Medical image reconstruction using adaptive signal models

Jeffrey A. Fessler

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

Machine learning for medical imaging (ML4MI) workshop
2018-10-05

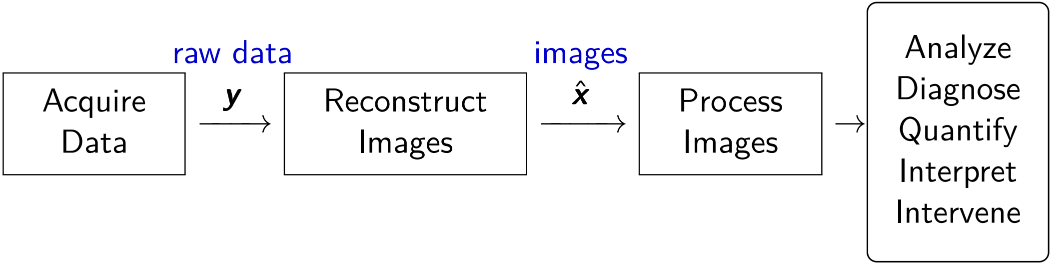
Acknowledgments: Doug Noll, Sai Ravishankar, Raj Nadakuditi,
Jon Nielsen, Gopal Nataraj, Il Yong Chun, Xuehang Zheng, ...

Declaration: No relevant financial interests or relationships to disclose

Outline

Medical imaging overview

Overview of medical imaging:



Most obvious place for machine learning is post-processing:

reconstructed images $\hat{\boldsymbol{x}}$ $\rightarrow$ $\boxed{\text{ML-based image processing (classification or regression)}}$ $\rightarrow$ interpretation

...

Machine learning in medical image interpretation

Most obvious place for machine learning is post-processing:

reconstructed images $\hat{\boldsymbol{x}}$ $\rightarrow$ $\boxed{\text{ML-based image processing (classification or regression)}}$ $\rightarrow$ interpretation

(Many conference sessions; special issue of IEEE Trans. on Med. Imaging in May 2016 [1], ...)

Another (initially less obvious?) place for machine learning (multiple conference sessions):

raw data $\boldsymbol{y}$ $\rightarrow$ $\boxed{\text{ML-based image reconstruction}}$ $\rightarrow$ images $\hat{\boldsymbol{x}}$

...

Another (initially less obvious?) place for machine learning (multiple conference sessions):

raw data $\boldsymbol{y}$ $\rightarrow$ $\boxed{\text{ML-based image reconstruction}}$ $\rightarrow$ images $\hat{\boldsymbol{x}}$

Possibly easier (than diagnosis) due to lower bar:

- current reconstruction methods based on simplistic image models;
- human eyes are better at detection than at solving inverse problems.

Machine learning in medical image reconstruction

Another (initially less obvious?) place for machine learning (multiple conference sessions):

raw data $\boldsymbol{y}$ $\rightarrow$ $\boxed{\text{ML-based image reconstruction}}$ $\rightarrow$ images $\hat{\boldsymbol{x}}$

Possibly easier (than diagnosis) due to lower bar:

- current reconstruction methods based on simplistic image models;
- human eyes are better at detection than at solving inverse problems.

June 2018 special issue of IEEE Trans. on Medical Imaging [2]:

IEEE TRANSACTIONS ON MEDICAL IMAGING, VOL. 37, NO. 6, JUNE 2018 1289

Image Reconstruction Is a New Frontier of Machine Learning

Ge Wang, *Fellow, IEEE*, Jong Chu Ye, *Senior Member, IEEE*, Klaus Mueller, *Senior Member, IEEE*, and Jeffrey A. Fessler, *Fellow, IEEE*

A more speculative opportunity for machine learning:

$$\begin{array}{c}\text{raw data}\\ \boldsymbol{y}\end{array} \rightarrow \boxed{\begin{array}{c}\text{ML-based}\\ \text{“magic”}\end{array}} \rightarrow \text{interpretation}$$

...

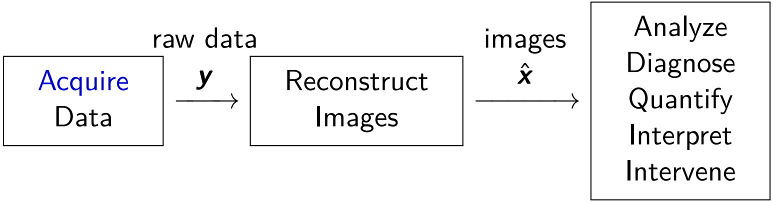
A more speculative opportunity for machine learning:

$$\begin{array}{c}\text{raw data}\\ \boldsymbol{y}\end{array} \rightarrow \boxed{\begin{array}{c}\text{ML-based}\\ \text{“magic”}\end{array}} \rightarrow \text{interpretation}$$

- CT sinogram to vessel diameter [3]
- k-space to ???

See Wiro Niessen's keynote...

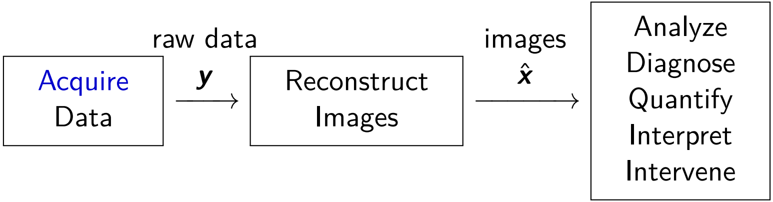
One more opportunity for ML in medical imaging:



...

Machine learning in medical imaging: scan design

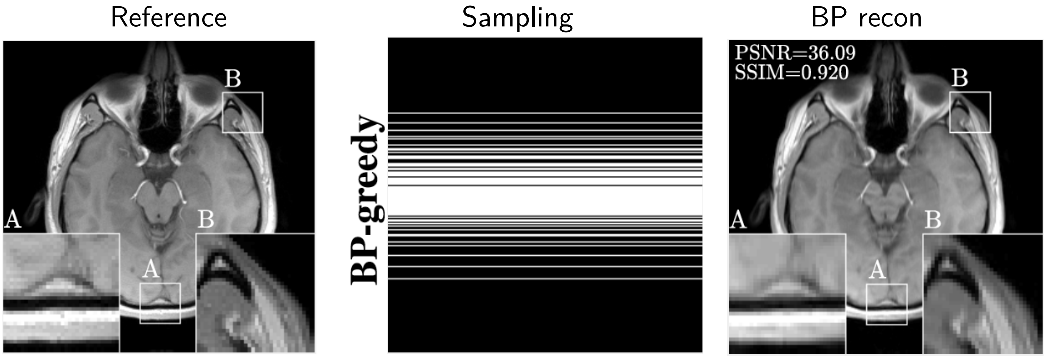
One more opportunity for ML in medical imaging:



Choose best k-space phase encoding locations based on training images:

- "Learning-based compressive MRI" [4, 5]
  (Volkan Cevher group, June 2018 IEEE T-MI)
  Single coil only so far; perhaps hard to generalize to parallel MRI?
- Yue Cao and David Levin, MRM Sep. 1993 "Feature recognizing MRI" [6–8]

Adaptive phase-encode selection



Sampling designed to optimize PSNR for basis pursuit (BP) reconstruction using shearlet transform, at 25% sampling rate.

Sampling design considers both the training data and the reconstruction method.

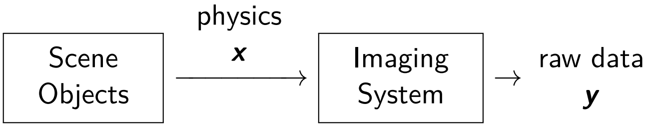
No high spatial frequencies!?

(Images from Gözcü et al. [5].)

Outline

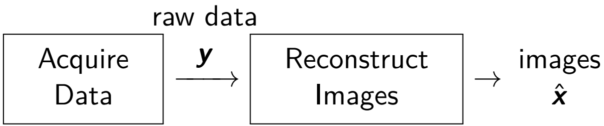
Image reconstruction background

- Forward problem (data acquisition):



SPECT, PET, X-ray CT, MRI, optical, ...

- Inverse problem (image formation):



- Image reconstruction topics: physics models, measurement statistical models, regularization / object priors, optimization...

