



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 chance of motion as time progresses

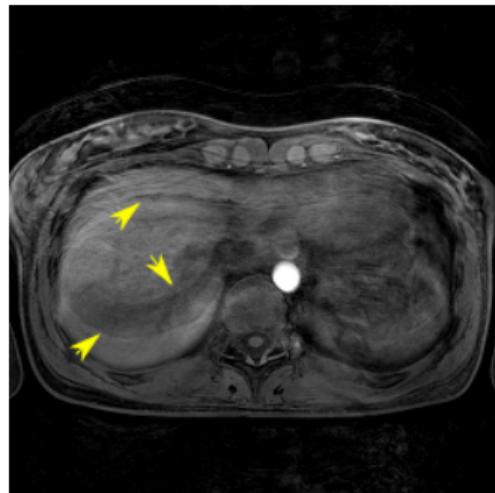


Figure: Example of motion artifacts in MRI²

¹ NHS: How it's performed - MRI scan (2018).

<https://www.nhs.uk/conditions/mri-scan/what-happens/>. Accessed: 2021-10-11.

²R. Grimm (2015). Reconstruction Techniques for Dynamic Radial MRI.
Friedrich-Alexander-Universitaet Erlangen-Nuernberg (Germany).

Our objective: accelerate MRI scans

1. Study and extension of the state of the art: unrolled models.
2. Tools to tackle deep learning blind spots: generalization and uncertainty quantification.
3. Methods to build even deeper models.

1. Magnetic Resonance Imaging (MRI)

What does an MRI look like?

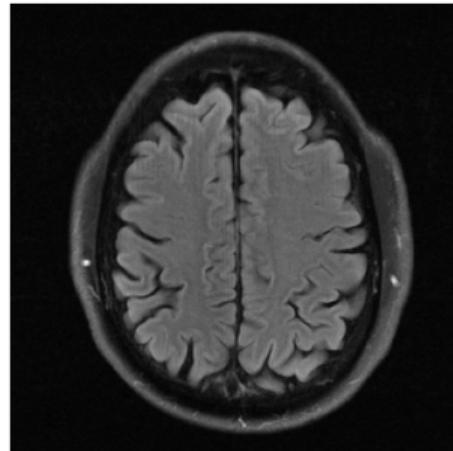
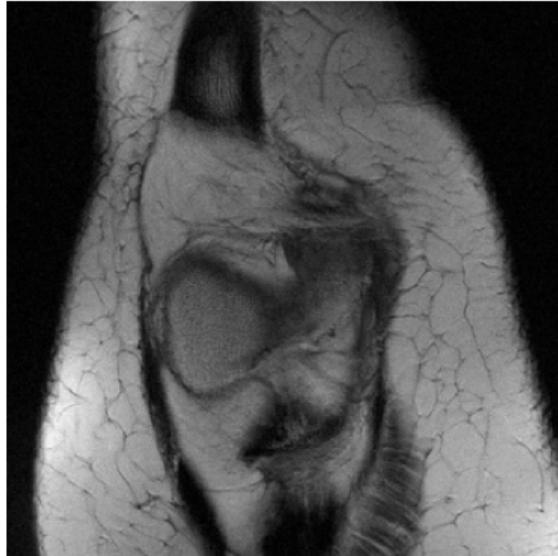
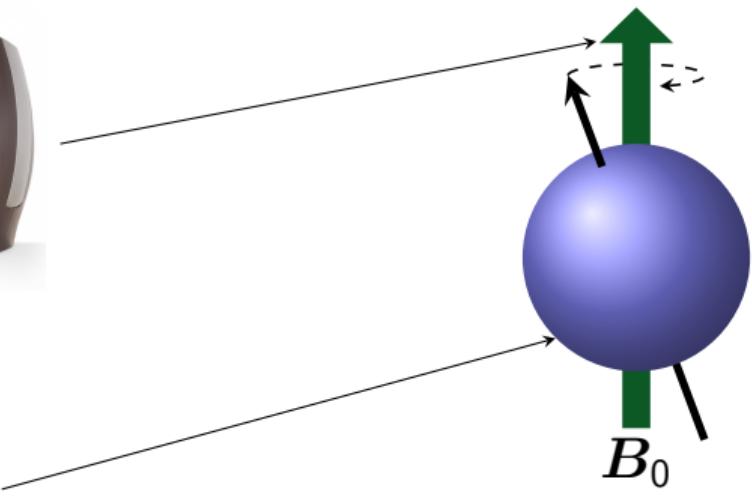
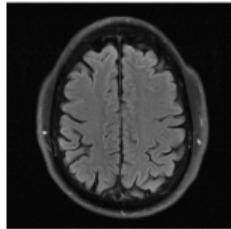


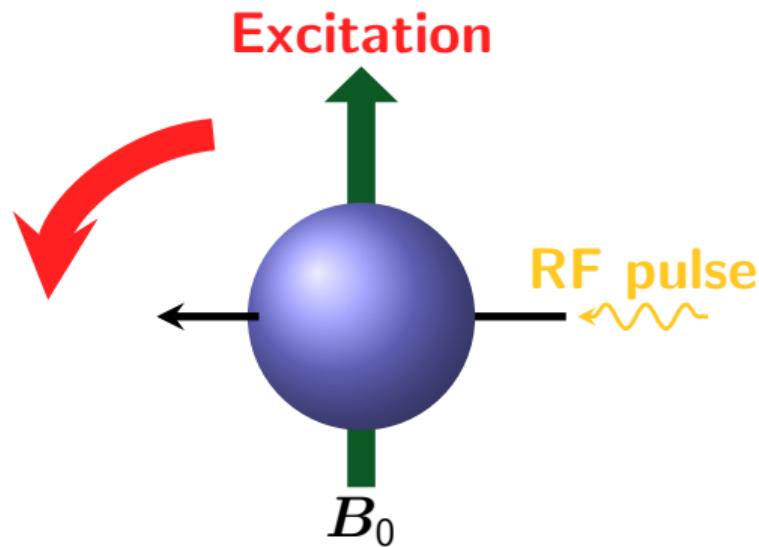
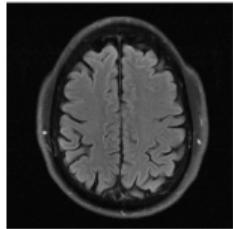
Figure: Examples of MR images: knee and brain taken from the fastMRI dataset.³

³J. Zbontar et al. (2018). *fastMRI: An Open Dataset and Benchmarks for Accelerated MRI*. Tech. rep.

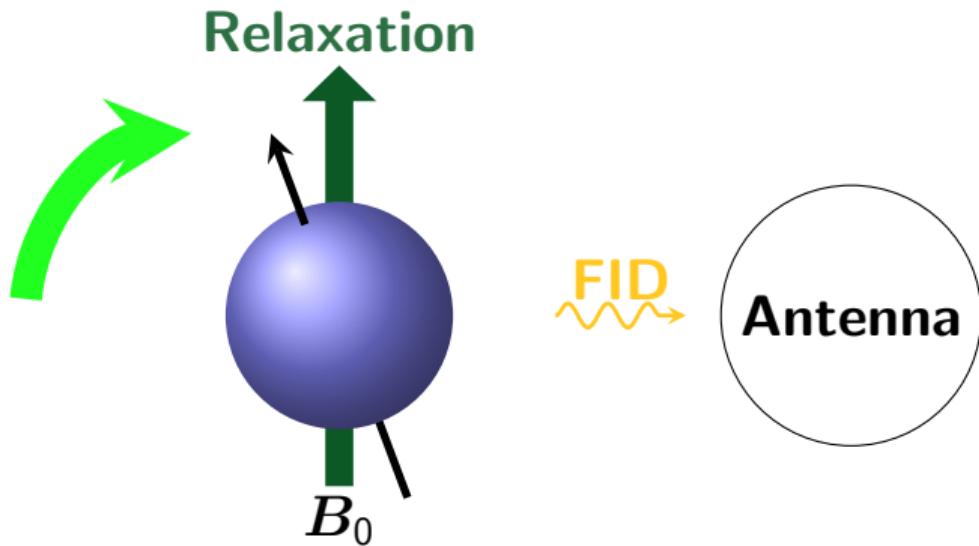
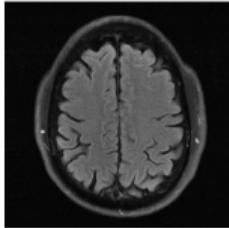
Physics of MRI: Nuclear Magnetic Resonance and k-space



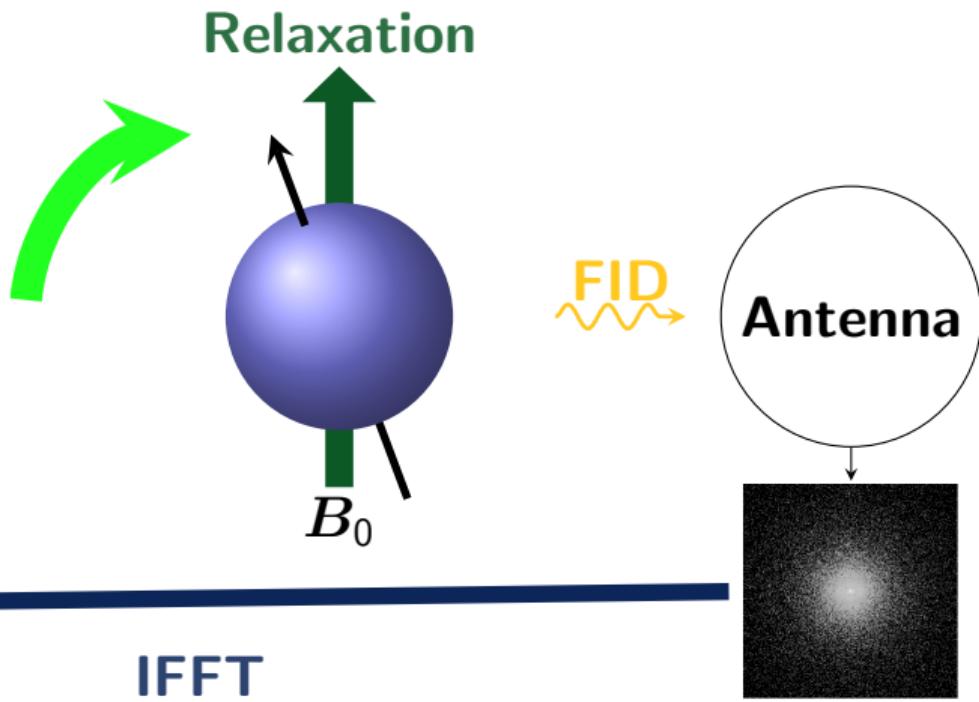
Physics of MRI: Nuclear Magnetic Resonance and k-space



Physics of MRI: Nuclear Magnetic Resonance and k-space

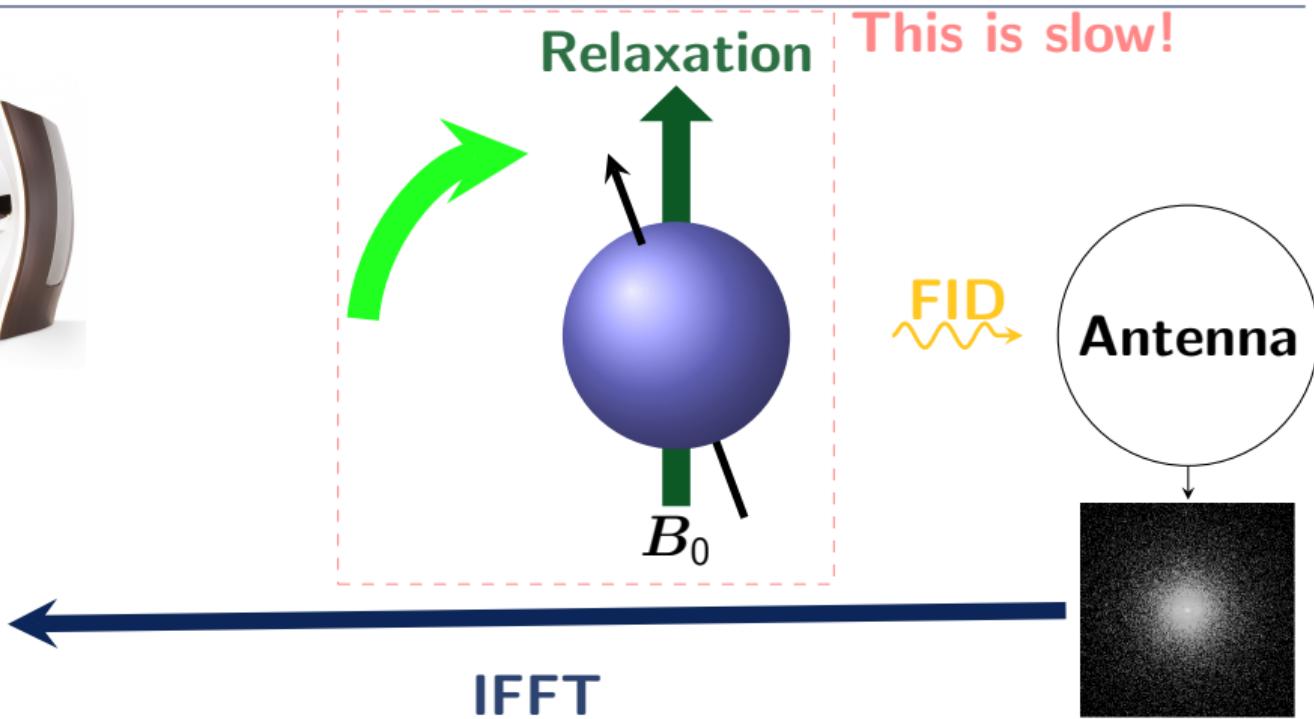
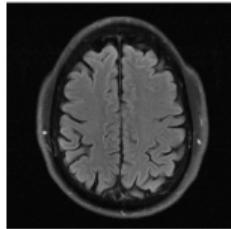


Physics of MRI: Nuclear Magnetic Resonance and k-space



FID: Free Induction Decay, IFFT: Inverse Fast Fourier Transform

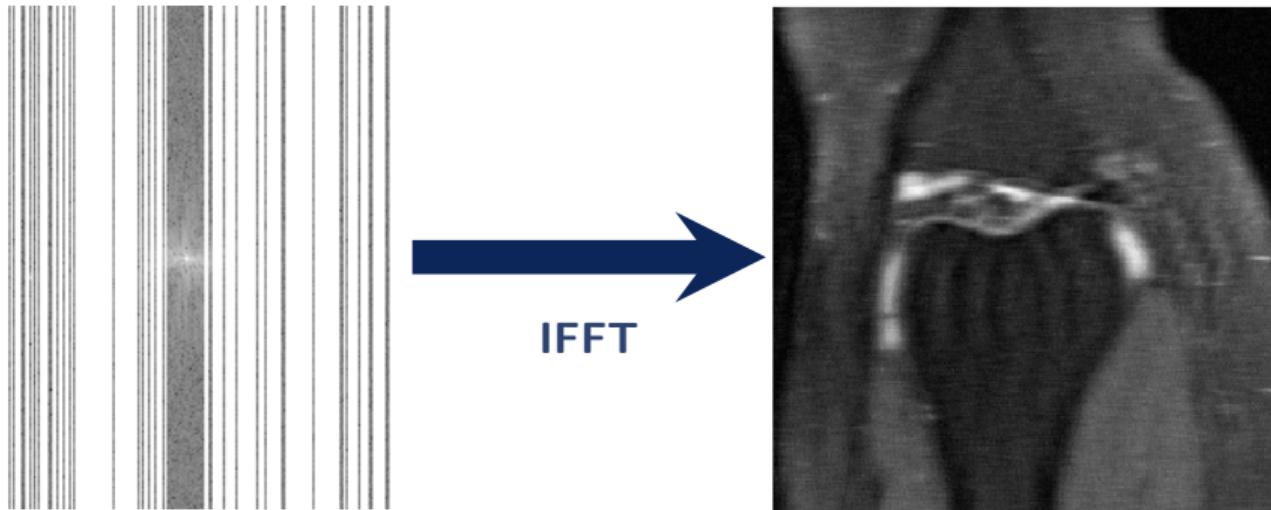
Physics of MRI: Nuclear Magnetic Resonance and k-space



FID: Free Induction Decay, IFFT: Inverse Fast Fourier Transform

Acceleration

Naively: sample fewer lines in the k-space.

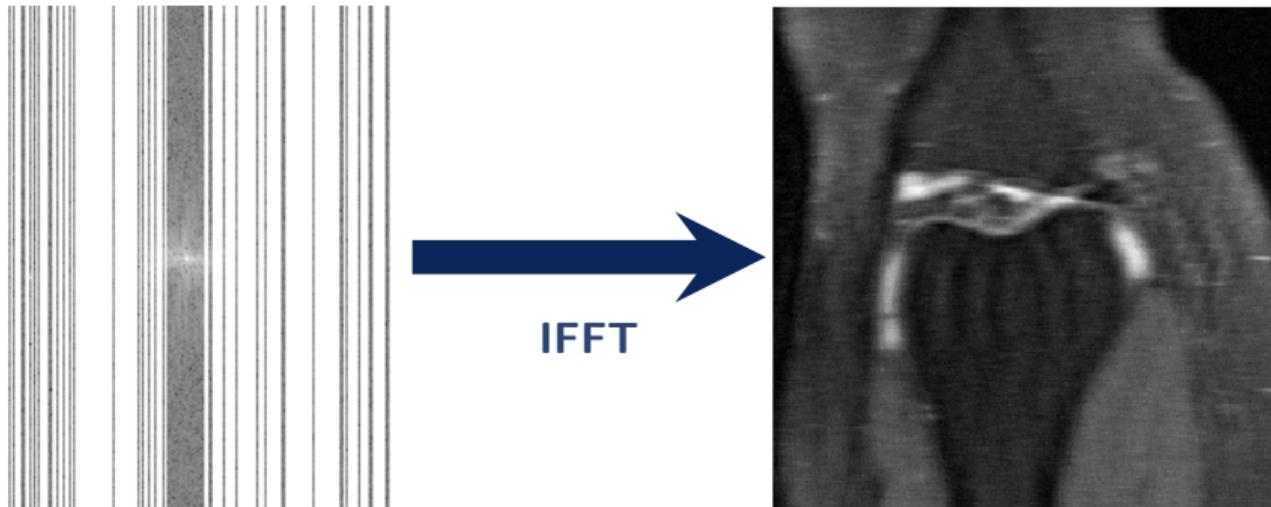


Undersampled k-space

Aliased image

Acceleration

Naively: sample fewer lines in the k-space.



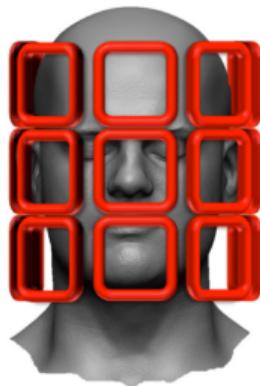
Undersampled k-space

Aliased image

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

Parallel imaging

More redundancy using **more antennas (called coils)** \Rightarrow **Parallel Imaging (PI)**



Multi-coil reconstruction algorithms: **SENSE**⁴ and **GRAPPA**⁵.

⁴K. P. Pruessmann et al. (Nov. 1999). "SENSE: Sensitivity encoding for fast MRI". In: *Magnetic Resonance in Medicine* 42.5, pp. 952–962.

⁵M. A. Griswold et al. (June 2002). "Generalized Autocalibrating Partially Parallel Acquisitions (GRAPPA)". In: *Magnetic Resonance in Medicine* 47.6, pp. 1202–1210.

Limits of Parallel Imaging

Acceleration function of the number of coils.

Resulting acceleration: 2 in 2D, 8 in 3D.

