Jeffrey A. Fessler

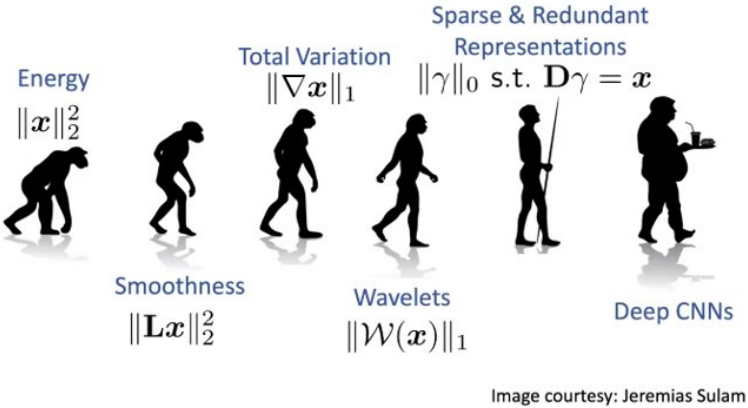
EECS Dept., BME Dept., Dept. of Radiology
University of Michigan

http://web.eecs.umich.edu/~fessler

ISBI 2023 Tutorial
2023-04-18

https://web.eecs.umich.edu/~fessler/talk/23/isbi.pdf

Overview



Energy
$\|x\|_2^2$
Smoothness
$\|\mathbf{L}x\|_2^2$
Total Variation
$\|\nabla x\|_1$
Wavelets
$\|\mathcal{W}(x)\|_1$
Sparse & Redundant Representations
$\|\gamma\|_0$ s.t. $\mathbf{D}\gamma = x$
Deep CNNs
Image courtesy: Jeremias Sulam

- Introduction
  - Measurement model review
- Data-driven regularizers
- Sparsity regularizers: Basic
- Sparsity regularizers: Advanced
- Denoising-based “regularization”
- Deep-learning approaches for image reconstruction
- Learning strategies
- Summary
- Bibliography

Measurement model in MRI

RF pulses
gradient waveforms
tissue properties $\rightarrow$ $\boxed{\text{Bloch Eqn} \\ \text{Nonlinear}}$ $\rightarrow$ transverse magnetization $\boldsymbol{x}$ $\rightarrow$ $\boldsymbol{A}$ encoding (coil, phase, ...) Linear $\rightarrow$ recorded data $\boldsymbol{y}$

Simplified data model [1, 2, 3]:

$$\begin{bmatrix} \boldsymbol{y}_1 \\ \vdots \\ \boldsymbol{y}_L \end{bmatrix} = \boldsymbol{y} = \boldsymbol{A}\boldsymbol{x} + \boldsymbol{\varepsilon}, \quad \boldsymbol{A} = (\boldsymbol{I}_L \otimes \boldsymbol{F})\boldsymbol{C}, \quad F_{ij} = \exp(-\imath 2\pi \vec{\nu}_i \cdot \vec{r}_j), \quad \boldsymbol{C} = \begin{bmatrix} \boldsymbol{C}_1 \\ \vdots \\ \boldsymbol{C}_L \end{bmatrix}$$

- $\boldsymbol{y}_l \in \mathbb{C}^M$: noisy samples recorded by the lth of of L receive coils
- $\boldsymbol{x} \in \mathbb{C}^N$: discretized version of the unknown transverse magnetization
- $\boldsymbol{\varepsilon} \in \mathbb{C}^{ML}$: complex white Gaussian noise [4]
- $\vec{\nu}_i$: k-space sample location of the ith sample (units cycles/cm)
- $\vec{r}_j$: spatial coordinates of the center of the jth pixel (units cm)
- $\boldsymbol{F} \in \mathbb{C}^{M \times N}$: Fourier encoding matrix; $\otimes$: Kronecker product
- $\boldsymbol{C}_l$: $N \times N$ diagonal matrix containing the lth coil sensitivity pattern.
- $\boldsymbol{A} \in \mathbb{C}^{LM \times N}$: system matrix

Extensions consider other physics effects like relaxation and field inhomogeneity [3].

Image reconstruction goal

- Data model:

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

- Goal: Estimate image $\boldsymbol{x}$ from data $\boldsymbol{y}$
-

Image reconstruction goal

- Data model:

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

- Goal: Estimate image $\boldsymbol{x}$ from data $\boldsymbol{y}$
- Regularization is essential for
  - under-sampled problems ($ML < N$) and compressed sensing ($M < N$)
  - poorly conditioned problems, *e.g.*, non-Cartesian sampling

Compressed sensing vs neural networks?

Received: 1 January 2020 | Revised: 28 April 2020 | Accepted: 30 April 2020

DOI: 10.1002/mrm.28338

FULL PAPER

Magnetic Resonance in Medicine

Advancing machine learning for MR image reconstruction with an open competition: Overview of the 2019 fastMRI challenge

Florian Knoll[1] | Tullie Murrell[2] | Anuroop Sriram[2] | Nafissa Yakubova[2] | Jure Zbontar[2] | Michael Rabbat[2] | Aaron Defazio[2] | Matthew J. Muckley[1] | Daniel K. Sodickson[1] | C. Lawrence Zitnick[2] | Michael P. Recht[1]

"the winners … chose approaches that used a combination of a learned prior and a data-fidelity term that encodes information about the MR physics of the acquisition, in line with approaches that can be seen as neural network extensions of classic iterative image reconstruction methods" [5]

Outline

Outline

- Synthesis model:
  Assume $\boldsymbol{x} = \boldsymbol{B}\boldsymbol{z}$
  $\boldsymbol{B}$: $N \times K$ matrix ("basis"), *e.g.*, wavelets, often wide (over complete)
  $\boldsymbol{z} \in \mathbb{C}^K$ sparse coefficient vector
  $\Longrightarrow$ use $\|\boldsymbol{z}\|_1$

- Analysis model:
  Assume $\boldsymbol{T}\boldsymbol{x}$ is sparse
  $\boldsymbol{T}$: $K \times N$ transformation matrix, usually tall,
  e.g., finite differences for total variation (TV)
  $\Longrightarrow$ use $\|\boldsymbol{T}\boldsymbol{x}\|_1$

- Equivalent if $\boldsymbol{B} = \boldsymbol{T}^{-1}$ (but usually both are non-square)
- Conventionally $\boldsymbol{B}$ and $\boldsymbol{T}$ are "hand crafted"
-

- Synthesis model:
  Assume $\boldsymbol{x} = \boldsymbol{B}\boldsymbol{z}$
  $\boldsymbol{B}$: $N \times K$ matrix ("basis"), *e.g.*, wavelets, often wide (over complete)
  $\boldsymbol{z} \in \mathbb{C}^K$ sparse coefficient vector
  $\Longrightarrow$ use $\|\boldsymbol{z}\|_1$

- Analysis model:
  Assume $\boldsymbol{T}\boldsymbol{x}$ is sparse
  $\boldsymbol{T}$: $K \times N$ transformation matrix, usually tall,
  e.g., finite differences for total variation (TV)
  $\Longrightarrow$ use $\|\boldsymbol{T}\boldsymbol{x}\|_1$

- Equivalent if $\boldsymbol{B} = \boldsymbol{T}^{-1}$ (but usually both are non-square)
- Conventionally $\boldsymbol{B}$ and $\boldsymbol{T}$ are "hand crafted"
- All models are wrong, but some models are useful...
  Most likely used in ≈2017 US FDA-approved CS methods [6, 7, 8, 9] .

Analysis regularization for image reconstruction

- Typical optimization problem for analysis sparsity model:

$$\hat{\boldsymbol{x}} = \arg\min_{\boldsymbol{x}} \frac{1}{2} \|\boldsymbol{A}\boldsymbol{x} - \boldsymbol{y}\|_2^2 + \beta \|\boldsymbol{T}\boldsymbol{x}\|_1 \tag{1}$$

  $\boldsymbol{T}$: sparsifying operator
  - wavelet transform
  - finite differences, aka total variation (TV) [10], or both [11]
- FDA-approved methods for compressed sensing MRI presumably related to (1).
- Non-trivial optimization problem due to the matrix $\boldsymbol{T}$ within 1-norm.

- Proximal gradient method (PGM) for analysis regularizer problem (1):

$$\tilde{\boldsymbol{x}}_k \triangleq \boldsymbol{x}_k - \frac{1}{L}\boldsymbol{A}'(\boldsymbol{A}\boldsymbol{x}_k - \boldsymbol{y}) \quad \text{(gradient step, aka ``data consistency'')}$$

$$\boldsymbol{x}_{k+1} = \arg\min_{\boldsymbol{x}} \frac{L}{2}\|\boldsymbol{x} - \tilde{\boldsymbol{x}}_k\|_2^2 + \beta\,\|\boldsymbol{T}\boldsymbol{x}\|_1 = \operatorname{prox}_{\frac{\beta}{L}\|\boldsymbol{T}\cdot\|_1}(\tilde{\boldsymbol{x}}_k) \quad \text{(denoising step)} \qquad (2)$$

$L = \|\boldsymbol{A}\|_2^2$: Lipschitz constant

-

- Proximal gradient method (PGM) for analysis regularizer problem (1):

$$\tilde{\boldsymbol{x}}_k \triangleq \boldsymbol{x}_k - \frac{1}{L} \boldsymbol{A}'(\boldsymbol{A}\boldsymbol{x}_k - \boldsymbol{y}) \quad \text{(gradient step, aka ``data consistency'')}$$

$$\boldsymbol{x}_{k+1} = \arg\min_{\boldsymbol{x}} \frac{L}{2} \|\boldsymbol{x} - \tilde{\boldsymbol{x}}_k\|_2^2 + \beta \|\boldsymbol{T}\boldsymbol{x}\|_1 = \operatorname{prox}_{\frac{\beta}{L}\|\boldsymbol{T}\cdot\|_1}(\tilde{\boldsymbol{x}}_k) \quad \text{(denoising step)} \qquad (2)$$

  $L = \|\boldsymbol{A}\|_2^2$: Lipschitz constant
- Many alternative algorithms (ADMM, POGM, primal-dual, ...). Survey: [12]
-

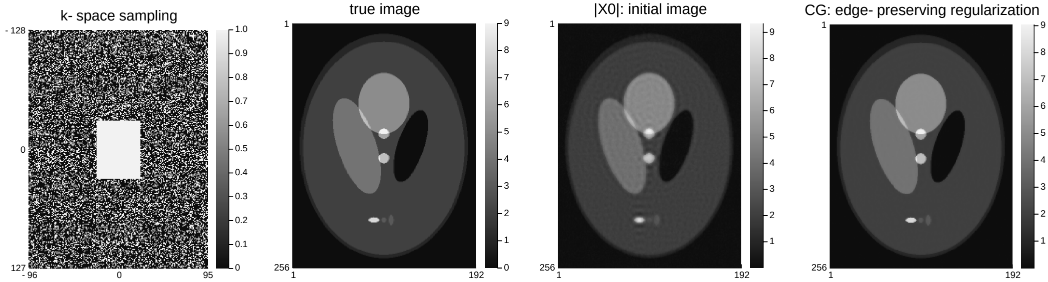
Proximal methods for analysis regularizer

- Proximal gradient method (PGM) for analysis regularizer problem (1):

$$\tilde{\boldsymbol{x}}_k \triangleq \boldsymbol{x}_k - \frac{1}{L} \boldsymbol{A}'(\boldsymbol{A}\boldsymbol{x}_k - \boldsymbol{y}) \quad \text{(gradient step, aka “data consistency”)}$$

$$\boldsymbol{x}_{k+1} = \arg\min_{\boldsymbol{x}} \frac{L}{2} \|\boldsymbol{x} - \tilde{\boldsymbol{x}}_k\|_2^2 + \beta \|\boldsymbol{T}\boldsymbol{x}\|_1 = \operatorname{prox}_{\frac{\beta}{L}\|\boldsymbol{T}\cdot\|_1}(\tilde{\boldsymbol{x}}_k) \quad \text{(denoising step)} \qquad (2)$$

  $L = \|\boldsymbol{A}\|_2^2$: Lipschitz constant
- Many alternative algorithms (ADMM, POGM, primal-dual, ...). Survey: [12]
- Common ingredients: data consistency and denoising
  cf. deep learning reconstruction



ψ: Fair potential, $\delta = 0.1$

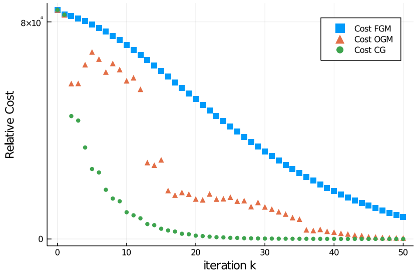
$\boldsymbol{T}$: finite differences

$=$ corner-rounded TV

Demo notebook: `01-recon`

`https://github.com/JeffFessler/mirt-demo`

Final NRMSE: 1.55%



Outline

Towards data-driven regularization

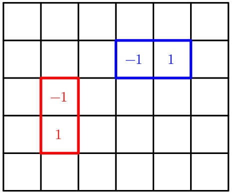
- Classical regularizers use “hand crafted” transform $\boldsymbol{T}$ or basis $\boldsymbol{B}$
- Learning $\boldsymbol{T}$ or $\boldsymbol{B}$ for entire image is impractical
-

Towards data-driven regularization

- Classical regularizers use "hand crafted" transform $\boldsymbol{T}$ or basis $\boldsymbol{B}$
- Learning $\boldsymbol{T}$ or $\boldsymbol{B}$ for entire image is impractical
- Learned regularizers are often patch based

Patch-based regularization

Using TV regularizer $R(\boldsymbol{x}) = \|\boldsymbol{T}\boldsymbol{x}\|_1$
where $\boldsymbol{T}$ is finite-differences
$\equiv$ patches of size 2×1.

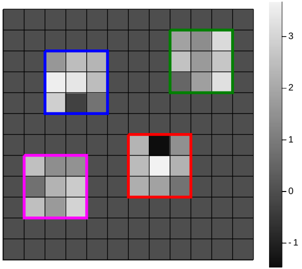


Larger patches provide more context for distinguishing signal from noise.

cf. CNN approaches

Patch-based regularizers:

- synthesis models
- analysis methods



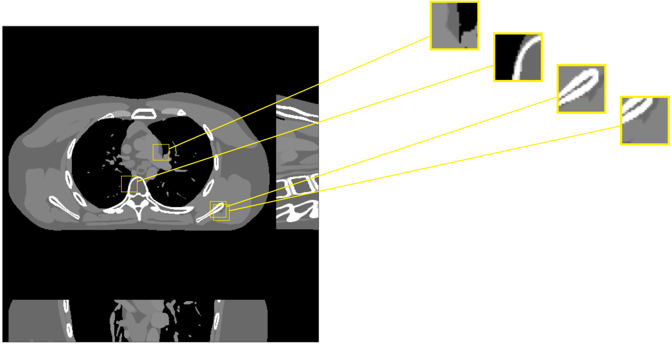
Patch-wise dictionary sparsity model

Assumption: if $\boldsymbol{x}$ is a plausible image, then each patch has

$$\boldsymbol{P}_p \boldsymbol{x} \approx \boldsymbol{D} \boldsymbol{z}_p,$$

for a sparse coefficient vector $\boldsymbol{z}_p$. (Synthesis approach.)

- $\boldsymbol{P}_p \boldsymbol{x}$ extracts the pth of P patches from $\boldsymbol{x}$
- $\boldsymbol{D}$ is a (typically overcomplete) dictionary for patches



Patch-based regularization: synthesis approach

- Patch synthesis model uses sparse linear combination of patch atoms: $\boldsymbol{P}_p \boldsymbol{x} \approx \boldsymbol{D} \boldsymbol{z}_p$
  $\boldsymbol{P}_p \in \{0,1\}^{d \times N}$: extracts pth of P d-pixel patches from image $\boldsymbol{x}$
  $\boldsymbol{D} \in \mathbb{C}^{d \times J}$: dictionary of J patch atoms
  $\boldsymbol{z}_p \in \mathbb{C}^{J}$: sparse coefficient vector for pth patch.
- Natural regularizer for patch synthesis sparsity model [13]:

$$\hat{\boldsymbol{x}} = \arg\min_{\boldsymbol{x}} \frac{1}{2} \|\boldsymbol{A}\boldsymbol{x} - \boldsymbol{y}\|_2^2 + \beta R(\boldsymbol{x}), \quad R(\boldsymbol{x}) = \min_{\{\boldsymbol{z}_p\}} \sum_{p=1}^{P} \frac{1}{2} \|\boldsymbol{P}_p \boldsymbol{x} - \boldsymbol{D}\boldsymbol{z}_p\|_2^2 + \alpha \|\boldsymbol{z}_p\|_1 .$$

- Three options for patch dictionary $\boldsymbol{D}$
  - Hand crafted
  - Learned from population training data images, *e.g.*, K-SVD [14], SOUP [15]
  -

Patch-based regularization: synthesis approach

- Patch synthesis model uses sparse linear combination of patch atoms: $\boldsymbol{P}_p \boldsymbol{x} \approx \boldsymbol{D} \boldsymbol{z}_p$
  $\boldsymbol{P}_p \in \{0,1\}^{d \times N}$: extracts pth of P d-pixel patches from image $\boldsymbol{x}$
  $\boldsymbol{D} \in \mathbb{C}^{d \times J}$: dictionary of J patch atoms
  $\boldsymbol{z}_p \in \mathbb{C}^{J}$: sparse coefficient vector for pth patch.
- Natural regularizer for patch synthesis sparsity model [13]:

$$\hat{\boldsymbol{x}} = \arg\min_{\boldsymbol{x}} \frac{1}{2} \|\boldsymbol{A}\boldsymbol{x} - \boldsymbol{y}\|_2^2 + \beta R(\boldsymbol{x}), \quad R(\boldsymbol{x}) = \min_{\{\boldsymbol{z}_p\}} \min_{\boldsymbol{D} \in \mathcal{D}} \sum_{p=1}^{P} \frac{1}{2} \|\boldsymbol{P}_p \boldsymbol{x} - \boldsymbol{D} \boldsymbol{z}_p\|_2^2 + \alpha \|\boldsymbol{z}_p\|_1 .$$

- Three options for patch dictionary $\boldsymbol{D}$
  - Hand crafted
  - Learned from population training data images, *e.g.*, K-SVD [14], SOUP [15]
  - Learn while reconstructing this patient ("blind")
-

Patch-based regularization: synthesis approach

- Patch synthesis model uses sparse linear combination of patch atoms: $\boldsymbol{P}_p \boldsymbol{x} \approx \boldsymbol{D} \boldsymbol{z}_p$
  $\boldsymbol{P}_p \in \{0,1\}^{d \times N}$: extracts pth of P d-pixel patches from image $\boldsymbol{x}$
  $\boldsymbol{D} \in \mathbb{C}^{d \times J}$: dictionary of J patch atoms
  $\boldsymbol{z}_p \in \mathbb{C}^{J}$: sparse coefficient vector for pth patch.
- Natural regularizer for patch synthesis sparsity model [13]:

$$\hat{\boldsymbol{x}} = \arg\min_{\boldsymbol{x}} \frac{1}{2} \|\boldsymbol{A}\boldsymbol{x} - \boldsymbol{y}\|_2^2 + \beta R(\boldsymbol{x}), \quad R(\boldsymbol{x}) = \min_{\{\boldsymbol{z}_p\}} \min_{\boldsymbol{D} \in \mathcal{D}} \sum_{p=1}^{P} \frac{1}{2} \|\boldsymbol{P}_p \boldsymbol{x} - \boldsymbol{D}\boldsymbol{z}_p\|_2^2 + \alpha \|\boldsymbol{z}_p\|_1 .$$

- Three options for patch dictionary $\boldsymbol{D}$
  - Hand crafted
  - Learned from population training data images, *e.g.*, K-SVD [14], SOUP [15]
  - Learn while reconstructing this patient (“blind”)
- Use alternating minimization algorithms for optimization

Properties of a convex regularizer

When learning the regularizer, convex functions allow most theory and avoid local minima!

