Data-driven methods for image reconstruction

Jeffrey A. Fessler

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

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

Emory University Scientific Computing Seminar
2020-11-06

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

Declaration: No relevant financial interests or relationships to disclose

Outline

Outline

Medical imaging overview

Medical imaging overview

Most obvious place for machine learning is in post-processing (image analysis).
Numerous special issues and surveys in medical imaging journals, *e.g.*, [1–9].

Machine learning for scan design

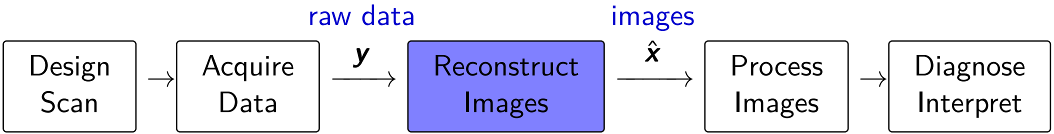
Choose best k-space phase encoding locations based on training images

Hot topic in MRI recently [10–15].

Precursor by Yue Cao and David Levin, MRM Sep. 1993 [16–18].

Medical imaging overview

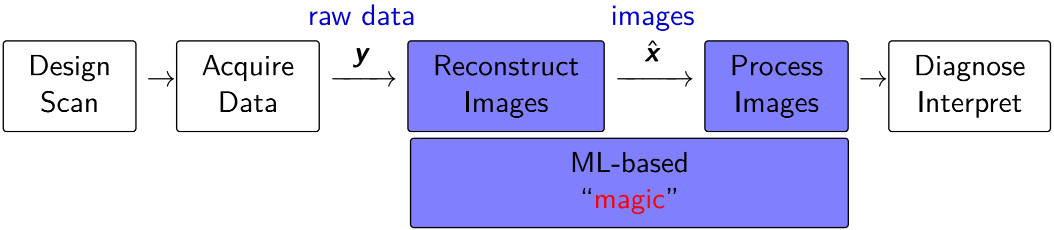
J. Fessler



Machine learning in medical image reconstruction
June 2018 special issue of IEEE Trans. on Medical Imaging [19].
Surveys: [20–27]
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.

Medical imaging overview



A holy grail for machine learning in medical imaging?

- CT sinogram to vessel diameter [28, 29]
- k-space to ???

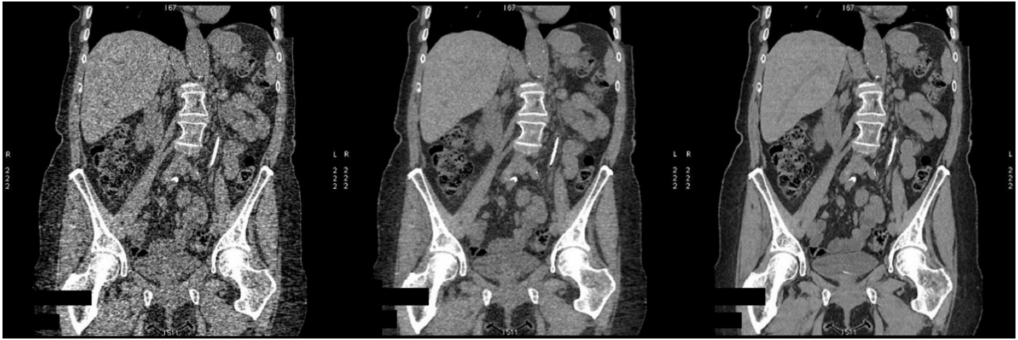
Generations of medical image reconstruction methods

J. Fessler

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 based on imaging physics ("model based")
   - 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, ...

Two important milestones for clinical CT

- Model-based image reconstruction (MBIR) FDA approved circa 2012 [30]



- Deep-learning image reconstruction FDA approved 2019 [31, 32]

Outline

Overview:

- image-domain learning [33–35]...
- k-space or data-domain learning
  e.g., [36], [37], [38]
- transform learning (direct from k-space to image)
  e.g., AUTOMAP [39], [40–42]
- hybrid-domain learning (unrolled loop, *e.g.*, variational network)
  alternate between denoising/dealiasing and reconstruction from k-space
  e.g., [37, 43–47] ...

