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

PhD defense, 18th February 2022

Zaccharie Ramzi

supervisors: Philippe Ciuciu and Jean-Luc Starck

Parietal team, Inria Saclay
NeuroSpin and Cosmostat, CEA Saclay

Credits: Getty Images / rubberball

MRI is slow

MRI (Magnetic Resonance Imaging) scan duration: 15 min (up to 90 min).[1]

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

[1] *NHS: How it's performed - MRI scan* (2018). https://www.nhs.uk/conditions/mri-scan/what-happens/. Accessed: 2021-10-11.

Our objective: accelerate MRI scans

Magnetic Resonance Imaging (MRI)

What does an MRI look like?

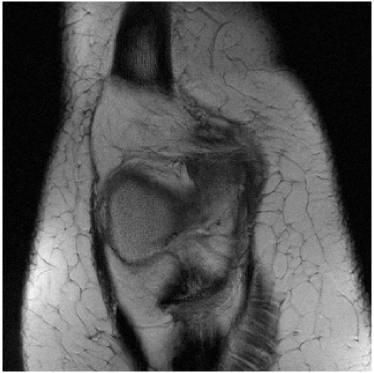


Figure: **Example of an MR image**: MR image of the knee taken from the fastMRI dataset.[2]

[2] J. Zbontar et al. (2018). *fastMRI: An Open Dataset and Benchmarks for Accelerated MRI*. Tech. rep.

Importance of MRI

99.9% chance you will get an MRI.

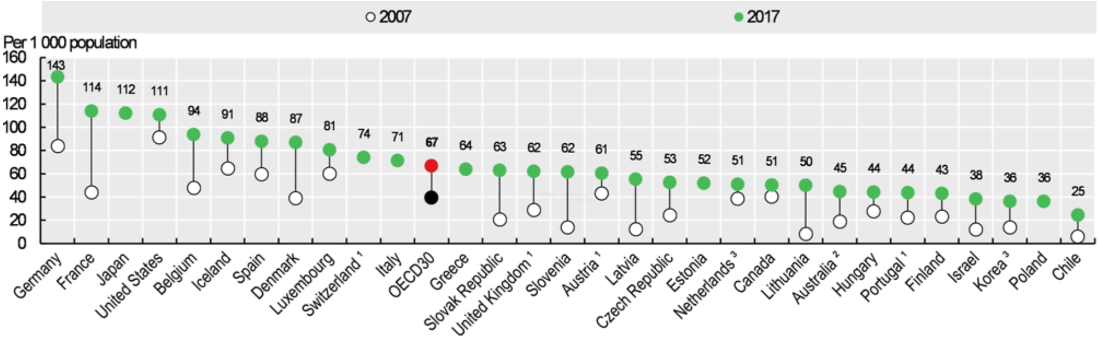
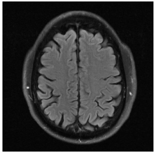


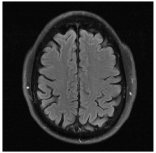
Figure: **Number of MRI scans per year per 1000 population**: figure courtesy of *Health at a Glance 2019: OECD Indicators - Medical technologies* (2019).

MRI players



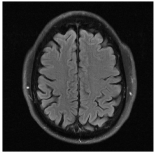
MRI players

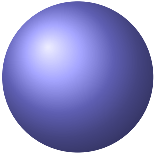
B_0



MRI players

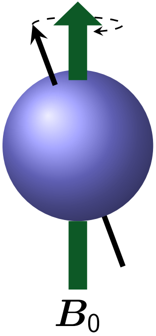
B_0



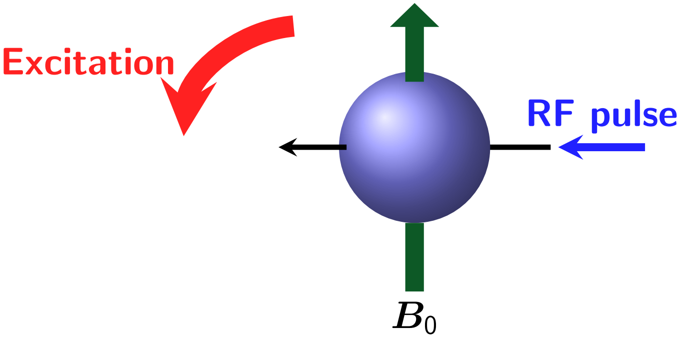


Proton

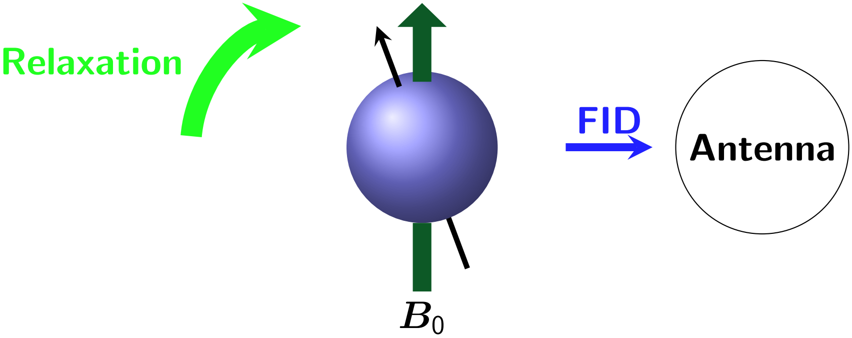
Nuclear Magnetic Resonance



Nuclear Magnetic Resonance



Nuclear Magnetic Resonance



Physics of MRI - 1

FID: global info.

Physics of MRI - 1

FID: global info.
Magnetic **gradients** $\Rightarrow$ change the magnetic field spatially.

Physics of MRI - 1

FID: global info.
Magnetic **gradients** $\Rightarrow$ change the magnetic field spatially.
Temporal signal:

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

Physics of MRI - 1

FID: global info.
Magnetic **gradients** $\Rightarrow$ change the magnetic field spatially.
Temporal signal:

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

Recorded MR signal

Physics of MRI - 1

FID: global info.
Magnetic **gradients** $\Rightarrow$ change the magnetic field spatially.
Temporal signal:

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

Recorded MR signal

Magnetic field in each location $\boldsymbol{r}$,
$\propto$ spin density $\rho(\boldsymbol{r})$