Generations of medical image reconstruction methods

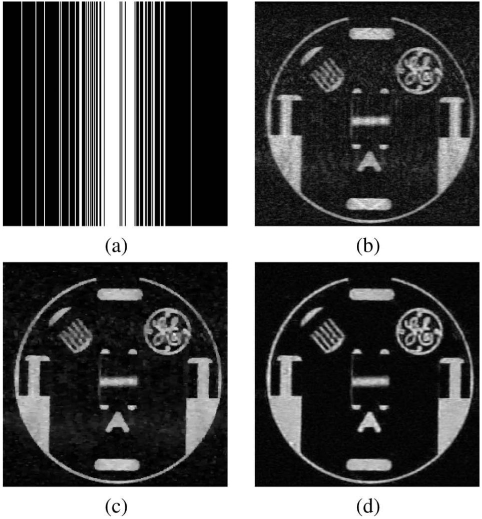
1. 70's "Analytical" methods (integral equations)
   FBP for SPECT / PET / X-ray CT, IFFT for MRI, ...
2. 80's Algebraic methods (as in "linear algebra")
   Solve $\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x}$
3. 90's Statistical methods
   - LS / ML methods
   - Bayesian methods (Markov random fields, ...)
   - regularized methods
4. 00's Compressed sensing methods
   (mathematical sparsity models)
5. 10's Adaptive / data-driven methods
   machine learning, deep learning, ...

Accelerating MR imaging using adaptive regularization

(a) $4\times$ under-sampled MR k-space
(b) zero-filled reconstruction
(c) "compressed sensing"
reconstruction with TV regularization
(d) adaptive regularization
using dictionary learning

Ravishankar & Bresler, DLMRI, T-MI, May 2011,
[9, Fig. 10]

DL = dictionary learning
(not "deep learning")

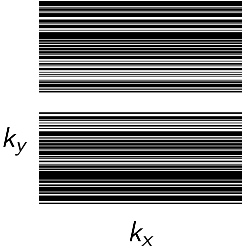


Ill-posed inverse problems

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

$\boldsymbol{y}$: measurements $\quad \varepsilon$: noise
$\boldsymbol{x}$: unknown image $\quad \boldsymbol{A}$: system matrix (typically wide)

- compressed sensing (*e.g.*, MRI) ($\boldsymbol{A}$ "random" rows of DFT)



- deblurring (restoration) ($\boldsymbol{A}$ Toeplitz)
- in-painting ($\boldsymbol{A}$ subset of rows of $\boldsymbol{I}$)
- denoising (not ill posed) ($\boldsymbol{A} = \boldsymbol{I}$)

Inverse problems via MAP estimation

$$\begin{array}{c} \text{Unknown} \\ \text{image} \\ \boldsymbol{x} \end{array} \rightarrow \boxed{\begin{array}{c} \text{System model} \\ \mathrm{p}(\boldsymbol{y} \,|\, \boldsymbol{x}) \end{array}} \rightarrow \begin{array}{c} \text{Data} \\ \boldsymbol{y} \end{array} \rightarrow \boxed{\begin{array}{c} \text{Estimator} \\ \end{array}} \rightarrow \begin{array}{c} \text{Recon.} \\ \text{image} \\ \hat{\boldsymbol{x}} \end{array}$$

If we have a prior $\mathrm{p}(\boldsymbol{x})$, then the MAP estimate is:

$$\hat{\boldsymbol{x}} = \arg\max_{\boldsymbol{x}} \mathrm{p}(\boldsymbol{x} \,|\, \boldsymbol{y}) = \arg\max_{\boldsymbol{x}} \log \mathrm{p}(\boldsymbol{y} \,|\, \boldsymbol{x}) + \log \mathrm{p}(\boldsymbol{x}) \,.$$

For gaussian measurement errors and a linear forward model:

$$-\log \mathrm{p}(\boldsymbol{y} \,|\, \boldsymbol{x}) \equiv \frac{1}{2} \|\boldsymbol{y} - \boldsymbol{A}\boldsymbol{x}\|_{\boldsymbol{W}}^2$$

where $\|\boldsymbol{y}\|_{\boldsymbol{W}}^2 = \boldsymbol{y}' \boldsymbol{W} \boldsymbol{y}$

and $\boldsymbol{W}^{-1} = \mathrm{Cov}\{\boldsymbol{y} \,|\, \boldsymbol{x}\}$ is known
($\boldsymbol{A}$ from physics, $\boldsymbol{W}$ from statistics)

Priors for MAP estimation

- If all images $\boldsymbol{x}$ are "plausible" (have non-zero probability) then

$$\mathrm{p}(\boldsymbol{x}) \propto \mathrm{e}^{-\mathsf{R}(\boldsymbol{x})} \Longrightarrow -\log \mathrm{p}(\boldsymbol{x}) \equiv \mathsf{R}(\boldsymbol{x})$$

(from fantasy / imagination / wishful thinking / data)

- MAP $\equiv$ regularized weighted least-squares (WLS) estimation:

$$\begin{aligned}\hat{\boldsymbol{x}} &= \arg\max_{\boldsymbol{x}} \log \mathrm{p}(\boldsymbol{y} \mid \boldsymbol{x}) + \log \mathrm{p}(\boldsymbol{x}) \\ &= \arg\min_{\boldsymbol{x}} \frac{1}{2} \|\boldsymbol{y} - \boldsymbol{A}\boldsymbol{x}\|_{\boldsymbol{W}}^2 + \mathsf{R}(\boldsymbol{x})\end{aligned}$$

- A regularizer $\mathsf{R}(\boldsymbol{x})$, aka log prior, is essential for high-quality solutions to ill-conditioned / ill-posed inverse problems.
- Why ill-posed? Often high ambitions...

Non-adaptive regularizers

- Tikhonov regularization (IID gaussian prior)
- Markov random field (MRF) models
- Roughness penalty (cf MRF prior)
- Edge-preserving regularization
- Total-variation (TV) regularization
- Black-box denoiser like NLM, *e.g.*, plug-and-play ADMM [10]
- Sparsity in ambient space
- Sparsifying transforms: wavelets, curvelets, ...
- Graphical models
- ...

All "hand crafted" from statistical / mathematical models ...

Simpler methods for ML in image reconstruction

Many possible ways to use ML ideas in image reconstruction.

Basic "fast" methods:

- Enhance raw data (k-space, sinogram, ...)
- Enhance poorly reconstructed image
  - patch-based
  - image-based

Computation / quality trade-offs ?

...

Simpler methods for ML in image reconstruction

Many possible ways to use ML ideas in image reconstruction.

Basic "fast" methods:

- Enhance raw data (k-space, sinogram, ...)
- Enhance poorly reconstructed image
  - patch-based
  - image-based

Computation / quality trade-offs ?

Basic "slow" methods:

- Auto-tune regularization parameter(s)
- Provide an initial image for "conventional" iterative reconstruction

May not fully exploit the potential of ML

- ML-based "prior" image for iterative reconstruction [11]:

$$\hat{\boldsymbol{x}} = \arg\min_{\boldsymbol{x}} \|\boldsymbol{A}\boldsymbol{x} - \boldsymbol{y}\|_2^2 + \beta \|\boldsymbol{x} - \boldsymbol{x}_{\text{prior}}\|_p^p$$

Fast for $p = 2$, but $p = 1$ more robust to errors in prior image
Reminiscent of U. Wisconsin's PICCS methods, *e.g.*, [12]

Advanced “fast” methods for ML-based IR

- ML-based “prior” image for iterative reconstruction [11]:

$$\hat{\boldsymbol{x}} = \arg\min_{\boldsymbol{x}} \|\boldsymbol{A}\boldsymbol{x} - \boldsymbol{y}\|_2^2 + \beta \|\boldsymbol{x} - \boldsymbol{x}_{\text{prior}}\|_p^p$$

  Fast for $p = 2$, but $p = 1$ more robust to errors in prior image
  Reminiscent of U. Wisconsin’s PICCS methods, *e.g.*, [12]

- Unrolled loop (recurrent NN) with learned components [13–16]

Nonlinear encoder methods for ML-based IR

- ML-based nonlinear encoder, *e.g.*, autoencoder or generative adversarial network (GAN) [17, 18]: nonlinear generalizations of subspace models
- learn G: maps low-dimensional latent parameter $\boldsymbol{z}$ into high-dimensional image $\boldsymbol{x}$

▶ Synthesis form [19]:

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

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

Nonlinear encoder methods for ML-based IR

- ML-based nonlinear encoder, *e.g.*, autoencoder or generative adversarial network (GAN) [17, 18]: nonlinear generalizations of subspace models
- learn G: maps low-dimensional latent parameter $\boldsymbol{z}$ into high-dimensional image $\boldsymbol{x}$

▶ Synthesis form [19]:

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

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

▶ Regularizer form:

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

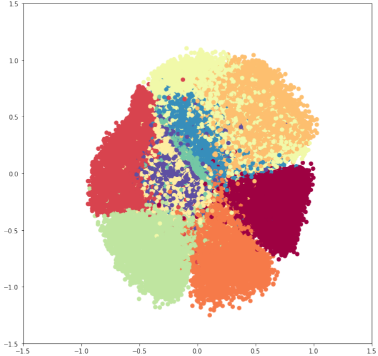
$$R_{\text{encoder}}(\boldsymbol{x}) = \min_{\boldsymbol{z}} \|\boldsymbol{x} - G(\boldsymbol{z})\|_p^p$$

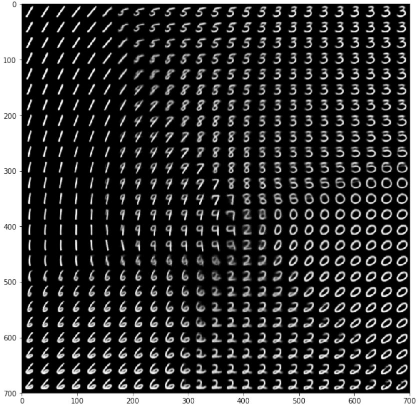
Expensive non-convex double minimization, but more robust to encoder?

