

# **Advanced deep neural networks for MRI image reconstruction from highly undersampled data in challenging acquisition settings**

PhD defense

Zaccharie Ramzi

Parietal team, Inria Saclay  
NeuroSpin and Cosmostat, CEA Saclay



# MRI is slow

---

Typical MRI (Magnetic Resonance Imaging) scan duration: 15 minutes (up to 90 minutes). Hence:

- discomfort & accessibility issues;
- reduced patient throughput;
- increased motion.

# Our objective: accelerate MRI scans

---

## 1. Introduction to MRI

- 1.1 Importance of MRI
- 1.2 Physics of MRI
- 1.3 Acceleration in MRI

## 2. Compressed Sensing

- 2.1 Linear Inverse Problems
- 2.2 Recovery Algorithms

## 3. Deep Learning

- 3.1 The power of Deep Learning
- 3.2 Requirements for Deep Learning

## 4. Deep Learning for MRI reconstruction

- 4.1 Simple models
- 4.2 Unrolled models
- 4.3 New unrolled models

## 5. Going even deeper

- 5.1 Implicit models
- 5.2 SHINE

## 6. Conclusion & Future works

# Magnetic Resonance Imaging (MRI)

# What does an MRI look like?

---

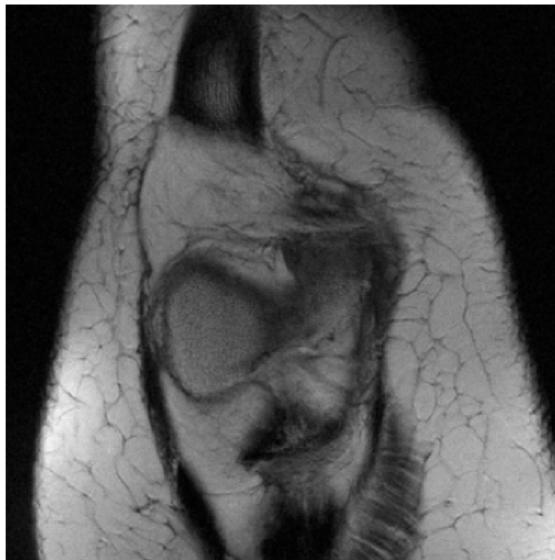


Figure: **Example of an MR image:** MR image of the knee taken from the fastMRI dataset.<sup>1</sup>

---

<sup>1</sup>J. Zbontar et al. (2018). *fastMRI: An Open Dataset and Benchmarks for Accelerated MRI*. Tech. rep.

# Importance of MRI - 1

Based on a rough extrapolation of the following figure, there is a 99.9% chance that you will get an MRI in your life in France.

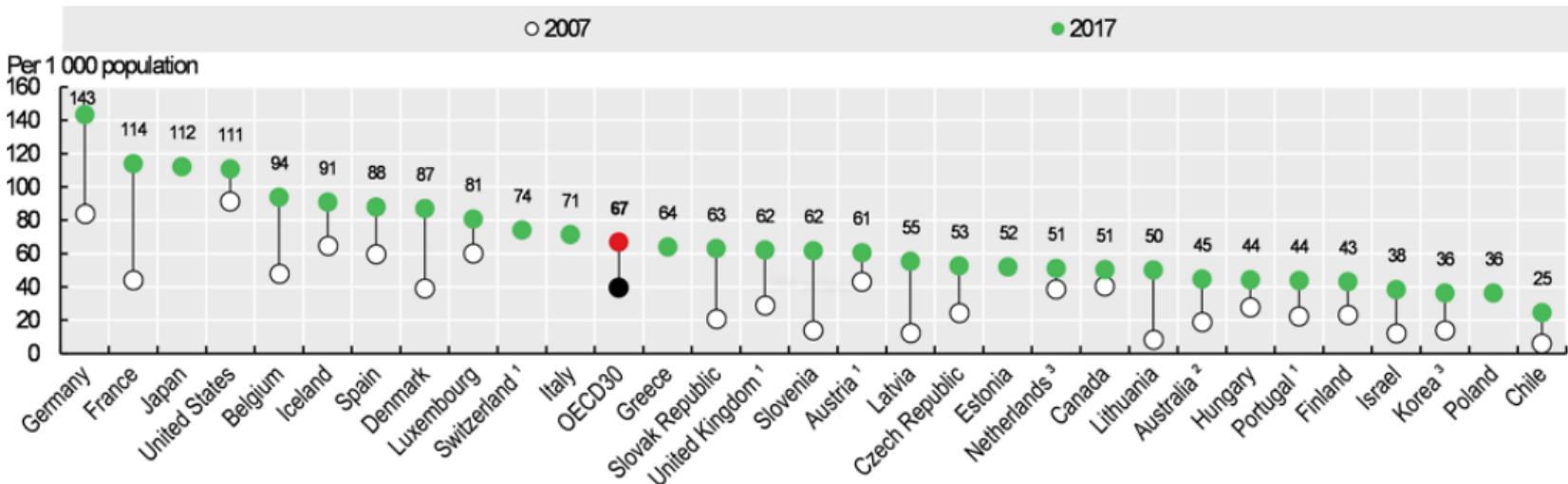
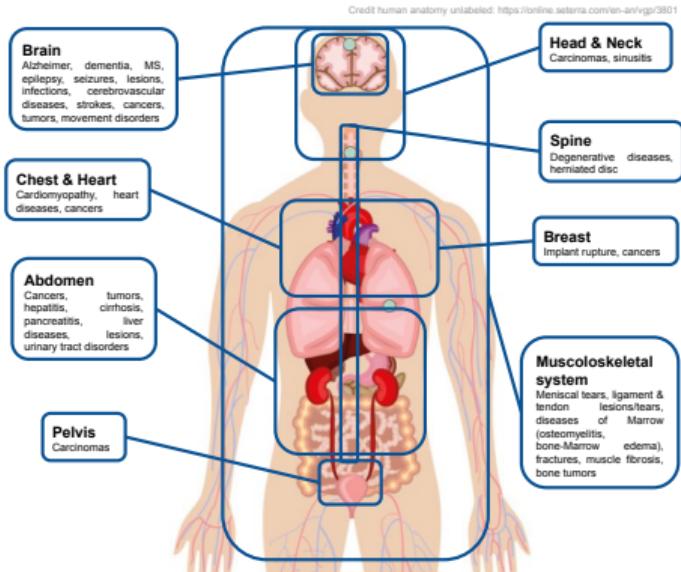


Figure: **Number of MRI scans per year per 1000 population:** figure courtesy of *Health at a Glance 2019: OECD Indicators - Medical technologies* (2019).

# Importance of MRI - 2



**Figure: What can we diagnose with MRI?** This illustration provides a non-exhaustive list of all the diagnoses that can be carried out with MRI. All the information was compiled from the works of Reimer et al. (2010) and Runge et al. (2019).

# Physics of MRI - 1

---

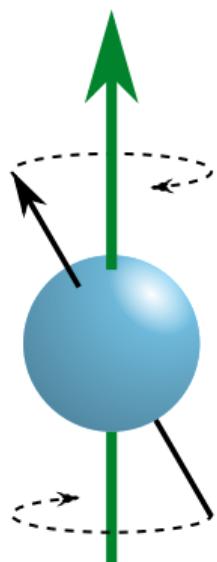


Figure: **Illustration of the precession of a spin in a magnetic field:** the green arrow represents the  $B_0$  magnetic field, while the black arrow represents the magnetic moment of the particle. Illustration courtesy of *Larmor precession Wikipedia page* (2012).

# Physics of MRI - 1

---

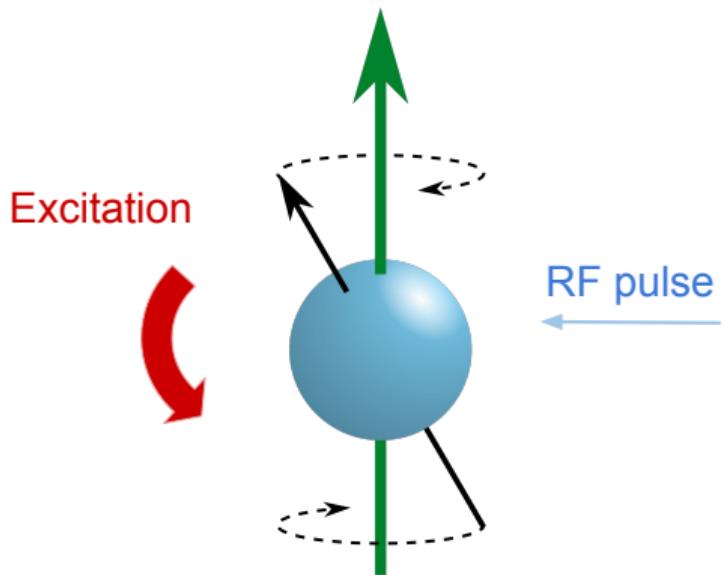


Figure: **Illustration of the excitation phenomenon:** the blue arrow represents an incoming RF (Radio Frequency) pulse. Illustration courtesy of *Larmor precession Wikipedia page* (2012).

# Physics of MRI - 1

---

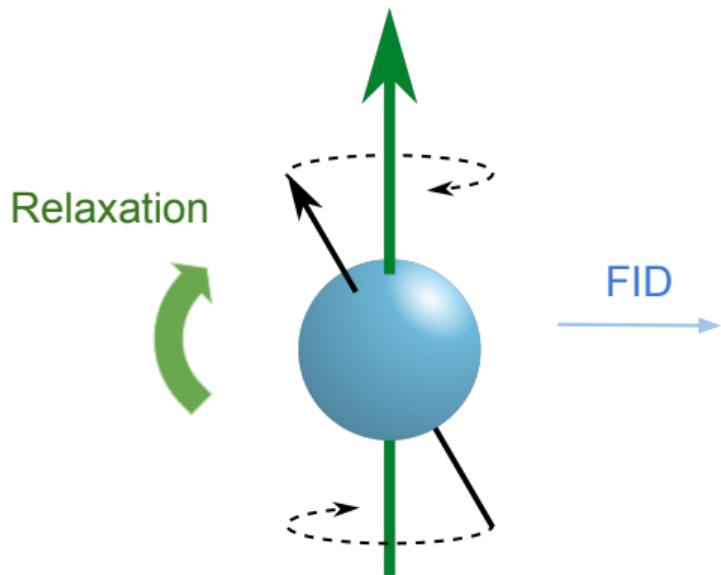


Figure: **Illustration of the relaxation phenomenon:** the blue arrow represents an outgoing FID (Free Induction Decay) pulse. Illustration courtesy of *Larmor precession Wikipedia page* (2012).

# Physics of MRI - 2

---

However, the FID carries global information.

# Physics of MRI - 2

---

However, the FID carries global information.

Using magnetic **gradients** enables changing a bit the magnetic field spatially, and therefore changing the global information depending on local factors.

# Physics of MRI - 2

---

However, the FID carries global information.

Using magnetic **gradients** enables changing a bit the magnetic field spatially, and therefore changing the global information depending on local factors.

This allows us to receive a temporal signal of the form:

$$S_{tr}(t) \propto \omega_0 \int_{V_s} B_{tr} M_{tr}(t, \mathbf{r}) e^{-i\gamma \mathbf{r} \cdot \int_0^t \mathbf{G}(\tau) d\tau} d\mathbf{r}$$

# Physics of MRI - 2

---

However, the FID carries global information.

Using magnetic **gradients** enables changing a bit the magnetic field spatially, and therefore changing the global information depending on local factors.

This allows us to receive a temporal signal of the form:

$$S_{tr}(t) \propto \omega_0 \int_{V_s} B_{tr} M_{tr}(t, \mathbf{r}) e^{-i\gamma \mathbf{r} \cdot \int_0^t \mathbf{G}(\tau) d\tau} d\mathbf{r}$$

Recorded MR signal

# Physics of MRI - 2

---

However, the FID carries global information.

Using magnetic **gradients** enables changing a bit the magnetic field spatially, and therefore changing the global information depending on local factors.

This allows us to receive a temporal signal of the form:

$$S_{tr}(t) \propto \omega_0 \int_{V_s} B_{tr} M_{tr}(t, \mathbf{r}) e^{-i\gamma \mathbf{r} \cdot \int_0^t \mathbf{G}(\tau) d\tau} d\mathbf{r}$$

Recorded MR signal

Magnetic field in each location  $\mathbf{r}$ ,  
proportional to the spin density  $\rho(\mathbf{r})$

# Physics of MRI - 2

However, the FID carries global information.

Using magnetic **gradients** enables changing a bit the magnetic field spatially, and therefore changing the global information depending on local factors.

This allows us to receive a temporal signal of the form:

$$S_{tr}(t) \propto \omega_0 \int_{V_s} B_{tr} M_{tr}(t, \mathbf{r}) e^{-i\gamma \mathbf{r} \cdot \int_0^t \mathbf{G}(\tau) d\tau} d\mathbf{r}$$

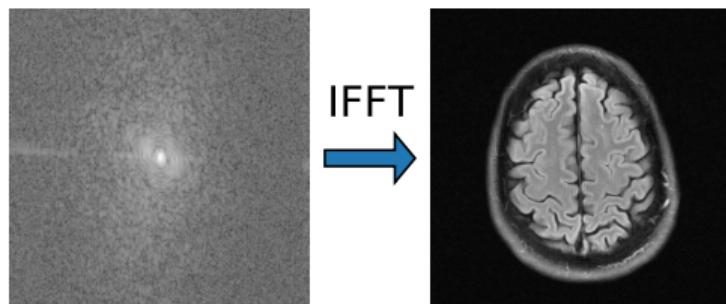
Recorded MR signal

Magnetic field in each location  $\mathbf{r}$ ,  
proportional to the spin density  $\rho(\mathbf{r})$

Temporal gradients,  
controlled by the operator

# Physics of MRI - 3

The **k-space** vector,  $k(t) = \frac{\gamma}{2\pi} \int_0^t G(\tau) d\tau$ , defines how we traverse the Fourier space of the anatomical image.



**Figure: Example of a k-space with its corresponding anatomical image:** The raw data is from the fastMRI dataset. The k-space is in log-scale and only the magnitude of the 2 images are represented. We selected only a single coil from the 16 coils available for illustrative purposes.