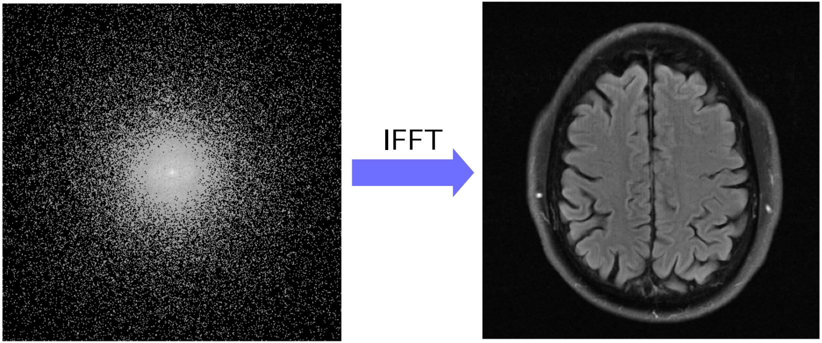
Physics of MRI - 1

FID: global info.
Magnetic **gradients** $\Rightarrow$ change the magnetic field spatially.
Temporal signal:

$$\underbrace{S_{tr}(t)}_{\text{Recorded MR signal}} \propto \omega_0 \int_{V_s} B_{tr} \underbrace{M_{tr}(t,\boldsymbol{r})}_{\text{Magnetic field in each location } \boldsymbol{r},\ \propto \text{ spin density } \rho(\boldsymbol{r})} e^{-\imath\gamma \boldsymbol{r} \cdot \int_0^t \overbrace{\boldsymbol{G}(\tau)}^{\text{Temporal gradients, controlled by the operator}} \,\mathrm{d}\tau} \,\mathrm{d}\boldsymbol{r}$$

Physics of MRI - 2

k-space vector: $\boldsymbol{k}(t) = \frac{\gamma}{2\pi} \int_0^t \boldsymbol{G}(\tau)\,\mathrm{d}\tau.$



$\boldsymbol{k}(t) \mapsto S_{tr}(t)$

$\propto \rho(\boldsymbol{r})$

Physics of MRI - 3

Recap

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

Physics of MRI - 3

Recap

MRI relies on the nuclear magnetic 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, or **sparsity, symmetry, structure or a priori information**, is the key.

Where is there room for acceleration?

Redundancy, or **sparsity, symmetry, structure or a priori information**, is the key.

Where is it in MRI?
Fourier Transform (FT) has a conjugate symmetry $\Rightarrow$ **Partial Fourier**
Resulting acceleration: 1.3

Parallel imaging

More redundancy using **more antennas (called coils)** $\Rightarrow$ **Parallel Imaging (PI)**
Multi-coil reconstruction algorithms: **SENSE**[3] and **GRAPPA**[4].

[3] K. P. Pruessmann et al. (Nov. 1999). "SENSE: Sensitivity encoding for fast MRI". In: *Magnetic Resonance in Medicine* 42.5, pp. 952–962.

[4] 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

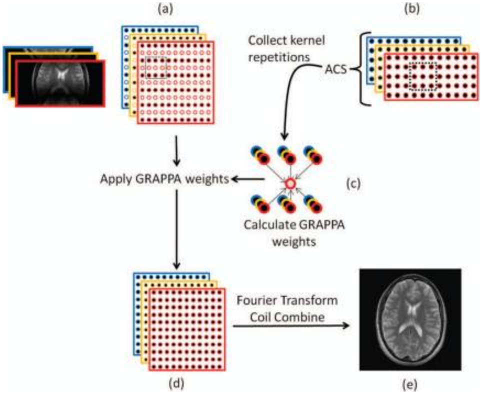


Figure: **GRAPPA illustration.** Image courtesy of Deshmane et al. (2012).

Limits of Parallel Imaging

Resulting acceleration: 2

Compressed Sensing

Linear Inverse Problems

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

Linear Inverse Problems

$$\boldsymbol{A}\,\boldsymbol{x} = \boldsymbol{y}$$

Linear Inverse Problems

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

😀 Signal to reconstruct

Linear Inverse Problems

$$\boldsymbol{A}\,\boldsymbol{x} = \boldsymbol{y}$$

Signal to reconstruct

Linear Inverse Problems

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

Signal to reconstruct

Measurements

Linear Inverse Problems

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

Signal to reconstruct

Measurements

Linear Inverse Problems

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

Measurement operator

Signal to reconstruct

Measurements

Linear Inverse Problems

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

Measurement operator

Signal to reconstruct

Measurements

Problem: what if $\operatorname{Ker} \boldsymbol{A} \neq \{0\}$? Multiple solutions!

Linear Inverse Problems

$$\boldsymbol{A}\,\boldsymbol{x} = \boldsymbol{y}$$

Measurement operator

Signal to reconstruct

Measurements

Problem: what if $\operatorname{Ker} \boldsymbol{A} \neq \{0\}$? Multiple solutions!

To select one of these solutions, we need a priori knowledge.

Another look at redundancy: the prior point of view

Redundancy is not always strict

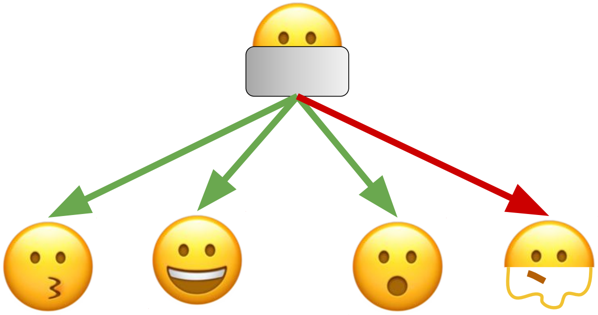
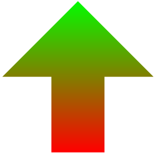


Figure: **A smiley example to a priori knowledge.**

Another look at redundancy: the prior point of view

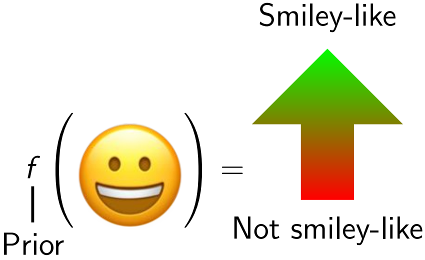
Redundancy is not always strict

$f\Bigg($  $\Bigg) =$



Another look at redundancy: the prior point of view

Redundancy is not always strict



Application to MRI

The Inverse Problem becomes:

$$(\boldsymbol{I}_L \otimes \mathcal{F}_\Omega)\,\mathbb{S}\,\boldsymbol{x} = \mathrm{y}$$

Application to MRI

The Inverse Problem becomes:

$$(\boldsymbol{I}_L \otimes \mathcal{F}_\Omega)\, \mathbb{S}\, \boldsymbol{x} = \mathrm{y}$$

2D or 3D MR image

Application to MRI

The Inverse Problem becomes:

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

$$(\boldsymbol{I}_L \otimes \mathcal{F}_\Omega)\,\mathbb{S}\,\boldsymbol{x} = \mathrm{y}$$