In the interpretation $R(u) = -\log(p(u))$, a convex regularizer makes no sense for images!

0.5  $+0.5$  $=$ 

$$\frac{1}{2}R(u_1) + \frac{1}{2}R(u_2) \geq R\left(\frac{1}{2}u_1 + \frac{1}{2}u_2\right)$$

Needs to have lower R(u) than one of the other images!

July 10 2020 Daniel Cremers, Michael Möller

- Dictionary-blind MR image reconstruction:

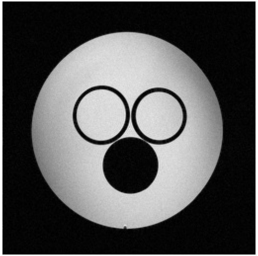
$$\hat{\boldsymbol{x}} = \arg\min_{\boldsymbol{x}} \frac{1}{2} \|\boldsymbol{A}\boldsymbol{x} - \boldsymbol{y}\|_2^2 + \beta\, \mathrm{R}(\boldsymbol{x})$$

$$\mathrm{R}(\boldsymbol{x}) = \min_{\boldsymbol{D} \in \mathcal{D}} \min_{\boldsymbol{Z}} \sum_{p=1}^{P} \left(\|\boldsymbol{P}_p \boldsymbol{x} - \boldsymbol{D}\boldsymbol{z}_p\|_2^2 + \lambda^2 \|\boldsymbol{z}_p\|_0 \right)$$

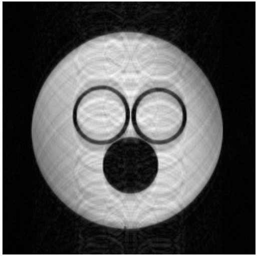
  $\boldsymbol{P}_p$: extracts pth of P image patches.
  $\mathcal{D}$: set of dictionaries with unit-norm atoms
  In words: of the many images...
- Alternating (nested) minimization:
  - Fixing $\boldsymbol{x}$ and $\boldsymbol{D}$, update each row of $\boldsymbol{Z} = [\boldsymbol{z}_1 \ \ldots \ \boldsymbol{z}_P]$ sequentially via hard-thresholding.
  - Fixing $\boldsymbol{x}$ and $\boldsymbol{Z}$, update $\boldsymbol{D}$ using SOUP-DIL [15].
  - Fixing $\boldsymbol{Z}$ and $\boldsymbol{D}$, updating $\boldsymbol{x}$ is a quadratic problem.
    - Efficient FFT solution for single-coil Cartesian MRI.
    - Use CG for non-Cartesian and/or parallel MRI.
- Non-convex due to $\mathcal{D}$, $\boldsymbol{D}\boldsymbol{z}_p$, 0-norm, but monotone decreasing and some convergence theory [15].

2D CS MRI results with blind DL I

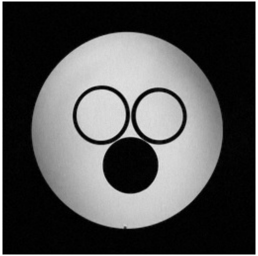
Fully Sampled



Zero-Filled



SOUP-DILLO-MRI

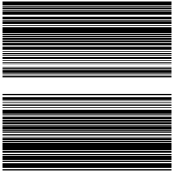


6×6 patches

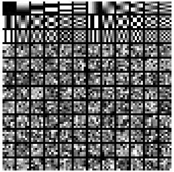
$\boldsymbol{D} \in \mathbb{C}^{6^2 \times 144}$

$\boldsymbol{D}_0$: [DCT | random] [15]

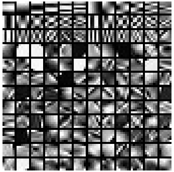
Sampling ($2.5\times$)



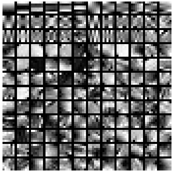
Initial $\boldsymbol{D}$



Learned real$\{\boldsymbol{D}\}$

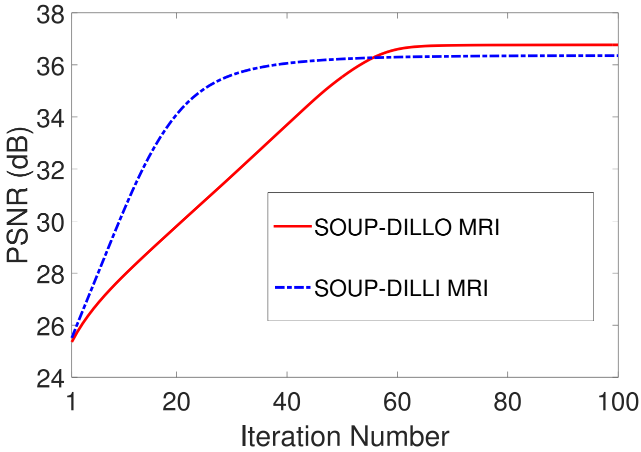


imag$\{\boldsymbol{D}\}$



todo: Would be interesting to see which atoms are most used.

2D CS MRI results with blind DL II

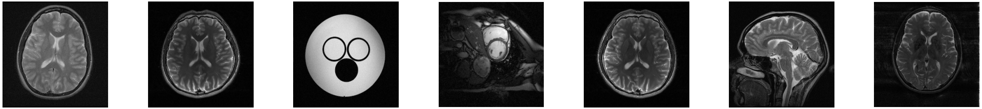


(SNR vs fully sampled image.)
Using $\|z_m\|_0$ leads to higher SNR than $\|z_m\|_1$.
Adaptive case is non-convex anyway...

Matlab code: http://web.eecs.umich.edu/~fessler/irt/reproduce/

https://gitlab.eecs.umich.edu/fessler/soupdil_dinokat

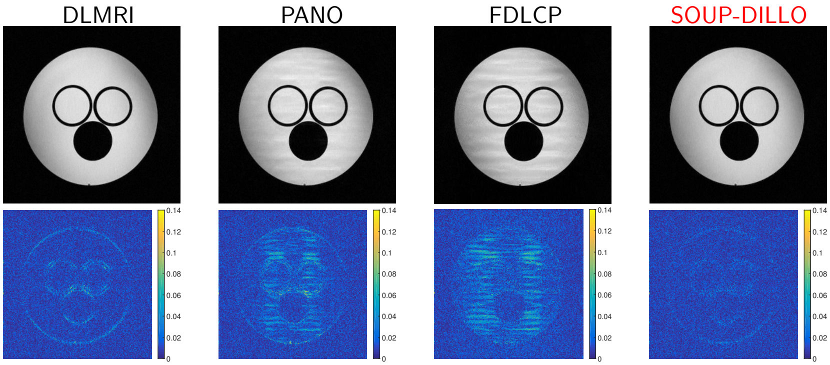
2D CS MRI results with blind DL III



(a) (b) (c) (d) (e) (f) (g)

PSNR:

Im.	Samp.	Acc.	0-fill	Sparse MRI	PANO	DLMRI	SOUP-DILLI	SOUP-DILLO
a	Cart.	7x	27.9	28.6	**31.1**	**31.1**	30.8	**31.1**
b	Cart.	2.5x	27.7	31.6	41.3	40.2	38.5	**42.3**
c	Cart.	2.5x	24.9	29.9	34.8	36.7	36.6	**37.3**
c	Cart.	4x	25.9	28.8	**32.3**	32.1	32.2	**32.3**
d	Cart.	2.5x	29.5	32.1	36.9	38.1	36.7	**38.4**
e	Cart.	2.5x	28.1	31.7	40.0	38.0	37.9	**41.5**
f	2D rand.	5x	26.3	27.4	30.4	30.5	30.3	**30.6**
g	Cart.	2.5x	32.8	39.1	41.6	41.7	42.2	**43.2**
Ref.				[16]	[17]	[13]	[15]	[15]



[13] [17] [18] [15]

Summary: 2D static MR reconstruction from under-sampled data with adaptive dictionary learning and convergent algorithm, faster than K-SVD approach of DLMRI.

Outline

Denoising based "regularization"

Patch-based and convolutional sparsity models lead to a denoising step for the current image estimate $\boldsymbol{x}_t$ at iteration t

Many alternative denoising methods:

- nonlocal means (NLM) [19]
- block-matching 3D (BM3D) [20]
- ...

To adapt most such denoising methods for image reconstruction:

- plug-and-play ADMM [21, 22]
- Regularization by denoising (RED) [23, 24, 25]

Plug-and-play ADMM

- Use auxiliary variable (variable splitting) to simplify optimization:

$$\begin{aligned}
\hat{\boldsymbol{x}} &= \arg\min_{\boldsymbol{x}} \frac{1}{2}\|\boldsymbol{A}\boldsymbol{x}-\boldsymbol{y}\|_2^2 + R(\boldsymbol{x}) && (\text{challenging \& unconstrained}) \\
&= \arg\min_{\boldsymbol{x}} \min_{\boldsymbol{z}} \frac{1}{2}\|\boldsymbol{A}\boldsymbol{x}-\boldsymbol{y}\|_2^2 + R(\boldsymbol{z}) \quad \text{s.t.} \quad \boldsymbol{x}=\boldsymbol{z} && (\text{constrained}) \\
&\approx \arg\min_{\boldsymbol{x}} \min_{\boldsymbol{z}} \frac{1}{2}\|\boldsymbol{A}\boldsymbol{x}-\boldsymbol{y}\|_2^2 + R(\boldsymbol{z}) + \frac{\mu}{2}\|\boldsymbol{x}-\boldsymbol{z}\|_2^2 && (\text{quadratic penalty})
\end{aligned}$$

-

- Use auxiliary variable (variable splitting) to simplify optimization:

$$
\begin{aligned}
\hat{\boldsymbol{x}} &= \arg\min_{\boldsymbol{x}} \frac{1}{2} \|\boldsymbol{A}\boldsymbol{x} - \boldsymbol{y}\|_2^2 + R(\boldsymbol{x}) && \text{(challenging \& unconstrained)} \\
&= \arg\min_{\boldsymbol{x}} \min_{\boldsymbol{z}} \frac{1}{2} \|\boldsymbol{A}\boldsymbol{x} - \boldsymbol{y}\|_2^2 + R(\boldsymbol{z}) \quad \text{s.t.} \quad \boldsymbol{x} = \boldsymbol{z} && \text{(constrained)} \\
&\approx \arg\min_{\boldsymbol{x}} \min_{\boldsymbol{z}} \frac{1}{2} \|\boldsymbol{A}\boldsymbol{x} - \boldsymbol{y}\|_2^2 + R(\boldsymbol{z}) + \frac{\mu}{2} \|\boldsymbol{x} - \boldsymbol{z}\|_2^2 && \text{(quadratic penalty)}
\end{aligned}
$$

- Simplified version of alternating direction method of multipliers (ADMM):

$$
\begin{aligned}
\boldsymbol{z}_k &= \arg\min_{\boldsymbol{z}} R(\boldsymbol{z}) + \frac{\mu}{2} \|\boldsymbol{x}_k - \boldsymbol{z}\|_2^2 && \text{proximal operation (denoising)} \\
\boldsymbol{x}_{k+1} &= \arg\min_{\boldsymbol{x}} \frac{1}{2} \|\boldsymbol{A}\boldsymbol{x} - \boldsymbol{y}\|_2^2 + \frac{\mu}{2} \|\boldsymbol{x} - \boldsymbol{z}_k\|_2^2 && \text{regularized data consistency (CG)}
\end{aligned}
$$

-

Plug-and-play ADMM

- Use auxiliary variable (variable splitting) to simplify optimization:

$$\hat{\boldsymbol{x}} = \arg\min_{\boldsymbol{x}} \frac{1}{2} \|\boldsymbol{A}\boldsymbol{x} - \boldsymbol{y}\|_2^2 + R(\boldsymbol{x}) \qquad \text{(challenging \& unconstrained)}$$

$$= \arg\min_{\boldsymbol{x}} \min_{\boldsymbol{z}} \frac{1}{2} \|\boldsymbol{A}\boldsymbol{x} - \boldsymbol{y}\|_2^2 + R(\boldsymbol{z}) \quad \text{s.t.} \quad \boldsymbol{x} = \boldsymbol{z} \qquad \text{(constrained)}$$

$$\approx \arg\min_{\boldsymbol{x}} \min_{\boldsymbol{z}} \frac{1}{2} \|\boldsymbol{A}\boldsymbol{x} - \boldsymbol{y}\|_2^2 + R(\boldsymbol{z}) + \frac{\mu}{2} \|\boldsymbol{x} - \boldsymbol{z}\|_2^2 \qquad \text{(quadratic penalty)}$$

- Simplified version of alternating direction method of multipliers (ADMM):

$$\boldsymbol{z}_k = \arg\min_{\boldsymbol{z}} R(\boldsymbol{z}) + \frac{\mu}{2} \|\boldsymbol{x}_k - \boldsymbol{z}\|_2^2 \qquad \text{proximal operation (denoising)}$$

$$\boldsymbol{x}_{k+1} = \arg\min_{\boldsymbol{x}} \frac{1}{2} \|\boldsymbol{A}\boldsymbol{x} - \boldsymbol{y}\|_2^2 + \frac{\mu}{2} \|\boldsymbol{x} - \boldsymbol{z}_k\|_2^2 \qquad \text{regularized data consistency (CG)}$$

- Replace denoising step with any denoiser, such as deep network

Outline

Machine-learning approaches to image reconstruction

- Learn models (sparsifying transform or dictionary) for image patches from training data
  - interpretable (?) optimization formulations
  - local prior information only (patch size)
  - perhaps slower computation due to optimization iterations
- Train neural network (aka deep learning)
  - less interpretable
  - possibly more global prior information
  - slow training, but perhaps faster computation after trained

Overview:

- image-domain learning [26, 27, 28]...
- k-space or data-domain learning
  e.g., [29], [30], [31]
- transform learning (direct from k-space to image)
  e.g., AUTOMAP [32], [33, 34, 35]
- hybrid-domain learning (unrolled loop, *e.g.*, variational network)
  alternate between denoising/dealiasing and reconstruction from k-space
  e.g., [36, 37, 38, 39, 40, 30] ...

DL for IR: image-domain learning

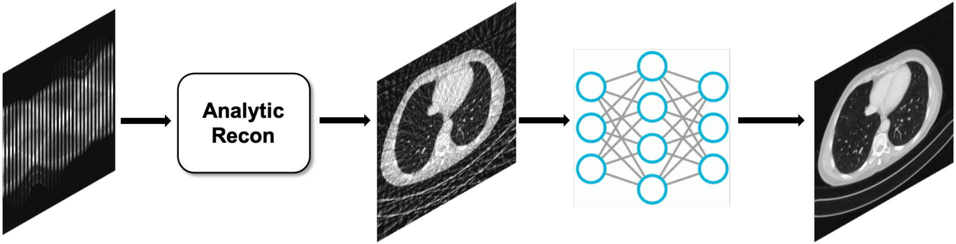


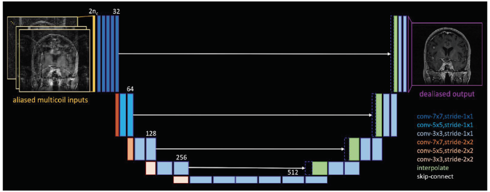
Figure courtesy of Jong Chul Ye, KAIST University.

+ simple and fast
– aliasing is spatially widespread, requires deep network

Investigating Robustness to Unseen Pathologies in Model-Free Deep Multicoil Reconstruction

Gopal Nataraj[1] and Ricardo Otazo[1,2]
[1]Dept. of Medical Physics, Memorial Sloan Kettering Cancer Center
[2]Dept. of Radiology, Memorial Sloan Kettering Cancer Center

Introduction
Speed is often claimed as a key advantage of deep learning (DL) for undersampled parallel MRI reconstruction [1]. However, the only DL approach that to our knowledge has studied generalizability to pathologies unseen in training [2] requires repeated application of the MR acquisition model and its adjoint, just as in iterative methods. In contrast, model-free DL reconstruction has the potential to be much faster. Prior model-free DL work [3] proposes to learn a mapping directly from k-space, but with



[41] ISMRM 2020 Workshop on Data Sampling & Image Reconstruction

Dangers of image-domain learning: Result

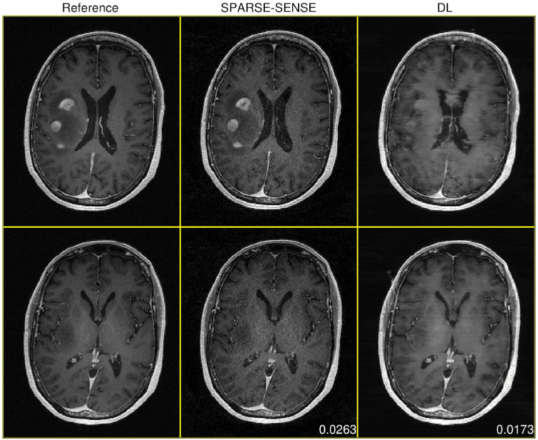


Figure 3: Reconstructions in a case of anaplastic astrocytoma, a rare malignant brain tumor. SPARSE-SENSE and DL reconstructions are from the same 4x-accelerated retrospectively undersampled acquisition. DL achieves lower whole-volume MAE than SPARSE-SENSE, but fails to properly reconstruct regions near the tumor.