2. Inverse Problems

Linear Inverse Problems

$$\textcolor{green}{A} \textcolor{purple}{x} = \textcolor{yellow}{y}$$

Linear Inverse Problems

$$A \ x = y$$


Linear Inverse Problems

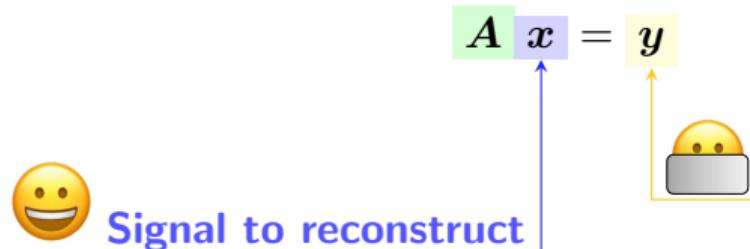
$$\mathbf{A} \mathbf{x} = \mathbf{y}$$



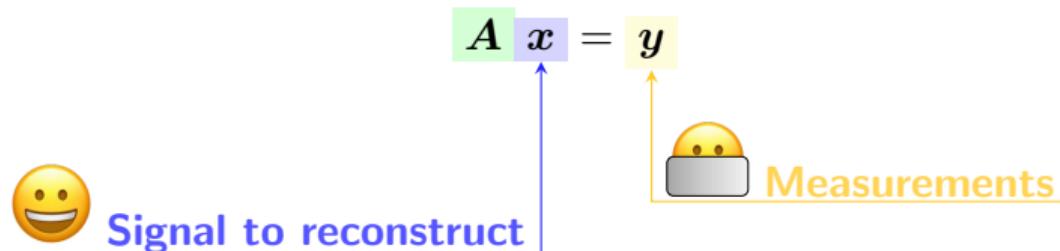
Signal to reconstruct



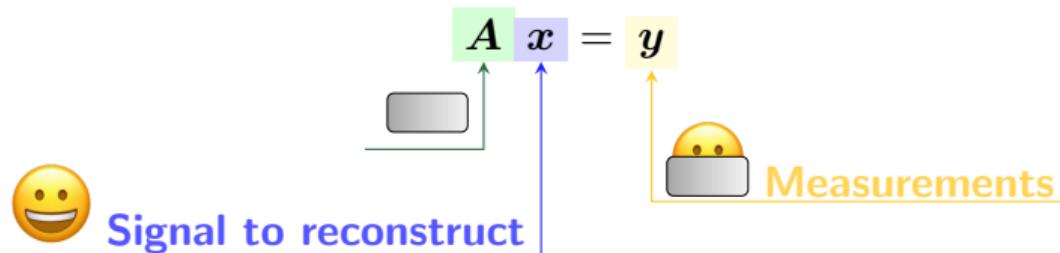
Linear Inverse Problems



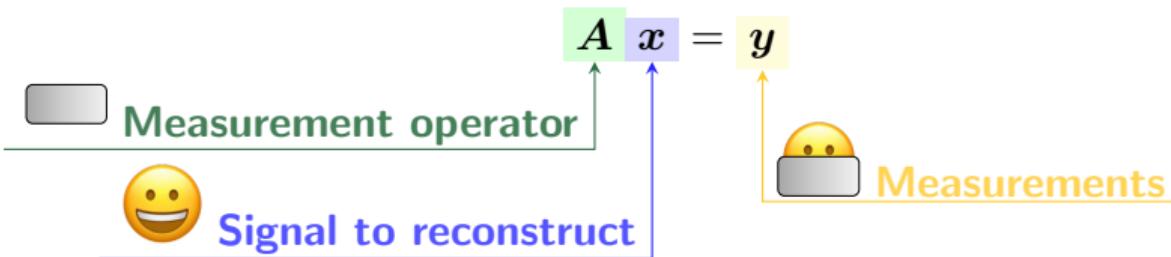
Linear Inverse Problems



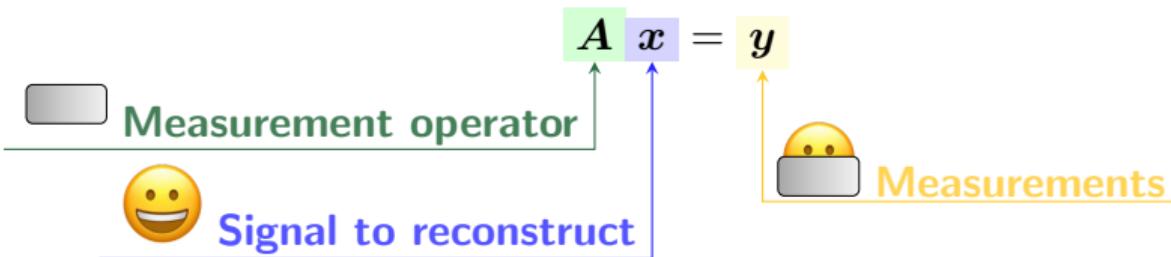
Linear Inverse Problems



Linear Inverse Problems

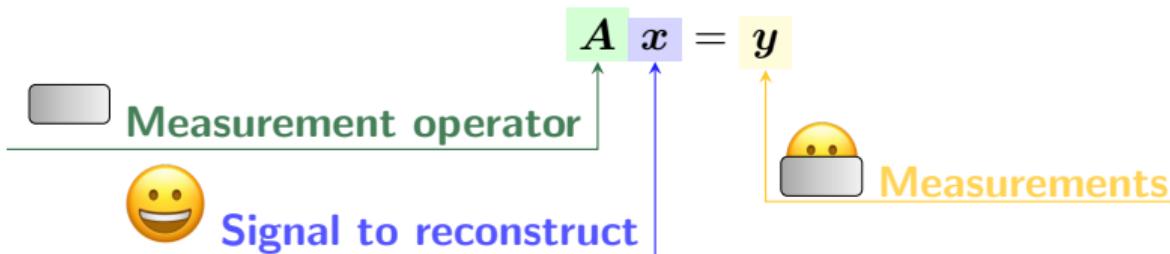


Linear Inverse Problems



Problem: Multiple solutions!

Linear Inverse Problems



Problem: Multiple solutions!

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

Another look at redundancy

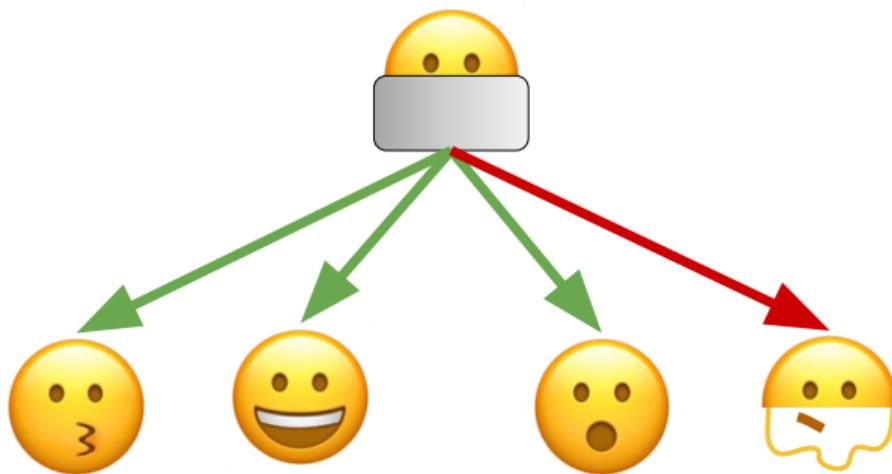


Figure: A smiley example to a priori knowledge.

Another look at redundancy

$$f \left(\begin{matrix} \text{Smiley-like} \\ \text{Regularizer} \end{matrix} \right) = \begin{matrix} \text{Not smiley-like} \\ \uparrow \text{Smiley-like} \end{matrix}$$

Application to MRI

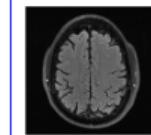
The Inverse Problem becomes:

$$\mathcal{F}_\Omega \$ \textcolor{purple}{x} = \textcolor{yellow}{y}$$

Application to MRI

The Inverse Problem becomes:

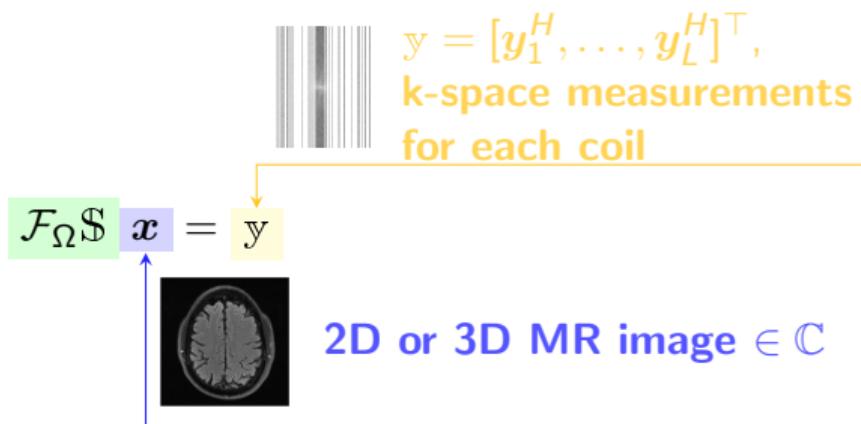
$$\mathcal{F}_\Omega \$ \quad \textcolor{blue}{x} = \textcolor{yellow}{y}$$



2D or 3D MR image $\in \mathbb{C}$

Application to MRI

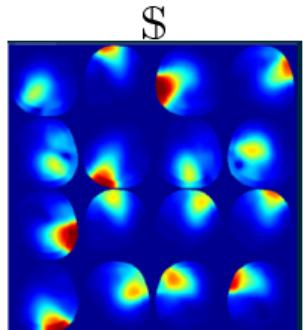
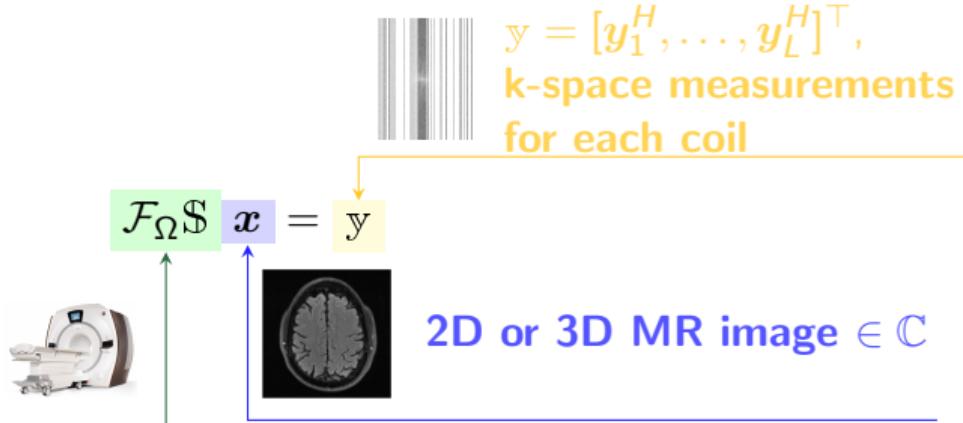
The Inverse Problem becomes:



Application to MRI

The Inverse Problem becomes:

\mathcal{F}_Ω : per-coil FT on the Ω set;
 $\$ = [S_1^H, \dots, S_L^H]^\top$: the sensitivity maps per coil



Cartesian



Ω
non-Cartesian



The canonical MRI reconstruction problem

$$\arg \min_{x \in \mathbb{C}^n} \underbrace{\frac{1}{2} \| \mathcal{A} x - y \|_2^2}_{= \mathcal{F}_\Omega \$} + \underbrace{\lambda \| \psi x \|_1}_{\begin{array}{l} \text{Regularization term, } \mathcal{R} \\ \text{Wavelet basis} \\ \text{Regularization hyperparameter} \end{array}}$$

⁵M. Lustig, 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

Limitations of classical recovery algorithms

Additional acceleration factor on top of PI: 1.5.⁶

⁶H. Koyasu et al. (Aug. 2018). "KNC can scan three more patients per day with Compressed SENSE". In: *FieldStrength*.

Limitations of classical recovery algorithms

Additional acceleration factor on top of PI: 1.5.⁶

The tool we used to express our prior knowledge, the wavelet basis, is limited:
handcrafted and linear.

⁶H. Koyasu et al. (Aug. 2018). “KNC can scan three more patients per day with Compressed SENSE”.
In: *FieldStrength*.

Inverse Problems

Recap

MRI is slow because of **relaxation**.

Inverse Problems

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.

Inverse Problems

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.

3. Deep Learning for MRI reconstruction

Model agnostic learning

Let's throw away all we know:⁷

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

⁷B. Zhu et al. (Mar. 2018). "Image reconstruction by domain-transform manifold learning". In: *Nature* 555.7697, pp. 487–492.

Model agnostic learning

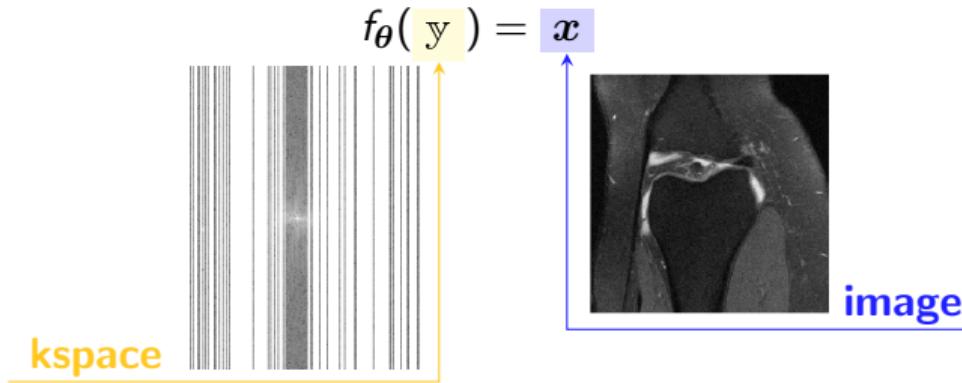
Let's throw away all we know:⁷



⁷B. Zhu et al. (Mar. 2018). “Image reconstruction by domain-transform manifold learning”. In: *Nature* 555.7697, pp. 487–492.

Model agnostic learning

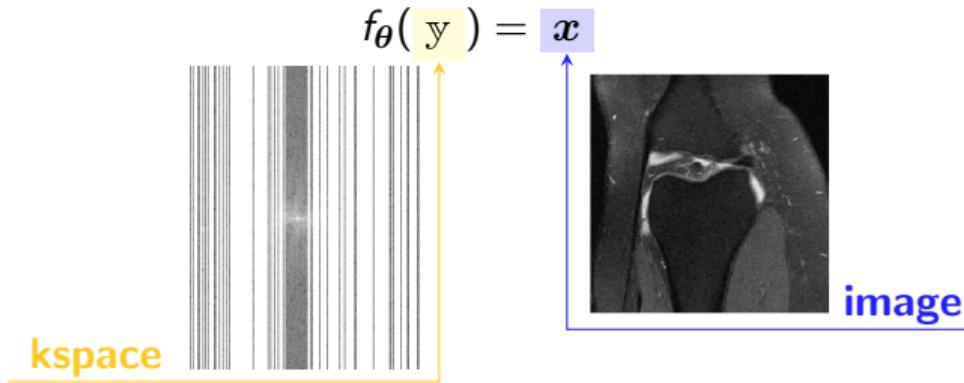
Let's throw away all we know:⁷



⁷B. 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:⁷



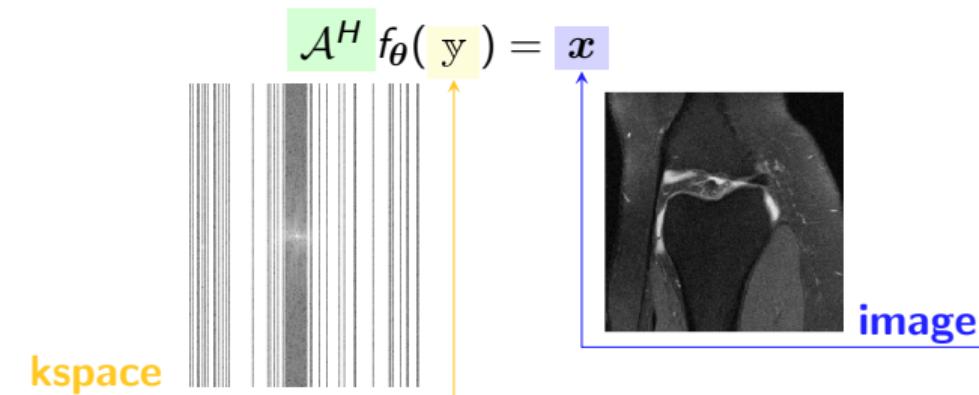
Cons:

- Not taking advantage of our knowledge of the physics
- No scaling (larger images, multi-coil, 3D)

⁷B. Zhu et al. (Mar. 2018). "Image reconstruction by domain-transform manifold learning". In: *Nature* 555.7697, pp. 487–492.

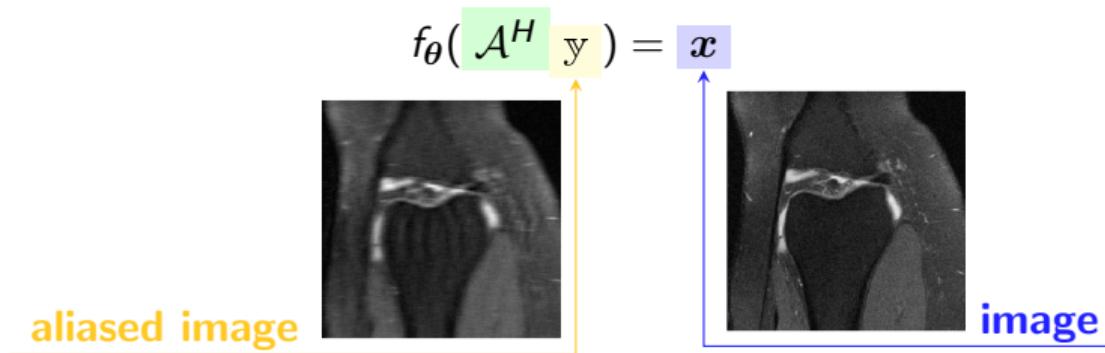
Single domain learning

Let's use \mathcal{A}^H to build a more informed model in the k-space:



Single domain learning

Let's use \mathcal{A}^H to build a more informed model in the image space:



Unrolled models - 1

We can mix the 2 single domain approaches, using the principled **optimization algorithm unrolling** method.⁸

A graph representation of ISTA:

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

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

$$\arg \min_{\mathbf{x} \in \mathbb{C}^n} \frac{1}{2} \|\mathcal{A}\mathbf{x} - \mathbf{y}\|_2^2 + \mathcal{R}(\mathbf{x})$$

↑
Proximity operator

⁸K. Gregor et al. (2010). “Learning fast approximations of sparse coding”. In: *ICML 2010 - Proceedings, 27th International Conference on Machine Learning*, pp. 399–406.

ISTA: Iterative Shrinkage Thresholding Algorithm

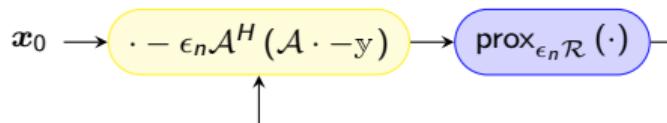
Unrolled models - 1

We can mix the 2 single domain approaches, using the principled **optimization algorithm unrolling** method.⁸

A graph representation of ISTA:

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

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



⁸K. Gregor et al. (2010). "Learning fast approximations of sparse coding". In: *ICML 2010 - Proceedings, 27th International Conference on Machine Learning*, pp. 399–406.

ISTA: Iterative Shrinkage Thresholding Algorithm

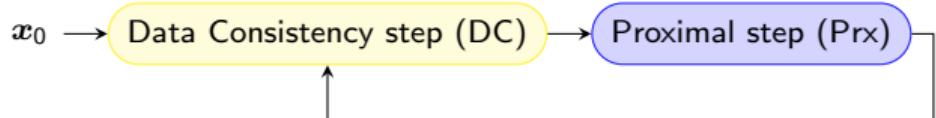
Unrolled models - 1

We can mix the 2 single domain approaches, using the principled **optimization algorithm unrolling** method.⁸

A graph representation of ISTA:

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

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



⁸K. Gregor et al. (2010). "Learning fast approximations of sparse coding". In: *ICML 2010 - Proceedings, 27th International Conference on Machine Learning*, pp. 399–406.

ISTA: Iterative Shrinkage Thresholding Algorithm

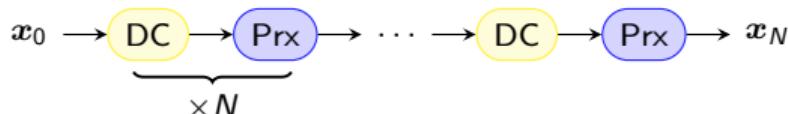
Unrolled models - 1

We can mix the 2 single domain approaches, using the principled **optimization algorithm unrolling** method.⁸

A graph representation of ISTA:

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

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



⁸K. Gregor et al. (2010). “Learning fast approximations of sparse coding”. In: *ICML 2010 - Proceedings, 27th International Conference on Machine Learning*, pp. 399–406.