From jupyter notebook for [20] (13 layer CNN with $\approx$ 300K learned parameters) at

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

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





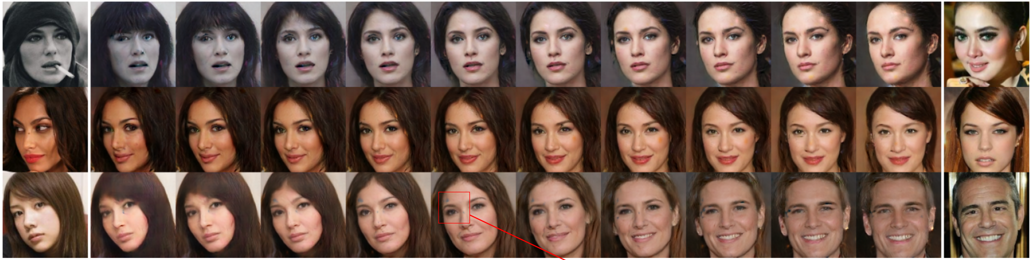
Where is 4?

Generative Adversarial Networks (GAN) example

From Google's [21]:

Much more realistic than linear interpolation (averaging).
"setting a new milestone in visual quality" [21].

From Google's [21]:



Non-physical output!

- Data
  - Population adaptive methods (*e.g.*, X-ray CT)
  - Patient adaptive methods (*e.g.*, dynamic MRI?)
- Spatial structure
  - Patch-based models
  - Convolutional models
- Regularizer formulation
  - Synthesis (dictionary) approach
  - Analysis (sparsifying transforms) approach

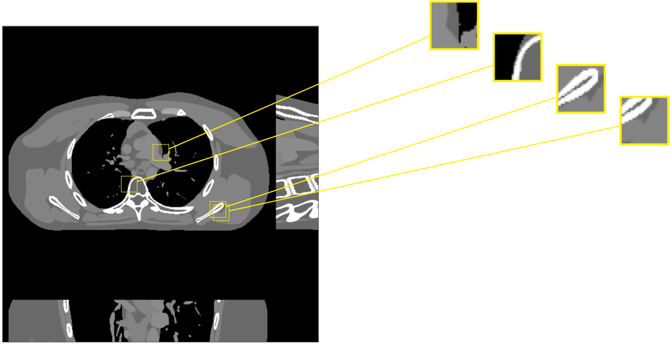
Many options...

Outline

- Data
  - Population adaptive methods
  - Patient adaptive methods
- Spatial structure
  - Patch-based models
  - Convolutional models
- Regularizer formulation
  - Synthesis (dictionary) approach
  - Analysis (sparsifying transform) approach

Assumption: if $\boldsymbol{x}$ is a plausible image, then each $\boldsymbol{\Omega} \boldsymbol{P}_m \boldsymbol{x}$ is sparse.

- $\boldsymbol{P}_m \boldsymbol{x}$ extracts the mth of M patches from $\boldsymbol{x}$
- $\boldsymbol{\Omega}$ is a square sparsifying transform matrix



Sparsifying transform learning (population adaptive)

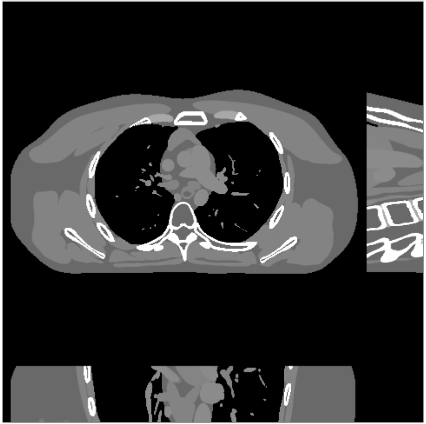
Given training images $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_L$ from a representative population, find transform $\boldsymbol{\Omega}_*$ that best sparsifies their patches:

$$\boldsymbol{\Omega}_* = \underset{\boldsymbol{\Omega} \text{ unitary}}{\arg\min} \min_{\{\boldsymbol{z}_{l,m}\}} \sum_{l=1}^{L} \sum_{m=1}^{M} \|\boldsymbol{\Omega} \boldsymbol{P}_m \boldsymbol{x}_l - \boldsymbol{z}_{l,m}\|_2^2 + \alpha \|\boldsymbol{z}_{l,m}\|_0$$

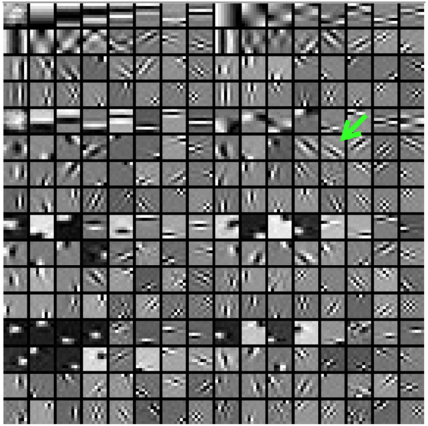
- Encourage aggregate sparsity, not patch-wise sparsity (cf K-SVD [22])
- Non-convex due to unitary constraint and $\|\cdot\|_0$
- Efficient alternating minimization algorithm [23]
  - $\boldsymbol{z}$ update is simply hard thresholding
  - $\boldsymbol{\Omega}$ update is an orthogonal Procrustes problem (SVD)
  - Subsequence convergence guarantees [23]

Example of learned sparsifying transform

3D X-ray training data



Parts of learned sparsifier $\boldsymbol{\Omega}_*$



(2D slices in x-y, x-z, y-z, from 3D image volume)

$8 \times 8 \times 8$ patches $\Longrightarrow \boldsymbol{\Omega}_*$ is $8^3 \times 8^3 = 512 \times 512$

top 8×8 slice of 256 of the 512 rows of $\boldsymbol{\Omega}_*$ $\uparrow$

Regularizer based on learned sparsifying transform

Regularized inverse problem [24]:

$$\hat{\boldsymbol{x}} = \arg\min_{\boldsymbol{x}} \|\boldsymbol{A}\boldsymbol{x} - \boldsymbol{y}\|_{\boldsymbol{W}}^2 + \beta\, \mathsf{R}(\boldsymbol{x})$$

$$\mathsf{R}(\boldsymbol{x}) = \min_{\{\boldsymbol{z}_m\}} \sum_{m=1}^{M} \|\boldsymbol{\Omega}_* \boldsymbol{P}_m \boldsymbol{x} - \boldsymbol{z}_m\|_2^2 + \alpha\, \|\boldsymbol{z}_m\|_0 \,.$$

$\boldsymbol{\Omega}_*$ adapted to population training data

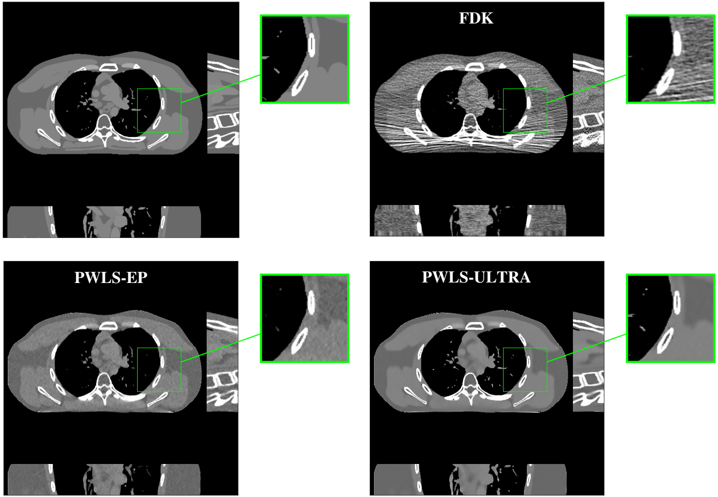
Alternating minimization optimizer:

