated with two independent physical processes.

Thus first, spins will rapidly dephase after the excitation occurs, pointing in all directions perpendicular to the static B_0 field and, as a result, removing the transverse component created by the RF pulse. These effects cause spins to precess at different frequencies that are related to the external B_0 depending on their position.

Second, when the RF pulse is removed, the B0 field is still present, thus the spins tend to point along the field's direction, rising the longitudinal component. Both processes occur at the same time and are basically independent, though the first one is generally much faster. Therefore, there exist two relaxation times:

T1: is the time for the longitudinal component M_z to return to its original state through the emission of electromagnetic energy. This is called the spin– lattice relaxation, corresponding to the exchange of energy between the spin system and its surroundings.

T2: is the time for the transverse component (M_{xy}) to return to its original state, associated with thermal equilibrium between spins. This is the spin–spin relaxation. Both relaxation times T1 and T2 refer to the time constant of the exponential laws ruling the two relaxation processes.

Measuring relaxation times of the longitudinal M_z and transverse components M_{xy} , different properties of the tissues may be inferred(Aja-Fernández et al., 2016): Contrast in morphological MRI acquisition are obtained by exploiting the difference in T1 and T2 of different tissues. This acquisition modality is called T1 (or T2) weighing.

6.2   Magnitude images and Rician Noise

From the simplified scheme in Figure 6.1 we can see how the process of acquiring an MR image starts in the frequency domain (the k-space) and it is acquired through a quadrature detector that provides the real and imaginary part of the signal. Each parts of the complex signal can be assumed affected by withe noise (i.e Gaussian distributed with zero mean), the main source of this noise is the RF coil resistance (Edelstein et al., 1986) and its final effect on the quality of the images depends on a

variety of factor such as the pixel dimension, the duration of the acquisition and the receiver bandwidth.

From the k-space, the real and imaginary images are reconstructed through the complex Fourier transform in the x-space. In the complex x-space the noise is still Gaussian and the real and imaginary part of the noise can be assumed uncorrelated, since the Fourier transform is a linear and orthogonal transform (Gudbjartsson and Patz, 1995). Then, the magnitude image is formed by calculating the magnitude of the signal pixel-by-pixel from the complex image. This step is crucial because in the magnitude image the noise is no longer additive and it is no longer Gaussian.

Clinical MRI are in most cases magnitude images but is possible to find image derived from the phase of the complex image. In any case, magnitude images are the most common, since phase artifact can be avoided by discarding the phase information (Gudbjartsson and Patz, 1995), so we will concentrate on them.

A noisy magnitude image can't be divided in a part of signal and a part of noise since, as we said, the noise in no longer additive: The probability distribution of intensities in a noisy magnitude image M, reconstructed from an image of signal I and Gaussian noise σ, assumed equal for both the real and imaginary parts, is given by

$$p(M) = \frac{M}{\sigma^2} e^{-\frac{(I^2 + M^2)}{2\sigma^2}} I_0\left(\frac{IM}{\sigma^2}\right) \tag{6.1}$$

Where I_0 is the modified $0 - th$ order Bessel function. This is the Rice distribution. A Gaussian approximation of this distribution is possible only if $I/\sigma \gg 1$(that is the Signal-To-Noise-Ratio (SNR) in the x-space). Magnitude image with high noise level will be far from the Gaussian approximation of their signal and they will suffer from the so-called "rician bias".

Thus, clinical MRI with low SNR, in addition of being difficult to read and interpreted, can also lead to the erroneous quantification of physical quantity since the noise has not zero means and it is not additive.

For example in T2 relaxation images, the accuracy and precision of the measured T2 may be substantially impaired by the low signal-to-noise ratio of images available from clinical examinations (Raya et al., 2010) or in diffusion weighted images the decreasing SNR at increasing diffusion weighting causes systematic errors when calculating apparent diffusion coefficients(Dietrich, Heiland, and Sartor, 2001).

The effect can be less important in morphological scan since new generation scanners can achieve excellent imaging quality but the effect of Rician noise is still a problem in many new acquisition modality (Diffusion - Kurtosis for example,(Glenn, Tabesh, and Jensen, 2015)) or when the low signal is given by the low concentration of the excited nuclei as is the case in fluorine magnetic resonance imaging (Taylor et al., 2016).

This phenomena are very well known and several correction schemes have been proposed to extract the real intensity I from the magnitude images and they often involve the estimation of the variance of the noise in the complex space. A simple review on the formulation and the proposition of an analytical correction scheme can be found in (Koay and Basser, 2006).

In this Chapter we are interested in verifying if a simple neural network with a supervised training approach can help mitigate this problem.
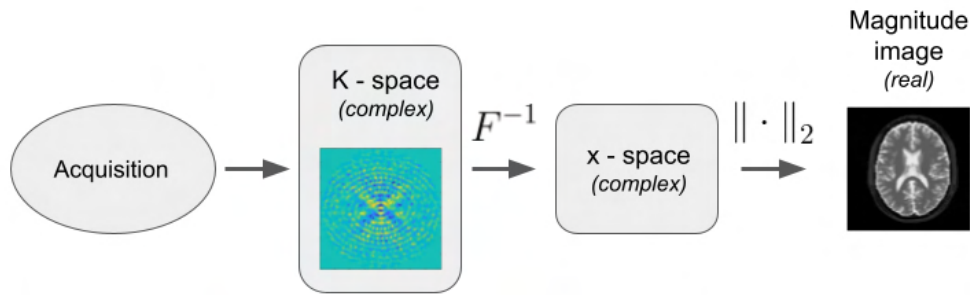
FIGURE 6.1: Magnetic Resonance Images are acquired in the frequency domain, then transformed to geometric space trough an inverse Fourier transform. The image usually used for medical application is the magnitude image, that is obtained calculating the magnitude pixel-by-pixel of the complex x-space image. Phase information are often discarded. Magnitude images are also referred as power images (improperly) or reconstructed images.

6.2.1 Denoising Magnitude Images with residual learning

Image denoising is defined as the task of removing the effect of external noise from an image. Thus, denoising an image should restore it (the noisy image) to the condition it was before the application of noise (the original image). The performance of an algorithm that performs this task (a denoiser) are evaluated on the closeness of the restored image to the original image. However, since noise, edge, and texture are high frequency components, it is difficult to distinguish them in the process of denoising and the denoised images could inevitably lose some details. (Fan et al., 2019)

In fact, image denoising is a classic problem and many solution has been proposed especially for the processing of natural images. The features of a good denoiser for natural images can be defined as: Flat areas should be smooth, edges should not be blurred, textures should be preserved, and new details absent in the original image (artifacts) should not be generated.

In practice, denoising algorithm are often evaluated on the base of a quantitative metric that measure the distance between the processed image and the original image, the Peak-Signal-to-Noise-Ratio (PSNR, See Section 6.3.1) and from visual inspection. Many other metrics have been proposed but the PSNR is the most adopted one.

This problem, namely the impossibility to directly use the desired performance metric as the loss function of a learning model, is discussed in more detail in Section 2.2.

In particular, the cited good feature of a denoised image are more qualitative than quantitative in nature so it is often very difficult to measure them correctly in all the possible applications.

Recently, deep learning methods have also proposed to denoise MR images by training different architectures with pairs of noisy and noise-free input-outputs (supervised learning approach).

The denoising of MRI has also an additional challenge: physical quantities that can be measured from the image should not be influenced by the effect of the denoising algorithm. This request may seems obvious if we want to exploit NMR imaging

as a quantitative tool: Quantitative MRI (qMRI) can be defined as maps of meaningful physical or chemical variables that can be measured in physical units and compared between tissue regions and among subjects (Pierpaoli, 2010).

If we follow this definition, still today most clinical MRI are not rigorously quantitative in nature thus, we should consider reasonable that most denoising algorithms do not explicitly measure quantitative based performance and rely only on standard denoising metrics, together with visual inspection, to evaluate their results. See for example (Manjón and Coupe, 2018) for a recent attempt.

Our experiment aims to evaluate the performance of a of feed-forward denoising convolutional neural networks (DnCNNs) that is applied to the denoising of magnitude images both directly on the image and on the raw frequency data in k-space.

The method chosen is derived from the work of Zhang and collaborators (Zhang et al., 2017) that adopt a strategy based on deep architecture, residual learning (RL), and regularization method into image denoising.

Specifically, residual learning and batch normalization (See Section 2.4.5) are utilized to speed up the training process as well as boost the denoising performance.

We are particularly interested in the residual learning approach that aims to gradually removing the latent clean image in the hidden layers to separate the noise contribution from the original image.

Using the same notation developed in Section 2.1 we can schematize Residual learning as follows: Focus on an image and its noisy approximation (y, X) where the noise model is assumed additive $X = y + \epsilon$. The exact form of ϵ may vary but for this example it may be thought as Gaussian with zero mean.

In discriminative denoising a neural network can learn to map the noisy example to the original by matching (finding the opportune set of weight w) its output $\hat{y}_{pred} = NN_{\mathbf{W}}(X)$ to the original image:

$$NN_{\mathbf{W}}(X) \sim y \tag{6.2}$$

The residual learning formulation instead aim to map the output of the network to the noise part of the input (i.e residual mapping). This is performed by subtracting the output to the noisy input

$$NN_{\mathbf{W}}(X) = X - \hat{y} \tag{6.3}$$

We can write the loss function, using a pixel-wise mean squared error for simplicity (i.e. each pixel is treated separately), of a denoising neural network with a fully convolutional architecture (as defined in Section 2.4.4) as:

$$L(\hat{y}, y) = \frac{1}{n} \sum_i (y_{i\,pred} - y_i)^2 \tag{6.4}$$

To simplify the notation we will not write explicitly the sum over the pixels in the next steps. In normal learning for which the output is $\hat{y} = NN_{\mathbf{W}}(X)$ the loss function is

$$L_{\mathbf{W}} = |\hat{y} - y|^2 \tag{6.5}$$

A good solution to minimize the loss is $\hat{y} = argmin(L) \rightarrow \hat{y} \sim y$

While for the residual learning approach $X - \hat{y} = NN_{\mathbf{W}}(X)$ the loss became

$$L(\hat{y}, y) = |X - \hat{y} - y|^2 \tag{6.6}$$

For the additive noise model, where we can write $X = y + \epsilon$, the loss reduce to

$$L(\tilde{y}, y) = |y + \epsilon - \hat{y} - y|^2 \tag{6.7}$$

That leads to $\hat{y} \sim \epsilon$. This slight modification in the loss function helps the neural network to find a solution that focuses on the noise part of the problem instead of learning features that depends on the image.

The authors of the original paper show that a simple neural network that applies this strategy can decrease training time, has a greater generalization ability (the training can be translated to related task), its training converge with a relatively small dataset and it is more stable and outperform many traditional algorithm in blind denoising, that is a denoising task where the noise level is unknown.

This is not surprising: Adding a residual layer means using prior information, in this case our knowledge of the additivity of the noise model, as an assumption to simplify the network task.

It is even less surprising that a residual learning approach will not excel in denoising tasks where the noise model is not additive since the effect of the residual learning will not remove the term that depends on the original image in equation 6.6.

From equation 6.1 is easy to see that the noise model for a MRI is not additive. To generate rician distributed values we can directly sum in quadrature two values sampled independently from a Gaussian distribution

$$x = \sqrt{(y + \epsilon_1)^2 + \epsilon_2^2} \rightarrow x^2 = y^2 + 2y\epsilon + R \tag{6.8}$$

Thus in quadrature, the noisy power image x^2 will have a non additive noise term $2y\epsilon$ and a additive noise term with non zero mean.

We tested how the performance of residual learning are affected when a DnCNN is trained with Rican distributed signal in presence of a high noise level for which the Gaussian approximation of equation 6.1 is invalid.

Since for power (and magnitude) images rewriting the loss function in a Residual Learning approach will not cancel out the terms relative to the image,

$$L(x^2 - \hat{y}, y^2) = \cancel{y^2} + 2y\epsilon_1 + R - \hat{y} - \cancel{y^2} \tag{6.9}$$

residual learning may not improve the denoising performance. Instead, it may impair the network ability to learn since the prior information given is misleading.

This test will be performed with a fully convolutional network architecture described in Section 2.4.4. This network consist only in convolutional operation, detector stages and batch normalization. Convolution operation is described in Section 2.4, the detector stage with ReLU activation in Section 2.1.1 and batch normalization in Section 2.4.5.

It is important to notice that this network has no subsamplig operation so its output units have a small receptive field, that means that only partial local information and not the entire input is used to compute (and train) the value of each output unit. In particular, for this architecture the effective patch size is 40 by 40 pixel. In Figure 2.6 we can see an example of how receptive field propagates trough layers in a FCNN architecture of this type.

The network is composed as follows:

- First Layer: 2D Convolutional with ReLU activation, 64 filter of size 3x3x1.

- Layer 2 to (D-1) : 2D Convolutional with Batch Normalization, ReLu activation, 64 filter of size 3x3x64.

- Output layer: 2D Convolutional with linear activation with a residual layer.

In figure 7.8 a schematic representation can be found.

The model was built in Tensorflow 2.x (Abadi et al., 2015) and the definition of the layers used can be found in the documentation. The model, code, trained weights and (hopefully) a working example can be found in my personal repository. The data used for training is the "BSDS500" dataset that is described in (Arbelaez et al., 2011) and it is often used as a benchmark for denoising tasks. We tested our performance on the "Set12" dataset (Fan et al., 2019). Both dataset can be freely downloaded. Images are converted to grayscale with a color depth of 8 bit before the noise addition.

The test will compare the performance of the denoiser in the task of denoising additive Gaussian noise to the performance obtained in denoising Rician distributed noise at the same image quality in terms of PSNR.

First, we generate noised data with additive Gaussian noise

$$M_{gauss} = I + \epsilon \tag{6.10}$$

where M_{gauss} is the resultant image affected by Gaussian noise, I is the original image and ϵ is a zero mean Gaussian noise with standard deviation σ.

Second, we generate Rician distributed noise to model an MRI following Eq. 6.8.

$$M_{rice} = \sqrt{(I + \epsilon_1)^2 + \epsilon_2^2} \tag{6.11}$$

where M_{rice} is the resultant image affected by Rician noise, I is the original image and ϵ_1, ϵ_2 are zero mean Gaussian noise with standard deviation σ. After the noise addition the images are again normalized to 8 bits.

Image are corrupted with an unknown level of noise (blind denoising task) with a standard deviation extracted randomly between $\sigma = [35, 60)$. To mach the sample size used in the original paper the model is trained on 400 images chosen randomly from the train dateset and validated on the remaining images.

The reported performance are calculated over the independent "Set12" test data. Each DnCNN is used to restore images affected only from the noise type it was trained for. Results for the denoising task on the Gaussian additive noise are reported in Figure 6.2 and 6.4a, while results for the denoising task on Rician noise in Figure 6.3 and 6.4b. The metrics used to evaluate the performance are defined in Section 6.3.1. We choose to compare our results to a standard image filtering technique. The application of wiener filter is common strategy adopted in image restoration task that is highly effective for white noise removal (Orieux, Giovannelli, and Rodet, 2010). In our case it is used as a baseline performance to justify the utilization of deep learning for the chosen task. The filter used can be found in image-restoration package of scikit-learn (Pedregosa et al., 2011).

Our results show that when we compare the restoration of images at the same level of degradation, expressed as PSNR with the original image, Gaussian denoising with residual learning is more effective than Rician denoising and, in particular, our DnCNN is always better than a wiener filter applied on the same image only in the case of Gaussian noise.

The performance shown at high noise level may seem low but is important to remember that in a 8 bit image, whose maximum intensity value is 255, a "bright scene" has a mean intensity around 125-200. This means that with a perturbation with $\sigma > 50$ the effective SNR is between 3 and 4.

For a natural image this means that the image is almost unrecognizable and it is often discarded in many applications, while for certain MRI acquisitions it may be an acceptable value.

In the next section we will propose and describe a new denoising methods based on the application of a denoiser directly on the frequency space of a MRI acquisition that will allow us to exploit the perks of residual learning even in magnetic resonance images with very low SNR where the hypothesis of Gaussian additive noise is invalid.
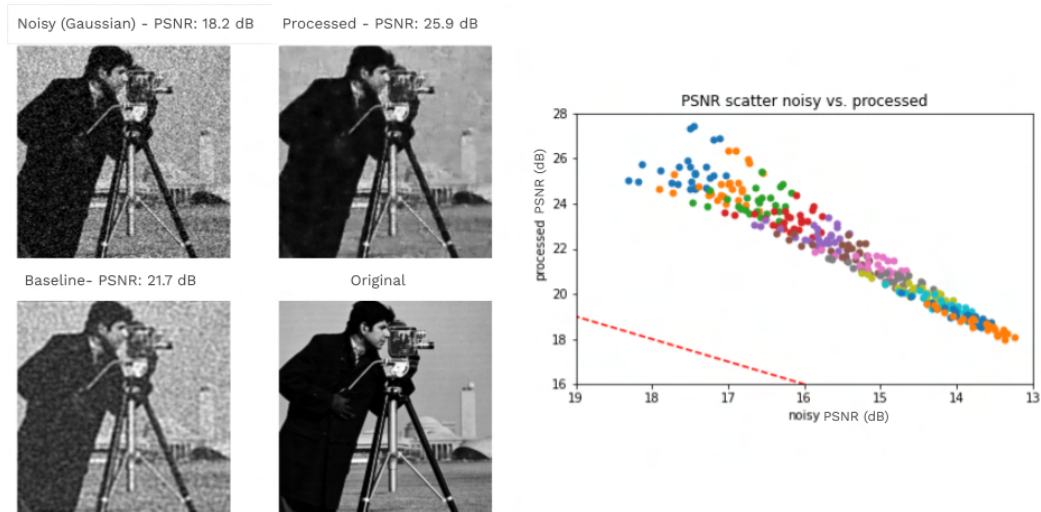
FIGURE 6.2: Example of the DnCNN on the Gaussian blind denoising task with $\sigma \in [35, 59]$.

Panel on the left: Clockwise, The nosy version of the image, the image processed by the denoiser, the original image, the image after the application of the wiener filter.

On the right: PSNR of the denoised images in function of the starting PSNR of the noisy images. Each point is an image in the test dataset, different colors represent different levels of noise applied. The dotted red line indicates that there is no improvement in PSNR after the application on the denoiser [PSNR(processed) = PSNR(noisy)].

6.3   Material and methods

6.3.1   Metrics

The quality, respect to noise in the acquisition, of magnitude images can be expressed as Signal-To-Noise-Ratio (SNR). For MRI this is often defined as the mean over the intensity of the pixels in a region of signal divided by the standard deviation calculated from a background region.

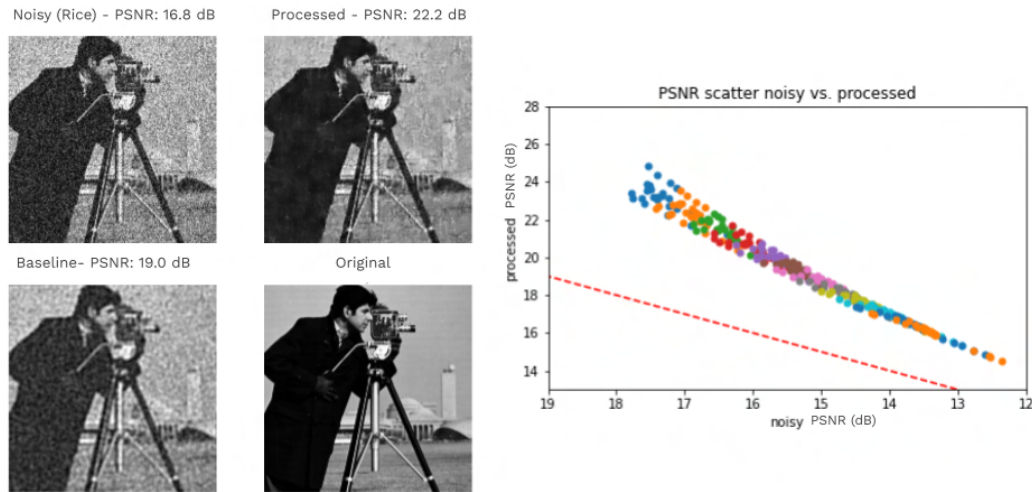$$SNR = \frac{\mu_{signal}}{\sigma_{background}} \tag{6.12}$$

FIGURE 6.3: Example of the DnCNN on the Rician blind denoising task with $\sigma \in [35, 59]$.
Panel on the left: Clockwise, The nosy version of the image, the image processed by the denoiser, the original image, the image after the application of the wiener filter.
On the right: PSNR of the denoised images in function of the starting PSNR of the noisy images. Each point is an image in the test dataset, different colors represent different levels of noise applied. The dotted red line indicates that there is no improvement in PSNR after the application on the denoiser [PSNR(processed) = PSNR(noisy)].

When the noise free image is available we will use the Peak Signal to Noise Ratio (PSNR) to quantify the quality of a magnitude image. PSNR is defined from the mean squared error (MSE) that, for a pair of NxM real images I and I^*, is written as:

$$MSE = \frac{1}{N \cdot M} \sum_{i=1}^{N} \sum_{j=1}^{M} [I(i,j) - I^*(i,j)]^2 \tag{6.13}$$

MSE is just the mean square euclidean distance in the pixel space. The PSNR is defined as:

$$PSNR = 20 \cdot log_{10} \left(\frac{MAX_I}{\sqrt{MSE}} \right) \tag{6.14}$$

and it is expressed in decibel (dB). It is the logarithm of the inverse of the euclidean distance between two images normalized to MAX_I, that is the maximum value that a pixel can assume (i.e. 255 for a 8bit image).

6.3.2 Phantom Generator and Simulation

In order to perform a supervised learning we need to provide the network with pairs of corrupted and noiseless images to be used as a ground truth. The ground truth will be use to provide an example of the expected output. We discussed supervised learning in more details in Section 2.2.

Unfortunately, dataset of clinical MRI images with pairs of high quality and low quality images are difficult to obtain. Even more difficult is to obtain the raw data
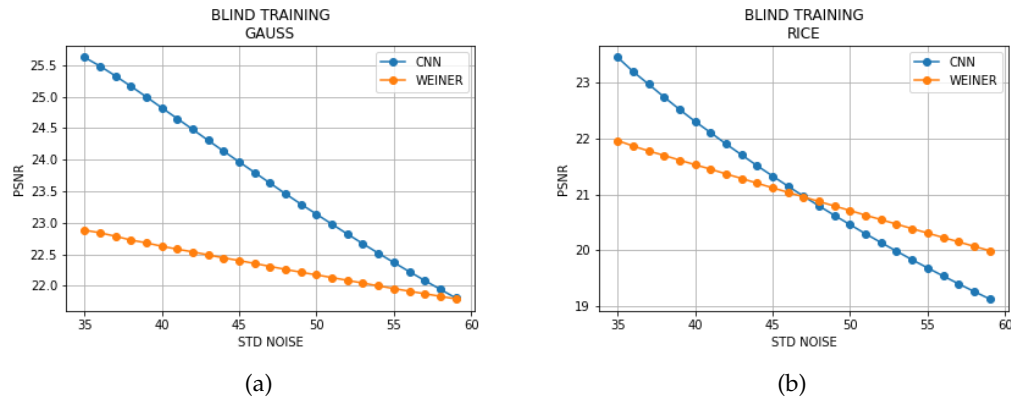
FIGURE 6.4: Average PSNR for the test dataset after the application of the DnCNN (Blue) and the wiener filter (Orange) in function of the standard deviation of the noise for the two noise models: (a) Gaussian noise (b) Rician noise
The DnCNN with residual learning is always better than a wiener filter for blind Gaussian denoising. In the case of Rician noise the baseline approach of a wiener filter has better performance in terms of PSNR at high noise levels. When the average SNR drops below 3 ($\sigma \sim 50$) the DnCNN is no longer a valid strategy for denoising images with Rician noise.

in the k-space that is usually discarded when the acquisition is completed and the image reconstructed.

To avoid this problem we decided to first validate our methods on simulated data. This means that our work should be considered a "proof of concept" and it may not be directly applicable to real data but we will be able to control every step of the pipeline.

To generate the dataset we performed MRI simulation in MRiLab (Liu et al., 2017) that is a comprehensive simulator for large-scale realistic MRI simulations. MRiLab combines realistic tissue modeling with numerical virtualization of an MRI system and scanning experiment to enable assessment of a broad range of MRI modalities.

Realistic simulation can be performed with biological plausible phantoms modelled as large 3D objects with biologically relevant tissue models. The computational power needed for simulation is gained using parallelized execution on GPU.

In addition to tissue and anatomical models, the realistic simulations require setting up a virtual MRI system and RF pulse sequence which conform to the existing physiological and technical limits of MRI scanning. MRiLab parameters specifying the scanning environment include the maps of main magnetic (B0) [to simulate artifacts derived from field dishomogeneity], transmit, and receive fields [to simulate artifacts derived from the coil geometry], and parameters of imaging gradients [to simulate artifacts from frequency shift]. Also resolution, number of acquisition average, noise level and common preprocessing steps can be selected.

To build our dataset we first prepared a phantom generator to increase the variability of our data. A phantom is a high resolution 3D model of biological plausible tissues with an anatomy that may resemble a clinical MRI.

Each phantom consist in a deformed cylindrical shape in which different element (that we will call "organs") may appear.
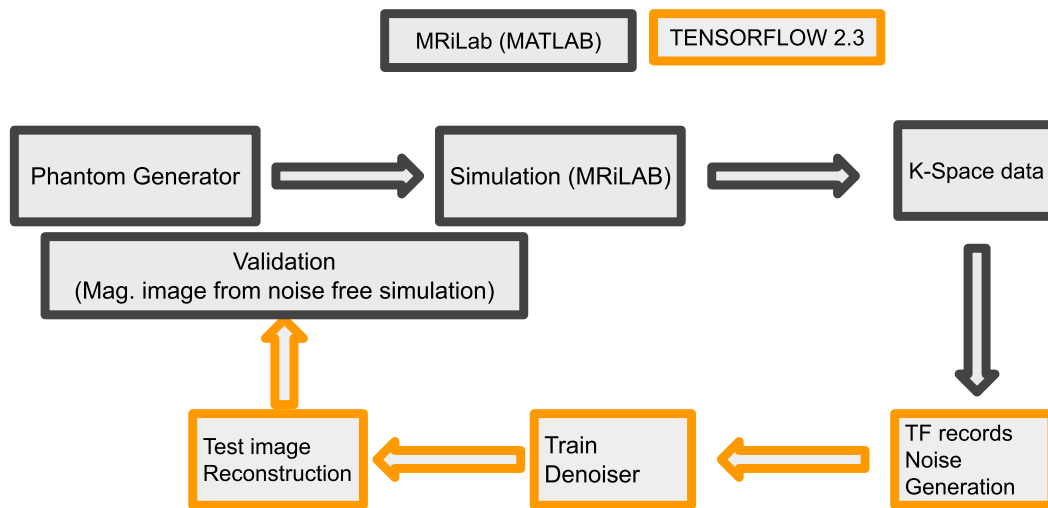
FIGURE 6.5: Schematic representation of the experimental pipeline.
1 - A phantom is randomly generated. 2 - A simulated magnetic
resonance image acquisition is performed in MRiLab with either a
Spin echo or Gradient echo sequence and a predefined set of acqui-
sition parameters. 3 - Before the reconstruction of the magnitude
image, data in the K-space are saved in a binary format compatible
with Python/Tensorflow. 4,5 - Once the train dataset is completed a
DnCNN is trained with pairs of original and corrupted images. Noise
power is chosen randomly (blind training). 6 - The DnCNN is applied
on noisy images, processed images are transformed in magnitude im-
ages and tested against the simulation noise free results. Steps in grey
rectangles are performed in Matlab 2018b. Steps in orange rectangles
are performed in Python/Tensorflow.

Shape, position, rotation and dimensions of organs can vary in a predefined in-
terval, an example can be seen in Figure 6.6. Each organ is provided with the phan-
tom properties that are needed in order to simulate the behavior of tissues in a mag-
netic resonance image acquisition as described in the user guide of the simulation
software (MRiLab, V 1.3, Section 3):

- Gyro (rad/s/T) : The gyromagnetic ratio of the spin

- Rho : A matrix for describing spin density

- T1 (s) : A matrix for describing T1 relaxation time

- T2 (s) : A matrix for describing T2 relaxation time

- T2Star (s) : A matrix for describing T2* relaxation time

- MassDen (kg/m3) : A matrixfor describing tissue mass density

- ECon (S/m) : A matrix for describing tissue electrical conductivity

This properties are sampled from a predefined interval with credible values for
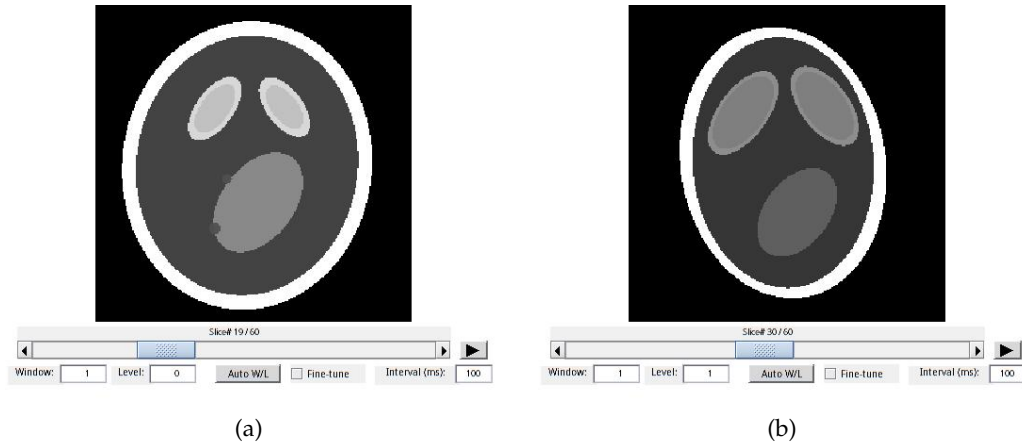a biological tissue.