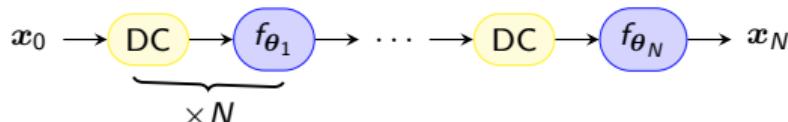
ISTA: Iterative Shrinkage Thresholding Algorithm

Unrolled models - 1

We can mix the 2 single domain approaches, using the principled **optimization algorithm unrolling** method.⁸

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

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



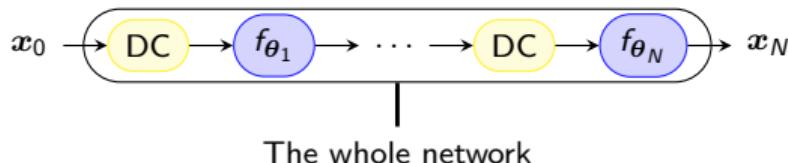
⁸K. Gregor et al. (2010). “Learning fast approximations of sparse coding”. In: *ICML 2010 - Proceedings, 27th International Conference on Machine Learning*, pp. 399–406.

Unrolled models - 1

We can mix the 2 single domain approaches, using the principled **optimization algorithm unrolling** method.⁸

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

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



⁸K. 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 [Ramzi et al. 2020a]

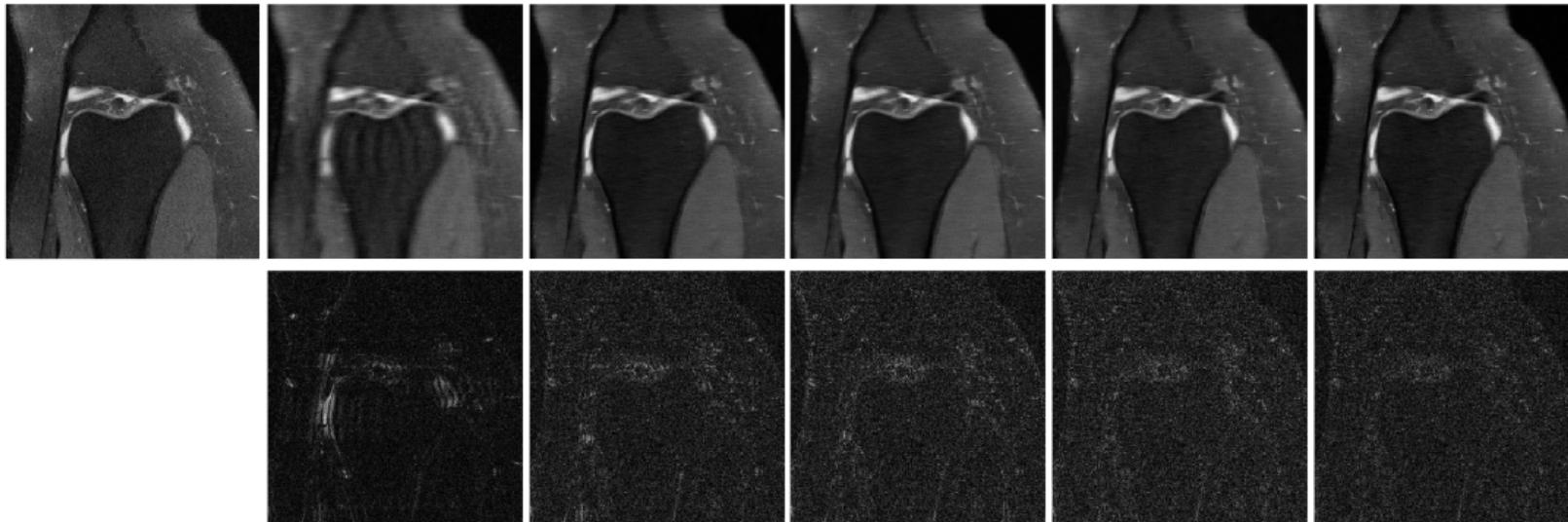
Contribution #1

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

Reference **Zero-filled** **KIKI-net** **U-net** **Cascade-net** **PD-net**



Unrolled models - 2 [Ramzi et al. 2020a]

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 ⁹
PSNR (dB)	29.61	31.38	31.78	31.97	32.15

⁹J. Adler and O. Öktem (2018). “Learned Primal-Dual Reconstruction”. In: *IEEE Transactions on Medical Imaging* 37.6, pp. 1322–1332

Unrolled models - 2 [Ramzi et al. 2020a]

Contribution #1

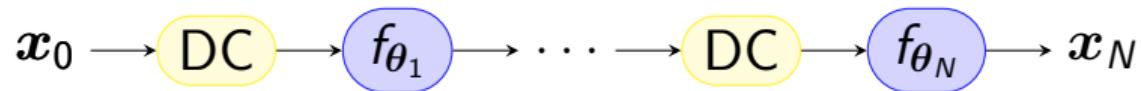
Zaccharie Ramzi, P. Ciuciu, and J. L. Starck (2020). “Benchmarking MRI reconstruction neural networks on large public datasets”. In: *Applied Sciences (Switzerland)* 10.5

Different models based on:

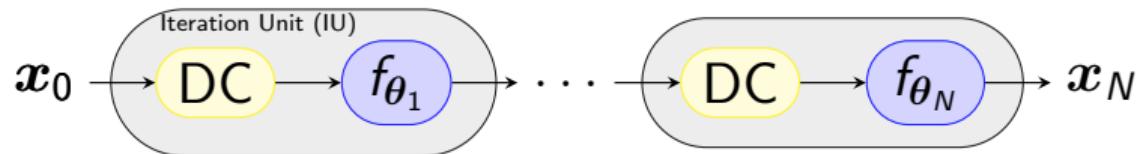
- optimization algorithm to unroll
- choice of f_θ
- N

-
- 🐾 Code available online:
github.com/zaccharieramzi/fastmri-reproducible-benchmark
 - 😊 Model weights available online: huggingface.co/zaccharieramzi

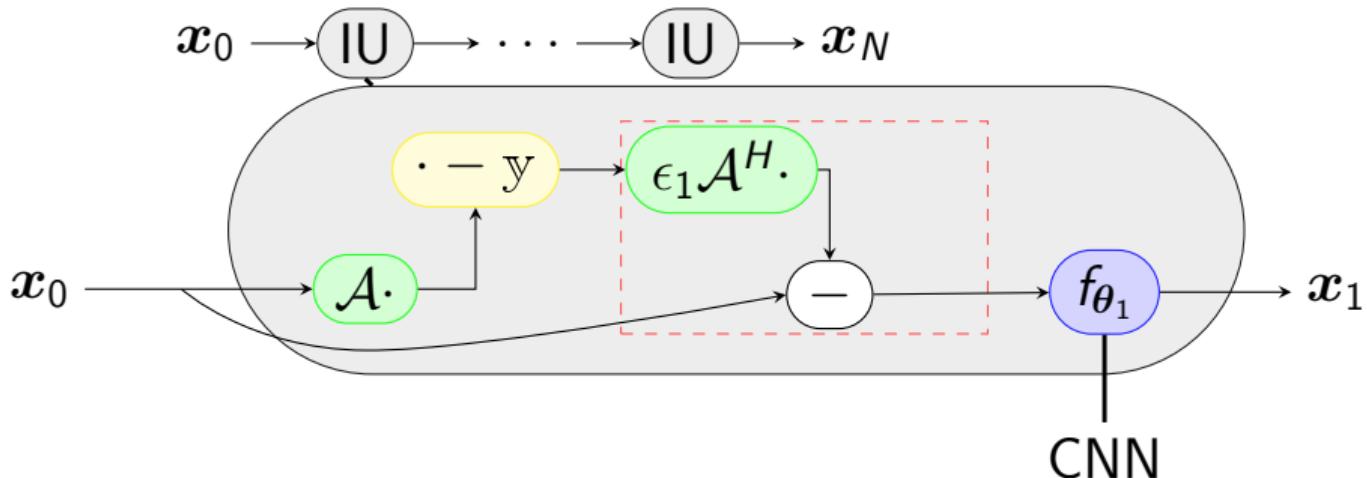
PDNet



PDNet



PDNet

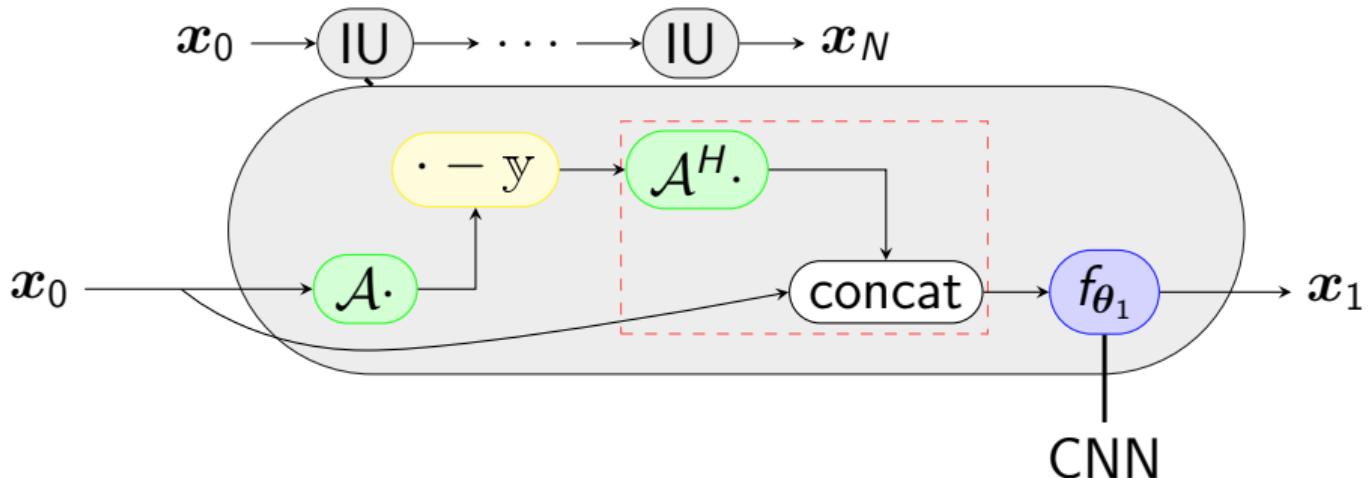


¹⁰P. Liu et al. (2018). "Multi-level Wavelet-CNN for Image Restoration". In: *CVPR NTIRE Workshop*.

¹¹A. Sriram et al. (2020). "End-to-End Variational Networks for Accelerated MRI Reconstruction". In: *MICCAI*.

¹²O. Ronneberger et al. (2015). "U-net: Convolutional networks for biomedical image segmentation". In: *International Conference on Medical image computing and computer-assisted intervention*.

PDNet

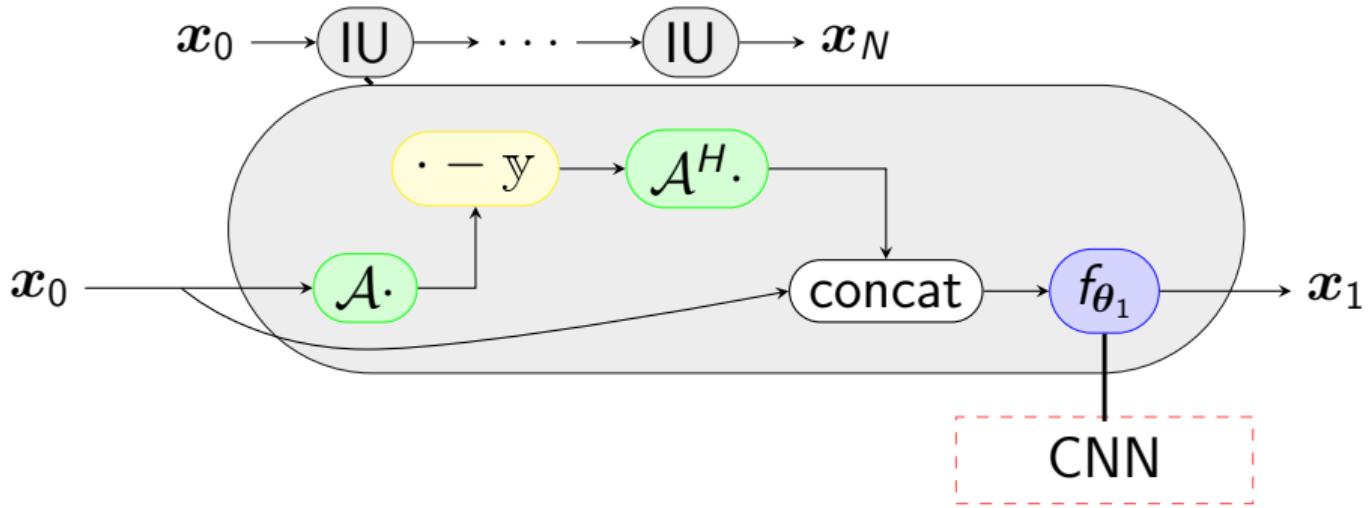


¹⁰P. Liu et al. (2018). "Multi-level Wavelet-CNN for Image Restoration". In: *CVPR NTIRE Workshop*.

¹¹A. Sriram et al. (2020). "End-to-End Variational Networks for Accelerated MRI Reconstruction". In: *MICCAI*.

¹²O. Ronneberger et al. (2015). "U-net: Convolutional networks for biomedical image segmentation". In: *International Conference on Medical image computing and computer-assisted intervention*.

XPDNet [Ramzi et al. 2020b]

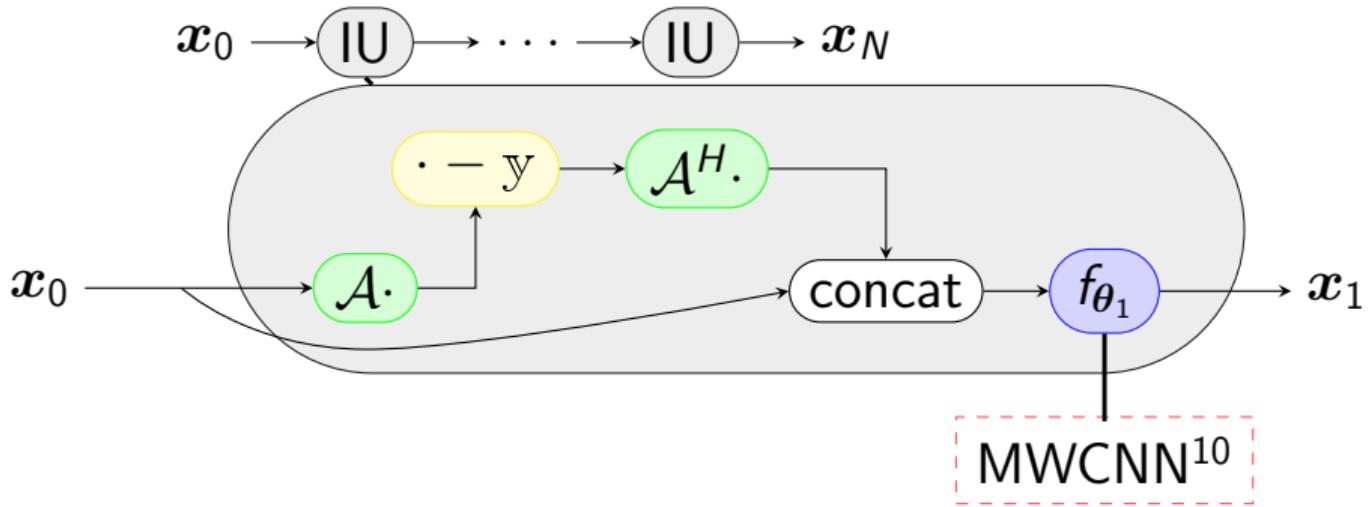


¹⁰P. Liu et al. (2018). "Multi-level Wavelet-CNN for Image Restoration". In: *CVPR NTIRE Workshop*.

¹¹A. Sriram et al. (2020). "End-to-End Variational Networks for Accelerated MRI Reconstruction". In: *MICCAI*.

¹²O. Ronneberger et al. (2015). "U-net: Convolutional networks for biomedical image segmentation". In: *International Conference on Medical image computing and computer-assisted intervention*.

XPDNet [Ramzi et al. 2020b]

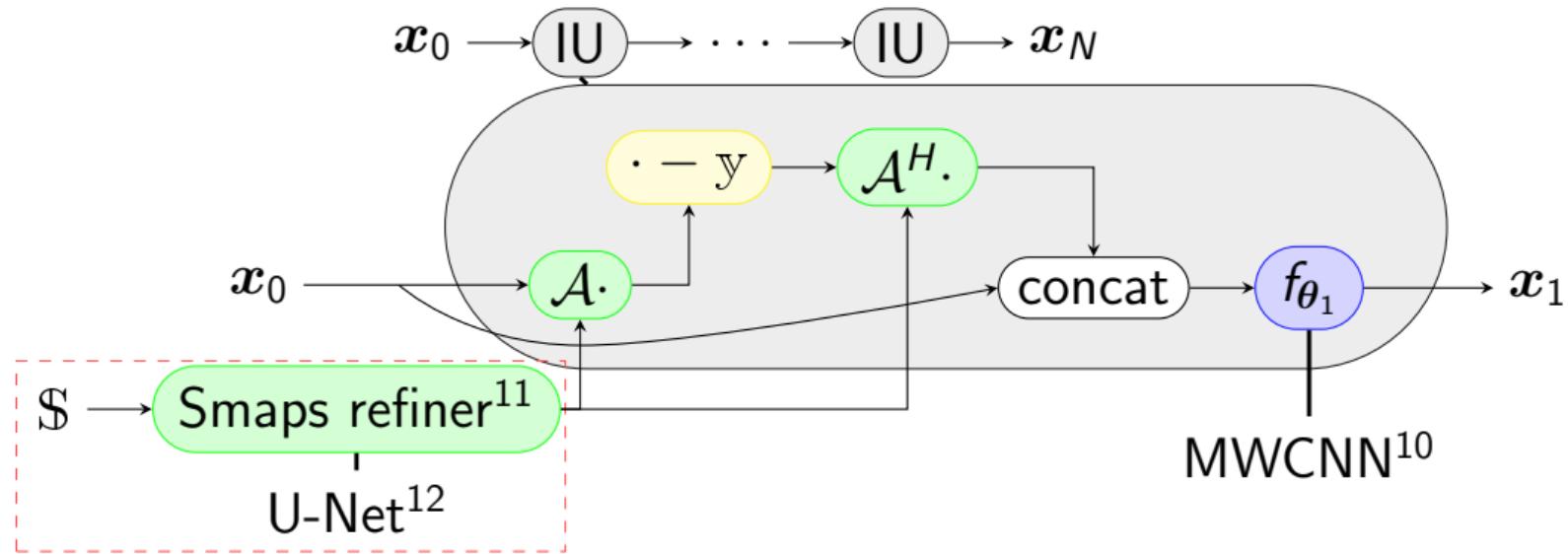


¹⁰P. Liu et al. (2018). "Multi-level Wavelet-CNN for Image Restoration". In: *CVPR NTIRE Workshop*.

¹¹A. Sriram et al. (2020). "End-to-End Variational Networks for Accelerated MRI Reconstruction". In: *MICCAI*.

¹²O. Ronneberger et al. (2015). "U-net: Convolutional networks for biomedical image segmentation". In: *International Conference on Medical image computing and computer-assisted intervention*.

XPDNet [Ramzi et al. 2020b]

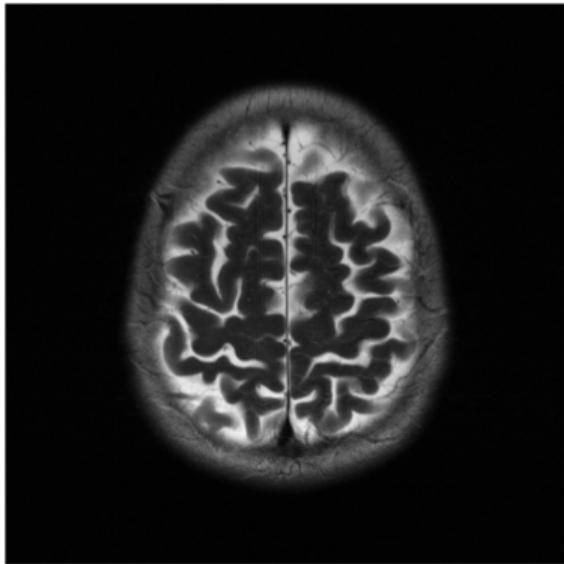


¹⁰P. Liu et al. (2018). "Multi-level Wavelet-CNN for Image Restoration". In: *CVPR NTIRE Workshop*.

¹¹A. Sriram et al. (2020). "End-to-End Variational Networks for Accelerated MRI Reconstruction". In: *MICCAI*.