- $\boldsymbol{z}_m$ update is simple hard thresholding
- $\boldsymbol{x}$ update is a quadratic problem: many options

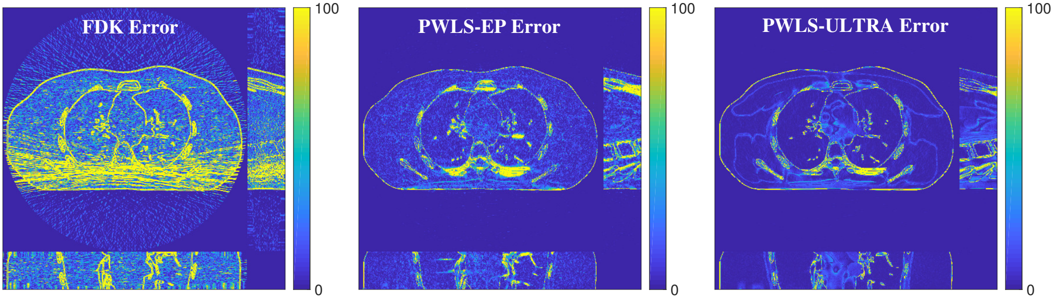
  Linearized augmented Lagrangian method (LALM) [25]

Example: low-dose 3D X-ray CT simulation

X. Zheng, S. Ravishankar, Y. Long, JF:
IEEE T-MI, June 2018 [24]



3D X-ray CT simulation Error maps



	X-ray Intensity	FDK	EP	ST Ω_*	ULTRA	ULTRA-$\{\tau_j\}$
RMSE in HU	1×10^4	67.8	34.6	32.1	30.7	**29.2**
	5×10^3	89.0	41.1	37.3	35.7	**34.2**

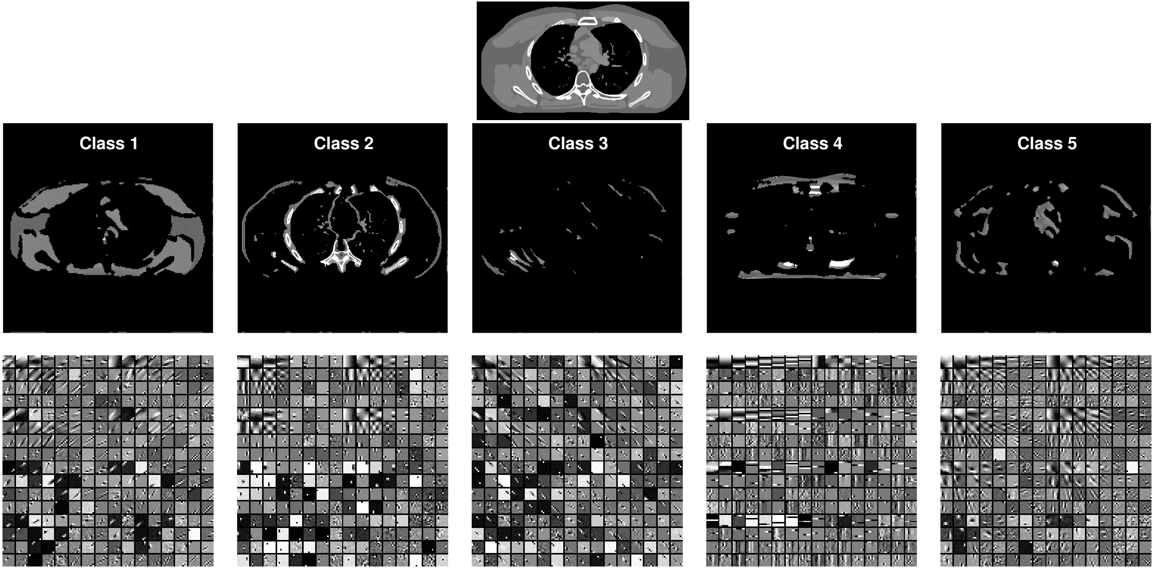
- Physics / statistics provides dramatic improvement
- Data adaptive regularization further reduces RMSE

Given training images $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_L$ from a representative population, find a set of transforms $\left\{\hat{\boldsymbol{\Omega}}_k\right\}_{k=1}^{K}$ that best sparsify image patches:

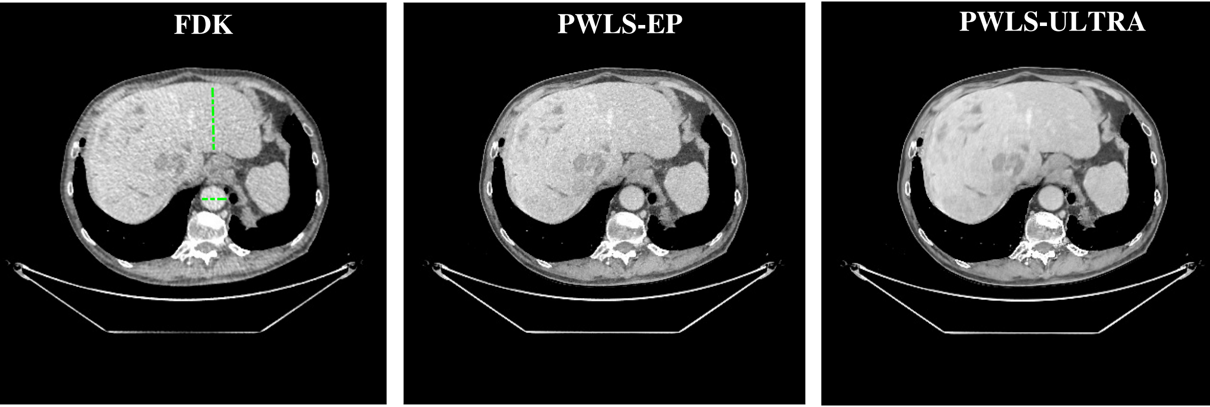
$$\left\{\hat{\boldsymbol{\Omega}}_k\right\} = \underset{\{\boldsymbol{\Omega}_k \text{ unitary}\}}{\arg\min} \min_{\left\{k_{l,m} \in \{1,\ldots,K\}\right\}} \min_{\left\{\boldsymbol{z}_{l,m}\right\}} \sum_{l=1}^{L} \sum_{m=1}^{M} \left\| \boldsymbol{\Omega}_{k_{l,m}} \boldsymbol{P}_m \boldsymbol{x}_l - \boldsymbol{z}_{l,m} \right\|_2^2 + \alpha \left\| \boldsymbol{z}_{l,m} \right\|_0$$

- Joint unsupervised clustering / sparsification
- Further nonconvexity due to clustering
- Efficient alternating minimization algorithm [26]

Example: 3D X-ray CT learned set of transforms



Example: 3D X-ray CT ULTRA for chest scan



Zheng et al., IEEE T-MI, June 2018 [24]

- Data
  - Population adaptive methods
  - Patient adaptive methods
- Spatial structure
  - Patch-based models
  - Convolutional models
- Regularizer formulation
  - Synthesis (dictionary) approach
  - Analysis (sparsifying transform) approach

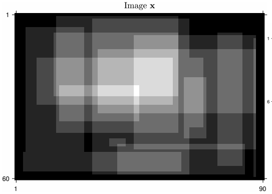
Drawback of basic patch-based methods:
$512 \times 512 \times 512$ 3D X-ray CT image volume
$8 \times 8 \times 8$ patches
$\Longrightarrow 512^3 \cdot 8^3 \cdot 4 = 256$ Gbyte of patch data for stride=1

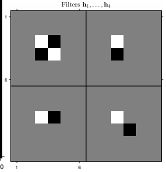
Convolutional sparsity model

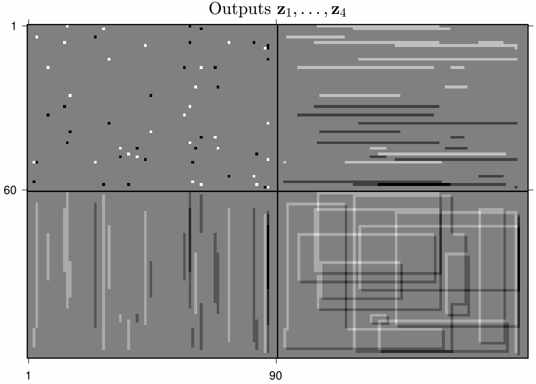
Assumption: There is a set of filters $\{\boldsymbol{h}_k\}_{k=1}^K$
such that the images $\{\boldsymbol{h}_k * \boldsymbol{x}\}$ are sparse for a plausible image $\boldsymbol{x}$.