# Physics of MRI - 4

---

## Recap

MRI relies on the nuclear resonance phenomenon. This enables us to sample the Fourier space of the anatomical object of interest.

# Physics of MRI - 4

---

## Recap

MRI relies on the nuclear resonance phenomenon. This enables us to sample the Fourier space of the anatomical object of interest.

MRI is slow, because the **relaxation** is slow!

# Where is there room for acceleration?

---

**Redundancy**, otherwise called **sparsity, symmetry, structure or a priori information**, is the core concept that will help us accelerate MRI.

# Where is there room for acceleration?

---

**Redundancy**, otherwise called **sparsity, symmetry, structure or a priori information**, is the core concept that will help us accelerate MRI.

Here is an illustrative example:



Figure: **The inpainting problem.** Even without access to all the pixel values directly, we can infer them, because the information in the image is redundant.

# Where is there room for acceleration?

---

**Redundancy**, otherwise called **sparsity, symmetry, structure or a priori information**, is the core concept that will help us accelerate MRI.

Here is an illustrative example:



Figure: **The inpainting problem.** Even without access to all the pixel values directly, we can infer them, because the information in the image is redundant.

# Where is there room for acceleration?

---

**Redundancy**, otherwise called **sparsity, symmetry, structure or a priori information**, is the core concept that will help us accelerate MRI.

Is there a similar thing in MRI?

Yes! The anatomical image is real-valued so its Fourier Transform (FT) has a conjugate symmetry. Using this redundancy to sample less points in the k-space (i.e. using the relaxation fewer times) is a technique called **Partial Fourier**.

But in practice it is still needed to sample 6/8 of the Fourier space (acceleration of 1.3).<sup>a</sup>

---

<sup>a</sup>e-MRI courses by IMAIOS (2008). <https://www.imaios.com/en/e-Courses/e-MRI>. Accessed: 2021-10-08.

# Parallel imaging

---

We can build more redundancy in the measuring system by using **more antennas (called coils)** to measure the magnetic signal.

This technique is called **Parallel Imaging (PI)**. A reconstruction algorithm is now needed to handle the multi-coil undersampled data. **SENSE**<sup>2</sup> and **GRAPPA**<sup>3</sup> are such algorithms.

---

<sup>2</sup>K. P. Pruessmann et al. (Nov. 1999). "SENSE: Sensitivity encoding for fast MRI". In: *Magnetic Resonance in Medicine* 42.5, pp. 952–962.

<sup>3</sup>M. A. Griswold et al. (June 2002). "Generalized Autocalibrating Partially Parallel Acquisitions (GRAPPA)". In: *Magnetic Resonance in Medicine* 47.6, pp. 1202–1210.

# The example of GRAPPA

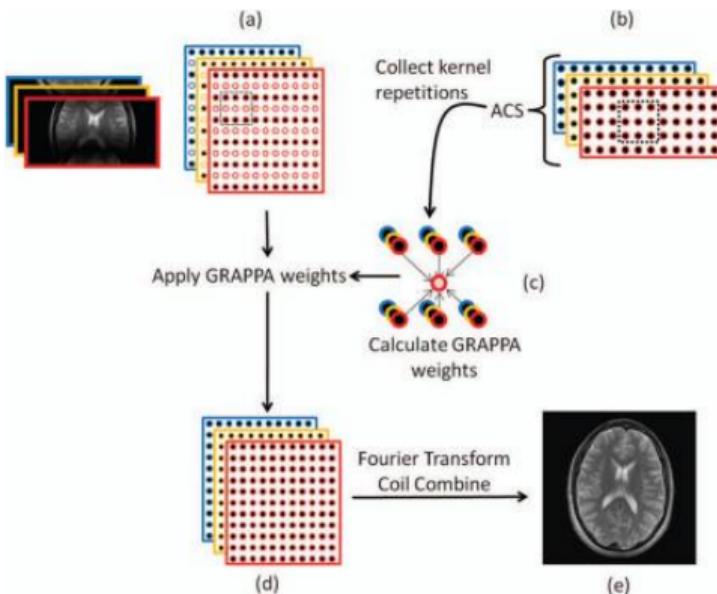


Figure: **GRAPPA illustration**. Image courtesy of Deshpande et al. (2012).

# Limits of Parallel Imaging

---

# Compressed Sensing

# Another look at redundancy: the prior point of view

Redundancy is not always strict: we may only have a strong correlation between 2 structures of the image.

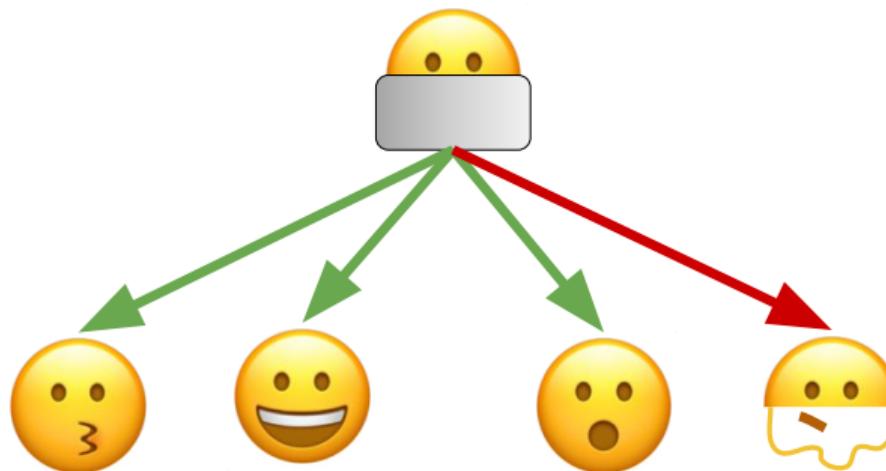


Figure: **A smiley example to a priori knowledge:** Even if we do not have access to the whole image, we still know which images are more *likely* to correspond to it.

# Linear Inverse Problems

---

To leverage this type of redundancy, we introduce the concept of **Linear Inverse Problems**:

$$\mathbf{A} \mathbf{x} = \mathbf{y}$$

# Linear Inverse Problems

---

To leverage this type of redundancy, we introduce the concept of **Linear Inverse Problems**:

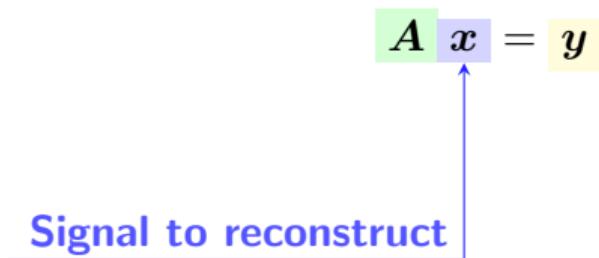
$$A \ x = y$$


# Linear Inverse Problems

---

To leverage this type of redundancy, we introduce the concept of **Linear Inverse Problems**:

$$\mathbf{A} \mathbf{x} = \mathbf{y}$$



Signal to reconstruct

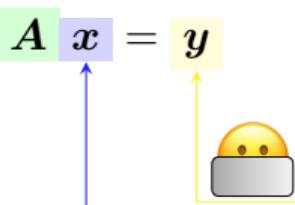
# Linear Inverse Problems

---

To leverage this type of redundancy, we introduce the concept of **Linear Inverse Problems**:

$$\mathbf{A} \mathbf{x} = \mathbf{y}$$

Signal to reconstruct

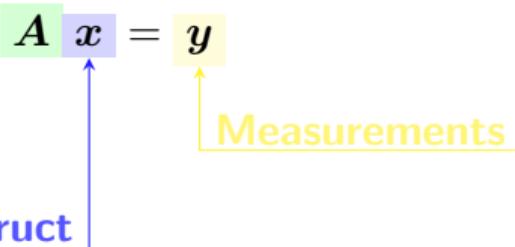


# Linear Inverse Problems

---

To leverage this type of redundancy, we introduce the concept of **Linear Inverse Problems**:

$$\mathbf{A} \mathbf{x} = \mathbf{y}$$



Signal to reconstruct

Measurements

# Linear Inverse Problems

---

To leverage this type of redundancy, we introduce the concept of **Linear Inverse Problems**:

$$\mathbf{A} \mathbf{x} = \mathbf{y}$$

The diagram shows a mathematical equation  $\mathbf{A} \mathbf{x} = \mathbf{y}$ . Above the equation, there is a gray rectangular block. A green arrow points from this block up to the matrix  $\mathbf{A}$ , which is colored green. A blue arrow points from the same gray block up to the vector  $\mathbf{x}$ , which is colored blue. To the right of the equation, a yellow arrow points from the word "Measurements" down to the vector  $\mathbf{y}$ , which is colored yellow. Below the equation, a blue box contains the text "Signal to reconstruct".

# Linear Inverse Problems

---

To leverage this type of redundancy, we introduce the concept of **Linear Inverse Problems**:

$$\mathbf{A} \mathbf{x} = \mathbf{y}$$

The diagram shows the equation  $\mathbf{A} \mathbf{x} = \mathbf{y}$ . Above the equation, there are three colored boxes: a green box containing  $\mathbf{A}$ , a purple box containing  $\mathbf{x}$ , and a yellow box containing  $\mathbf{y}$ . Below the equation, three labels with arrows point to these boxes: "Measurement operator" points to the  $\mathbf{A}$  box, "Signal to reconstruct" points to the  $\mathbf{x}$  box, and "Measurements" points to the  $\mathbf{y}$  box.

# Linear Inverse Problems

---

To leverage this type of redundancy, we introduce the concept of **Linear Inverse Problems**:

$$\mathbf{A} \mathbf{x} = \mathbf{y}$$

Diagram illustrating the components of the linear inverse problem equation:

- Measurement operator** ( $\mathbf{A}$ ): Represented by a green box containing the letter  $A$ .
- Signal to reconstruct** ( $\mathbf{x}$ ): Represented by a blue box containing the letter  $x$ .
- Measurements** ( $\mathbf{y}$ ): Represented by a yellow box containing the letter  $y$ .

Arrows point from the labels to their corresponding colored boxes.

Problems arise when  $\text{Ker } \mathbf{A} \neq \{0\}$ , i.e. when there are multiple solutions to this equation.

# Linear Inverse Problems

---

To leverage this type of redundancy, we introduce the concept of **Linear Inverse Problems**:

$$\mathbf{A} \mathbf{x} = \mathbf{y}$$

Measurement operator      Signal to reconstruct      Measurements

Problems arise when  $\text{Ker } \mathbf{A} \neq \{0\}$ , i.e. when there are multiple solutions to this equation.

In order to select one of these solutions, we need to use a priori knowledge.

# Sparsity and Inverse Problems

## Definition (Sparsity)

A vector  $\mathbf{x} \in \mathbb{C}^n$  is called  $s$ -sparse if it contains at most  $s$  non-zero entries.

## Lemma (Optimization reformulation of sparse vector recovery (Foucart et al., 2013))

For a given sparsity  $s$ , and  $s$ -sparse vector  $\mathbf{x}$ :

- (a) The vector  $\mathbf{x}$  is the unique  $s$ -sparse solution of  $\mathbf{A}\mathbf{x} = \mathbf{y}$ , that is  
$$\{\mathbf{z} \in \mathbb{C}^n : \mathbf{A}\mathbf{z} = \mathbf{A}\mathbf{x}, \|\mathbf{z}\|_0 \leq s\} = \{\mathbf{x}\}$$
- (b) The vector  $\mathbf{x}$  can be reconstructed as the unique solution of:

$$\min_{\mathbf{z} \in \mathbb{C}^n} \|\mathbf{z}\|_0 \quad \text{subject to} \quad \mathbf{A}\mathbf{z} = \mathbf{y}$$

# Recovery guarantees

---

Theorem ((Foucart et al., 2013, Theorem 2.13))

*The following properties are equivalent:*

- (a) *Every  $s$ -sparse vector  $\mathbf{x} \in \mathbb{C}^n$  is the unique  $s$ -sparse solution of  $\mathbf{A}\mathbf{z} = \mathbf{A}\mathbf{x}$ , that is, if  $\mathbf{A}\mathbf{x} = \mathbf{A}\mathbf{z}$  and both  $\mathbf{x}$  and  $\mathbf{z}$  are  $s$ -sparse, then  $\mathbf{x} = \mathbf{z}$ .*
- (b) *The null space  $\text{Ker}(\mathbf{A})$  does not contain any  $2s$ -sparse vector other than the zero.*
- (c) *Every set of  $2s$  columns of  $\mathbf{A}$  is linearly independent.*

# Application to MRI

---

MR images themselves cannot be represented as sparse vectors directly, we need a way to express them as such. Lustig et al. (2007) did that by using the fact that MR images can be represented sparsely in a **wavelet** basis.

The Inverse Problem becomes:

$$(\mathbf{I}_L \otimes \mathcal{F}_{\Omega}) \mathbf{x} = \mathbf{y}$$

# Application to MRI

---

MR images themselves cannot be represented as sparse vectors directly, we need a way to express them as such. Lustig et al. (2007) did that by using the fact that MR images can be represented sparsely in a **wavelet** basis.

The Inverse Problem becomes:

$$(\mathbf{I}_L \otimes \mathcal{F}_{\Omega}) \mathbf{x} = \mathbf{y}$$

 2D or 3D MR image

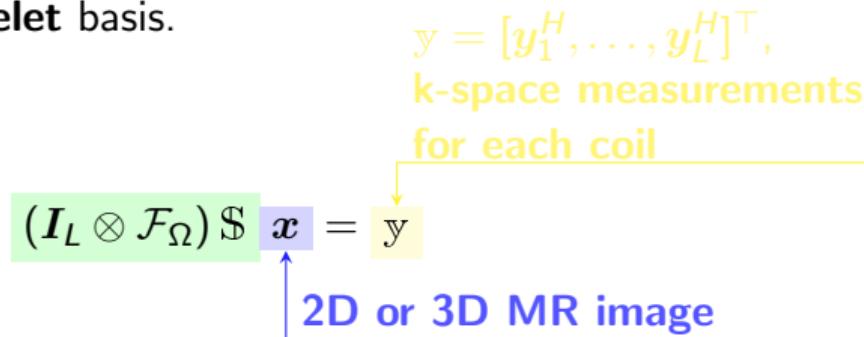
# Application to MRI

---

MR images themselves cannot be represented as sparse vectors directly, we need a way to express them as such. Lustig et al. (2007) did that by using the fact that MR images can be represented sparsely in a **wavelet** basis.