2D or 3D MR image

Application to MRI

The Inverse Problem becomes:

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

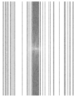
$$\underbrace{(\boldsymbol{I}_L \otimes \mathcal{F}_\Omega)\,\mathbb{S}}\;\boldsymbol{x} = \mathrm{y}$$

$\mathcal{F}_\Omega$ **: FT on the Ω set;**
$\mathbb{S} = [S_1^H, \ldots, S_L^H]^\top$**: the sensitivity maps per coil**

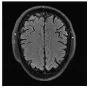
2D or 3D MR image

Application to MRI

The Inverse Problem becomes:



$$(\boldsymbol{I}_L \otimes \mathcal{F}_\Omega)\,\mathbb{S}\,\boldsymbol{x} = \mathrm{y}$$



The canonical MRI reconstruction problem

$$\min_{\boldsymbol{x}\in\mathbb{C}^n} \overbrace{\|\mathcal{A}\boldsymbol{x} - \boldsymbol{y}\|_2^2}^{\text{Noisy data consistency}} + \overbrace{\lambda\|\boldsymbol{\psi}\boldsymbol{x}\|_1}^{\text{Regularization term}}$$

$= (\boldsymbol{I}_L \otimes \mathcal{F}_\Omega)\,\mathbb{S}$

Wavelet basis

Regularization hyperparameter

ISTA

Iterative Shrinkage-Thresholding Algorithm (ISTA):

$$\boldsymbol{x}_{n+1} = \boxed{\boldsymbol{x}_n - \epsilon_n \mathcal{A}^H \left(\mathcal{A}\boldsymbol{x}_n - \mathbb{y} \right)}$$

$$\boldsymbol{x}_{n+1} = \boxed{\mathrm{prox}_{\epsilon_n \mathcal{R}} \left(\boldsymbol{x}_{n+1} \right)}$$

Proximity operator $\quad = \|\psi \cdot\|_1$

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

The prior is a complicated visual function.

The power of Deep Learning

The prior is a complicated visual function.

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

$$f_\theta \left(\right.$$

$$\left. \right) = \text{"DOG"}$$

The power of Deep Learning

The prior is a complicated visual function.

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

f_θ () $=$ "DOG"

Neural network:
a chain of elementary linear & nonlinear functions

Formalism - 1

Supervised learning:

$$\arg\min_{\boldsymbol{\theta} \in \Theta} \sum_{(\boldsymbol{x}_i, \boldsymbol{y}_i) \in \mathcal{D}} \mathcal{L}(f_{\boldsymbol{\theta}}(\boldsymbol{x}_i), \boldsymbol{y}_i, \boldsymbol{\theta})$$

Formalism - 1

Supervised learning:

input

$$\underset{\boldsymbol{\theta} \in \Theta}{\arg\min} \sum_{(\boldsymbol{x}_i, \boldsymbol{y}_i) \in \mathcal{D}} \mathcal{L}(f_{\boldsymbol{\theta}}(\boldsymbol{x}_i), \boldsymbol{y}_i, \boldsymbol{\theta})$$

Formalism - 1

Supervised learning:

input

$$\underset{\boldsymbol{\theta}\in\Theta}{\arg\min} \sum_{(\boldsymbol{x}_i,\boldsymbol{y}_i)\in\mathcal{D}} \mathcal{L}\left(f_{\boldsymbol{\theta}}(\boldsymbol{x}_i), \boldsymbol{y}_i, \boldsymbol{\theta}\right)$$

"DOG", label

Formalism - 1

Supervised learning:

$$\underset{\boldsymbol{\theta}\in\Theta}{\arg\min} \sum_{(\boldsymbol{x}_i,\boldsymbol{y}_i)\in\mathcal{D}} \mathcal{L}\left(\underbrace{f_{\boldsymbol{\theta}}}_{\text{neural network}}(\overbrace{\boldsymbol{x}_i}^{\text{input}}), \underbrace{\boldsymbol{y}_i}_{\text{"DOG", label}}, \boldsymbol{\theta}\right)$$

Formalism - 1

Supervised learning:

$$\arg\min_{\boldsymbol{\theta} \in \Theta} \sum_{(\boldsymbol{x}_i, \boldsymbol{y}_i) \in \mathcal{D}} \overset{\text{loss}}{\mathcal{L}}\Big(\underset{\text{neural network}}{f_{\boldsymbol{\theta}}}(\overset{\text{input}}{\boldsymbol{x}_i}), \underset{\text{"DOG", label}}{\boldsymbol{y}_i}, \boldsymbol{\theta}\Big)$$

Formalism - 1

Supervised learning:

$$\underset{\boldsymbol{\theta}\in\Theta}{\arg\min} \underbrace{\sum_{(\boldsymbol{x}_i,\boldsymbol{y}_i)\in\mathcal{D}}}_{\text{Estimator of the expected value}} \overbrace{\mathcal{L}}^{\text{loss}}\Big(\underbrace{f_{\boldsymbol{\theta}}}_{\text{neural network}}\big(\overbrace{\boldsymbol{x}_i}^{\text{input}}\big), \underbrace{\boldsymbol{y}_i}_{\text{"DOG", label}}, \boldsymbol{\theta}\Big)$$

Formalism - 1

Supervised learning:

$$\underset{\underbrace{\boldsymbol{\theta} \in \Theta}_{\text{Parameters}}}{\arg\min} \underbrace{\sum_{(\boldsymbol{x}_i,\boldsymbol{y}_i)\in\mathcal{D}}}_{\text{Estimator of the expected value}} \overbrace{\mathcal{L}}^{\text{loss}}\Big(\underbrace{f_{\boldsymbol{\theta}}}_{\text{neural network}}(\overbrace{\boldsymbol{x}_i}^{\text{input}}), \underbrace{\boldsymbol{y}_i}_{\text{"DOG", label}}, \boldsymbol{\theta}\Big)$$

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 functions.

Deep Learning for MRI reconstruction

Model agnostic learning

Let's throw away all we know:[5]

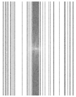
$$f_{\boldsymbol{\theta}}(\,\mathbb{y}\,) = \boldsymbol{x}$$

[5] B. Zhu et al. (Mar. 2018). "Image reconstruction by domain-transform manifold learning". In: *Nature* 555.7697, pp. 487–492.

Model agnostic learning

Let's throw away all we know:[5]

$$f_{\boldsymbol{\theta}}(\mathbb{y}) = \boldsymbol{x}$$



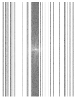
kspace

[5] B. Zhu et al. (Mar. 2018). "Image reconstruction by domain-transform manifold learning". In: *Nature* 555.7697, pp. 487–492.