(a)   (b)

FIGURE 6.6: Examples of phantoms generated. Each phantom is 3D
model with "organs" of different size, shape, orientation and posi-
tion. Physical quantities of the "organs" can also vary randomly in
a predetermined interval. (a) Example of a phantom that is used for
training. Random size sphere (Cysts) of different intensities and re-
laxation times can appear randomly inside the body. (b) Example of
a phantom that is used for testing in relaxation experiments. "Cysts"
can not appear.

In figure 6.5 is shown the schematic pipeline on the analysis. After the phantom
is generated a MRI exam is simulated with a spin echo 2D pulse sequence. A spin
echo is a simple acquisition modality where two RF pulses are used to generate the
signal. The peak of this signal is found at the echo time (TE). Many other pulse
sequence can be used in the simulation but we will limit to spin echo for simplicity.
This choice should impact on the generality of our results but in future more complex
acquisitions will be simulated.

Plane resolution is fixed at 128x128 pixel while slice thickness and the number
of slices in each exam can vary. After the simulation, raw data in k-space are saved
before the magnitude images are reconstructed. Afterwards, also the corresponding
magnitude images are saved.

Each saved example is a simulated MRI 2D image in the k-space composed by a
two channel matrix (128x128x2) that store the real and imaginary part of the signal
in k-space and the resulting magnitude image of size 128x128. Figure 6.9 shows an
example of the simulated data on a brain phantom.

To increase the variability of the dataset and to test if effectively the residual learn-
ing approach limits the formation of representations derived from the images in or-
der to concentrate on the noise form, some of the phantoms used in training will also
have signals sphere shaped, that we will call Cysts, of different random intensities
and relaxation times that can appear in random positions inside the phantom area.

In the examples used in testing and validation this additional variability is never
present. An example in shown n Figure 6.6 a, cysts are visible in the in the left part
of the lower organ. In simulation cysts appear as hyper-intense or ipo-intense spots.
These formations are confounding signal only and we will use them to test if these
details are correctly never proposed when the network is tested.
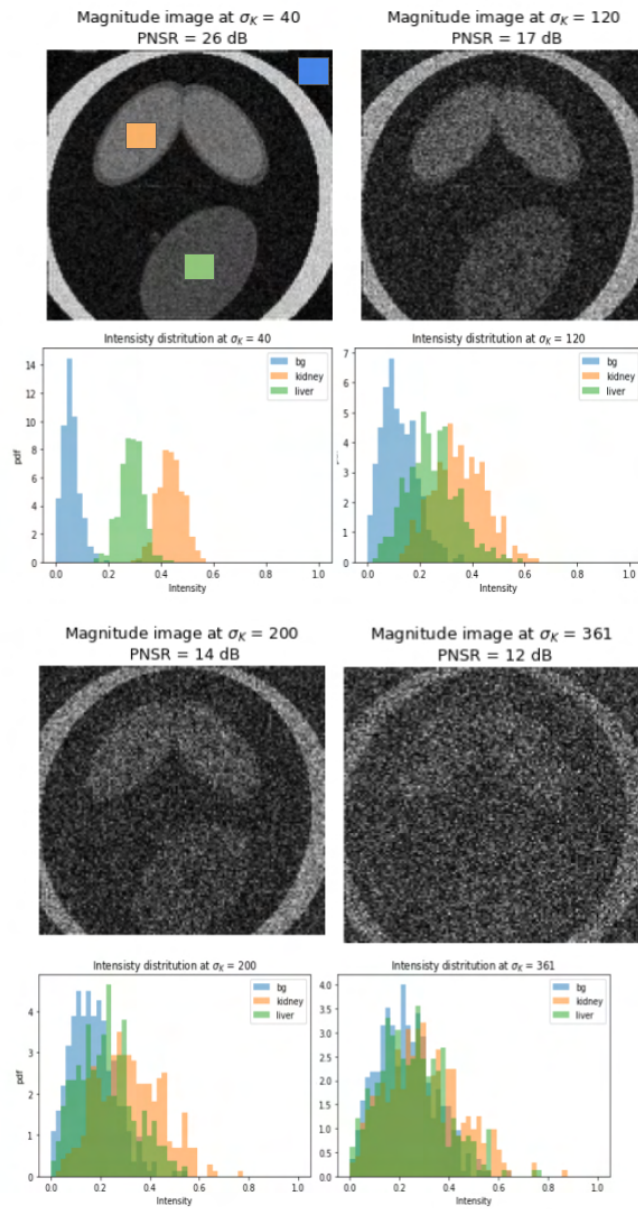
FIGURE 6.7: Noise Effect on simulated MRI. Magnitude image reconstructed after the addition of noise. Examples at $\sigma_K = 40$, 120, 200, 361. Resulting PSNR respect to the noise free simulation are 26 dB, 17 dB, 14 dB, 12 dB
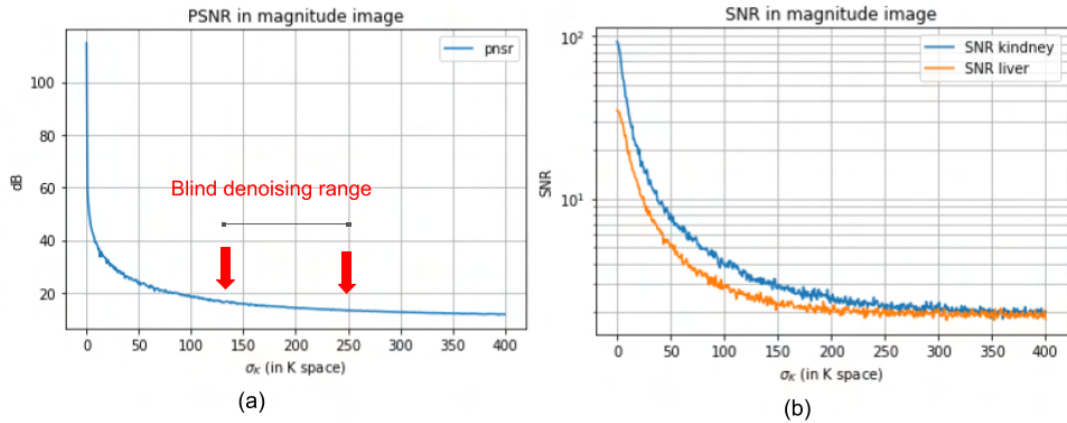
FIGURE 6.8: Noise Effect applied in k-space on magnitude images.
(a) Average PSNR of the noisy magnitude image respect to the noise
free magnitude image. PSNR in function of σ_k the standard deviation
of the noise applied in k-space. Red arrows indicate the range for
blind denoising training (b) Average SNR of two homogeneous signal
zone in the image in function of the standard deviation of the noise
applied in k-space. "liver" and "kidney" refer respectively to the single
"organ" that appears in the lower part of the phantom and to the pair
of "organs" that appears in the upper part.

6.3.3 Training and noise generation

After the noise free dataset is formed, noised data is generated by the addition of
complex white noise in the frequency domain. For a single coil acquisition, white
noise is a reasonable model (Aja-Fernández et al., 2016) under the assumptions that
the noise affects equally all the sampled frequencies and it is sample and source
independent.

The noise standard deviation in k-space σ_k is chosen to generate images with a
SNR between 5 and 2 when measured on the corresponding magnitude images. Fig-
ure 6.7 shows the effect of the addition of the noise in the k-space on the magnitude
images and Figure 6.8 shows the range of σ_k selected.

We trained a DnCNN for the task of denoising directly the k-space. We will refer
to this network as Kspace-Dn. Training is performed on pairs of clear and noisy
images minimizing the loss defined in Section 6.3.4. Also, to increase our statistic
we performed data augmentation on the k-space data.

Data augmentation in K-space is defined as a transformation that maintain the
same reconstructed magnitude image or produce a simple geometrical effect on the
magnitude image (i.e. rotation) Thus. given the k-space signal $S(x,y) = S_R(x,y) +
S_I(x,y)$ we will apply randomly the following transformations:

- Reverse axis:

    - $S(x,y) \rightarrow S(-x,y)$
    - $S(x,y) \rightarrow S(x,-y)$
    - $S(x,y) \rightarrow S(-x,-y)$

- Transpose axis:

  - $S(x, y) \rightarrow S(y, x)$ [Rotation]

- Phase multiplication:

  - $S(x, y) \rightarrow S(x, y) * \exp\{(i\theta)\}$ with $\theta \in [0, 2\pi]$

To compare our results to a network with the same complexity (architecture and number of parameters) we also trained a DnCNN on the noisy magnitude images reconstructed from the noisy k-space. We will refer to this network as the M-Dn. The training is performed in parallel to make sure that the two network can access the same dataset statistics.

The network architecture is the same defined in Section 6.2.1 with a two channel input and output for the Kspace-Dn that are needed because k-data are complex number. Each channel represent the real and the imaginary part of the signal and they are treated as colors in our image. We can not work directly with complex number because activation function and convolutional operation are not defined in the complex space.

Both networks have a deepness of 20 layers thus their effective receptive field is 41x41 pixels. The receptive field of a fully convolutional network is discussed in Section 2.4.4.

Thus, the Dn-CNN are effectively denoising areas of the image based only on partial context. This helps to reduce the number of parameters of the model in order to limit overfitting even when the network is trained on a small dataset and it will also contribute to avoid the learning of complex anatomical structure since only a patch on the input is seen at each output unit.

Both the networks are trained for 300 epochs with Adam (Kingma and Ba, 2017) as optimization and a learning rate $l_R = 10^{-3}$. Afterwards, the learning rate is reduced to $l_R = 10^{-4}$ and the the network is trained for another 100 epochs.

We will train our models on 10^3 images generates from the simulation on the phantoms defined in Section 6.3.2 to evaluate our performance with a dataset of the size that is commonly available with clinical MRI.

During training the images will be augmented with the transformations described earlier in this Section, so the actual instances of the signal examples will be higher than the simulated scans.

Geometrical transformation will augment the dataset size by a factor 4 and forcing the phase invariance, the random multiplication for an unitary complex number that loosely model the coil orientation, will also increase the number of independent samples.

Also, the corrupted images are generated on the fly so at every epoch the noise is independently sampled from its distribution before the addition. This training strategy allows to perform more training steps while limiting the overfitting risk since the exact pair of original and corrupted example will be never shown more than once.

In any case, for our simple simulation, where the only source of corruption is the acquisition noise, the size of the dataset should be appropriate to learn the denoising task.

Probably the limiting factor in the size of our dataset is our phantom generation algorithm: even if a large sample is generated, its effective variability may be low since the generator itself depends on a small number of parameters that can not mimic the great variability observed in real biological tissues.
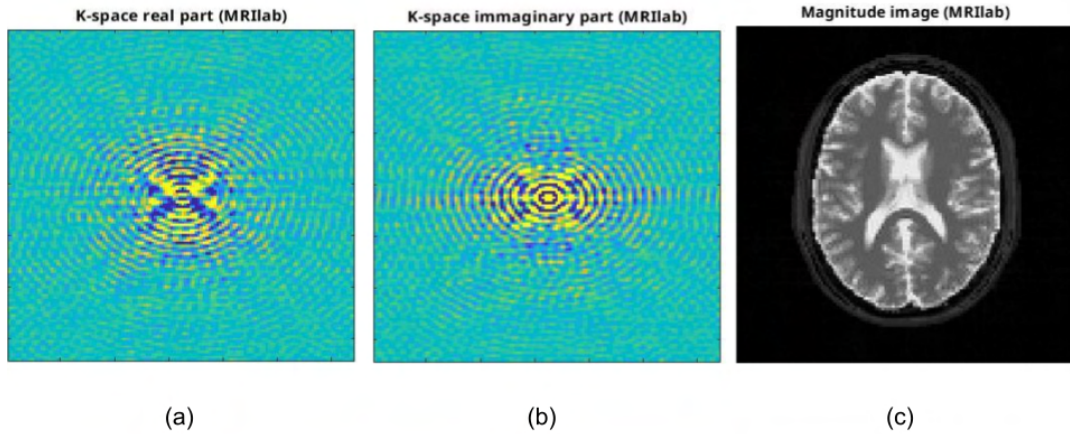
FIGURE 6.9: Example data obtained from a simulation that is used as the ground truth during training. An acquisition in k-space from a 2D spin echo pulse sequence: (a) Real part of the signal (S_R) (b) Imaginary part of the signal (S_I) (c) Magnitude image (M)

6.3.4 Loss function in K-space

The network that processes data in k-space will try to map a noisy version of an input to its ground truth. The ground truth is the clean version of the k-space so the minimization of the loss should reproduce it but since our metrics will be based only on the reconstructed version of the image we can add a term in the loss function to help the network to concentrate also on the final image.

The input of the network is a two channels image S representing the real (S_R) and imaginary (S_I) parts of the signal in the k-space: $S = (S_R, S_I)$.

The magnitude image M is reconstructed by taking the modulus of the 2D Inverse Discrete Fourier Transform (iDFT) applied over the signal in k-space. Lets define this procedure as:

$$M := reco(S) \tag{6.15}$$

The loss used for in k-space denoiser is composed by two term:

$$L_k = MSE(S_y, S) + \beta * MSE(reco(S_y), M) \tag{6.16}$$

Where S_y is the two channel output of the network. S is ground truth signal in the k-space, M is the ground truth magnitude image.

The first term is just the mean square error for the two channels signal in k-space and the second term is the mean square error for the magnitude image reconstructed on the fly from the predicted k-space.

The parameter β is used to balance the two terms. If beta is kept small ($\beta = 10^{-2}$) it has no impact on the performance of the network (See Figure 6.11) but it helps to avoid artifact in the magnitude image.

The loss function used to train the denoiser on the magnitude images is a standard MSE loss function.

6.3.5 Testing data

Residual learning has the great benefit of generalize to different types of additive models but as we have seen in Section 6.2.1 it won't perform well at very low SNR on magnitude images. We will test if working on frequency data improves our performances in the denoising task since in this case we will have Gaussian noise instead of rician.

To test the networks we will first use data from a realistic brain phantom, available in the simulation software (Liu et al., 2017), to check the noise removal capabilities on fine details that are never seen during training. This test will be a measure of the generalization capability of the networks.

Secondly, we will simulate a T2 measure experiment in which a series of acquisitions is performed with RF pulses at increasing echo times (TE). We will use those images to measure transverse relaxation. This magnetic resonance signal is expected to follow an exponential decay with increasing echo time (TE) as we saw in Section 3.1.1.

$$S(TE) = S_0 \exp\left\{ \left(-\frac{TE}{T2}\right) \right\} \tag{6.17}$$

The time constant is T2 relaxation time that affects the coherent components of M perpendicular to B_0.

The nonzero mean distribution of noise in magnitude images causes the signal to not decay exponentially with increasing TE in low signal-to-noise-ratio (SNR) images.

Thus, calculating T2 from low-SNR images by fitting is prone to overestimation of T2. This is a well know effect ((Raya et al., 2010)) and often prevents the use of low SNR images for quantitative purposes.

We will then estimate the T2 from the denoised images to check if the Dn-CNN restore the expected behavior since the diagnostic value of T2 critically depends on how exactly T2 can be measured.

6.4 Results

6.4.1 Results on Realistic Phantom

In this test three networks were trained as described in Section 6.3.3 with pairs of original and noisy simulated MRIs of the simple phantom described in Section 6.3.2: a network with a reconstruction term with $\beta = 10^{-2}$ as defined in Section 6.3.4 will be trained on data in k-space (Kspace-Dn), a network with $\beta = 0$ also trained on frequency data (Ks no RECO) and a network trained on the same dataset on the resulting magnitude images (M-Dn). From Figure 6.11 we can see how the networks that works on the k-space always outperform the network trained with magnitude images in terms of image quality. Especially at high noise levels, the improvement in PSNR from the noisy images is always significantly better for Kspace-Dn than for M-Dn.

This was expected from our early experiences reported in Section 6.2.1 that were based only on a different noise generation model and not from a simulation:

For a Dn-CNN with residual learning is inherently more difficult to work with MRI magnitude images so it should be avoided.

In Figure 6.10 a few compared images denoised with Ks-Dn and M-Dn are shown for visual inspection. It seems that M-Dn tends to favor smoother surfaces at the

expense of losing details. This may be caused from an incorrect noise estimate in the blind denoising task.

Ks no RECO is discussed in Section 6.4.3.

6.4.2   Relaxometry experiments

The aim of this experiment was to verify the ability to remove the bias in the T2 estimate from noisy magnitude images. These images are acquired at growing TE and as expected from Eq. 6.17 the signal will decreases exponentially, so at fixed noise standard deviation, the SNR of the image will also decrease exponentially.

We tested K-Dn and M-Dn (defined in Section 6.4.1) to the same task. It is important to remember that these networks were not directly trained to denoise a set of images thus images are processed independently: The denoiser is able to perform blind denoising so it should be able to estimate the effect of the noise and remove it regardless.

Since we are not interested in reproducing correctly the shapes of the objects but only in reproducing the average intensities in an area, for the purpose of this experiment we used $\beta = 0$ in the loss function (Equation 6.16). The beta parameter controls how much the reconstruction in magnitude space is weighed in the loss function.

Test data were simulated at increasing TE with a fixed noise level of $\sigma_k = 30, 100$

T2 estimate is derived from a fit with an exponential function on the average intensity of a patch of 15x15 pixels of the "organ" with no additional preprocessing.

Results from this test are shown in Figure 6.12. At high noise level M-Dn overestimate, as expected, the real T2 and K-Dn gives a better estimate.

A probable explanation is that the most prominent effect of rician noise at low SNR is the non-zero term that increase the average intensity over the background. This effect depends on both the noise variance and the original intensity so it is difficult to remove.

Estimating and correcting this effect on magnitude images is more difficult respect to perform the same task with gaussian noise in the k-space. The M-Dn fails to remove the Rician bias giving smooth surface but at the wrong intensity while K-Dn remove the effect of the non zero mean noise at the expense of higher variance in the image.

This effect can be seen in Figure 6.13 where the intensity at high echo time is plotted for the original image and the denoised ones in frequency and magnitude.

Qualitatively the M-Dn better reconstructs the smooth surface of the original image but fails to reproduce the correct intensity on average. In morphological images this behaviour may not be a problem but for quantitative MRI it is more important to correctly measure the average signal in an area respect to obtain a sharp contrast.

It would be interesting to see how the performance compare to analytical methods such as(Raya et al., 2010) and (Koay and Basser, 2006) that explicitly use the information about noise distribution in the estimation of the real intensity.

From this experiment we can also see that a M-dn may not correct the relative pixel intensities in a T2 weighted image distorting the resulting contrast.

6.4.3   Artifact in Reconstruction

One of the main advantage with working with residual learning is that it should help avoiding the creation of artifacts in the denoised image. A we said in Section 6.2.1, during learning with residual learning the noisy image is subtracted to the
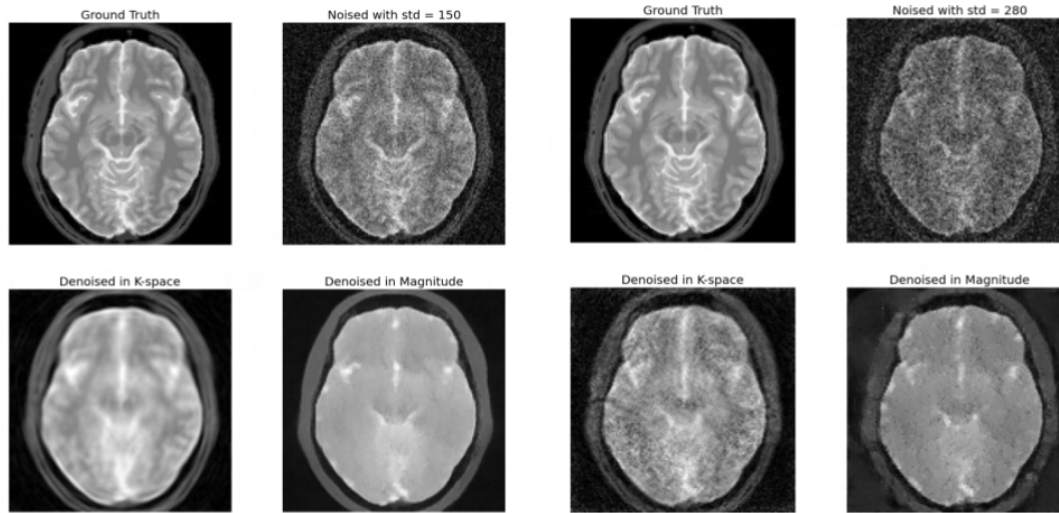
(a)
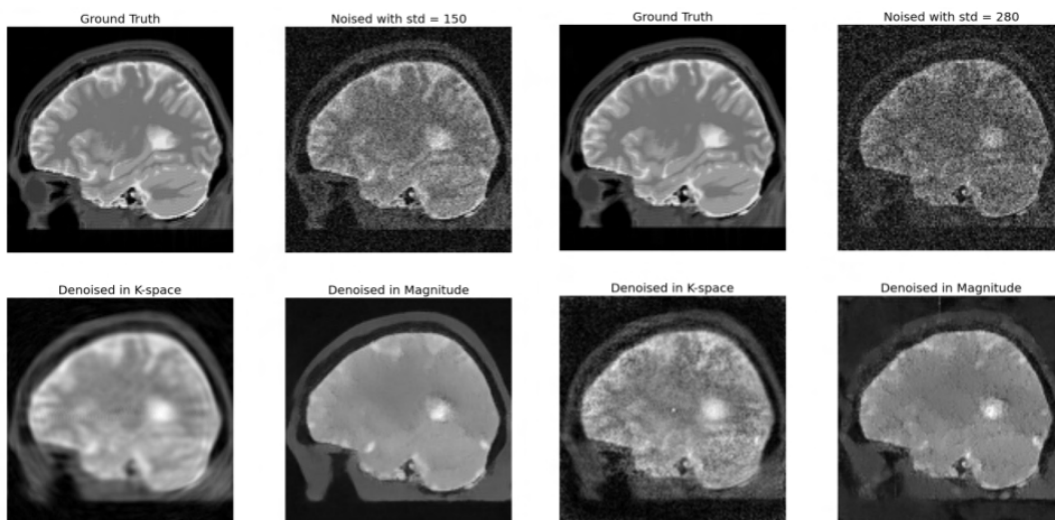
(b)

FIGURE 6.10: (a) Axial and (b) Coronal view of a the brain phantom. Denoising in magnitude space tends to create smoother surfaces at the cost of losing important details.

FIGURE 6.11: Performance of the denoisers on the complex brain blind denoising task for four value of noise level in the k-space $\s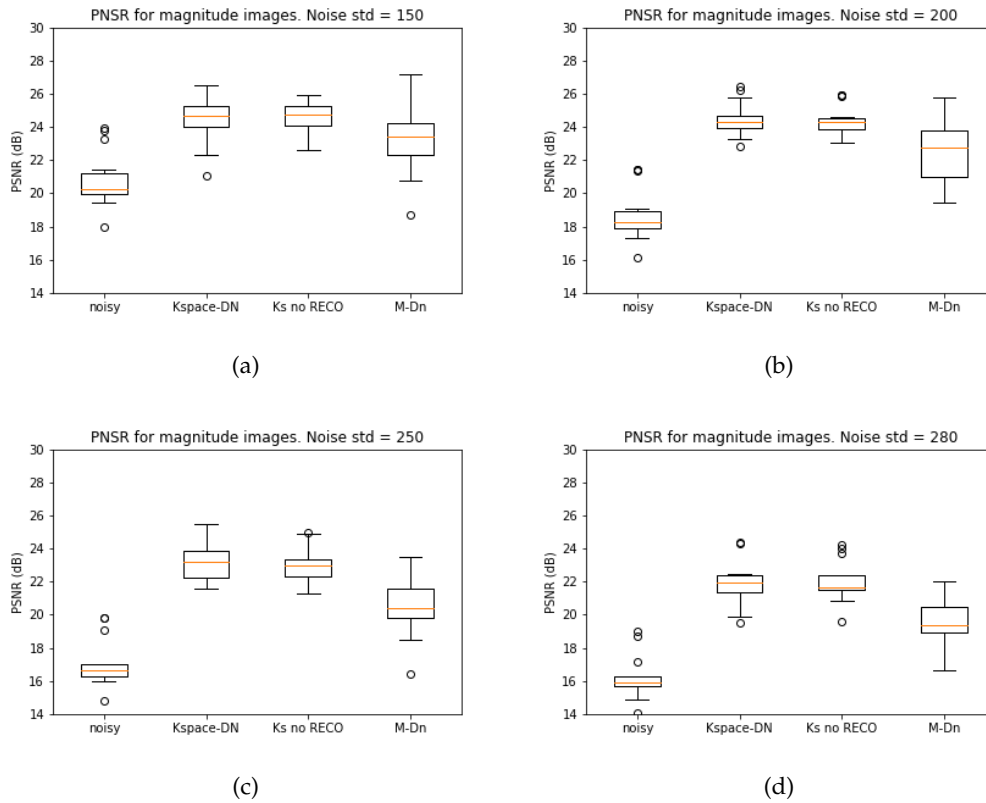igma_k$. The box-plots show: PSNR of the noisy images before the application of the denoiser (noisy), PSNR of the images denoised in k-space with the complete loss function (Kspace-Dn), PSNR of the images denoised in k-space with $\beta = 0$ (Ks no reco, no reconstruction term in the loss function), SNR of the images denoised as magnitude images (M-Dn) (a) $\sigma_k = 150$ (b) $\sigma_k = 200$ (c) $\sigma_k = 250$ (d) $\sigma_k = 280$ Performance of Kspace-Dn is always better than the M-Dn, especially at high noise levels. The presence of a reconstruction term do not impairs results in the denoising task. Orange lines are the median of the distribution, boxes are the 25 and 75 percentiles, whiskers are the 5 and 95 percentiles. Outliers are indicated with a dot.
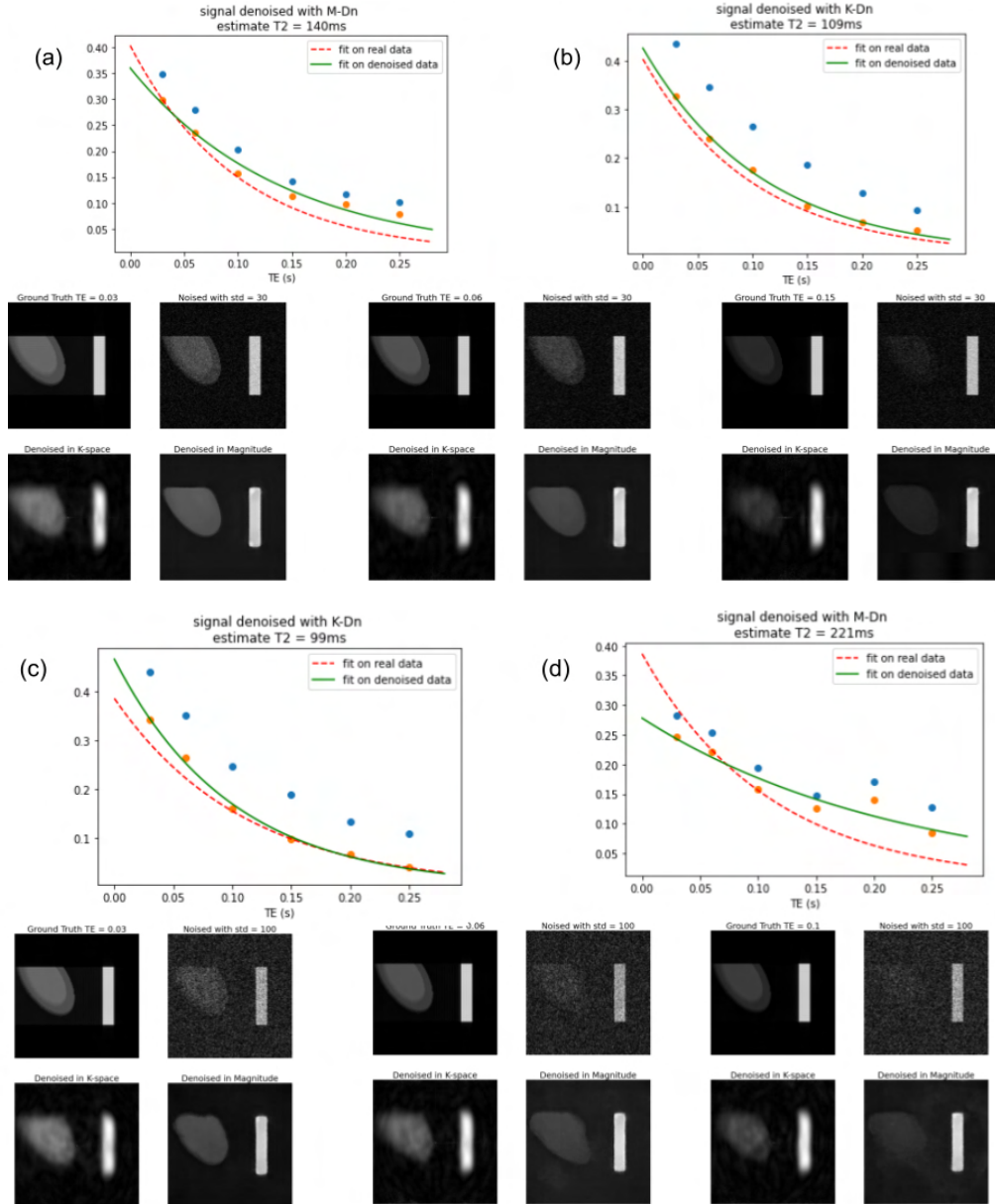
FIGURE 6.12: Noise effect on T2 estimate. "kidney" organ with "cortex" (Blu) and "Medulla" (Orange) average intensity at increasing TE. Dotted red line: real relaxation with T2 = 120 ms. Green line: Fit on denoised data. (a) $\sigma_k = 30$ denoised in magnitude. Estimate $T2_M = 140ms$ (b) $\sigma_k = 30$ denoised in k-space. Estimate $T2_k = 109ms$. (c) $\sigma_k = 100$ denoised in k-space. Estimate $T2_k = 99ms$. (d) $\sigma_k = 100$ denoised in magnitude. Estimate $T2_M = 221ms$.

network input to force the algorithm to reproduce the residual image instead of the clean one.