The Inverse Problem becomes:

$$(\mathbf{I}_L \otimes \mathcal{F}_{\Omega}) \$ \mathbf{x} = \mathbf{y}$$



$\mathbf{y} = [\mathbf{y}_1^H, \dots, \mathbf{y}_L^H]^\top,$   
**k-space measurements  
for each coil**

**2D or 3D MR image**

# Application to MRI

MR images themselves cannot be represented as sparse vectors directly, we need a way to express them as such. Lustig et al. (2007) did that by using the fact that MR images can be represented sparsely in a **wavelet** basis.

The Inverse Problem becomes:

$$(\mathbf{I}_L \otimes \mathcal{F}_{\Omega}) \$ \mathbf{x} = \mathbf{y}$$

$\mathbf{y} = [y_1^H, \dots, y_L^H]^\top,$   
**k-space measurements  
for each coil**

**2D or 3D MR image**

$\mathcal{F}_{\Omega}$  : FT on the  $\Omega$  set;

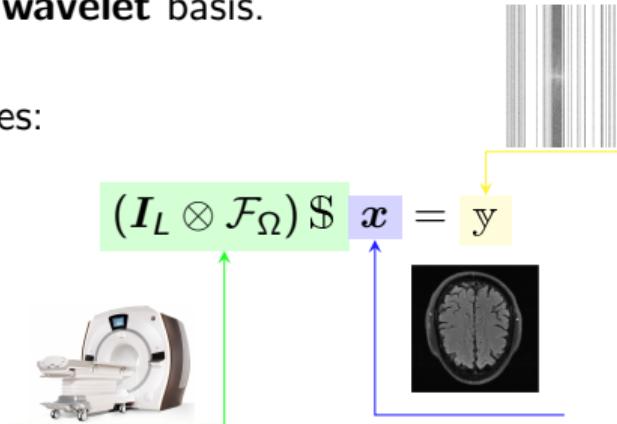
$\$ = [S_1^H, \dots, S_L^H]^\top$ : the sensitivity maps per coil

# Application to MRI

---

MR images themselves cannot be represented as sparse vectors directly, we need a way to express them as such. Lustig et al. (2007) did that by using the fact that MR images can be represented sparsely in a **wavelet** basis.

The Inverse Problem becomes:



# Relaxation

---

Now we have an optimization problem to solve, but can we?

# Relaxation

---

Now we have an optimization problem to solve, but can we? No; we need to relax it using the basis pursuit:

$$\min_{\mathbf{x} \in \mathbb{C}^n} \|\mathbf{x}\|_1 \quad \text{subject to} \quad \mathbf{A}\mathbf{x} = \mathbf{y}$$

# Relaxation

---

Now we have an optimization problem to solve, but can we? No; we need to relax it using the basis pursuit:

$$\min_{\mathbf{x} \in \mathbb{C}^n} \|\mathbf{x}\|_1 \quad \text{subject to} \quad \mathbf{A}\mathbf{x} = \mathbf{y}$$

For this problem to have the same solutions as the non-relaxed one, we need coherence-based constraints on the measurement operator  $\mathbf{A}$ .

# The canonical MRI reconstruction problem

---

We introduce the notion of a **sparsity** basis  $\psi$  (typically wavelets) and the fact that the measurements can be noisy to obtain the canonical MRI reconstruction problem:

$$\min_{\mathbf{x} \in \mathbb{C}^n} \underbrace{\|\mathcal{A}\mathbf{x} - \mathbf{y}\|_2^2}_{= (\mathcal{I}_L \otimes \mathcal{F}_\Omega) \$} + \underbrace{\lambda \|\psi \mathbf{x}\|_1}_{\text{Regularization hyperparameter}}$$

Data consistency      Regularization term

# ISTA

---

The Iterative Shrinkage-Thresholding Algorithm (ISTA) can be used to solve the canonical MRI reconstruction problem:

$$\mathbf{x}_{n+1} = \mathbf{x}_n - \epsilon_n \mathcal{A}^H (\mathcal{A}\mathbf{x}_n - \mathbf{y})$$

$$\mathbf{x}_{n+1} = \text{prox}_{\epsilon_n \mathcal{R}} (\mathbf{x}_{n+1})$$

Proximity operator

$$= \|\psi \cdot\|_1$$

# Limitations of classical recovery algorithms

---

Additional acceleration factor on top of PI: 1.5.

# Limitations of classical recovery algorithms

---

Additional acceleration factor on top of PI: 1.5. The prior knowledge expressed by the wavelet basis (or other basis) is limited: handcrafted and linear.

# Compressed Sensing

---

## Recap

MRI is slow because of **relaxation**.

# Compressed Sensing

---

## Recap

MRI is slow because of **relaxation**.

We can use **redundancy** in many forms to reduce the amount of samples we need in the Fourier space, and therefore the number of relaxations.

# Compressed Sensing

---

## Recap

MRI is slow because of **relaxation**.

We can use **redundancy** in many forms to reduce the amount of samples we need in the Fourier space, and therefore the number of relaxations.

But we are limited by simple forms of redundancy.

# Deep Learning

# The power of Deep Learning

---

We want to learn a complicated function that tells us whether a complex-valued vector is an MR image or not.

# The power of Deep Learning

---

We want to learn a complicated function that tells us whether a complex-valued vector is an MR image or not.

Deep Learning (DL) has been used to build such functions:

$$f_{\theta} \left( \begin{array}{c} \text{Image of a dog} \end{array} \right) = \text{"DOG"}$$

# Formalism - 1

---

The classical framework for DL is supervised learning:

$$\arg \min_{\theta \in \Theta} \sum_{(x_i, y_i) \in \mathcal{D}} \mathcal{L}(f_{\theta}(x_i), y_i, \theta)$$

# Formalism - 1

---

The classical framework for DL is supervised learning:

$$\arg \min_{\theta \in \Theta} \sum_{(x_i, y_i) \in \mathcal{D}} \mathcal{L}(f_{\theta}(x_i), y_i, \theta)$$


A blue bracket labeled "input" points from the left side of the equation to the variable  $x_i$  in the term  $f_{\theta}(x_i)$ .

# Formalism - 1

---

The classical framework for DL is supervised learning:

$$\arg \min_{\theta \in \Theta} \sum_{(x_i, y_i) \in \mathcal{D}} \mathcal{L}(f_{\theta}(x_i), y_i, \theta)$$

A diagram illustrating the components of the supervised learning equation. A blue bracket labeled "input" points to the term  $f_{\theta}(x_i)$ . A red bracket labeled "label" points to the term  $y_i$ .

# Formalism - 1

---

The classical framework for DL is supervised learning:

$$\arg \min_{\theta \in \Theta} \sum_{(x_i, y_i) \in \mathcal{D}} \mathcal{L}(f_{\theta}(x_i), y_i, \theta)$$

The diagram illustrates the supervised learning framework. It shows the flow of data through a neural network. An input arrow points from the input  $x_i$  to the function  $f_{\theta}$ . A label arrow points from the label  $y_i$  to the loss function  $\mathcal{L}$ . The neural network is represented by the function  $f_{\theta}$ .

# Formalism - 1

---

The classical framework for DL is supervised learning:

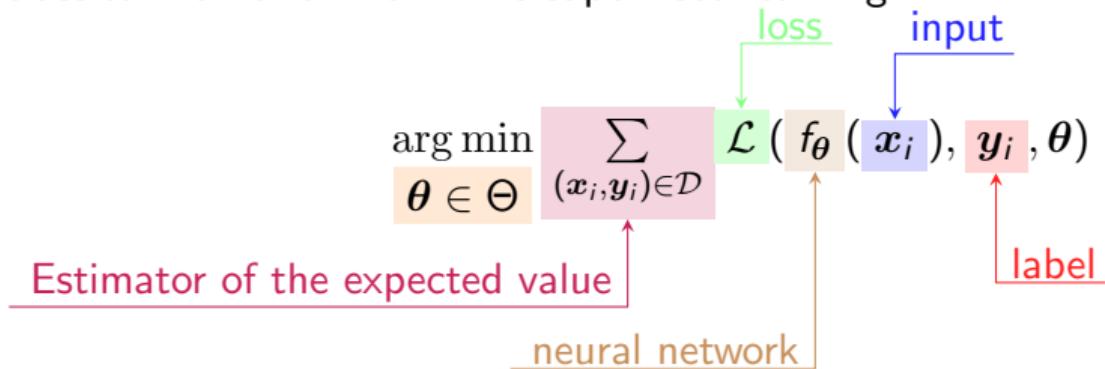
$$\arg \min_{\theta \in \Theta} \sum_{(x_i, y_i) \in \mathcal{D}} \mathcal{L}(f_{\theta}(x_i), y_i, \theta)$$

loss  
input  
label  
neural network

# Formalism - 1

---

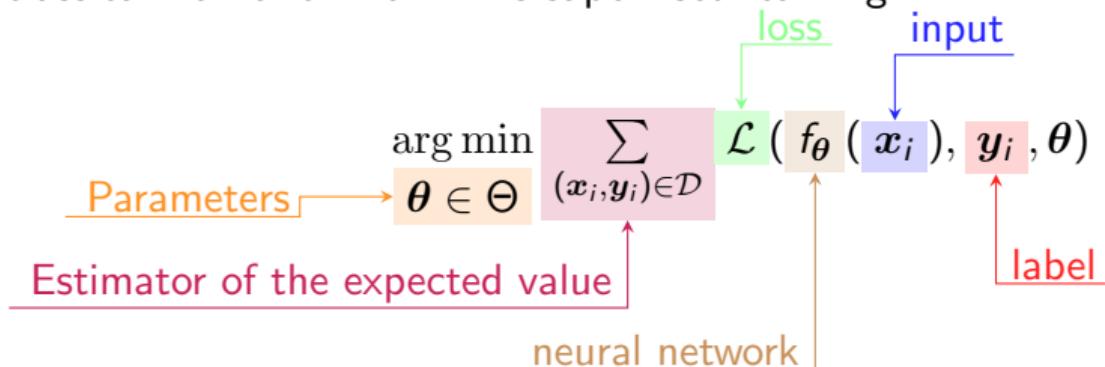
The classical framework for DL is supervised learning:



# Formalism - 1

---

The classical framework for DL is supervised learning:



# Formalism - 2

---

To solve the previous equation we will use two main tools:

1. Stochastic Gradient Descent (SGD) ;

## Definition

An algorithm to solve the previous optimization problem based on first order derivatives.

# Formalism - 2

---

To solve the previous equation we will use two main tools:

1. Stochastic Gradient Descent (SGD);
2. Chain rule .

## Definition

A property allowing us to compute easily derivatives of compound functions.

# Requirements for Deep Learning

---

What does it take to use DL in a problem?

- data;

# Requirements for Deep Learning

---

What does it take to use DL in a problem?

- data;
- compute & memory;

# Requirements for Deep Learning

---

What does it take to use DL in a problem?

- data;
- compute & memory;
- development framework;

# Requirements for Deep Learning

---

What does it take to use DL in a problem?

- data;
- compute & memory;
- development framework;
- accepting that it's "black-box".

# Introduction Recap

---

## Recap

MRI is slow because of **relaxation**.

# Introduction Recap

---

## Recap

MRI is slow because of **relaxation**.

If we want to do fewer relaxations, we need to exploit some **redundancy** in MR images.

# Introduction Recap

---

## Recap

MRI is slow because of **relaxation**.

If we want to do fewer relaxations, we need to exploit some **redundancy** in MR images.  
But this redundancy is not easy to express with handcrafted linear functions.

# Introduction Recap

---

## Recap

MRI is slow because of **relaxation**.

If we want to do fewer relaxations, we need to exploit some **redundancy** in MR images.  
But this redundancy is not easy to express with handcrafted linear functions.

This is why we want to use **Deep Learning** which enables the calibration of complicated function.

# Deep Learning for MRI reconstruction

# Model agnostic learning

---

The first use of DL for MRI reconstruction is to actually throw away most of what we just saw:<sup>4</sup>

$$f_{\theta}(\text{y}) = \text{x}$$

---

<sup>4</sup>B. Zhu et al. (Mar. 2018). "Image reconstruction by domain-transform manifold learning". In: *Nature* 555.7697, pp. 487–492.

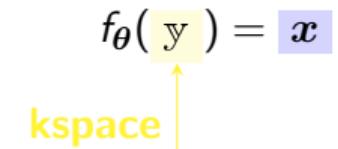
# Model agnostic learning

---

The first use of DL for MRI reconstruction is to actually throw away most of what we just saw:<sup>4</sup>

$$f_{\theta}(\text{y}) = \text{x}$$

kspace



---

<sup>4</sup>B. Zhu et al. (Mar. 2018). "Image reconstruction by domain-transform manifold learning". In: *Nature* 555.7697, pp. 487–492.

# Model agnostic learning

---

The first use of DL for MRI reconstruction is to actually throw away most of what we just saw:<sup>4</sup>

$$f_{\theta}(\text{y}) = \text{x}$$

A diagram illustrating a function  $f_{\theta}$ . On the left, the word "kspace" is written in yellow, with a blue arrow pointing upwards to a yellow rectangular box containing the variable "y". On the right, the word "image" is written in blue, with a blue arrow pointing upwards to a blue rectangular box containing the variable "x". The function itself is  $f_{\theta}(y) = x$ .

---

<sup>4</sup>B. Zhu et al. (Mar. 2018). “Image reconstruction by domain-transform manifold learning”. In: *Nature* 555.7697, pp. 487–492.

# Single domain learning

---

We actually have access to the backward operator  $\mathcal{A}^H$ , the inverse FT composed with the sensitivity maps.

Let's use that to build a more informed model:

$$\mathcal{A}^H f_\theta(\textcolor{yellow}{y}) = \textcolor{blue}{x}$$

# Single domain learning

---

We actually have access to the backward operator  $\mathcal{A}^H$ , the inverse FT composed with the sensitivity maps.

Let's use that to build a more informed model:

$$f_{\theta}(\mathcal{A}^H \textcolor{yellow}{y}) = \textcolor{blue}{x}$$

# Single domain learning

---

We actually have access to the backward operator  $\mathcal{A}^H$ , the inverse FT composed with the sensitivity maps.