- For more plausible images, $\{\boldsymbol{h}_k * \boldsymbol{x}\}$ is more sparse.
- $*$ denotes convolution
- Inherently shift invariant and no patches

Example (hand crafted filters):







Sparsifying filter learning (population adaptive)

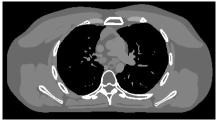
Given training images $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_L$ from a representative population, find filters $\left\{\hat{\boldsymbol{h}}_k\right\}_{k=1}^K$ that best sparsify them:

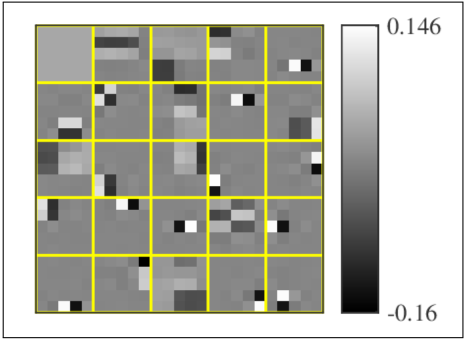
$$\left\{\hat{\boldsymbol{h}}_k\right\} = \underset{\{\boldsymbol{h}_k\}\in\mathcal{H}}{\arg\min} \min_{\{\boldsymbol{z}_{l,k}\}} \sum_{l=1}^{L} \sum_{k=1}^{K} \|\boldsymbol{h}_k * \boldsymbol{x}_l - \boldsymbol{z}_{l,k}\|_2^2 + \alpha \|\boldsymbol{z}_{l,k}\|_0$$

- To encourage filter diversity:
  - $\mathcal{H} = \{\boldsymbol{H} : \boldsymbol{H}\boldsymbol{H}' = \boldsymbol{I}\}, \ \boldsymbol{H} = [\boldsymbol{h}_1 \ \ldots \ \boldsymbol{h}_K]$
  - *cf.* tight-frame condition $\sum_{k=1}^K \|\boldsymbol{h}_k * \boldsymbol{x}\|_2^2 \propto \|\boldsymbol{x}\|_2^2$
- Encourage aggregate sparsity, period
- Non-convex due to constraint $\mathcal{H}$ and $\|\cdot\|_0$
- Efficient alternating minimization algorithm [27]
  - $\boldsymbol{z}$ update is simply hard thresholding
  - Filter update uses diagonal majorizer, proximal map (SVD)
  - Subsequence convergence guarantees [27]

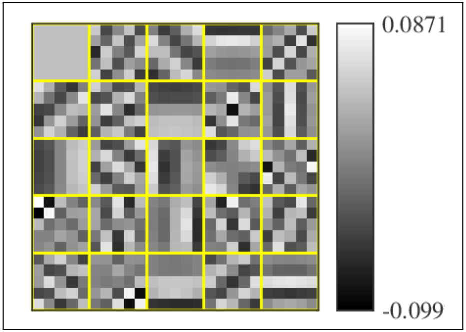
Examples of learned sparsifying filters

2D X-ray CT training data and learned 5×5 sparsifying filters $\left\{ \hat{h}_k \right\}$ [27]:





$\alpha = 10^{-4}$



$\alpha = 2 \times 10^{-3}$

Regularized inverse problem [27]:

$$\hat{\boldsymbol{x}} = \arg\min_{\boldsymbol{x} \succeq \boldsymbol{0}} \|\boldsymbol{A}\boldsymbol{x} - \boldsymbol{y}\|_{\boldsymbol{W}}^2 + \beta\, \mathsf{R}(\boldsymbol{x})$$

$$\mathsf{R}(\boldsymbol{x}) = \arg\min_{\{\boldsymbol{z}_k\}} \sum_{k=1}^{K} \left\| \hat{\boldsymbol{h}}_k * \boldsymbol{x} - \boldsymbol{z}_k \right\|_2^2 + \alpha \|\boldsymbol{z}_k\|_0 .$$

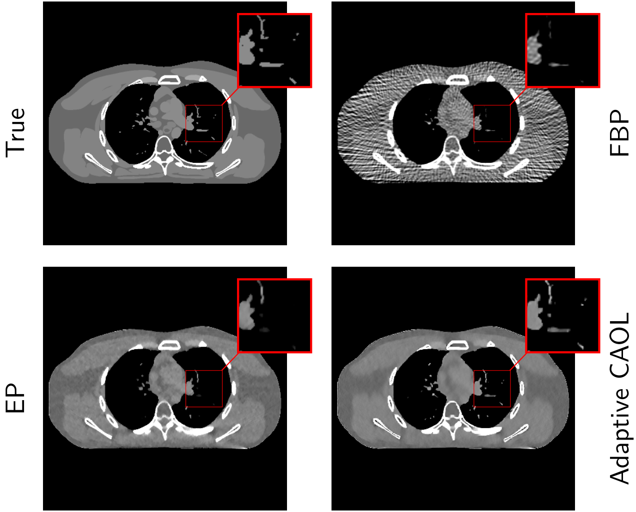
$\left\{ \hat{\boldsymbol{h}}_k \right\}$ adapted to population training data

Block proximal gradient with majorizer (BPG-M) optimizer:

- $\boldsymbol{z}_k$ update is simple hard thresholding
- $\boldsymbol{x}$ update is a quadratic problem: diagonal majorizer

I. Y. Chun, JF, 2018, arXiv 1802.05584 [27]

Example: sparse-view 2D X-ray CT simulation



Quantitative results

123 views (out of usual 984) $\Longrightarrow$ $8\times$ dose reduction

RMSE (in HU):

FBP	82.8
EP	40.8
Adaptive filters	**35.2**

- Physics / statistics provides dramatic improvement
- Data-adaptive regularization further reduces RMSE

Extension to multiple layers (cf CNN) I

Convolutional sparsity model: $\boldsymbol{h}_k * \boldsymbol{x}$ is sparse for $k = 1, \ldots, K_1$
Learning 1 "layer" of filters:

$$\{\hat{\boldsymbol{h}}_k^{[1]}\} = \underset{\{\boldsymbol{h}_k^{[1]}\} \in \mathcal{H}}{\arg\min} \min_{\{\boldsymbol{z}_{l,k}^{[1]}\}} \sum_{l=1}^{L} \sum_{k=1}^{K_1} \left\| \boldsymbol{h}_k^{[1]} * \boldsymbol{x}_l - \boldsymbol{z}_{l,k}^{[1]} \right\|_2^2 + \alpha \left\| \boldsymbol{z}_{l,k}^{[1]} \right\|_0$$

Learning 2 layers of filters [27]:

$$\left(\{\hat{\boldsymbol{h}}_k^{[1]}\}, \{\hat{\boldsymbol{h}}_k^{[2]}\} \right) = \mathop{\arg\min}_{\{\boldsymbol{h}_k^{[1]}\},\{\boldsymbol{h}_k^{[2]}\} \in \mathcal{H}} \min_{\{\boldsymbol{z}_{l,k}^{[1]}\}} \min_{\{\boldsymbol{z}_{l,k}^{[2]}\}} \sum_{l=1}^{L} \sum_{k=1}^{K_1} \left\| \boldsymbol{h}_k^{[1]} * \boldsymbol{x}_l - \boldsymbol{z}_{l,k}^{[1]} \right\|_2^2 + \alpha \left\| \boldsymbol{z}_{l,k}^{[1]} \right\|_0 + \sum_{l=1}^{L} \sum_{k=1}^{K_2} \left\| \boldsymbol{h}_k^{[2]} * \left(\boldsymbol{P}_k \boldsymbol{z}_l^{[1]} \right) - \boldsymbol{z}_{l,k}^{[2]} \right\|_2^2 + \alpha \left\| \boldsymbol{z}_{l,k}^{[2]} \right\|_0$$

Here $\boldsymbol{P}_k$ is a pooling operator for the output of first layer

Block proximal gradient with majorizer (BPG-M) optimizer

I. Y. Chun, JF, 2018, arXiv 1802.05584 [27]

Use multi-level learned filters as (interpretable?) regularizer for CT.