- Use NN output as a "prior" for iterative reconstruction [26, 42]:

$$\hat{\boldsymbol{x}}_\beta = \arg\min_{\boldsymbol{x}} \|\boldsymbol{A}\boldsymbol{x} - \boldsymbol{y}\|_2^2 + \beta \|\boldsymbol{x} - \boldsymbol{x}_{\mathrm{NN}}\|_2^2 = (\boldsymbol{A}'\boldsymbol{A} + \beta \boldsymbol{I})^{-1}(\boldsymbol{A}'\boldsymbol{y} + \beta \boldsymbol{x}_{\mathrm{NN}})$$

  - For single-coil Cartesian case:
    - no iterations are needed (solve with FFTs)
    - $\lim_{\beta \to 0} \hat{\boldsymbol{x}}_\beta$ replaces missing k-space data with FFT of $\boldsymbol{x}_{\mathrm{NN}}$
  - Iterations needed for parallel MRI and/or non-Cartesian sampling (PCG)

- Learn residual (aliasing artifacts), then subtract [43, 44]

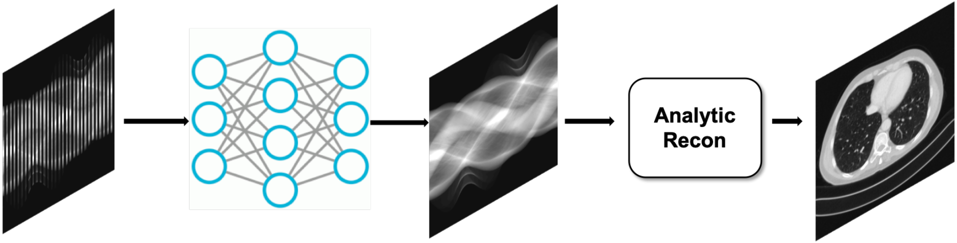


Figure courtesy of Jong Chul Ye, KAIST University.

+ simple and fast ("nonlinear GRAPPA")
+ "database-free" : learn from auto-calibration data [29], [30], [31]
− perhaps harder to represent local image features?

DL for IR: transform learning

Figure courtesy of Jong Chul Ye, KAIST University.

- \+ in principle, purely data driven; potential to avoid model mismatch
- − high memory requirement for fully connected layers [32]

DL for IR: hybrid domain learning (unrolled loop)

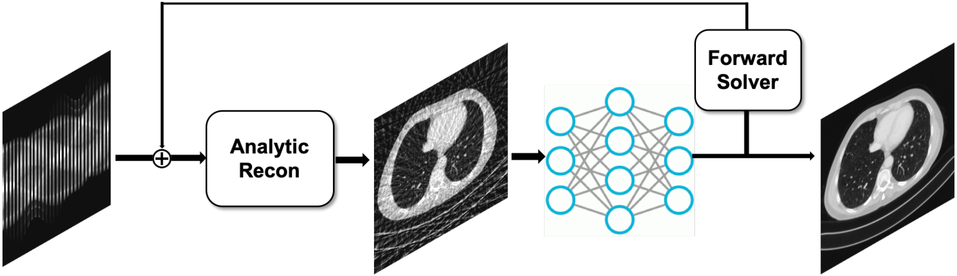
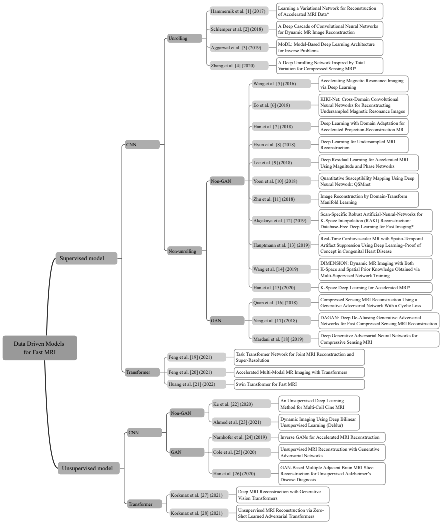


Figure courtesy of Jong Chul Ye, KAIST University.

+ physics-based use of k-space data & image-domain priors, *e.g.*, [36, 37, 38, 39, 40, 30, 45, 46] ...
+ interpretable connections to optimization approaches
+ best results in MRI recon challenges [47, 5, 48]
− more computation to due to "iterations" (hyper-layers) and repeated $\boldsymbol{A}\boldsymbol{x}$, $\boldsymbol{A}'\boldsymbol{r}$

DL for MRI: a taxonomy



Huang et al.., arXiv 2204.01706, Apr. 2022 [49]

Unrolled / unfolded loops

- learned ISTA (LISTA) [50]
  aka proximal gradient method / forward-backward splitting [51]
- half-quadratic [52]
- reaction-diffusion (GD) [53, 54]
- gradient descent / Landweber [55, 37]
- ADMM [36, 56]
- iterative hard thresholding (IHT) [57]
- approximate message passing (AMP) [58]
- accelerated gradient method [59]
- primal dual [60]
- primal dual with line search [61]
- alternating minimization [62]
- block coordinate descent (BCD-Net) [63, 64, 65, 66]
- block proximal gradient with momentum (BPGM: Momentum-Net) [67, 68, 46]
- And more [69, 45, 70, 71, 72, 73]

Surveys: [74, 75]

Zaccharie Ramzi, Philippe Ciuciu, Jean-Luc Starck Appl. Sci. 2020 [76]

Different models based on:

- optimization algorithm to unroll
- choice of f_θ
- N

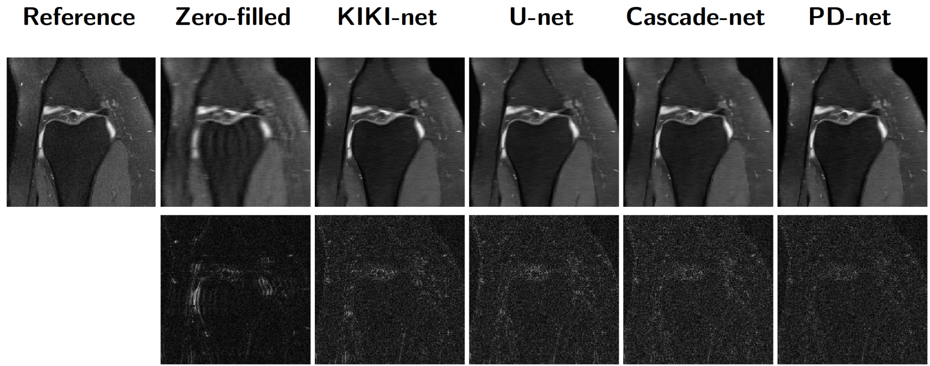
Table: **Quantitative results for the fastMRI dataset.** The PSNR is computed over the 200 validation volumes.

Network	**Zero-filled**	**KIKI-net**	**U-net**	**Cascade net**	**PD-net**[8]
PSNR	29.61	31.38	31.78	31.97	**32.15**

[8] **Adler2018**

Adler & Öktem, IEEE T-MI, 2018 [60] Learned primal-dual reconstruction

Learned PD vs alternatives



Figures courtesy Zaccharie Ramzi & Philippe Ciuciu.
https://github.com/zaccharieramzi/fastmri-reproducible-benchmark

Extension of PD from NeuroSpin

SUMMARY OF QUALITY RANKS AND LIKERT SCORES

Team	Rank	Artifacts	Sharpness	CNR
4X Track				
AIRS	**1.36 ± 0.64**	**1.53 ± 0.70**	**1.53 ± 0.51**	**1.53 ± 0.51**
Nspin	1.94 ± 0.86	1.81 ± 1.01	1.72 ± 0.66	1.75 ± 0.84
ATB	2.22 ± 0.87	1.75 ± 0.97	1.97 ± 0.65	1.86 ± 0.80
8X Track				
AIRS	**1.28 ± 0.64**	**1.67 ± 0.68**	**1.89 ± 0.75**	**1.94 ± 0.75**
Nspin	2.25 ± 0.77	1.86 ± 0.83	2.72 ± 0.81	2.28 ± 0.81
ATB	2.28 ± 0.70	1.92 ± 0.94	2.56 ± 0.77	2.42 ± 0.84

- XPDNet Ramzi et al., arXiv [77] 2010.07290
- 2nd place in radiologist ratings in 2020 fastMRI challenge [48]
- Replaced plain CNN with multi-scale wavelet CNN; sensitivity map refiner network; 25 unrolled iterations
- AIRS and ATB were also unrolled networks

- ML-based nonlinear encoder, *e.g.*, autoencoder or generative adversarial network (GAN) [78, 79]: nonlinear generalizations of subspace models
- learn G: maps low-dimensional latent parameter $\boldsymbol{z}$ into high-dimensional image $\boldsymbol{x}$

- Synthesis form [80]:

$$\hat{\boldsymbol{x}} = G(\hat{\boldsymbol{z}}), \qquad \hat{\boldsymbol{z}} = \arg\min_{\boldsymbol{z}} \|\boldsymbol{A} G(\boldsymbol{z}) - \boldsymbol{y}\|_2^2$$

  Caveat: $\hat{\boldsymbol{x}} \in \text{Range}(G)$, non-convex minimization

-

- ML-based nonlinear encoder, *e.g.*, autoencoder or generative adversarial network (GAN) [78, 79]: nonlinear generalizations of subspace models
- learn G: maps low-dimensional latent parameter $\boldsymbol{z}$ into high-dimensional image $\boldsymbol{x}$

- Synthesis form [80]:

$$\hat{\boldsymbol{x}} = G(\hat{\boldsymbol{z}}), \qquad \hat{\boldsymbol{z}} = \arg\min_{\boldsymbol{z}} \|\boldsymbol{A} G(\boldsymbol{z}) - \boldsymbol{y}\|_2^2$$

Caveat: $\hat{\boldsymbol{x}} \in \text{Range}(G)$, non-convex minimization

- Regularizer form:

$$\hat{\boldsymbol{x}} = \arg\min_{\boldsymbol{x}} \|\boldsymbol{A}\boldsymbol{x} - \boldsymbol{y}\|_2^2 + \beta R_{\text{encoder}}(\boldsymbol{x})$$

$$R_{\text{encoder}}(\boldsymbol{x}) = \min_{\boldsymbol{z}} \|\boldsymbol{x} - G(\boldsymbol{z})\|_p^p$$

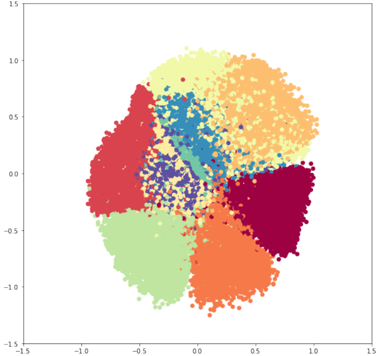
Caveat: expensive non-convex double minimization, but more robust to encoder

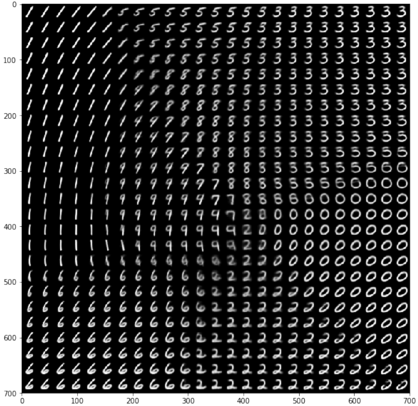
Nonlinear encoder illustration

From jupyter notebook for [81] (13 layer CNN with ≈ 300K learned parameters) at

https://github.com/skolouri/swae/blob/master/MNIST_SlicedWassersteinAutoEncoder_Circle.ipynb

$\boldsymbol{z} \in \mathbb{R}^2$ $\mapsto$ $\boldsymbol{x} = G(\boldsymbol{z}) \in \mathbb{R}^{28\times 28}$



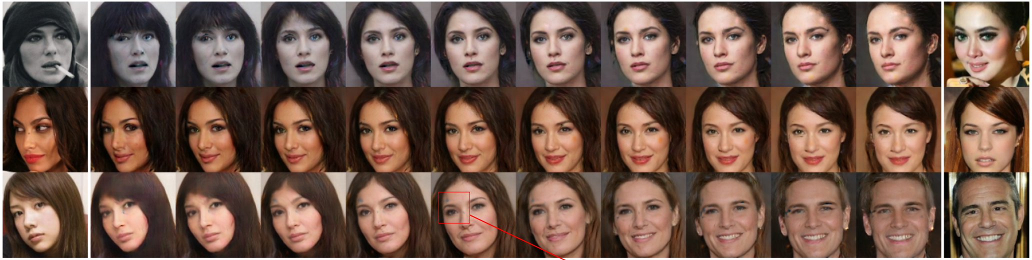


Caveat: Where is 4?

From Google's [82]:

Much more realistic than linear interpolation (averaging)
"setting a new milestone in visual quality" [82]

From Google's [82]:



Caveat: non-physical output

Deep-learned prior (MODL)

Model based image reconstruction using deep learned priors (MODL) [70, 45]

$$\hat{\boldsymbol{x}} = \operatorname*{arg\,min}_{\boldsymbol{x}} \frac{1}{2} \|\boldsymbol{A}\boldsymbol{x} - \boldsymbol{y}\|_2^2 + \|\mathrm{CNN}(\boldsymbol{x})\|_2^2$$

- $\mathrm{CNN}(\boldsymbol{x}) = \boldsymbol{x} - \mathrm{denoise}(\boldsymbol{x})$ predicts noise and aliasing patterns (*cf.* ResNet principle [43])
- Demonstrated robustness to changes in acceleration factors

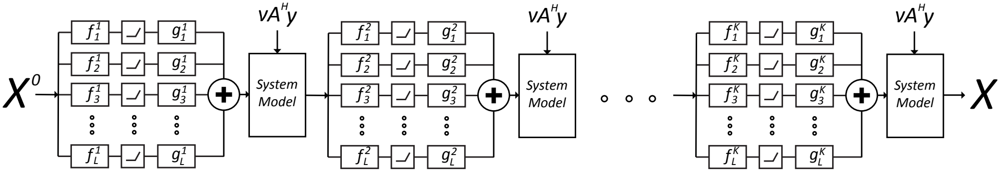
Outline

Caveats to NN methods

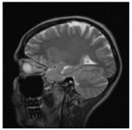
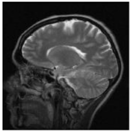
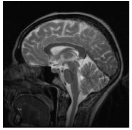
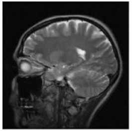
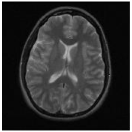
- Training data size (but self supervision [83])
- Local minimizers of training loss functions
- Sensitivity to adversarial examples (for classification problems)
- Enormous design space (architectures, parameters)
- Training loss functions, evaluation metrics vs clinical tasks
- Generalizability
  - noise level
  - coil sensitivity
  - k-space sampling
- Stability [84]
- Memory (especially 3D and dynamic)
- ...

Caveat: careful comparisons needed I

Unrolled loop method with 20 layers trained with $1.3 \cdot 10^6$ MR image 8×8 patches [62]



Tested with 5 different images:

Results:

UF	Image	Zero-filled	Sparse MRI	UTMRI	Proposed
3.3×	1	25.6	26.7	28.3	28.2
	2	25.2	26.6	27.9	27.8
	3	26.0	27.3	29.3	28.9
	4	25.4	26.7	28.2	28.1
	5	27.2	28.9	30.6	30.3
Avg. PSNR change	-	-	**1.36**	**2.98**	**2.78**
5×	1	24.7	25.9	27.6	27.5
	2	24.2	25.5	27.2	27.0
	3	24.9	26.3	28.5	28.0
	4	24.4	25.7	27.6	27.4
	5	26.2	27.9	29.8	29.5
Avg. PSNR change	-	-	**1.38**	**3.26**	**3.0**
Approx recon time	-	-	**100s**	**240s**	**50s**

Sparse MRI [85] total variation and wavelets
UTMRI [86] (union of learned sparsifying transforms): adaptive, not "deep"

Challenges: 3D and beyond

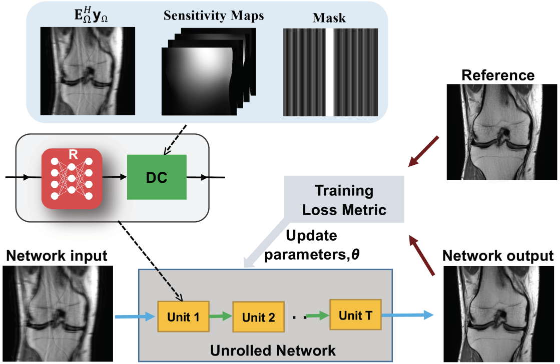
- Deep networks can require lots of memory to train
- Mitigation strategies:
  - gradient checkpointing [87]
  - invertible / reversible networks [88, 89, 47, 90, 91, 92]
  - 2.5D models for 3D images [93, 94]
  - implicit models (neural fields, neural ODEs...) [95, 96, 97, 98, 99, 100]
  - deep equilibrium models [101, 102, 103, 104, 105]
  - monotone operator learning [106]
  - ...

Deep network learning strategies

Unrolled deep networks: physics-driven deep learning

- Supervised learning: learning from large labeled data
- Self-supervised learning: learning from large unlabeled data
- Zero-shot learning: learning from a single sample

(The terminology is non-intuitive.)