One way to test this aspect is to feed the network with only noise to verify if random activations of the weights produce in output only noise or structures that are related to what the network had experienced during the training phase.

During training, cyst like objects were introduced in images as a confounding signal (they are described in Section 6.3.2) , since they are random in intensity and position, random fluctuations can be confused with them.
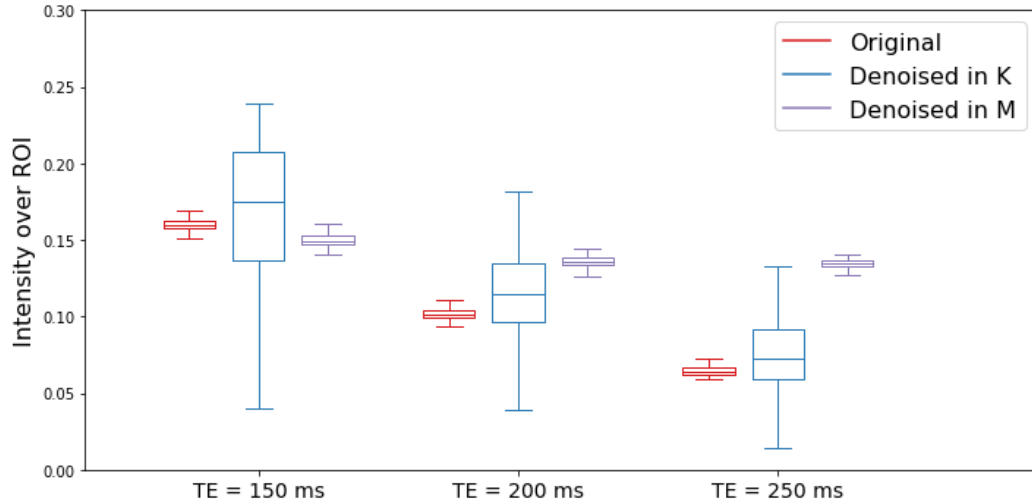
FIGURE 6.13: Intensity in images at high echo time in the ROI for the original, noise free image (red), the image denoised with K-Dn (blue) and the image denoised with M-Dn (purple). K-Dn denoising remove the effect that impairs the correct estimation of the relaxation time since on average it reproduces a value closer to the original. M-Dn instead reproduces the smooth surface of the original image but at the wrong intensity level. Horizontal lines are the median of the distribution, boxes are the 25 and 75 percentiles, whiskers are the 5 and 95 percentiles.

We trained two networks with the same dataset with RL and without it modifying the loss function as explained in Eq. 6.6 and we tested the output when a image with $SNR << 1$ is provided as input.

We tested 100 random images, in 77 of them structures resembling the cysts in shape and intensity were present when normal learning was used, while networks trained with RL never shown any recognizable form. An example of the output of the network when random noise is provided in input is shown in Figure 6.14.

Even in residual learning artifact are possible but they are probably derived from the incorrect processing of the k-space with patch operation such as convolution that may introduce border effect that are more visible in the reconstructed image since an error of a few pixels in frequency space may introduce large effects in the corresponding magnitude space.

Our early tests were made with denoisers trained with $\beta = 0$ (as the Ks no RECO network shown in Section 6.4.1) that showed better results in terms of PSNR respect to M-Dn but in some occasion artifact was present in the reconstructed images. Since we are interested only in the final magnitude image (that has clinical relevance) we decide to add the reconstruction term to the loss function in equation 6.16.

If this term is small there is no loss of performances and artifacts are greatly reduced. We are still developing a methodology to quantify the production of this type of artifacts for a more rigorous analysis.

From Figure 6.15 we can see that there are artifacts that we think are derived from a few pixels errors in k-space. In the example a dotted horizontal line may be generated from a single pixel at low frequency in a coordinate (the line) and at high frequency in the other coordinate. It is likely a boundary artifact that are common in CNN but it has a greater effect on the resulting magnitude image. Unfortunately,
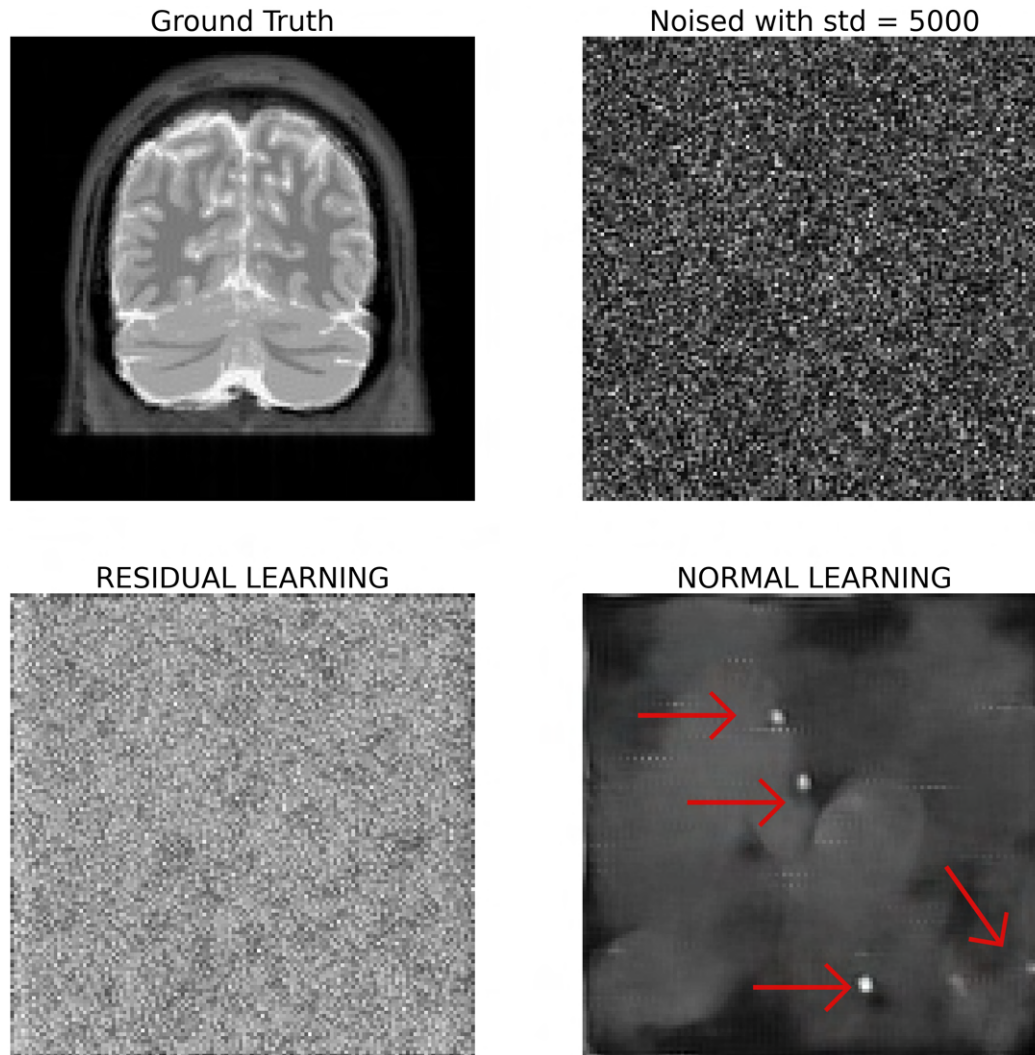
FIGURE 6.14: Effect of the denoiser trained with and without residual learning on an image with $SNR << 1$. Normal learning seems to had learn the structure of the cysts that are included in training as a confounding signal and it proposes these structure when there is no signal. This behavior is absent when the denoiser is trained with residual learning. Red arrows signal the positions of the artifacts

single pixel errors are difficult to avoid but our mitigation strategy may ease the problem. Still, further investigations are needed.

6.5 Discussion

In this Chapter we introduced the problem of denoising a magnitude MRI with a Deep CNN approach.

We implemented two solutions: A novel approach based on the raw frequency data and a more traditional one that works on the already reconstructed image. The advantage of working in the k-space is that the noise model is simpler since it can be rightfully modelled as additive Gaussian while images in magnitude have a more complex signal distributed with a Rician distribution.

FIGURE 6.15: Example of an artifact probably caused by a few mis-
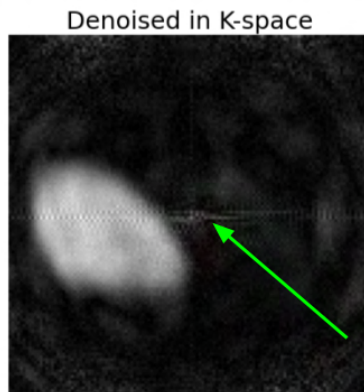placed pixels in k-space. A dotted line has high frequency in a direc-
tion and low frequency in the other. It may have been generated by a
pixel with one high coordinate and one low coordinate in k-space.

As expected, the difference in performance of the two approaches start to be
relevant at very high level of noise that is usually not a problem in clinical morpho-
logical MRI but it is a major problem in other acquisition modalities such as DWI
and acquisitions on etereonucleai (19F or 13C, for example).

Another perk of working with frequency data is that the residual learning for-
mulation is exact since only in this case the noise is additive. Residual learning is
a powerful framework that helps the neural network forming representation of the
noise model instead of the underling clean image, allowing the development of a
denoising strategy with far less parameters respect to normal learning at the same
level of performance.

Residual learning is also less prone to include artifacts in the denoised image
since the internal representations of the network should contain less information
about the images structures. Our preliminary results suggest that when a network
trained with normal learning is exposed to a random input it may produce details
that are absent but that are present in the training set. A network trained with resid-
ual learning does not exhibit this behavior in the same conditions.

Also, since at low SNR the limit in the quantification of area intensities to extract
quantitative NMR parameter such as the relaxation times or the spin density, is given
by the Rician bias our K-space based method might be the most suitable choice since
it seems to better preserve the average signal intensity over the ROI, at the expense
of visual clarity, respect to the denoising in magnitude space.

It is worth notice here that our method doesn't resolve the problem, common
to all standard denoising algorithms, that any unwanted effect in MRI, other than
fast Rician noise, that most of the times are the real problem in imaging such as any
movement artifact or field inhomogenities will be treated as signal and maintained
(or amplified) by our method. This is also a problem with standard denoising algo-
rithm, as a matter of fact these problems are usually treated by separated algorithms
in all other denoising solutions.

Removing the "fast" noise of acquisition should be thought of as the first step of a general MRI enhancement pipeline since it will help all the other steps eliminating the acquisition noise that often limits the applicability of other correction algorithms.

Regarding the poor visual presentation of our denoised images, we need to say that in the case of the tested phantoms where we were interest only in the conservation of the integral of the signal in an area, so we do not enforce any reconstruction metric.

We don't believe it is a problem in our specific case but we are aware that in morphological images the visual inspection is a fundamental part of their examination so we should aim to obtain better results in shapes and border sharpness.

This can be accomplished by tuning the loss function to weight more the reconstruction term. The application to a morphological dataset already has better results in this context as we will show in the next Chapter.

Furthermore, the availability of frequency data is scarce at best since they are routinely discarded after the acquisition is completed. Thus, if we want to collect them we will have to make a specific deal with the operators to save them. Open dataset are, for now, also very limited with the notable exception of the FastMRI dataset (Zbontar et al., 2019).

Also, while MRI images are saved in the DICOM standard, so they are easily opened and processed almost independently from the scanner manufacturer and software, there is no analogue for the raw data that often is saved only in binary format and every manufacturer store it at different level of preprocessing.

Since we worked on simulated data we can't produce general results valid on real data but our work is useful to lay the foundation for future applications of denoiser neural networks to magnetic resonance frequency data. In particular, the presented tests do not aim to be definitive but we will like to propose this two parameter as essential in evaluating the role of denoising in MRI: First, the ability to generalize to different datasets and acquisitions modality and secondly the ability to restore the physical quantities that can be extracted from images.

We believe that, in a general application of denoising with CNN, processing the raw data in the k-space should be preferred and additional test with real data and real noise and artifacts should be made.

In particular it is important to validate any denoising approach on other source of noise that are present in clinical application of MRI such as movement artifact, chemical shift artifact and hardware/acquisition related ones.

Currently we are applying our methods on real data acquired in k-space and a more realistic model of noise. Also, a more realistic simulation to pre-train the model and limit the scarceness of data, that is often a problem in clinical imaging, is an interesting development that we are examining.

In the near future we would like to try this denoising methods in situation where low SNR is inevitable such as fluorine magnetic resonance. As we can see from Chapter 5 the fluorine concentration reached in organs in a real application is very low. This will produce images with very low SNR and where noise will be a dominant source of signal. It will be interesting if a deep learning approach may work in this situation.

6.5.1 Limits of performance evaluation

Whenever a filter or a processing algorithm is applied to medical images there is an inherent risk of loosing important information or worse, of creating a confounding effect.

This risk is critical in the situation in which the automated operation that manipulate or modify the medical image is performed before it is inspected by the operator.

In this context, we can safely say that denoising is more dangerous than automated image analysis like computer aided diagnosis (CAD) or automated segmentation, since it is more difficult for the operator to exercise control or to refute the computer proposal.

Low level denoising algorithm are already implemented in most commercial image acquisition scanners but they mostly consist in traditional methods, usually based on a statistical model of the noise, that has been extensively validated and, in some situation, they may perform poorly but is difficult for them to create artifacts that are not clearly recognizable by an expert operator.

When a learning algorithm is used there is another intrinsic risk that can have highly unwanted effect for medical imaging, that is the risk of information transfer between patients.

The biological variability in anatomical structures is typically very high but medical images are in general very stereotyped since the general anatomy is the same in all the subjects.

In this case overfitting, where we use the term to indicate the possibility to learn the structure of the single example instead of the general rule of the task and not to indicate the generic effect of diminishing return in longer training, can hypothetically transfer information gained from an example to another.

Research in image restoration usually care less of this particular effect since it is more interested in the visual effect of the restored image. Often the border between denoising (i.e removing the effect of a noise source) and image enhancing (i.e improve images visual effect) is thin in natural image denoising: One of possible example is the denoiser variational autoencoder (Im et al., 2016) that implements a very narrow latent space description between the downscale path of the encoder and upscale path of the decoder. This kind of architecture may introduce unwanted details in the final image since it performs an extreme compression and decompression of the input.

Because of these reason there is the necessity to define a performance metric to evaluate the clinical applicability of deep learning based MRI denoising that take into account both the quality (visual and quantitative) and the risk is details modification.

Our proposed denoising method, although for now only applied to simulated data, increases both SNR, PSNR and helps to retain the real integral count of the signal for noised images.

Still, even if the performance based of these metrics will be confirmed of real data, that would not be enough to ensure the safety of the application in a clinical setting.

This is a general problem, to our knowledge even when image super resolution, that is the task to create higher resolution images from lower ones, there is no explicit quantitative metrics to measure the possibility of this phenomena. See for example Chapter 10 of this collection (Lu et al., 2017) for a survey of super resolution in cardiac MRI.

Our proposed method tries to indirectly limit this possibility:

The use of residual learning and its application at raw signal level where its exact formulation is valid can help avoiding learning the images structures.

Additionally, the small receptive field are effectively a "pin hole" so retaining coherence for large anatomical structure is difficult. This is the opposite approach that is used for example in super resolution on denoising autoencoders (Im et al., 2016)

that usually have large downsampling branches in their architectures and are often implemented since their performance can be superior to the simple CNN architecture.

Also, the relatively simple architecture with only 0.5 M parameters is smaller than the standard U-net with 30 M parameters (Ronneberger, Fischer, and Brox, 2015) that we can think of as the unofficial fully convolutional network benchmark. The relative small size has benefits in computational and memory requirement, it needs less data to be trained and it is less prone to overfitting.

Still, this is not definitive and more testing should be done and we should strive for a quantitative measure of the risk in the application of any deep learning method to medical imaging.

The most important requirement for any application to medical images should be to ensure that the action of the denoiser does not produce effects that can lead to an incorrect interpretation of the image by the clinician. This scenario would be at the same time a technical failure and a risk of legal repercussions: It must be avoided but the first step is to quantify it, if it is present.

The most robust method of testing it would be to prepare a set of original, noise free, diagnostic images and a set on the same images with noise applied and then removed with the algorithm, then to test if the application of the denoiser influences the results of a human operator.

This proposal is difficult to implement in practice since it would require a large dataset and a lot of works by skilled personnel. Also, it would not be general and it would have to be repeated on a case-by-case basis.

An easier and more realistic alternative can be performed with an already validated CAD system on a dataset with established good performance. The experiment would be to retest the same data after noise application and denoising to see if the results of the analysis are influenced by the denoising operation.

For example, brain segmentation can be performed automatically reasonably well and it is impaired by high level of noise so it may be a good candidate for this test.

This scenario would also be important to develop a denoiser that removes the effect of the noise and not the noise itself.

6.5.2 Limits for the application of our method on real data

The next step of our work is the validation of this method on real data acquired in frequency.

We will have to face two main problems: First, as already mentioned we will need to acquire the data. The first validation of our method will be done on morphological acquisitions that are already available but afterwards we might want to apply our solution to imaging acquisition of flurinated tracers. As we measured in our ex-vivo studies (See Chapter 5) the time needed for an acquisition in our experimental condition may be several hours for a spectroscopy experiment that are far more sensitive than imaging since they measure all the spin in a volume many order of magnitude larger than a pixel. So acquiring a large dataset is impossible.

Probably with a complete revisitation of our protocol, in the best possible conditions, we might be able to reach a scan time of 10 to 20 minutes, that is also the limit for an acquisition on an anesthetized mouse, that will still limits our dataset to a few hundred maximum examples.

A possible solution will be to adopt a transfer learning strategy, training first on a larger (for example simulated) dataset and then fine tune on the smaller dataset. If

this strategy works, it may solve our particular problem and the underlying question about the reasonableness of the application of deep learning in this case:

The time spent in acquiring a small dataset, that will always be a compromise with respect to the real data needed to perform the training , is worth the probably low performance CNN that we will implement? Probably not. Unless we also provide a strategy to generalize the training so that even an experiment with a low data throughput may eventually benefit from it.

Second, especially for the fluorine case, we may have problems in the definition of a ground truth since a noiseless scan is impossible in practice to obtain. With morphological imaging we can have an almost perfect acquisition in reasonable time while with very low fluorine concentration our dataset may be just the best possible measure that can be acquired with specialized hardware and it may not be useful as a ground truth.

In this case our only hope will be to work with unsupervised learning that have the advantage of not needing a ground truth. A possible approach may be the so called Noise2Noise framework (Lehtinen et al., 2018) that, if the condition of applicability are met, allows to restore images training with only low signal samples.

Finally, real MRIs have many other defects introduced by the acquisition or reconstruction protocol that may increase the variability of our data to a value too large to be effectively sampled in our dataset.

In any case this field of study is still open and many possible solutions that already works for natural image processing may be successfully implemented in our case study.

Chapter 7

Preliminary application of K-DnCNN to the FastMRI dataset

Chapter Abstract

In this Chapter, our proposed denoising neural network is applied to a large dataset of parallel multicoil acquisitions. We performed the training with both the supervised learning and, exploiting a recently proposed technique (Noise2Noise), with the unsupervised learning. The possibility to adopt the Noise2Noise framework is directly derived from the choice of working with raw frequency data instead of reconstructed images. Both strategies produce interesting results and they are comparable in performance. At the end, we discuss how to improve our results and how to step up from the development phase of the method to the optimization phase.

7.1 Introduction

In this chapter we are going to apply the deep learning based denoiser that we had introduced in Chapter 3 to the task of denoising parallel multi coils acquisitions.

Our proposed method is based on the denoising of the raw data of an MRI in the k-space instead of directly trying to denoise the reconstructed image.

The neural network was originally test on simulated data of single coils acquisition at very low SNR to test the possibility to remove the rician bias and restore the quantitative relation between pixel intensities and physical quantities.

Now, we will show how with little modification it can be easily applied to a different imaging modalities that is instead used for accelerating morphological imaging.

The possibility of quickly transferring the use of the Dn-CNN network to this type of data is given by our idea of working in k-space because, independently from the acquisition modality, the noise model is simpler in the frequency domain . We will see how also in this case this property will be useful in the task of denoising.

This should not surprise since we exploit a fundamental relation in MRI acquisition that is common to many different modalities.

The availability of raw data in k-space is scarce and in particular we don't have access to 19F-MRI that are our final target, so to test the performance of our method we decided to apply it on the denoising of the FastMRI dataset (Zbontar et al., 2019) that contains raw k-data in the order of 10^4 examples.

The FastMRI dataset is not originally conceived as a denoising task: It is used to test reconstruction algorithms in parallel acquisitions with subsampling of the k-space, that means that not every point in the frequency space is acquired to speed up the acquisition process.

Still, it is a perfect candidate for our test because it consists of fully sampled, high definition, k-space data of images with generally high SNR to which we can add realistic simulated noise to train the Dn-CNN. Furthermore, this dataset provides examples from two different anatomical district: The brain and the knee. This will allows us to test if our network can maintain the same level of performance if it is tested on data that comes from a different body part acquired with different NMR sequences.

Together with the supervised learning strategy that we implemented in the previous chapter we will also perform an unsupervised training with this dataset.

In the denoising task, training with an unsupervised approach consists in developing a strategy that does not need the noise free ground truth image.

We will implement the Noise2Noise framework that substitutes the ground truth images with multiple examples of pairs of the same images affected with different sampling of the noise, thus in our case multiple low signal acquisition of the same subject.

In MRI this strategy can be extremely useful since it is often easier and less time consuming to perform multiple scans at low SNR instead of a single high quality acquisition.

We will also discuss how the fact that this strategy of unsupervised learning works without much modification is a direct consequence of working in the k-space and how this can reinforce the motivation behind our proposed method.

7.1.1 The Noise2Noise framework

In general, the denoising task with deep learning is performed by constructing a map between a corrupted version of an image and the corresponding clean one. This map is usually learned directly by matching pairs of noisy images and clean ones in a supervised training approach.

When the clean image is not available it is usually more difficult to train a neural network with acceptable performance for the task.

One possible strategy to overcome this limit is to use an unsupervised learning approach: The Noise2Noise framework was recently proposed by Lehtinen et al. (Lehtinen et al., 2018) and it offers the possibility to train a neural network using only multiple corrupted examples of the same target without explicitly using the target itself.

Their method allows them to exploit the general purpose deep CNN model to unsupervised denoising and to reach significant performance, very close to supervised learning, but without the problem of collecting a ground truth.

The idea is based on the fact that, if we think of the loss function L as a generalized point estimator, so an operator that involves the use of sample data to calculate a single value which is to serve as a best guess of an unknown distribution parameter, then we can write the expected value over distribution of the pair of noisy / clean image (X, y)

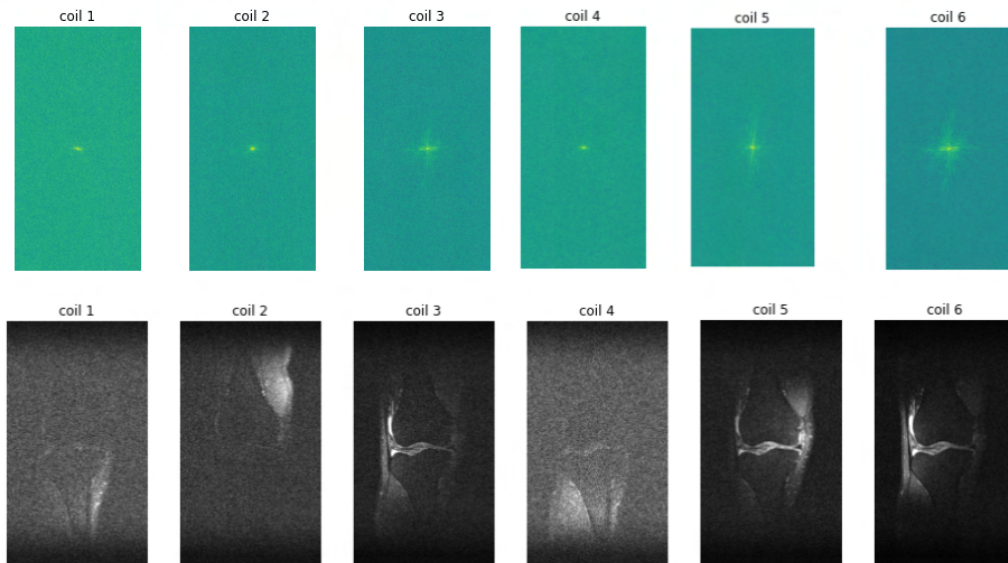$$\underset{\mathbf{W}}{argmin}\, E_{(X,y)}\{L(NN_{\mathbf{W}}(X), y)\} \tag{7.1}$$

FIGURE 7.1: Example of the multiple coils signal in k-space for a element of the FastMRI dataset. Each exam is composed by 15 acquisition from parallel coils, the first 6 are shown.
Top row: modulus of the complex signal in k-space. Bottom row: Magnitude images reconstructed from the k-space signal for each coil. Notice how each coil capture the whole volume with spatial dependent sensitivity. The noise variance in each coil can be assumed constant so the images that seems more noisy have a lower signal. The strength of the signal is higher if the source is close to the coil.

To which we know a possible solution is the set of **W** that gives

$$NN_{\mathbf{W}}(X) = E_{(y)}\{y\} \tag{7.2}$$

We can explicitly write the conditional distribution over all the possible clean images and the average over the sample of the noisy images.

$$\underset{\mathbf{W}}{argmin}\, E_{(X)}\{E_{(y|X)}\{L(NN_{\mathbf{W}}(X), y)\}\} \tag{7.3}$$

In this case $P(y|X)$ means all the possible correct explanations that we can match to a noisy sample. To visualize it lets take for example all the possible positions of an edge when the borders are noisy or the exact position of a boundary between two noisy surfaces.

We can notice that if we replace the $P(y|X)$ with another distribution with the same expectation value the relation 7.2 is still valid independently on the exact distribution that we are using. In particular, the parameters **W** found that solve the task also remain unchanged. So we may reuse them to perform also the standard task.

If we replace the target with another noisy observation that has the same expectation value (for example we add additive white noise) then the solution still hold and for example in the case of the L_2 loss we can train using target corrupted samples if their noise has zero mean since the L_2 distance is the corrected point estimator for noisy observation with additive noise with zero mean. In the loss function both the input and the target $\tilde{y}_i$ are drawn from a corrupted distribution conditioned to the (possibly unknown) clear target y_i:

$$\underset{\mathbf{W}}{argmin} \sum_i L(NN_{\mathbf{W}}(X_i), \tilde{y}_i) \tag{7.4}$$

Now, until $E\{\tilde{y}_i|X_i\} = y_i$ we don't need to know the explicit distribution of the original images nor of the noise but having data distributed accordingly to them is enough.

We will see that this relation also holds in the residual learning approach and if we are working in the k-space where the noise model can be correctly represented as additive we can apply this method to multicoil MRI. Doing so on the already reconstructed images wouldn't be that straightforward.

We will compare the performance to the supervised learning with the same number of examples in the blind denoising task to see if convergence and training time are comparable since denoising performance is the more important but not the only criteria to select a ML method.

The Noise2Noise approach can greatly improve the possibility of application in a real task setting because in MRI is usually faster and more practical to perform multiple scans of the same subject at low quality instead of making an almost noiseless acquisition. The main reasons is the presence of other problems such as movement that impairs long scans and the fact also that, in MRI, SNR goes with the root of acquisition time so acquiring multiple low quality scans may also be time conserving.

7.2   Methods

7.2.1   Dataset description

The FastMRI dataset (Zbontar et al., 2019) contains fully sampled k-space acquisition of knees and brains consisting of raw multicoil k-space data of unprocessed complex valued NMR measurement.