Outline

MR with adapted dictionary

- Data
  - Population adaptive methods
  - Patient adaptive methods
- Spatial structure
  - Patch-based models
  - Convolutional models
- Regularizer formulation
  - Synthesis (dictionary) approach
  - Analysis (sparsifying transform) approach

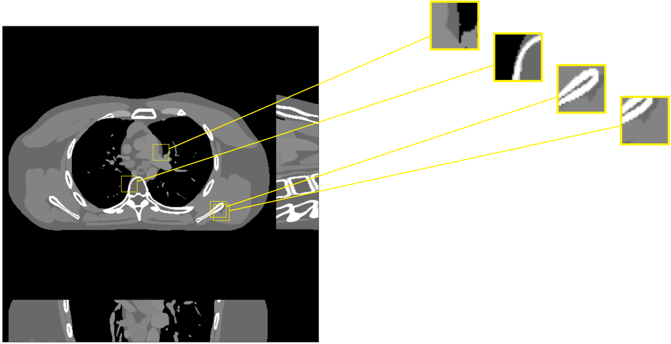
Patch-wise dictionary sparsity model

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

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

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

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



MR reconstruction using adaptive dictionary regularizer

Dictionary-blind MR image reconstruction:

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

$$\mathsf{R}(\boldsymbol{x}) = \min_{\boldsymbol{D} \in \mathcal{D}} \min_{\boldsymbol{Z}' \in \mathcal{C}} \sum_{m=1}^{M} \left(\|\boldsymbol{P}_m \boldsymbol{x} - \boldsymbol{D} \boldsymbol{z}_m\|_2^2 + \lambda^2 \|\boldsymbol{z}_m\|_0 \right)$$

where $\boldsymbol{P}_m$ extracts mth of M image patches.

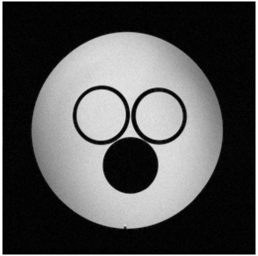
In words: of the many images...

Alternating (nested) minimization:

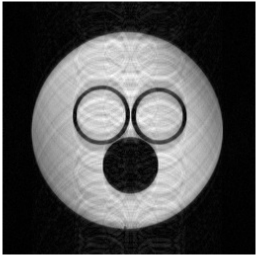
- Fixing $\boldsymbol{x}$ and $\boldsymbol{D}$, update each row of $\boldsymbol{Z} = [\boldsymbol{z}_1 \; \ldots \; \boldsymbol{z}_M]$ sequentially via hard-thresholding.
- Fixing $\boldsymbol{x}$ and $\boldsymbol{Z}$, update $\boldsymbol{D}$ using SOUP-DIL [28].
- Fixing $\boldsymbol{Z}$ and $\boldsymbol{D}$, 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, but monotone decreasing and some convergence theory [28].

2D CS MRI results I

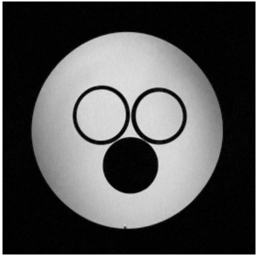
Fully Sampled



Zero-Filled



SOUP-DILLO-MRI

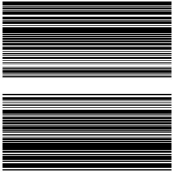


6×6 patches

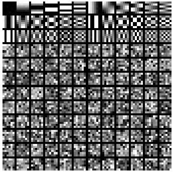
$\boldsymbol{D} \in \mathbb{C}^{6^2 \times 144}$

[28]

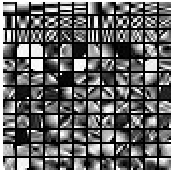
Sampling $(2.5\times)$



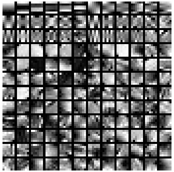
Initial $\boldsymbol{D}$



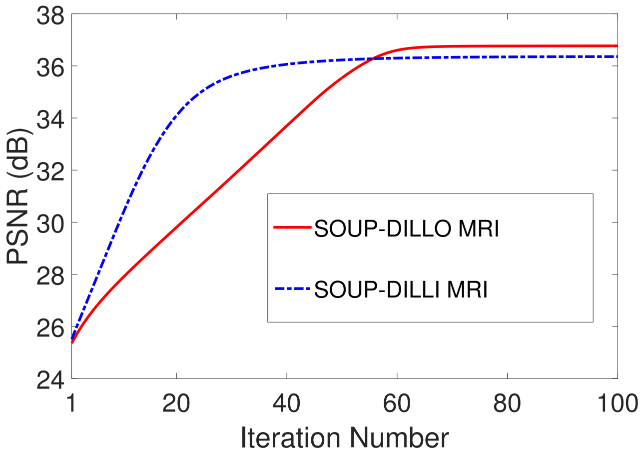
Learned real$\{\boldsymbol{D}\}$



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

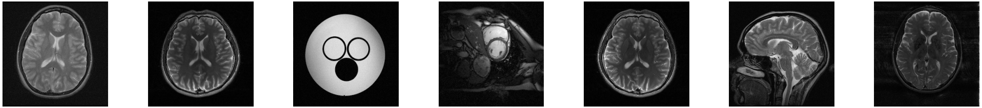


2D CS MRI results II



(SNR compared to fully sampled image.)
Using $\|\boldsymbol{z}_m\|_0$ leads to higher SNR than $\|\boldsymbol{z}_m\|_1$.
Adaptive case is non-convex anyway...

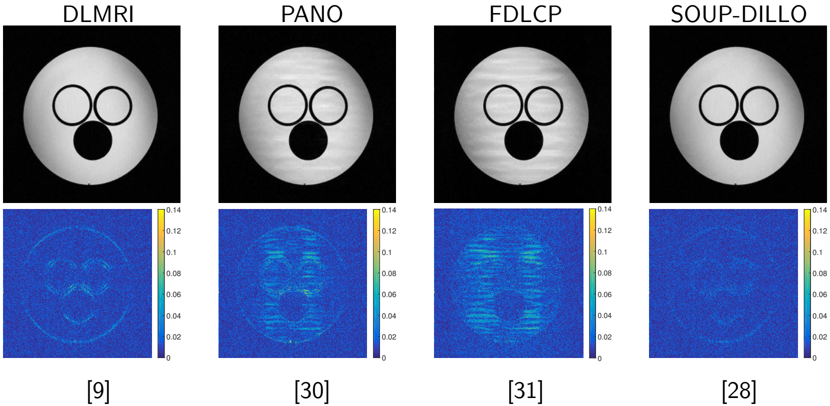
J. Fessler
ML for IR



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

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.				[29]	[30]	[9]	[28]	[28]

2D CS MRI results IV



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

Summary of patch-based, data-driven adaptive regularizers

Use training data to learn:

- dictionary $\boldsymbol{D}$ (for patches)
- sparsifying transform(s) $\boldsymbol{\Omega}$ (for patches)
- or convolutional versions thereof [32, 33]

ML-based regularized optimization problem using M image patches:

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

$$R_{\mathrm{ML-DL}}(\boldsymbol{x}) = \min_{\{\boldsymbol{z}_m\}} \sum_{m=1}^{M} \|\boldsymbol{P}_m\boldsymbol{x} - \boldsymbol{D}\boldsymbol{z}_m\|_2^2 + \alpha \|\boldsymbol{z}_m\|_0$$

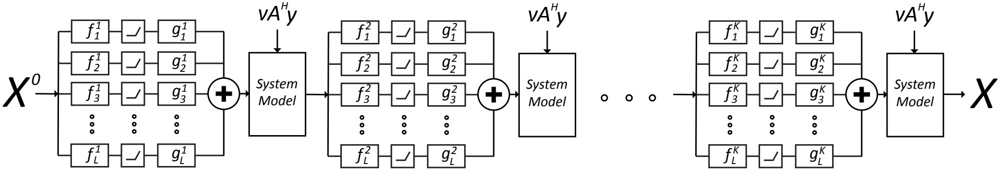
$$R_{\mathrm{ML-ST}}(\boldsymbol{x}) = \min_{\{\boldsymbol{z}_m\}} \sum_{m=1}^{M} \|\boldsymbol{\Omega}\boldsymbol{P}_m\boldsymbol{x} - \boldsymbol{z}_m\|_2^2 + \alpha \|\boldsymbol{z}_m\|_0$$

Alternative: blind adaptive learned dictionary [9] or learned sparsifying transform [34].
Double minimization (so very "deep?") More interpretable than CNNs?