Training process:

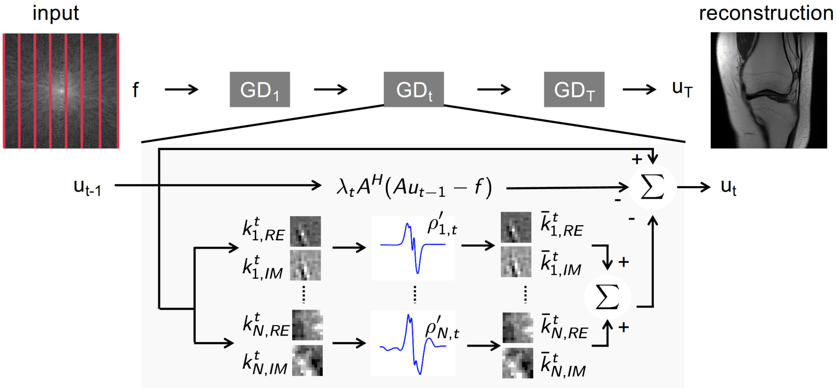
$$\arg\min_{\boldsymbol{\theta}} \frac{1}{N} \sum_{n=1}^{N} \mathcal{L}\left(\boldsymbol{x}_n^{\mathrm{ref}}, f_T(\boldsymbol{y}_n, \boldsymbol{A}_n; \boldsymbol{\theta})\right)$$

$\boldsymbol{\theta}$: network parameters
$f_T(\cdot)$: network output iteration T
N : number of training samples
$\mathcal{L}$: loss function
$\boldsymbol{x}_n^{\mathrm{ref}}$: nth "ground truth" image

$\hat{\boldsymbol{\theta}}$ is specific to T

Figures for next many slides courtesy of Burhan Yaman.

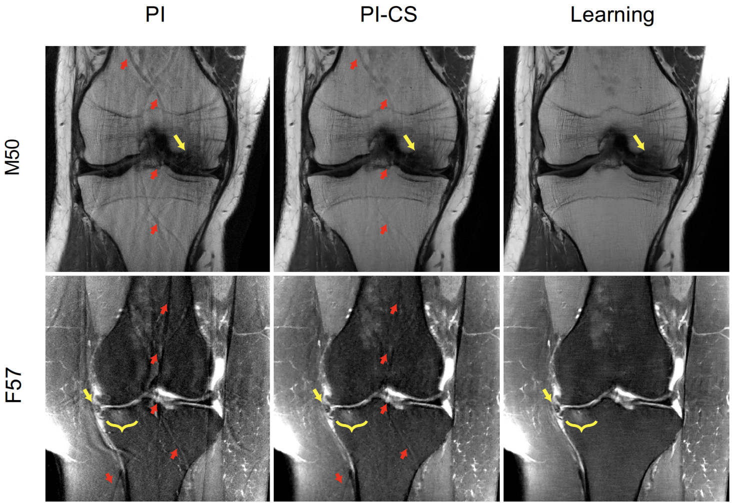
- Example[1]: Variational network with Fields of Experts model $\mathcal{R}(\mathbf{u}) = \sum_{i=1}^{N_k} \langle \Phi_i(\mathbf{K}_i \mathbf{u}), \mathbf{1} \rangle.$



$$u_t = u_{t-1} - \sum_i^N K_{i,t}^T \rho'_{i,t}(K_{i,t} u_{t-1}) - \lambda_t A^H (A u_{t-1} - f)$$

[1]Hammernik et al, MRM, 2018

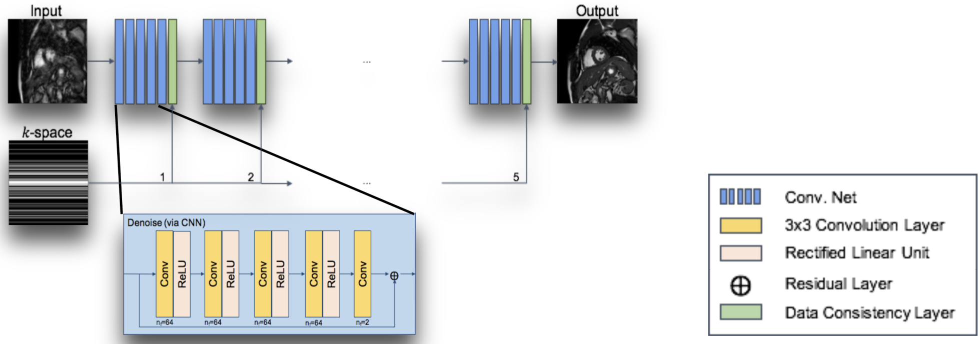
Slide courtesy of F. Knoll



Hammernick et al., MRM, 2018 [37]

Comparisons to dictionary learning and total generalized variation (TGV) in paper.

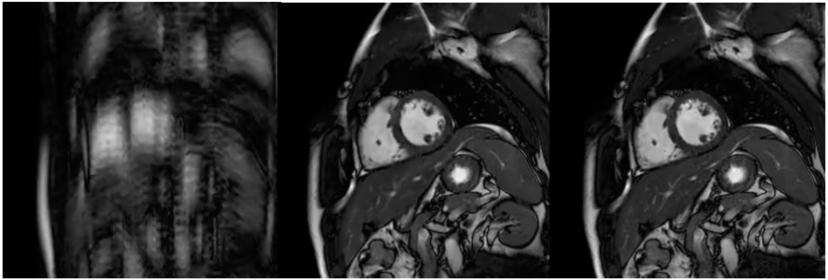
- Example[2]: Cascade of CNNs



[2]Schlemper et al, IEEE TMI, 2018

Slide courtesy of D. Rueckert

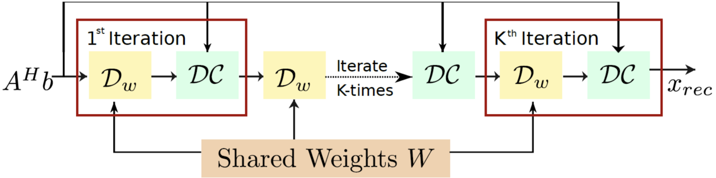
Cascade of CNNs: results

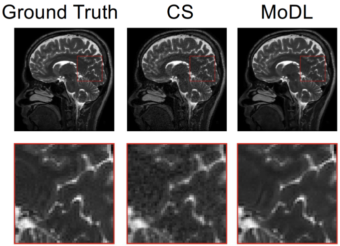


(a) 6x Undersampled **(b) CNN reconstruction** **(c) Ground Truth**

Schlemper et al., IEEE T-MI, 2018 [38]
Comparisons with dictionary learning (DLMRI) in paper.

- Example[3]: Model-Based Deep Learned Priors (MoDL)

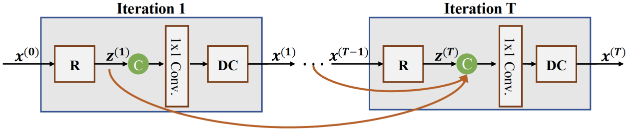


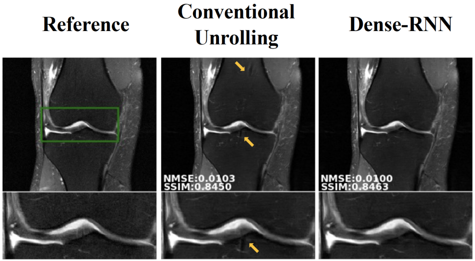


[3]Aggarwal et al, IEEE TMI, 2019

Slide courtesy of M. Jacobs

- Example[4]: Dense Recurrent Neural Network (~Nesterov unrolling)





[4]Hosseini et al, IEEE JSTSP, 2020

Challenges of end-to-end supervised training

- GPU memory for 3D and beyond
  -

Challenges of end-to-end supervised training

- GPU memory for 3D and beyond
  - plug-and-play methods [25]
  - deep equilibrium models [101, 102, 108, 103]
  - neural fields [100]
  - monotone operator learning [109, 110]
  - ...
- Generalizability / robustness to distribution shift
-

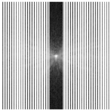
Challenges of end-to-end supervised training

- GPU memory for 3D and beyond
  - plug-and-play methods [25]
  - deep equilibrium models [101, 102, 108, 103]
  - neural fields [100]
  - monotone operator learning [109, 110]
  - ...
- Generalizability / robustness to distribution shift
- Availability of fully sampled training data?
  - High resolution MRI
  - Organ motion / dynamic MRI
  - Signal decay
  - ...

Slides from Burnam Yaman
(2023 ISMRM Sedona Workshop)
https://www.ismrm.org/workshops/2023/Data

Self-Supervised Learning

Acquired k-space locations: Ω

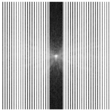


Yaman et al, MRM, 2020

Self-Supervised Learning

- Acquired k-space locations Ω, split into two sets

Acquired
k-space locations: Ω

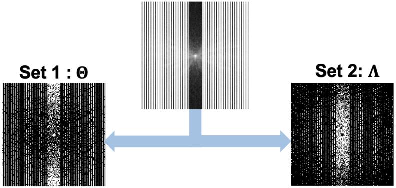


Yaman et al, MRM, 2020

Self-Supervised Learning

- Acquired k-space locations Ω, split into two sets

$$\Omega = \Theta \cup \Lambda$$



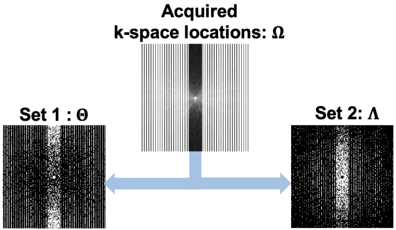
Yaman et al, MRM, 2020

Self-Supervised Learning

- Acquired k-space locations Ω, split into two sets

$$\Omega = \Theta \cup \Lambda$$

Data consistency
in unrolled network



Yaman et al, MRM, 2020

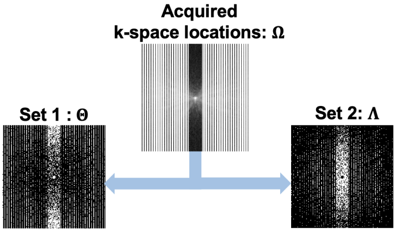
Self-Supervised Learning

- Acquired k-space locations Ω, split into two sets

$$\Omega = \Theta \cup \Lambda$$

Data consistency in unrolled network

Define network loss in k-space



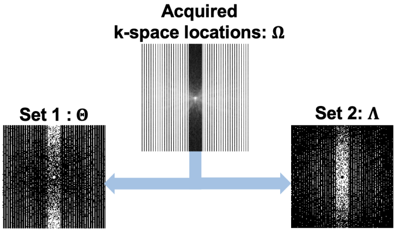
Yaman et al, MRM, 2020

Self-Supervised Learning

- Acquired k-space locations Ω, split into two sets

$$\Omega = \Theta \cup \Lambda$$
$$\Theta = \Omega \backslash \Lambda$$

DC units in unrolled
network only sees data at Θ

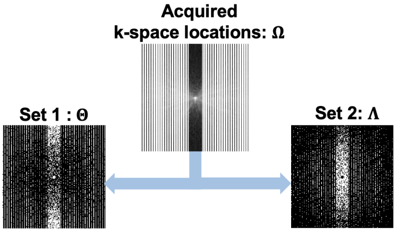


Yaman et al, MRM, 2020

Self-Supervised Learning

- Acquired k-space locations Ω, split into two sets

$$\Omega = \Theta \cup \Lambda$$
$$\Theta = \Omega \backslash \Lambda$$



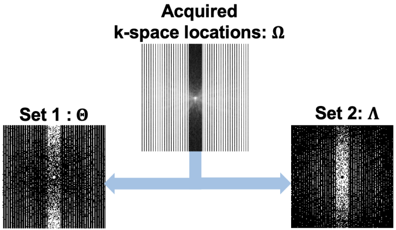
- Self-supervision via data undersampling (SSDU)

Yaman et al, MRM, 2020

Self-Supervised Learning

- Acquired k-space locations Ω, split into two sets

$$\Omega = \Theta \cup \Lambda$$
$$\Theta = \Omega \backslash \Lambda$$



- Self-supervision via data undersampling (SSDU)

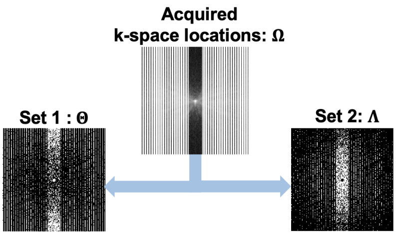
- End-to-end minimization

Yaman et al, MRM, 2020

Self-Supervised Learning

- Acquired k-space locations Ω, split into two sets

$$\Omega = \Theta \cup \Lambda$$
$$\Theta = \Omega \backslash \Lambda$$



- Self-supervision via data undersampling (SSDU)

- End-to-end minimization

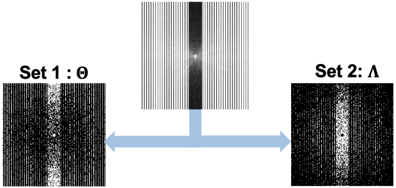
$$\min_{\boldsymbol{\theta}} \frac{1}{N} \sum_{i=1}^{N} \mathcal{L}\left(\mathbf{y}_{\Lambda}^{i}, \mathbf{E}_{\Lambda}^{i}\left(f\left(\mathbf{y}_{\Theta}^{i}, \mathbf{E}_{\Theta}^{i}; \boldsymbol{\theta}\right)\right)\right)$$

Yaman et al, MRM, 2020

Self-Supervised Learning

- Acquired k-space locations Ω, split into two sets

$$\Omega = \Theta \cup \Lambda$$
$$\Theta = \Omega \backslash \Lambda$$



- Self-supervision via data undersampling (SSDU)

- End-to-end minimization

$$\min_{\boldsymbol{\theta}} \frac{1}{N} \sum_{i=1}^{N} \mathcal{L}\left(\mathbf{y}_{\Lambda}^{i}, \mathbf{E}_{\Lambda}^{i}\left(f\left(\mathrm{y}_{\Theta}^{i}, \mathbf{E}_{\Theta}^{i}; \boldsymbol{\theta}\right)\right)\right)$$

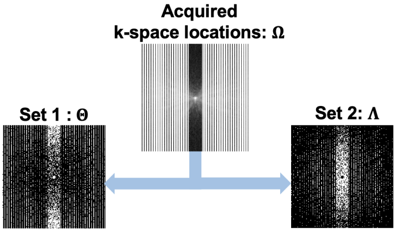
Loss is measured on k-space at unseen locations in training, Λ

Yaman et al, MRM, 2020

Self-Supervised Learning

- Acquired k-space locations Ω, split into two sets

$$\Omega = \Theta \cup \Lambda$$
$$\Theta = \Omega \backslash \Lambda$$



- Self-supervision via data undersampling (SSDU)

- End-to-end minimization

$$\min_{\boldsymbol{\theta}} \frac{1}{N} \sum_{i=1}^{N} \mathcal{L}\left(\mathbf{y}_{\Lambda}^{i}, \mathbf{E}_{\Lambda}^{i}\left(f\left(\mathbf{y}_{\Theta}^{i}, \mathbf{E}_{\Theta}^{i}; \boldsymbol{\theta}\right)\right)\right)$$

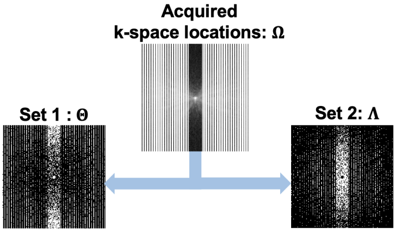
Network input

Yaman et al, MRM, 2020

Self-Supervised Learning

- Acquired k-space locations Ω, split into two sets

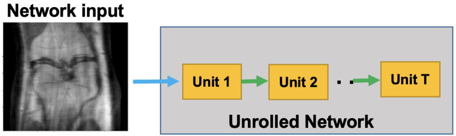
$$\Omega = \Theta \cup \Lambda$$
$$\Theta = \Omega \backslash \Lambda$$



- Self-supervision via data undersampling (SSDU)

- End-to-end minimization

$$\min_{\boldsymbol{\theta}} \frac{1}{N} \sum_{i=1}^{N} \mathcal{L}\left(\mathbf{y}_{\Lambda}^{i}, \mathbf{E}_{\Lambda}^{i}\left(f\left(\mathrm{y}_{\Theta}^{i}, \mathbf{E}_{\Theta}^{i}; \boldsymbol{\theta}\right)\right)\right)$$

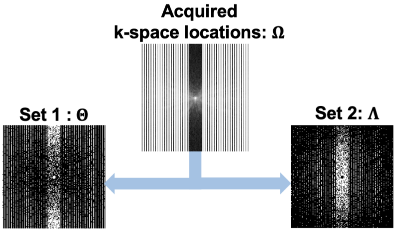


Yaman et al, MRM, 2020

Self-Supervised Learning

- Acquired k-space locations Ω, split into two sets

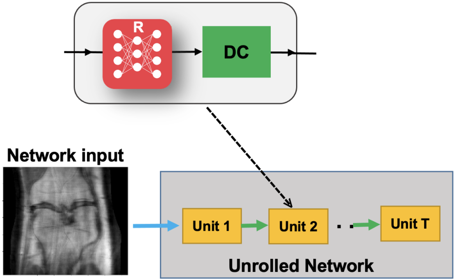
$$\Omega = \Theta \cup \Lambda$$
$$\Theta = \Omega \backslash \Lambda$$



- Self-supervision via data undersampling (SSDU)

- End-to-end minimization

$$\min_{\boldsymbol{\theta}} \frac{1}{N} \sum_{i=1}^{N} \mathcal{L}\left(\mathbf{y}_{\Lambda}^{i}, \mathbf{E}_{\Lambda}^{i}\left(f\left(\mathbf{y}_{\Theta}^{i}, \mathbf{E}_{\Theta}^{i}; \boldsymbol{\theta}\right)\right)\right)$$