This dataset is thought of as a benchmark for algorithms that perform image reconstruction of frequency subsampled data and it is not originally used for denoising tasks. Still, we believe that its characteristics make it perfect for our task:

The set of fully sampled images will be the ground truth of our training, they represent an acquisition modality of higher quality than what is commonly used in clinical practice so they are also a perfect, almost noise free, example that can be used to train the denoiser. To be precise, some of the acquisitions can be noisy, as we can see from Figure 7.1 , but they are a minority and usually the resulting image is still of great quality.

In any case, the noise already present in the acquisition will be negligible with respect to the artificial noise that we will add for training purposes and its presence is natural and expected in real data so any denoising methods should be robust to labels that are not perfect.

Given that the original task is reconstruction in presence of frequency subsampling, the dataset contains also subsampled data to be used in the task for which it is originally intended, but it is not useful in our case so we will discard it.

The dataset is formed by examples of multicoil acquisitions of the knee and the brain. Initially we will train our neural networks on the knee data and we will test its performance on both the brain and the knee scans.

Multi coil raw data from 1594 knee scans acquired for diagnostic purposes are present in the dataset. There are multiple scanners both at field strength 3T (Siemens Magnetom Skyram, Prisma, and Biograph) and at 1.5 T (Siemens Magneton Aera), all of them employ a 15 coils array.

The acquisition is performed with a turbo spin echo sequence with a pixel resolution of $0.5mm$ and a thickness of $3mm$. The parameters used between the scanner are homogeneous but noticeably the scans were performed with the fat suppression tecniques (See for example Figure 7.3) or without (See for example Figure 7.4). This setup results in visually different resulting images.

The total number of patients used for training and testing are shown in Table 7.1.

The dataset also provides 6970 fully sampled brains scans. A portion of 255 of them was used for additional testing of our solution but we did not perform any training on it. A future version of our denoiser will be trained on the whole dataset. This dataset provide examples from multiple sequence and acquisition modalities. Both T1 and T2 weighted images are present and there are also contrast medium enhanced acquisitions.

In Figure 7.1 an example of a multicoil knee scan is shown with both the raw data and the corresponding reconstruction. The top row shows the modulus of the complex signal in k-space and the bottom row the reconstructed image. The final image is obtained from the combination of the images captured by each coil trough the root sum of their square intensities as will be shown in equation 7.5.

TABLE 7.1: Description of the FastMRI dataset. In the left column is reported the total size of the dataset. In the Right column is reported the number of volumes (patients) used to train, validate, and test the model. The results shown for the supervised and unsupervised training are based on the same data split.

	FastMRI Patients		Patients	Slices
FastMRI Train Set	973			
		Used in Train :	350	10236
		Used in Test :	100	2959
FastMRI Val. Set	197			
		Used in Validation :	80	2380

7.2.2 Data preprocessing

One of the main perks of our method is that it can be applied with minimal data preprocessing. Still, manly to reduce the computational load during training, we performed the following operations: We applied a low-pass filter to the frequency data reducing the acquisition size from $640x372$ or $640x368$ to $320x184$. This has two main motivations. First, the k-space is sampled at a very high frequency that in a noisy acquisition that we want to simulate wouldn't be acquired in the first place since the high frequency component encodes both the details and both the vast majority of the noise. This provides us with cleaner ground truth images and secondarily we obtain a smaller example reducing the memory needed for training.

Visually, this filtering produces only minor effects on the final image.

Then, the complex k-space signal of each coil is normalized to the maximum value of the modulus on a slice by slice basis.

This step allows the standardization of the examples and helps convergence during training. However, it introduces a bias since not every coil has the same sensitivity in each point of the volume so a more natural solution would be to normalize by the maximum value over the whole volume. The effect on image reconstruction in this case is a minor difference in contrast in slices over the periphery of the scanned volume where coils may have a large difference in sensitivity. In future tests we will implement the normalization over the volume.

Finally, we do not use all the coils but we selected only 8 of 15 with a fixed choice equal for all the examples. This step diminishes the overall quality of the final image since less data is used in reconstruction but it is needed to perform the training in a more reasonable time . We will avoid this step in future versions when the development phase of this work will end.

All these preprocessing steps should not influence the overall performance metrics since they are made before the ground truth reconstructed image is produced so both the tested and ground truth images are processed the same way.

Working with Mixed Precision

To speed up the training during the development phase of this study we performed the learning with the Tensorflow Mixed Precision[1] framework that helps the model run faster and use less memory.

Mixed precision used both 16bit and 32bit variables during training. Lower precision data types in the model weights uses less memory and can exploit the presence of specialized hardware in GPU to perform faster operation: Modern accelerators can run operations faster in the 16bit data types, as they have specialized hardware to run 16bit computations and 16-bit data types can be read from memory faster.

The most important benefit is that we can double the size of the mini batch at the same memory cost and doing so double the rate of examples processed at each training step.

Implementing this technique gives us a three fold decrease in training time for an epoch with little to none decrease in performance.

This package works better only if hardware Tensor Cores are used that can multiply float16 matrices very quickly. However, Tensor Cores requires certain dimensions of tensors to be a multiple of 8.

For this reason we perform the crop over the data matrix and we select only a part of the available coil channels.

Also the network parameters that are used in Tensor Cores are chosen to be multiple of eighth as shown for example in Figure 7.8.

7.2.3   Noise and signal model for multiple correlated coils

In parallel MRI the image is reconstructed from multiple k-space, each acquired from a separate coil. Each coil captures the MR signal from the whole volume but it is

---

[1]https://www.tensorflow.org/guide/mixed_precision

typically more sensitive in one region. Each extra coil produces an additional measurement with a different sensitivity pattern that are afterward recombined together.

The final image is obtained from the root of the sum of square of the single magnitude images reconstructed independently from each coil. This technique is called the Sum of Square (SoS) reconstruction and it is one of the most commonly implemented is clinical practice (Roemer et al., 1990)

$$M_T^2 = \sum_{l=1}^{L} |M_l(\vec{x})|^2 \tag{7.5}$$

In a multiple coils system the additive Gaussian noise model is still valid at the coil level but the presence of multiple coils, since there is always mutual inductance between them, introduces a strong correlation in the additive noise term.

In general, the complex signal s_l in a system of L coils can be modelled as follows:

$$s_l(\vec{k}) = a_l(\vec{k}) + n_l(\vec{k}; 0, \sigma_{K_l}(\vec{k})) \tag{7.6}$$

Where l is the index that goes over the coils, a_l in the noise free signal generated by the RF pulse and n_l is an additive Gaussian noise term with zero mean and standard deviation $\sigma_{K_l}(\vec{k})$ that in general depends on the coil and on the frequency.

In the approximation of stationary noise, variance does not depends on $\vec{k}$ and is equal on all the k-space. This approximation is realistic in most modern scanners if the amplifier is well tuned. Thus we can drop the dependency from the point in k-space in the noise model.

$$s_l(\vec{k}) = a_l(\vec{k}) + n_l(\vec{k}; 0, \sigma_{K_l}) \tag{7.7}$$

Since all the coils are spatially close there will be a strong correlation in the noise between them. In particular geometrically closer coils will be more correlated. Our multivariate Gaussian will have a non diagonal covariance matrix where we will indicate the covariace between the i, j coil as $\rho_{i,j}$.

$$C = \sigma_{k0} \begin{pmatrix} 1 & \rho_{12} & ... & \rho_{1L} \\ \rho_{21} & 1 & ... & \rho_{2L} \\ ... & ... & ... & ... \\ \rho_{L1} & \rho_{L2} & ... & 1 \end{pmatrix} \tag{7.8}$$

Usually $\rho_{i,j}$ is significant in multi coils system but their value is defined by how the antenna is built and it will be a specific characteristic on the particular model of MRI scanner.

We don't know the exact value of the correlation matrix for the scanners that are present in the dataset, although we may be able to estimate it from the few background scans that are present in it, so we will propose a dummy coils geometry and a realistic correlation matrix associated.

We choose a circular geometry where correlation between coils is $\rho_{i,j} = 0.3$ if the coils are first neighbours and $\rho_{i,j} = 0.15$ if they are second neighbours. In all other cases $\rho_{i,j} = 0.05$. In Figure 7.2 there is representation of the chosen geometry and a the correlation matrix is reported. As we already mentioned our system has originally 15 coils that we subsampled to 8 coils. In the dataset there are 3 different scanners but in our approximation they will have the same correlation matrix.

The correlation of noise on the coils is particularly important in the magnitude space. When the signal is high a Gaussian approximation of the signal area is possible and the effect of correlation are left aside as minor visual effects. However,

the effect of the correlation strongly influences the underling statistical model of the noise and signal and, at low SNR, it makes difficult to adopt many statistical based processing methods.

For the SoS reconstruction scheme that means that the effective noise value in signal areas and background will be different so the noise in the background will no longer be representative of the noise in the signal areas. This makes more difficult the estimation of the noise. In figure 7.7 we can see how an additive noise model in k-space as the one described in Eq. 7.7 can generate complex residual (i.e quadratic difference between the original and noisy image) patterns from the simple and spatially uncorrelated noise model in Eq. 7.7 .

It is clear that we can see the anatomy in the residual map and it is a direct effect of the SoS reconstruction.

The reason is easy to see. If we recall Eq. 6.1, each image reconstructed from a coil will have a rician distributed signal: the distribution of the square sum of rician distributed signal can be approximated as a non central chi-squared distribution.

Thus we can say that the non linear process of image reconstruction modifies the noise introducing complex non linear terms that makes more difficult the denoising respect to the k-space.
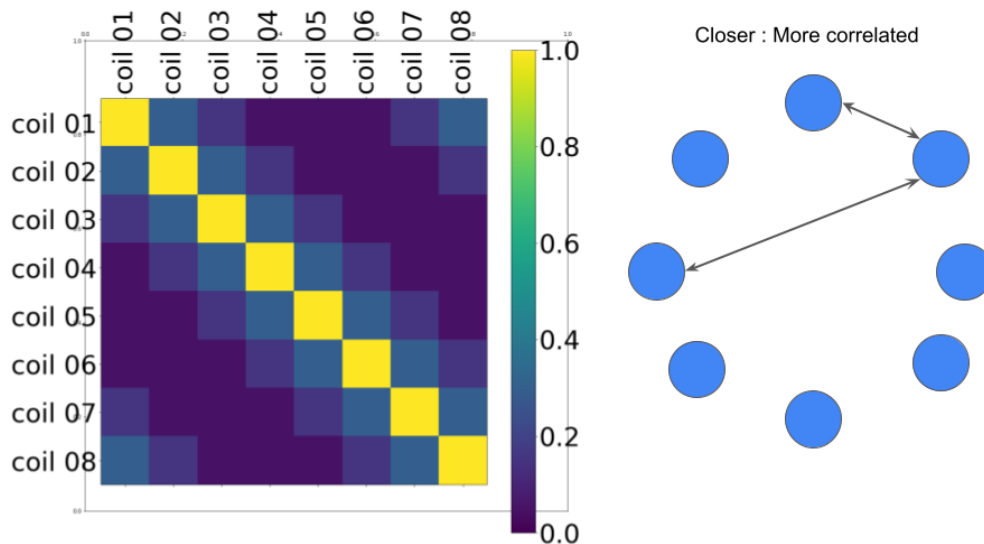

FIGURE 7.2: Noise correlation matrix and proposed dummy coils geometry. The values of correlation can be considered realistic with the approximation of equal variance in the noise for each coil. The level of correlation between two coils depends on their distance. Closer coils are more correlated.

7.2.4   Metrics

The denoising performance of the FastMRI dataset will be evaluated with the following metrics:

The first one is the PSNR that was introduced in Section 6.3.1 and the second one is the structural similarity index that is the average value calculated over the image where the structural similarity measure (SSIM) was applied with a sliding windows.

The SSIM between the original image patch m and the patch $\hat{m}$ of the denoised image is defined as

$$SSIM(m,\hat{m}) = \frac{(2\mu_{\hat{m}}\mu_m + c_1)(2\sigma_{\hat{m}m} + c_2)}{(\mu_{\hat{m}}^2 + \mu_m^2 + c_1)(\sigma_{\hat{m}}^2 + \sigma_m^2 + c_2)} \tag{7.9}$$

where μ is the intensity mean of the patch, σ is the standard deviation, $\sigma_{\hat{m}m}$ is the covariance and c_1, c_2 are small regularizing constant that we set to 0.01 and 0.03. Images are normalized between 0 and 1 before the calculation. The filter size is set to 11x11 pixels.

When computed over a sliding windows SSIM produces maps that are used to check if the interdependency of nearby pixels are restored after the denoising.

Also, residual maps will be used. They are computed as the square differences between pixels intensity of the original and restored image so each value of the residual map $D_{i,j}$ is

$$D_{i,j} = (I_{i,j} - \hat{I}_{i,j})^2 \tag{7.10}$$

Both I and $\hat{I}$ are normalized between 0 and 1.

We also reports the average of the residual map over the image. This metric is related to the PSNR (See Eq. 6.13) that is proportional to the logarithm of its inverse.

Both metrics are computed only on the center part of the image to avoid that many background pixels are considered in the metrics since they are easy to treat and they would give a bias in the score in images with large zones of background.

Examples of these maps are shown in Figures 7.11, 7.10, 7.14, 7.13, 7.17

Notice how these two metrics are complementary since SSI is computed over a small region and gives information about the relation between close pixels while the residual maps depends only on a difference between single pixel.

It is important to understand that the reconstruction term of the loss function is directly correlated to the euclidean distance of the residual map so the network is explicitly trained to minimize this value while the SSI is an independent measure. This aspect is important to verify that the training helps restoring the image also at the structural level.

7.2.5 Training

The model used in this test is the same described in Chapter 6. It is a fully convolutional neural network with 18 convolutional layers regularized with batch normalization, without any downsampling path and with a residual learning connection at the output layer implemented as in Eq. 6.6.

This network has a receptive field, as defined in Equation 2.12, of 41 x 41 pixels, this it works with local context only.

The network input is formed by the real and imaginary part of the complex k-space data for each coil. Thus, it is an image with 16 channels (i.e 2 for each of the eighth coils). The diagram of the network used is show in Figure 7.8.

We will now describe the two proposed learning strategies: the supervised learning that we already implemented with the simulated data in Section 6.3.3 and the new approach the unsupervised Noise2Noise. Both these training strategies utilize the same data for training and testing as described in Table 7.1.

A subset of 350 patients with 10^4 slices was selected at random for training from the official train data split. We do not utilize the whole dataset to reduce the training time as these results are preliminary and we still are in the optimization phase. When the correct strategy will be derived we will make use of the whole dataset and we will reports any gain in performance.

In any case, 350 patients and 10^4 slices is a reasonable size for a clinical dataset that may also be reached in an hospital setting.

We trained the network for a blind denoising task varying the standard deviation of the noise between $\sigma \in [5, 20] \cdot 10^{-3}$ in the k-space. Since the starting quality of the original images is not homogeneous and in the dataset are present both scan acquired with and without fat suppression that generate brighter or darker images, it is difficult to measure the effect of this noise on the whole dataset in terms of SNR.

Instead, we will use PSNR to define the degradation of noised images with respect to the original ones and we will use the gain in PNSR as a measure of improvements.

In general, this random noise addition with this range of standard deviation will generate corrupted images with PSNR in a range between 10 to 35 that means their visual quality is distributed from a reasonable to deeply damaged.

The network will then learn to estimate the noise content and remove it. In multi coils scan with strong correlation between noise in the coils this task is particularly difficult at the image level because the noise is not stationary and will depend on the intensity of the image.

We will now discuss the supervised and unsupervised training and in the following section we will report the results obtained.

The trained network obtained with these procedures is still not optimized so this is the description of the experiment conducted and they may not be the best possible set of hyper-parameters and training strategy.

Supervised learning

The supervised learning approach needs ground truth images and the learning process is used to find the best set of weights that minimize the loss function.

The loss function used is a modification of the one presented for the denoising of the single coil scan in Equation 6.16 that introduces the reconstruction with the Sos (Eq. 7.5) for the final image.

$$L_k = MSE(S_y, S) + \beta * MSE(Sos(S_y), M) \qquad (7.11)$$

Where S_y is the 16 channels output of the network. S is ground truth signal in the k-space, M is the ground truth reconstructed image.

The main difference is that now the coefficient of the term that directly checks for the resulting image β has a larger value that makes the reconstruction term of the same order of magnitude with respect to the one in k-space, since we are more interested in the visual quality of the images.

Also, β is not a constant but it will start at zero and slowly increase to its maximum value.

The network is trained with the Adam (Kingma and Ba, 2017) optimizer for 300 epochs with a learning rate of $3 \cdot 10^{-3}$ then the learning rate is reduced to $3 \cdot 10^{-4}$ and the network is trained until the validation loss decrease.

The loss coefficient β is initialized at 0 and then it linearly increases from 0 to $5 \cdot 10^3$ between epoch 50 and epoch 200. This final value is chosen to set the order of magnitude of the reconstruction term to be of the same order of magnitude with respect to the loss in k-space.

Unsupervised learning

For the unsupervised training we are using the Noise2Noise framework introduced in Section 7.1.1. In this case there is no ground truth so the loss function is modified to contain only the corrupted version of the images.

We can write the unsupervised loss explicitly, remembering the output of the network in residual learning in Eq. 6.6, as

$$L_k = MSE(S_n - N_{\mathbf{W}}(S_n), S_n^*) \tag{7.12}$$

That with the additive noise model described in Eq. 7.7 became

$$L_k = |\mathcal{S} + n - N_{\mathbf{W}}(S_n) - \mathcal{S} - n^*|^2 \tag{7.13}$$

Where $S_n = s + n$ and $S_n^* = S + n^*$ are both independently corrupted versions of the same original k-space data S as described in Section 7.2.3. This task is essentially to create a map between two independent samplings of the noise.

We must say that for the blind denoising task there is no need to use the same standard deviation of the noise (until the noise model remains the same) and the input to the network does not need to be of better quality with respect to the target. So it is possible to ask the network to map an image with higher PSNR to another with lower PSNR without damaging the training. This is extremely important for a true unsupervised blind denoising since also the level of the noise can be unknown at every step.

We are exploiting the fact that the L_2 distance is the correct estimator of the mean in the pointwise estimation when the noise term has zero mean. For this reason only the term in k-space can be used in the training loss since this relation is invalid for the noise in the Sos image.

If we want to retain the adoption of the reconstruction term we should mix the supervised and unsupervised learning with the possibility of implementing the supervised term only for a small fraction of the dataset.

In contrast to the training in the supervised approach the aim here is not to minimize the training loss since the task is impossible to complete. In residual learning this unsupervised task consists in transforming one instance of the noise to the another one independently sampled.

Fortunately, the gradients are also very large and they point in very different directions at each step but on average during training the effect of the weight will point on the correct solution.

This is possible because we are effectively averaging the gradient effect on the weights over all the pixels in the batch.

For this reason while the training loss after a rapid transient will be constant the validation loss computed on a network that is performing the original task (restoring the clean images and not mapping noise to noise representation) will decrease normally.

In Figure 7.9 is shown the loss function for the training and the validation of the supervised and unsupervised training: In supervised training, the train and validation loss decrease together and they differ from a small amount that is called the training bias. In unsupervised learning instead, we have a training loss that is almost flat and a validation loss that rapidly decreases. Be careful that since parameters are not the same the losses absolute values are not directly comparable.

The network is trained with the Adam (Kingma and Ba, 2017) optimizer for 300 epochs with a learning rate of $3 \cdot 10^{-3}$, afterwards, the learning rate is reduced to $3 \cdot 10^{-4}$ and the network is trained for another 100 epochs and then the learning rate is exponentially decreased to zero.
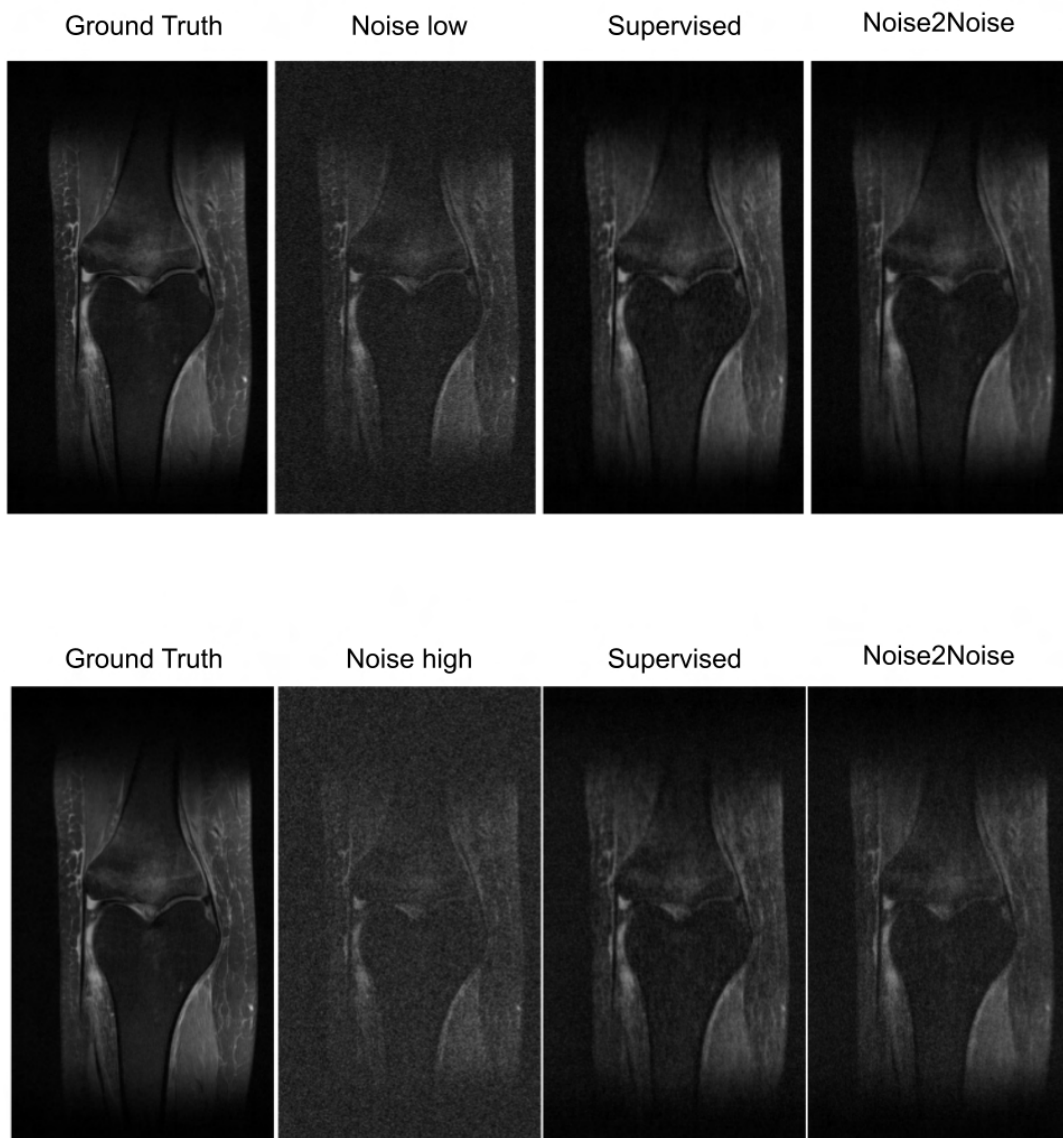
FIGURE 7.3: Noise effect and denoising with K-DnCNN for supervised and unsupervised (Noise2Noise) training at $\sigma = 8 \cdot 10^{-3}$ (Low) and $\sigma = 16 \cdot 10^{-3}$ (High) noise level. This scan is an example of acquisition with fat suppression.
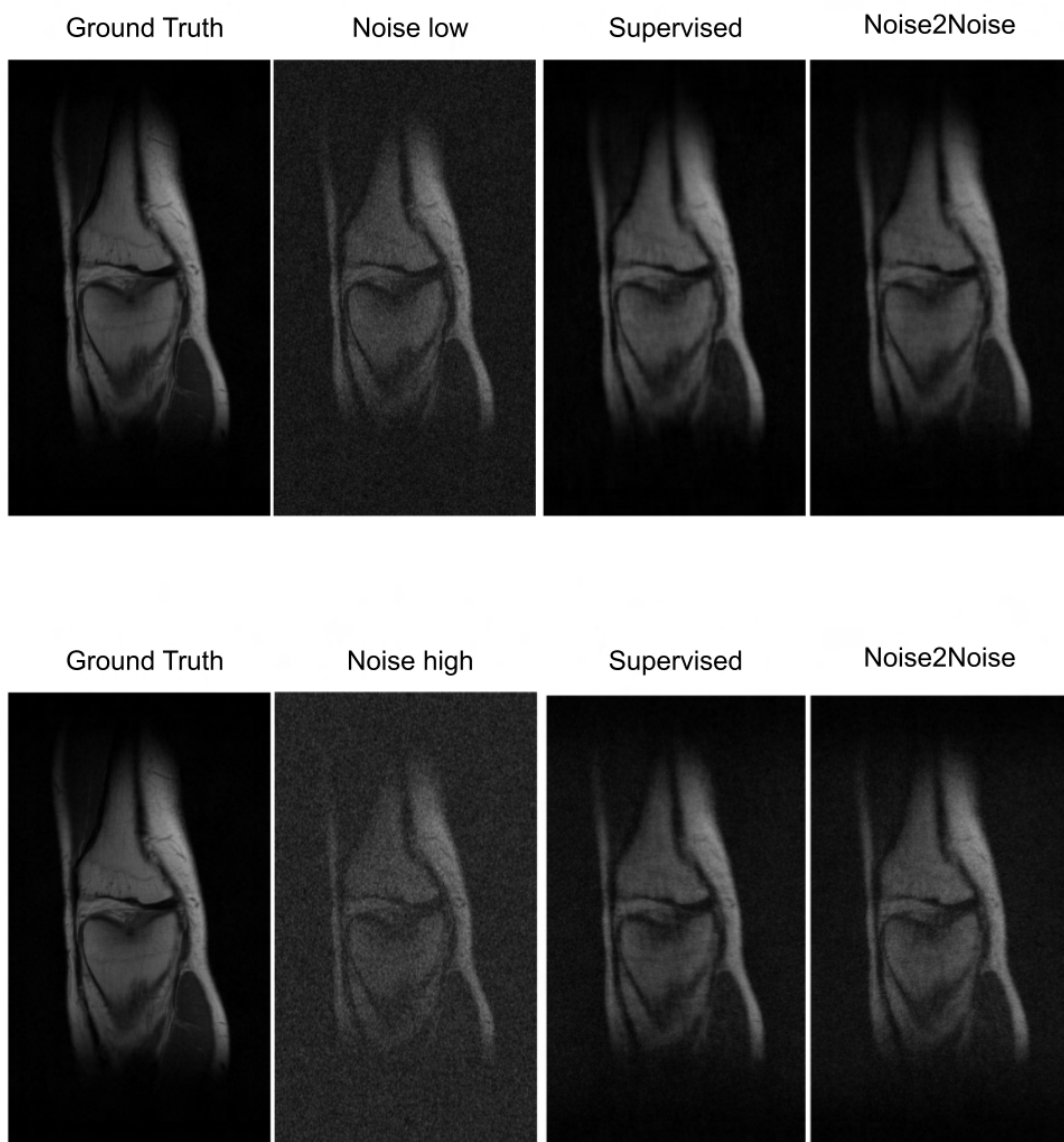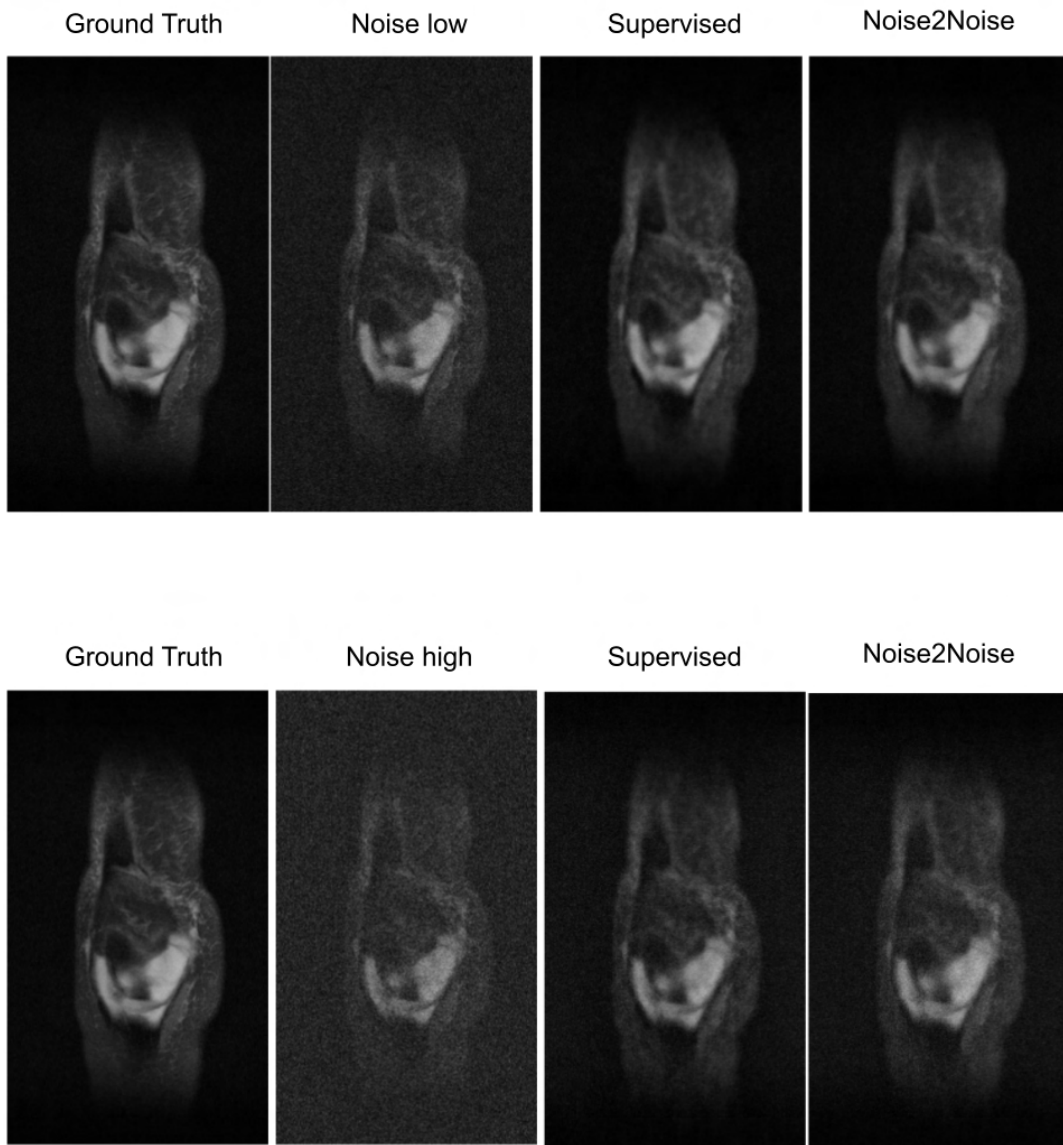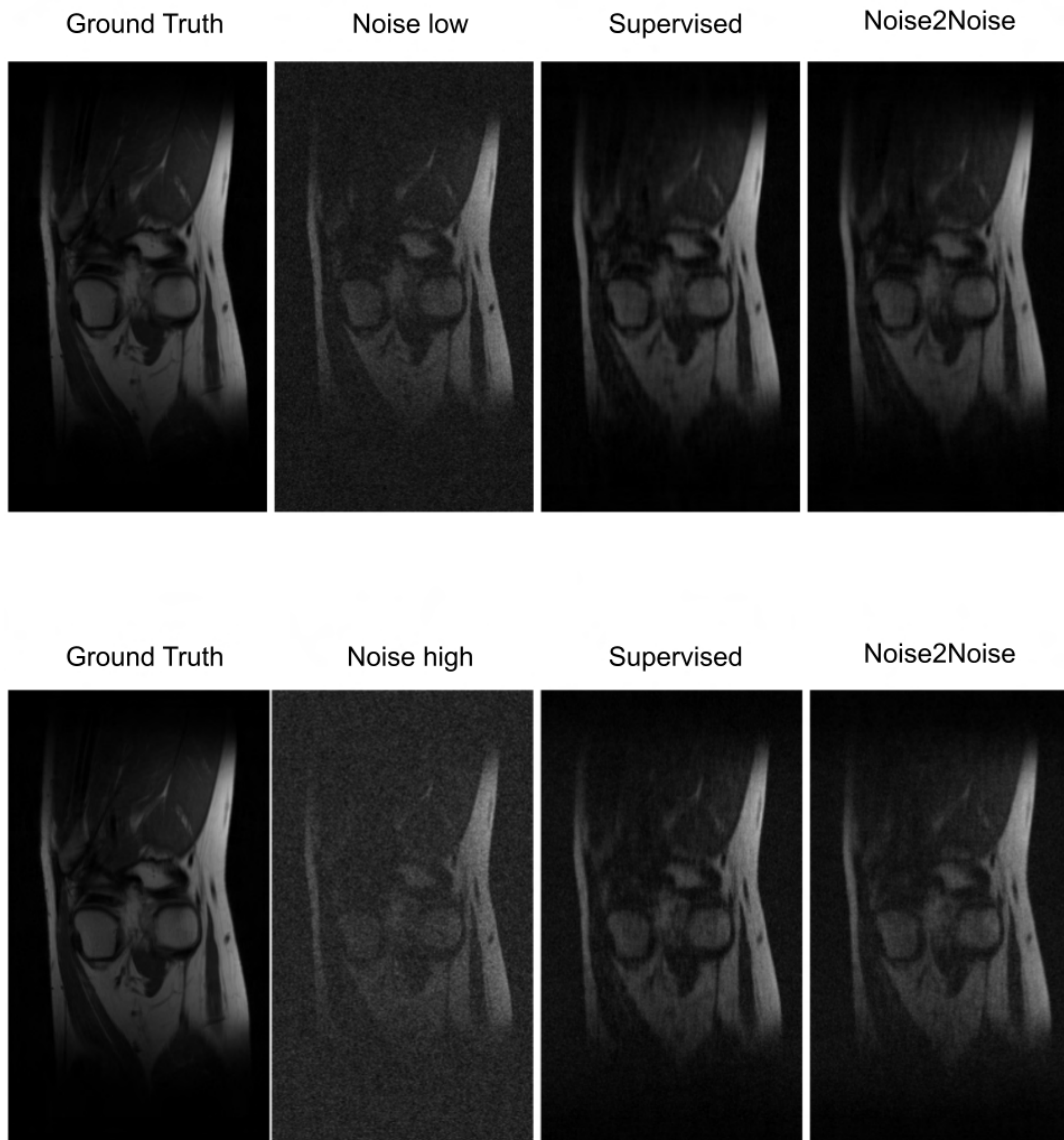
FIGURE 7.4: Noise effect and denoising with K-DnCNN for supervised and unsupervised (Noise2Noise) training at $\sigma = 8 \cdot 10^{-3}$ (Low) and $\sigma = 16 \cdot 10^{-3}$ (High) noise level. This scan is an example of acquisition without fat suppression.
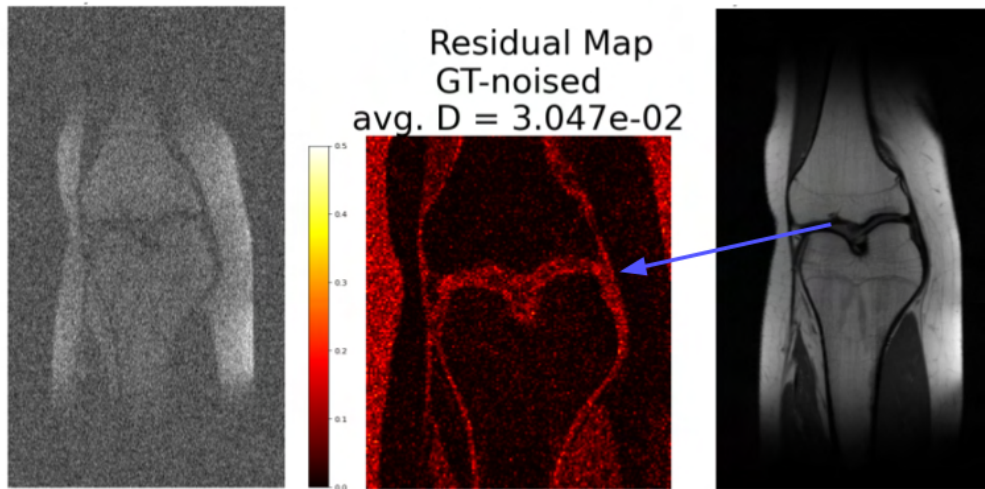
FIGURE 7.5: Noise effect and denoising with K-DnCNN for supervised and unsupervised (Noise2Noise) training at $\sigma = 8 \cdot 10^{-3}$ (Low) and $\sigma = 16 \cdot 10^{-3}$ (High) noise level.