Let's use that to build a more informed model:

$$f_{\theta}(\mathcal{A}^H \textcolor{yellow}{y}) = \textcolor{blue}{x}$$

  
aliased image

# Single domain learning

---

We actually have access to the backward operator  $\mathcal{A}^H$ , the inverse FT composed with the sensitivity maps.

Let's use that to build a more informed model:

$$f_{\theta}(\mathcal{A}^H \textcolor{yellow}{y}) = \textcolor{blue}{x}$$



# Unrolled models - 1

---

We can mix the 2 single domain approaches, using the principled **optimization algorithm unrolling** method.<sup>5</sup> Let's look at a graph representation of the ISTA algorithm:

$$\mathbf{x}_{n+1} = \mathbf{x}_n - \epsilon_n \mathcal{A}^H (\mathcal{A}\mathbf{x}_n - \mathbf{y})$$

$$\mathbf{x}_{n+1} = \text{prox}_{\epsilon_n \mathcal{R}} (\mathbf{x}_{n+1})$$

---

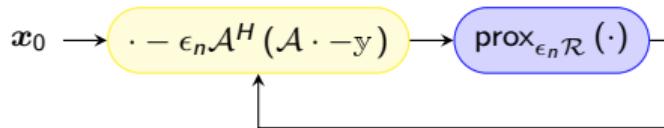
<sup>5</sup>K. Gregor et al. (2010). "Learning fast approximations of sparse coding". In: *ICML 2010 - Proceedings, 27th International Conference on Machine Learning*, pp. 399–406.

# Unrolled models - 1

We can mix the 2 single domain approaches, using the principled **optimization algorithm unrolling** method.<sup>5</sup> Let's look at a graph representation of the ISTA algorithm:

$$x_{n+1} = x_n - \epsilon_n \mathcal{A}^H (\mathcal{A}x_n - y)$$

$$x_{n+1} = \text{prox}_{\epsilon_n \mathcal{R}} (x_n)$$



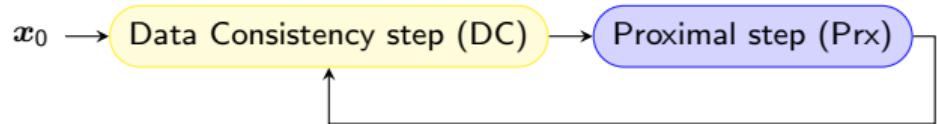
<sup>5</sup>K. Gregor et al. (2010). "Learning fast approximations of sparse coding". In: *ICML 2010 - Proceedings, 27th International Conference on Machine Learning*, pp. 399–406.

# Unrolled models - 1

We can mix the 2 single domain approaches, using the principled **optimization algorithm unrolling** method.<sup>5</sup> Let's look at a graph representation of the ISTA algorithm:

$$x_{n+1} = x_n - \epsilon_n \mathcal{A}^H (\mathcal{A}x_n - y)$$

$$x_{n+1} = \text{prox}_{\epsilon_n \mathcal{R}} (x_{n+1})$$



<sup>5</sup>K. Gregor et al. (2010). “Learning fast approximations of sparse coding”. In: *ICML 2010 - Proceedings, 27th International Conference on Machine Learning*, pp. 399–406.

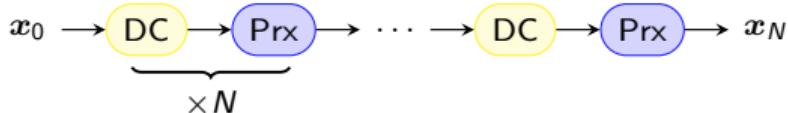
# Unrolled models - 1

---

We can mix the 2 single domain approaches, using the principled **optimization algorithm unrolling** method.<sup>5</sup> Let's look at a graph representation of the ISTA algorithm:

$$x_{n+1} = x_n - \epsilon_n \mathcal{A}^H (\mathcal{A}x_n - \mathbf{y})$$

$$x_{n+1} = \text{prox}_{\epsilon_n \mathcal{R}} (x_{n+1})$$



---

<sup>5</sup>K. Gregor et al. (2010). “Learning fast approximations of sparse coding”. In: *ICML 2010 - Proceedings, 27th International Conference on Machine Learning*, pp. 399–406.

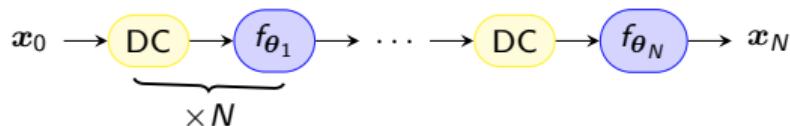
# Unrolled models - 1

---

We can mix the 2 single domain approaches, using the principled **optimization algorithm unrolling** method.<sup>5</sup>

$$\mathbf{x}_{n+1} = \mathbf{x}_n - \epsilon_n \mathcal{A}^H (\mathcal{A}\mathbf{x}_n - \mathbf{y})$$

$$\mathbf{x}_{n+1} = f_{\theta_n}(\mathbf{x}_{n+1})$$



---

<sup>5</sup>K. Gregor et al. (2010). "Learning fast approximations of sparse coding". In: *ICML 2010 - Proceedings, 27th International Conference on Machine Learning*, pp. 399–406.

# Unrolled models - 2

---

## Contribution

**Zaccharie Ramzi**, P. Ciuciu, and J. L. Starck (2020). “Benchmarking MRI reconstruction neural networks on large public datasets”. In: *Applied Sciences (Switzerland)* 10.5

We can build different models depending on the optimization algorithm we unroll, the choice of  $f_\theta$  and the number of iterations.

# Unrolled models - 2

## Contribution

Zaccharie Ramzi, P. Ciuciu, and J. L. Starck (2020). “Benchmarking MRI reconstruction neural networks on large public datasets”. In: *Applied Sciences (Switzerland)* 10.5

We can build different models depending on the optimization algorithm we unroll, the choice of  $f_\theta$  and the number of iterations.

Table: Quantitative results for the fastMRI dataset. The PSNR is computed over the 200 validation volumes.

Network	Zero-filled	KIKI-net	U-net	Cascade net	PD-net <sup>a</sup>
PSNR	29.61	31.38	31.78	31.97	32.15

<sup>a</sup>J. Adler et al. (2018). “Learned Primal-Dual Reconstruction”. In: *IEEE Transactions on Medical Imaging* 37.6, pp. 1322–1332.

# Unrolled models - 2

## Contribution

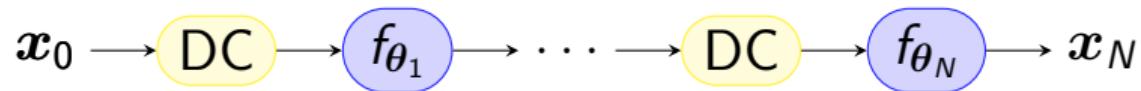
**Zaccharie Ramzi**, P. Ciuciu, and J. L. Starck (2020). “Benchmarking MRI reconstruction neural networks on large public datasets”. In: *Applied Sciences (Switzerland)* 10.5

We can build different models depending on the optimization algorithm we unroll, the choice of  $f_\theta$  and the number of iterations.

- 🎨 Code available online:  
[github.com/zaccharieramzi/fastmri-reproducible-benchmark](https://github.com/zaccharieramzi/fastmri-reproducible-benchmark)
- 😊 Model weights available online: [huggingface.co/zaccharieramzi](https://huggingface.co/zaccharieramzi)

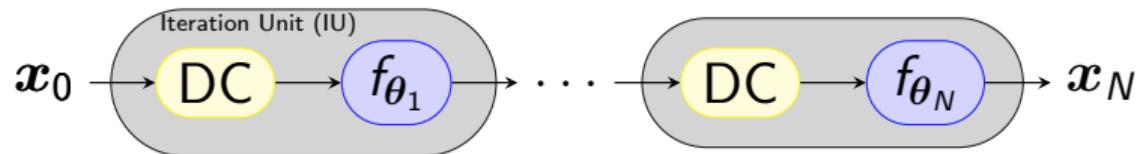
# XPDNet

---

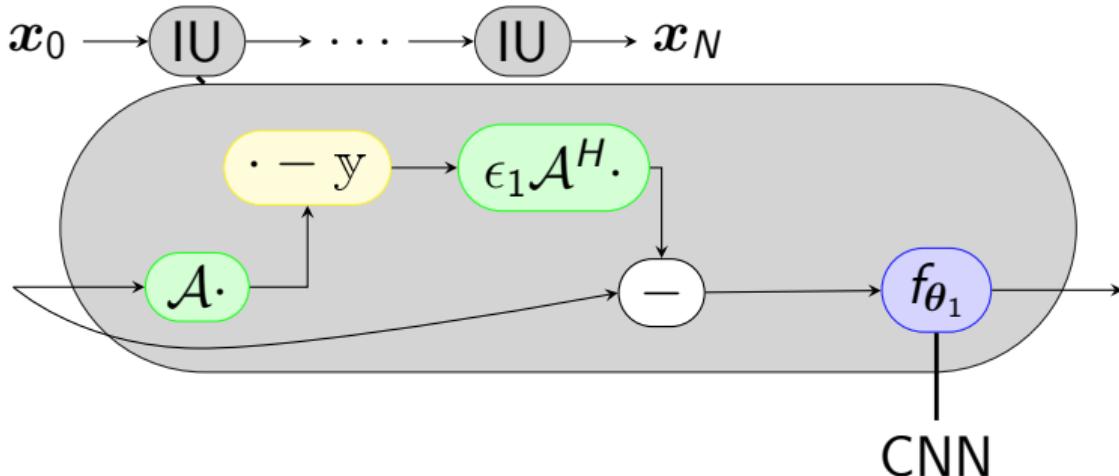


# XPDNet

---



# XPDNet

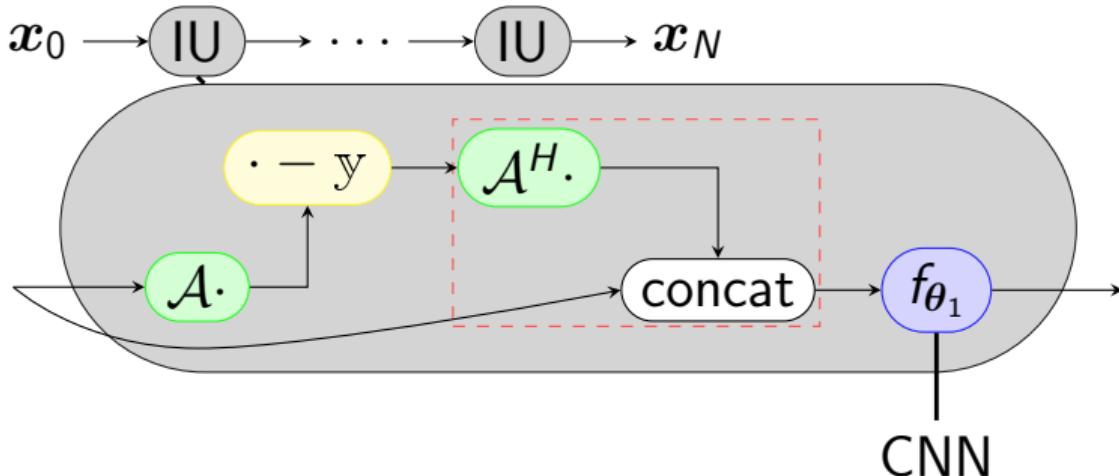


<sup>6</sup>P. Liu et al. (2018). "Multi-level Wavelet-CNN for Image Restoration". In: *CVPR NTIRE Workshop*.

<sup>7</sup>A. Sriram et al. (2020). "End-to-End Variational Networks for Accelerated MRI Reconstruction". In: *MICCAI*.

<sup>8</sup>O. Ronneberger et al. (2015). "U-net: Convolutional networks for biomedical image segmentation". In: *International Conference on Medical image computing and computer-assisted intervention*.

# XPDNet

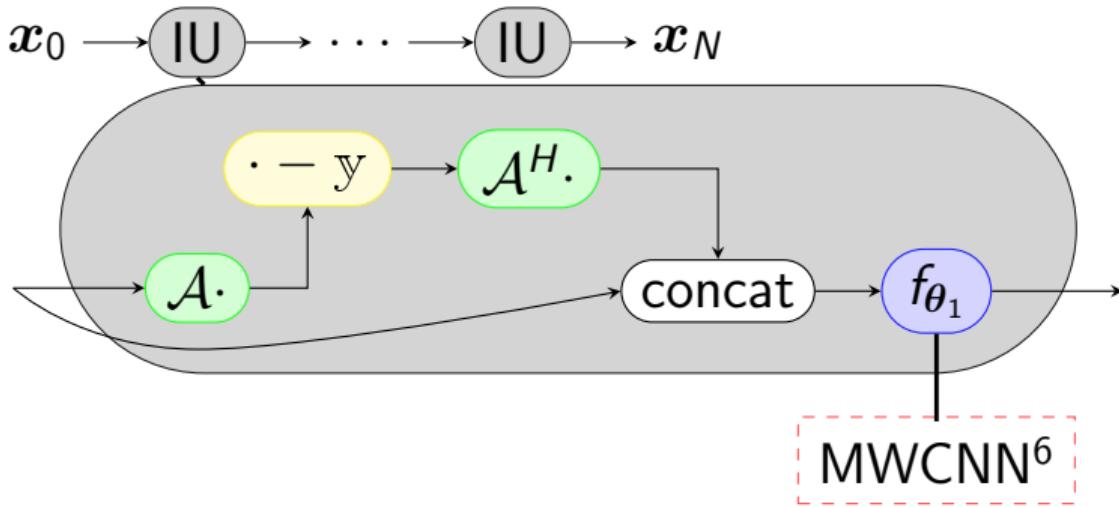


<sup>6</sup>P. Liu et al. (2018). "Multi-level Wavelet-CNN for Image Restoration". In: *CVPR NTIRE Workshop*.

<sup>7</sup>A. Sriram et al. (2020). "End-to-End Variational Networks for Accelerated MRI Reconstruction". In: *MICCAI*.

<sup>8</sup>O. Ronneberger et al. (2015). "U-net: Convolutional networks for biomedical image segmentation". In: *International Conference on Medical image computing and computer-assisted intervention*.