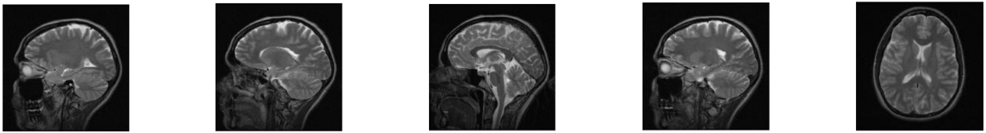
Training an unrolled loop I

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

Ravishankar et al., ISBI 2018 [15]



Tested with 5 different MR images:



Results:

Undersampling	Image	Zero-filled	Sparse MRI	UTMRI	Unrolled
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**

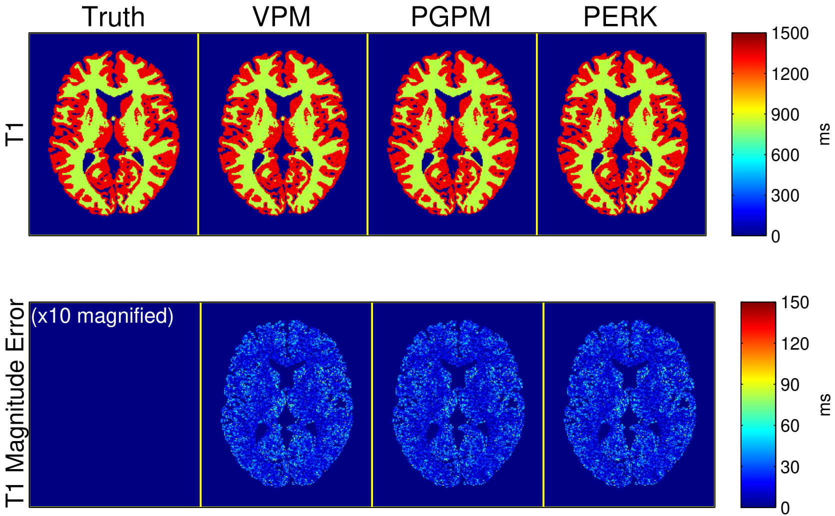
Sparse MRI [35] total variation (TV) and wavelets
UTMRI [26] (union of learned sparsifying transforms): adaptive, not "deep"

Shallow machine learning for qMRI

Quantitative MRI: images $\rightarrow$ $\boxed{\text{estimation}}$ $\rightarrow$ parameters (T1,T2,...)

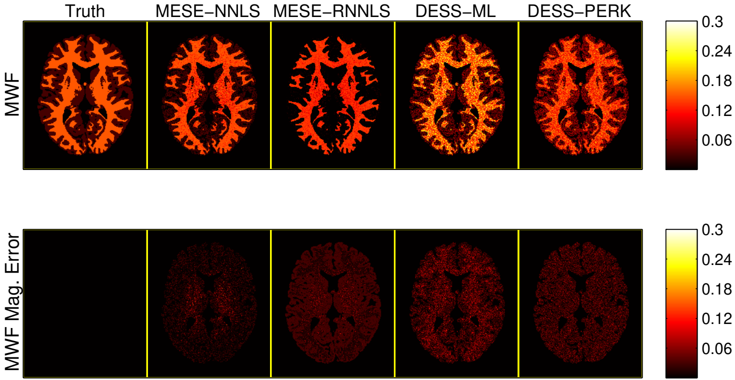
- Traditional nonlinear estimation methods:
  - nonlinear least squares
  - dictionary matching (quantized maximum likelihood via variable projection)
- Machine-learning methods
  - deep neural network regression [36–39]
    Requires long training times
  - parameter estimation via kernel regression (PERK)
    Gopal Nataraj et al., ISBI 2017, IEEE T-MI 2018 [40, 41]

Parameter estimation via kernel regression (PERK) example



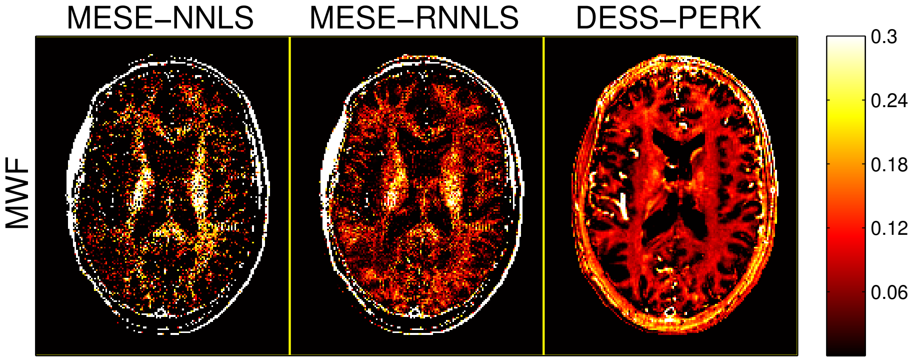
PERK applied to myelin water imaging

6 parameters (T1 slow/fast, T2 slow/fast, M_0, fast fraction)
Estimated from 3 optimized dual-echo steady state (DESS) scans [42]



PERK training: 33.8s, testing 0.99s / slice

PERK in vivo myelin water fraction results



MESE scan took 32m (16m $\times$2)
DESS scan took 3m15s
Take away: "traditional" machine learning is still useful...

Outline

Summary

- Machine learning has great potential for medical imaging
- Much excitement but many challenges
- Image reconstruction seems especially suitable for ML ideas
- Data-driven, adaptive regularizers beneficial for low-dose CT and under-sampled MRI
- More comparisons between model-based methods with adaptive regularizers and CNN-based methods needed
- Machine learning tools like kernel regression remain useful

Recommended reading (incomplete lists)

- Overviews: [43, 45, 46]
- Generative models: [20, 47]:
- Deep learning myths [48]
- NN complexity analysis / function approximation [49–51] [52]
- Application to MR fingerprinting [36, 39]
- MR reconstruction / enhancement using CNN [16, 53–60]
- Dynamic MR reconstruction using CNN [61]
- ...

Resources

Talk and code available online at
`http://web.eecs.umich.edu/~fessler`

[1] H. Greenspan, B. van Ginneken, and R. M. Summers. "Guest editorial deep learning in medical imaging: overview and future promise of an exciting new technique." In: *IEEE Trans. Med. Imag.* 35.5 (May 2016), 1153–9.

[2] G. Wang, J. C. Ye, K. Mueller, and J. A. Fessler. "Image reconstruction is a new frontier of machine learning." In: *IEEE Trans. Med. Imag.* 37.6 (June 2018), 1289–96.

[3] E. Haneda, B. Claus, P. FitzGerald, G. Wang, and B. De Man. "CT sinogram analysis using deep learning." In: *Proc. 5th Intl. Mtg. on Image Formation in X-ray CT.* 2018, 419–22.

[4] L. Baldassarre, Y-H. Li, J. Scarlett, B. Gozcu, I. Bogunovic, and V. Cevher. "Learning-based compressive subsampling." In: *IEEE J. Sel. Top. Sig. Proc.* 10.4 (June 2016), 809–22.

[5] B. Gozcu, R. K. Mahabadi, Y-H. Li, E. Ilicak, T. Cukur, J. Scarlett, and V. Cevher. "Learning-based compressive MRI." In: *IEEE Trans. Med. Imag.* 37.6 (June 2018), 1394–406.

[6] Y. Cao and D. N. Levin. "Feature-recognizing MRI." In: *Mag. Res. Med.* 30.3 (Sept. 1993), 305–17.

[7] Y. Cao, D. N. Levin, and L. Yao. "Locally focused MRI." In: *Mag. Res. Med.* 34.6 (Dec. 1995), 858–67.

[8] Y. Cao and D. N. Levin. "Using an image database to constrain the acquisition and reconstruction of MR images of the human head." In: *IEEE Trans. Med. Imag.* 14.2 (June 1995), 350–61.

[9] 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.

[10] 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.

[11] G. Yang, S. Yu, H. Dong, G. Slabaugh, P. L. Dragotti, X. Ye, F. Liu, S. Arridge, J. Keegan, Y. Guo, and D. Firmin. "DAGAN: Deep de-aliasing generative adversarial networks for fast compressed sensing MRI reconstruction." In: *IEEE Trans. Med. Imag.* 37.6 (June 2018), 1310–21.

[12] G-H. Chen, J. Tang, and S. Leng. "Prior image constrained compressed sensing (PICCS): A method to accurately reconstruct dynamic CT images from highly undersampled projection data sets." In: *Med. Phys.* 35.2 (Feb. 2008), 660–3.

[13] K. Gregor and Y. LeCun. "Learning fast approximations of sparse coding." In: *Proc. Intl. Conf. Mach. Learn.* 2010.

[14] 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.

[15] S. Ravishankar, A. Lahiri, C. Blocker, and J. A. Fessler. "Deep dictionary-transform learning for image reconstruction." In: *Proc. IEEE Intl. Symp. Biomed. Imag.* 2018, 1208–12.