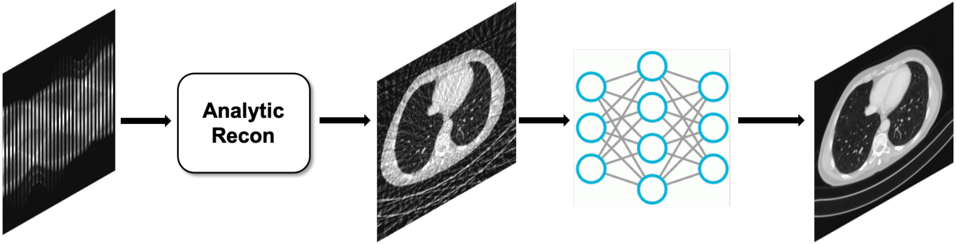


Figure courtesy of Jong Chul Ye, KAIST University.

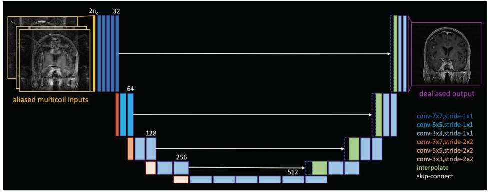
+ simple and fast
− aliasing is spatially widespread, requires deep network

Dangers of image-domain learning: Method

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



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

Dangers of image-domain learning: Result

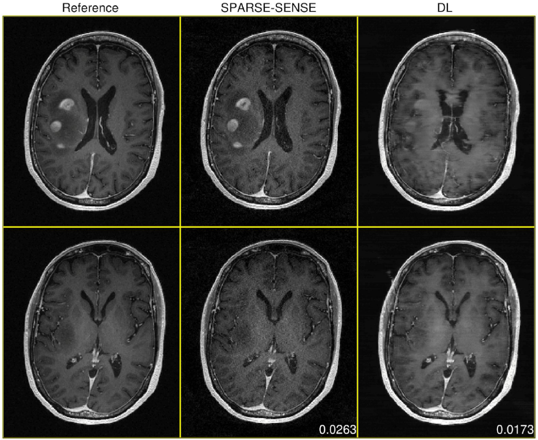


Figure 3: Reconstructions in a case of anaplastic astrocytoma, a rare malignant brain tumor. SPARSE-SENSE and DL reconstructions are from the same 4x-accelerated retrospectively undersampled acquisition. DL achieves lower whole-volume MAE than SPARSE-SENSE, but fails to properly reconstruct regions near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

Image-domain learning variations

- Use NN output as a "prior" for iterative reconstruction [33, 49]:

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

  - 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 $\boldsymbol{x}_{\text{NN}}$
  - Iterations needed for parallel MRI and/or non-Cartesian sampling (PCG)

- Learn residual (aliasing artifacts), then subtract [50, 51]

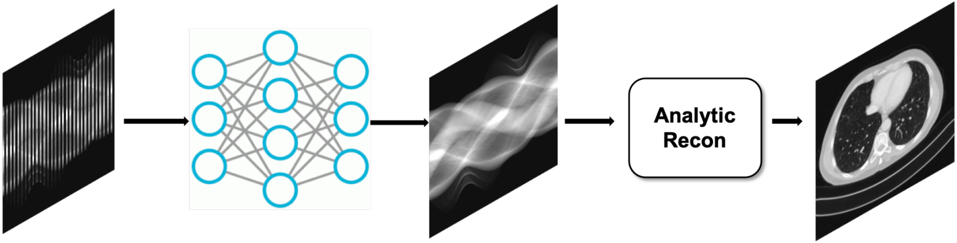


Figure courtesy of Jong Chul Ye, KAIST University.