# XPDNet

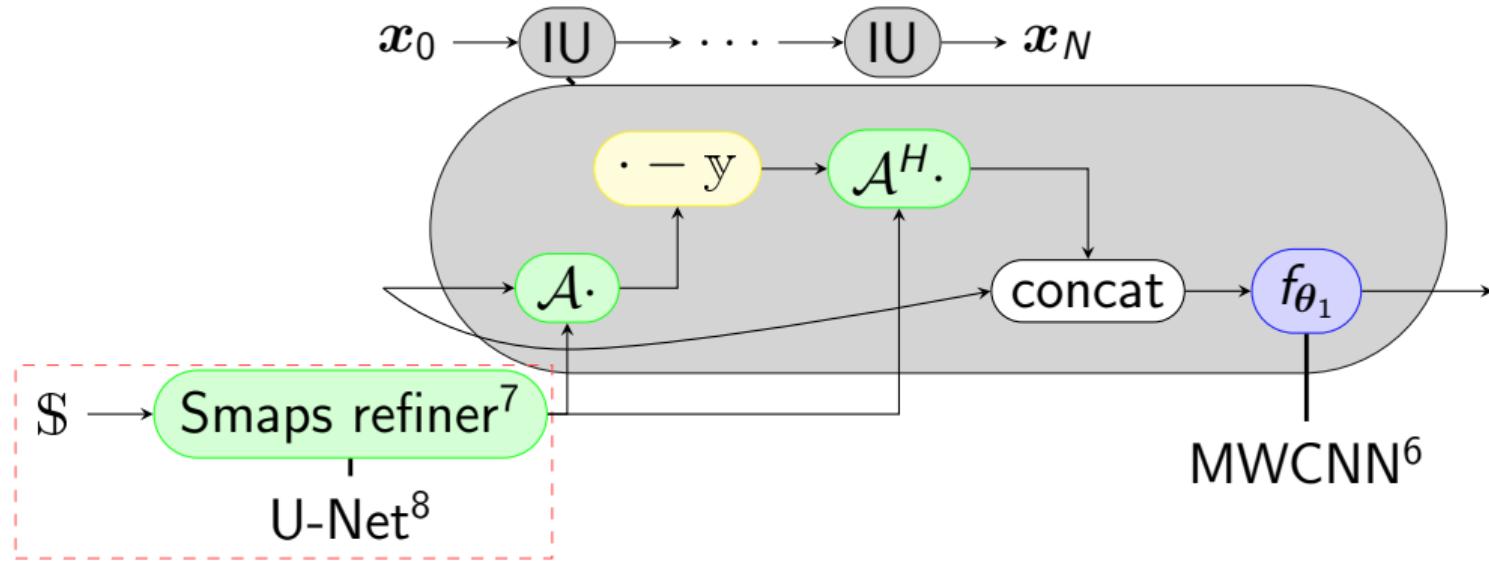


<sup>6</sup>P. Liu et al. (2018). "Multi-level Wavelet-CNN for Image Restoration". In: *CVPR NTIRE Workshop*.

<sup>7</sup>A. Sriram et al. (2020). "End-to-End Variational Networks for Accelerated MRI Reconstruction". In: *MICCAI*.

<sup>8</sup>O. Ronneberger et al. (2015). "U-net: Convolutional networks for biomedical image segmentation". In: *International Conference on Medical image computing and computer-assisted intervention*.

# XPDNet



<sup>6</sup>P. Liu et al. (2018). "Multi-level Wavelet-CNN for Image Restoration". In: *CVPR NTIRE Workshop*.

<sup>7</sup>A. Sriram et al. (2020). "End-to-End Variational Networks for Accelerated MRI Reconstruction". In: *MICCAI*.

<sup>8</sup>O. Ronneberger et al. (2015). "U-net: Convolutional networks for biomedical image segmentation". In: *International Conference on Medical image computing and computer-assisted intervention*.

# fastMRI challenge

## Contributions

- M. J. Muckley, ..., **Zaccharie Ramzi**, P. Ciuciu, J. L. Starck, ..., and F. Knoll (2021). “Results of the 2020 fastMRI Challenge for Machine Learning MR Image Reconstruction”. In: *IEEE Transactions on Medical Imaging* 40.9, pp. 2306–2317
- **Zaccharie Ramzi**, P. Ciuciu, and J.-L. Starck (2020). “XPDNet for MRI Reconstruction: an application to the 2020 fastMRI challenge”. In: *ISMRM*. Oral

Table: fastMRI challenge radiologist evaluation.

Team	Rank 4X	Rank 8X
AIRS	1.36	1.28
NeuroSpin	1.94	2.25
ATB	2.22	2.28

# Robustness test

---

How does the XPDNet fare in a prospective out-of-distribution setting:  
different orientation, higher resolution, higher field strength, lower acceleration factor,  
presence of the cerebellum.<sup>9</sup>

---

<sup>9</sup>For anonymity reasons, the cerebellum is not present in the fastMRI dataset.

<sup>10</sup>L. Marrakchi-Kacem et al. (2016). "Robust imaging of hippocampal inner structure at 7T: in vivo acquisition protocol and methodological choices". In: *Magnetic Resonance Materials in Physics, Biology and Medicine* 29.3, pp. 475–489.

# Robustness test

---

How does the XPDNet fare in a prospective out-of-distribution setting:  
different orientation, higher resolution, higher field strength, lower acceleration factor,  
presence of the cerebellum.<sup>9</sup>

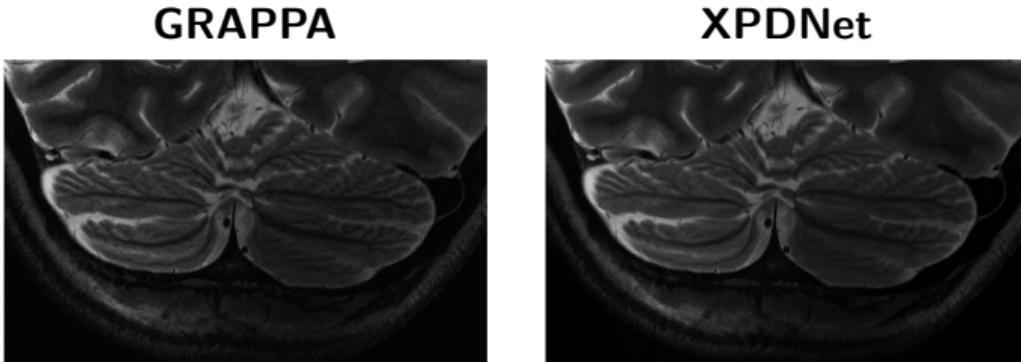


Figure: XPDNet reconstruction on a brain prospectively accelerated.<sup>10</sup> (zoom on the cerebellum)

---

<sup>9</sup>For anonymity reasons, the cerebellum is not present in the fastMRI dataset.

<sup>10</sup>L. Marrakchi-Kacem et al. (2016). "Robust imaging of hippocampal inner structure at 7T: in vivo acquisition protocol and methodological choices". In: *Magnetic Resonance Materials in Physics, Biology and Medicine* 29.3, pp. 475–489.

# Non-Cartesian acquisitions

---

We need non-Cartesian acquisitions to better cover the k-space.

# Non-Cartesian acquisitions

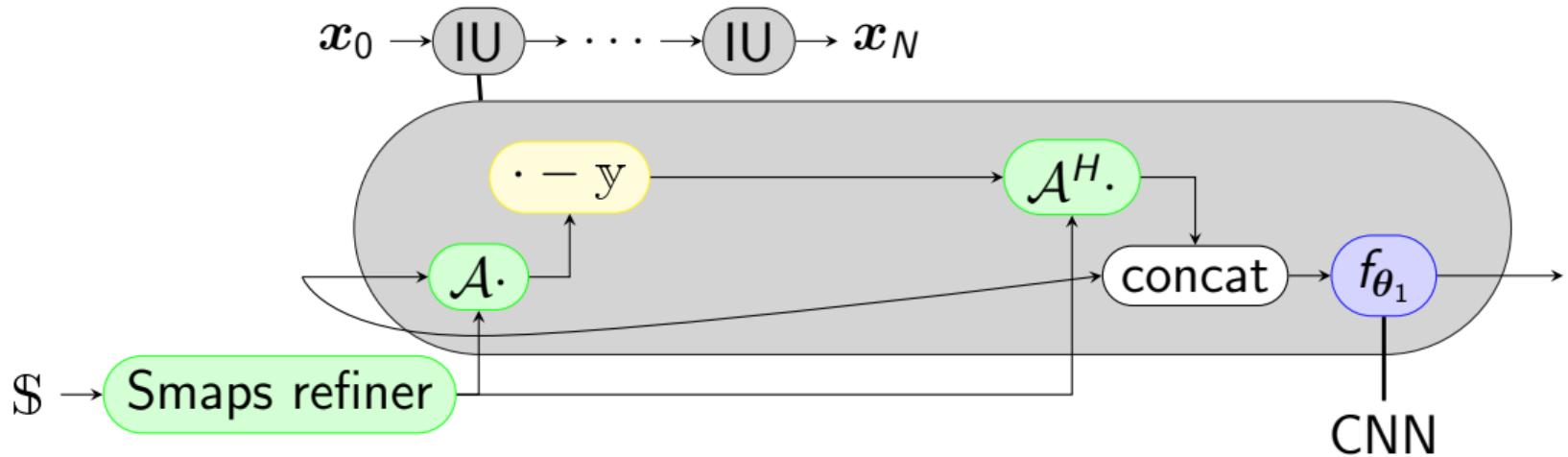
---

We need non-Cartesian acquisitions to better cover the k-space.

The difficulty from a computational point of view is that we need to now use the **Nonuniform Fourier Transform (NDFT)**. We resort to using its fast approximation, the NUFFT, which we implemented in TensorFlow, to enable gradient-based learning:

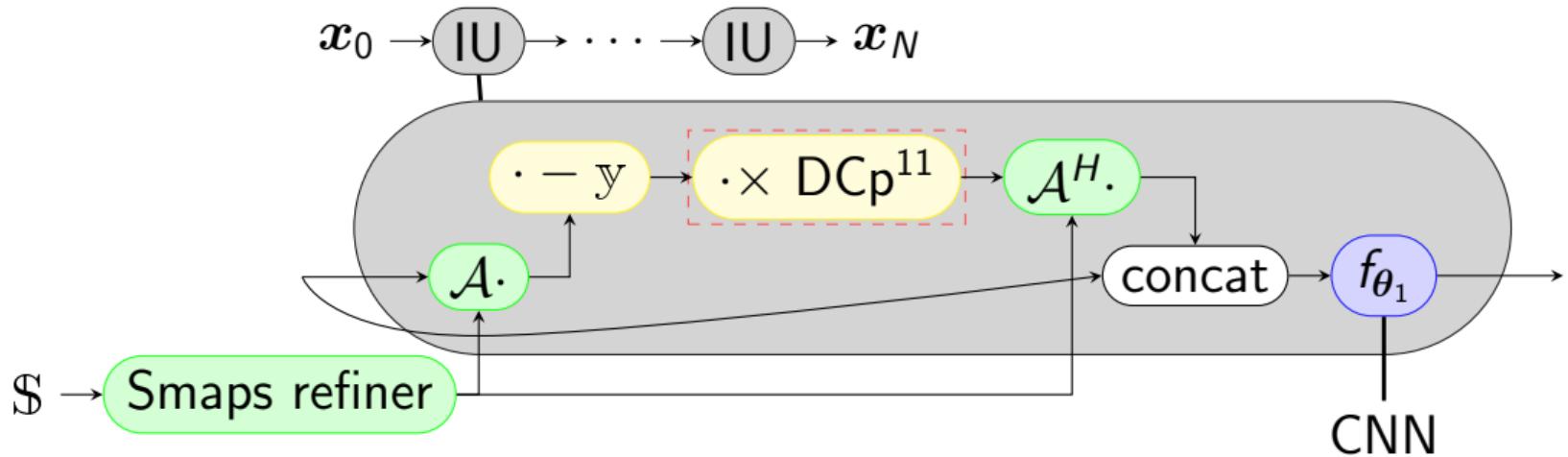
-  Code available online: [github.com/zaccharieramzi/tfkbnufft](https://github.com/zaccharieramzi/tfkbnufft)

# NC-PDNet - 1



<sup>11</sup>J. G. Pipe et al. (1999). "Sampling density compensation in MRI: Rationale and an iterative numerical solution". In: *Magnetic Resonance in Medicine* 41.1, pp. 179–186.

# NC-PDNet - 1



<sup>11</sup>J. G. Pipe et al. (1999). "Sampling density compensation in MRI: Rationale and an iterative numerical solution". In: *Magnetic Resonance in Medicine* 41.1, pp. 179–186.

# NC-PDNet - 2

## Contribution

Zaccharie Ramzi, J.-L. Starck, C. G R, and P. Ciuciu (2021). “NC-PDNet: a Density-Compensated Unrolled Network for 2D and 3D non-Cartesian MRI Reconstruction”. Accepted to IEEE Transactions on Medical Imaging

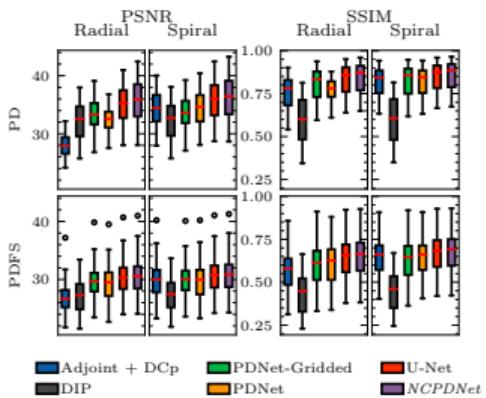


Figure: 2D single-coil reconstruction quantitative results on the fastMRI knee dataset for non-Cartesian trajectories.