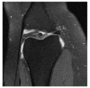
Model agnostic learning

Let's throw away all we know:[5]

$$f_{\boldsymbol{\theta}}(\mathbb{y}) = \boldsymbol{x}$$



kspace



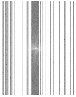
image

[5] B. Zhu et al. (Mar. 2018). "Image reconstruction by domain-transform manifold learning". In: *Nature* 555.7697, pp. 487–492.

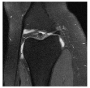
Single domain learning

Let's use $\mathcal{A}^H$ to build a more informed model:

$$\mathcal{A}^H f_{\boldsymbol{\theta}}(\mathrm{y}) = \boldsymbol{x}$$



kspace



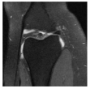
image

Single domain learning

Let's use $\mathcal{A}^H$ to build a more informed model:

$$f_{\boldsymbol{\theta}}(\mathcal{A}^H \, \mathbb{y}) = \boldsymbol{x}$$

aliased image



image

Unrolled models - 1

We can mix the 2 single domain approaches, using the principled **optimization algorithm unrolling** method.[6]
A graph representation of ISTA:

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

$$\boldsymbol{x}_{n+1} = \mathrm{prox}_{\epsilon_n \mathcal{R}} \left(\boldsymbol{x}_{n+1}\right)$$

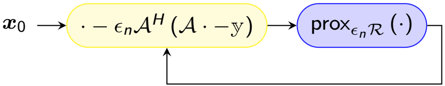
[6] K. Gregor et al. (2010). "Learning fast approximations of sparse coding". In: *ICML 2010 - Proceedings, 27th International Conference on Machine Learning*, pp. 399–406.

Unrolled models - 1

We can mix the 2 single domain approaches, using the principled **optimization algorithm unrolling** method.[6]
A graph representation of ISTA:

$$\boldsymbol{x}_{n+1} = \boxed{\boldsymbol{x}_n - \epsilon_n \mathcal{A}^H \left(\mathcal{A} \boldsymbol{x}_n - \mathrm{y} \right)}$$

$$\boldsymbol{x}_{n+1} = \boxed{\mathrm{prox}_{\epsilon_n \mathcal{R}} \left(\boldsymbol{x}_{n+1} \right)}$$



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

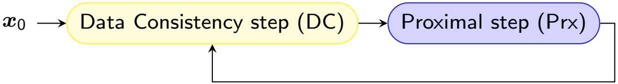
Unrolled models - 1

We can mix the 2 single domain approaches, using the principled **optimization algorithm unrolling** method.[6]

A graph representation of ISTA:

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

$$\boldsymbol{x}_{n+1} = \mathrm{prox}_{\epsilon_n \mathcal{R}} \left(\boldsymbol{x}_{n+1}\right)$$



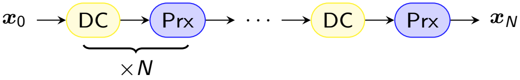
[6] K. Gregor et al. (2010). "Learning fast approximations of sparse coding". In: *ICML 2010 - Proceedings, 27th International Conference on Machine Learning*, pp. 399–406.

Unrolled models - 1

We can mix the 2 single domain approaches, using the principled **optimization algorithm unrolling** method.[6]
A graph representation of ISTA:

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

$$\boldsymbol{x}_{n+1} = \mathrm{prox}_{\epsilon_n \mathcal{R}} \left(\boldsymbol{x}_{n+1} \right)$$



[6] 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.[6]

$$\boldsymbol{x}_{n+1} = \boxed{\boldsymbol{x}_n - \epsilon_n \mathcal{A}^H \left(\mathcal{A}\boldsymbol{x}_n - \mathbb{y}\right)}$$

$$\boldsymbol{x}_{n+1} = \boxed{f_{\boldsymbol{\theta}_n}\left(\boldsymbol{x}_{n+1}\right)}$$

$$\boldsymbol{x}_0 \longrightarrow \underbrace{\text{DC} \longrightarrow f_{\boldsymbol{\theta}_1}}_{\times N} \longrightarrow \cdots \longrightarrow \text{DC} \longrightarrow f_{\boldsymbol{\theta}_N} \longrightarrow \boldsymbol{x}_N$$

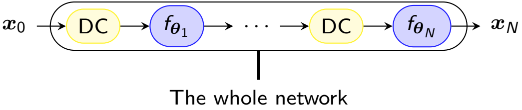
[6] K. Gregor et al. (2010). "Learning fast approximations of sparse coding". In: *ICML 2010 - Proceedings, 27th International Conference on Machine Learning*, pp. 399–406.

Unrolled models - 1

We can mix the 2 single domain approaches, using the principled **optimization algorithm unrolling** method.[6]

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

$$\boldsymbol{x}_{n+1} = f_{\boldsymbol{\theta}_n} \left(\boldsymbol{x}_{n+1} \right)$$



[6] 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 #1

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

Different models based on:

- optimization algorithm to unroll
- choice of f_{θ}
- N

Unrolled models - 2

Contribution #1

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

Different models based on:

- optimization algorithm to unroll
- choice of f_θ
- N

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[7]
PSNR	29.61	31.38	31.78	31.97	**32.15**

[7] J. Adler and O. Öktem (2018). "Learned Primal-Dual Reconstruction". In: *IEEE Transactions on Medical Imaging* 37.6, pp. 1322–1332

Unrolled models - 2

Contribution #1

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

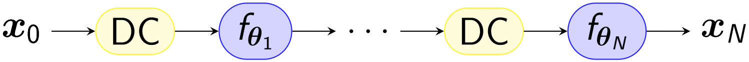
Different models based on:

- optimization algorithm to unroll
- choice of f_θ
- N

- Code available online: `github.com/zacchariermazi/fastmri-reproducible-benchmark`
- 🤗 Model weights available online: `huggingface.co/zaccharieramzi`

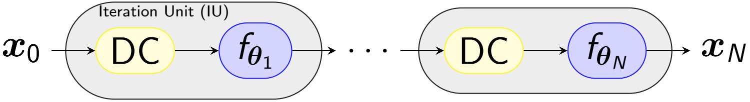
XPDNet

XPDNet

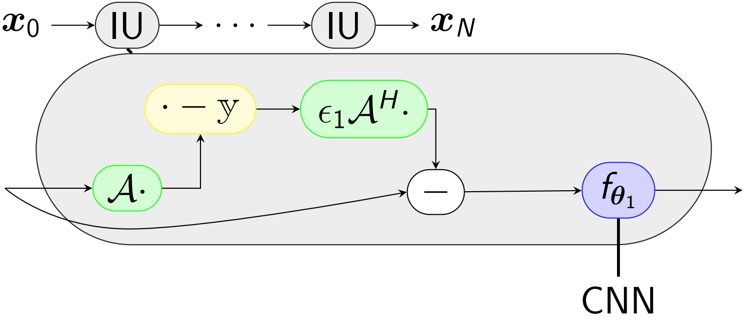


$\boldsymbol{x}_0$
DC
$f_{\boldsymbol{\theta}_1}$
$\cdots$
DC
$f_{\boldsymbol{\theta}_N}$
$\boldsymbol{x}_N$