+ simple and fast (“nonlinear GRAPPA”)
+ “database-free” : learn from auto-calibration data
− 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

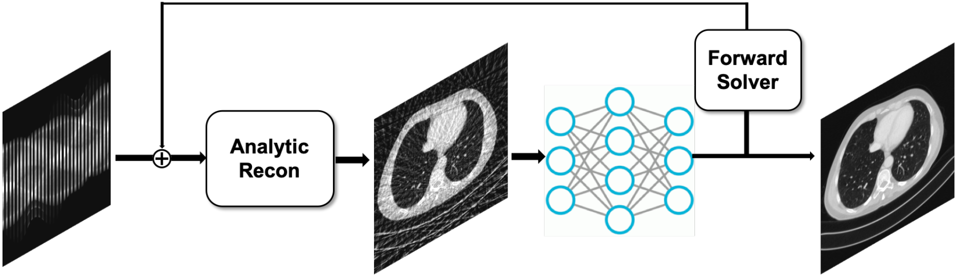


Figure courtesy of Jong Chul Ye, KAIST University.

+ physics-based use of k-space data & image-domain priors
+ interpretable connections to optimization approaches
− more computation to due to "iterations" (layers) and repeated $\boldsymbol{A}\boldsymbol{x}$, $\boldsymbol{A}'\boldsymbol{r}$

Outline

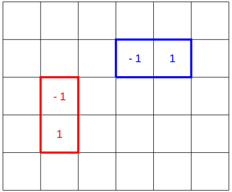
- Data
  - Population adaptive methods (*e.g.*, X-ray CT)
  - Patient adaptive methods (*e.g.*, dynamic MRI?)
- Spatial structure
  - Patch-based models
  - Convolutional models
- Regularizer formulation
  - Synthesis (dictionary) approach
  - Analysis (sparsifying transforms) approach

Many options...

Outline

Patch-based regularization and TV

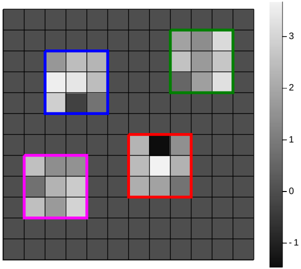
Anisotropic discrete TV regularizer:
$R(\boldsymbol{x}) = \|\boldsymbol{T}\boldsymbol{x}\|_1$
where $\boldsymbol{T}$ is finite-differences
$\equiv$ patches of size 2×1.



Larger patches provide more context for distinguishing signal from noise.

cf. CNN approaches

Patch-based regularizers:
- synthesis models
- analysis methods



X-ray CT with learned sparsifying transforms

J. Fessler

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

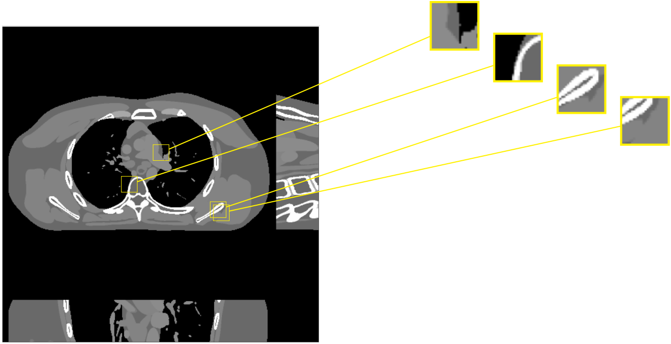
Patch-wise transform sparsity model

J. Fessler

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

- $\boldsymbol{P}_m\boldsymbol{x}$ extracts the mth of M patches from $\boldsymbol{x}$
- $\boldsymbol{T}$ is a (often square) sparsifying transform matrix.

What $\boldsymbol{T}$?