FIGURE 7.6: Noise effect and denoising with K-DnCNN for supervised and unsupervised (Noise2Noise) training at $\sigma = 8 \cdot 10^{-3}$ (Low) and $\sigma = 16 \cdot 10^{-3}$ (High) noise level.

FIGURE 7.7: Right: Ground truth image. Left: Noisy version of it. Center: Example of a residual map. Each pixel is the square difference between the ground truth and the noisy image. This is a clear example of how noise in MRI is not additive and it is not stationary. We can see the anatomy in the residual image, this means that the noise term depend on the intensity value of the image. This effect derives from the reconstruction process of the image since in the k-space the noise is Gaussian additive noise.
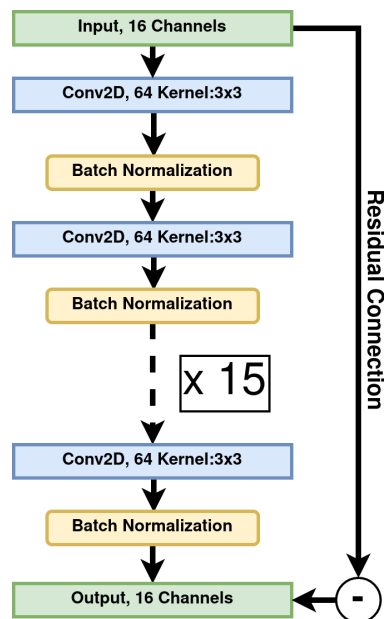

FIGURE 7.8: K-DnCNN model for multichannel MRI denoising.
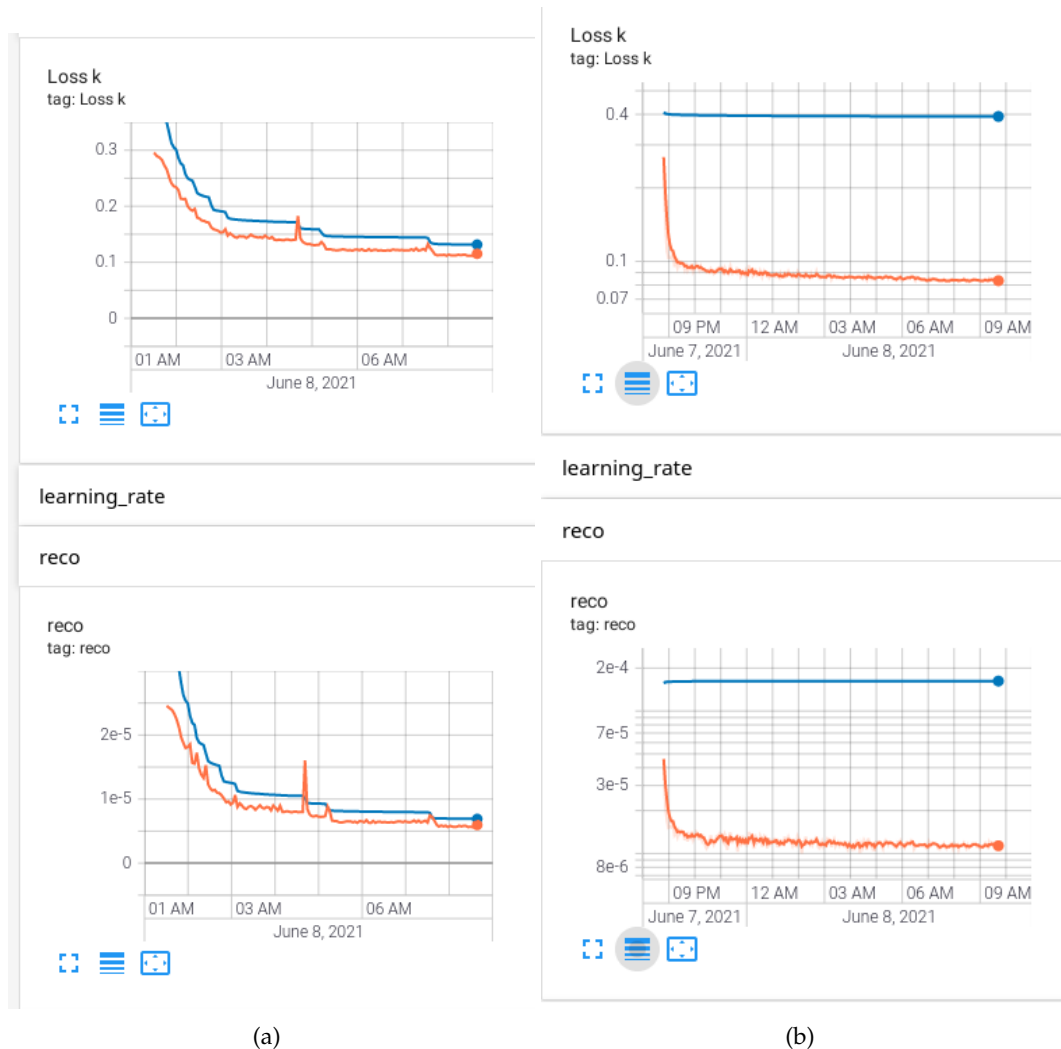
(a)   (b)

FIGURE 7.9: Example of the qualitative differences in training for the (a) supervised and (b) unsupervised approaches.
Train loss (blue) and validation loss (orange) during live training. In the unsupervised Noise2Noise the train loss does not decrease since the train task is effectively impossible. Still, the weight gradients are the correct ones so when the network is validated for the original task the loss correctly decreases.
Since the coefficient β is different is the two cases the absolute values are not comparable. The reconstruction term (reco) is explicitly part of the loss function only in supervised training.

7.3   Results

The Dn-CNN is trained for a blind denoising task so to test its performance we will use different levels of image corruption. The test will be performed using a noise standard deviation of $\sigma = 8 \cdot 10^{-3}$, $16 \cdot 10^{-3}$ that, since the initial image quality is not homogeneous, will produce an average PSNR of the noisy images respectively of (24.3 ± 2.9) dB and (18.9 ± 2.5) dB.

All the results obtained are measured on the reconstructed images only. The denoising in performed in the k-space but since it is never shown to the operator it would be useless to base our results on it.

The final reconstructed image from the denoised k-space, without any post processing (with the exception of a 0 to 1 normalization of pixel intensities), is used in all the tests performed.

The metric used to evaluate the restored image are the PSNR and the SSI and to avoid that large background areas may contribute too much in the calculation, each image is processed entirely but the metrics are computed only on a large central region of signal as shown for example in Figure 7.4 for the whole image and in Figures 7.10 for the central patch.

This is very important since the background is very easy to treat so its presence will give us exceedingly high scores that are not representative of the real performances.

We tested on a hold off portion of data of 100 patients and 2959 scans comprising both acquisition with and without fat suppression.

In Table 7.2 the results for the supervised and unsupervised training are reported. Both approaches consistently improve the image quality at high level of noise and at low level and remarkably, at high noise level, their performance matches.

This is not trivial since the unsupervised approach did not explicitly show the target image but it is nevertheless correctly reproduced.

In Figures 7.12 and 7.15 are reported the distributions of PSNR and SSI for the high and low noise level for the supervised learning while in Figures 7.18 and 7.19 the results for the unsupervised case are reported.

TABLE 7.2: Average results on the test Dataset composed by 100 patients with 2959 slices that were processed separately. The results is obtained at two levels on noise $\sigma = 8 \cdot 10^{-3}$, $16 \cdot 10^{-3}$ that produces an average PSNR of the noisy images respectively of (24.3 ± 2.9) dB and (18.9 ± 2.5) dB. We report the results for both the K-DnCNN trained with supervised learning and unsupervised learning (Noise2Noise) for the PSNR and structural similarity index (SSI). Also, the gain (Δ) respect to the noisy image is reported. At the higher level of noise the two methods are comparable while at low level the supervised training has better performance. Results are reported as mean $\pm$ std.

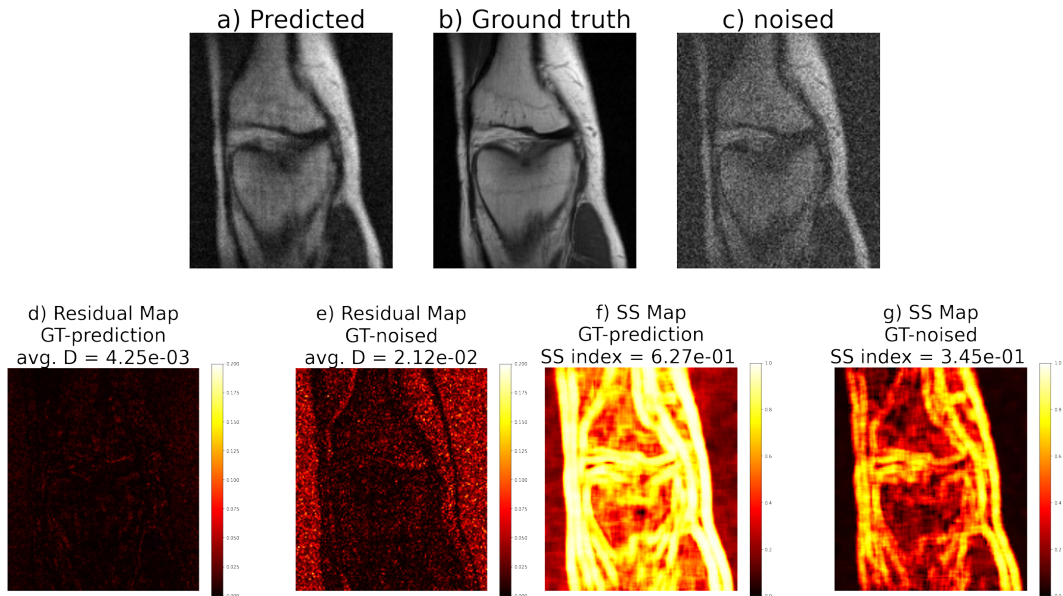|	Noise std	PSNR (dB)	Δ PSNR (dB)	SSI	Δ SSI
Supervised					
	$\sigma = 8 \cdot 10^{-3}$	29.3 ± 3.1	5.1 ± 2.1	0.80 ± 0.06	0.2 ± 0.1
	$\sigma = 16 \cdot 10^{-3}$	25.6 ± 2.6	6.7 ± 1.8	0.60 ± 0.10	0.3 ± 0.1
Noise2Noise					
	$\sigma = 8 \cdot 10^{-3}$	28.3 ± 3.4	4.0 ± 3.1	0.80 ± 0.06	0.2 ± 0.1
	$\sigma = 16 \cdot 10^{-3}$	25.6 ± 2.8	6.7 ± 2.4	0.61 ± 0.10	0.3 ± 0.1

FIGURE 7.10: a-c) Example of denoising on the celtral image patch.
d-e) Residual maps.
f-g) Structural similarity maps.
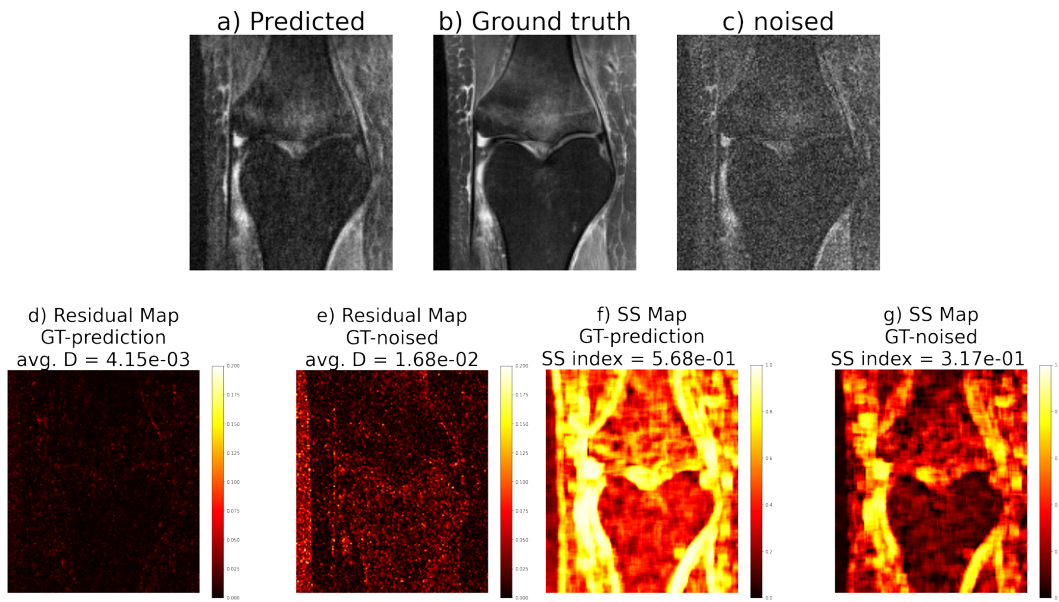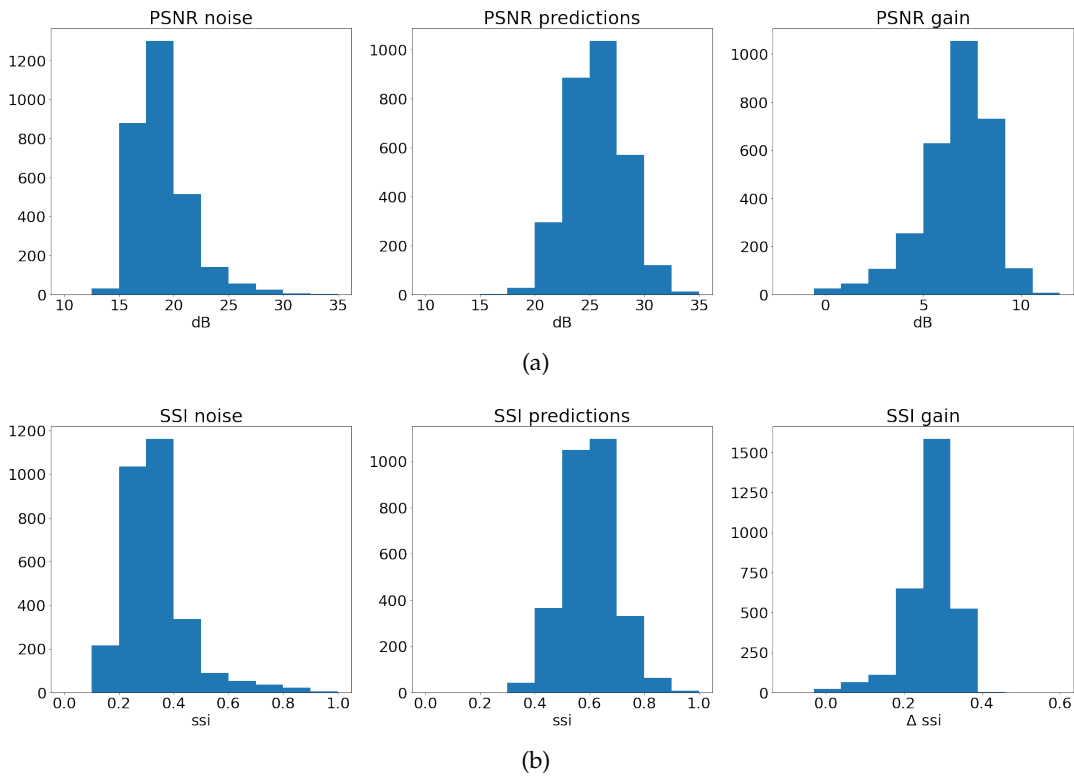Results for the supervised training at noise level $\sigma = 16$

FIGURE 7.11: a-c) Example of denoising on the celtral image patch.
d-e) Residual maps.
f-g) Structural similarity maps.
Results for the supervised training at noise level $\sigma = 16$

FIGURE 7.12: Distibution of the a) PSNR and b) SSI of the value computed at slice level on the noisy image (Left), on the restored image (center),and the slice-wise gain. Results for the supervised training with at noise level $\sigma = 16$
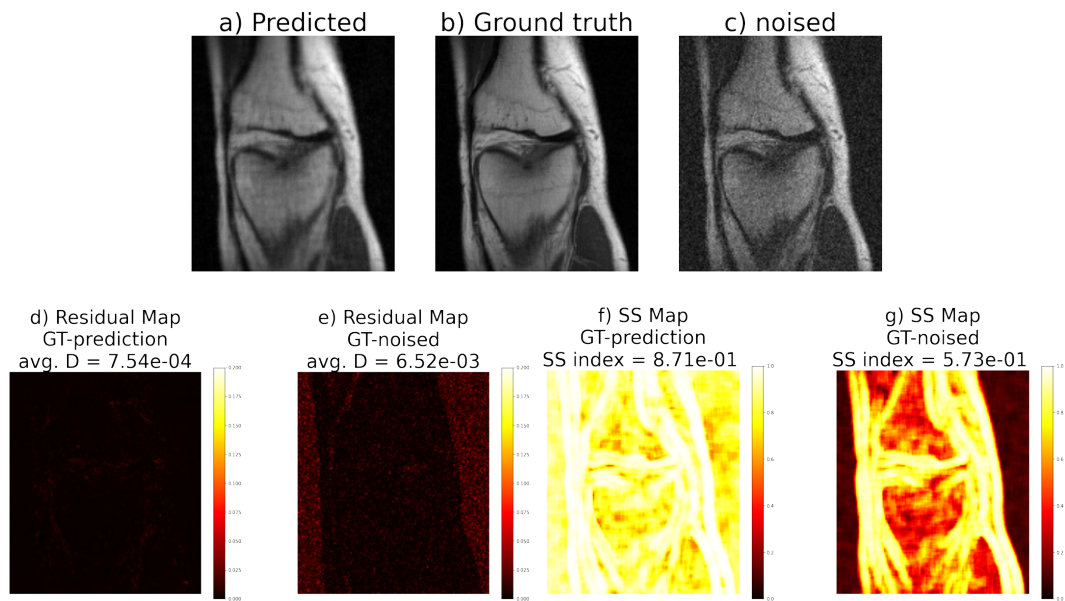
FIGURE 7.13: a-c) Example of denoising on the celtral image patch.
d-e) Residual maps.
f-g) Structural similarity maps.
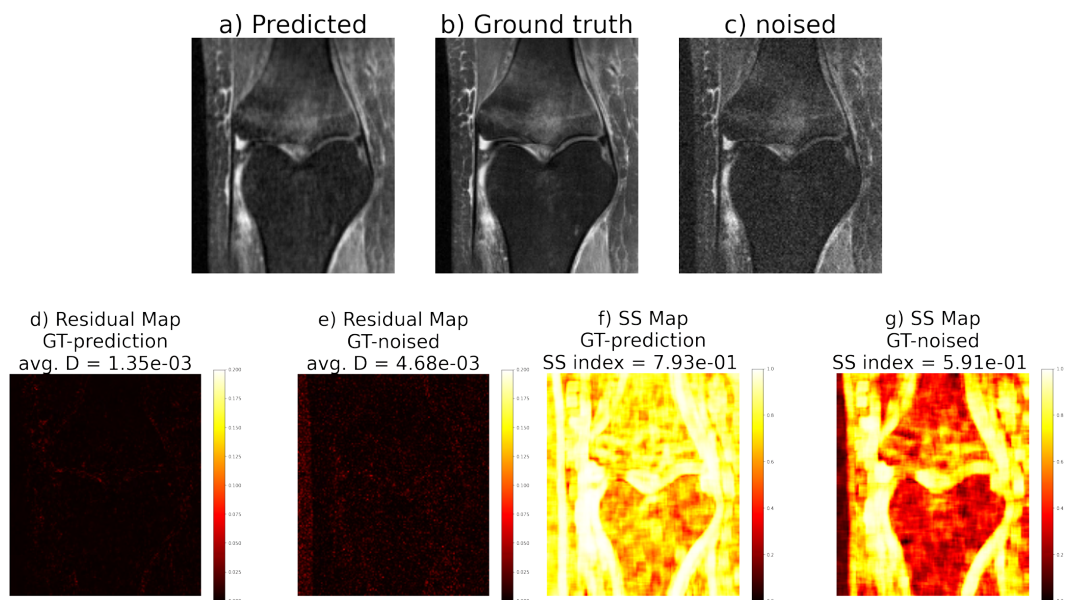Results for the supervised training at noise level $\sigma = 8$

FIGURE 7.14: a-c) Example of denoising on the celtral image patch.
d-e) Residual maps.
f-g) Structural similarity maps.
Results for the supervised training at noise level $\sigma = 8$
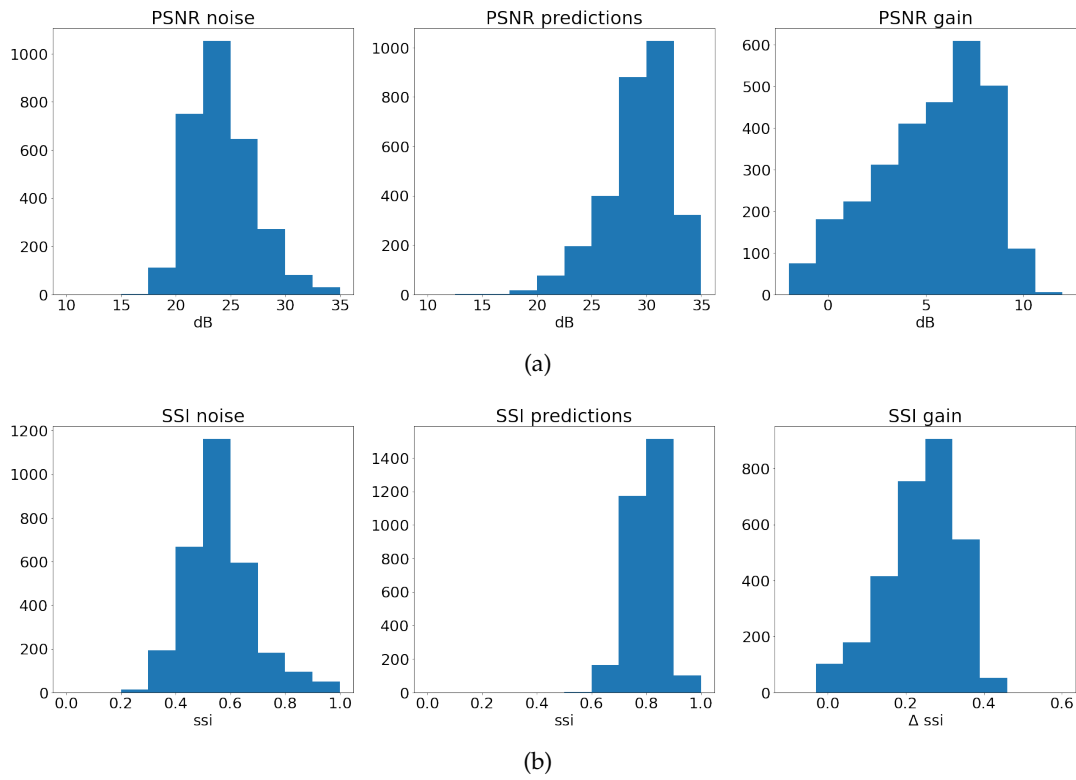
(a)