# NC-PDNet - 2

## Contribution

Zaccharie Ramzi, J.-L. Starck, C. G R, and P. Ciuciu (2021). “NC-PDNet: a Density-Compensated Unrolled Network for 2D and 3D non-Cartesian MRI Reconstruction”. Accepted to IEEE Transactions on Medical Imaging

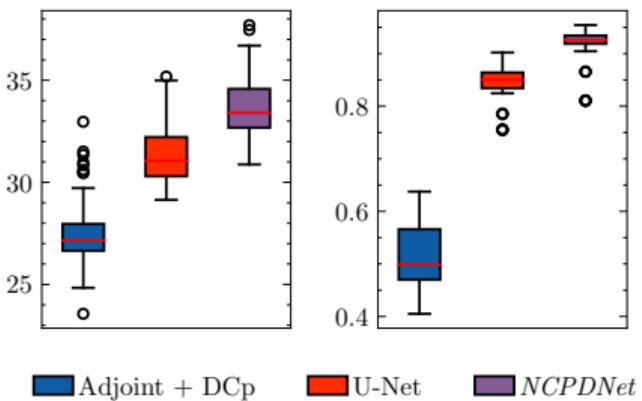


Figure: 3D single-coil reconstruction quantitative results on the OASIS dataset for a radial trajectory.

# NC-PDNet - 2

## Contribution

Zaccharie Ramzi, J.-L. Starck, C. G R, and P. Ciuciu (2021). "NC-PDNet: a Density-Compensated Unrolled Network for 2D and 3D non-Cartesian MRI Reconstruction". Accepted to IEEE Transactions on Medical Imaging

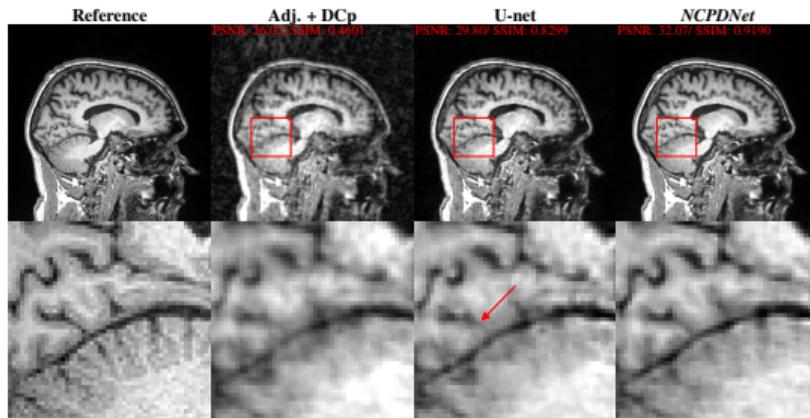


Figure: 3D single-coil reconstruction qualitative results on the OASIS dataset for a radial trajectory.

# Unrolled models for MRI reconstruction

---

## Recap

MRI is slow because of **relaxation**.

# Unrolled models for MRI reconstruction

---

## Recap

MRI is slow because of **relaxation**.

If we want to do fewer relaxations, we need to exploit some **redundancy** in MR images.

# Unrolled models for MRI reconstruction

---

## Recap

MRI is slow because of **relaxation**.

If we want to do fewer relaxations, we need to exploit some **redundancy** in MR images.

**Deep Learning** allows us to learn more complex structures in MR images than Compressed Sensing. We showcased 2 instances of unrolled models, **XPDNet** and **NC-PDNet**, which can perform really well in challenging acquisition settings.

# Unrolled models for MRI reconstruction

---

## Recap

MRI is slow because of **relaxation**.

If we want to do fewer relaxations, we need to exploit some **redundancy** in MR images.

**Deep Learning** allows us to learn more complex structures in MR images than Compressed Sensing. We showcased 2 instances of unrolled models, **XPDNet** and **NC-PDNet**, which can perform really well in challenging acquisition settings.

But we needed to trade off some model capacity for memory, in order to train the models in the 3D single-coil case. How will this fare going to 3D multi-coil?

Going even deeper

# Why should we go deep?

---

The general empirical gist of DL is that with deeper models comes better performance.

# Why should we go deep?

The general empirical gist of DL is that with deeper models comes better performance.



Figure: Credits: reddit.com/r/ProgrammerHumor/comments/5si1f0/machine\_learning\_approaches/

# Why should we go deep?

The general empirical gist of DL is that with deeper models comes better performance.

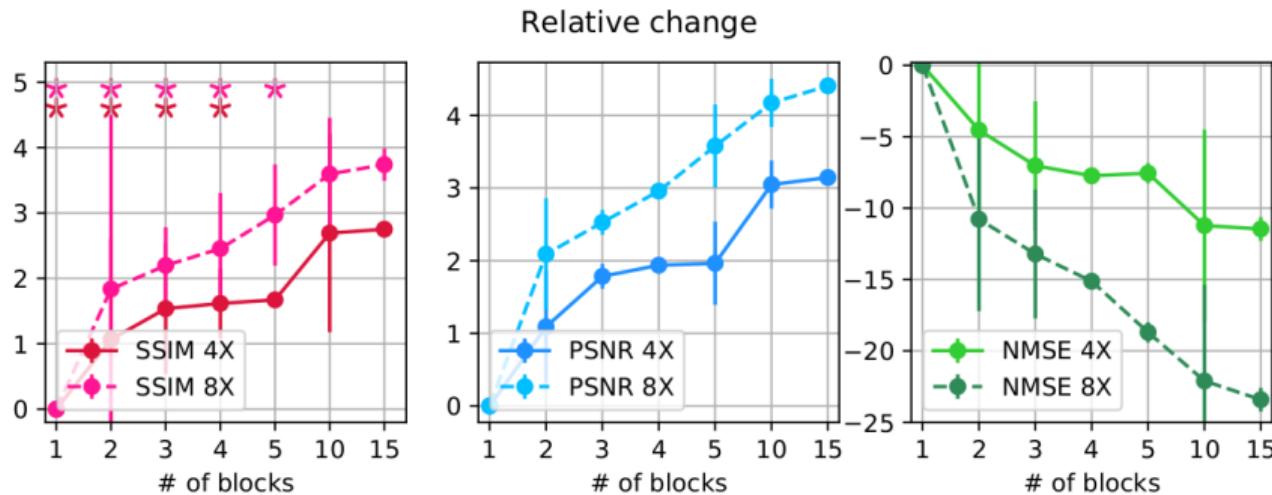


Figure: Performance of an unrolled MRI reconstruction network function of the number of iteration units (blocks).<sup>b</sup>

<sup>a</sup>N. Pezzotti et al. (2020). "An adaptive intelligence algorithm for undersampled knee MRI reconstruction". In: *IEEE Access* 8, pp. 204825–204838.

# Can we go deeper?

---

The big problem at hand is that deeper models need more **memory** to train, and the memory we have is constrained.

# Can we go deeper?

---

The big problem at hand is that deeper models need more **memory** to train, and the memory we have is constrained.

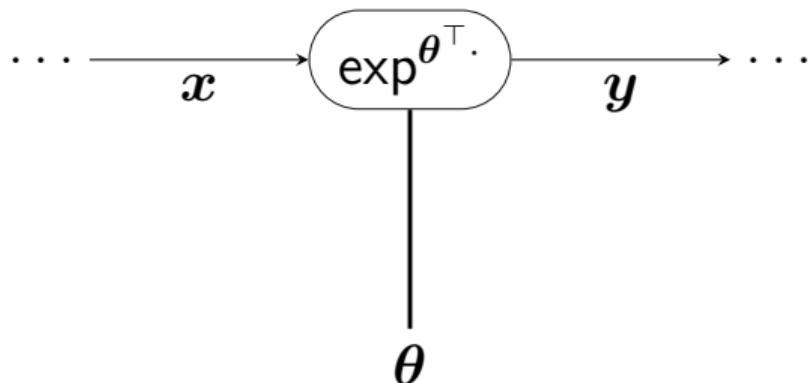
This memory requirement comes primarily from the **activations** of the model, not from the model weights.

# Can we go deeper?

---

The big problem at hand is that deeper models need more **memory** to train, and the memory we have is constrained.

This memory requirement comes primarily from the **activations** of the model, not from the model weights.

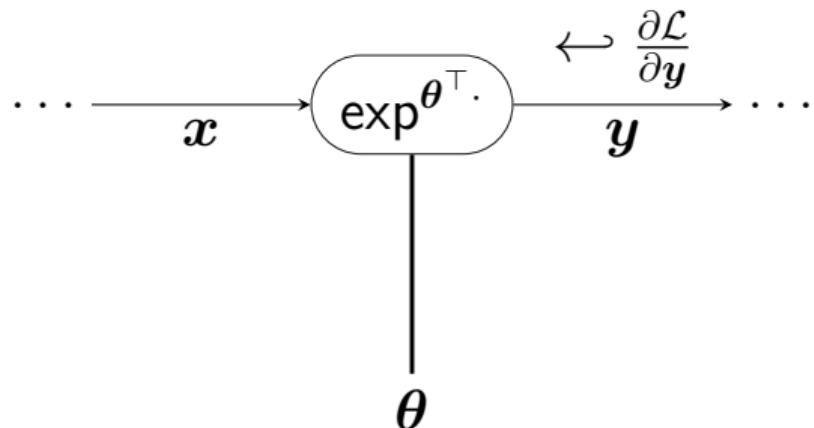


# Can we go deeper?

---

The big problem at hand is that deeper models need more **memory** to train, and the memory we have is constrained.

This memory requirement comes primarily from the **activations** of the model, not from the model weights.

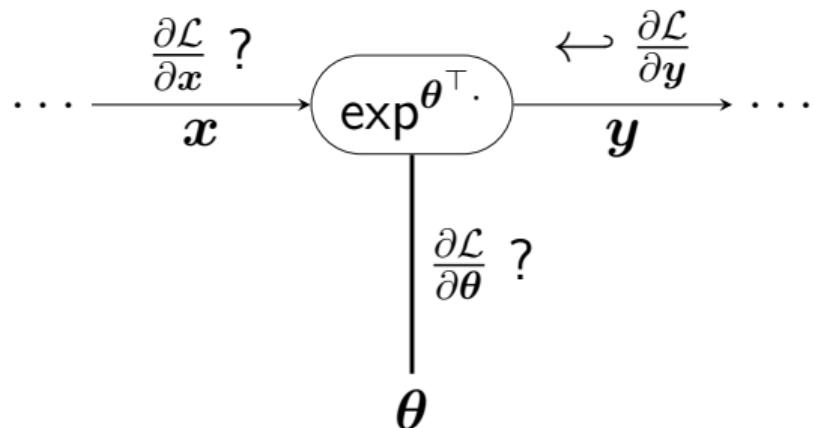


# Can we go deeper?

---

The big problem at hand is that deeper models need more **memory** to train, and the memory we have is constrained.

This memory requirement comes primarily from the **activations** of the model, not from the model weights.

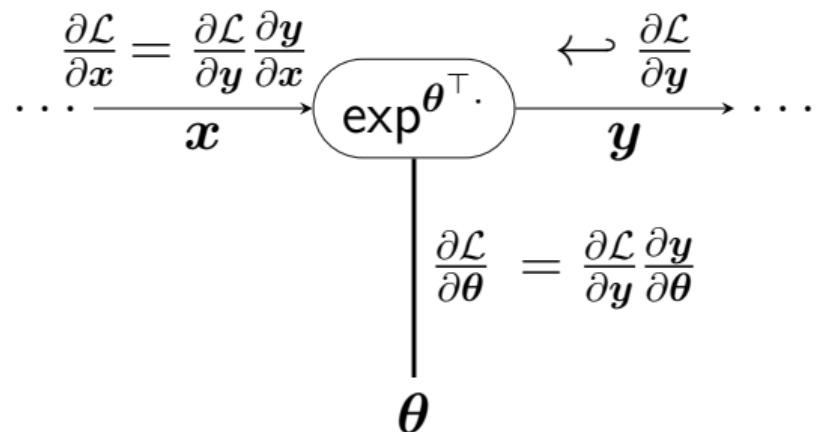


# Can we go deeper?

---

The big problem at hand is that deeper models need more **memory** to train, and the memory we have is constrained.

This memory requirement comes primarily from the **activations** of the model, not from the model weights.

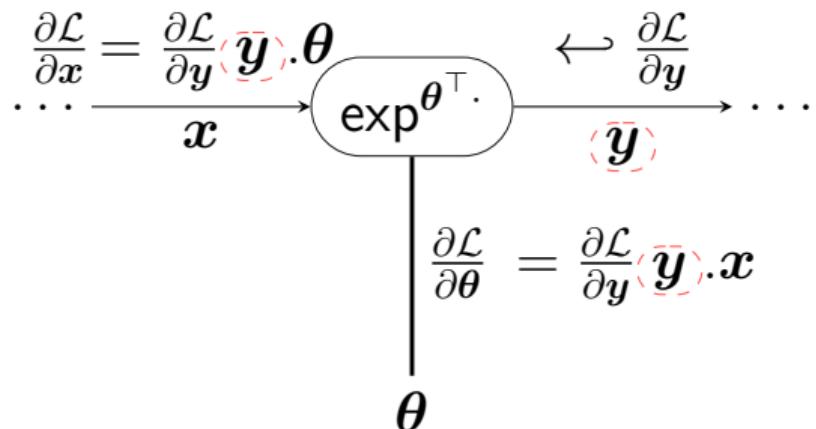


# Can we go deeper?

---

The big problem at hand is that deeper models need more **memory** to train, and the memory we have is constrained.

This memory requirement comes primarily from the **activations** of the model, not from the model weights.



# The modeling solutions

---

On the modeling side, several solutions exist:

- Gradient checkpointing (T. Chen et al., 2016)

# The modeling solutions

---

On the modeling side, several solutions exist:

- Gradient checkpointing (T. Chen et al., 2016)
- Invertible networks (Gomez et al., 2017; Sander et al., 2021)

# The modeling solutions

---

On the modeling side, several solutions exist:

- Gradient checkpointing (T. Chen et al., 2016)
- Invertible networks (Gomez et al., 2017; Sander et al., 2021)
- Implicit models (Bai, Kolter, et al., 2019; R. T. Chen et al., 2018)

# Deep Equilibrium networks - 1

---

Bai, Kolter, et al. (2019) introduced **Deep Equilibrium networks (DEQs)**, a type of implicit model. The output of DEQs is defined implicitly as the solution to a fixed-point equation.

$$h_{\theta}(x) = z^*, \text{ where } z^* = f_{\theta}(z^*, x)$$

# Deep Equilibrium networks - 1

---

Bai, Kolter, et al. (2019) introduced **Deep Equilibrium networks (DEQs)**, a type of implicit model. The output of DEQs is defined implicitly as the solution to a fixed-point equation.

$$h_{\theta}(x) = z^*, \text{ where } z^* = f_{\theta}(z^*, x)$$

In practice, we solve this equation with a **quasi-Newton method**.