Sparsifying transform learning (population adaptive)

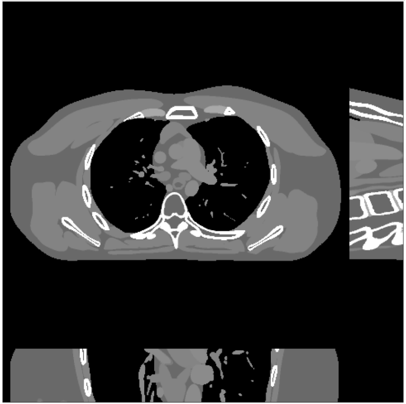
Given training images $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_L$ from a representative population, find transform $\boldsymbol{T}_*$ that best sparsifies their patches:

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

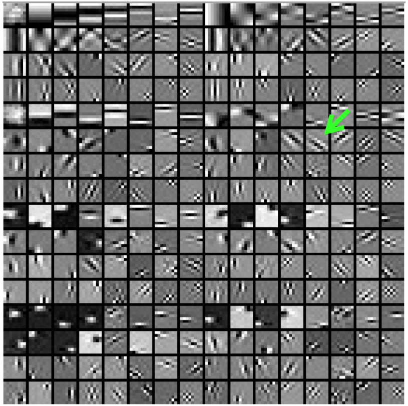
- Encourage aggregate sparsity, not patch-wise sparsity (cf K-SVD [52])
- Non-convex due to unitary constraint and $\|\cdot\|_0$
- Efficient alternating minimization algorithm [53]
  - $\boldsymbol{z}$ update : simple hard thresholding
  - $\boldsymbol{T}$ update : orthogonal Procrustes problem (SVD)
  - Subsequence convergence guarantees [53]

Example of learned sparsifying transform

3D X-ray training data (XCAT phantom)



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



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

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

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

Regularizer based on learned sparsifying transform

Regularized inverse problem [54]:

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

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

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

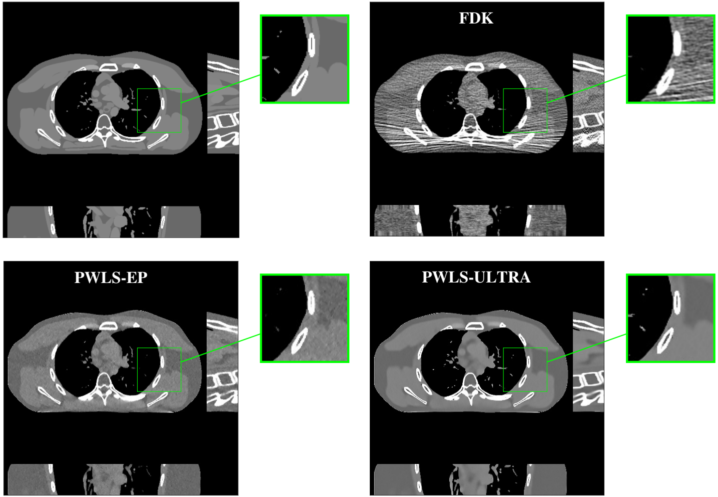
Alternating minimization optimizer:

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

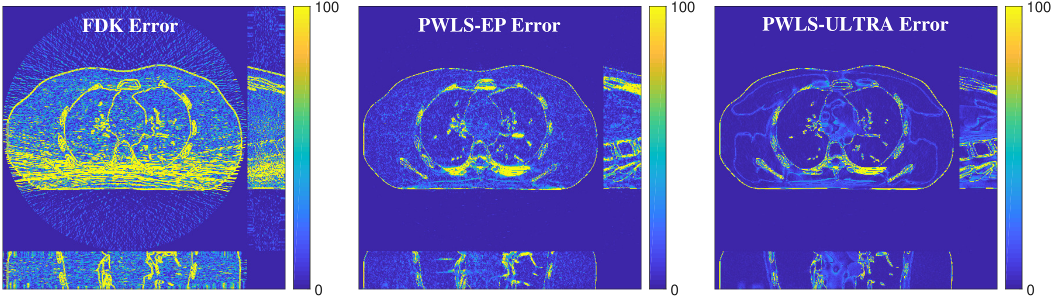
  Linearized augmented Lagrangian method (LALM) [55]

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

X. Zheng, S. Ravishankar, Y. Long, JF: IEEE T-MI, June 2018 [54].



3D X-ray CT simulation Error maps



	X-ray Intensity	FDK	EP	ST T_*	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