(b)

FIGURE 7.15: Distibution of the a) PSNR and b) SSI of the value com-
puted at slice level on the noisy image (Left), on the restored image
(center),and the slice-wise gain. Results for the supervised training
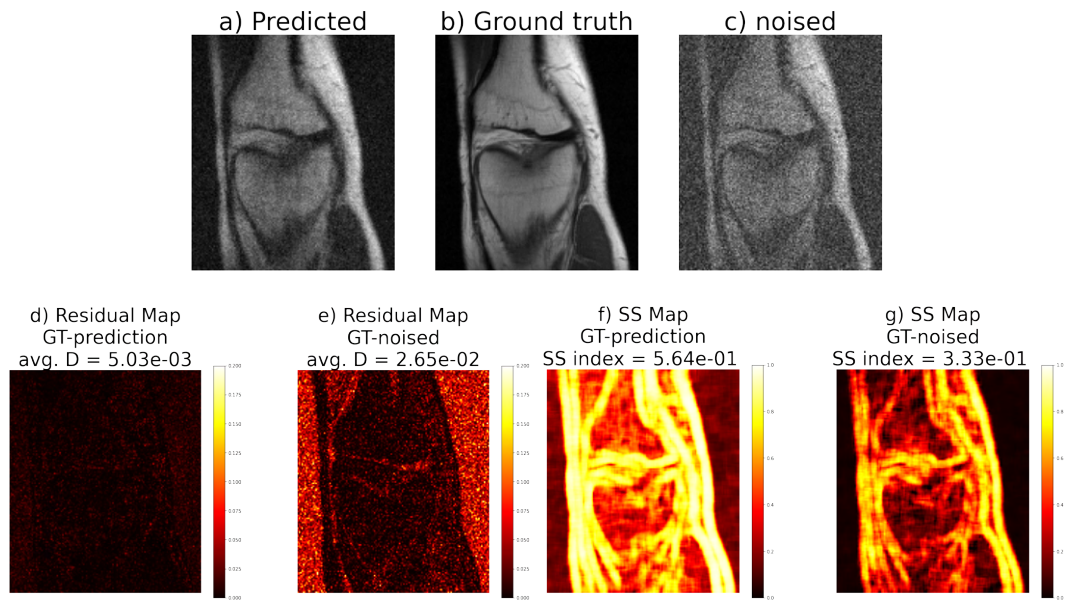with at noise level $\sigma = 8$

FIGURE 7.16: a-c) Example of denoising on the celtral image patch.
d-e) Residual maps.
f-g) Structural similarity maps.
Results for the unsupervised Noise2Noise training at noise level $\sigma =$
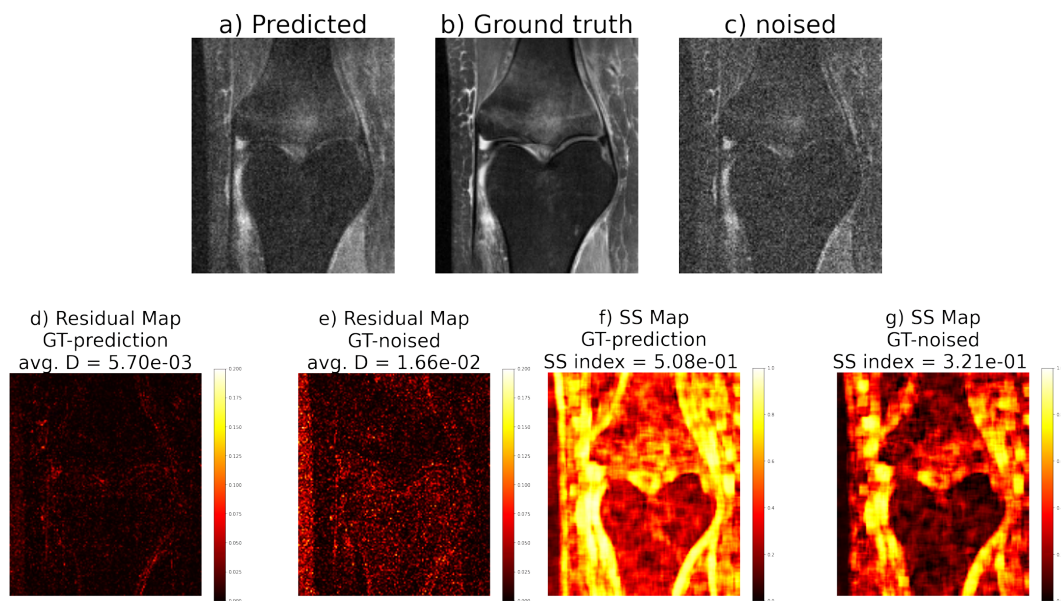16

FIGURE 7.17: a-c) Example of denoising on the celtral image patch.
d-e) Residual maps.
f-g) Structural similarity maps.
Results for the unsupervised Noise2Noise training at noise level $\sigma = 16$

(a)

(b)

FIGURE 7.18: Distibution of the a) PSNR and b) SSI of the value computed at slice level on the noisy image (Left), on the restored image (center),and the slice-wise gain. Results for the unsupervised Noise2Noise training with at noise level $\sigma = 16$
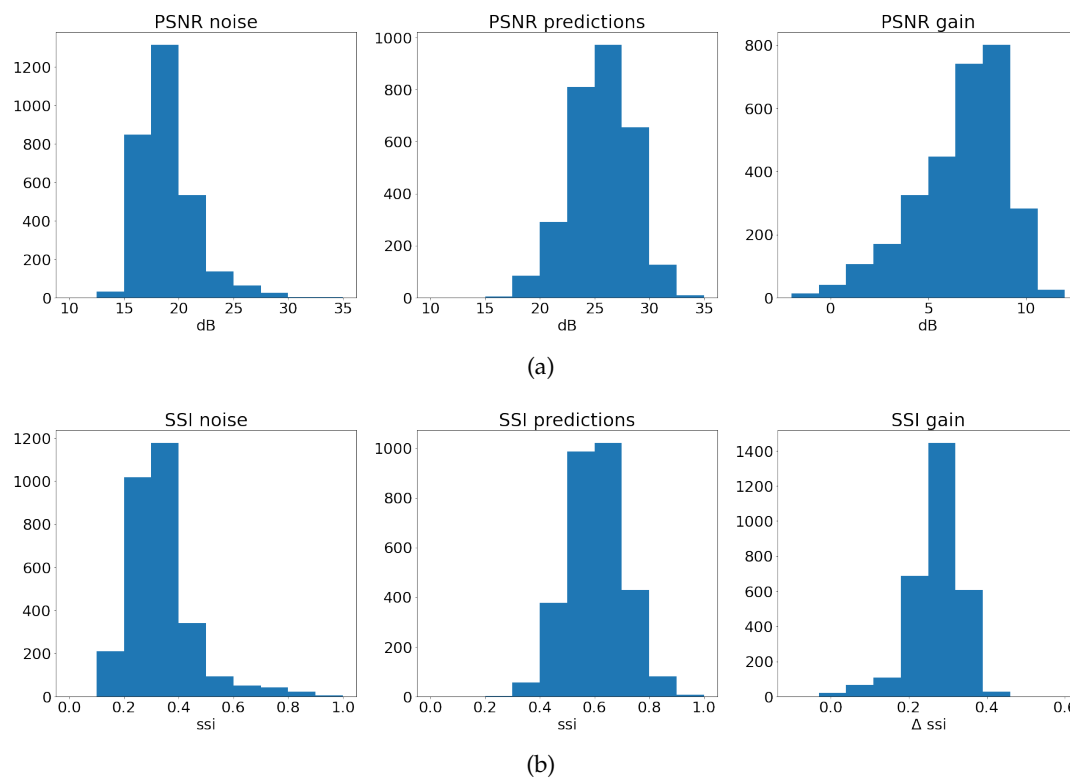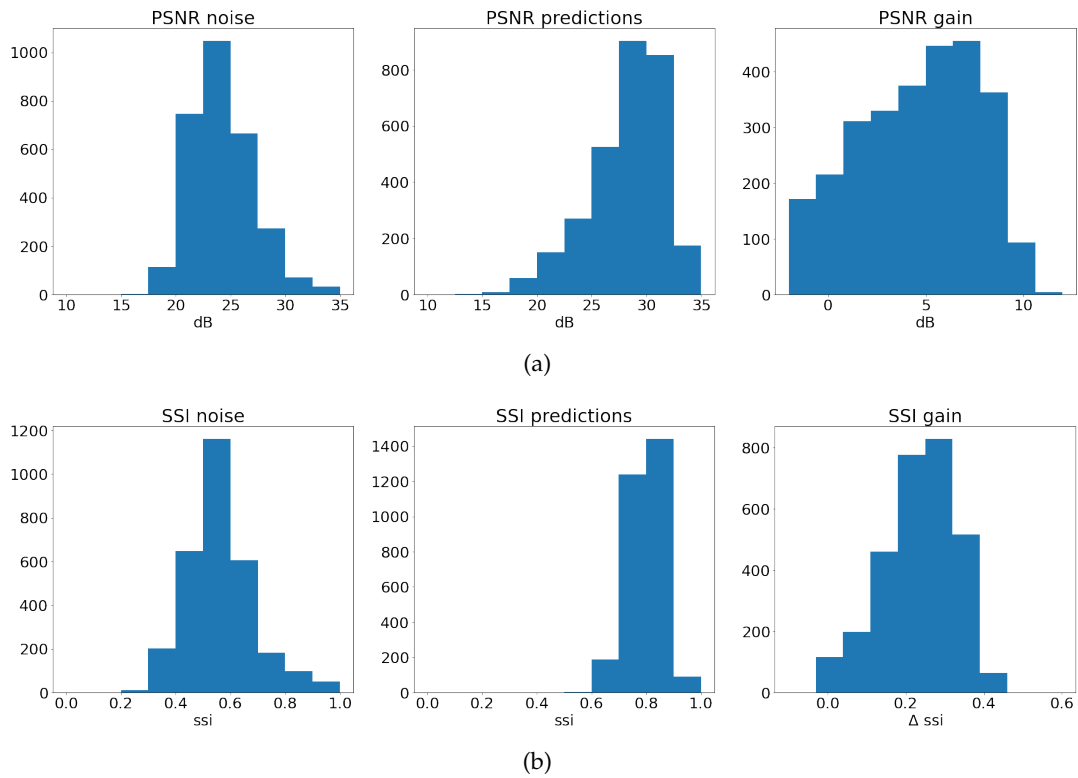
(a)

(b)

FIGURE 7.19: Distibution of the a) PSNR and b) SSI of the value computed at slice level on the noisy image (Left), on the restored image (center),and the slice-wise gain. Results for the unsupervised Noise2Noise training with at noise level $\sigma = 8$

7.3.1 Application to the brain dataset

The possibility to perform the denoising task on data of the same type but with meaningful differences respect to the one used during training is a sought after feature for a method that may be implemented in real practice.

To test if we reach reasonable performance on k-space data derived from different NMR acquisition sequences and from a different body area, we applied the Dn-CNN trained on the knee dataset to the denoising of brain data that are also present in the FastMRI dataset.

MRI of the brain differs profoundly in both acquisition parameters and general visual effect, also the brain dataset contains a larger variety of acquisition modality since both T2 weighted, T1 weighted with and without contrast medium and FLAIR (Fluid Attenuated Inversion Recovery, a heavily T2 weighted acquisition) are present. Additionally they were acquired with different hardware and different scanners.

Compared to the knee dataset, which is mostly homogeneous in acquisition parameters, this brain dataset is definitely more challenging for denoising especially when it is not directly used for training.

Also, the shapes, contrasts, and average intensities of a brain scan are very visually different from the ones that are commonly found on an arthroscopic acquisition such as the one on the knee.

For these reasons, this preliminary task of denoising based on training on a different dataset , is difficult to overcome and the results are an important test of the generalizability of the method .

To test the Dn-CNN denoiser trained for the blind denoising task with supervised learning, we selected 637 slices of brain scans from 255 patients present in the validation set of the Brain FastMRI dataset and we add noise with a $\sigma = 16 \cdot 10^{-3}$ and with the correlation between coils defined in 7.2.3 to them to simulate an highly noisy acquisition.

Since the starting quality of the scans are very different because it depends on the acquisition used and the noise already present, this noise injection produces a varying effect on the images in terms of quality.

The range of PSNR and SSI of the corrupted version of the images and of the predictions is shown in Figure 7.22. The mean PNSR is $(23.6 \pm 4.5) dB$ that usually correspond to highly noisy images with clearly visible fluctuations in intensity areas and background as shown in Figures 7.21c and 7.20c .

After the application of the denoiser the average gain of image quality is 4.6 ± 2.7 dB for the PSNR and 0.2 ± 0.1 for the SSI that means that on average the image is improved both in its original intensity restoration and pixel correlation.

An example of processed images and residual and SS maps can be found in figures 7.21d-f and 7.20d-f.

The results are a bit inferior but in line with the ones obtained earlier with the knee dataset. This is not trivial since this kind of images are never shown during the training and the variability in the dataset is very large.

To be more reasonable this test should at least be conducted with acquisition of different body parts that are acquired with the same sequence but unfortunately they are not available.
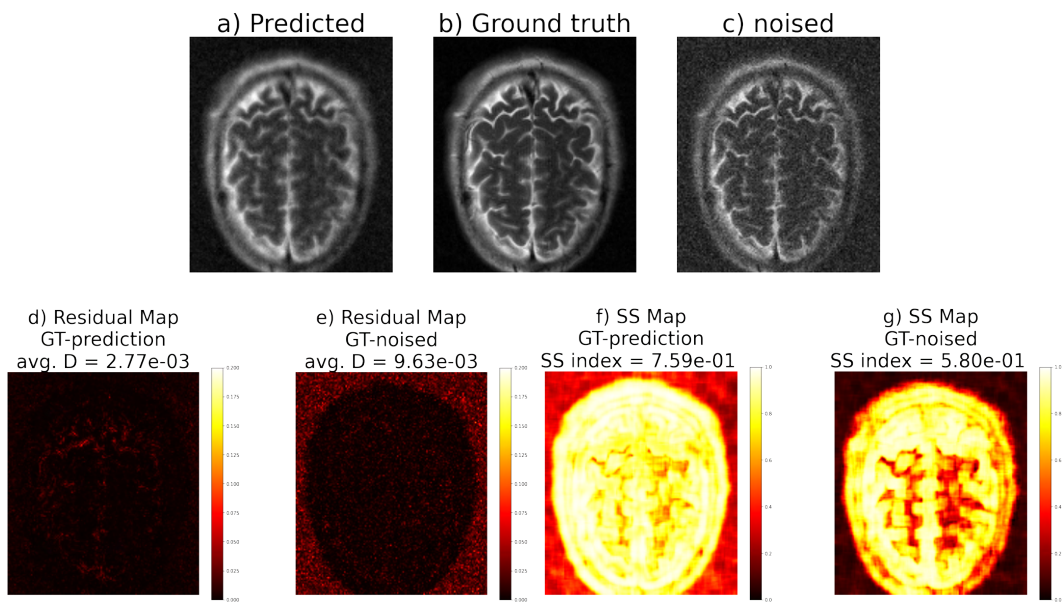
FIGURE 7.20: Brain - std noise $sigma = 16$
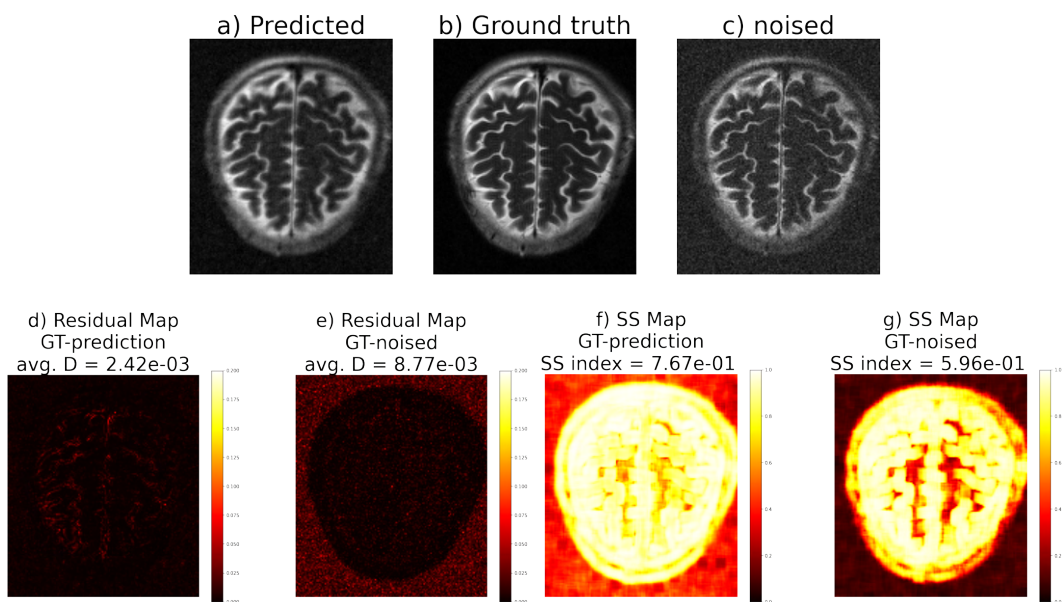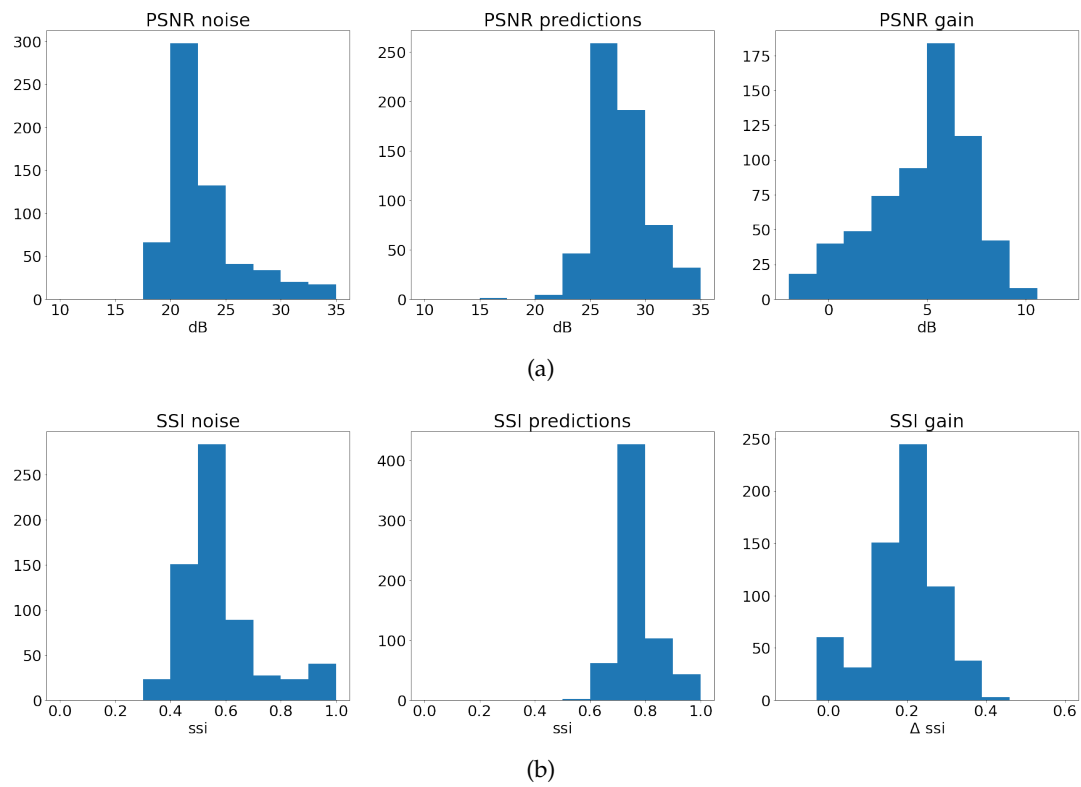

FIGURE 7.21: Brain - std noise $sigma = 16$

(a)

(b)

FIGURE 7.22: Results for the brain dataset: Distibution of the a) PSNR and b) SSI of the value computed at slice level on the noisy image (Left), on the restored image (center),and the slice-wise gain. Results for the unsupervised Noise2Noise training with at noise level $\sigma = 16$

7.3.2   Comparison with non local means denoising

To validate our denoising strategy we need to motivate our choice to develop a deep learning based method by conducting an analysis of its performance and points of strength with respect to already proposed, state of the art, denoisers.

As a first step, we must compare the performance achieved, in terms of our metrics the PSNR and SSI, with other MRI denoising solutions that are not based on deep learning.

In the previous Chapter we applied to the same denoising task the wiener filter and an image based Dn-CNN. The comparison was useful to analyze the behavior of the network in the high noise regime but for a fair assessment of the network capabilities, a more performant method should be chosen.

Some of the first approaches in MRI filtering were simple kernel based convolutions with a gaussian assumption of the noise model but they are not implemented in practice anymore.

The reasons behind the needs of developing new methods of MRI denoising is that simple spatial filters do reduce the effect of noise, generally increasing the SNR of images, but these methods also increase the blurr of edges, and deform the structure of small details depending on the size of the filter.

Moreover, in the high noise regime, they introduce a bias in the processed image since usually they do not correctly model the noise form.

In order to avoid these limitations, one common strategy is to adopt filters based on adaptive windows that try to estimate the correct pixel intensity from its neighborhood while avoiding pixels that come from other tissues to be in the same estimation window.

This is the case for the so-called *non local neighbors* methodology in which the adaptive windows filters search the optimal neighbours across the entire image with the aim of finding similar structural neighbors of the voxel of interest. The choice of the optimal neighbours is performed using a similarity measure between local patches.

This is the main idea behind the non-locals-means (NLM) algorithm initially proposed by Buades et al. (Buades, Coll, and Morel, 2005) for natural image denoising that was later introduced to MRI images and now is one of the most popular and adopted methods.

The NLM algorithm is a non linear filter based on a weighted average of pixels inside a large search window so a non local search, in terms of spatial distance from the center pixel. The weights are selected to preserve the structure of the image so pixels in the average have a larger weight if their neighbours are similar to the neighbours of the pixel of interest. The similarity between neighbour patches is defined, for example, as the MSE between the two.

In other words, in a large block every small patch is confronted with the neighborhood of the pixel of interest and this pixel intensity is corrected with a weighted average of all pixels that share a similar neighborhood thus with the assumption that they are in the same context: a background pixel, a border pixel or a tissue pixel).

The simpler formulation of NLM presented in (Buades, Coll, and Morel, 2005) is the following:

The corrected intensity of the pixel $\tilde{A}(p)$ is the average of the intensity $M(q)$ of pixels in a large windows Ω with some weights $w(p,q)$ that depends on the distance

between the context of p and the context of q.

$$\tilde{A}(p) = \sum_{q \in \Omega} w(p,q)M(q) \tag{7.14}$$

And the set of weights is

$$w(p,q) = \frac{1}{\zeta(p)} exp(-\frac{d(p,q)}{h^2}) \tag{7.15}$$

With $\zeta(p)$ a normalization constant ,

$$\zeta(p) = \sum_{q \in \Omega} exp(-\frac{d(p,q)}{h^2}) \tag{7.16}$$

And $d(p,q)$ is a similarity measure at patch level between the neighborhood of p and the neighborhood of q, while h is related to the noise power in the image and it is a parameter of the algorithm.

The distance $d(p,q)$ can vary between different implementations but usually is a distance in the distribution of gray levels between the two patches.

For the performance comparison with our proposed method, we will adopt the implementation specifically derived to denoise 3D MRI, the adaptive optimized non-local means (AONLM), by coupe et al (Coupe et al., 2008; Manjón et al., 2010).

In AONLM, the restored intensities of a small block are the weighted average of the intensities of all similar blocks within a predefined neighborhood. The 3 dimensional AONLM extends the 2 dimensional NLM filtering by 3D using a block-wise implementation.

This algorithm is a standard in the MRI denoising and it is also adopted as a comparison with a deep learning method in the recent work of Tian et al. (Tian et al., 2021) that proposes a RL CNN method for the denoising of magnitude images.

An implementation of the algorithm is present in the widely used Dipy package[2]. The AONLM was performed assuming Gaussian noise with $3 \times 3 \times 3$ patch radius and $5 \times 5 \times 5$ search volume (block). These parameters are selected by a grid search algorithm with the aim of maximizing the PSNR gain over 5 volumes of the validation set.

The NLM does not perform blind denoising, thus an estimate of σ of the noise has to be used as a parameter for the algorithm.

The Dipy implementation of the algorithm estimates the noise power using the method presented in the work of Koay et al. (Koay and Basser, 2006) that correct the estimate for the multicoil acquisition but do not treat the correlation between coils introducing a bias in the estimate that leads to an overestimation of the true noise variance.

Following the procedure reported in (Aja-Fernández and Vegas-Sánchez-Ferrero, 2016), if the noise correlation matrix **æ** is known, for the noise level used in the test, it is possible to introduce a correction to the estimate multiplying by a factor $(1 + \bar{\rho}(L-1))^{-1/2}$ where $\bar{\rho}$ is the average of the off-diagonal terms of the matrix and L is the number of coils.

---

[2]https://dipy.org/documentation/1.4.1./interfaces/denoise_flow/#coupe11

For the noise model defined in equation 7.7 this factor is 0.65. To check if other contributions has to be taken into account, we also performed an optimization of the noise estimation by a grid search of the best parameter to be used in terms of PSNR gain.

In the regime of noise used, a correction to the estimation of the noise power of 0.68 ± 0.04 was found to be optimal and it is in agreement with the effect of the correlation between coils reported above. In figure 7.23a is shown the best value of the noise power used as a parameter in the algorithm in function of the uncorrected estimation. In figure 7.23b the improvement of performance adopting the correction is reported. Both results was obtained using 5 patients from the validation set of the Knee dataset that were not used in the test.

We compared the PSNR and SSI results obtained with the K-DnCNN with the supervised and unsupervised training and the NLM. We tested 195 volumes of test set the knee data and 95 volumes of test set the brain data. As specified in section 7.3.1 the networks are trained only on the knee dataset. All the metrics are computed only on the central part of the image and values calculated over slices are averaged in each volume.

In Figure 7.24 is shown the PSNR and the SSI of the processed image in function of the original quality of the noisy image. Each volume is corrupted with a noise with standard deviation $\sigma_K \in [6 \cdot 10^{-3}, 16 \cdot 10^{-3}]$ at steps of $2 \cdot 10^{-3}$ and then denoised with both the K-DnCNN and the NLM algorithm.

Our proposed method gives on average better or equal results than the NLM for all the noise levels tested for both the metrics used as reported also in figure 7.25

On the brain dataset we also obtain better or equal results at high noise level but in this case the difference in performance is less pronounced. The results are shown in Figure 7.26 and 7.27.

We expected different results in the two datasets since the denoising task learned from the knee data is probably only in part transferable to the brain dataset than is extremely more complex and it does not contain the same acquisition modality found in the knee dataset.

These results are encouraging but it is important to stress two points to put them in the right perspective.

First, the NLM does not maximize the PSNR as an objective function as the K-DnCNN (that minimize the MSE in the k-space) so we are comparing a specialized tool to a generic one. For this purpose it is probably better to take more into account the SSI metric instead.

Second, the NLM implementation used is not optimized for the acquisition modality in exam. In particular, it does not directly treat the noise correlation in coils.

Also, we can say that the neural network has access to more information to perform the denoising task since it can process the information acquired from the multiple coils to leverage the redundancy in the acquisition while the NLM algorithm works only on the final reconstructed image.