# Deep Equilibrium networks - 1

---

Bai, Kolter, et al. (2019) introduced **Deep Equilibrium networks (DEQs)**, a type of implicit model. The output of DEQs is defined implicitly as the solution to a fixed-point equation.

$$h_{\theta}(x) = z^*, \text{ where } g_{\theta}(z^*, x) = z^* - f_{\theta}(z^*, x) = 0$$

In practice, we solve this equation with a **quasi-Newton method**.

# Deep Equilibrium networks - 2

---

The big question is now: how do I compute the gradient  $\frac{\partial \mathcal{L}}{\partial \theta}$ ?

# Deep Equilibrium networks - 2

The big question is now: how do I compute the gradient  $\frac{\partial \mathcal{L}}{\partial \theta}$ ?

The **Implicit Function Theorem** gives us just that:

Theorem (Hypergradient (Bai, Kolter, et al., 2019; Krantz et al., 2013))

Let  $\theta \in \mathbb{R}^p$  be a set of parameters, let  $\mathcal{L} : \mathbb{R}^d \rightarrow \mathbb{R}$  be a loss function and  $g_\theta : \mathbb{R}^d \rightarrow \mathbb{R}^d$  be a root-defining function. Let  $z^* \in \mathbb{R}^d$  such that  $g_\theta(z^*) = 0$  and  $J_{g_\theta}(z^*) = \left. \frac{\partial g_\theta}{\partial z} \right|_{z^*}$  is invertible, then the gradient of the loss  $\mathcal{L}$  wrt.  $\theta$ , called Hypergradient, is given by

$$\left. \frac{\partial \mathcal{L}}{\partial \theta} \right|_{z^*} = \nabla_z \mathcal{L}(z^*)^\top J_{g_\theta}(z^*)^{-1} \left. \frac{\partial g_\theta}{\partial \theta} \right|_{z^*}.$$

# Deep Equilibrium networks - 2

The big question is now: how do I compute the gradient  $\frac{\partial \mathcal{L}}{\partial \theta}$ ?

The **Implicit Function Theorem** gives us just that:

Theorem (Hypergradient (Bai, Kolter, et al., 2019; Krantz et al., 2013))

Let  $\theta \in \mathbb{R}^p$  be a set of parameters, let  $\mathcal{L} : \mathbb{R}^d \rightarrow \mathbb{R}$  be a loss function and  $g_\theta : \mathbb{R}^d \rightarrow \mathbb{R}^d$  be a root-defining function. Let  $z^* \in \mathbb{R}^d$  such that  $g_\theta(z^*) = 0$  and  $J_{g_\theta}(z^*) = \left. \frac{\partial g_\theta}{\partial z} \right|_{z^*}$  is invertible, then the gradient of the loss  $\mathcal{L}$  wrt.  $\theta$ , called Hypergradient, is given by

$$\left. \frac{\partial \mathcal{L}}{\partial \theta} \right|_{z^*} = \nabla_z \mathcal{L}(z^*)^\top J_{g_\theta}(z^*)^{-1} \left. \frac{\partial g_\theta}{\partial \theta} \right|_{z^*}.$$

Key for our memory problem: it does not rely on any activations you could have when solving the fixed-point equation.

# The limits of DEQs

DEQs achieve excellent results in NLP (Natural Language Processing) and CV (Computer Vision) tasks, but they are slow to train.

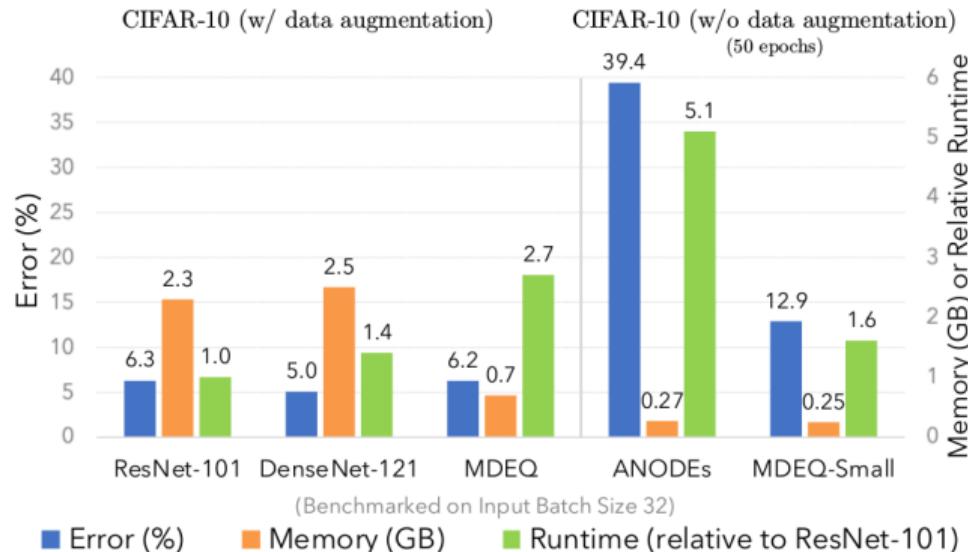


Figure: Performance, memory and training speed of DEQs. (Bai, Koltun, et al., 2020)

# Why are DEQs slow?

---

If we look back the equation used to compute the gradient of DEQs:

$$\frac{\partial \mathcal{L}}{\partial \theta} \Big|_{z^*} = \nabla_z \mathcal{L}(z^*)^\top J_{g_\theta}(z^*)^{-1} \frac{\partial g_\theta}{\partial \theta} \Big|_{z^*},$$

we see that we need to invert a huge matrix  $J_{g_\theta}(z^*)$  in a certain direction  $\nabla_z \mathcal{L}(z^*)$ .

# Why are DEQs slow?

---

If we look back the equation used to compute the gradient of DEQs:

$$\frac{\partial \mathcal{L}}{\partial \theta} \Big|_{z^*} = \nabla_z \mathcal{L}(z^*)^\top J_{g_\theta}(z^*)^{-1} \frac{\partial g_\theta}{\partial \theta} \Big|_{z^*},$$

we see that we need to invert a huge matrix  $J_{g_\theta}(z^*)$  in a certain direction  $\nabla_z \mathcal{L}(z^*)$ .  
In practice this is done using an iterative algorithm.

# Can we avoid the Jacobian inversion?

## Contribution

Zaccharie Ramzi, F. Mannel, S. Bai, J.-L. Starck, P. Ciuciu, and T. Moreau (2022). “SHINE: SHaring the INverse Estimate from the forward pass for bi-level optimization and implicit models”. In: *ICLR*. (Spotlight)

We introduced **SHINE: SHaring the INverse Estimate**.

$$B^{-1} \approx J_{g_\theta}(z^*)^{-1}$$

# Can we avoid the Jacobian inversion?

## Contribution

Zaccharie Ramzi, F. Mannel, S. Bai, J.-L. Starck, P. Ciuciu, and T. Moreau (2022). "SHINE: SHaring the INverse Estimate from the forward pass for bi-level optimization and implicit models". In: *ICLR*. (Spotlight)

We introduced **SHINE: SHaring the INverse Estimate**.

$$B^{-1} \approx J_{g_\theta}(z^*)^{-1}$$

  
True Jacobian inverse

# Can we avoid the Jacobian inversion?

## Contribution

Zaccharie Ramzi, F. Mannel, S. Bai, J.-L. Starck, P. Ciuciu, and T. Moreau (2022). "SHINE: SHaring the INverse Estimate from the forward pass for bi-level optimization and implicit models". In: *ICLR*. (Spotlight)

We introduced **SHINE: SHaring the INverse Estimate**.

$$B^{-1} \approx J_{g_\theta}(z^*)^{-1}$$

quasi-Newton matrix      True Jacobian inverse

```
graph TD; B_inv[B^{-1}] <-->|blue arrow| quasi[quasi-Newton matrix]; J_inv[J_{g_\theta}(z^*)^{-1}] <-->|red arrow| true[True Jacobian inverse]
```

# Can we avoid the Jacobian inversion?

## Contribution

Zaccharie Ramzi, F. Mannel, S. Bai, J.-L. Starck, P. Ciuciu, and T. Moreau (2022). "SHINE: SHaring the INverse Estimate from the forward pass for bi-level optimization and implicit models". In: *ICLR*. (Spotlight)

We introduced **SHINE: SHaring the INverse Estimate**.

$$B^{-1} \approx J_{g_\theta}(z^*)^{-1}$$

quasi-Newton matrix      True Jacobian inverse

Properties of  $B$ :

- It is computed when solving  $g_\theta(z^*, x)$  using a quasi-Newton method.

# Can we avoid the Jacobian inversion?

## Contribution

Zaccharie Ramzi, F. Mannel, S. Bai, J.-L. Starck, P. Ciuciu, and T. Moreau (2022). "SHINE: SHaring the INverse Estimate from the forward pass for bi-level optimization and implicit models". In: *ICLR*. (Spotlight)

We introduced **SHINE: SHaring the INverse Estimate**.

$$B^{-1} \approx J_{g_\theta}(z^*)^{-1}$$

quasi-Newton matrix      True Jacobian inverse

Properties of  $B$ :

- It is computed when solving  $g_\theta(z^*, x)$  using a quasi-Newton method.
- It is easily invertible using the Sherman-Morrison formula.

# Application to Hyperparameter optimization - 1

---

Interestingly, other problems can benefit from this idea, notably Hyperparameter optimization.

$$\begin{aligned} & \arg \min_{\lambda} \mathcal{L}_{\text{val}}(\boldsymbol{x}^*) \\ \text{s.t. } & \boldsymbol{x}^* = \arg \min_{\boldsymbol{x}} \mathcal{L}_{\text{train}}(\boldsymbol{x}) + \exp^{\lambda} \|\boldsymbol{x}\|_2^2 \end{aligned}$$

# Application to Hyperparameter optimization - 1

---

Interestingly, other problems can benefit from this idea, notably Hyperparameter optimization.

$$\begin{aligned} & \arg \min_{\lambda} \mathcal{L}_{\text{val}}(\boldsymbol{x}^*) \\ \text{s.t. } & \boldsymbol{x}^* = \arg \min_{\boldsymbol{x}} \mathcal{L}_{\text{train}}(\boldsymbol{x}) + \exp^{\lambda} \|\boldsymbol{x}\|_2^2 \end{aligned}$$

The IFT can also be applied, and when a quasi-Newton method is used to solve  $\arg \min_{\boldsymbol{x}} \mathcal{L}_{\text{train}}(\boldsymbol{x}) + \exp^{\lambda} \|\boldsymbol{x}\|_2^2$ , we may use SHINE.

# Application to Hyperparameter optimization - 2

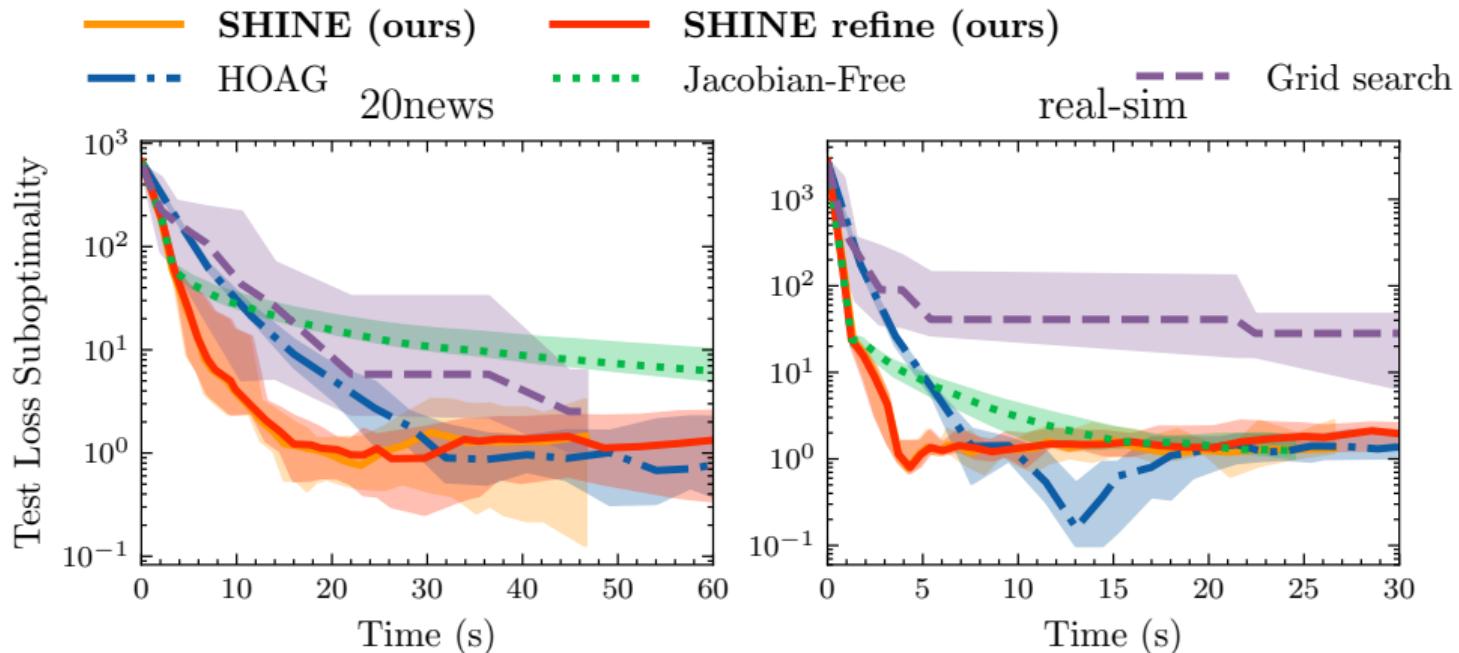


Figure: Bilevel optimization (Pedregosa, 2016) with SHINE: convergence of held-out test loss.