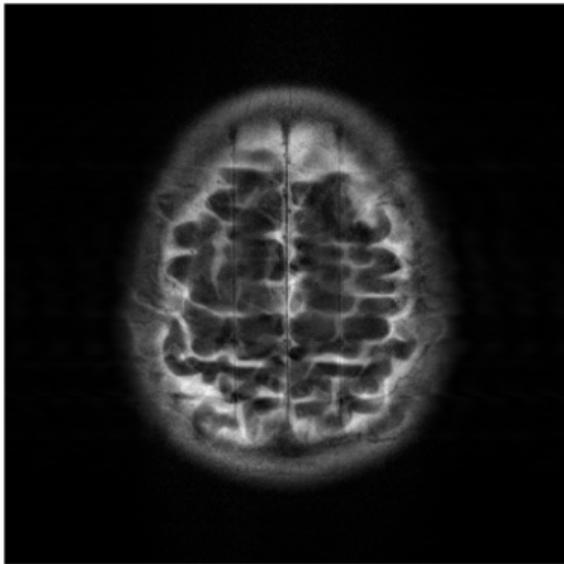
¹²O. Ronneberger et al. (2015). "U-net: Convolutional networks for biomedical image segmentation". In: *International Conference on Medical image computing and computer-assisted intervention*.

Reference



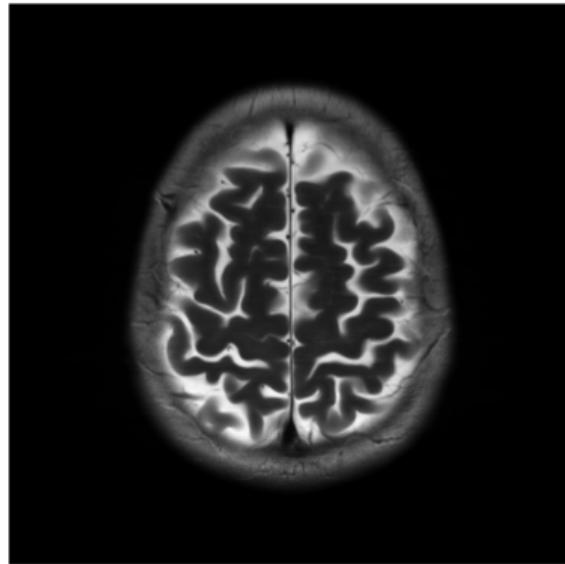
GRAPPA

PSNR: 26 dB, SSIM: 0.77



XPDNet

PSNR: 36 dB, SSIM: 0.96



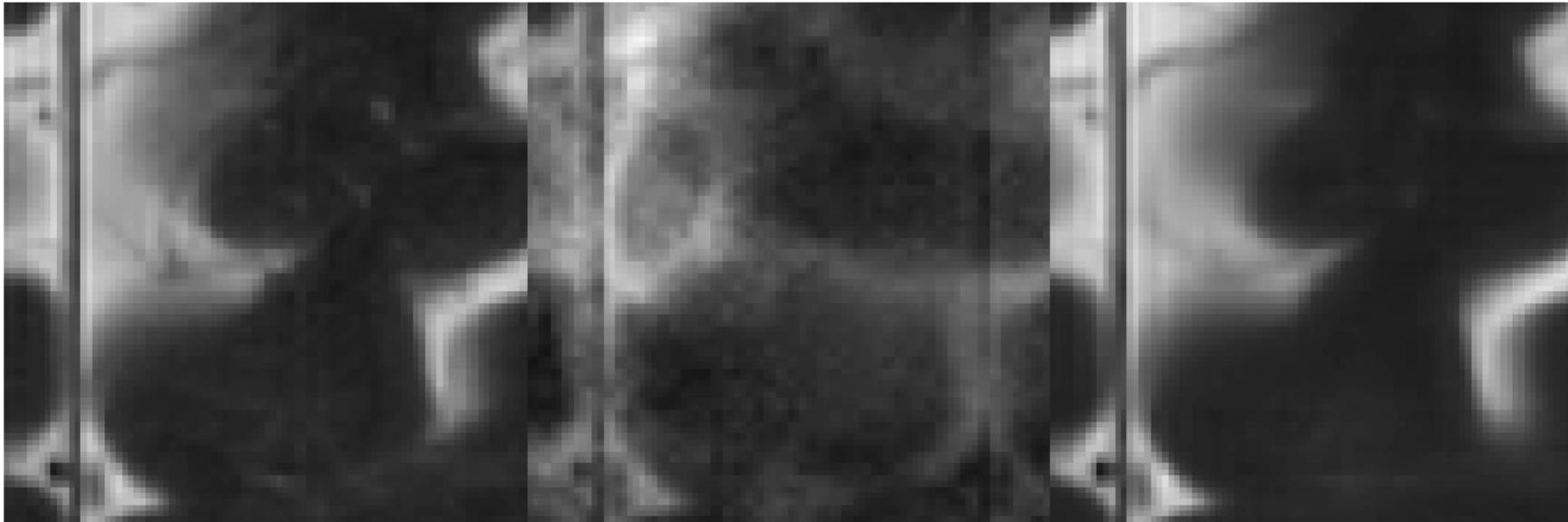
Reference

GRAPPA

PSNR: 26 dB, SSIM: 0.77

XPDNet

PSNR: 36 dB, SSIM: 0.96



fastMRI challenge [Ramzi et al. 2020b], [Muckley et al. 2021]

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 [Ramzi et al. 2020b], [Muckley et al. 2021]

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 [Ramzi et al. 2020b], [Muckley et al. 2021]

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

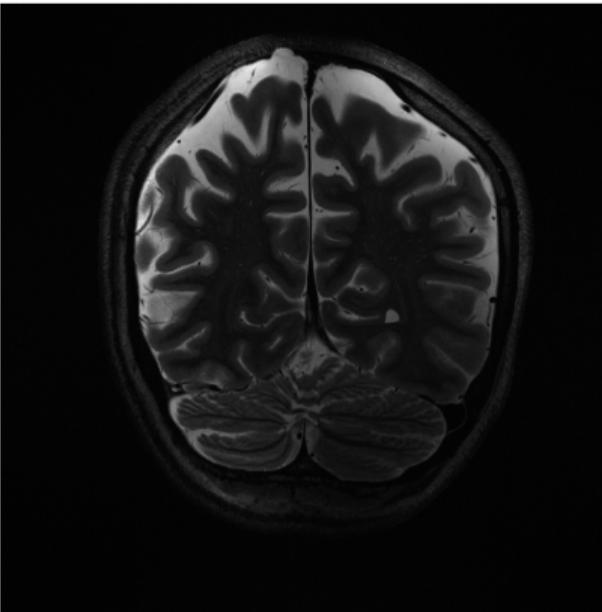
Robustness test [Ramzi et al. 2020c]

XPDNet in a prospective out-of-distribution setting:¹³
different orientation, higher resolution, higher field strength, lower acceleration factor,
presence of the cerebellum.¹⁴

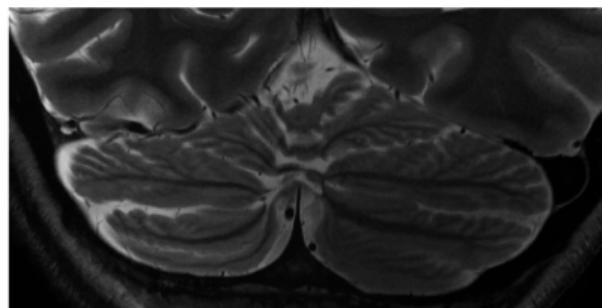
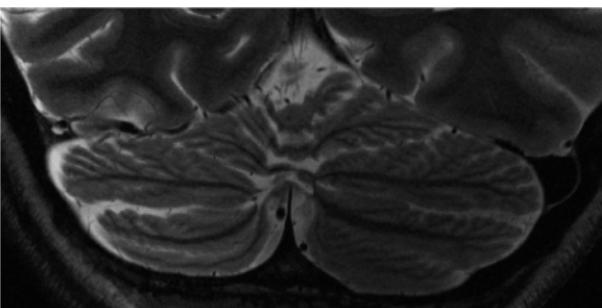
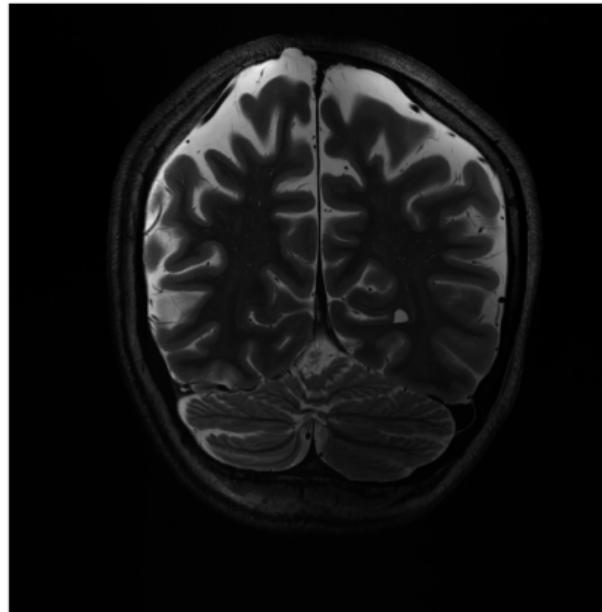
¹³L. Marrakchi-Kacem et al. (2016). "Robust imaging of hippocampal inner structure at 7T: in vivo acquisition protocol and methodological choices". In: *Magnetic Resonance Materials in Physics, Biology and Medicine* 29.3, pp. 475–489.

¹⁴For anonymity reasons, the cerebellum is not present in the fastMRI dataset.

GRAPPA



XPDNet



Non-Cartesian acquisitions

Non-Cartesian acquisitions more robust to motion.

Non-Cartesian acquisitions

Non-Cartesian acquisitions more robust to motion.

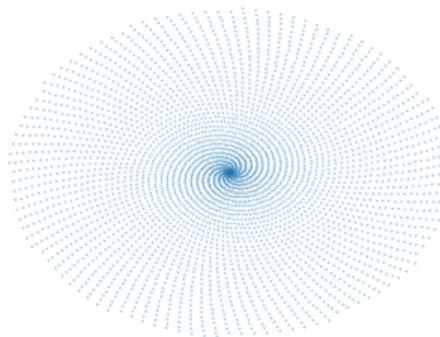
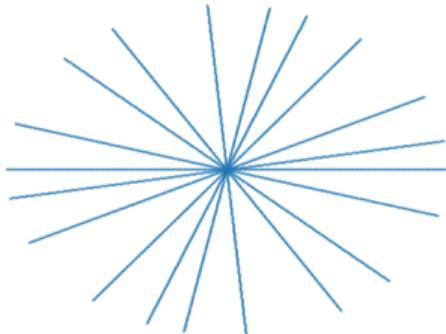


Figure: **Radial and spiral undersampled trajectories.**

Non-Cartesian acquisitions

Non-Cartesian acquisitions more robust to motion.

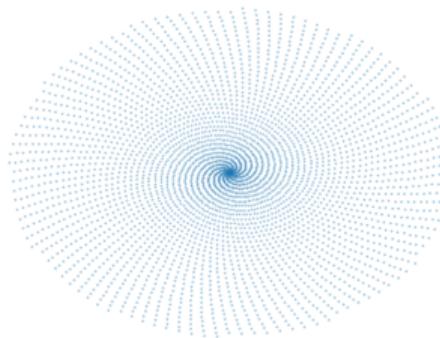
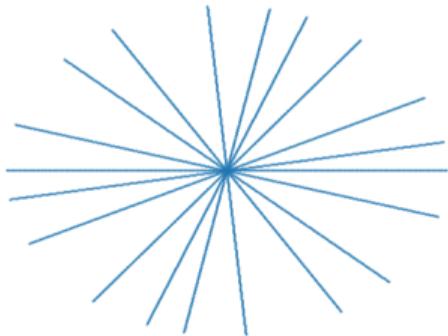
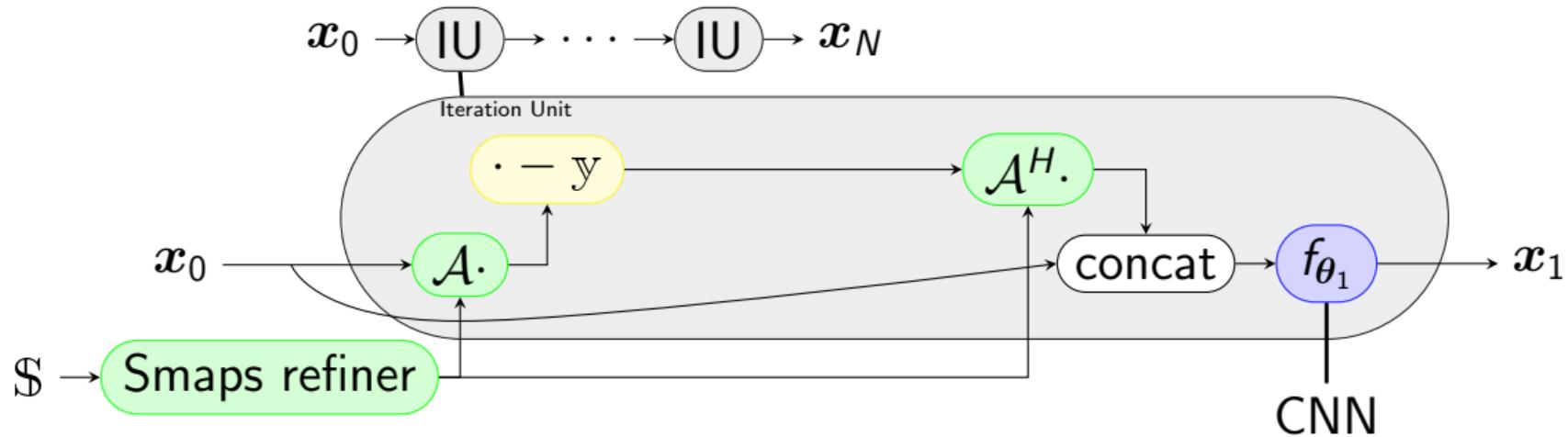


Figure: Radial and spiral undersampled trajectories.

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

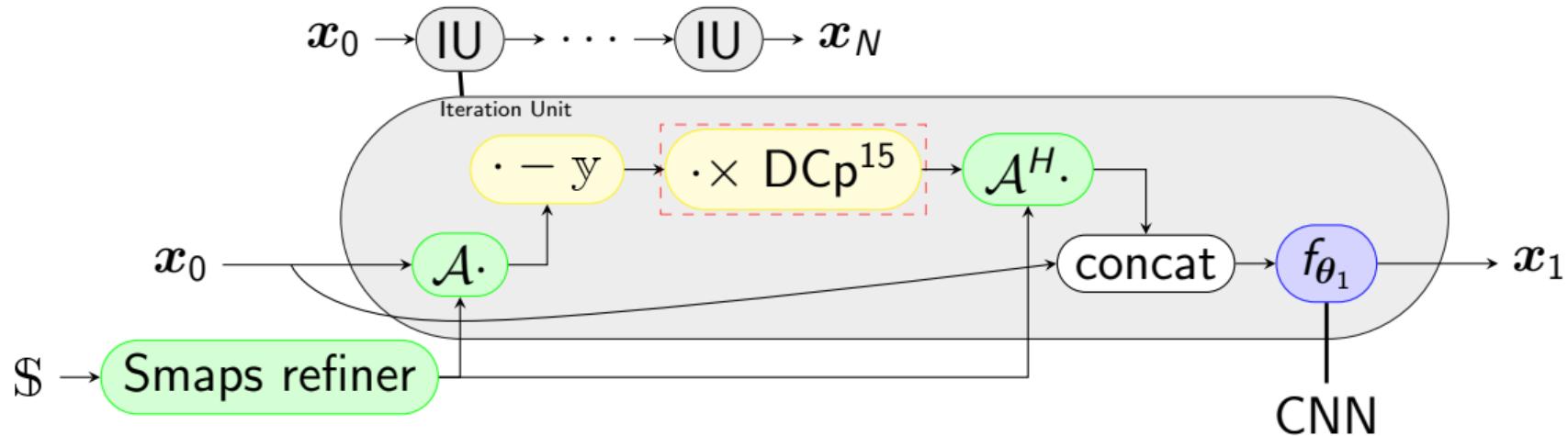
- pip install tfkbnufft

NC-PDNet - 1 [Ramzi et al. 2022a]



¹⁵J. 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 [Ramzi et al. 2022a]



¹⁵J. 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 [Ramzi et al. 2022a]

Contribution #3

Z. Ramzi, C. G R, J.-L. Starck, and P. Ciuciu (2022). “NC-PDNet: a Density-Compensated Unrolled Network for 2D and 3D non-Cartesian MRI Reconstruction”. In: *IEEE Transactions on Medical Imaging*

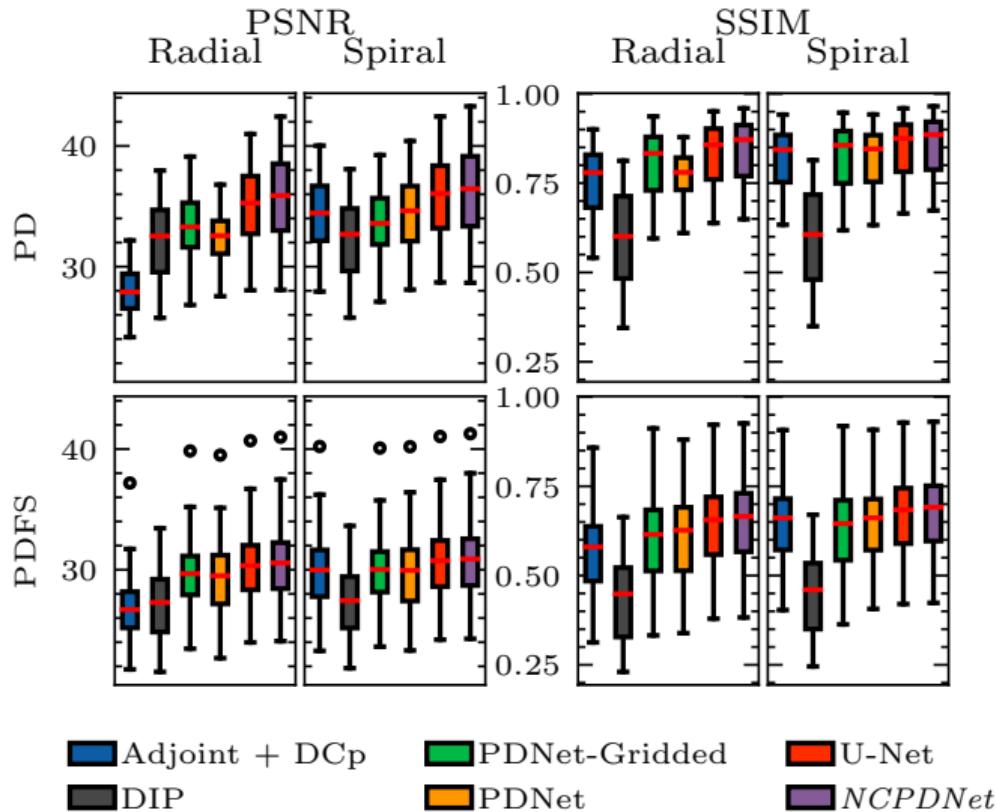


Figure: 2D single-coil reconstruction quantitative results on the fastMRI knee dataset.