Still, NLM is considered one of the state of the art methods for MRI denoising thus this comparison produces a valid benchmark from which to start for future developments of our denoising method.
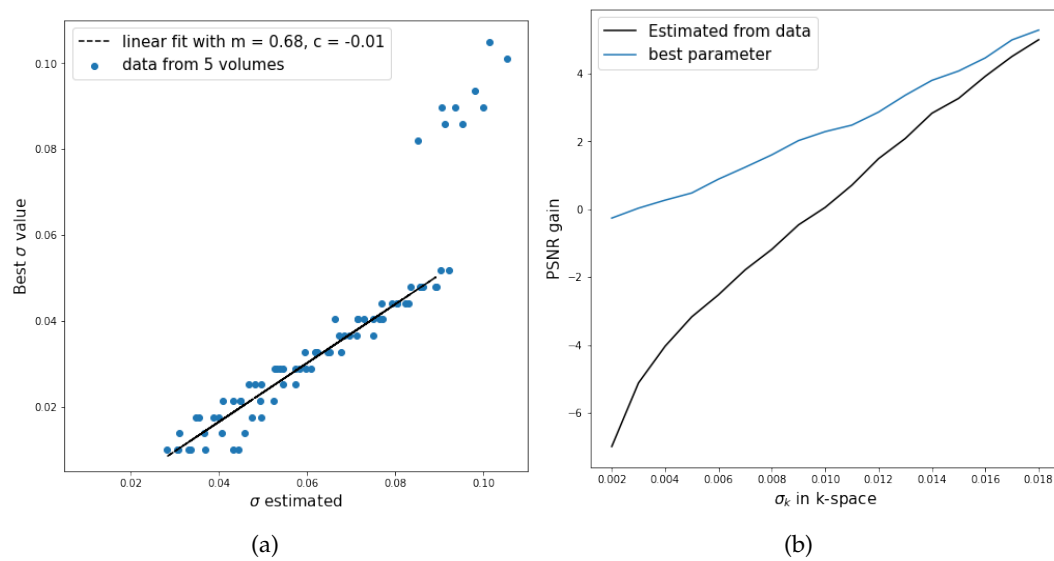
FIGURE 7.23: (a) Best noise level estimation for the maximization of PSNR gain in function of the estimation performed in the Dipy package (see text). The discrepancy in the two values can be attributed to the effect of noise correlation in the coil array. The Dotted line is a linear fit performed in the range of noise level used in the test. (b) PSNR gain after denoising when the noise estimation is performed with the Dipy package (Black line) and with the corrected value (Blue Line) in function of the noise level applied to the K-space.
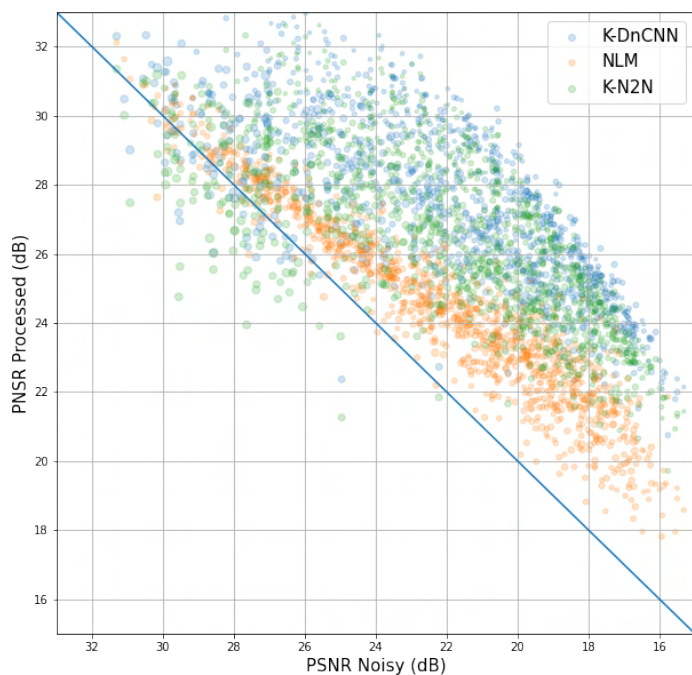
(a)
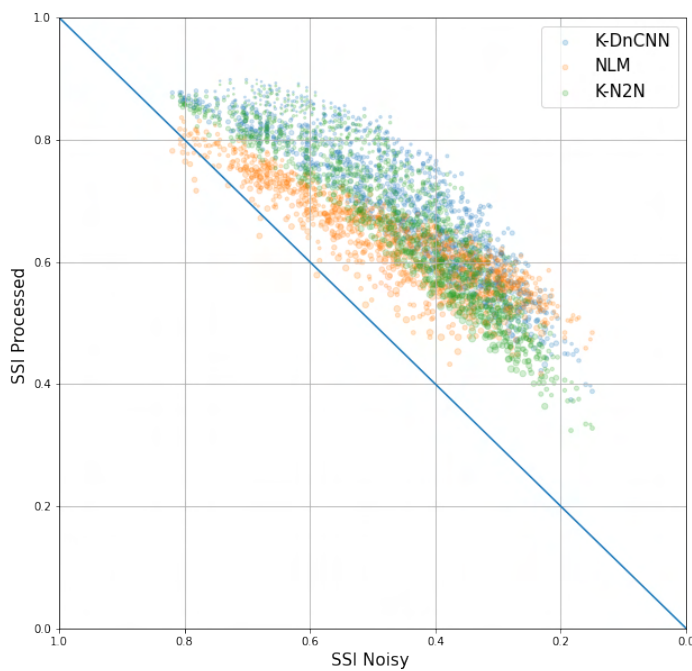
(b)

FIGURE 7.24: Performance metrics on the knee dataset. PSNR (a) and SSI (b) of the denoised image for the neural network with the supervised approach (Blue), the unsupervised approach (Green) and the NLM (Orange) in function of the quality of the starting noised image.
The quality of the starting image is ordered from high initial (a) PSNR or (b) SSI (low noise) to low initial quality (high noise) The diagonal indicates no gain in terms of (a) PSNR or (b) SSI. The Deep learning approaches, at high noise level, obtain better or equal results than the NLM for both the supervised and unsupervised approach. Each point is the average over all the slices in a volume taking into account only the central image patch for the images in the test set of the knee dataset. The noise levels used cover all the interval the network is trained for.

(a)   (b)

FIGURE 7.25: (a) Gain in PSNR (PSNR of the denoised image - Initial PSNR of the noisy image) and (b) gain in SSI (SSI of the denoised image - Initial SSI of the noisy image) averaged over the 195 volumes of the test set of the knee data for the neural network with the supervised approach (Blue), the unsupervised approach (Green) and the NLM (Orange) in function of the noise standard deviation in the k-space.

(a)
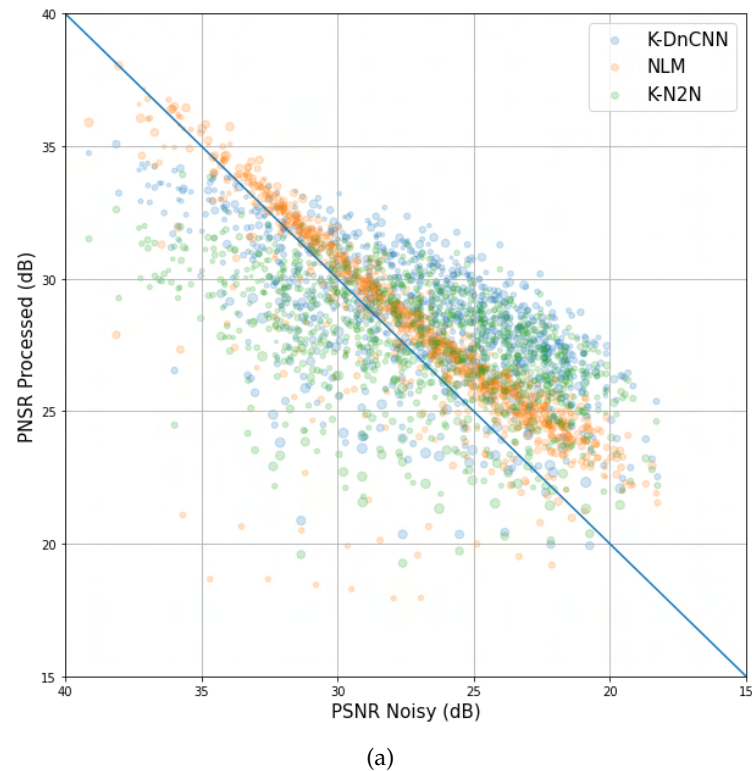
FIGURE 7.26: Performance metrics on the brain dataset. PSNR (a) and SSI (b) of the denoised image for the neural network with the supervised approach (Blue), the unsupervised approach (Green) and the NLM (Orange) in function of the quality of the starting noised image.

The quality of the starting image is ordered from high initial (a) PSNR or (b) SSI (low noise) to low initial quality (high noise) The diagonal indicates no gain in terms of (a) PSNR or (b) SSI. The Deep learning approaches, at high noise level, obtain better or equal results than the NLM for both the supervised and unsupervised approach. Each point is the average over all the slices in a volume taking into account only the central image patch for the images in the test set of the brain dataset. The noise levels used cover all the interval the network is trained for.
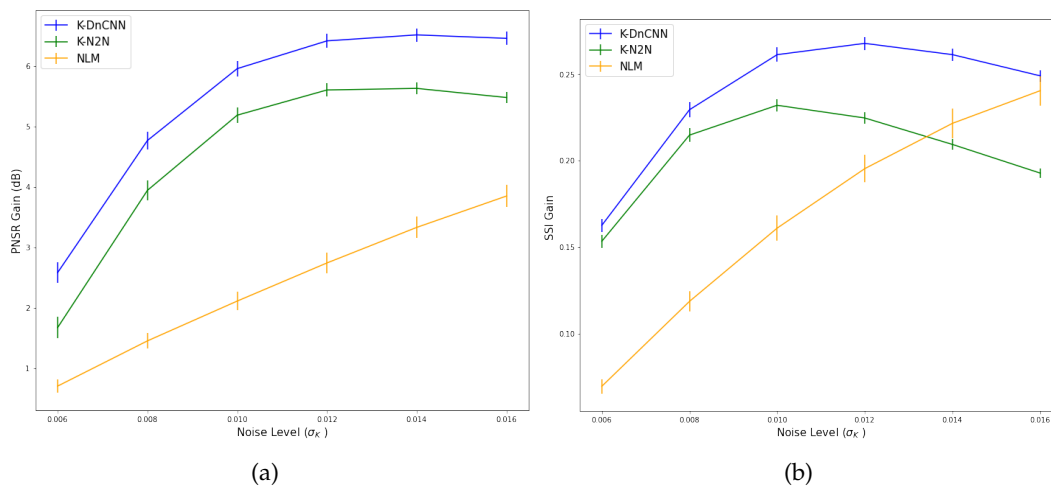
FIGURE 7.27: (a) Gain in PSNR (PSNR of the denoised image - Initial PSNR of the noisy image) and (b) gain in SSI (SSI of the denoised image - Initial SSI of the noisy image) averaged over the 95 volumes of the test set of the brain data for the neural network with the supervised approach (Blue), the unsupervised approach (Green) and the NLM (Orange) in function of the noise standard deviation in the k-space.
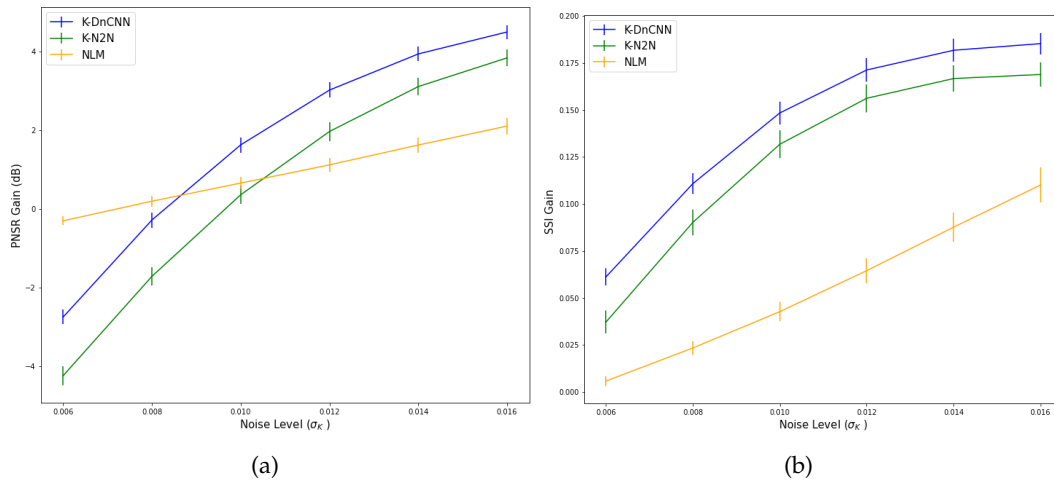
7.4   Discussion

In this Chapter we performed the denoising task on the k-space raw data of the FastMRI dataset that is the largest and more complete one of its type.

The method we proposed is based on the residual learning denoising applied to the frequency data instead of denoising directly the reconstructed images.

Residual learning is a powerful framework that in presence of additive noise allows the network to concentrate in building high level representation of the noise component instead of the clean image.

In MRI, and in particular in multicoil acquisition, the noise can not be assumed additive in the reconstructed image since the process to obtain it from the raw data is non linear and introduces mixed terms and non additive terms is the noise form.

In Chapter 6 we discuss how in a simple simulation performing the denoising task with residual learning over the frequency data produces superior results than applying a network of the same complexity directly on the image data.

Now we have applied the same method to the denoising of multicoil data for morphological imaging. In this case the most important result is the correct reproduction of the anatomical parts of the knee such as the muscle, the bones and the cartilages.

For this reason we quantify our results in terms of PSNR and SSI: the first one measures the pixel wise restoration of the true intensity and second the correct reproduction in a small window of the correlation between and contrast between the original and predicted images.

In our test both metrics improved substantially after the action of the denoiser in a blind denoising task both at a high level of noise and at a lower level. This is an important result since the starting quality on a clinical MRI depends on a large variety of factors and a denoising solution must be applicable in most cases.

We also tested the denoiser on data acquired on brains which is both a different anatomical district, where the shape and size of the details are different from the ones that the denoiser was trained with, and more importantly it consisted in different MRI sequences.

We know that the k-space data depends on the acquisition sequence chosen so this ability to generalize cannot be taken for granted. The results obtained were encouraging and the ability to generalize to different sequence and anatomical subjects is probably one of the most important perks to develop in view of the applicability in a clinical setting.

Another important result is that in our experimental pipeline, the supervised and unsupervised training reach comparable performance with the same set of examples. This is not trivial since in general unsupervised learning needs a larger amount of data.

This is the result of the implementation of the Noise2Noise approach that, in turn, is easily applicable only because we decide to work in the k-space where the noise form is simpler than the one present in images.

The possibility to perform unsupervised learning really widens the applicability of the method since it is often impractical or impossible to acquire ground truth images for a supervised training.

Since our final aim is the application to denoising MRI scans from acquisition of etereonuclei the possibility to reach a good performance with unsupervised learning

is mandatory to the successful application since the images are noisy due to the low concentration of resonant molecules so the quality of imaging is always intrinsically low making the definition of a ground truth very difficult.

We also performed a preliminary comparison of our proposed method with a state of the art denoising algorithm that is widely adopted for MRI denoising, the NLM.

Both the supervised and unsupervised approach performed better or at the same level respect to the NLM algorithm and this result should motivate us to continue the development of this methodology since it has the potential to be competitive with state of the art methods with a relatively smaller, once it is trained, computational cost.

The direction of improvements of these results are many: In the short term we should train the network on the whole dataset and check if we reach gains in performance. In general, a study of performance in function of the training size is always useful to put a low-end to the dataset size needed for training in a real case scenario.

Again, in the short term we should restore the original data to which we cut the higher frequency part that produced a somewhat easier task since we lowered the noise content but probably erased also some small details in the image that are the most difficult to retain when the network is trained with a distance based loss function such as the MSE: A few pixels do not impact the loss function much and the network is prone to concentrate on the general intensity.

The visually most relevant effect of this behaviour is the smoothness of the surfaces that is usually present in the processed images.

In our case details are usually restored, at least in the low noise case, but we can still improve the sharpness of the final results. However, we should be careful to maintain the important aspect of our solution, that is to retain the quantitative information of the original image, and we shouldn't sacrifice this ability to force a better visual effect.

Another improvement will be to utilize all the available coils in the denoising. This will surely improve the results since more signal from the same volume is present in each coil acquisition and a neural network can easily exploit this information.

We discard a part of the coils to obtain a number of channels multiple of eighth in our input images to take full advantage of the Tesor Cores. Probably the best solution will be to create a dummy coil to be added to the original 15 with useful information such as the average absolute value of the signal to both introduce additional information and reach a suitable number of channels.

Also, training with the brain dataset and cross checking if performance is maintained when testing on the knee dataset will be another useful information. And since the brain dataset consists of many different acquisition modality we can also check if training with the right sequence example is needed to reach the best performance or training can effectively be generalized across acquisition modality.

Instead, with regard to unsupervised training specifically, we are now limited by an unrealistic scenario where the number of noisy examples of a real exam is theoretically infinite since we are generating pairs of corrupted samples automatically.

This is not a realistic method since the number of noisy acquisition of a subject is usually finite and relatively small (tens of copies). As the number of ordered paired sample instead is large (i.e $N * (N - 1)$) it is possible to try the Noise2Noise with finite noise realization in MRI and it may still be a viable approach. In this context we should not forget that for the supervised approach the limit is even more constraining since if the noise is not generated from a model then the number of possible pairs noise / ground truth is just N.

From the point of view of selecting the best neural network approach for the MRI denoising task, the architecture proposed has many advantages as we deeply discussed in the conclusion of Chapter 6:

It works on partial context only since its receptive field is smaller than the image and it effectively processes patches of the input. Thus, it should reduce the leakage of information derived from learning to the denoised example since it lacks the spatial coherence to create large details.

Furthermore , it uses a residual learning approach that, in addition to helping in the performance increase, it may also contribute to the reduction of unwanted artifacts in the final image.

Also, the network has a small size in terms of number of parameters and a relatively simple architecture that should help make it robust to overfitting.

Still, many possible new architectures that maintain the original spirit can be tested to find the optimal solution within these constraints.

For example a layer block derived from the popular ResNet (He et al., 2016) is often employed with success in denoising and can substitute our simple convolutional layer maintaining the general strategy intact while helping the deployment of deeper solutions.

Finally, after the architecture and parameter optimization will be completed we must compare the final performance reached with the state of the art in MRI denoising that is not based on deep learning and it is optimized for multicoil acquisition with strong noise correlation.

We already tested the performance of one of the most successful image based denoising methods that is also used to improve quantitative measurement in MRI affected by rician noise is the Non-Local-Means (Manjón et al., 2008) and it was be a fitting benchmark for our proposed solution but we should strive to a more fair comparison with newly proposed method that are specifically optimized for our task.

In a completely different direction, we can think of merging the task for which the FastMRI dataset was collected for, the reconstruction of frequency subsampled data, with the denoising task. These two problems are independent but related since they both influenced the quality of MRI when the acquisition time must be short.

It is theoretically possible to solve both problems with the same network and our architecture can be modified accordingly.

Incidentally, a variance of the noise2noise framework that we are implementing was proposed for subsampled MRI data (even if not in the case of interest in the FastMRI challenge (Lehtinen et al., 2018)) so we may also be able to perform unsupervised learning in this scenario.

If this solution is implemented and it proves to be competitive it may be a step forward in a general porpoise deep learning based MRI acquisition enhancement in the frequency space.

Chapter 8

Conclusions

In this concluding section I will try to put the results achieved in my work into context.

This work has carried out some necessary steps in order to make possible the experimentation on animal model of magnetic resonance imaging as a quantification tool of fluorinated tracers for proton therapy enhancement applications.

This work is composed of two parts: a section in which I concentrate on the experimental work that was needed to perform the preliminary measures to validate our choice of the tracer for the cancer model our research is aimed at.

These experiments are directly moved by the interest of collecting necessary preliminary data prior to the quantification experiment in imaging of this tracer for in in-vivo applications.

A second part deals instead with the development of a denoising system for magnetic resonance imaging based on deep learning. This part also is theoretically addressed to the same problem, because the rationale behind its development was that when MRI is performed on nuclei that are not hydrogen the signal is inevitably low and therefore the measurements are noisy.

Nonetheless, it can be considered a method that has its own function as it can generally be applied to magnetic resonance image acquisitions regardless of the acquisition modality since it is a general denoising technique that works on raw frequency data that is collected in every MRI scan.

The first point addressed is the internalization of BPA inside the cells. This point is fundamental because the accumulation in the cells is an essential step for the BPFT to have an effect and therefore any future experiment, for this to be useful to contribute to the development of this therapy, must be carried out with a molecule that is able to deposit boron atoms inside cells.

This experiment led us to define the next course of the work because initially the possible candidates were many and the choice of F-BPA as a model molecule influenced the subsequent experiments. In particular the proposed B12F12, that had the advantage of 12 identical fluorine atoms, was discarded because it does not internalize in PANC1 cells.

The technique we adopted was aimed to measure the internalization of F-BPA but it is a general methodology that can be applied to every molecule that contains both boron and fluorine atoms.

The strength of our measurement is the fact that three different techniques based on orthogonal physical principles were compared under the same experimental conditions in the quantitative measure of the internalization fraction of F-BPA.

In particular, the internalization fraction of F-BPA in the PANC1 cell line was determined by us for the first time.

The second point addressed was the measure of the in vivo distribution of the drug in the tissues of our animal model after the administration at the dose that is used in the proposed therapy.

As we have seen, it is clear in the literature that BPA does not distribute itself homogeneously in the tissues and it is equally clear that its distribution strongly depends on the protocol used for the somministration that itself is strongly dependent on the operators.

Before the conduction of the imaging experiment we needed to test our protocol to set a ground truth to validate the results.

For this reason, we have tried to measure with an alternative technique but more sensitive than imaging, quantitative NMR spectroscopy, the content of fluorine in the different tissues.

To accomplish this measure we have performed 19F MRS in samples extracted ex vivo and compared our results with neutron autoradiography.

This experiment was necessary in order to have a starting point on which to program the parameters of the imaging experiment both to evaluate which is the best acquisition strategy given the real signal concentration, both to fine-tune our experimental pipeline.

The last point especially proved crucial because the encountered problems in our experimental protocol, that greatly influenced our results, would have been more difficult to be verified indirectly with an in vivo experiment in MRI.

This experiment will also have fundamental consequences on how continuing our research.

Beyond the problems encountered, we realize that probably we should somehow increase the fluorine concentration in the tissue since the measured fluorine content was at the limit of our detection capabilities.

This suggests different paths to take, each one valid and promising: we can try to find another boron carrier, we can change the delivery system or we can try to improve the sensibility of our measure.

This brings us to the third point, the development of the denoiser.

It was expected that very low concentrations would cause a very weak signal, thus very noisy images, and we dedicated ourselves to studying a software solution to increase our ability to reduce the effect of the noise.

We then developed a deep learning-based denoiser that could be implemented for this purpose.

We started from the very simple yet very important consideration that the form of the noise in the images already reconstructed is more complex than the one in the raw data acquired in frequency. While our conclusion is almost trivial from a statistical perspective, it significantly eases the task of denoising.

In particular the choice of working with raw data in the k-space allowed us to implement the residual learning strategy, that in image denoising has many advantages, to our problem and that would have been more difficult, or impossible without important modification, to use it directly on the images.

The first step was to test our idea on simulated data which allowed us to appreciate the strengths of our solution in a controlled environment in which we compared it against another neural network model that performed the same denoising task on the corresponding reconstructed images and we found that denoising in k-space allows us to have better performance in image restoration and the restored image, while less visually attractive, keeps the average value of the intensity over an area equal to the intensity value measured on the original image.

This property is particularly interesting when the denoising has to be performed before a quantitative study over the image.

Additionally, our method is very robust to overfitting since it has a deep architecture but with a relatively low number of parameters and it works on partial context only, that means that the noise removal is made at small patches level and that is a valid strategy to avoid the creation of artifacts.

The second step was to validate our system on real data. We choose to work with the FastMRI dataset that is the largest collection of MRI frequency raw data available.

With this dataset we proved that we can in fact properly denoise reconstructed images from the k-space with good performance at high noise level.

In this task we also adopted an unsupervised approach in training. This means that we can perform the learning on the neural networks without providing a ground truth. It is important to notice that the framework that we adopted, the Noise2Noise, that allows the training with pairs of corrupted images instead the usual pair on corrupted and clean images is applicable in our case only because we work in the k-space.

This result is quite interesting in the task of denoising the FastMRI dataset but more importantly it is extremely significant for our original goal of applying it to denoise quantitative fluorine MRI.

To perform the training in normal learning we have to provide the network with noise free high quality images that in our case, especially in in-vivo applications, may be impossible to obtain in practice.

Since the expected concentration is so low we may be in the situation in which our acquisition would already be the best possible measure so a ground truth may be difficult to define.

The Noise2Noise approach will provide the means of training the network with multiple noisy acquisitions of the same subject instead of providing a ground truth of high quality.

If we would be able to confirm the applicability of this strategy with phantom study in 19F MRI it may be possible to transfer the application in-vivo.

As far as future developments are concerned, there are many directions to explore.

From the experimental point of view we still have to optimize our delivery of the drug, it may be the most rapid solution to increase the concentration in tissues.

Another solution would be to abandon the use of F-BPA and test new tracers, in this case the know-how developed can be easily transferred to other possible molecules.

As for the imaging with 19F part instead, once we will determine the molecule of choice and the concentration at which we will have to work given the adopted delivery system, we will need to develop the best pulse sequence and acquisition method that is controlled by the MR properties of the nuclei being measured.

In particular, Relaxation Times determine the time frames of signal losses, so the time scale for data acquisition should be assigned correctly.

Other parameters, such as acquisition bandwidths, should also be properly planned to avoid recording unnecessary noise ,to achieve the highest possible SNR and to optimize the acquisition method to promote signal sensitivity.

We already discussed the problem of noise in MRI but also other effects come in place to disrupt the quality of the measure and the quantitative estimation of concentration. Particularly for quantitative 19F MRI, the dishomogeneity in the RF pulse

and in the receiver field will impact the performance when comparing concentrations of with an external standard of known 19F concentration.

This effect may also be treated with a deep learning approach but it will need a dedicated study and solution.

Finally, the direction of development of the denoiser can be pursued along different paths: Probably the data needed for training will eventually be proved prohibitive for the application to the denoising in our case. Even with the unsupervised training we will need thousands of acquisitions to reach acceptable results.

Thus the first possible development is to work toward a realistic simulation of the acquisition to pre train the denoiser on simulated realistic data to reduce the number of real acquisition for training.

Also, we should consider the denoiser to be a separate part of the work that can be deployed also to other problems. We had good performance with morphological multiple coils acquisition and optimizing the method to be reliably adopted in this field may be a valid choice.

The first step in this process is to establish a benchmark to compare our performance to other methods that are not based on deep learning. This step will provide the necessary boundaries to understand if our methods may be of general applicability.

Also, at the moment we only test simulated noise even with the real data. This is a major weakness since the noise model adopted is realistic but we will have to test if performance is maintained on real low SRN acquisitions.

The field of deep learning denoising methods is in rapid development for MRI acquisition and it still lacks a reliable all-purpose solution. The particular acquisition process of an MRI makes it often difficult to adopt common strategy used in natural images denoising and for what the raw data solutions are concerned, they are difficult to generalize since every scanner manufacturer performs the acquisitions and the reconstructions in different ways.

In any case there is plenty of room for future proposals and a properly tuned method that can handle multiple imaging modalities may also be directly implemented in the scanning software.

Bibliography

Abadi, Martín et al. (2015). *TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems*. Software available from tensorflow.org. URL: https://www.tensorflow.org/.

Aja-Fernández, Santiago and Gonzalo Vegas-Sánchez-Ferrero (2016). "Noise Estimation in Multiple-Coil MR Data". In: *Statistical Analysis of Noise in MRI: Modeling, Filtering and Estimation*. Ed. by Santiago Aja-Fernández and Gonzalo Vegas-Sánchez-Ferrero. Cham: Springer International Publishing, pp. 173–186. URL: https://doi.org/10.1007/978-3-319-39934-8_8.

Aja-Fernández, Santiago et al. (2016). "Acquisition and Reconstruction of Magnetic Resonance Imaging". In: *Statistical Analysis of Noise in MRI: Modeling, Filtering and Estimation*. Cham: Springer International Publishing, pp. 9–29. DOI: 10.1007/978-3-319-39934-8_2. URL: https://doi.org/10.1007/978-3-319-39934-8_2.

Arbelaez, Pablo et al. (May 2011). "Contour Detection and Hierarchical Image Segmentation". In: *IEEE Trans. Pattern Anal. Mach. Intell.* 33.5, pp. 898–916. ISSN: 0162-8828. DOI: 10.1109/TPAMI.2010.161. URL: http://dx.doi.org/10.1109/TPAMI.2010.161.

Athersuch, T. J. et al. (July 2010). "UPLC-MS, HPLC-radiometric, and NMR-spectroscopic studies on the metabolic fate of 3-fluoro-[U-14C]-aniline in the bile-cannulated rat". In: *Xenobiotica* 40.7, pp. 510–523. ISSN: 0049-8254. DOI: 10.3109/00498254.2010.483294. URL: https://doi.org/10.3109/00498254.2010.483294.

Barth, Rolf F., Peng Mi, and Weilian Yang (June 2018). "Boron delivery agents for neutron capture therapy of cancer". In: *Cancer Communications* 38.1, p. 35. ISSN: 2523-3548. URL: https://doi.org/10.1186/s40880-018-0299-7.