Yaman et al, MRM, 2020

Self-Supervised Learning

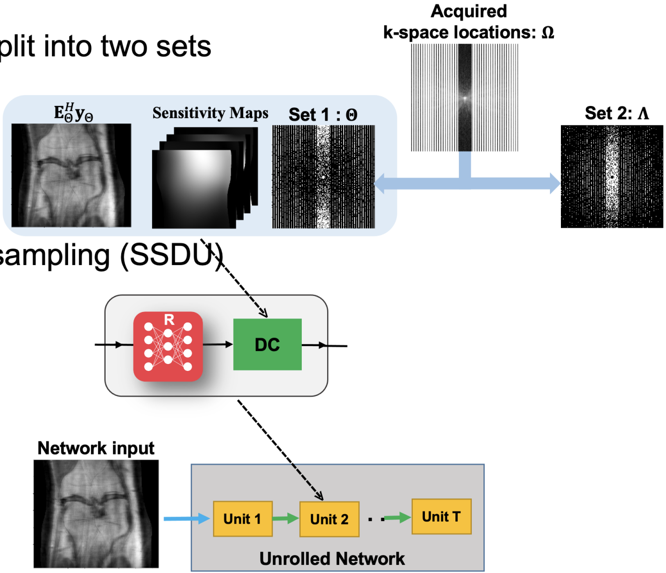
- Acquired k-space locations Ω, split into two sets

$$\Omega = \Theta \cup \Lambda$$
$$\Theta = \Omega \backslash \Lambda$$

- Self-supervision via data undersampling (SSDU)

- End-to-end minimization

$$\min_{\boldsymbol{\theta}} \frac{1}{N}\sum_{i=1}^{N}\mathcal{L}\left(\mathbf{y}^i_{\Lambda}, \mathbf{E}^i_{\Lambda}\left(f\left(\mathbf{y}^i_{\Theta}, \mathbf{E}^i_{\Theta}; \boldsymbol{\theta}\right)\right)\right)$$



Yaman et al, MRM, 2020

Self-Supervised Learning

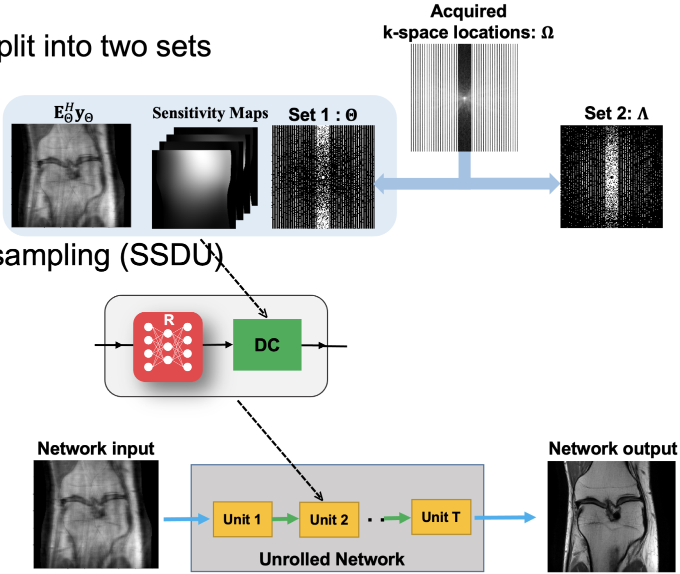
- Acquired k-space locations Ω, split into two sets

$$\Omega = \Theta \cup \Lambda$$
$$\Theta = \Omega \backslash \Lambda$$

- Self-supervision via data undersampling (SSDU)

- End-to-end minimization

$$\min_{\boldsymbol{\theta}} \frac{1}{N} \sum_{i=1}^{N} \mathcal{L}\left(\mathbf{y}_{\Lambda}^{i}, \mathbf{E}_{\Lambda}^{i}\left(f\left(\mathbf{y}_{\Theta}^{i}, \mathbf{E}_{\Theta}^{i}; \boldsymbol{\theta}\right)\right)\right)$$

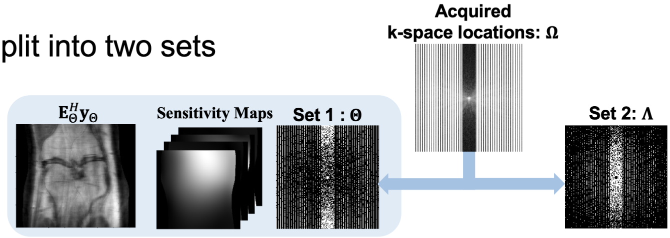


Yaman et al, MRM, 2020

Self-Supervised Learning

- Acquired k-space locations Ω, split into two sets

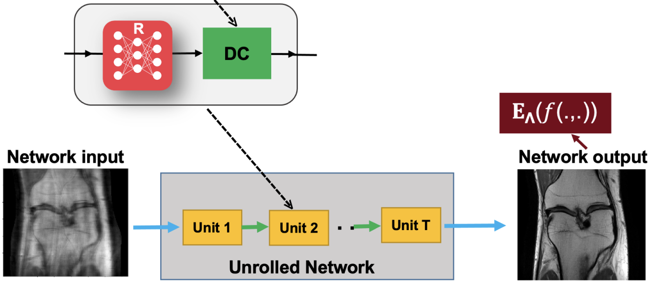
$$\Omega = \Theta \cup \Lambda$$
$$\Theta = \Omega \backslash \Lambda$$



- Self-supervision via data undersampling (SSDU)

- End-to-end minimization

$$\min_{\boldsymbol{\theta}} \frac{1}{N} \sum_{i=1}^{N} \mathcal{L}\left(\mathbf{y}_\Lambda^i, \mathbf{E}_\Lambda^i\left(f\left(\mathbf{y}_\Theta^i, \mathbf{E}_\Theta^i; \boldsymbol{\theta}\right)\right)\right)$$



Yaman et al, MRM, 2020

Self-Supervised Learning

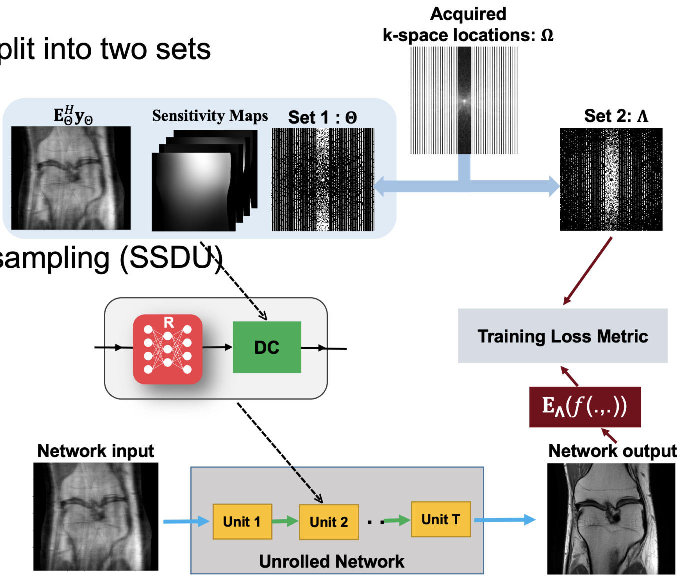
- Acquired k-space locations Ω, split into two sets

$$\Omega = \Theta \cup \Lambda$$
$$\Theta = \Omega \backslash \Lambda$$

- Self-supervision via data undersampling (SSDU)

- End-to-end minimization

$$\min_{\boldsymbol{\theta}} \frac{1}{N} \sum_{i=1}^{N} \mathcal{L}\left(\mathbf{y}_{\Lambda}^{i}, \mathbf{E}_{\Lambda}^{i}\left(f\left(\mathbf{y}_{\Theta}^{i}, \mathbf{E}_{\Theta}^{i}; \boldsymbol{\theta}\right)\right)\right)$$



Yaman et al, MRM, 2020

Self-Supervised Learning

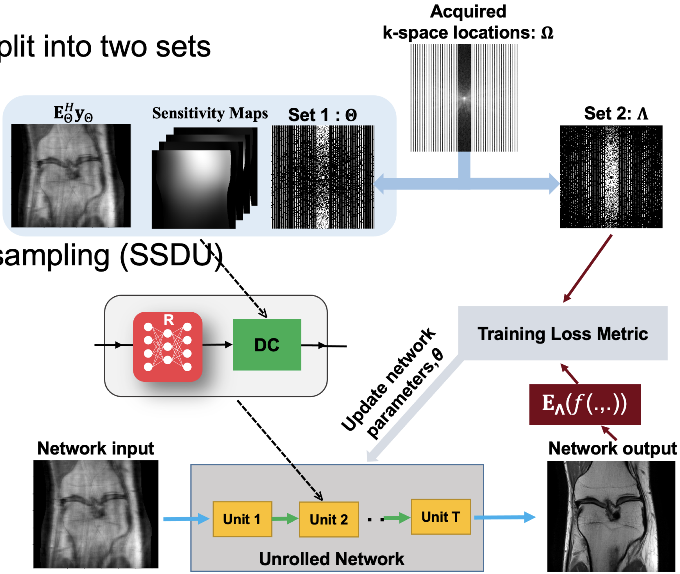
- Acquired k-space locations Ω, split into two sets

$$\Omega = \Theta \cup \Lambda$$
$$\Theta = \Omega \backslash \Lambda$$

- Self-supervision via data undersampling (SSDU)

- End-to-end minimization

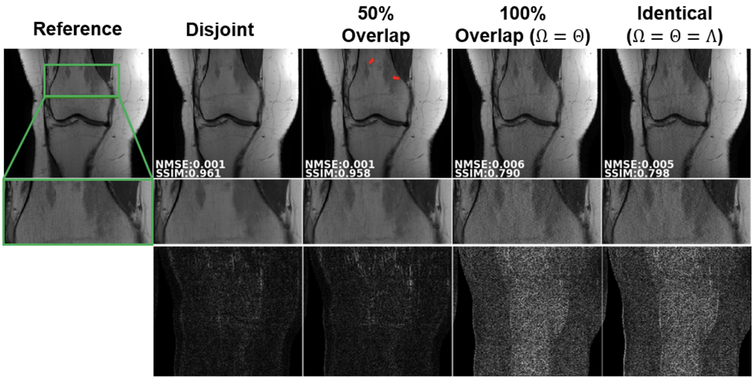
$$\min_{\boldsymbol{\theta}} \frac{1}{N} \sum_{i=1}^{N} \mathcal{L}\left(\mathbf{y}_{\Lambda}^{i}, \mathbf{E}_{\Lambda}^{i}\left(f\left(\mathbf{y}_{\Theta}^{i}, \mathbf{E}_{\Theta}^{i}; \boldsymbol{\theta}\right)\right)\right)$$



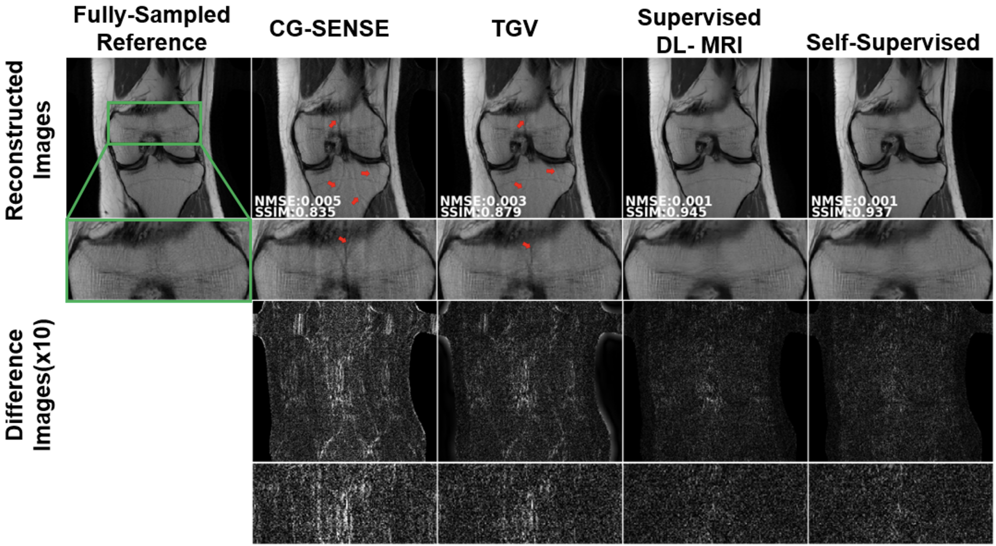
Yaman et al, MRM, 2020

Overlapped Sampling Points

- Overlap %= $|\Lambda \cap \Omega|/|\Lambda|$
  Amount of data in Λ that was also included in Θ
- Identical set suffers from noise amplification
- As overlap between two sets increase, performance degrades
- Disjoint sets outperform overlapping and identical sets

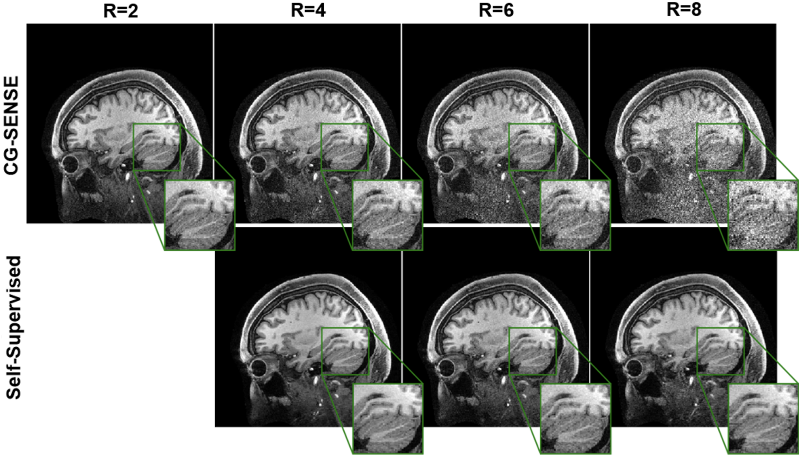


Physics-Driven DL Reconstruction



Physics-Driven DL Reconstruction

- Prospectively subsampled (R = 2)
- Supervised DL MRI not available (no ref data)
- Self-supervised successful reconstruction at high rates



Zero-shot self-supervised learning (ZS-SSL)

Database deep learning

- Lack of large datasets
  Motion, 3D/
- Trained model may not generalize well if the test data differs
  contrast / coils / sampling / anatomy / FOV / vendor / SNR . . .

ZS-SSL

-

Zero-shot self-supervised learning (ZS-SSL)

Database deep learning

- Lack of large datasets
  Motion, 3D/
- Trained model may not generalize well if the test data differs
  contrast / coils / sampling / anatomy / FOV / vendor / SNR ...

ZS-SSL

- No training data required
- train and test on *single* case
- agnostic to distribution
- compute expensive
-

Zero-shot self-supervised learning (ZS-SSL)

Database deep learning

- Lack of large datasets
  Motion, 3D/
- Trained model may not generalize well if the test data differs
  contrast / coils / sampling / anatomy / FOV / vendor / SNR ...

ZS-SSL

- No training data required
- train and test on *single* case
- agnostic to distribution
- compute expensive
- combine with pretrained models via transfer learning to reduce computation

Yaman et al., ICLR 2022 [111]

Deep image prior (DIP)

- An early zero-shot method Ulyanov et al., CVPR 2018 [112]
- Recall CS-GAN approach of Bora et al., ICML 2017 [80]:

$$\hat{\boldsymbol{x}} = G_{\hat{\boldsymbol{\theta}}}(\hat{\boldsymbol{z}}), \qquad \hat{\boldsymbol{z}} = \arg\min_{\boldsymbol{z}} \left\| \boldsymbol{A} G_{\hat{\boldsymbol{\theta}}}(\boldsymbol{z}) - \boldsymbol{y} \right\|_2^2$$

- DIP approach using a random latent parameter $\boldsymbol{z}_0$ [112]:

$$\hat{\boldsymbol{x}} = f_{\hat{\boldsymbol{\theta}}}(\boldsymbol{z}_0), \qquad \hat{\boldsymbol{\theta}} = \arg\min_{\boldsymbol{\theta}} \left\| \boldsymbol{A} f_{\boldsymbol{\theta}}(\boldsymbol{z}_0) - \boldsymbol{y} \right\|_2^2$$

  Akin to a very nonlinear form of blind dictionary learning (also expensive)
- Applied to dynami MRI Yoo et al., IEEE T-MI 2021 [113]
  (no comparison to blind dictionary learning)

Zero-Shot Learning

Database Deep Learning

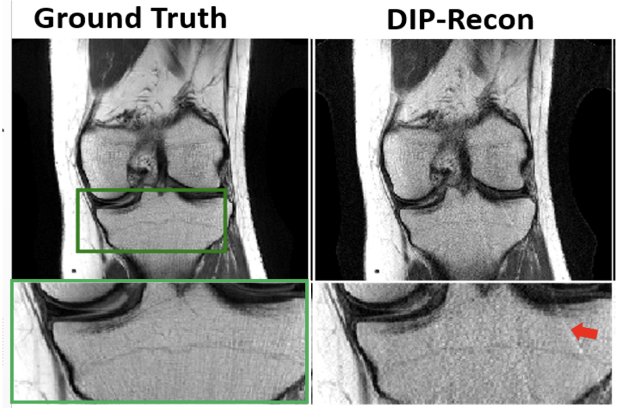
- Lack of large datasets due to physiological and physical constraints
  - Contrast uptake, breathing patterns…
  - Move towards processing larger 3D/4D volumes
- Trained model may not generalize well if the test data is out-of-distribution
  - Vendor/SNR/Mask&Rate/Anatomy…
- Retraining is computationally expensive

Zero-Shot Self-Supervised Learning (ZS-SSL)

- Does not require any external dataset
- Training & testing on a **single** image
- Agnostic to changes in distribution
- Potential high quality reconstruction for *every individual*
- Combined with pretrained models via transfer learning for computational efficiency

Deep Image Prior - MRI

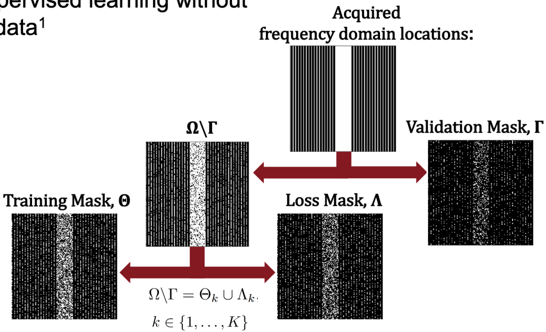
- DIP Reconstruction[1] $(\boldsymbol{\Omega} = \boldsymbol{\Theta} = \boldsymbol{\Lambda})$
- Performs training and testing on a single slice
- No stopping criterion → Overfitting



[1]Ulyanov et al, CVPR, 2017.