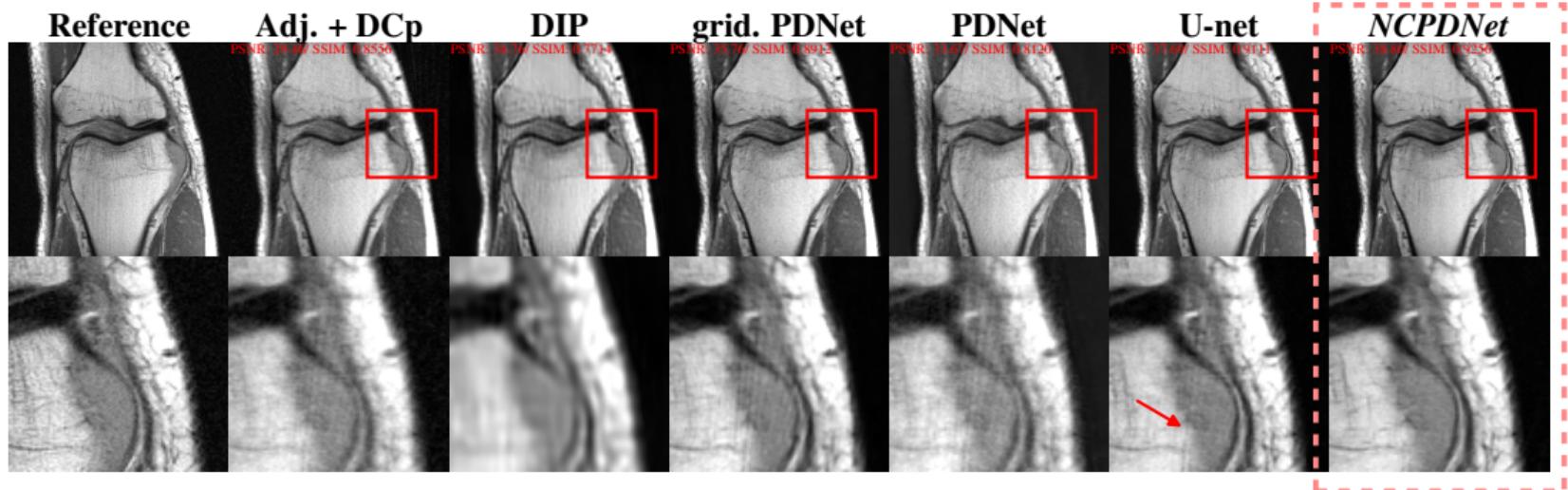


Figure: 2D single-coil reconstruction qualitative results on the fastMRI dataset for a radial trajectory.

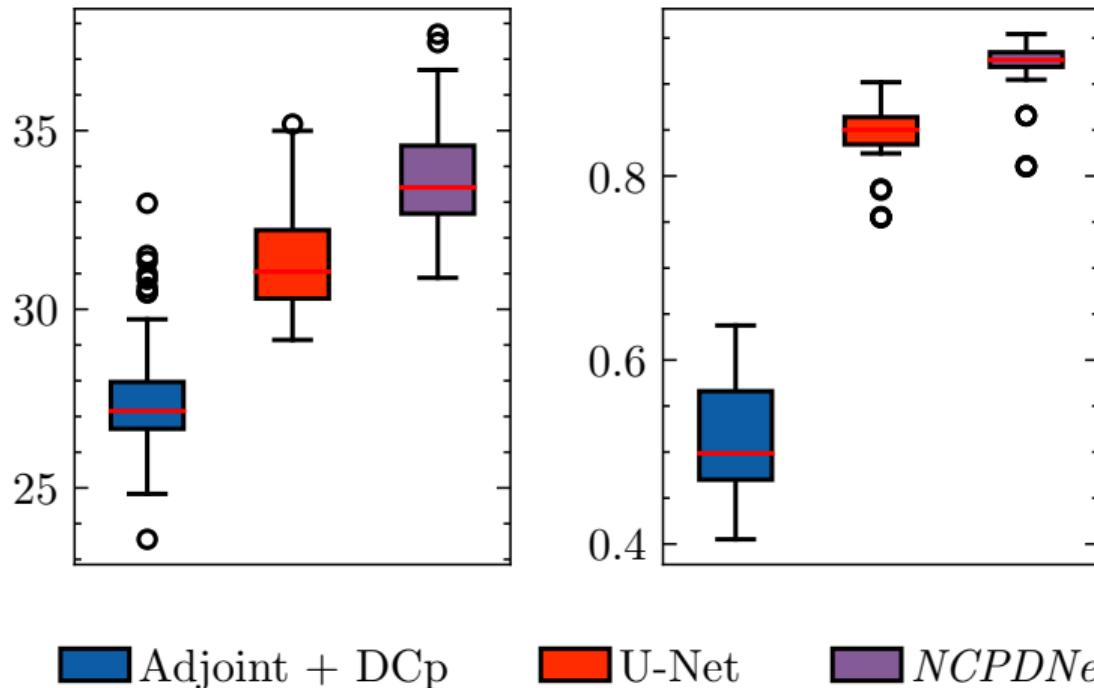
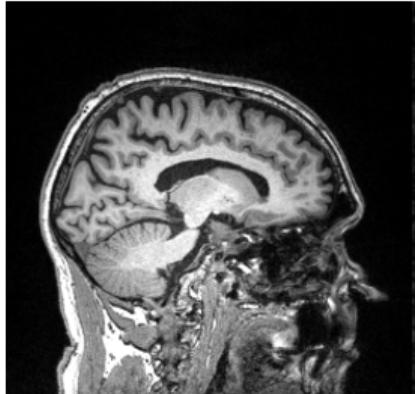


Figure: 3D single-coil reconstruction quantitative results on the OASIS dataset for a radial trajectory.

Reference

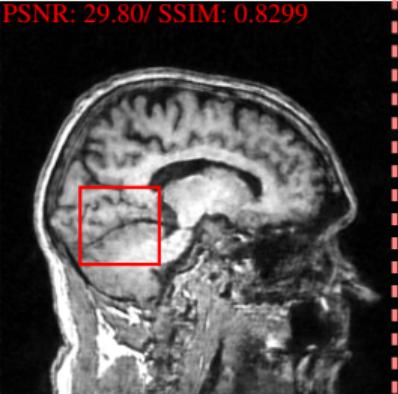


Adj. + DCp



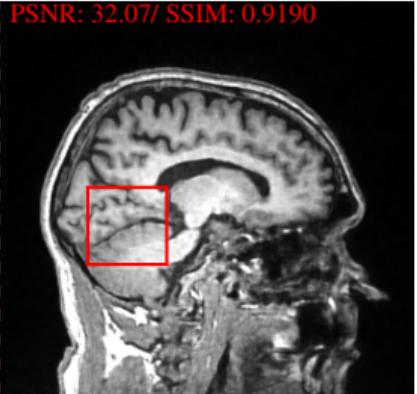
PSNR: 26.02/ SSIM: 0.4601

U-net



PSNR: 29.80/ SSIM: 0.8299

NCPDNet



PSNR: 32.07/ SSIM: 0.9190

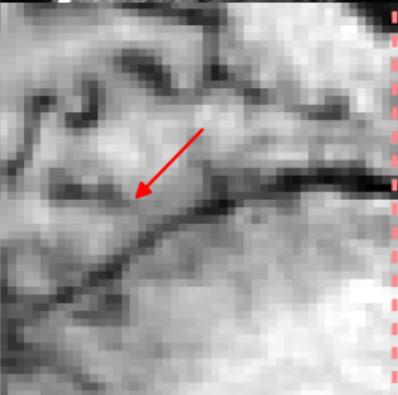


Figure: 3D single-coil reconstruction qualitative results on the OASIS dataset for a radial trajectory.

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 complex structures in MR images. 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 complex structures in MR images. We showcased 2 instances of unrolled models, **XPDNet** and **NC-PDNet**, which can perform really well in challenging acquisition settings.

Challenges:

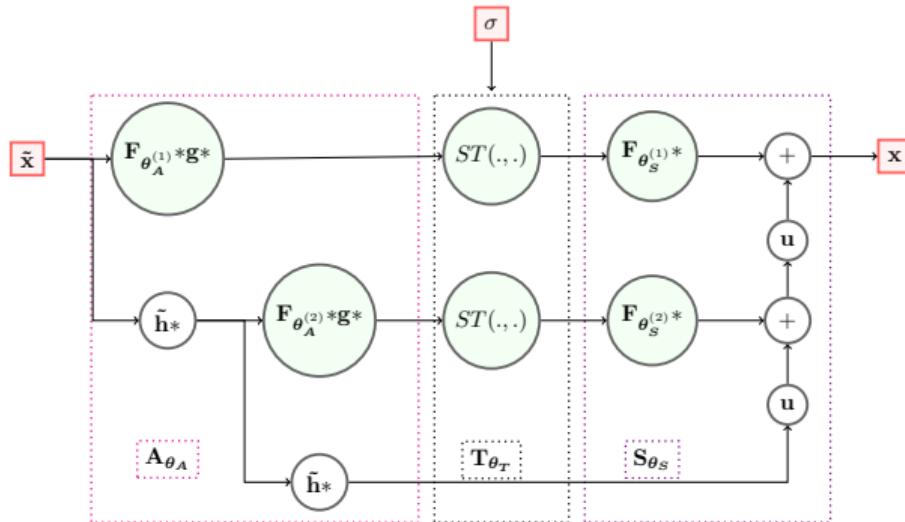
- generalization
- uncertainty quantification
- training memory requirements

4. Clinical applicability

Learnlets - 1 [Ramzi et al. 2021a]

Contribution #4

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



ST: Soft Thresholding

Learnlets - 2 [Ramzi et al. 2021a]

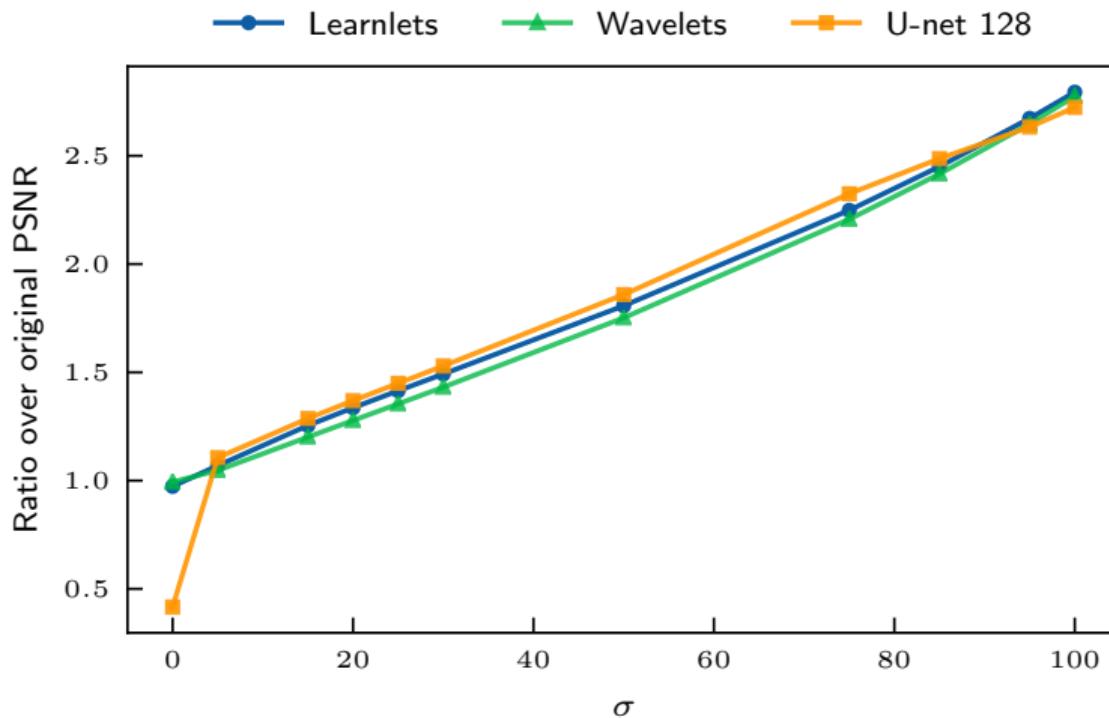


Figure: Denoising performances.

Learnlets - 2 [Ramzi et al. 2021a]

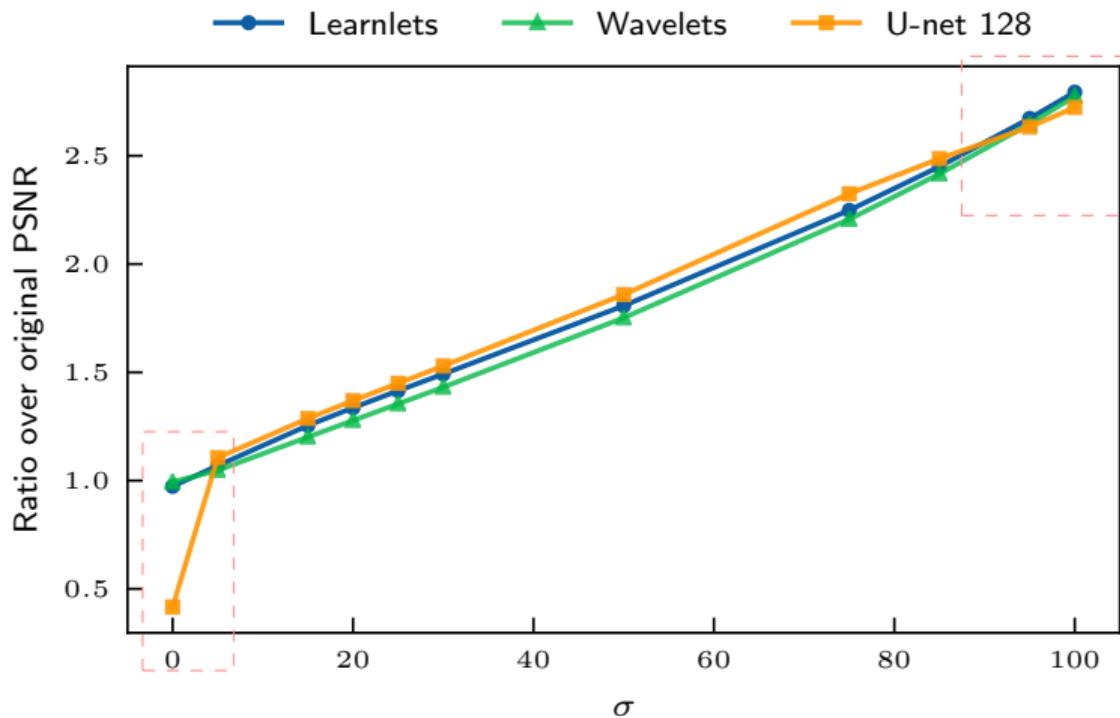


Figure: Denoising performances.

Denoising Score Matching - 1 [Ramzi et al. 2020d]

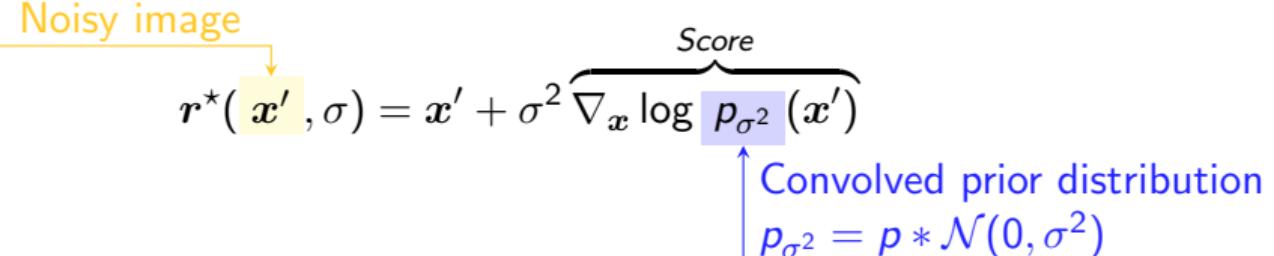
Contribution #5

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*

Optimal denoiser r^* :¹⁶

$$r^*(x', \sigma) = x' + \sigma^2 \underbrace{\nabla_{x'} \log p_{\sigma^2}(x')}_{\text{Score}}$$

\uparrow Convolved prior distribution
 $p_{\sigma^2} = p * \mathcal{N}(0, \sigma^2)$



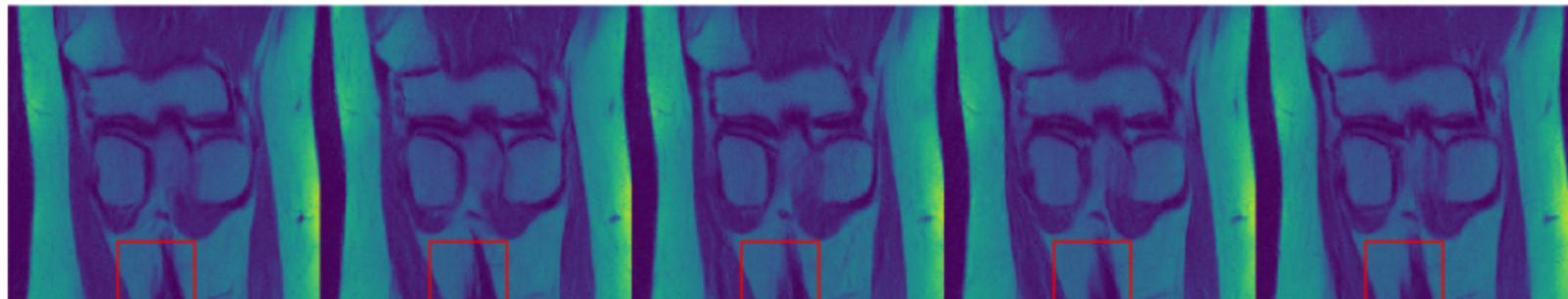
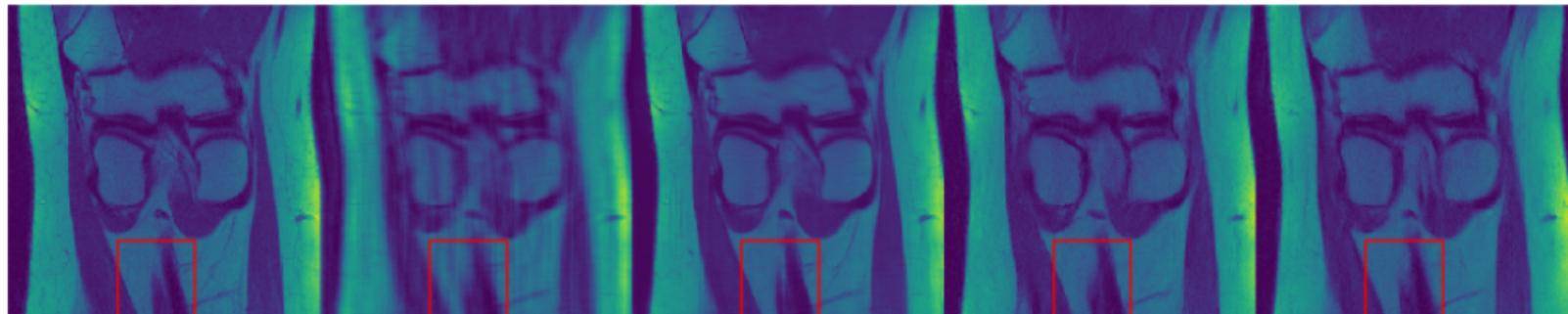
¹⁶G. Alain et al. (2013). “What regularized auto-encoders learn from the data generating distribution”. In: *ICLR 2013*. Vol. 15, pp. 3743–3773; P. Vincent (2011). “A connection between scorematching and denoising autoencoders”. In: *Neural Computation* 23.7, pp. 1661–1674.

Denoising Score Matching - 2 [Ramzi et al. 2020d]

Ground truth Zero-filled

UPDNet

Samples → ...



5. 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.

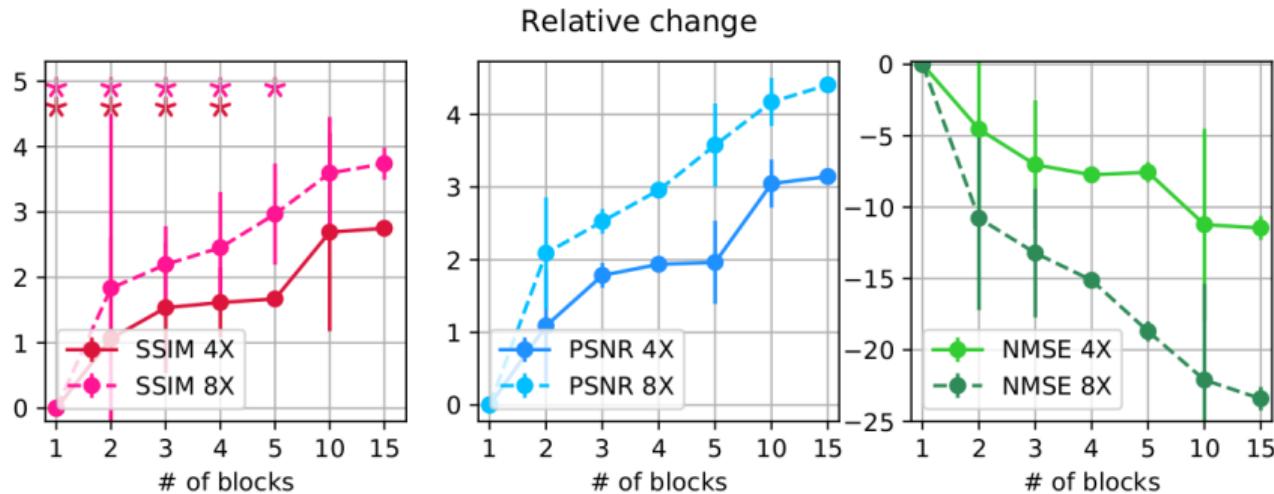


Figure: Performance of an unrolled MRI reconstruction network function of the number of iteration units (blocks).¹⁷

¹⁷ 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?