# Results on DEQs

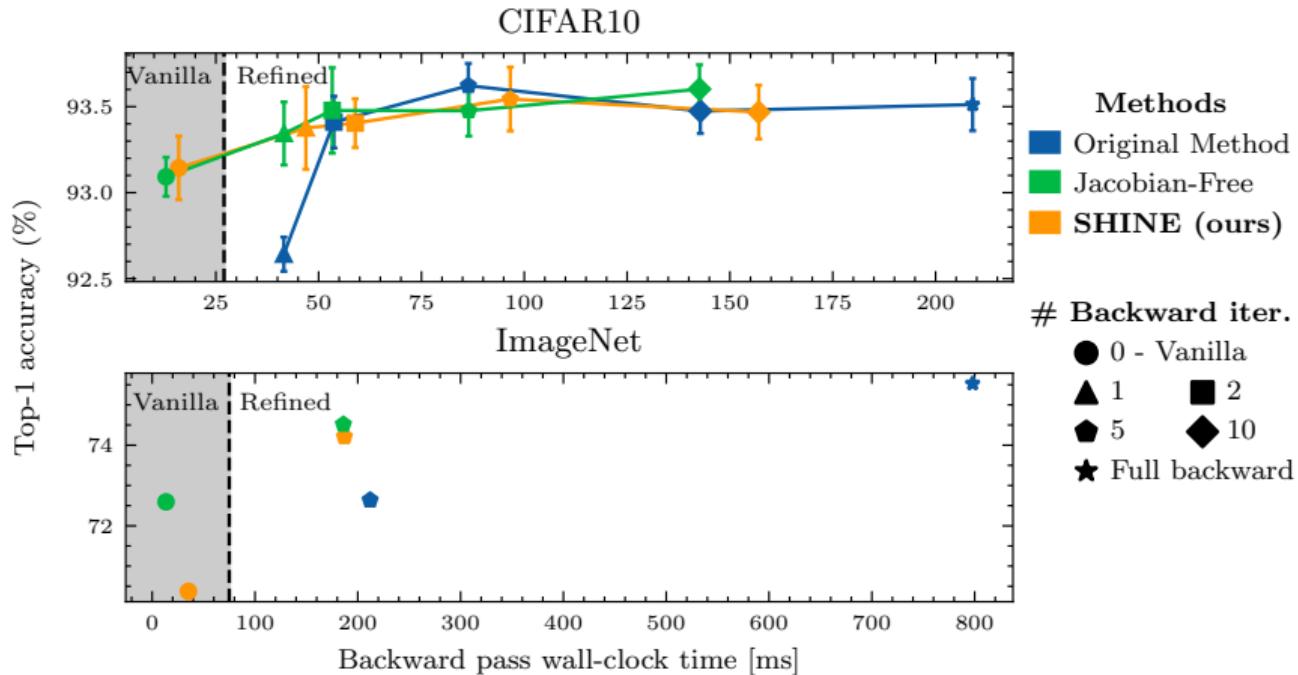


Figure: MDEQs (Bai, Koltun, et al., 2020) with SHINE.

# Conclusion

---

1. MRI is an important medical imaging modality, but it is inherently slow due to the Nuclear Magnetic Resonance phenomenon it relies on.

# Conclusion

---

1. MRI is an important medical imaging modality, but it is inherently slow due to the Nuclear Magnetic Resonance phenomenon it relies on.
2. Using redundancy is our main tool to accelerate MRI scans, but exploiting redundancy is not always straightforward.

# Conclusion

---

1. MRI is an important medical imaging modality, but it is inherently slow due to the Nuclear Magnetic Resonance phenomenon it relies on.
2. Using redundancy is our main tool to accelerate MRI scans, but exploiting redundancy is not always straightforward.
3. Deep Learning can allow us to express structure in a more principled way, and we saw two examples of this.

# Conclusion

---

1. MRI is an important medical imaging modality, but it is inherently slow due to the Nuclear Magnetic Resonance phenomenon it relies on.
2. Using redundancy is our main tool to accelerate MRI scans, but exploiting redundancy is not always straightforward.
3. Deep Learning can allow us to express structure in a more principled way, and we saw two examples of this.
  - The XPDNet, a deep learning network that secured the 2nd spot of the 2020 fastMRI challenge.

# Conclusion

---

1. MRI is an important medical imaging modality, but it is inherently slow due to the Nuclear Magnetic Resonance phenomenon it relies on.
2. Using redundancy is our main tool to accelerate MRI scans, but exploiting redundancy is not always straightforward.
3. Deep Learning can allow us to express structure in a more principled way, and we saw two examples of this.
  - The XPDNet, a deep learning network that secured the 2nd spot of the 2020 fastMRI challenge.
  - The NC-PDNet, a deep learning network that can reconstruct single-coil 3D non-Cartesian data.

# Conclusion

---

1. MRI is an important medical imaging modality, but it is inherently slow due to the Nuclear Magnetic Resonance phenomenon it relies on.
2. Using redundancy is our main tool to accelerate MRI scans, but exploiting redundancy is not always straightforward.
3. Deep Learning can allow us to express structure in a more principled way, and we saw two examples of this.
  - The XPDNet, a deep learning network that secured the 2nd spot of the 2020 fastMRI challenge.
  - The NC-PDNet, a deep learning network that can reconstruct single-coil 3D non-Cartesian data.
4. In order to prepare for even deeper networks, with the promise of even better results, we proposed SHINE, a method to accelerate DEQs, which are memory-efficient models.

## Future works

---

- Applying DEQs to MRI reconstruction.
- Refine the measurement operator even more, for example with  $B_0$  corrections (pursued by G. Daval-Frérot).
- Learn better k-space acquisition trajectories (pursued by Chaithya G R).

# Additional contributions in clinical applicability

---

## Contributions

- **Zaccharie Ramzi**, K. Michalewicz, J. L. Starck, T. Moreau, and P. Ciuciu (2021). “Wavelets in the deep learning era”. Under review in *Journal of Mathematical Imaging and Vision*
- **Zaccharie Ramzi**, B. Remy, F. Lanusse, J.-L. Starck, and P. Ciuciu (2020). “Denoising Score-Matching for Uncertainty Quantification in Inverse Problems”. In: *NeurIPS 2020 Deep Learning and Inverse Problems workshop*

# References |

---

-  Adler, J. and O. Öktem (2018). "Learned Primal-Dual Reconstruction". In: *IEEE Transactions on Medical Imaging* 37.6, pp. 1322–1332.
-  Bai, S., J. Z. Kolter, and V. Koltun (2019). "Deep equilibrium models". In: *Advances in Neural Information Processing Systems*. Vol. 32.
-  Bai, S., V. Koltun, and J. Z. Kolter (2020). "Multiscale deep equilibrium models". In: *Advances in Neural Information Processing Systems*. Vol. 2020-Decem.
-  Chen, R. T., Y. Rubanova, J. Bettencourt, and D. Duvenaud (2018). "Neural Ordinary differential equations". In: *NeurIPS*.
-  Chen, T., B. Xu, C. Zhang, and C. Guestrin (2016). *Training Deep Nets with Sublinear Memory Cost*. Tech. rep., pp. 1–12.
-  Deshmane, A., V. Gulani, M. A. Griswold, and N. Seiberlich (2012). "Parallel MR imaging". In: *Journal of Magnetic Resonance Imaging* 36.1, pp. 55–72.

## References II

---

-  *e-MRI courses by IMAIOS* (2008). <https://www.imaios.com/en/e-Courses/e-MRI>. Accessed: 2021-10-08.
-  Foucart, S. and H. Rauhut (2013). "Sparse Solutions of Underdetermined Systems". In: *A Mathematical Introduction to Compressive Sensing*. Birkhauser. Chap. 2, pp. 41–59.
-  Gomez, A. N., M. Ren, R. Urtasun, and R. B. Grosse (2017). "The Reversible Residual Network: Backpropagation Without Storing Activations". In: *NIPS*.
-  Gregor, K. and Y. LeCun (2010). "Learning fast approximations of sparse coding". In: *ICML 2010 - Proceedings, 27th International Conference on Machine Learning*, pp. 399–406.
-  Griswold, M. A., P. M. Jakob, R. M. Heidemann, M. Nittka, V. Jellus, J. Wang, B. Kiefer, and A. Haase (June 2002). "Generalized Autocalibrating Partially Parallel Acquisitions (GRAPPA)". In: *Magnetic Resonance in Medicine* 47.6, pp. 1202–1210.

## References III

---

-  Krantz, S. G. and H. R. Parks (Jan. 2013). *The Implicit Function Theorem: History, Theory, and Applications*. Springer New York, pp. 1–163.
-  *Larmor precession Wikipedia page* (2012).  
[https://en.wikipedia.org/wiki/Larmor\\_precession](https://en.wikipedia.org/wiki/Larmor_precession). Accessed: 2021-10-09.
-  Liu, P., H. Zhang, K. Zhang, L. Lin, and W. Zuo (2018). “Multi-level Wavelet-CNN for Image Restoration”. In: *CVPR NTIRE Workshop*.
-  Lustig, M., D. Donoho, and J. M. Pauly (2007). “Sparse MRI: The application of compressed sensing for rapid MR imaging”. In: *Magnetic Resonance in Medicine* 58.6, pp. 1182–1195.
-  Marrakchi-Kacem, L., A. Vignaud, J. Sein, J. Germain, T. R. Henry, C. Poupon, L. Hertz-Pannier, S. Lehéricy, O. Colliot, P. F. Van de Moortele, and M. Chupin (2016). “Robust imaging of hippocampal inner structure at 7T: in vivo acquisition protocol and methodological choices”. In: *Magnetic Resonance Materials in Physics, Biology and Medicine* 29.3, pp. 475–489.

# References IV

---

-  Muckley, M. J., ..., **Zaccharie Ramzi**, P. Ciuciu, J. L. Starck, ..., and F. Knoll (2021). "Results of the 2020 fastMRI Challenge for Machine Learning MR Image Reconstruction". In: *IEEE Transactions on Medical Imaging* 40.9, pp. 2306–2317.
-  *Health at a Glance 2019: OECD Indicators - Medical technologies* (2019).  
<https://www.oecd-ilibrary.org/sites/eadc0d9d-en/index.html?itemId=/content/component/eadc0d9d-en>. Accessed: 2021-10-06.
-  Pedregosa, F. (2016). "Hyperparameter optimization with approximate gradient". In: *33rd International Conference on Machine Learning, ICML 2016* 2, pp. 1150–1159.
-  Pezzotti, N., S. Yousefi, M. S. Elmahdy, J. H. F. van Gemert, C. Schuelke, M. Doneva, T. Nielsen, S. Kastrulyin, B. P. Lelieveldt, M. J. van Osch, E. D. Weerdt, and M. Staring (2020). "An adaptive intelligence algorithm for undersampled knee MRI reconstruction". In: *IEEE Access* 8, pp. 204825–204838.

# References V

---

-  Pipe, J. G. and P. Menon (1999). "Sampling density compensation in MRI: Rationale and an iterative numerical solution". In: *Magnetic Resonance in Medicine* 41.1, pp. 179–186.
-  Pruessmann, K. P., M. Weiger, M. B. Scheidegger, and P. Boesiger (Nov. 1999). "SENSE: Sensitivity encoding for fast MRI". In: *Magnetic Resonance in Medicine* 42.5, pp. 952–962.
-  Reimer, P., P. M. Parizel, J. F. Meaney, and F. A. Stichnoth (2010). *Clinical MR imaging*. Springer.
-  Ronneberger, O., P. Fischer, and T. Brox (2015). "U-net: Convolutional networks for biomedical image segmentation". In: *International Conference on Medical image computing and computer-assisted intervention*.
-  Runge, V. M., H. von Tengg-Kobligk, and J. Heverhagen (2019). *Essentials of Clinical MR*. Thieme.

# References VI

---

-  Sander, M. E., P. Ablin, M. Blondel, and G. Peyré (2021). "Momentum Residual Neural Networks". In: *ICML*.
-  Sriram, A., J. Zbontar, T. Murrell, A. Defazio, C. L. Zitnick, N. Yakubova, F. Knoll, and P. Johnson (2020). "End-to-End Variational Networks for Accelerated MRI Reconstruction". In: *MICCAI*.
-  **Zaccharie Ramzi**, P. Ciuciu, and J. L. Starck (2020). "Benchmarking MRI reconstruction neural networks on large public datasets". In: *Applied Sciences (Switzerland)* 10.5.
-  **Zaccharie Ramzi**, P. Ciuciu, and J.-L. Starck (2020). "XPDNet for MRI Reconstruction: an application to the 2020 fastMRI challenge". In: *ISMRM*. Oral.
-  **Zaccharie Ramzi**, F. Mannel, S. Bai, J.-L. Starck, P. Ciuciu, and T. Moreau (2022). "SHINE: SHaring the INverse Estimate from the forward pass for bi-level optimization and implicit models". In: *ICLR*. (Spotlight).

# References VII

---

-  **Zaccharie Ramzi**, K. Michalewicz, J. L. Starck, T. Moreau, and P. Ciuciu (2021). “Wavelets in the deep learning era”. Under review in *Journal of Mathematical Imaging and Vision*.
-  **Zaccharie Ramzi**, B. Remy, F. Lanusse, J.-L. Starck, and P. Ciuciu (2020). “Denoising Score-Matching for Uncertainty Quantification in Inverse Problems”. In: *NeurIPS 2020 Deep Learning and Inverse Problems workshop*.
-  **Zaccharie Ramzi**, J.-L. Starck, C. G R, and P. Ciuciu (2021). “NC-PDNet: a Density-Compensated Unrolled Network for 2D and 3D non-Cartesian MRI Reconstruction”. Accepted to *IEEE Transactions on Medical Imaging*.

# References VIII

---

-  Zbontar, J., F. Knoll, A. Sriram, T. Murrell, Z. Huang, M. J. Muckley, A. Defazio, R. Stern, P. Johnson, M. Bruno, M. Parente, K. J. Geras, J. Katsnelson, H. Chandarana, Z. Zhang, M. Drozdzal, A. Romero, M. Rabbat, P. Vincent, N. Yakubova, J. Pinkerton, D. Wang, E. Owens, C. L. Zitnick, M. P. Recht, D. K. Sodickson, and Y. W. Lui (2018). *fastMRI: An Open Dataset and Benchmarks for Accelerated MRI*. Tech. rep.
-  Zhu, B., J. Z. Liu, S. F. Cauley, B. R. Rosen, and M. S. Rosen (Mar. 2018). “Image reconstruction by domain-transform manifold learning”. In: *Nature* 555.7697, pp. 487–492.

Backup slides