Zero-Shot Learning

- Zero-shot self-supervised learning without external training data[1]

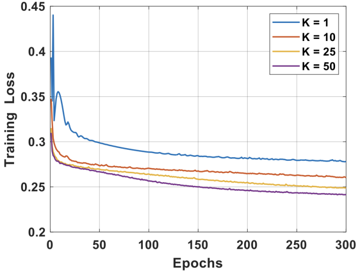


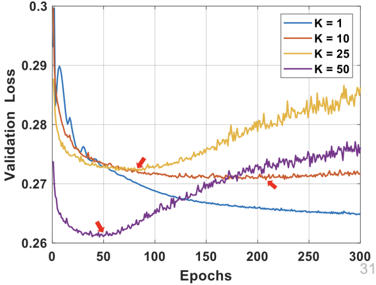
Training Loss

$$\min_{\theta} \frac{1}{K} \sum_{k=1}^{K} \mathcal{L}\Big(\mathbf{y}_{\Lambda_k},\ \mathbf{E}_{\Lambda_k}\big(f\big(\mathbf{y}_{\Theta_k}, \mathbf{E}_{\Theta_k}; \theta\big)\big)\Big)$$

Validation Loss

$$\mathcal{L}\Big(\mathbf{y}_{\Gamma},\ \mathbf{E}_{\Gamma}\big(f\big(\mathbf{y}_{\Omega\backslash\Gamma}, \mathbf{E}_{\Omega\backslash\Gamma}; \theta^{(l)}\big)\big)\Big)$$

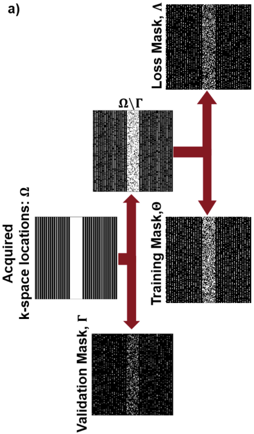




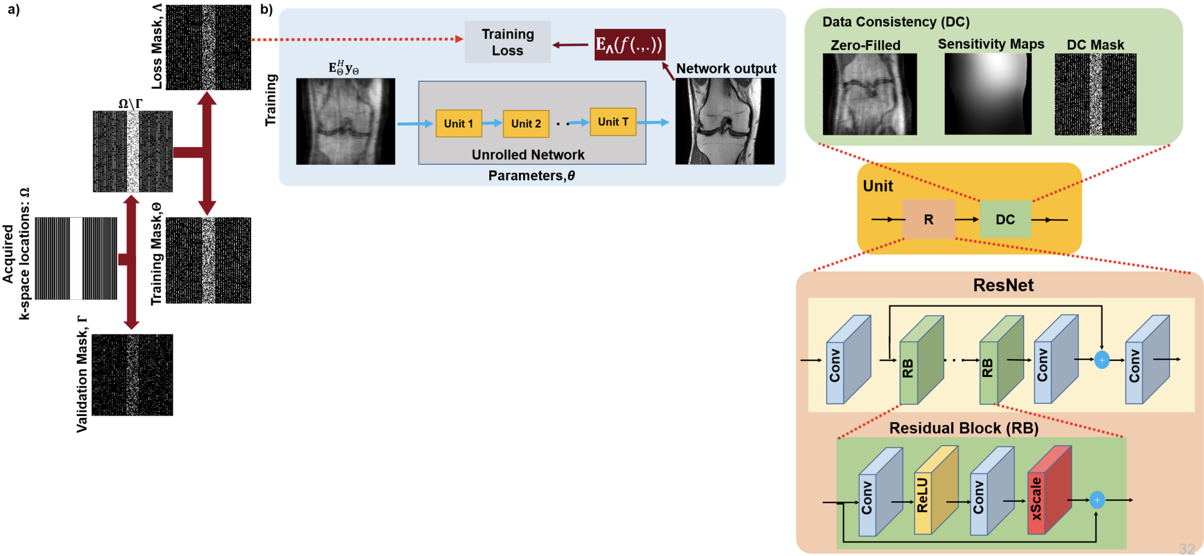
[1]Yaman et al, ICLR, 2022.

Zero-Shot Learning

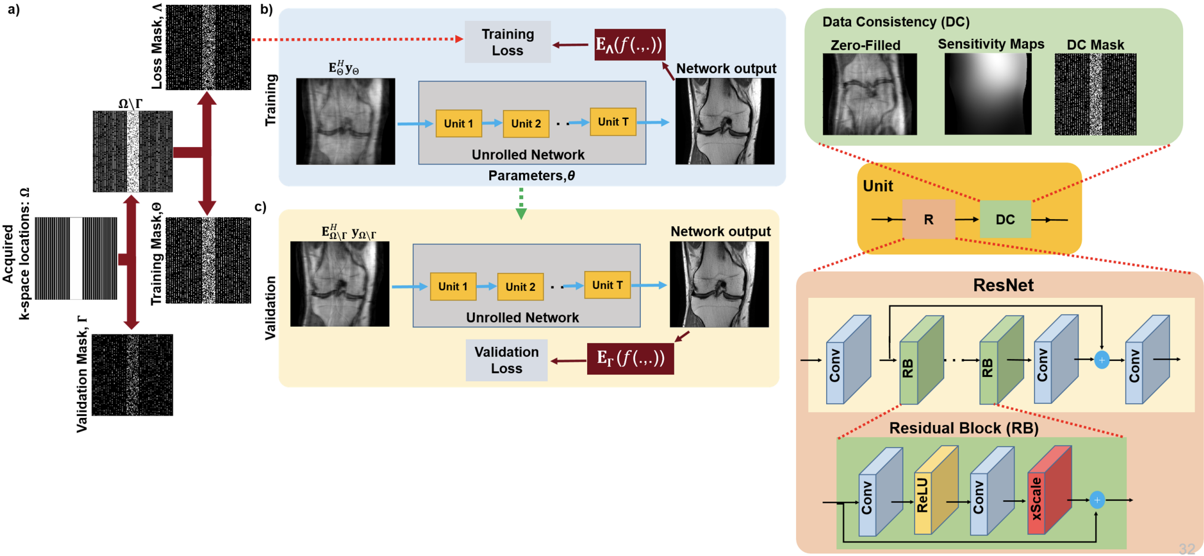
a)



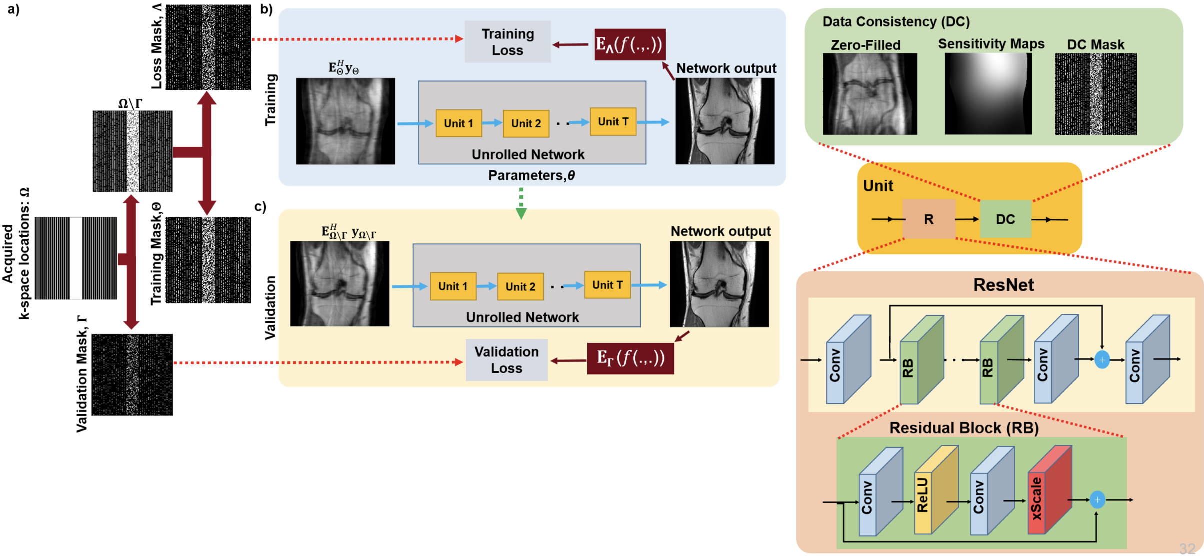
Zero-Shot Learning



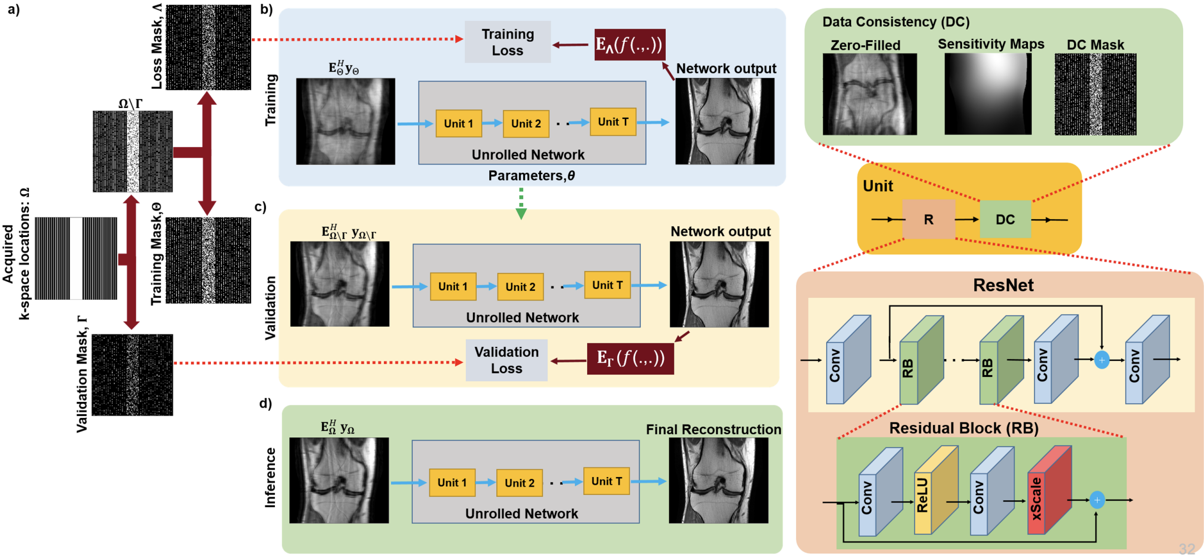
Zero-Shot Learning



Zero-Shot Learning

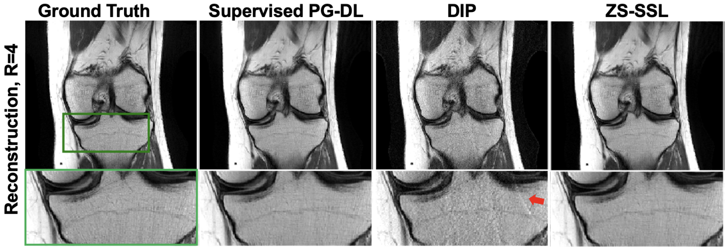


Zero-Shot Learning



Zero-Shot Learning

- DIP: Deep Image Prior
- ZS-SSL: Zero-Shot Self-Supervised Learning
- DIP and ZS-SSL performs training on a single slice
- Supervised PG-DL is a database deep learning approach

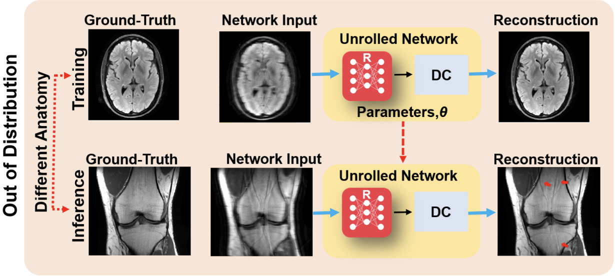


Trained on a database

Trained & tested on a single sample

Zero-shot + Transfer Learning

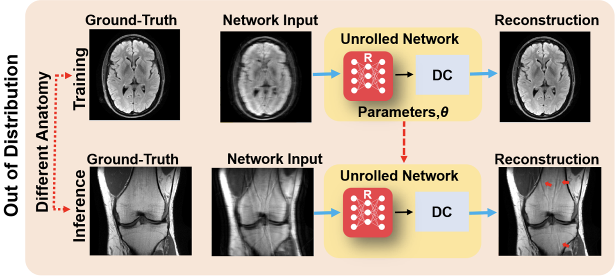
- Pretrained models performance degrades in presence of mismatch between training and test data



Zero-shot + Transfer Learning

- Pretrained models performance degrades in presence of mismatch between training and test data
- Combine pretrained models with ZS-SSL via transfer learning to improve:

a) accuracy, robustness and generalization

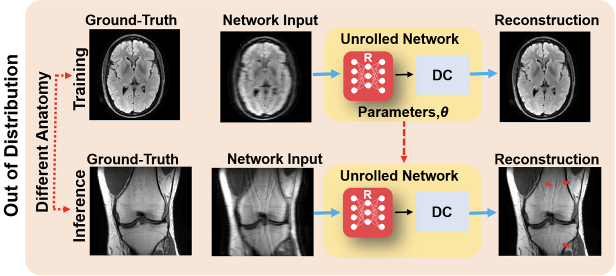


Zero-shot + Transfer Learning

- Pretrained models performance degrades in presence of mismatch between training and test data
- Combine pretrained models with ZS-SSL via transfer learning to improve:

a) accuracy, robustness and generalization

b) computational efficiency

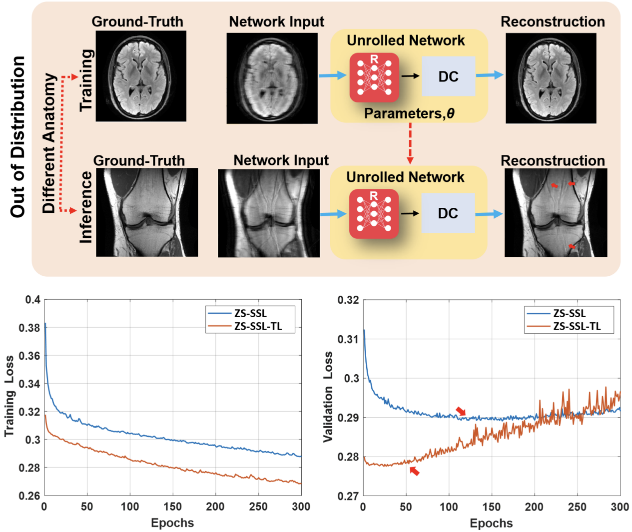


Zero-shot + Transfer Learning

- Pretrained models performance degrades in presence of mismatch between training and test data
- Combine pretrained models with ZS-SSL via transfer learning to improve:

a) accuracy, robustness and generalization

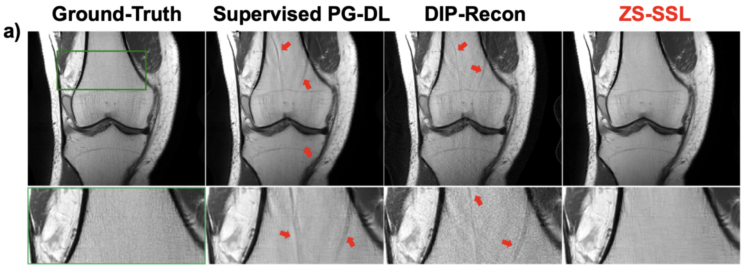
b) computational efficiency



In-Domain Challenges: Sampling& Acc. Rate

Supervised PG-DL was trained with

a) random mask and tested on uniform mask, both at R = 4;

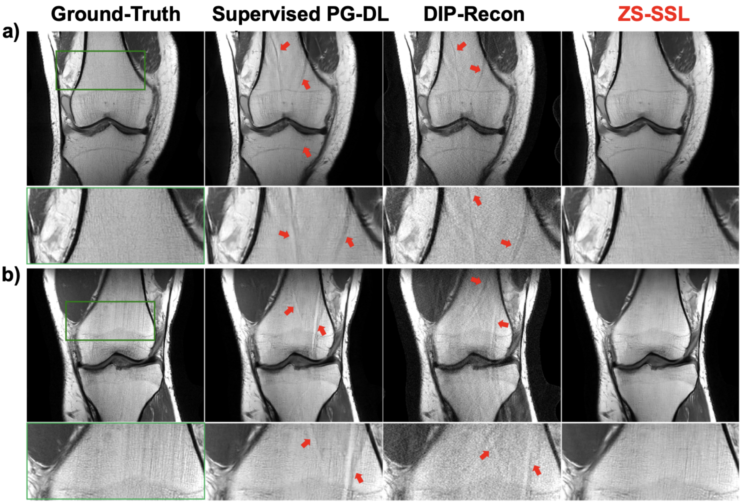


In-Domain Challenges: Sampling& Acc. Rate

Supervised PG-DL was trained with

a) random mask and tested on uniform mask, both at R = 4;

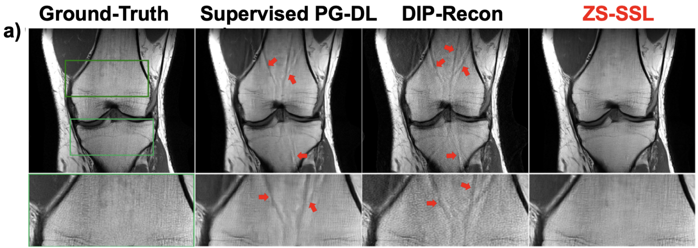
b) uniform mask at R = 4 and tested on uniform mask at R = 6



Cross-Domain Challenges: Anatomy

Supervised PG-DL was trained with

a) Ax-FLAIR (brain) model and tested on Cor-PD (knee)

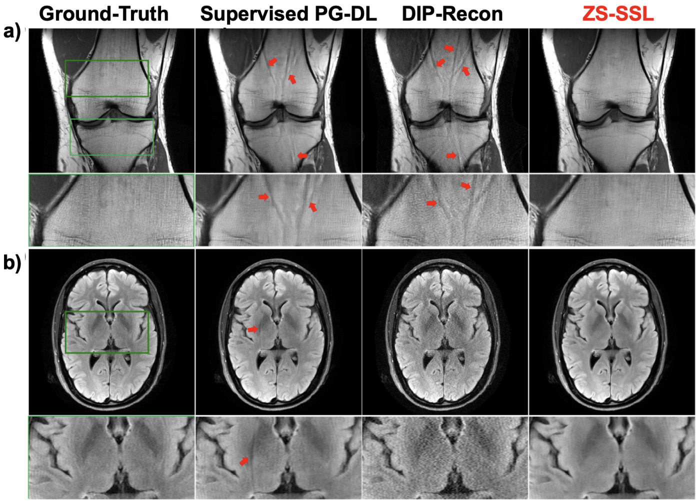


Cross-Domain Challenges: Anatomy

Supervised PG-DL was trained with

a) Ax-FLAIR (brain) model and tested on Cor-PD (knee)

b) Cor-PD model and tested on Ax-FLAIR



Unrolled Networks: Practical Considerations

- Weight sharing
  - Regularizer units may share weights or may be different
  - Unrolling iterative algorithms suggests sharing weights[1] → also fewer parameters[2]
- Loss functions
  - Typically: l_1, l_2 losses[3]
  - Adversarial/Perceptual losses also receiving attention[4,]
- Metrics
  - SSIM/NMSE
  - Reader Study
  - New metrics[5] → Precision, Recall

[1]Monga et al, IEEE SPM,2021; [2]Aggarwal et al, IEEE TMI, 2019 [3]Hammernik et al, MRM, 2018 [4]Hammernik et al, IEEE SPM, 2023; [6]Zhao et al, Nature, 2022

Summary

- Deep learning based image reconstruction research is exploding
- US FDA has approved DL recon for MRI [114] and X-ray CT [115, 116]
- Many omissions...
- Survey papers: [117, 118, 74, 119, 120, 121, 49, 122]
- Other topics:
  - robustness / stability with adversarial noise [84, 123, 124]
  - score-based diffusion models (and uncertainty quantification) [125, 126, 127, 128]
  - quantitative MRI [129]

Acknowledgements

Thanks to numerous graduate students, postdocs, collaborators.