XPDNet



XPDNet

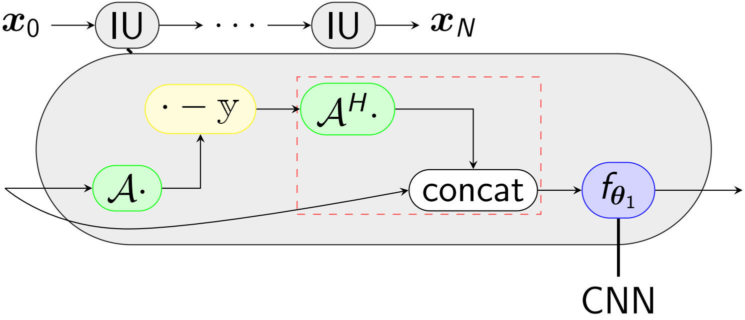


[8] P. Liu et al. (2018). “Multi-level Wavelet-CNN for Image Restoration”. In: *CVPR NTIRE Workshop*.

[9] A. Sriram et al. (2020). “End-to-End Variational Networks for Accelerated MRI Reconstruction”. In: *MICCAI*.

[10] 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

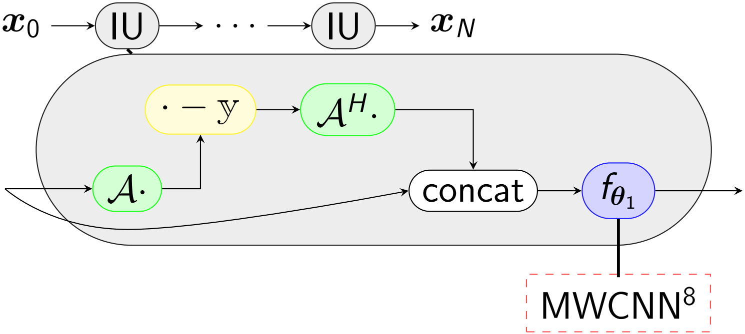


[8] P. Liu et al. (2018). "Multi-level Wavelet-CNN for Image Restoration". In: *CVPR NTIRE Workshop*.

[9] A. Sriram et al. (2020). "End-to-End Variational Networks for Accelerated MRI Reconstruction". In: *MICCAI*.

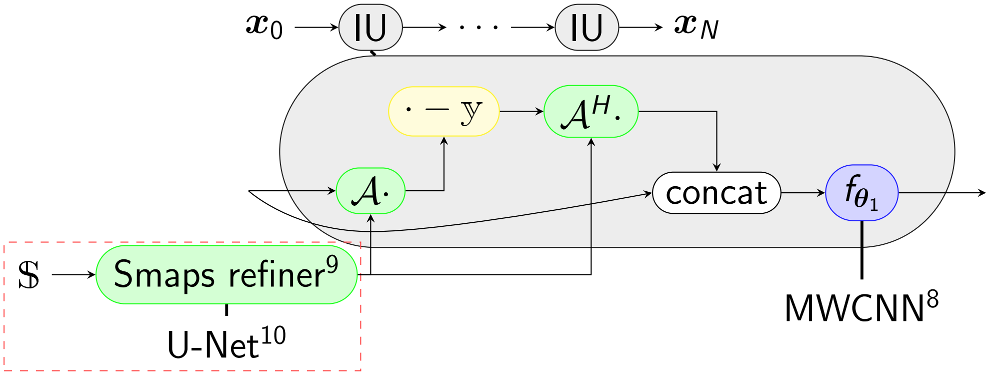
[10] O. Ronneberger et al. (2015). "U-net: Convolutional networks for biomedical image segmentation". In: *International Conference on Medical image computing and computer-assisted intervention*.

XPDNet



[8] P. Liu et al. (2018). "Multi-level Wavelet-CNN for Image Restoration". In: *CVPR NTIRE Workshop*.
[9] A. Sriram et al. (2020). "End-to-End Variational Networks for Accelerated MRI Reconstruction". In: *MICCAI*.
[10] 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



[8] P. Liu et al. (2018). "Multi-level Wavelet-CNN for Image Restoration". In: *CVPR NTIRE Workshop*.

[9] A. Sriram et al. (2020). "End-to-End Variational Networks for Accelerated MRI Reconstruction". In: *MICCAI*.

[10] 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 #2

- 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

fastMRI challenge

Contributions #2

- 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

---

- Data: fastMRI
- Compute: Jean Zay

fastMRI challenge

Contributions #2

- 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

XPDNet in a prospective out-of-distribution setting:
different orientation, higher resolution, higher field strength, lower acceleration factor, presence of the cerebellum.[11]

[11] For anonymity reasons, the cerebellum is not present in the fastMRI dataset.

[12] 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

XPDNet in a prospective out-of-distribution setting:
different orientation, higher resolution, higher field strength, lower acceleration factor, presence of the cerebellum.[11]

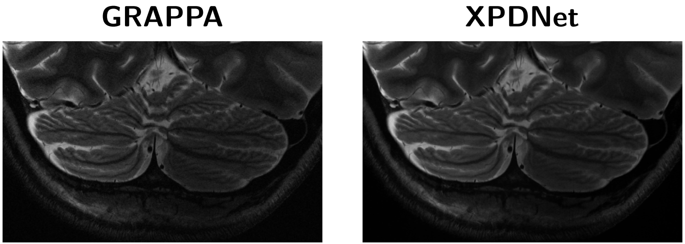


Figure: **XPDNet reconstruction on a brain prospectively accelerated.**[12] (zoom on the cerebellum)

---

[11] For anonymity reasons, the cerebellum is not present in the fastMRI dataset.

[12] 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

Non-Cartesian acquisitions better cover the k-space.

Non-Cartesian acquisitions

Non-Cartesian acquisitions better cover the k-space.

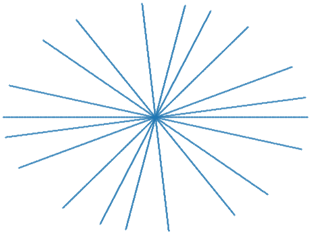


Figure: **Radial undersampled trajectory.**

Non-Cartesian acquisitions

Non-Cartesian acquisitions better cover the k-space.