[16] K. Hammernik, T. Klatzer, E. Kobler, M. P. Recht, D. K. Sodickson, T. Pock, and F. Knoll. "Learning a variational network for reconstruction of accelerated MRI data." In: *Mag. Res. Med.* 79.6 (June 2018), 3055–71.

[17] I. J. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio. *Generative adversarial networks*. 2014.

[18] X. Chen, Y. Duan, R. Houthooft, J. Schulman, I. Sutskever, and P. Abbeel. "InfoGAN: interpretable representation learning by information maximizing generative adversarial nets." In: *Neural Info. Proc. Sys.* 2016, 2172–80.

[19] A. Bora, A. Jalal, E. Price, and A. G. Dimakis. "Compressed sensing using generative models." In: *Proc. Intl. Conf. Mach. Learn.* Vol. 70. 2017, 537–46.

[20] S. Kolouri, P. E. Pope, C. E. Martin, and G. K. Rohde. *Sliced-Wasserstein autoencoder: an embarrassingly simple generative model*. 2018.

[21] D. Berthelot, T. Schumm, and L. Metz. *BEGAN: boundary equilibrium generative adversarial networks*. 2017.

[22] 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.

[23] S. Ravishankar and Y. Bresler. "l_0 sparsifying transform learning with efficient optimal updates and convergence guarantees." In: *IEEE Trans. Sig. Proc.* 63.9 (May 2015), 2389–404.

[24] X. Zheng, S. Ravishankar, Y. Long, and J. A. Fessler. “PWLS-ULTRA: An efficient clustering and learning-based approach for low-dose 3D CT image reconstruction.” In: *IEEE Trans. Med. Imag.* 37.6 (June 2018), 1498–510.

[25] H. Nien and J. A. Fessler. “Relaxed linearized algorithms for faster X-ray CT image reconstruction.” In: *IEEE Trans. Med. Imag.* 35.4 (Apr. 2016), 1090–8.

[26] 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.

[27] I. Y. Chun and J. A. Fessler. “Convolutional analysis operator learning: acceleration, convergence, application, and neural networks.” In: *IEEE Trans. Im. Proc.* (2018). Submitted.

[28] 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.

[29] 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.

[30] X. Qu, Y. Hou, F. Lam, D. Guo, J. Zhong, and Z. Chen. “Magnetic resonance image reconstruction from undersampled measurements using a patch-based nonlocal operator.” In: *Med. Im. Anal.* 18.6 (Aug. 2014), 843–56.

[31] Z. Zhan, J-F. Cai, D. Guo, Y. Liu, Z. Chen, and X. Qu. “Fast multiclass dictionaries learning with geometrical directions in MRI reconstruction.” In: *IEEE Trans. Biomed. Engin.* 63.9 (Sept. 2016), 1850–61.

[32] I. Y. Chun and J. A. Fessler. “Convolutional dictionary learning: acceleration and convergence.” In: *IEEE Trans. Im. Proc.* 27.4 (Apr. 2018), 1697–712.

[33] I. Y. Chun and J. A. Fessler. *Convolutional analysis operator learning: acceleration, convergence, application, and neural networks.* 2018.

[34] S. Ravishankar and Y. Bresler. “Efficient blind compressed sensing using sparsifying transforms with convergence guarantees and application to MRI.” In: *SIAM J. Imaging Sci.* 8.4 (2015), 2519–57.

[35] 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.

[36] P. Virtue, S. X. Yu, and M. Lustig. "Better than real: Complex-valued neural nets for MRI fingerprinting." In: *Proc. IEEE Intl. Conf. on Image Processing*. 2017, 3953–7.

[37] A. Lahiri, J. A. Fessler, and L. Hernandez-Garcia. "Optimized design of MRF scan parameters for ASL signal acquisition." In: *ISMRM Workshop on MR Fingerprinting*. 2017.

[38] A. Lahiri, J. A. Fessler, and L. Hernandez-Garcia. "Optimized scan design for ASL fingerprinting and multiparametric estimation using neural network regression." In: *Proc. Intl. Soc. Mag. Res. Med.* 2018, p. 309.

[39] O. Cohen, B. Zhu, and M. S. Rosen. "MR fingerprinting Deep RecOnstruction NEtwork (DRONE)." In: *Mag. Res. Med.* 80.3 (Sept. 2018), 885–94.

[40] G. Nataraj, J-F. Nielsen, and J. A. Fessler. "Dictionary-free MRI parameter estimation via kernel ridge regression." In: *Proc. IEEE Intl. Symp. Biomed. Imag.* 2017, 5–9.

[41] G. Nataraj, J-F. Nielsen, C. D. Scott, and J. A. Fessler. "Dictionary-free MRI PERK: Parameter estimation via regression with kernels." In: *IEEE Trans. Med. Imag.* 37.9 (Sept. 2018), 2103–14.

[42] G. Nataraj, J-F. Nielsen, M. Gao, and J. A. Fessler. *Fast, precise myelin water quantification using DESS MRI and kernel learning*. Submitted. 2018.

[43] G. Wang, M. Kalra, and C. G. Orton. "Machine learning will transform radiology significantly within the next five years." In: *Med. Phys.* 44.6 (June 2017), 2041–4.

[44] A. Esteva, B. Kuprel, R. A. Novoa, J. Ko, S. M. Swetter, H. M. Blau, and S. Thrun. "Dermatologist-level classification of skin cancer with deep neural networks." In: *Nature* 542.7639 (2017), 115–8.

[45] G. Wang. "A perspective on deep imaging." In: *IEEE Access* 4 (Nov. 2016), 8914–24.

[46] M. T. McCann, K. H. Jin, and M. Unser. "Convolutional neural networks for inverse problems in imaging: A review." In: *IEEE Sig. Proc. Mag.* 34.6 (Nov. 2017), 85–95.

[47] I. Deshpande, Z. Zhang, and A. Schwing. "Generative modeling using the sliced Wasserstein distance." In: *Proc. IEEE Conf. on Comp. Vision and Pattern Recognition*. 2018.

[48] S. Rakhlin. *MythBusters: A Deep Learning Edition*. Slides dated Jan 18-19, 2018. 2018.

[49] N. Golowich, A. Rakhlin, and O. Shamir. *Size-independent sample complexity of neural networks*. 2017.

[50] T. Liang, T. Poggio, A. Rakhlin, and J. Stokes. *Fisher-Rao metric, geometry, and complexity of neural networks*. 2017.

[51] M. Raghu, B. Poole, J. Kleinberg, S. Ganguli, and J. Sohl-Dickstein. "On the expressive power of deep neural networks." In: *Proc. Intl. Conf. Mach. Learn.* Vol. 70. 2017, 2847–54.

[52] S. Liang and R. Srikant. "Why deep neural networks for function approximation?" In: *Proc. Intl. Conf. on Learning Representations*. 2017.

[53] 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.

[54] M. Mardani, E. Gong, J. Y. Cheng, S. Vasanawala, G. Zaharchuk, M. Alley, N. Thakur, S. Han, W. Dally, J. M. Pauly, and L. Xing. *Deep generative adversarial networks for compressed sensing automates MRI*. 2017.

[55] B. Zhu, J. Z. Liu, S. F. Cauley, B. R. Rosen, and M. S. Rosen. "Image reconstruction by domain-transform manifold learning." In: *Nature* 555 (Mar. 2018), 487–92.

[56] Y. Han, J. Yoo, H. H. Kim, H. J. Shin, K. Sung, and J. C. Ye. "Deep learning with domain adaptation for accelerated projection-reconstruction MR." In: *Mag. Res. Med.* 80.3 (Sept. 2018), 1189–205.

[57] K. H. Jin and M. Unser. "3D BPConvNet to reconstruct parallel MRI." In: *Proc. IEEE Intl. Symp. Biomed. Imag.* 2018, 361–4.

[58] H. Jeelani, J. Martin, F. Vasquez, M. Salerno, and D. S. Weller. "Image quality affects deep learning reconstruction of MRI." In: *Proc. IEEE Intl. Symp. Biomed. Imag.* 2018, 357–60.

[59] 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.

[60] T. Eo, Y. Jun, T. Kim, J. Jang, H-J. Lee, and D. Hwang. "KIKI-net: cross-domain convolutional neural networks for reconstructing undersampled magnetic resonance images." In: *Mag. Res. Med.* (2018).

[61] J. Schlemper, J. Caballero, J. V. Hajnal, A. N. Price, and D. Rueckert. "A deep cascade of convolutional neural networks for dynamic MR image reconstruction." In: *IEEE Trans. Med. Imag.* 37.2 (Feb. 2018), 491–503.