Special thanks for slides to Burhan Yaman and Zaccharie Ramzi!

Resources

Talk: https://web.eecs.umich.edu/~fessler/talk/23/isbi.pdf
Code: https://github.com/JeffFessler/MIRT.jl
https://github.com/JuliaImageRecon

Bibliography I

[1] G. A. Wright. "Magnetic resonance imaging." In: *IEEE Sig. Proc. Mag.* 14.1 (Jan. 1997), 56–66. DOI: 10.1109/79.560324 (cit. on p. 5).

[2] M. Doneva. "Mathematical models for magnetic resonance imaging reconstruction: an overview of the approaches, problems, and future research areas." In: *IEEE Sig. Proc. Mag.* 37.1 (Jan. 2020), 24–32. DOI: 10.1109/MSP.2019.2936964 (cit. on p. 5).

[3] J. A. Fessler. "Model-based image reconstruction for MRI." In: *IEEE Sig. Proc. Mag.* 27.4 (July 2010). Invited submission to special issue on medical imaging, 81–9. DOI: 10.1109/MSP.2010.936726 (cit. on p. 5).

[4] A. Macovski. "Noise in MRI." In: *Mag. Res. Med.* 36.3 (Sept. 1996), 494–7. DOI: 10.1002/mrm.1910360327 (cit. on p. 5).

[5] F. Knoll et al. "Advancing machine learning for MR image reconstruction with an open competition: Overview of the 2019 fastMRI challenge." In: *Mag. Res. Med.* 84.6 (Dec. 2020), 3054–70. DOI: 10.1002/mrm.28338 (cit. on pp. 8, 47).

[6] FDA. *510k premarket notification of HyperSense (GE Medical Systems)*. 2017. URL: https://www.accessdata.fda.gov/scripts/cdrh/cfdocs/cfpmn/pmn.cfm?ID=K162722 (cit. on pp. 11, 12).

[7] FDA. *510k premarket notification of Compressed Sensing Cardiac Cine (Siemens)*. 2017. URL: https://www.accessdata.fda.gov/scripts/cdrh/cfdocs/cfpmn/pmn.cfm?ID=K163312 (cit. on pp. 11, 12).

[8] FDA. *510k premarket notification of Compressed SENSE*. 2018. URL: https://www.accessdata.fda.gov/cdrh_docs/pdf17/K173079.pdf (cit. on pp. 11, 12).

[9] L. Geerts-Ossevoort et al. *Compressed SENSE*. Philips white paper 4522 991 31821 Nov. 2018. 2018. URL: https://philipsproductcontent.blob.core.windows.net/assets/20180109/619119731f2a42c4acd4a863008a46c7.pdf (cit. on pp. 11, 12).

[10] K. T. Block, M. Uecker, and J. Frahm. "Undersampled radial MRI with multiple coils. Iterative image reconstruction using a total variation constraint." In: *Mag. Res. Med.* 57.6 (June 2007), 1086–98. DOI: 10.1002/mrm.21236 (cit. on p. 13).

[11] M. Lustig et al. "Compressed sensing MRI." In: *IEEE Sig. Proc. Mag.* 25.2 (Mar. 2008), 72–82. DOI: 10.1109/MSP.2007.914728 (cit. on p. 13).

[12] J. A. Fessler. "Optimization methods for MR image reconstruction." In: *IEEE Sig. Proc. Mag.* 37.1 (Jan. 2020), 33–40. DOI: 10.1109/MSP.2019.2943645 (cit. on pp. 14–16).

[13] S. Ravishankar and Y. Bresler. "MR image reconstruction from highly undersampled k-space data by dictionary learning." In: *IEEE Trans. Med. Imag.* 30.5 (May 2011), 1028–41. DOI: 10.1109/TMI.2010.2090538 (cit. on pp. 23–25, 31, 32).

[14] M. Aharon, M. Elad, and A. Bruckstein. "K-SVD: an algorithm for designing overcomplete dictionaries for sparse representation." In: *IEEE Trans. Sig. Proc.* 54.11 (Nov. 2006), 4311–22. DOI: 10.1109/TSP.2006.881199 (cit. on pp. 23–25).

[15] S. Ravishankar, R. R. Nadakuditi, and J. A. Fessler. "Efficient sum of outer products dictionary learning (SOUP-DIL) and its application to inverse problems." In: *IEEE Trans. Computational Imaging* 3.4 (Dec. 2017), 694–709. DOI: 10.1109/TCI.2017.2697206 (cit. on pp. 23–25, 28, 29, 31, 32).

[16] M. Lustig and J. M. Pauly. "SPIRiT: Iterative self-consistent parallel imaging reconstruction from arbitrary k-space." In: *Mag. Res. Med.* 64.2 (Aug. 2010), 457–71. DOI: 10.1002/mrm.22428 (cit. on p. 31).

[17] X. Qu et al. "Magnetic resonance image reconstruction from undersampled measurements using a patch-based nonlocal operator." In: *Med. Im. Anal.* 18.6 (Aug. 2014), 843–56. DOI: 10.1016/j.media.2013.09.007 (cit. on pp. 31, 32).

[18] Z. Zhan et al. "Fast multiclass dictionaries learning with geometrical directions in MRI reconstruction." In: *IEEE Trans. Biomed. Engin.* 63.9 (Sept. 2016), 1850–61. DOI: 10.1109/tbme.2015.2503756 (cit. on p. 32).

[19] A. Buades, B. Coll, and J-M. Morel. "The staircasing effect in neighborhood filters and its solution." In: *IEEE Trans. Im. Proc.* 15.6 (June 2006), 1499–505. DOI: 10.1109/TIP.2006.871137 (cit. on p. 34).

[20] K. Dabov et al. "Image denoising by sparse 3-D transform-domain collaborative filtering." In: *IEEE Trans. Im. Proc.* 16.8 (Aug. 2007), 2080–95. DOI: 10.1109/TIP.2007.901238 (cit. on p. 34).

[21] S. H. Chan, X. Wang, and O. A. Elgendy. "Plug-and-play ADMM for image restoration: fixed-point convergence and applications." In: *IEEE Trans. Computational Imaging* 3.1 (Mar. 2017), 84–98. DOI: 10.1109/tci.2016.2629286 (cit. on p. 34).

[22] G. T. Buzzard et al. "Plug-and-play unplugged: optimization-free reconstruction using consensus equilibrium." In: *SIAM J. Imaging Sci.* 11.3 (Jan. 2018), 2001–20. DOI: 10.1137/17m1122451 (cit. on p. 34).

[23] Y. Romano, M. Elad, and P. Milanfar. "The little engine that could: Regularization by denoising (RED)." In: *SIAM J. Imaging Sci.* 10.4 (2017), 1804–44. DOI: 10.1137/16M1102884 (cit. on p. 34).

[24] E. T. Reehorst and P. Schniter. *Regularization by denoising: clarifications and new interpretations*. 2018. URL: http://arxiv.org/abs/1806.02296 (cit. on p. 34).

[25] R. Ahmad et al. "Plug and play methods for magnetic resonance imaging." In: *IEEE Sig. Proc. Mag.* 37.1 (Jan. 2020), 105–16. DOI: 10.1109/MSP.2019.2949470 (cit. on pp. 34, 72–74).

[26] S. Wang et al. "Exploiting deep convolutional neural network for fast magnetic resonance imaging." In: *Proc. Intl. Soc. Mag. Res. Med.* 2016, p. 1778. URL: http://archive.ismrm.org/2016/1778.html (cit. on pp. 40, 44).

[27] D. Lee, J. Yoo, and J. C. Ye. *Deep artifact learning for compressed sensing and parallel MRI*. 2017. URL: http://arxiv.org/abs/1703.01120 (cit. on p. 40).

[28] K. H. Jin et al. "Deep convolutional neural network for inverse problems in imaging." In: *IEEE Trans. Im. Proc.* 26.9 (Sept. 2017), 4509–22. DOI: 10.1109/TIP.2017.2713099 (cit. on p. 40).

[29] M. Akcakaya et al. "Scan-specific robust artificial-neural-networks for k-space interpolation (RAKI) reconstruction: Database-free deep learning for fast imaging." In: *Mag. Res. Med.* 81.1 (Jan. 2019), 439–53. DOI: 10.1002/mrm.27420 (cit. on pp. 40, 45).

[30] Y. Han and J. C. Ye. "K-space deep learning for accelerated MRI." In: *IEEE Trans. Med. Imag.* 39.2 (Feb. 2020), 377–86. DOI: 10.1109/TMI.2019.2927101 (cit. on pp. 40, 45, 47).

[31] M. U. Ghani and W. C. Karl. *Data and image prior integration for image reconstruction using consensus equilibrium*. 2020. URL: http://arxiv.org/abs/2009.00092 (cit. on pp. 40, 45).

[32] B. Zhu et al. “Image reconstruction by domain-transform manifold learning.” In: *Nature* 555 (Mar. 2018), 487–92. DOI: 10.1038/nature25988 (cit. on pp. 40, 46).

[33] I. Haggstrom et al. “DeepPET: A deep encoder-decoder network for directly solving the PET image reconstruction inverse problem.” In: *Med. Im. Anal.* 54 (May 2019), 253–62. DOI: 10.1016/j.media.2019.03.013 (cit. on p. 40).

[34] W. Whiteley, W. K. Luk, and J. Gregor. “DirectPET: full-size neural network PET reconstruction from sinogram data.” In: *J. Med. Im.* 7.3 (Feb. 2020), 1–16. DOI: 10.1117/1.JMI.7.3.032503 (cit. on p. 40).

[35] W. Whiteley et al. “FastPET: near real-time reconstruction of PET histo-image data using a neural network.” In: *IEEE Trans. Radiation and Plasma Med. Sci.* 5.1 (Jan. 2021), 65–77. DOI: 10.1109/TRPMS.2020.3028364 (cit. on p. 40).

[36] Y. Yang et al. “Deep ADMM-net for compressive sensing MRI.” In: *Neural Info. Proc. Sys.* 2016, 10–18. URL: https://papers.nips.cc/paper/6406-deep-admm-net-for-compressive-sensing-mri (cit. on pp. 40, 47, 49).

[37] K. Hammernik et al. “Learning a variational network for reconstruction of accelerated MRI data.” In: *Mag. Res. Med.* 79.6 (June 2018), 3055–71. DOI: 10.1002/mrm.26977 (cit. on pp. 40, 47, 49, 67).

[38] J. Schlemper et al. “A deep cascade of convolutional neural networks for dynamic MR image reconstruction.” In: *IEEE Trans. Med. Imag.* 37.2 (Feb. 2018), 491–503. DOI: 10.1109/tmi.2017.2760978 (cit. on pp. 40, 47, 69).

[39] T. M. Quan, T. Nguyen-Duc, and W-K. Jeong. “Compressed sensing MRI reconstruction using a generative adversarial network with a cyclic loss.” In: *IEEE Trans. Med. Imag.* 37.6 (June 2018), 1488–97. DOI: 10.1109/TMI.2018.2820120 (cit. on pp. 40, 47).

[40] D. Lee et al. “Deep residual learning for accelerated MRI using magnitude and phase networks.” In: *IEEE Trans. Biomed. Engin.* 65.9 (Sept. 2018), 1985–95. DOI: 10.1109/TBME.2018.2821699 (cit. on pp. 40, 47).

[41] G. Nataraj and R. Otazo. “Investigating robustness to unseen pathologies in model-free deep multicoil reconstruction.” In: *ISMRM Workshop on Data Sampling and Image Reconstruction*. 2020 (cit. on p. 42).

[42] G. Yang et al. "DAGAN: Deep de-aliasing generative adversarial networks for fast compressed sensing MRI reconstruction." In: *IEEE Trans. Med. Imag.* 37.6 (June 2018), 1310–21. DOI: 10.1109/TMI.2017.2785879 (cit. on p. 44).

[43] K. He et al. "Deep residual learning for image recognition." In: *Proc. IEEE Conf. on Comp. Vision and Pattern Recognition*. 2016, 770–8. DOI: 10.1109/CVPR.2016.90 (cit. on pp. 44, 58).

[44] K. Zhang et al. "Beyond a Gaussian denoiser: residual learning of deep CNN for image denoising." In: *IEEE Trans. Im. Proc.* 26.7 (July 2017), 3142–55. DOI: 10.1109/tip.2017.2662206 (cit. on p. 44).

[45] H. K. Aggarwal, M. P. Mani, and M. Jacob. "MoDL: model-based deep learning architecture for inverse problems." In: *IEEE Trans. Med. Imag.* 38.2 (Feb. 2019), 394–405. DOI: 10.1109/tmi.2018.2865356 (cit. on pp. 47, 49, 58).

[46] I. Y. Chun et al. "Momentum-Net: Fast and convergent iterative neural network for inverse problems." In: *IEEE Trans. Patt. Anal. Mach. Int.* 45.4 (Apr. 2023), 4915–31. DOI: 10.1109/TPAMI.2020.3012955 (cit. on pp. 47, 49).

[47] P. Putzky et al. *i-RIM applied to the fastMRI challenge*. 2019. URL: http://arxiv.org/abs/1910.08952 (cit. on pp. 47, 63).

[48] M. J. Muckley et al. "Results of the 2020 fastMRI Challenge for Machine Learning MR Image Reconstruction." In: *IEEE Trans. Med. Imag.* 40.9 (Sept. 2021), 2306–17. DOI: 10.1109/TMI.2021.3075856 (cit. on pp. 47, 52).

[49] J. Huang et al. *Data and physics driven learning models for fast MRI – fundamentals and methodologies from CNN, GAN to attention and transformers*. Submitted to ieee-spmag. 2022. URL: http://arxiv.org/abs/2204.01706 (cit. on pp. 48, 119).

[50] K. Gregor and Y. LeCun. "Learning fast approximations of sparse coding." In: *Proc. Intl. Conf. Mach. Learn.* 2010. URL: http://yann.lecun.com/exdb/publis/pdf/gregor-icml-10.pdf (cit. on p. 49).

[51] T. Meinhardt et al. "Learning proximal operators: using denoising networks for regularizing inverse imaging problems." In: *Proc. Intl. Conf. Comp. Vision*. 2017, 1799–808. DOI: 10.1109/ICCV.2017.198 (cit. on p. 49).

[52] U. Schmidt and S. Roth. "Shrinkage fields for effective image restoration." In: *Proc. IEEE Conf. on Comp. Vision and Pattern Recognition*. 2014, 2774–81. DOI: 10.1109/CVPR.2014.349 (cit. on p. 49).

[53] Y. Chen, W. Yu, and T. Pock. "On learning optimized reaction diffusion processes for effective image restoration." In: *Proc. IEEE Conf. on Comp. Vision and Pattern Recognition*. 2015, 5261–9. DOI: 10.1109/CVPR.2015.7299163 (cit. on p. 49).

[54] Y. Chen and T. Pock. "Trainable nonlinear reaction diffusion: A flexible framework for fast and effective image restoration." In: *IEEE Trans. Patt. Anal. Mach. Int.* 39.6 (June 2017), 1256–72. DOI: 10.1109/TPAMI.2016.2596743 (cit. on p. 49).

[55] K. Hammernik et al. "Learning a variational model for compressed sensing MRI reconstruction." In: *Proc. Intl. Soc. Mag. Res. Med.* 2016, p. 1088. URL: http://archive.ismrm.org/2016/1088.html (cit. on p. 49).

[56] Y. Yang et al. *ADMM-net: A deep learning approach for compressive sensing MRI*. 2017. URL: http://arxiv.org/abs/1705.06869 (cit. on p. 49).

[57] B. Xin et al. *Maximal sparsity with deep networks?* 2016. URL: http://arxiv.org/abs/1605.01636 (cit. on p. 49).

[58] P. Schniter. "Recent advances in approximate message passing." In: *spars-17*. 2017, plenary (cit. on p. 49).

[59] H-Y. Liu et al. "Compressive imaging with iterative forward models." In: *Proc. IEEE Conf. Acoust. Speech Sig. Proc.* 2017 (cit. on p. 49).