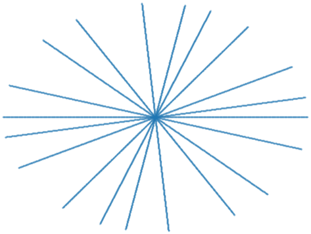
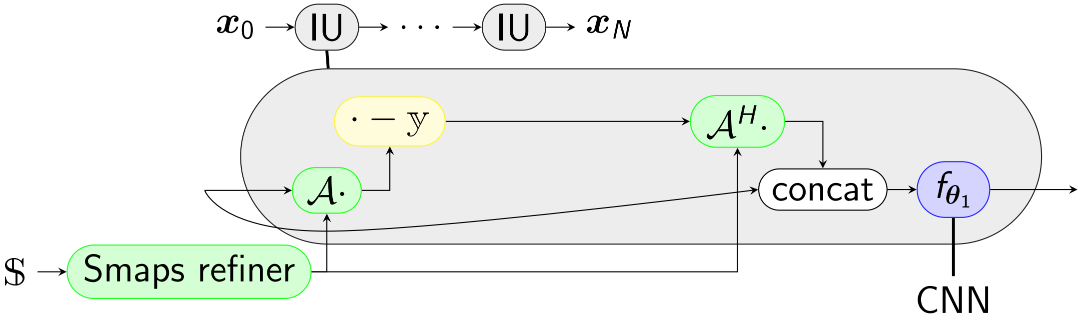


Figure: **Radial undersampled trajectory.**

Nonuniform Fourier Transform (NDFT) too costly $\Rightarrow$ NUFFT, with the TensorFlow implementation:

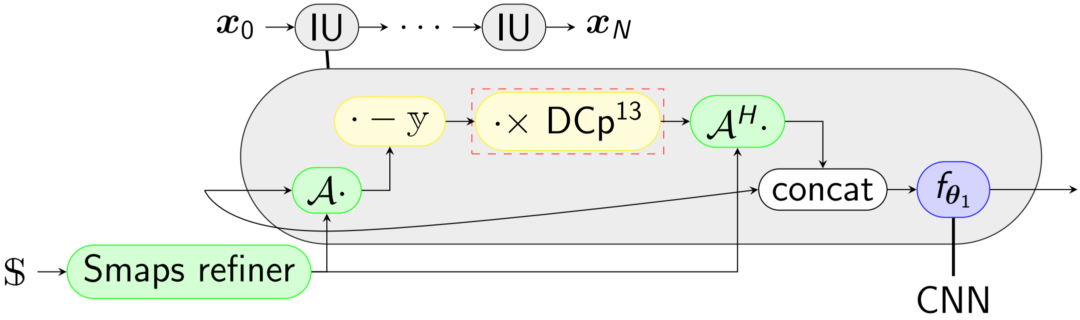
- Code available online: `github.com/zacchariemazi/tfkbnufft`

NC-PDNet - 1



[13] 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



[13] 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 #3

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

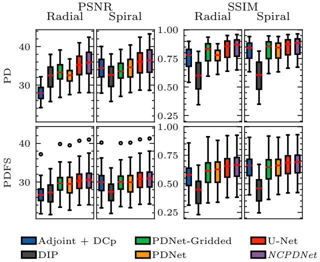


Figure: **2D single-coil reconstruction quantitative results on the fastMRI knee dataset for non-Cartesian trajectories.**

NC-PDNet - 2

Contribution #3

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

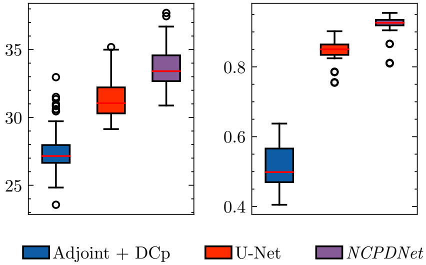


Figure: **3D single-coil reconstruction quantitative results on the OASIS dataset for a radial trajectory.**

NC-PDNet - 2

Contribution #3

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

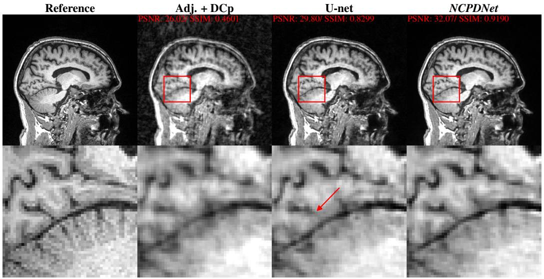


Figure: **3D single-coil reconstruction qualitative results on the OASIS dataset for a radial trajectory.**

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?

With deeper models comes better performance.

Why should we go deep?

With deeper models comes better performance.

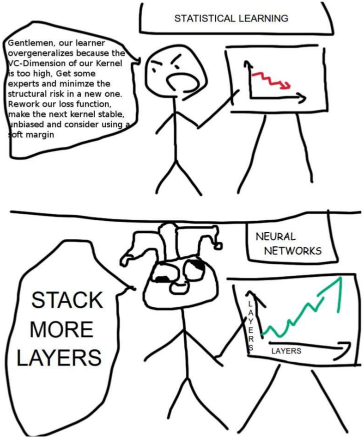


Figure: Credits: reddit.com/r/ProgrammerHumor/comments/5si1f0/machine_learning_approaches/

Why should we go deep?

With deeper models comes better performance.

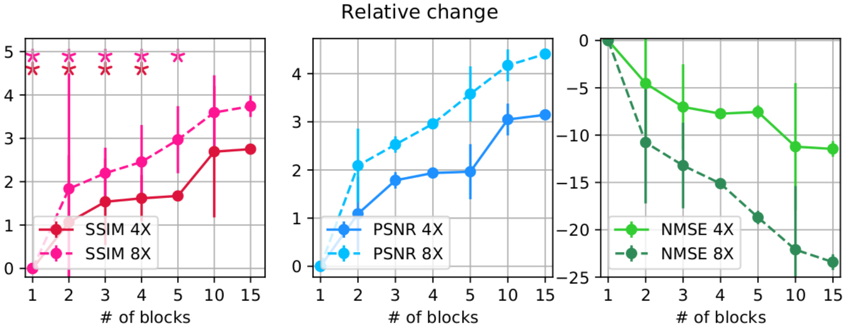


Figure: **Performance of an unrolled MRI reconstruction network function of the number of iteration units (blocks).**[15]

[15] 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?

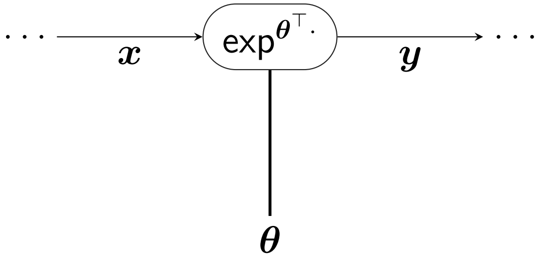
Deeper models $\Rightarrow$ more **memory** to train. Memory we have is constrained.