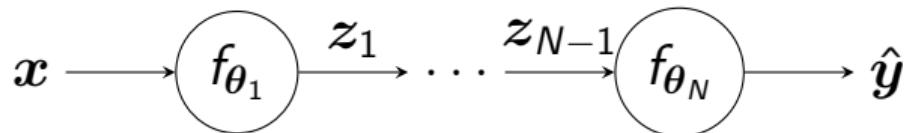
Deeper models \Rightarrow more **memory** to train. Memory we have is constrained.

Can we go deeper?

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

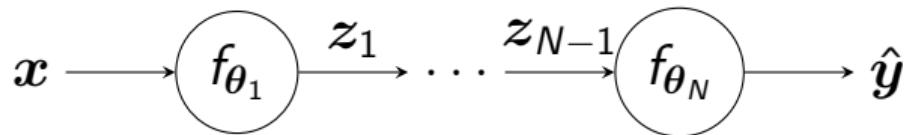
Can we go deeper?

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



Can we go deeper?

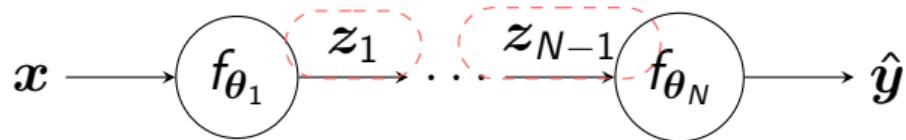
Deeper models \Rightarrow more **memory** to train. Memory we have is constrained.
Why more memory? **Activations!**



$$\frac{\partial \mathcal{L}}{\partial \theta_1} = \frac{\partial \mathcal{L}}{\partial \hat{y}} \Big|_{\hat{y}} \frac{\partial \hat{y}}{\partial z_{N-1}} \Big|_{z_{N-1}} \cdots \frac{\partial z_2}{\partial z_1} \Big|_{z_1} \frac{\partial z_1}{\partial \theta_1} \Big|_{\theta_1}$$

Can we go deeper?

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



$$\frac{\partial \mathcal{L}}{\partial \theta_1} = \frac{\partial \mathcal{L}}{\partial \hat{y}} \Big|_{\hat{y}} \frac{\partial \hat{y}}{\partial z_{N-1}} \Big|_{z_{N-1}} \cdots \frac{\partial z_2}{\partial z_1} \Big|_{z_1} \frac{\partial z_1}{\partial \theta_1} \Big|_{\theta_1}$$

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)

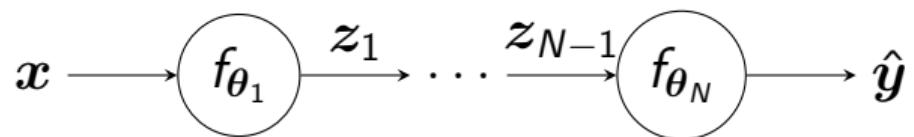
The modeling solutions

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)

Infinite depth neural networks

A recurrent expression of classical, explicit networks:



Infinite depth neural networks

A recurrent expression of classical, explicit networks:

$$z_n = f_{\theta_n}(z_{n-1}), \quad \forall n < N$$

Infinite depth neural networks

A recurrent expression of classical, explicit networks:

$$z_n = f_{\theta_n}(z_{n-1}), \quad \forall n < N$$

What if $N \rightarrow \infty$?

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.

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

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.

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

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.

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

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

Deep Equilibrium networks - 2

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

Deep Equilibrium networks - 2

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. θ , called Hypergradient, is given by

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

Deep Equilibrium networks - 2

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. θ , called Hypergradient, is given by

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

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.

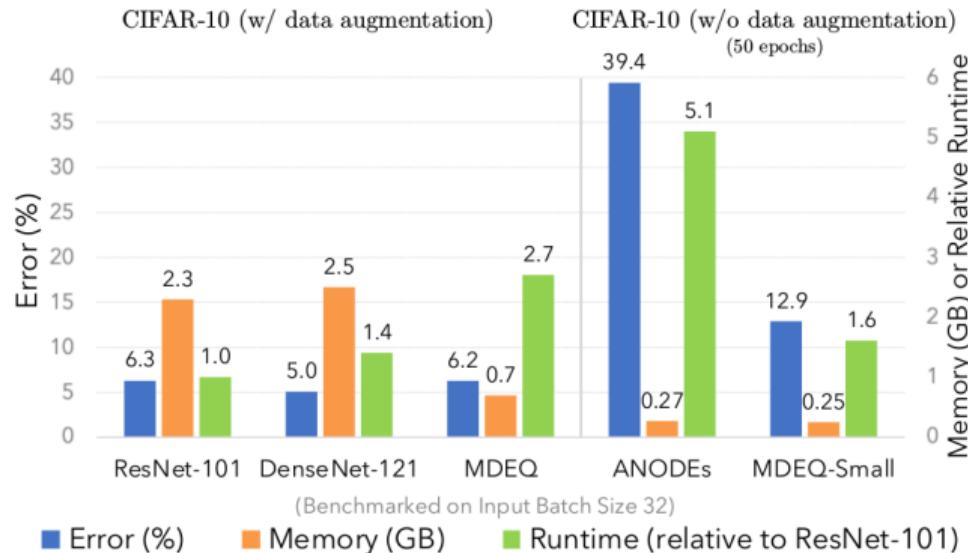


Figure: Performance, memory and training speed of DEQs. (Bai, Koltun, et al., 2020)

Why are DEQs slow?

DEQs gradient computation:

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

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

Why are DEQs slow?

DEQs gradient computation:

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

we 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? [Ramzi et al. 2022b]

Contribution #6

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 2022*. (Spotlight)

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

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

Can we avoid the Jacobian inversion? [Ramzi et al. 2022b]

Contribution #6

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 2022*. (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? [Ramzi et al. 2022b]

Contribution #6

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 2022*. (Spotlight)

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

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

quasi-Newton matrix True Jacobian inverse

Can we avoid the Jacobian inversion? [Ramzi et al. 2022b]

Contribution #6

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 2022*. (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 $g_\theta(z^*, x) = 0$ using a quasi-Newton method.

Can we avoid the Jacobian inversion? [Ramzi et al. 2022b]

Contribution #6

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 2022*. (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 $g_\theta(z^*, x) = 0$ using a quasi-Newton method.
- It is easily invertible using the Sherman-Morrison formula, because low-rank.

Can we avoid the Jacobian inversion? [Ramzi et al. 2022b]

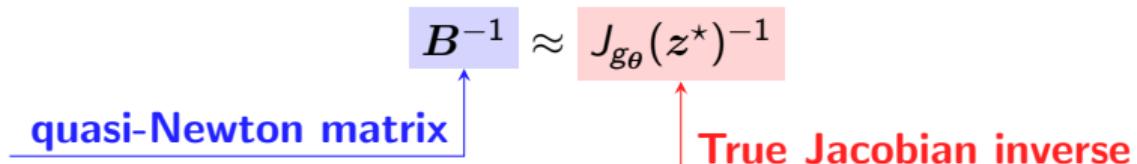
Contribution #6

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 2022*. (Spotlight)

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

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

quasi-Newton matrix True Jacobian inverse



- We have asymptotic correctness of the approximation!

Computer vision results [Ramzi et al. 2022b]

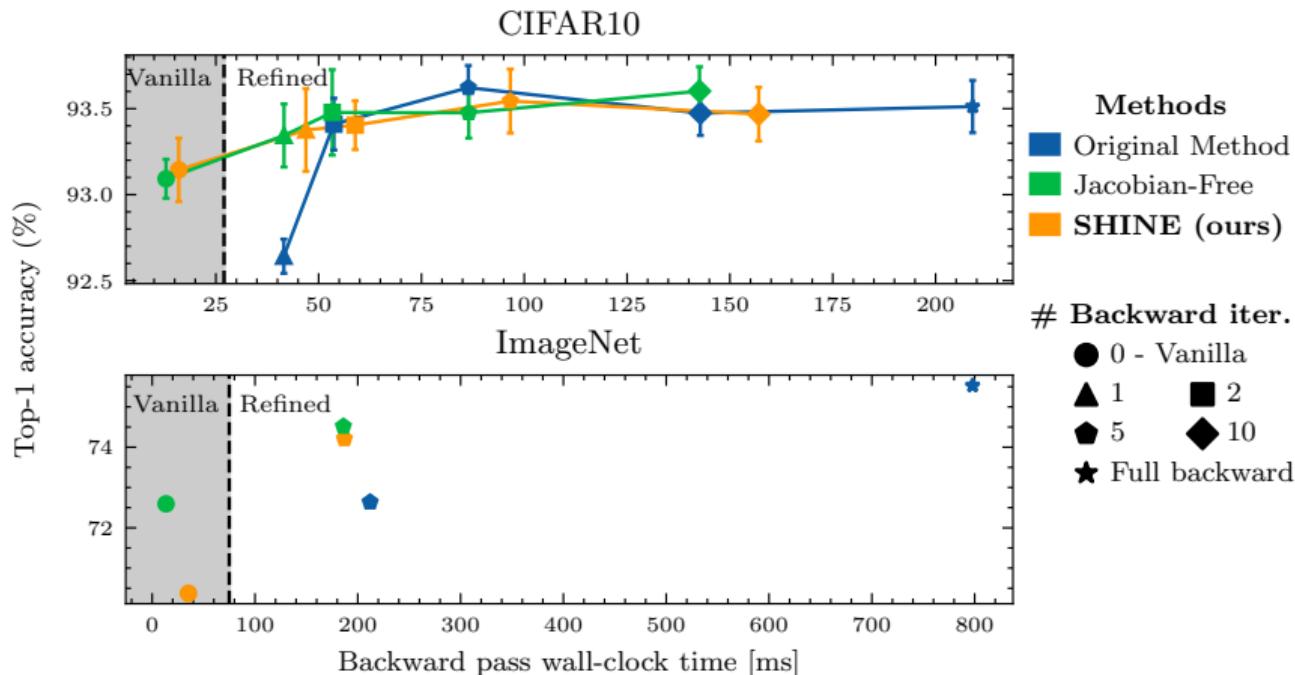


Figure: MDEQs (Bai, Koltun, et al., 2020) with SHINE.

Application to Hyperparameter optimization - 1

[Ramzi et al. 2022b]

Hyperparameter optimization can benefit from SHINE.

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

Application to Hyperparameter optimization - 1

[Ramzi et al. 2022b]

Hyperparameter optimization can benefit from SHINE.

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

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

[Ramzi et al. 2022b]

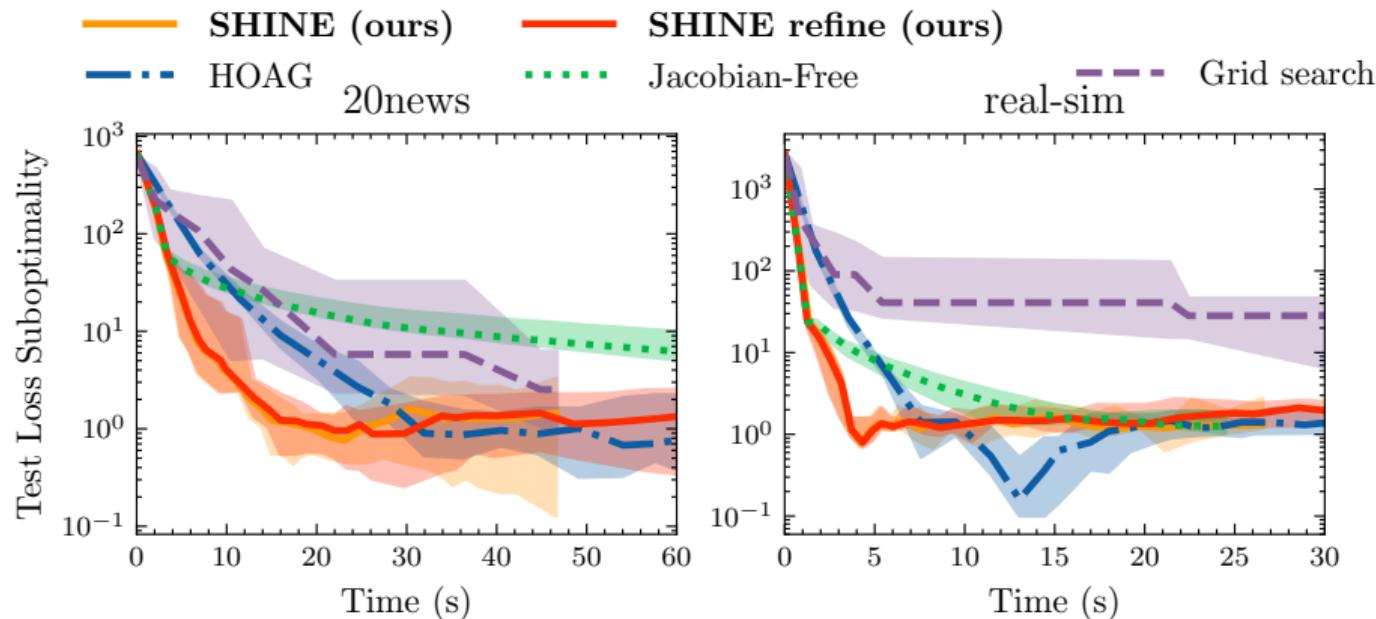


Figure: Bilevel optimization (Pedregosa, 2016) with SHINE: convergence of held-out test loss.

Conclusions

1. In order to accelerate MRI, I showed how to further adapt unrolled models to MRI reconstruction:

Conclusions

1. In order to accelerate MRI, I showed how to further adapt unrolled models to MRI reconstruction:
 - 1.1 2D multi-coil: [XPDNet](#)

Conclusions

1. In order to accelerate MRI, I showed how to further adapt unrolled models to MRI reconstruction:
 - 1.1 2D multi-coil: XPDNet
 - 1.2 non-Cartesian: NC-PDNet

Conclusions

1. In order to accelerate MRI, I showed how to further adapt unrolled models to MRI reconstruction:
 - 1.1 2D multi-coil: XPDNet
 - 1.2 non-Cartesian: NC-PDNet
2. I also tackled 3 challenges associated with the application of deep learning to MRI reconstruction:

Conclusions

1. In order to accelerate MRI, I showed how to further adapt unrolled models to MRI reconstruction:
 - 1.1 2D multi-coil: XPDNet
 - 1.2 non-Cartesian: NC-PDNet
2. I also tackled 3 challenges associated with the application of deep learning to MRI reconstruction:
 - 2.1 generalization: Learnlets

Conclusions

1. In order to accelerate MRI, I showed how to further adapt unrolled models to MRI reconstruction:
 - 1.1 2D multi-coil: XPDNet
 - 1.2 non-Cartesian: NC-PDNet
2. I also tackled 3 challenges associated with the application of deep learning to MRI reconstruction:
 - 2.1 generalization: Learnlets
 - 2.2 uncertainty quantification: Denoising Score Matching

Conclusions

1. In order to accelerate MRI, I showed how to further adapt unrolled models to MRI reconstruction:
 - 1.1 2D multi-coil: XPDNet
 - 1.2 non-Cartesian: NC-PDNet
2. I also tackled 3 challenges associated with the application of deep learning to MRI reconstruction:
 - 2.1 generalization: Learnlets
 - 2.2 uncertainty quantification: Denoising Score Matching
 - 2.3 training memory and speed: SHINE

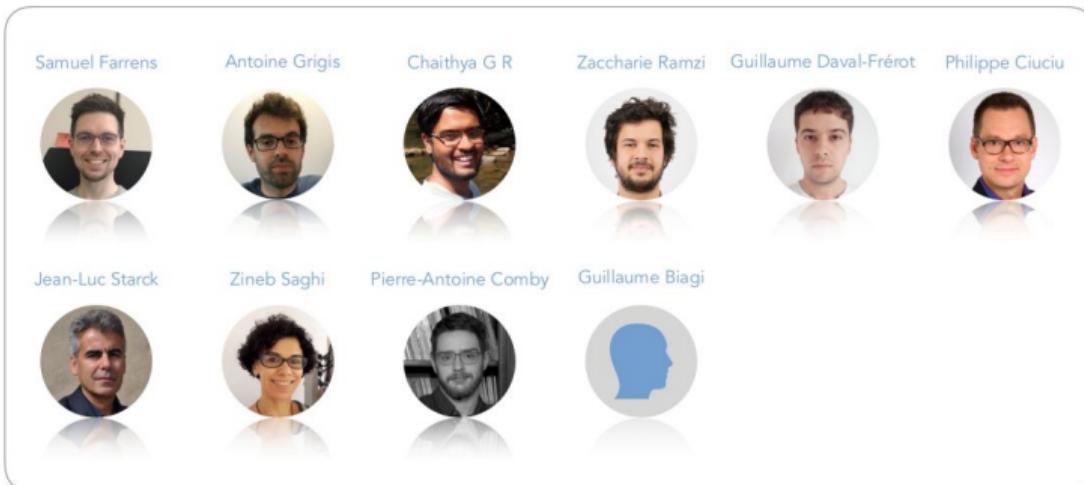
Future works

- Refine the measurement operator even more, for example with B_0 inhomogeneity corrections.
- Extend the NC-PDNet to more dimensions, like in fMRI.
- Learn better k-space acquisition trajectories.
- Applying DEQs to MRI reconstruction.
- Understand better whether the subnetworks learn something far from the score of the prior distribution.

Additional contribution

Contribution #7

- S. Farrens et al. (2020). “PySAP: Python Sparse Data Analysis Package for multidisciplinary image processing”. In: *Astronomy and Computing* 32



PySAP
Python Sparse data Analysis Package

pip install python-pysap

Miscellaneous contributions

Contributions

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

Thank you all!



Backup slides

Importance of MRI - 1

99.9% chance you will get an MRI.

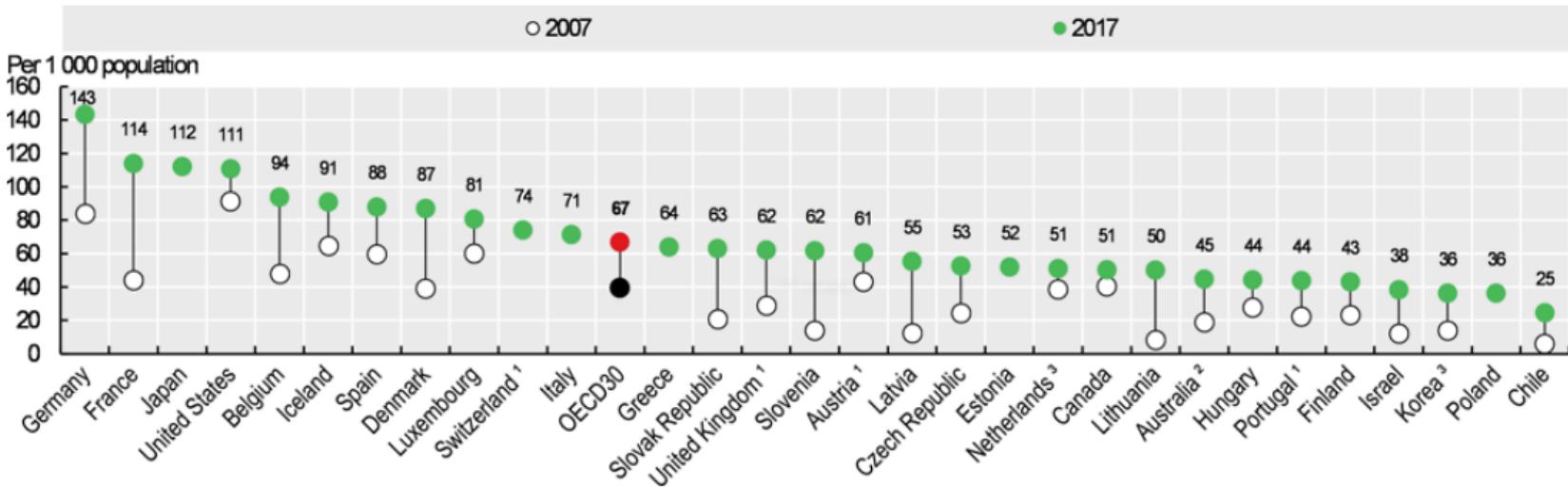


Figure: Number of MRI scans per year per 1000 population: figure courtesy of *Health at a Glance 2019: OECD Indicators - Medical technologies* (2019).