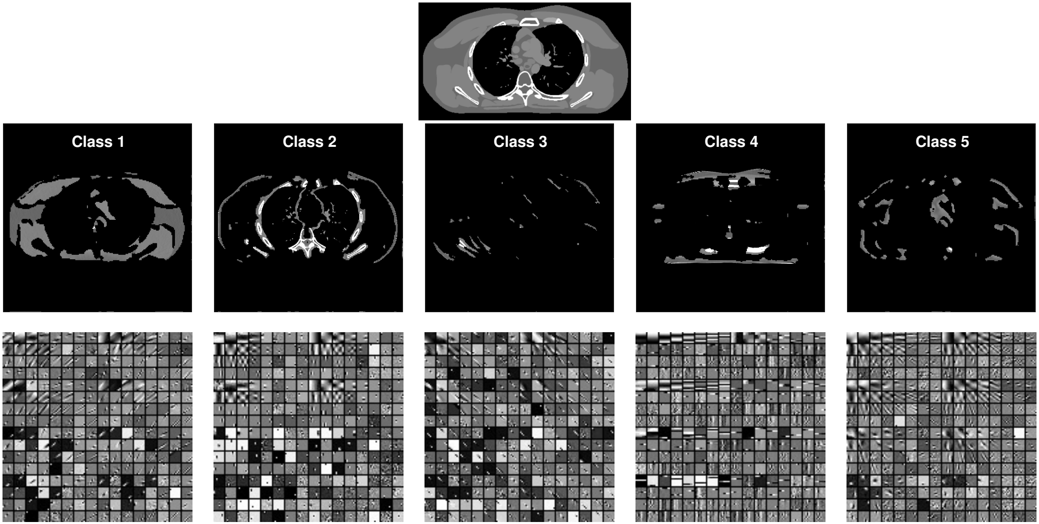
Union of Learned TRAnsforms (ULTRA)

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

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

- Joint unsupervised clustering / sparsification
- Further nonconvexity due to clustering
- Efficient alternating minimization algorithm [56]

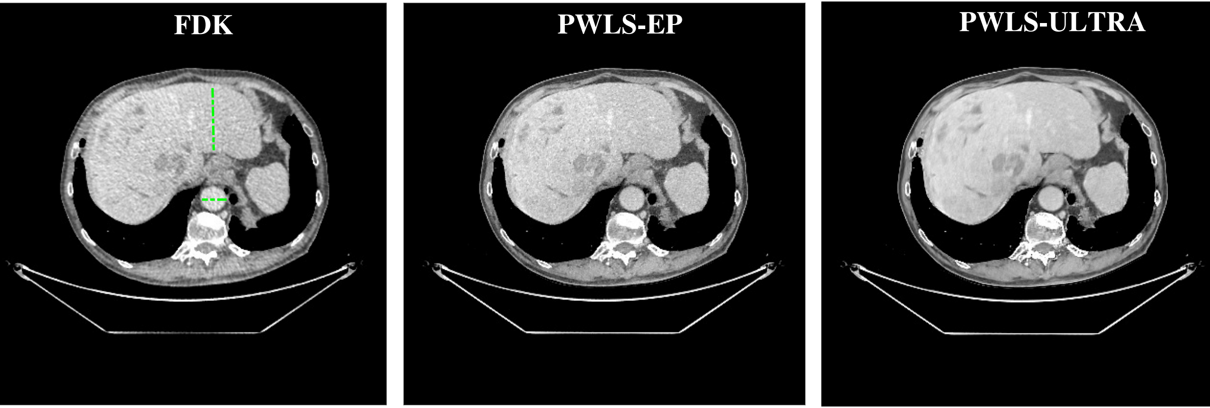
Example: 3D X-ray CT learned set of transforms



X. Zheng, S. Ravishankar, Y. Long, JF: IEEE T-MI, June 2018 [54]

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

J. Fessler



Zheng et al., IEEE T-MI, June 2018 [54] (Special issue on machine learning for image reconstruction)

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

`https://github.com/xuehangzheng/PWLS-ULTRA-for-Low-Dose-3D-CT-Image-Reconstruction`