Can we go deeper?

Deeper models $\Rightarrow$ more **memory** to train. Memory we have is constrained.
Why more memory? **activations**!

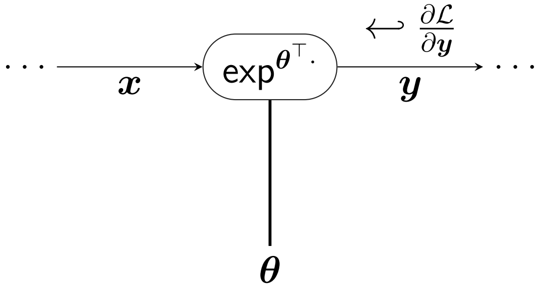
Can we go deeper?

Deeper models $\Rightarrow$ more **memory** to train. Memory we have is constrained.
Why more memory? **activations**!



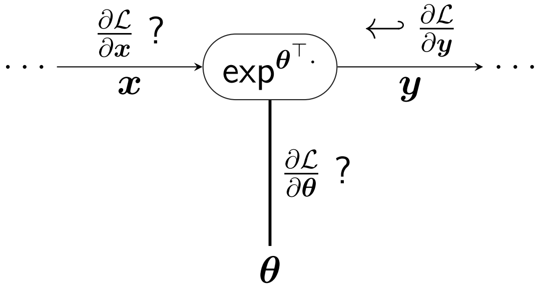
Can we go deeper?

Deeper models $\Rightarrow$ more **memory** to train. Memory we have is constrained.
Why more memory? **activations**!



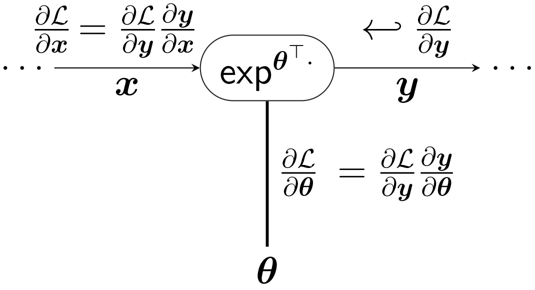
Can we go deeper?

Deeper models $\Rightarrow$ more **memory** to train. Memory we have is constrained.
Why more memory? **activations**!



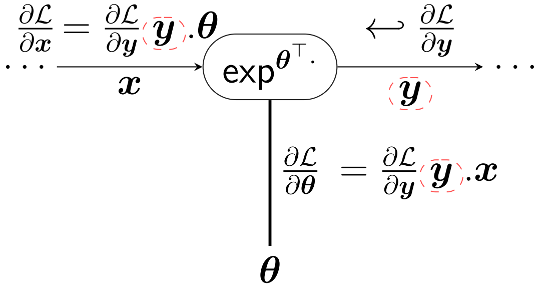
Can we go deeper?

Deeper models $\Rightarrow$ more **memory** to train. Memory we have is constrained.
Why more memory? **activations**!



Can we go deeper?

Deeper models $\Rightarrow$ more **memory** to train. Memory we have is constrained.
Why more memory? **activations**!



The modeling solutions

Memory-efficient training:

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

The modeling solutions

Memory-efficient training:

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

The modeling solutions

Memory-efficient training:

- 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

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

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

Deep Equilibrium networks - 1

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

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

In practice, solved with a **quasi-Newton method**.

Deep Equilibrium networks - 1

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

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

In practice, solved with a **quasi-Newton method**.

Deep Equilibrium networks - 2

How do I compute the gradient $\frac{\partial \mathcal{L}}{\partial \boldsymbol{\theta}}$?

Deep Equilibrium networks - 2

How do I compute the gradient $\frac{\partial \mathcal{L}}{\partial \boldsymbol{\theta}}$?
The **Implicit Function Theorem** gives us just that:

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

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

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

Deep Equilibrium networks - 2

How do I compute the gradient $\frac{\partial \mathcal{L}}{\partial \boldsymbol{\theta}}$?
The **Implicit Function Theorem** gives us just that:

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

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

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

Does not rely on activations!

The limits of DEQs

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

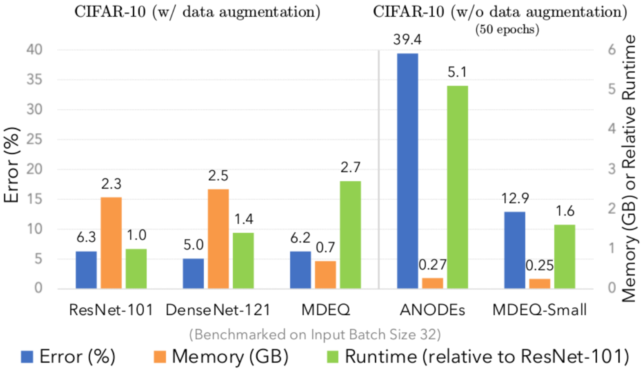


Figure: **Performance, memory and training speed of DEQs.** (Bai, Koltun, et al., 2020)

Why are DEQs slow?

DEQs gradient computation:

$$\frac{\partial \mathcal{L}}{\partial \boldsymbol{\theta}}\bigg|_{\boldsymbol{z}^\star} = \nabla_{\boldsymbol{z}} \mathcal{L}(\boldsymbol{z}^\star)^\top \, J_{g_{\boldsymbol{\theta}}}(\boldsymbol{z}^\star)^{-1} \frac{\partial g_{\boldsymbol{\theta}}}{\partial \boldsymbol{\theta}}\bigg|_{\boldsymbol{z}^\star},$$

we need to invert a huge matrix $J_{g_{\boldsymbol{\theta}}}(\boldsymbol{z}^\star)$ in a certain direction $\nabla_{\boldsymbol{z}} \mathcal{L}(\boldsymbol{z}^\star)$.

Why are DEQs slow?

DEQs gradient computation:

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

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

Can we avoid the Jacobian inversion?

Contribution #4

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**.

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

Can we avoid the Jacobian inversion?

Contribution #4

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**.

$$\boldsymbol{B}^{-1} \approx \underbrace{J_{g_{\boldsymbol{\theta}}}(\boldsymbol{z}^\star)^{-1}}_{\text{True Jacobian inverse}}$$

Can we avoid the Jacobian inversion?

Contribution #4

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**.

$$\underbrace{\boldsymbol{B}^{-1}}_{\text{quasi-Newton matrix}} \approx \underbrace{J_{g_{\boldsymbol{\theta}}}(\boldsymbol{z}^\star)^{-1}}_{\text{True Jacobian inverse}}$$

Can we avoid the Jacobian inversion?