Importance of MRI - 2

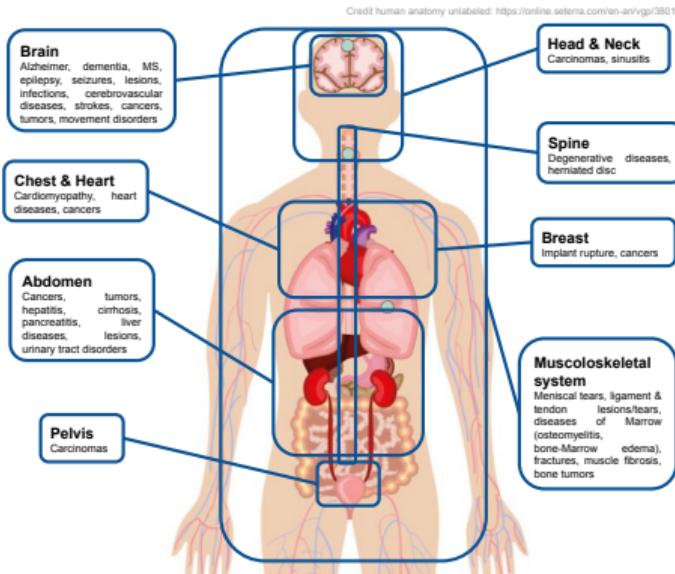


Figure: What can we diagnose with MRI? Info compiled from Reimer et al. (2010) and Runge et al. (2019).

Physics of MRI - 1

FID: global info.

Physics of MRI - 1

FID: global info.

Magnetic **gradients** ⇒ change the magnetic field spatially.

Physics of MRI - 1

FID: global info.

Magnetic **gradients** \Rightarrow change the magnetic field spatially.

Temporal signal:

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

Physics of MRI - 1

FID: global info.

Magnetic **gradients** \Rightarrow change the magnetic field spatially.

Temporal signal:

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

Recorded MR signal

Physics of MRI - 1

FID: global info.

Magnetic **gradients** \Rightarrow change the magnetic field spatially.

Temporal signal:

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

Recorded MR signal

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

Physics of MRI - 1

FID: global info.

Magnetic **gradients** \Rightarrow change the magnetic field spatially.

Temporal signal:

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

Recorded MR signal

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

Temporal gradients,
controlled by the operator

MR signal derivation

Bloch equations:

$$\frac{dM_{tr}}{dt} = -\frac{M_{tr}}{T_2}$$

$$\frac{dM_l}{dt} = \frac{M_0 - M_l}{T_1}$$

Transverse component of M

Equilibrium state

Longitudinal component of M

MR signal derivation

Solution:

$$\begin{aligned} |M_{tr}(0, r)| &= \frac{1}{4}\rho(r) \frac{\gamma^2 \hbar^2}{kT} B_0 \\ M_{tr}(t, r) &= M_{tr}(0, r) e^{-\frac{t}{T_2}} \\ M_I(t, r) &= M_I(0, r) e^{-\frac{t}{T_1}} + M_0(1 - e^{-\frac{t}{T_1}}) \end{aligned}$$

MR signal derivation

EF force in antenna:

$$S(t) = -\frac{d}{dt} \int_{V_s} \mathbf{B}_1 \cdot \mathbf{M}(t, \mathbf{r}) d\mathbf{r}$$

Fourier Transform and MRI

MR signal:

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

Fourier Transform of a signal:

$$\hat{f}(\omega) = \int_{-\infty}^{\infty} f(x) e^{-2\pi i \omega x} dx$$

The example of GRAPPA

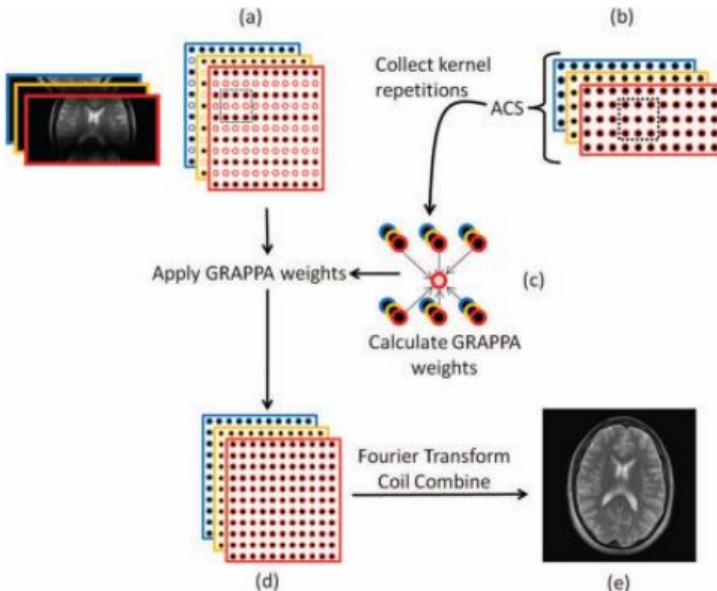


Figure: **GRAPPA illustration.** Image courtesy of Deshpande et al. (2012).

Noise model for MRI

Bayesian view of MRI reconstruction:

$$\arg \max_{\mathbf{x}} p(\mathbf{x}|\mathbf{y}) \propto p(\mathbf{y}|\mathbf{x})p(\mathbf{x})$$

If we consider an additive white Gaussian noise model:

$$p(\mathbf{y}|\mathbf{x}) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2\sigma^2}\|\mathcal{A}\mathbf{x} - \mathbf{y}\|^2}$$

We retrieve:

$$-\log p(\mathbf{y}|\mathbf{x}) \propto \|\mathcal{A}\mathbf{x} - \mathbf{y}\|^2 + \text{cst}$$

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.

Proximity operator

Definition:

$$\text{prox}_{\mathcal{R}}(\mathbf{x}) = \arg \min_{\mathbf{z} \in \mathcal{H}} \mathcal{R}(\mathbf{z}) + \frac{1}{2} \|\mathbf{z} - \mathbf{x}\|_2^2$$

2 intuitions:

- Prox. of indicator of \mathcal{C} , a convex set, is the projection onto \mathcal{C} .
- Prox. of a smooth function is its gradient step.

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:

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

The power of Deep Learning

The prior is a complicated visual function.

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

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

Neural network:
a chain of elementary linear & nonlinear functions

Formalism - 1

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

Supervised learning:

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



input



Formalism - 1

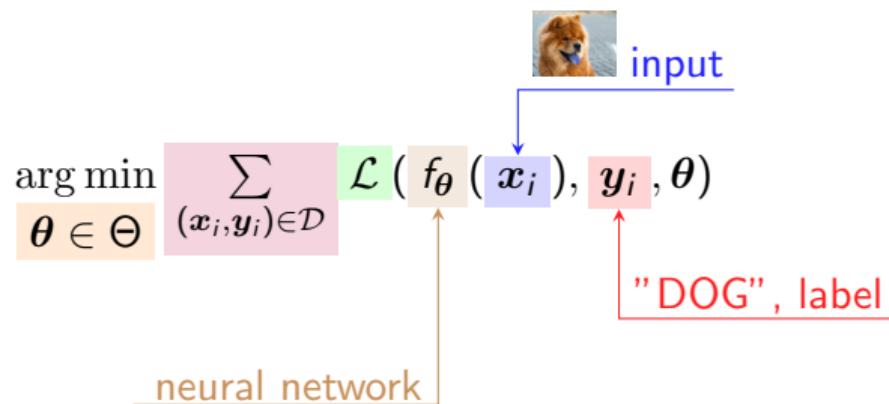
Supervised learning:

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

The diagram illustrates the supervised learning process. An input image of a dog is shown at the top, with a blue arrow labeled "input" pointing to the term $f_{\theta}(x_i)$ in the equation. A red arrow labeled "\"DOG\", label" points from the same term to the variable y_i , indicating that the input image corresponds to the label "DOG".

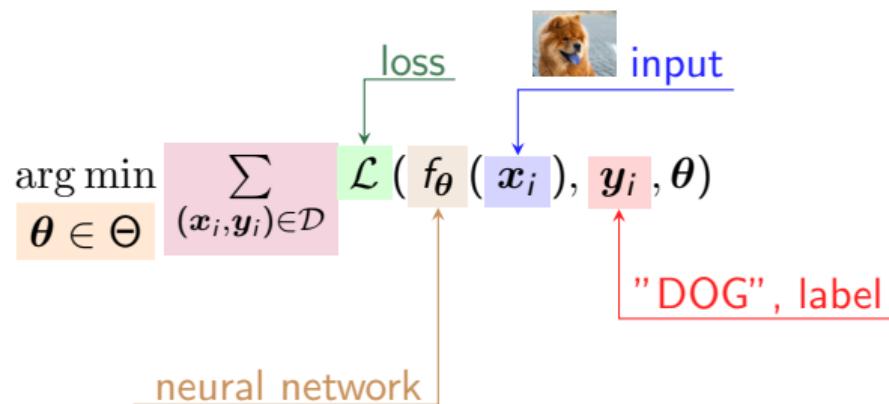
Formalism - 1

Supervised learning:



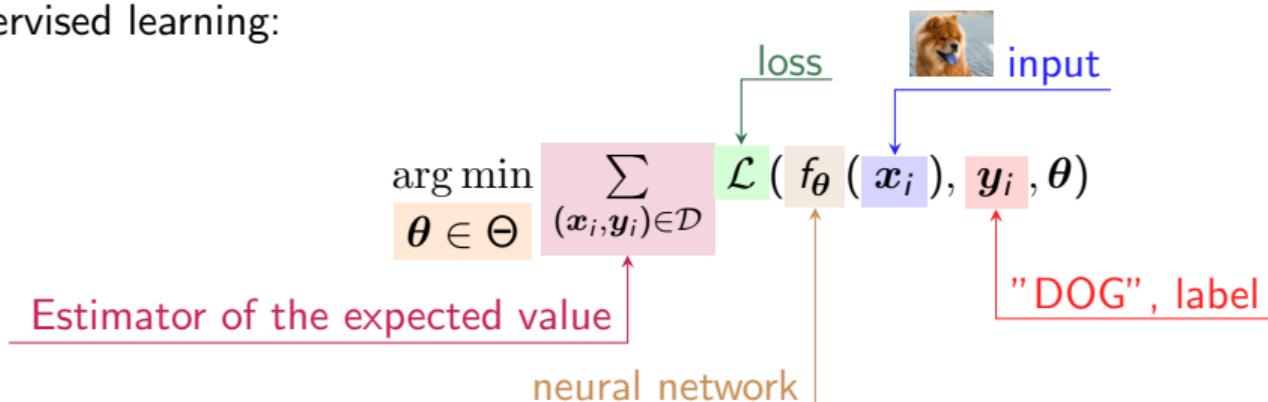
Formalism - 1

Supervised learning:



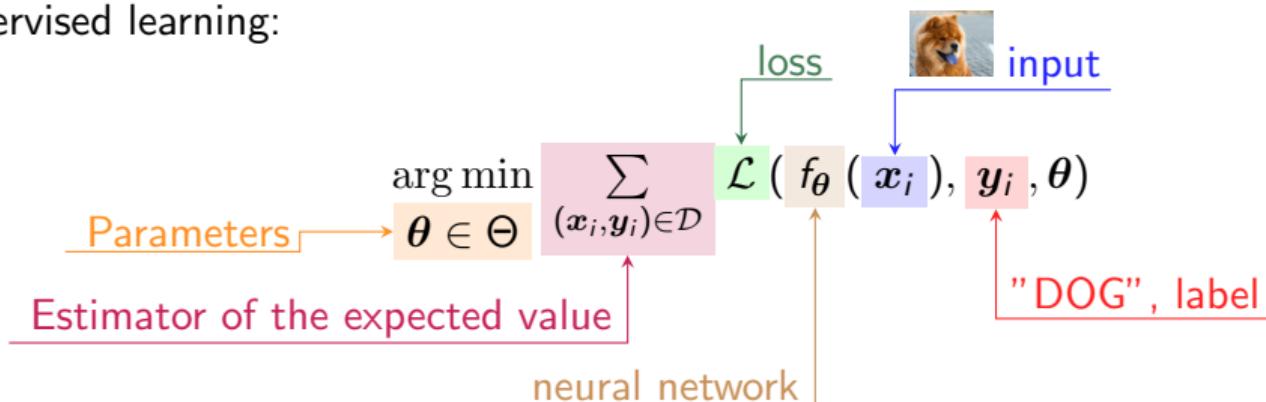
Formalism - 1

Supervised learning:



Formalism - 1

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.

$$\frac{\partial f}{\partial y} = \frac{\partial f}{\partial x} \frac{\partial x}{\partial y}$$

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"

CNN

Convolutional Neural Network (CNN): chain of Convolution + Nonlinearity.

U-net

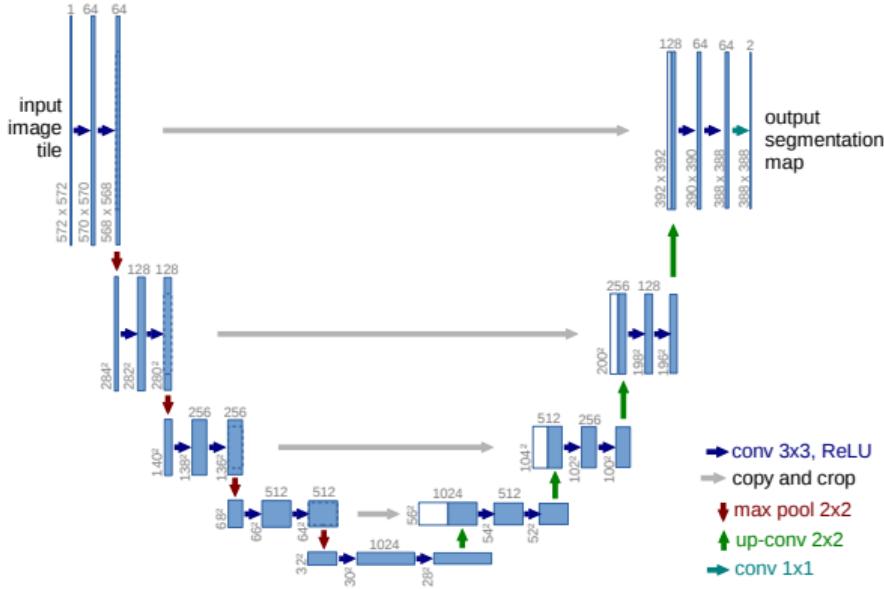


Figure: Illustration from the original paper.¹⁸

¹⁸O. Ronneberger et al. (2015). "U-net: Convolutional networks for biomedical image segmentation". In: *International Conference on Medical image computing and computer-assisted intervention*.

MWCNN

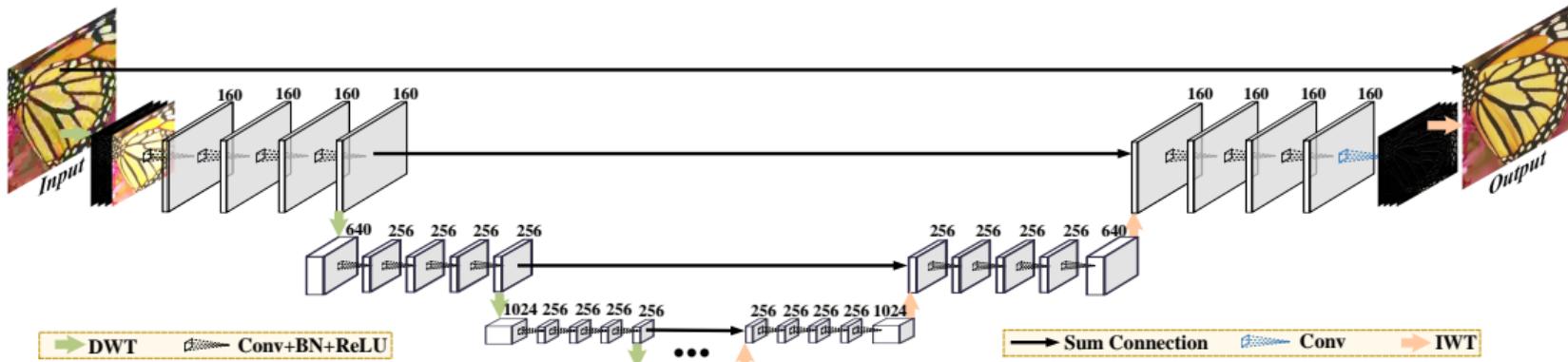


Figure: Illustration from the original paper.¹⁹

¹⁹P. Liu et al. (2018). "Multi-level Wavelet-CNN for Image Restoration". In: *CVPR NTIRE Workshop*.

fastMRI dataset

fastmri.org

Knee

- 973 train volumes, 199 validation volumes
- 2 contrasts: PD and PDFS
- 15 coils, 1.5T/3T, 320 x 320, 0.5 mm x 0.5 mm, Cartesian 2D TSE

Brain

- 4469 train volumes, 1378 validation volumes
- 4 contrasts: T1, T1 post injection, T2, FLAIR
- different locations, different coil architecture
- 1.5T/3T, 320 x 320 (with exceptions), 0.5 mm x 0.5 mm, Cartesian 2D TSE

2020 fastMRI challenge

- 2nd edition
- 8 teams
- Brain data
- 3 tracks: 4X, 8X, transfer

What did the winners do that I did not

- Unclear feature multi-domain learning
- 3D Post-processing (main network is 2D)
- Distributed training (4 GPUs)

autoMAP

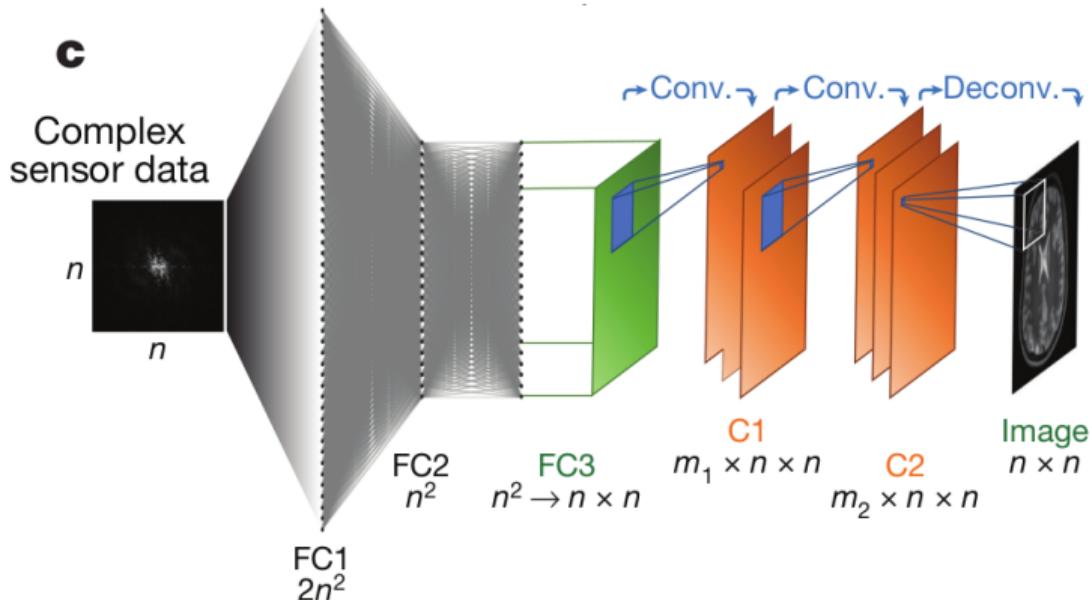


Figure: Illustration from the original paper.²⁰

²⁰B. Zhu et al. (Mar. 2018). "Image reconstruction by domain-transform manifold learning". In: *Nature* 555.7697, pp. 487–492.

XPDNet - ct'ed

Algorithm 1: XPDNet.