Barth, Rolf F. et al. (June 2014). "Evaluation of unnatural cyclic amino acids as boron delivery agents for treatment of melanomas and gliomas". In: *15th International Congress on Neutron Capture Therapy Impact of a new radiotherapy against cancer* 88, pp. 38–42. ISSN: 0969-8043. URL: https://www.sciencedirect.com/science/article/pii/S0969804313005824.

Bi, Huichang et al. (June 2013). "Optimization of harvesting, extraction, and analytical protocols for UPLC-ESI-MS-based metabolomic analysis of adherent mammalian cancer cells". In: *Analytical and Bioanalytical Chemistry* 405.15, pp. 5279–5289. ISSN: 1618-2650. URL: https://doi.org/10.1007/s00216-013-6927-9.

Bortolussi, S. et al. (Feb. 2011). "Boron uptake measurements in a rat model for Boron Neutron Capture Therapy of lung tumours". In: *Applied Radiation and Isotopes* 69.2, pp. 394–398. ISSN: 0969-8043. URL: https://www.sciencedirect.com/science/article/pii/S0969804310004471.

Brown, Francis F., Iain D. Campbell, and Philip W. Kuchel (Oct. 1977). "Human erythrocyte metabolism studies by 1H spin echo NMR". In: *FEBS Letters* 82.1, pp. 12–16. ISSN: 0014-5793. URL: https://doi.org/10.1016/0014-5793(77)80875-2.

Buades, A., B. Coll, and J. . Morel (2005). "A non-local algorithm for image denoising". In: *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*. Vol. 2, 60–65 vol. 2.

Callaghan, Paul (1991). *Principles of Nuclear Magnetic Resonance Microscopy*. illustrated edition. Clarendon Press; Oxford University Press. ISBN: 0198539444,9780198539445.

Capuani, Silvia et al. (Oct. 2008). "L-DOPA Preloading Increases the Uptake of Borophenylalanine in C6 Glioma Rat Model: A New Strategy to Improve BNCT Efficacy". In: *International Journal of Radiation Oncology • Biology • Physics* 72.2, pp. 562–567. ISSN: 0360-3016. DOI: 10.1016/j.ijrobp.2008.06.1493. URL: https://doi.org/10.1016/j.ijrobp.2008.06.1493.

Chapelin, Fanny et al. (2017). "Fluorine-19 nuclear magnetic resonance of chimeric antigen receptor T cell biodistribution in murine cancer model". In: *Scientific Reports* 7.1, p. 17748. ISSN: 2045-2322. DOI: 10.1038/s41598-017-17669-4. URL: https://doi.org/10.1038/s41598-017-17669-4.

Chen, M. et al. (May 1992). "An in vivo 19F NMR study of isoflurane elimination as a function of age in rat brain". In: *NMR Biomed.* 5.3, pp. 121–126. ISSN: 0952-3480. DOI: 10.1002/nbm.1940050304. URL: https://doi.org/10.1002/nbm.1940050304.

Chong, Edwin and Stanislaw Zak (2001). "An Introduction to Optimization". In: 2nd. SERIES IN DISCRETE MATHEMATICS AND OPTIMIZATION. WILEY-INTERSCIENCE. Chap. 8th. URL: https://www.wiley.com/en-us/An+Introduction+to+Optimization%2C+4th+Edition-p-9781118279014.

Chou, F. I. et al. (July 2009). "Suitability of boron carriers for BNCT: Accumulation of boron in malignant and normal liver cells after treatment with BPA, BSH and BA". In: *Applied Radiation and Isotopes* 67.7, Supplement, S105–S108. ISSN: 0969-8043. DOI: https://doi.org/10.1016/j.apradiso.2009.03.025. URL: http://www.sciencedirect.com/science/article/pii/S0969804309002589.

Cirrone, G. A. P. et al. (Jan. 2018). "First experimental proof of Proton Boron Capture Therapy (PBCT) to enhance protontherapy effectiveness". In: *Scientific Reports* 8.1, p. 1141. ISSN: 2045-2322. URL: https://doi.org/10.1038/s41598-018-19258-5.

Coderre, Jeffrey A. et al. (May 1997). "Boron neutron capture therapy for glioblastoma multiforme using p-boronophenylalanine and epithermal neutrons: Trial design and early clinical results". In: *Journal of Neuro-Oncology* 33.1, pp. 141–152. ISSN: 1573-7373. URL: https://doi.org/10.1023/A:1005741919442.

Coupe, P. et al. (Apr. 2008). "An optimized blockwise nonlocal means denoising filter for 3-D magnetic resonance images". eng. In: *IEEE transactions on medical imaging* 27.4, pp. 425–441. ISSN: 0278-0062. URL: https://pubmed.ncbi.nlm.nih.gov/18390341.

Cullen, Christopher H., G. Joseph Ray, and Christina M. Szabo (Nov. 2013). "A comparison of quantitative nuclear magnetic resonance methods: internal, external, and electronic referencing". In: *Magn. Reson. Chem.* 51.11, pp. 705–713. ISSN: 0749-1581. DOI: 10.1002/mrc.4004. URL: https://doi.org/10.1002/mrc.4004.

Dell'Oro, Mikaela et al. (Jan. 2020). "Clinical Limitations of Photon, Proton and Carbon Ion Therapy for Pancreatic Cancer". eng. In: *Cancers* 12.1, p. 163. ISSN: 2072-6694. URL: https://pubmed.ncbi.nlm.nih.gov/31936565.

Dietrich, Olaf, Sabine Heiland, and Klaus Sartor (Mar. 2001). "Noise correction for the exact determination of apparent diffusion coefficients at low SNR". In: *Magn. Reson. Med.* 45.3, pp. 448–453. ISSN: 0740-3194. URL: https://doi.org/10.1002/1522-2594(200103)45:3<448::AID-MRM1059>3.0.CO;2-W.

Edelstein, W. A. et al. (Aug. 1986). "The intrinsic signal-to-noise ratio in NMR imaging". In: *Magn. Reson. Med.* 3.4, pp. 604–618. ISSN: 0740-3194. DOI: 10.1002/mrm.1910030413. URL: https://doi.org/10.1002/mrm.1910030413.

Fan, Linwei et al. (Dec. 2019). "Brief review of image denoising techniques". In: *Visual Computing for Industry, Biomedicine, and Art* 2. DOI: 10.1186/s42492-019-0016-7.

Ferroni, Edward L. and Katherine M. Letang (July 1989). "Calibration of air displacement pipets". In: *J. Chem. Educ.* 66.7, p. 596. ISSN: 0021-9584. DOI: 10.1021/ed066p596. URL: https://doi.org/10.1021/ed066p596.

Garabalino, Marcela A. et al. (2011). "Boron neutron capture therapy (BNCT) for the treatment of liver metastases: biodistribution studies of boron compounds in an experimental model". In: *Radiation and Environmental Biophysics* 50.1, pp. 199–207. ISSN: 1432-2099. DOI: 10.1007/s00411-010-0345-6. URL: https://doi.org/10.1007/s00411-010-0345-6.

Giraudeau, Patrick et al. (2014). "Reference and normalization methods: Essential tools for the intercomparison of NMR spectra". In: *Journal of Pharmaceutical and Biomedical Analysis* 93, pp. 3–16. ISSN: 0731-7085. DOI: 10.1016/j.jpba.2013.07.020. URL: https://www.sciencedirect.com/science/article/pii/S0731708513003269.

Glenn, G. Russell, Ali Tabesh, and Jens H. Jensen (2015). "A simple noise correction scheme for diffusional kurtosis imaging". In: *Magnetic Resonance Imaging* 33.1, pp. 124–133. ISSN: 0730-725X. DOI: 10.1016/j.mri.2014.08.028. URL: https://www.sciencedirect.com/science/article/pii/S0730725X14002598.

Goldman, M. and G. A. Webb (May 1989). "M. Goldman. Quantum description of high-resolution NMR in liquids. Oxford University Press, 1988. £35.00". In: *Magn. Reson. Chem.* 27.5, pp. 507–507. ISSN: 0749-1581. URL: https://doi.org/10.1002/mrc.1260270519.

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

Gradiz, Rui et al. (Feb. 2016). "MIA PaCa-2 and PANC-1 - pancreas ductal adenocarcinoma cell lines with neuroendocrine differentiation and somatostatin receptors". eng. In: *Scientific reports* 6, pp. 21648–21648. ISSN: 2045-2322. URL: https://pubmed.ncbi.nlm.nih.gov/26884312.

Grand Challenge Biomedical challenge database. https://grand-challenge.org/challenges/. Accessed: 2021-05-31.

Grunewald, Catrin et al. (Jan. 2017). "On the applicability of [18F]FBPA to predict L-BPA concentration after amino acid preloading in HuH-7 liver tumor model and the implication for liver boron neutron capture therapy". In: *Nuclear Medicine and Biology* 44, pp. 83–89. ISSN: 0969-8051. URL: http://www.sciencedirect.com/science/article/pii/S096980511630141X.

Gudbjartsson, H. and S. Patz (Dec. 1995). "The Rician distribution of noisy MRI data". eng. In: *Magnetic resonance in medicine* 34.8598820, pp. 910–914. ISSN: 1522-2594. DOI: 10.1002/mrm.1910340618. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2254141/.

Hanaoka, Kohei et al. (Dec. 2014). "FBPA PET in boron neutron capture therapy for cancer: prediction of 10B concentration in the tumor and normal tissue in a rat xenograft model". In: *EJNMMI Research* 4.1, p. 70. ISSN: 2191-219X. DOI: 10.1186/s13550-014-0070-2. URL: https://doi.org/10.1186/s13550-014-0070-2.

He, Kaiming et al. (2016). *Identity Mappings in Deep Residual Networks*. arXiv: 1603.05027 [cs.CV].

Holzgrabe, Ulrike (Aug. 2010). "Quantitative NMR spectroscopy in pharmaceutical applications". In: *Progress in Nuclear Magnetic Resonance Spectroscopy* 57.2,

pp. 229–240. ISSN: 0079-6565. URL: https://www.sciencedirect.com/science/article/pii/S0079656510000506.

Holzgrabe, Ulrike (2015). "Quantitative NMR Spectroscopy in Pharmaceutical R&D". In: *eMagRes*. American Cancer Society, pp. 45–56. ISBN: 9780470034590. DOI: https://doi.org/10.1002/9780470034590.emrstm1399. eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/9780470034590.emrstm1399. URL: https://onlinelibrary.wiley.com/doi/abs/10.1002/9780470034590.emrstm1399.

Im, Daniel Jiwoong et al. (2016). *Denoising Criterion for Variational Auto-Encoding Framework*. arXiv: 1511.06406 [cs.LG].

Ioffe, Sergey and Christian Szegedy (2015). "Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift". In: *Proceedings of the 32nd International Conference on Machine Learning*. Ed. by Francis Bach and David Blei. Vol. 37. Proceedings of Machine Learning Research. Lille, France: PMLR, pp. 448–456. URL: http://proceedings.mlr.press/v37/ioffe15.html.

Ishiwata, Kiichi (2019a). "4-Borono-2-18F-fluoro-l-phenylalanine PET for boron neutron capture therapy-oriented diagnosis: overview of a quarter century of research". In: *Annals of Nuclear Medicine* 33.4, pp. 223–236. ISSN: 1864-6433. DOI: 10.1007/s12149-019-01347-8. URL: https://doi.org/10.1007/s12149-019-01347-8.

— (2019b). "4-Borono-2-18F-fluoro-l-phenylalanine PET for boron neutron capture therapy-oriented diagnosis: overview of a quarter century of research". In: *Annals of Nuclear Medicine* 33.4, pp. 223–236. ISSN: 1864-6433. DOI: 10.1007/s12149-019-01347-8. URL: https://doi.org/10.1007/s12149-019-01347-8.

James, Alexander David et al. (Aug. 2017). "Comparison of <sup>19</sup>F NMR and <sup>14</sup>C Measurements for the Assessment of ADME of BYL719 (Alpelisib) in Humans". In: *Drug Metab Dispos* 45.8, p. 900. URL: http://dmd.aspetjournals.org/content/45/8/900.abstract.

Kabalka, G. W. et al. (June 2000a). "SYNTHESIS OF 4-BORONO-2-FLUOROPHENYLALANINE". In: *organic Preparations and Procedures International,* 32.3, pp. 290–293. ISSN: 0030-4948. DOI: 10.1080/00304940009355929. URL: https://doi.org/10.1080/00304940009355929.

— (June 2000b). "SYNTHESIS OF 4-BORONO-2-FLUOROPHENYLALANINE". In: *orThe New Journal for Organic Synthesis* 32.3, pp. 290–293. ISSN: 0030-4948. DOI: 10.1080/00304940009355929. URL: https://doi.org/10.1080/00304940009355929.

Kingma, Diederik P. and Jimmy Ba (2017). *Adam: A Method for Stochastic Optimization*. arXiv: 1412.6980 [cs.LG].

Koay, Cheng Guan and Peter J. Basser (2006). "Analytically exact correction scheme for signal extraction from noisy magnitude MR signals". In: *Journal of Magnetic Resonance* 179.2, pp. 317–322. ISSN: 1090-7807. DOI: 10.1016/j.jmr.2006.01.016. URL: https://www.sciencedirect.com/science/article/pii/S109078070600019X.

Kostidis, Sarantos et al. (2017). "Quantitative NMR analysis of intra- and extracellular metabolism of mammalian cells: A tutorial". In: *Analytica Chimica Acta* 980, pp. 1–24. ISSN: 0003-2670. DOI: 10.1016/j.aca.2017.05.011. URL: https://www.sciencedirect.com/science/article/pii/S0003267017306037.

Krizhevsky, Alex, Ilya Sutskever, and Geoffrey E. Hinton (2012). "ImageNet Classification with Deep Convolutional Neural Networks". In: *Proceedings of the 25th International Conference on Neural Information Processing Systems - Volume 1*. NIPS'12. Lake Tahoe, Nevada: Curran Associates Inc., 1097–1105.

Lehtinen, Jaakko et al. (2018). "Noise2Noise: Learning Image Restoration without Clean Data". In: *Proceedings of the 35th International Conference on Machine Learning*. Ed. by Jennifer Dy and Andreas Krause. Vol. 80. Proceedings of Machine Learning Research. PMLR, pp. 2965–2974. URL: http://proceedings.mlr.press/v80/lehtinen18a.html.

LIN, YU-CHUAN et al. (2012). "Macro- and Microdistributions of Boron Drug for Boron Neutron Capture Therapy in an Animal Model". In: *Anticancer Research* 32.7, pp. 2657–2664. eprint: http://ar.iiarjournals.org/content/32/7/2657.full.pdf+html. URL: http://ar.iiarjournals.org/content/32/7/2657.abstract.

Lin, Yu-Chuan et al. (2020). "Similar T/N ratio between 18F-FBPA diagnostic and BPA therapeutic dosages for boron neutron capture therapy in orthotropic tongue cancer model". In: *Annals of Nuclear Medicine* 34.1, pp. 58–64. ISSN: 1864-6433. DOI: 10.1007/s12149-019-01415-z. URL: https://doi.org/10.1007/s12149-019-01415-z.

Liu, F. et al. (2017). "Fast Realistic MRI Simulations Based on Generalized Multi-Pool Exchange Tissue Model". In: *IEEE Transactions on Medical Imaging* 36.2, pp. 527–537. ISSN: 1558-254X. DOI: 10.1109/TMI.2016.2620961.

Long, Jonathan, Evan Shelhamer, and Trevor Darrell (2015). "Fully convolutional networks for semantic segmentation". In: *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 3431–3440. DOI: 10.1109/CVPR.2015.7298965.

Lu, Le et al. (2017). *Deep Learning and Convolutional Neural Networks for Medical Image Computing: Precision Medicine, High Performance and Large-Scale Datasets*. 1st. Springer Publishing Company, Incorporated. ISBN: 3319429981.

Luo, Xian, Xinyun Gu, and Liang Li (Dec. 2018). "Development of a simple and efficient method of harvesting and lysing adherent mammalian cells for chemical isotope labeling LC-MS-based cellular metabolomics". In: *Analytical Metabolomics* 1037, pp. 97–106. ISSN: 0003-2670. URL: https://www.sciencedirect.com/science/article/pii/S0003267017313326.

Mandal, Pravat K. and Jay W. Pettegrew (2008). "Clinically Relevant Concentration Determination of Inhaled Anesthetics (Halothane, Isoflurane, Sevoflurane, and Desflurane) by 19F NMR". In: *Cell Biochemistry and Biophysics* 52.1, p. 31. ISSN: 1559-0283. DOI: 10.1007/s12013-008-9022-7. URL: https://doi.org/10.1007/s12013-008-9022-7.

Manjón, José V. and Pierrick Coupe (2018). "MRI Denoising Using Deep Learning". In: *Patch-Based Techniques in Medical Imaging*. Cham: Springer International Publishing, pp. 12–19.

Manjón, José V. et al. (2008). "MRI denoising using Non-Local Means". In: *Medical Image Analysis* 12.4, pp. 514–523. ISSN: 1361-8415. DOI: 10.1016/j.media.2008.02.004. URL: https://www.sciencedirect.com/science/article/pii/S1361841508000248.

Manjón, José V. et al. (Jan. 2010). "Adaptive non-local means denoising of MR images with spatially varying noise levels". In: *J. Magn. Reson. Imaging* 31.1, pp. 192–203. ISSN: 1053-1807. URL: https://doi.org/10.1002/jmri.22003.

Mattes, Allison O. et al. (Sept. 2016). "Application of 19F quantitative NMR to pharmaceutical analysis". In: *Concepts Magn Reson Part A* 45A.5, e21422. ISSN: 1546-6086. DOI: 10.1002/cmr.a.21422. URL: https://doi.org/10.1002/cmr.a.21422.

McRobbie, Donald W. et al. (2017). "The Basic Stuff". In: 3rd ed. Cambridge: Cambridge University Press, pp. 9–182. URL: https://www.cambridge.org/core/books/mri-from-picture-to-proton/basic-stuff/8D50FA264E1CF75C782813434015DEBC.

Moss, Raymond L. (June 2014). "Critical review, with an optimistic outlook, on Boron Neutron Capture Therapy (BNCT)". In: *15th International Congress on Neutron Capture Therapy Impact of a new radiotherapy against cancer* 88, pp. 2–11. ISSN: 0969-8043. URL: https://www.sciencedirect.com/science/article/pii/S0969804313005587.

Orieux, François, Jean-François Giovannelli, and Thomas Rodet (June 2010). "Bayesian estimation of regularization and point spread function parameters for Wiener-Hunt deconvolution". In: *Journal of the Optical Society of America. A Optics, Image Science, and Vision*, p. 1593. DOI: 10.1364/JOSAA.27.001593. URL: https://hal.archives-ouvertes.fr/hal-00674508.

Orlandi, Simonetta et al. (2020). "Synthesis and 19F NMR parameters of a perfluoro-tert-butoxy tagged L-DOPA analogue". In: *Journal of Fluorine Chemistry* 237, p. 109596. ISSN: 0022-1139. DOI: 10.1016/j.jfluchem.2020.109596. URL: https://www.sciencedirect.com/science/article/pii/S0022113920302943.

Pedregosa, F. et al. (2011). "Scikit-learn: Machine Learning in Python". In: *Journal of Machine Learning Research* 12, pp. 2825–2830.

Pierpaoli, Carlo (Apr. 2010). "Quantitative brain MRI". eng. In: *Topics in magnetic resonance imaging : TMRI* 21.2, pp. 63–63. ISSN: 0899-3459. URL: https://pubmed.ncbi.nlm.nih.gov/21613871.

Pinelli, T. (1996). "Cancer Neutron Capture Therapy". In: ed. by Y. Mishima. Springer, Boston, MA. Chap. Development of a Method to Use Boron Neutron Capture Therapy for Diffused Tumours of Liver. DOI: doi.org/10.1007/978-1-4757-9567-7_111.

Porcari, P. et al. (July 2009). "In vivo 19F MR imaging and spectroscopy for the BNCT optimization". In: *Applied Radiation and Isotopes* 67.7, Supplement, S365–S368. ISSN: 0969-8043. URL: https://www.sciencedirect.com/science/article/pii/S0969804309003273.

Porcari, Paola et al. (2006). "Multi-nuclear MRS and 19 F MRI of 19 F-labelled and 10 B-enriched p -boronophenylalanine-fructose complex to optimize boron neutron capture therapy: phantom studies at high magnetic fields". In: *Physics in Medicine & Biology* 51.12, p. 3141. URL: http://stacks.iop.org/0031-9155/51/i=12/a=010.

Porcari, Paola et al. (Nov. 2008). "In vivo19F MRI and19F MRS of19F-labelled boronophenylalanine-fructose complex on a C6 rat glioma model to optimize boron neutron capture therapy (BNCT)". In: *Physics in Medicine and Biology* 53.23, pp. 6979–6989. ISSN: 1361-6560. URL: http://dx.doi.org/10.1088/0031-9155/53/23/021.

Portu, Agustina et al. (Nov. 2015). "Inter-comparison of boron concentration measurements at INFN-University of Pavia (Italy) and CNEA (Argentina)". In: *Applied Radiation and Isotopes* 105, pp. 35–39. ISSN: 0969-8043. URL: https://www.sciencedirect.com/science/article/pii/S0969804315301032.

Postuma, Ian et al. (2016). "An improved neutron autoradiography set-up for (10)B concentration measurements in biological samples". eng. In: *Reports of practical oncology and radiotherapy : journal of Greatpoland Cancer Center in Poznan and Polish Society of Radiation Oncology* 21.2, pp. 123–128. ISSN: 2083-4640. URL: https://pubmed.ncbi.nlm.nih.gov/26933395.

Puris, Elena et al. (May 2020). "L-Type amino acid transporter 1 as a target for drug delivery". In: *Pharmaceutical Research* 37.5, p. 88. ISSN: 1573-904X. URL: https://doi.org/10.1007/s11095-020-02826-8.

Rawla, Prashanth, Tagore Sunkara, and Vinaya Gaduputi (Feb. 2019). "Epidemiology of Pancreatic Cancer: Global Trends, Etiology and Risk Factors". eng. In: *World journal of oncology* 10.1, pp. 10–27. ISSN: 1920-4531. URL: https://pubmed.ncbi.nlm.nih.gov/30834048.

Raya, José G. et al. (Jan. 2010). "T2 measurement in articular cartilage: Impact of the fitting method on accuracy and precision at low SNR". In: *Magn. Reson. Med.* 63.1, pp. 181–193. ISSN: 0740-3194. DOI: 10.1002/mrm.22178. URL: https://doi.org/10.1002/mrm.22178.

Roemer, P. B. et al. (Nov. 1990). "The NMR phased array". In: *Magn Reson Med* 16.2, pp. 192–225. ISSN: 0740-3194. DOI: 10.1002/mrm.1910160203. URL: https://doi.org/10.1002/mrm.1910160203.

Romanov, Victor et al. (Mar. 2020). "Evaluation of the total distribution volume of 18F-FBPA in normal tissues of healthy volunteers by non-compartmental kinetic modeling". In: *Annals of Nuclear Medicine* 34.3, pp. 155–162. ISSN: 1864-6433. URL: https://doi.org/10.1007/s12149-019-01427-9.

Ronneberger, Olaf, Philipp Fischer, and Thomas Brox (2015). "U-Net: Convolutional Networks for Biomedical Image Segmentation". In: *Medical Image Computing and Computer-Assisted Intervention - MICCAI 2015*. Cham: Springer International Publishing, pp. 234–241.

Rosenau, Carl Philipp et al. (July 2018). "Exposing the Origins of Irreproducibility in Fluorine NMR Spectroscopy". In: *Angew. Chem. Int. Ed.* 57.30, pp. 9528–9533. ISSN: 1433-7851. URL: https://doi.org/10.1002/anie.201802620.

Ross, Alfred et al. (Jan. 2007). "Chapter 3 - NMR Spectroscopy Techniques for Application to Metabonomics". In: *The Handbook of Metabonomics and Metabolomics*. Ed. by John C. Lindon, Jeremy K. Nicholson, and Elaine Holmes. Amsterdam: Elsevier Science B.V., pp. 55–112. URL: https://www.sciencedirect.com/science/article/pii/B9780444528414500047.

Rumelhart, David E., Geoffrey E. Hinton, and Ronald J. Williams (Oct. 1986). "Learning representations by back-propagating errors". In: *Nature* 323.6088, pp. 533–536. ISSN: 1476-4687. URL: https://doi.org/10.1038/323533a0.

Sereti, Evangelia et al. (Sept. 2018). "Patient Derived Xenografts (PDX) for personalized treatment of pancreatic cancer: emerging allies in the war on a devastating cancer?" In: *The key-role of multi-omics in predictive, preventive and personalised medicine as the medicine of the future* 188, pp. 107–118. ISSN: 1874-3919. URL: https://www.sciencedirect.com/science/article/pii/S1874391918300435.

Seshacharyulu, Parthasarathy et al. (Aug. 2017). "Biological determinants of radioresistance and their remediation in pancreatic cancer". In: *Biochimica et Biophysica Acta (BBA) - Reviews on Cancer* 1868.1, pp. 69–92. ISSN: 0304-419X. URL: https://www.sciencedirect.com/science/article/pii/S0304419X16301238.

Simonyan, Karen and Andrew Zisserman (2015). *Very Deep Convolutional Networks for Large-Scale Image Recognition*. arXiv: 1409.1556 [cs.CV].

Taylor, Alexander J. et al. (Oct. 2016). "Probe-Specific Procedure to Estimate Sensitivity and Detection Limits for 19F Magnetic Resonance Imaging". In: *PLOS ONE* 11.10, e0163704. DOI: 10.1371/journal.pone.0163704. URL: https://doi.org/10.1371/journal.pone.0163704.

Tian, Qiyuan et al. (June 2021). "Improved cortical surface reconstruction using submillimeter resolution MPRAGE by image denoising". In: *NeuroImage* 233, p. 117946. ISSN: 1053-8119. URL: https://www.sciencedirect.com/science/article/pii/S1053811921002238.

Tirotta, Ilaria et al. (Jan. 2015). "19F Magnetic Resonance Imaging (MRI): From Design of Materials to Clinical Applications". In: *Chem. Rev.* 115.2, pp. 1106–1129.

ISSN: 0009-2665. DOI: 10.1021/cr500286d. URL: https://doi.org/10.1021/cr500286d.

Uekama, Kaneto, Fumitoshi Hirayama, and Tetsumi Irie (July 1998). "Cyclodextrin Drug Carrier Systems". In: *Chem. Rev.* 98.5, pp. 2045–2076. ISSN: 0009-2665. DOI: 10.1021/cr970025p. URL: https://doi.org/10.1021/cr970025p.

Warneke, Jonas et al. (2016). "Evidence for an intrinsic binding force between dodecaborate dianions and receptors with hydrophobic binding pockets". In: *Chem. Commun.* 52.37, pp. 6300–6303. ISSN: 1359-7345. DOI: 10.1039/C6CC01233G. URL: http://dx.doi.org/10.1039/C6CC01233G.

Watanabe, Miki et al. (Apr. 2012). "Metabolic Profiling Comparison of Human Pancreatic Ductal Epithelial Cells and Three Pancreatic Cancer Cell Lines using NMR Based Metabonomics". eng. In: *Journal of molecular biomarkers & diagnosis* 3.2. ISSN: 2155-9929. URL: https://europepmc.org/articles/PMC4655885?pdf=render.

Watanabe, Tsubasa et al. (2016). "Comparison of the pharmacokinetics between L-BPA and L-FBPA using the same administration dose and protocol: a validation study for the theranostic approach using [18F]-L-FBPA positron emission tomography in boron neutron capture therapy". In: *BMC Cancer* 16.1, p. 859. ISSN: 1471-2407. DOI: 10.1186/s12885-016-2913-x. URL: https://doi.org/10.1186/s12885-016-2913-x.

Watanabe, Yoshiaki et al. (Jan. 2017). "Relationship between the uptake of 18F-borono-L-phenylalanine and L-[methyl-11C] methionine in head and neck tumors and normal organs". In: *Radiation Oncology* 12.1, p. 17. ISSN: 1748-717X. URL: https://doi.org/10.1186/s13014-017-0763-6.

Wittig, Andrea et al. (Oct. 2008). "Boron analysis and boron imaging in biological materials for Boron Neutron Capture Therapy (BNCT)". In: *Critical Reviews in Oncology/Hematology* 68.1, pp. 66–90. ISSN: 1040-8428. URL: https://www.sciencedirect.com/science/article/pii/S1040842808000565.

Zbontar, Jure et al. (2019). *fastMRI: An Open Dataset and Benchmarks for Accelerated MRI*. arXiv: 1811.08839 [cs.CV].

Zhang, K. et al. (2017). "Beyond a Gaussian Denoiser: Residual Learning of Deep CNN for Image Denoising". In: *IEEE Transactions on Image Processing* 26.7, pp. 3142–3155. ISSN: 1941-0042. DOI: 10.1109/TIP.2017.2662206.

Zhou and Chellappa (24-2). "Computation of optical flow using a neural network". In: *IEEE 1988 International Conference on Neural Networks*, 71–78 vol.2.

Ziessow, D. (Dec. 1988). "R. R. Ernst, G. Bodenhausen u. A. Wokaun: Principles of Nuclear Magnetic Resonance in One and Two Dimensions, Clarendon Press, Oxford 1987. 610 Seiten". In: *Berichte der Bunsengesellschaft für physikalische Chemie* 92.12, pp. 1563–1565. ISSN: 0005-9021. URL: https://doi.org/10.1002/bbpc.198800370.