[60] J. Adler and O. Oktem. "Learned primal-dual reconstruction." In: *IEEE Trans. Med. Imag.* 37.6 (June 2018), 1322–32. DOI: 10.1109/tmi.2018.2799231 (cit. on pp. 49, 50).

[61] Y. Malitsky and T. Pock. "A first-order primal-dual algorithm with linesearch." In: *SIAM J. Optim.* 28.1 (Jan. 2018), 411–32. DOI: 10.1137/16m1092015 (cit. on p. 49).

[62] S. Ravishankar et al. "Deep dictionary-transform learning for image reconstruction." In: *Proc. IEEE Intl. Symp. Biomed. Imag.* 2018, 1208–12. DOI: 10.1109/ISBI.2018.8363788 (cit. on pp. 49, 61).

[63] S. Ravishankar, I. Y. Chun, and J. A. Fessler. "Physics-driven deep training of dictionary-based algorithms for MR image reconstruction." In: *Proc., IEEE Asilomar Conf. on Signals, Systems, and Comp.* Invited. 2017, 1859–63. DOI: 10.1109/ACSSC.2017.8335685 (cit. on p. 49).

Bibliography VII

[64] I. Y. Chun and J. A. Fessler. "Deep BCD-net using identical encoding-decoding CNN structures for iterative image recovery." In: *Proc. IEEE Wkshp. on Image, Video, Multidim. Signal Proc.* 2018, 1–5. DOI: 10.1109/IVMSPW.2018.8448694 (cit. on p. 49).

[65] I. Y. Chun and J. A. Fessler. *Deep BCD-net using identical encoding-decoding CNN structures for iterative image recovery*. 2018. URL: http://arxiv.org/abs/1802.07129 (cit. on p. 49).

[66] H. Lim et al. "Application of trained deep BCD-net to iterative low-count PET image reconstruction." In: *Proc. IEEE Nuc. Sci. Symp. Med. Im. Conf.* 2018, 1–4. DOI: 10.1109/NSSMIC.2018.8824563 (cit. on p. 49).

[67] I. Y. Chun et al. "Fast and convergent iterative image recovery using trained convolutional neural networks." In: *Allerton Conf. on Comm., Control, and Computing*. Invited. 2018, 155–9. DOI: 10.1109/ALLERTON.2018.8635932 (cit. on p. 49).

[68] I. Y. Chun et al. *Momentum-Net: Fast and convergent iterative neural network for inverse problems*. 2019. URL: http://arxiv.org/abs/1907.11818 (cit. on p. 49).

[69] J. R. Hershey, J. L. Roux, and F. Weninger. *Deep unfolding: Model-based inspiration of novel deep architectures*. 2014. URL: http://arxiv.org/abs/1409.2574 (cit. on p. 49).

[70] H. K. Aggarwal, M. P. Mani, and M. Jacob. "Model based image reconstruction using deep learned priors (MODL)." In: *Proc. IEEE Intl. Symp. Biomed. Imag.* 2018, 671–4. DOI: 10.1109/ISBI.2018.8363663 (cit. on pp. 49, 58).

[71] K. H. Jin, M. McCann, and M. Unser. "BPConvNet for compressed sensing recovery in bioimaging." In: *spars-17*. 2017 (cit. on p. 49).

[72] H. Chen et al. "LEARN: Learned experts: assessment-based reconstruction network for sparse-data CT." In: *IEEE Trans. Med. Imag.* 37.6 (June 2018), 1333–47. DOI: 10.1109/TMI.2018.2805692 (cit. on p. 49).

[73] D. Wu et al. *End-to-end abnormality detection in medical imaging*. 2018. URL: http://arxiv.org/abs/1711.02074 (cit. on p. 49).

[74] D. Liang et al. "Deep MRI reconstruction: Unrolled optimization algorithms meet neural networks." In: *IEEE Sig. Proc. Mag.* 37.1 (Jan. 2020), 141–51. DOI: 10.1109/MSP.2019.2950557 (cit. on pp. 49, 119).

[75] V. Monga, Y. Li, and Y. C. Eldar. "Algorithm unrolling: interpretable, efficient deep learning for signal and image processing." In: *IEEE Sig. Proc. Mag.* 38.2 (Mar. 2021), 18–44. DOI: 10.1109/MSP.2020.3016905 (cit. on p. 49).

[76] Z. Ramzi, P. Ciuciu, and J-L. Starck. "Benchmarking MRI reconstruction neural networks on large public datasets." In: *Appl. Sci.* 10.5 (2020), p. 1816. DOI: 10.3390/app10051816 (cit. on p. 50).

[77] Z. Ramzi, P. Ciuciu, and J-L. Starck. *XPDNet for MRI Reconstruction: an application to the 2020 fastMRI challenge*. 2020. URL: http://arxiv.org/abs/2010.07290 (cit. on p. 52).

[78] I. J. Goodfellow et al. *Generative adversarial networks*. 2014. URL: http://arxiv.org/abs/1406.2661 (cit. on pp. 53, 54).

[79] X. Chen et al. "InfoGAN: interpretable representation learning by information maximizing generative adversarial nets." In: *Neural Info. Proc. Sys.* 2016, 2172–80. URL: https://papers.nips.cc/paper/6399-infogan-interpretable-representation-learning-by-information-maximizing-generative-adversarial-nets (cit. on pp. 53, 54).

[80] A. Bora et al. "Compressed sensing using generative models." In: *Proc. Intl. Conf. Mach. Learn.* Vol. 70. 2017, 537–46. URL: http://proceedings.mlr.press/v70/bora17a.html (cit. on pp. 53, 54, 100).

[81] S. Kolouri et al. *Sliced-Wasserstein autoencoder: an embarrassingly simple generative model*. 2018. URL: https://openreview.net/forum?id=H1xaJn05FQ (cit. on p. 55).

[82] D. Berthelot, T. Schumm, and L. Metz. *BEGAN: boundary equilibrium generative adversarial networks*. 2017. URL: http://arxiv.org/abs/1703.10717 (cit. on pp. 56, 57).

[83] B. Yaman et al. "Self-supervised learning of physics-based reconstruction neural networks without fully-sampled reference data." In: *Mag. Res. Med.* 84.6 (Dec. 2020), 3172–91. DOI: 10.1002/mrm.28378 (cit. on p. 60).

[84] V. Antun et al. "On instabilities of deep learning in image reconstruction and the potential costs of AI." In: *Proc. Natl. Acad. Sci.* 117.48 (Dec. 2020), 30088–95. DOI: 10.1073/pnas.1907377117 (cit. on pp. 60, 119).

[85] M. Lustig, D. Donoho, and J. M. Pauly. "Sparse MRI: The application of compressed sensing for rapid MR imaging." In: *Mag. Res. Med.* 58.6 (Dec. 2007), 1182–95. DOI: 10.1002/mrm.21391 (cit. on p. 62).

[86] S. Ravishankar and Y. Bresler. "Data-driven learning of a union of sparsifying transforms model for blind compressed sensing." In: *IEEE Trans. Computational Imaging* 2.3 (Sept. 2016), 294–309. DOI: 10.1109/TCI.2016.2567299 (cit. on p. 62).

[87] T. Chen et al. *Training deep nets with sublinear memory cost*. 2016. URL: http://arxiv.org/abs/1604.06174 (cit. on p. 63).

[88] A. N. Gomez et al. "The reversible residual network: backpropagation without storing activations." In: *NeurIPS*. Vol. 30. 2017. URL: https://proceedings.neurips.cc/paper/2017/hash/f9be311e65d81a9ad8150a60844bb94c-Abstract.html (cit. on p. 63).

[89] J. Behrmann et al. "Invertible residual networks." In: *Proc. Intl. Conf. Mach. Learn.* Vol. 97. 2019, 573–82. URL: http://proceedings.mlr.press/v97/behrmann19a.html (cit. on p. 63).

[90] P. Putzky and M. Welling. "Invert to learn to invert." In: *NeurIPS*. Vol. 32. 2019, 446–56. URL: https://papers.nips.cc/paper/2019/hash/ac1dd209cbcc5e5d1c6e28598e8cbbe8-Abstract.html (cit. on p. 63).

[91] V. A. Kelkar, S. Bhadra, and M. A. Anastasio. *Compressible latent-space invertible networks for generative model-constrained image reconstruction*. 2020. URL: http://arxiv.org/abs/2007.02462 (cit. on p. 63).

[92] M. E. Sander et al. "Momentum residual neural networks." In: *Proc. Intl. Conf. Mach. Learn.* Vol. 139. 2021, 9276–87. URL: https://proceedings.mlr.press/v139/sander21a.html (cit. on p. 63).

[93] A. Ziabari et al. "2.5D deep learning for CT image reconstruction using A multi-GPU implementation." In: *asccs*. 2018, 2044–9. DOI: 10.1109/ACSSC.2018.8645364 (cit. on p. 63).

[94] S. Lee et al. *Improving 3D imaging with pre-trained perpendicular 2D diffusion models*. 2023. URL: http://arxiv.org/abs/2303.08440 (cit. on p. 63).

[95] R. T. Q. Chen et al. "Neural ordinary differential equations." In: *NeurIPS*. Vol. 31. 2018. URL: https://papers.nips.cc/paper/2018/hash/69386f6bb1dfed68692a24c8686939b9-Abstract.html (cit. on p. 63).

[96] J. J. Park et al. "DeepSDF: learning continuous signed distance functions for shape representation." In: *Proc. IEEE Conf. on Comp. Vision and Pattern Recognition*. 2019, 165–74. DOI: 10.1109/CVPR.2019.00025 (cit. on p. 63).

[97] B. Mildenhall et al. "NeRF: representing scenes as neural radiance fields for view synthesis." In: *Proc. European Comp. Vision Conf.* 2020. URL: http://arxiv.org/abs/2003.08934 (cit. on p. 63).

[98] W. Huang et al. *Neural implicit k-space for binning-free non-cartesian cardiac MR imaging*. 2022. URL: http://arxiv.org/abs/2212.08479 (cit. on p. 63).

[99] W. Peng et al. "Learning optimal K-space acquisition and reconstruction using physics-informed neural networks." In: *Proc. IEEE Conf. on Comp. Vision and Pattern Recognition*. 2022, 20762–71. DOI: 10.1109/CVPR52688.2022.02013 (cit. on p. 63).

[100] L. Lozenski, M. A. Anastasio, and U. Villa. "A memory-efficient self-supervised dynamic image reconstruction method using neural fields." In: *IEEE Trans. Computational Imaging* 8 (2022), 879–92. DOI: 10.1109/TCI.2022.3208511 (cit. on pp. 63, 72–74).

[101] S. Bai, J. Z. Kolter, and V. Koltun. "Deep equilibrium models." In: *NeurIPS*. Vol. 32. 2019. URL: https://papers.nips.cc/paper/2019/hash/01386bd6d8e091c2ab4c7c7de644d37b-Abstract.html (cit. on pp. 63, 72–74).

[102] D. Gilton, G. Ongie, and R. Willett. "Deep equilibrium architectures for inverse problems in imaging." In: *IEEE Trans. Computational Imaging* 7 (2021), 1123–33. DOI: 10.1109/TCI.2021.3118944 (cit. on pp. 63, 72–74).

[103] S. W. Fung et al. "JFB: Jacobian-free backpropagation for implicit networks." In: *Proc. AAAI Conf. on Artificial Intell.* 2022. URL: https://aaai-2022.virtualchair.net/poster_aaai10048 (cit. on pp. 63, 72–74).

[104] W. Gan et al. *Self-supervised deep equilibrium models for inverse problems with theoretical guarantees*. submitted to CoRR. 2022. URL: https://openreview.net/forum?id=WjjkInTQn6 (cit. on p. 63).

[105] Z. Ramzi et al. "SHINE: SHaring the INverse Estimate from the forward pass for bi-level optimization and implicit models." In: *Proc. Intl. Conf. on Learning Representations*. 2022. URL: https://openreview.net/forum?id=-ApAkox5mp (cit. on p. 63).

[106] A. Pramanik, M. B. Zimmerman, and M. Jacob. "Memory-efficient model-based deep learning with convergence and robustness guarantees." In: *IEEE Trans. Computational Imaging* 9 (2023), 260–75. DOI: 10.1109/TCI.2023.3252268 (cit. on p. 63).

[107] S. A. H. Hosseini et al. "Dense recurrent neural networks for inverse problems: History-cognizant unrolling of optimization algorithms." In: *IEEE J. Sel. Top. Sig. Proc.* 14.6 (Oct. 2020), 1280–91. DOI: 10.1109/JSTSP.2020.3003170.

[108] J. Liu et al. *Online deep equilibrium learning for regularization by denoising*. 2022. URL: http://arxiv.org/abs/2205.13051 (cit. on pp. 72–74).

[109] A. Pramanik and M. Jacob. *Stable and memory-efficient image recovery using monotone operator learning (MOL)*. 2022. URL: http://arxiv.org/abs/2206.04797 (cit. on pp. 72–74).

[110] A. Pramanik and M. Jacob. *Accelerated parallel MRI using memory efficient and robust monotone operator learning (MOL)*. 2023. URL: http://arxiv.org/abs/2304.01351 (cit. on pp. 72–74).

[111] B. Yaman, S. A. H. Hosseini, and M. Akcakaya. "Zero-shot self-supervised learning for MRI reconstruction." In: *Proc. Intl. Conf. on Learning Representations*. 2022. URL: https://openreview.net/forum?id=085y6YPaYjP (cit. on pp. 97–99).

[112] D. Ulyanov, A. Vedaldi, and V. Lempitsky. "Deep image prior." In: *Proc. IEEE Conf. on Comp. Vision and Pattern Recognition*. 2018, 9446–54. URL: http://openaccess.thecvf.com/content_cvpr_2018/html/Ulyanov_Deep_Image_Prior_CVPR_2018_paper.html (cit. on p. 100).

[113] J. Yoo et al. "Time-dependent deep image prior for dynamic MRI." In: *IEEE Trans. Med. Imag.* 40.12 (Dec. 2021), 3337–48. DOI: 10.1109/TMI.2021.3084288 (cit. on p. 100).

[114] FDA. *510k premarket notification of GE AIR Recon DL*. 2020. URL: https://www.accessdata.fda.gov/scripts/cdrh/cfdocs/cfpmn/pmn.cfm?ID=K193282 (cit. on p. 119).

[115] FDA. *510k premarket notification of AiCE Deep Learning Reconstruction (Canon)*. 2019. URL: https://www.accessdata.fda.gov/scripts/cdrh/cfdocs/cfpmn/pmn.cfm?ID=K183046 (cit. on p. 119).

Bibliography XII

[116] FDA. *510k premarket notification of Deep Learning Image Reconstruction (GE Medical Systems)*. 2019. URL: https://www.accessdata.fda.gov/scripts/cdrh/cfdocs/cfpmn/pmn.cfm?ID=K183202 (cit. on p. 119).

[117] P. M. Johnson, M. P. Recht, and F. Knoll. "Improving the speed of MRI with artificial intelligence." In: *Semin Musculoskelet Radiol* 24.1 (2020), 12–20. DOI: 10.1055/s-0039-3400265 (cit. on p. 119).

[118] F. Knoll et al. "Deep learning methods for parallel magnetic resonance image reconstruction." In: *IEEE Sig. Proc. Mag.* 37.1 (Jan. 2020), 128–40. DOI: 10.1109/MSP.2019.2950640 (cit. on p. 119).

[119] C. M. Sandino et al. "Compressed sensing: From research to clinical practice with deep neural networks: shortening scan times for magnetic resonance imaging." In: *IEEE Sig. Proc. Mag.* 37.1 (Jan. 2020), 117–27. DOI: 10.1109/MSP.2019.2950433 (cit. on p. 119).

[120] D. J. Lin et al. "Artificial intelligence for MR image reconstruction: An overview for clinicians." In: *J. Mag. Res. Im.* 53.4 (Apr. 2021), 1015–28. DOI: 10.1002/jmri.27078 (cit. on p. 119).

[121] M. Akcakaya et al. "Unsupervised deep learning methods for biological image reconstruction and enhancement: an overview from a signal processing perspective." In: *IEEE Sig. Proc. Mag.* 39.2 (2022), 28–44. DOI: 10.1109/MSP.2021.3119273 (cit. on p. 119).

[122] K. Hammernik et al. "Physics-driven deep learning for computational magnetic resonance imaging: combining physics and machine learning for improved medical imaging." In: *IEEE Sig. Proc. Mag.* 40.1 (2023), 98–114. DOI: 10.1109/MSP.2022.3215288 (cit. on p. 119).

[123] J. Jia et al. "On the robustness of deep learning-based MRI reconstruction to image transformations." In: *NeurIPS Wkshp TSRML*. 2022. URL: https://openreview.net/forum?id=guu52Gtj1B (cit. on p. 119).

[124] H. Li et al. "SMUG: Towards robust MRI reconstruction by smoothed unrolling." In: *Proc. IEEE Conf. Acoust. Speech Sig. Proc.* 2023. URL: http://arxiv.org/abs/2303.12735 (cit. on p. 119).

[125] Z. Ramzi et al. "Denoising score-matching for uncertainty quantification in inverse problems." In: *NeurIPS 2020 Workshop on Deep Learning and Inverse Problems*. 2020. URL: https://openreview.net/forum?id=GpwoGZNeUC (cit. on p. 119).

[126] H. Chung and J. C. Ye. "Score-based diffusion models for accelerated MRI." In: *Med. Im. Anal.* 80 (Aug. 2022), p. 102479. DOI: 10.1016/j.media.2022.102479 (cit. on p. 119).

[127] Y. Song et al. "Solving inverse problems in medical imaging with score-based generative models." In: *Proc. Intl. Conf. on Learning Representations.* 2022. URL: https://openreview.net/forum?id=vaRCHVj0uGI (cit. on p. 119).

[128] G. Luo et al. "Bayesian MRI reconstruction with joint uncertainty estimation using diffusion models." In: *Mag. Res. Med.* (2023). DOI: 10.1002/mrm.29624 (cit. on p. 119).

[129] Y. Zhu et al. "Physics-driven deep learning methods for fast quantitative magnetic resonance imaging: performance improvements through integration with deep neural networks." In: *IEEE Sig. Proc. Mag.* 40.2 (2023), 116–28. DOI: 10.1109/MSP.2023.3236483 (cit. on p. 119).