Data: y the k-space data, Ω the Cartesian trajectory, $\$$ the coarse estimates of the sensitivity maps

Result: x , the reconstructed magnitude MR image

```
1  $\$ = g_{\theta_r}(\$)$ ; // Sensitivity maps refinement
2 Update  $\mathcal{A}$  and  $\mathcal{A}^H$ ;
3  $x = \mathcal{A}^H y$ ;
4  $x_b = [x, x, x, x, x]$ ; // Buffer creation, in practice concatenation along the channel dimension
5 for  $i \leftarrow 1$  to  $N_C$  do
6    $\mathbf{y}_{res} = \mathcal{A} x_b[0] - y$ ; // Data consistency
7    $x_{dc} = \mathcal{A}^H \mathbf{y}_{res}$ ; // Density compensation
8    $x_b = x_b + f_{\theta_i}([x_b, x_{dc}]))$ ; // Proximity operator learning and nonlinear acceleration scheme
9  $x = |x_b[0]|$  // Magnitude computation
```

PDNet ops

Concatenation

We change from

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

to

$$\mathbf{x}_{n+1} = \text{prox}([\mathbf{x}_n; \mathcal{A}^H(\mathcal{A}\mathbf{x}_n - \mathbf{y})])$$

Buffer

We change from

$$\mathbf{x}_{n+1} = \nu \mathbf{x}_{n+1} + (1 - \nu) \mathbf{x}_n$$

to

$$\mathbf{x}_b = [\mathbf{x}, \mathbf{x}, \mathbf{x}, \mathbf{x}, \mathbf{x}]$$

PDNet & Recurrent Inference Machines (RIMs)

Bayesian Inverse Problem formulation:

$$\arg \max_x p(x|y) \propto p(y|x)p(x)$$

Updates:

$$x_{n+1} = x_n + \epsilon_n \nabla_x (\log p(y|x) + \log p(x))(x_n)$$

RIMs generalize to:²¹

$$x_{n+1} = x_n + g(\nabla_x (\log p(y|x))(x_n), x_n)$$

²¹P. Putzky et al. (2017). *Recurrent Inference Machines for Solving Inverse Problems*. Tech. rep.

NC-PDNet - ct'ed

Algorithm 2: NC-PDNet: Density compensated Primal Dual unrolled neural network over N_C iterations.

Data: y the k-space data, Ω the non-Cartesian trajectories, d the pre-computed DCp weights, $\$$ the coarse estimates of the sensitivity maps

Result: x , the reconstructed magnitude MR image

```
1 \$ =  $g_{\theta_r}(\$)$ ; // Sensitivity maps refinement
2 Update  $\mathcal{A}$  and  $\mathcal{A}^H$ ;
3  $x = \mathcal{A}^H y$ ;
4  $x_b = [x, x, x, x, x]$ ; // Buffer creation, in practice concatenation along the channel dimension
5 for  $i \leftarrow 1$  to  $N_C$  do
6    $y_{res} = \mathcal{A} x_b[0] - y$ ; // Data consistency
7    $x_{dc} = \mathcal{A}^H d y_{res}$ ; // Density compensation
8    $x_b = x_b + f_{\theta_i}([x_b, x_{dc}])$ ; // Proximity operator learning and nonlinear acceleration scheme
9  $x = |x_b[0]|$  // Magnitude computation
```

Density Compensation

$$d_{n+1} = \frac{d_n}{\mathcal{F}_\Omega \mathcal{F}_\Omega^H d_n}$$

Image quality metrics

Peak Signal to Noise Ratio (PSNR):

$$\text{PSNR}(x, \hat{x}) = 10 \log_{10} \left(\frac{\max_i x_i}{\frac{1}{n} \|x - \hat{x}\|_2^2} \right)$$

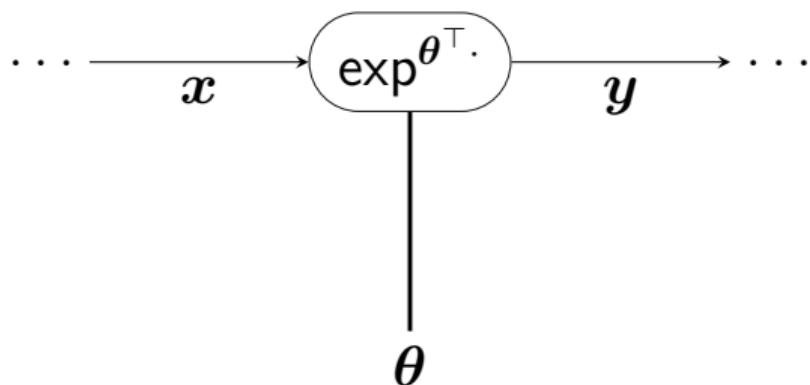
Structural Similarity Index (SSIM):

$$\text{SSIM}(x, \hat{x}) = [l(x, \hat{x})]^\alpha \cdot [c(x, \hat{x})]^\beta \cdot [s(x, \hat{x})]^\gamma$$

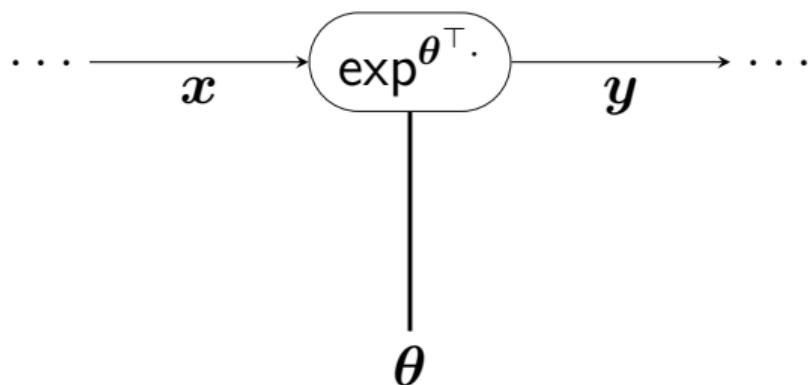
- l : luminance
- c : contrast
- s : structure

Problem: not a good correlation with the actual visual quality.

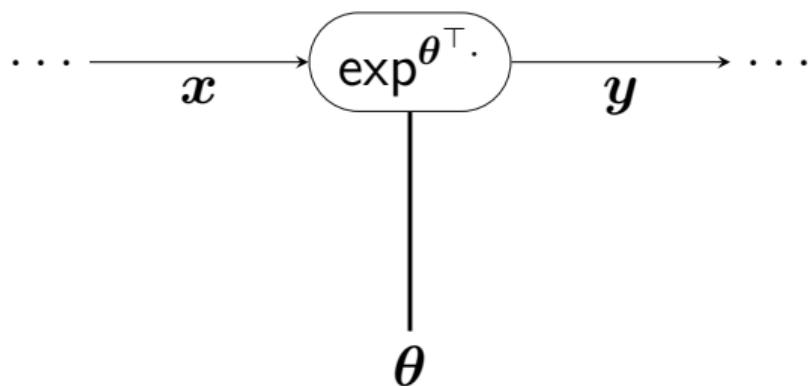
Example activation



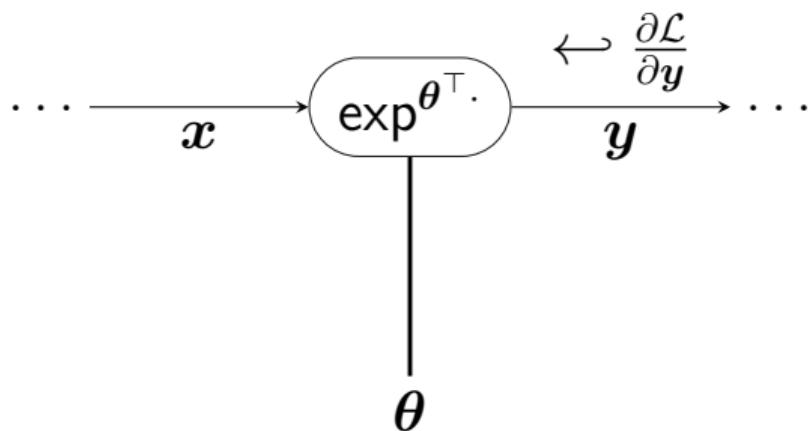
Example activation



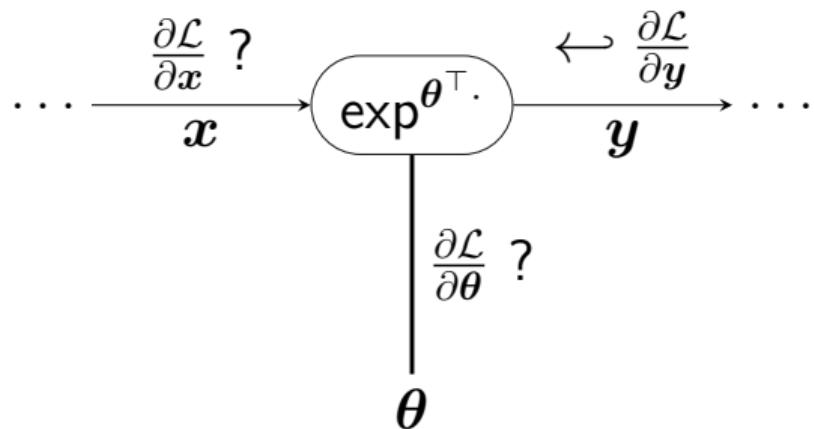
Example activation



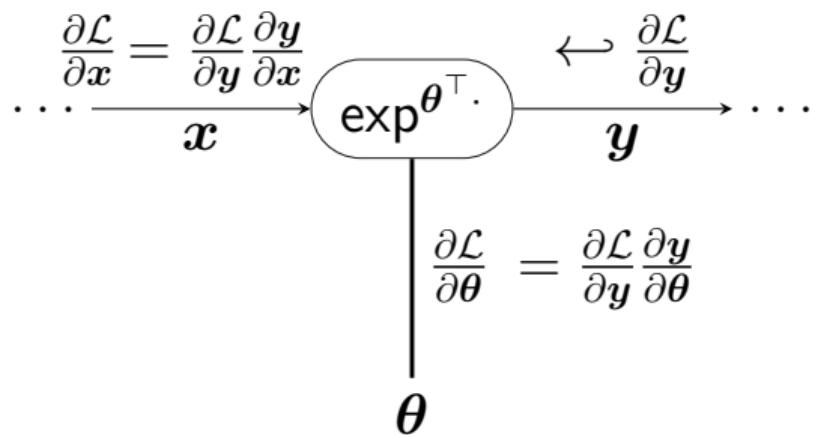
Example activation



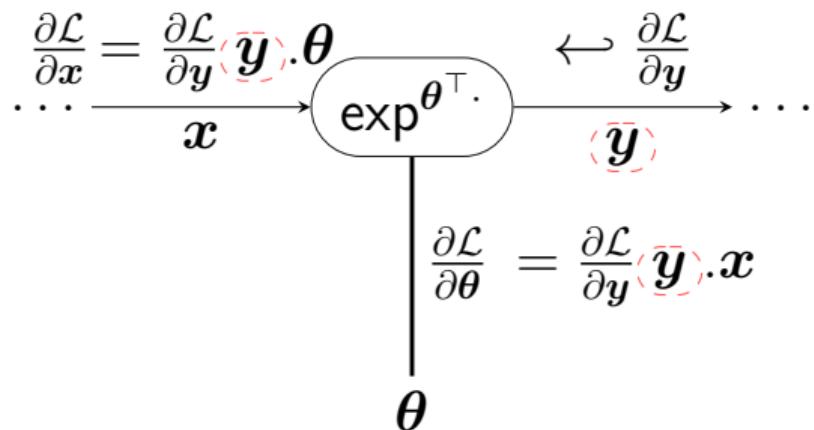
Example activation



Example activation



Example activation



quasi-Newton methods - 1

Idea: replace the costly Jacobian inverse with a qN matrix \mathbf{B}^{-1} .

$$f(x^*) = 0$$

Newton Methods

$$x_{n+1} = x_n - \frac{\partial f}{\partial x}(x_n)^{-1} f(x_n)$$

Quasi-Newton Methods

$$x_{n+1} = x_n - \mathbf{B}_n^{-1} f(x_n)$$

Update \mathbf{B}_n and its inverse with the
Sherma-Morrison formula.

quasi-Newton methods - 2

Secant conditions: set of conditions B must verify.

Typically: $B_n(x_n - x_{n-1}) = f(x_n) - f(x_{n-1})$.

Multiple solutions \Rightarrow regularization with $B_n = \arg \min_{B: B\Delta x_n = \Delta f_n} \|B - B_{n-1}\|$

SHINE direction convergence

Theorem (Convergence of SHINE to the Hypergradient using ULI)

Let us denote $p_{\theta}^{(n)}$, the SHINE direction for iterate n . Under the Uniform Linear Independence (ULI) assumption and some additional smoothness and convexity assumptions, for a given parameter θ , (z_n) converges q -superlinearly to z^* and

$$\lim_{n \rightarrow \infty} p_{\theta}^{(n)} = \frac{\partial \mathcal{L}}{\partial \theta} \Big|_{z^*}.$$

OPA - 1

Outer Problem Awareness: modify the inner problem resolution depending on the outer problem.

Additional updates of \mathbf{B} with the OPA direction: $e_n = t_n \mathbf{B}_n^{-1} \frac{\partial g_\theta}{\partial \theta} \Big|_{z_n}$.

OPA - 2

Algorithm LBFGS: (Limited memory) BFGS method with OPA.

Input: initial guess (z_0, B_0^{-1}) , where B_0^{-1} is symmetric and positive definite,
tolerance $\epsilon > 0$, frequency of additional updates $M \in \mathbb{N}$, memory limit
 $L \in \mathbb{N} \cup \{\infty\}$, (t_n) a null sequence of positive numbers with $\sum_n t_n < \infty$

```
1 Let  $F := \nabla_z g_\theta$ 
2 for  $n = 0, 1, 2, \dots$  do
3   if  $\|F(z_n)\| \leq \epsilon$  then let  $z^* := z_n$  and let  $B := B_n$ ; STOP
4   Let  $\hat{B}_n^{-1} := B_n^{-1}$ 
5   If  $(n \bmod M) = 0$  let  $e_n := t_n B_n^{-1} \frac{\partial g_\theta}{\partial \theta} \Big|_{z_n}$ ,  $\hat{y}_n := F(z_n + e_n) - F(z_n)$  and
        $\hat{r}_n := (e_n)^\top \hat{y}_n$ 
6   If  $\hat{r}_n > 0$  let  $\hat{a}_n := e_n - B_n^{-1} \hat{y}_n$  and let
       
$$\hat{B}_n^{-1} := B_n^{-1} + \frac{\hat{a}_n (e_n)^\top + e_n (\hat{a}_n)^\top}{\hat{r}_n} - \frac{(\hat{a}_n)^\top \hat{y}_n}{(\hat{r}_n)^2} e_n (e_n)^\top$$

       Let  $B_n^{-1} := \hat{B}_n^{-1}$ 
7   if  $n \geq L$  then remove update  $n - L$  from  $B_n^{-1}$ 
8   Let  $p_n := -B_n^{-1} F(z_n)$ 
9   Obtain  $\alpha_n$  via line-search and let  $s_n := \alpha_n p_n$ 
10  Let  $z_{n+1} := z_n + s_n$ ,  $y_n := F(z_{n+1}) - F(z_n)$  and  $r_n := (s_n)^\top y_n$ 
11  if  $r_n > 0$  then
12    let  $a_n := s_n - B_n^{-1} y_n$  and let
        
$$B_{n+1}^{-1} := B_n^{-1} + \frac{a_n (s_n)^\top + s_n (a_n)^\top}{r_n} - \frac{(a_n)^\top y_n}{(r_n)^2} s_n (s_n)^\top$$

13  else let  $B_{n+1}^{-1} := B_n^{-1}$ 
14  if  $n \geq L$  then remove update  $n - L$  from  $B_{n+1}^{-1}$ 
Output:  $z^*, B$ 
```

Adjoint Broyden

For DEQs, vanilla OPA direction involves $\frac{\partial g_\theta}{\partial \theta} \Big|_{z_n} \Rightarrow$ inefficient.

Recall that the SHINE direction is: $p_\theta = \nabla_z \mathcal{L}(z^*) B^{-1} \frac{\partial g_\theta}{\partial \theta} \Big|_{z^*}$.

Other direction, $\nabla_z \mathcal{L}(z^*)$, in left-multiplication \Rightarrow Adjoint Broyden for the secant condition to work.

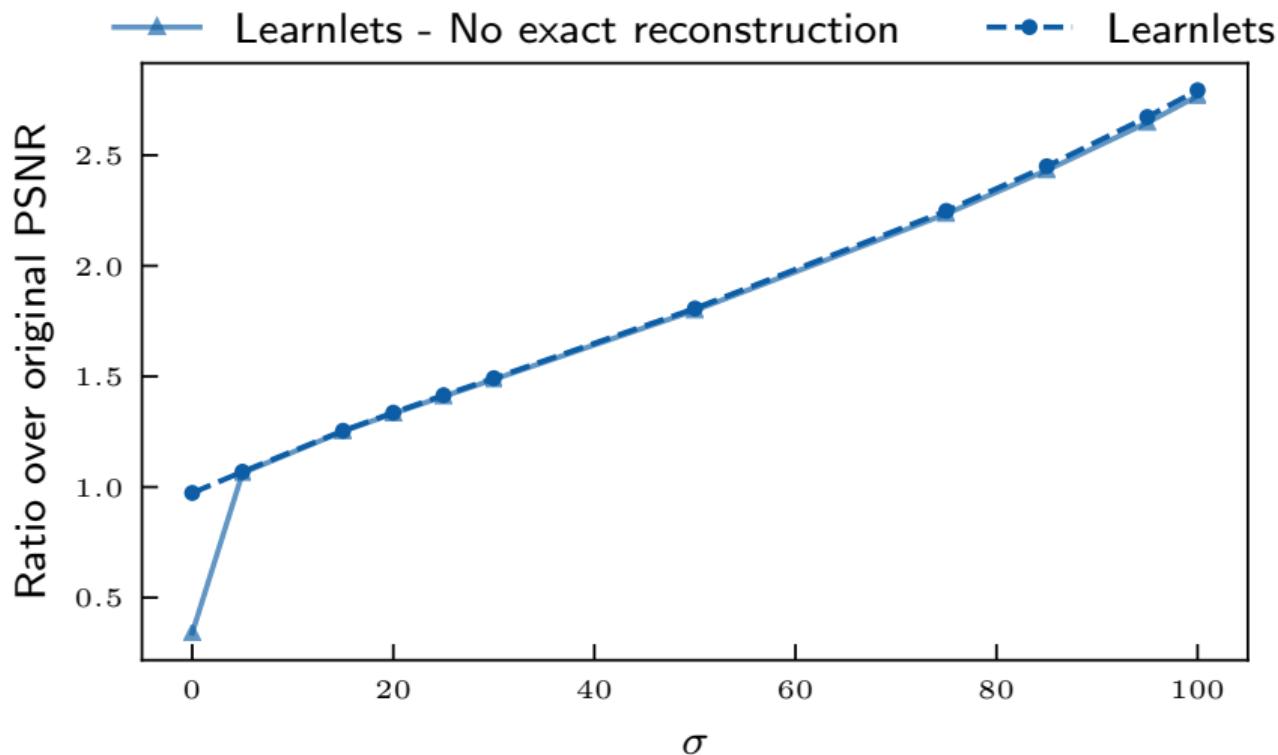
Jacobian-Free

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

Jacobian-Free

$$\mathbf{I} \approx J_{g_\theta}(z^\star)^{-1}$$

Learnlets - exact reconstruction



References |

-  Adler, J. and O. Öktem (2018). "Learned Primal-Dual Reconstruction". In: *IEEE Transactions on Medical Imaging* 37.6, pp. 1322–1332.
-  Alain, G. and Y. Bengio (2013). "What regularized auto-encoders learn from the data generating distribution". In: *ICLR 2013*. Vol. 15, pp. 3743–3773.
-  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.

References II

-  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.
-  Farrens, S., A. Grigis, L. El Gueddari, **Zaccharie Ramzi**, C. G R, S. Starck, B. Sarthou, H. Cherkaoui, P. Ciuciu, and J.-L. Starck (2020). "PySAP: Python Sparse Data Analysis Package for multidisciplinary image processing". In: *Astronomy and Computing* 32.
-  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.

References III

-  Grimm, R. (2015). *Reconstruction Techniques for Dynamic Radial MRI*. Friedrich-Alexander-Universitaet Erlangen-Nuernberg (Germany).
-  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.
-  Koyasu, H. and K. Shiba (Aug. 2018). "KNC can scan three more patients per day with Compressed SENSE". In: *FieldStrength*.
-  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*.
-  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.

References IV

-  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 V

-  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.
-  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.
-  Putzky, P. and M. Welling (2017). *Recurrent Inference Machines for Solving Inverse Problems*. Tech. rep.

References VI

-  Ramzi, Z., C. G R, J.-L. Starck, and P. Ciuciu (2022). "NC-PDNet: a Density-Compensated Unrolled Network for 2D and 3D non-Cartesian MRI Reconstruction". In: *IEEE Transactions on Medical Imaging*.
-  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.
-  Sander, M. E., P. Ablin, M. Blondel, and G. Peyré (2021). "Momentum Residual Neural Networks". In: *ICML*.

References VII

-  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.
-  Vincent, P. (2011). "A connection between scorematching and denoising autoencoders". In: *Neural Computation* 23.7, pp. 1661–1674.
-  **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 VIII

-  **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 2022*. (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*.

References IX

-  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.