Contribution #4

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**.

$$\underbrace{\boldsymbol{B}^{-1}}_{\text{quasi-Newton matrix}} \approx \underbrace{J_{g_{\boldsymbol{\theta}}}(\boldsymbol{z}^\star)^{-1}}_{\text{True Jacobian inverse}}$$

Properties of $\boldsymbol{B}$:

- It is computed when solving $g_{\boldsymbol{\theta}}(\boldsymbol{z}^\star, \boldsymbol{x}) = 0$ using a quasi-Newton method.

Can we avoid the Jacobian inversion?

Contribution #4

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**.

$$\underbrace{\boldsymbol{B}^{-1}}_{\text{quasi-Newton matrix}} \approx \underbrace{J_{g_{\boldsymbol{\theta}}}(\boldsymbol{z}^{\star})^{-1}}_{\text{True Jacobian inverse}}$$

Properties of $\boldsymbol{B}$:

- It is computed when solving $g_{\boldsymbol{\theta}}(\boldsymbol{z}^{\star}, \boldsymbol{x}) = 0$ using a quasi-Newton method.
- It is easily invertible using the Sherman-Morrison formula.

Application to Hyperparameter optimization - 1

Hyperparameter optimization can benefit from SHINE.

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

Application to Hyperparameter optimization - 1

Hyperparameter optimization can benefit from SHINE.

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

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

Application to Hyperparameter optimization - 2

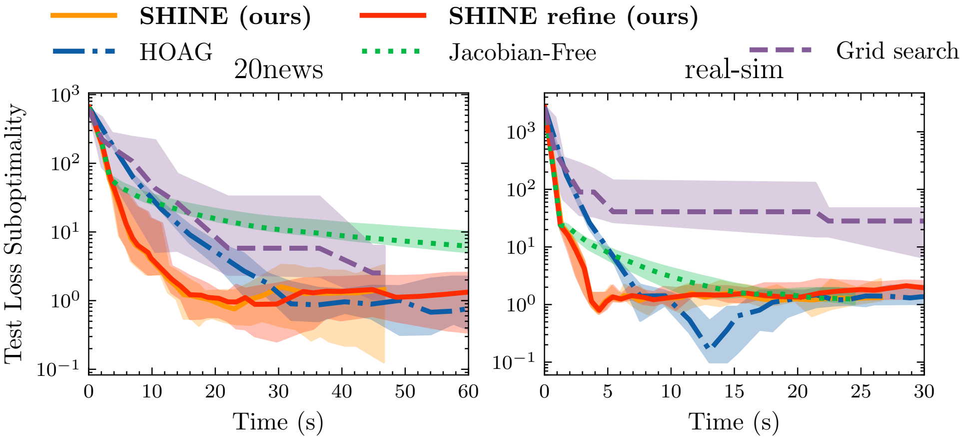


Figure: **Bilevel optimization (Pedregosa, 2016) with SHINE:** convergence of held-out test loss.

Results on DEQs

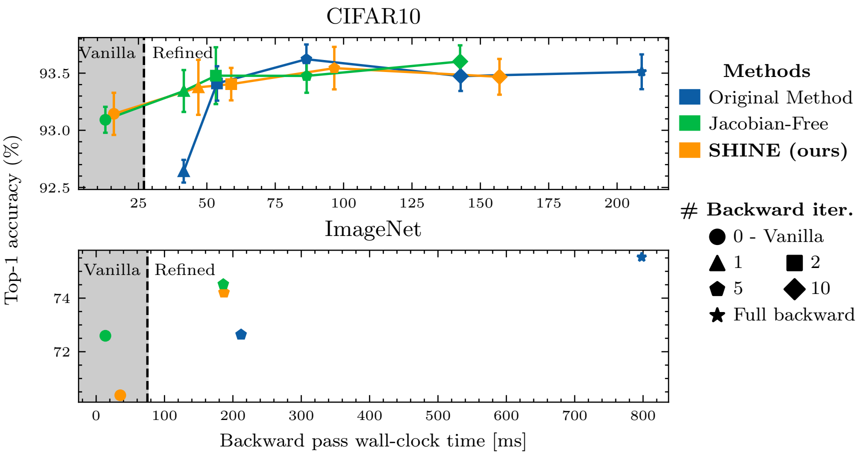


Figure: **MDEQs (Bai, Koltun, et al., 2020) with SHINE.**

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.
   3.1 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.
   3.1 The XPDNet, a deep learning network that secured the 2nd spot of the 2020 fastMRI challenge.
   3.2 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.
   3.1 The XPDNet, a deep learning network that secured the 2nd spot of the 2020 fastMRI challenge.
   3.2 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

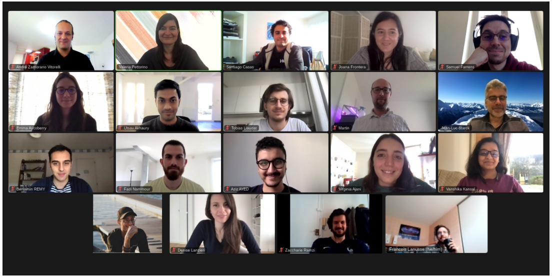
- **Zaccharie Ramzi**, K. Michalewicz, J. L. Starck, T. Moreau, and P. Ciuciu (2021). "Wavelets in the deep learning era". Under review in Journal of Mathematical Imaging and Vision
- **Zaccharie Ramzi**, B. Remy, F. Lanusse, J.-L. Starck, and P. Ciuciu (2020). "Denoising Score-Matching for Uncertainty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

Miscellanous contributions

Contributions

- Jean Zay user doc: jean-zay-doc.readthedocs.io
- NeuroSpin Deep Learning lecture group

Thank you all!



Backup slides

quasi-Newton methods

References I

- 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

- 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.
- Krantz, S. G. and H. R. Parks (Jan. 2013). *The Implicit Function Theorem: History, Theory, and Applications*. Springer New York, pp. 1–163.
- Liu, P., H. Zhang, K. Zhang, L. Lin, and W. Zuo (2018). "Multi-level Wavelet-CNN for Image Restoration". In: *CVPR NTIRE Workshop*.

References III

- 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.
- 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.

References IV

Pezzotti, N., S. Yousefi, M. S. Elmahdy, J. H. F. van Gemert, C. Schuelke, M. Doneva, T. Nielsen, S. Kastryulin, 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.

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.

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.*

References V

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*.

NHS: How it's performed - MRI scan (2018). https://www.nhs.uk/conditions/mri-scan/what-happens/. Accessed: 2021-10-11.

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.

References VI

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).

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 VII

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.