Outline

X-ray CT with learned convolutional filters

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

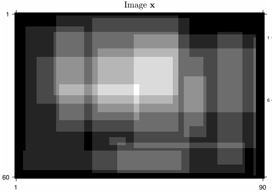
Drawback of basic patch-based methods:
$512 \times 512 \times 512$ 3D X-ray CT image volume
$8 \times 8 \times 8$ patches
$\Longrightarrow 512^3 \cdot 8^3 \cdot 4 = 256$ Gbyte of patch data for stride=1

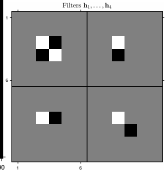
Convolutional sparsity: analysis model

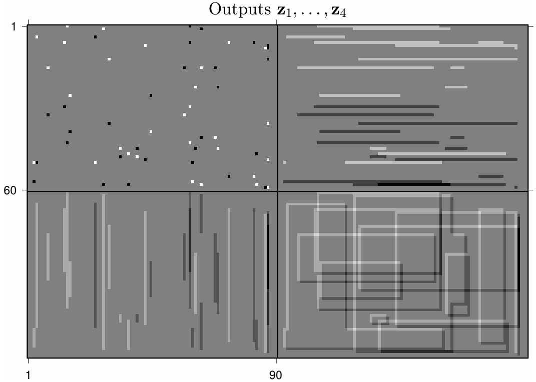
Assumption: For a plausible image $\boldsymbol{x}$, the filter outputs $\{\boldsymbol{h}_k * \boldsymbol{x}\}$ are sparse, for some filters $\{\boldsymbol{h}_k\}_{k=1}^K$ [57]

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

Example (hand crafted filters):







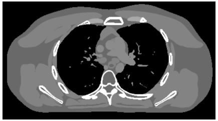
Given training images $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_L$ from a representative population, find filters $\left\{\hat{\boldsymbol{h}}_k\right\}_{k=1}^{K}$ that best sparsify them:

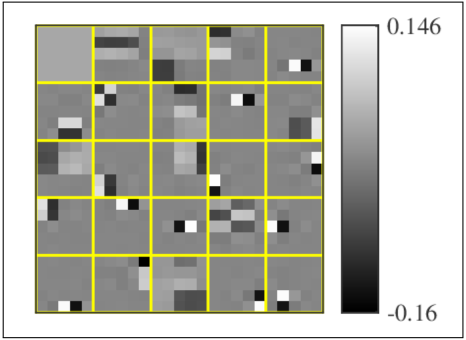
$$\left\{\hat{\boldsymbol{h}}_k\right\} = \operatorname*{arg\,min}_{\{\boldsymbol{h}_k\}\in\mathcal{H}} \min_{\{\boldsymbol{z}_{l,k}\}} \sum_{l=1}^{L} \sum_{k=1}^{K} \|\boldsymbol{h}_k * \boldsymbol{x}_l - \boldsymbol{z}_{l,k}\|_2^2 + \alpha \|\boldsymbol{z}_{l,k}\|_0$$

- 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 [58]
  - $\boldsymbol{z}$ update is simply hard thresholding
  - Filter update uses diagonal majorizer, proximal map (SVD)
  - Subsequence convergence guarantees [58]

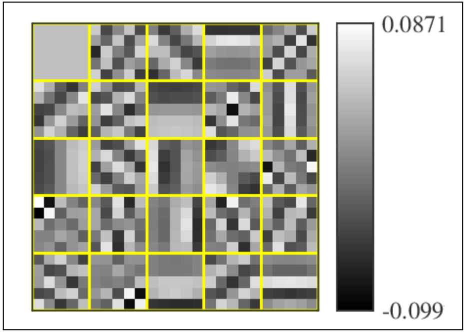
Examples of learned sparsifying filters

2D X-ray CT training data and learned 5×5 sparsifying filters $\{\hat{h}_k\}$ [58]:





$\alpha = 10^{-4}$



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

Regularizer based on learned sparsifying filters

Regularized inverse problem [58]:

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

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

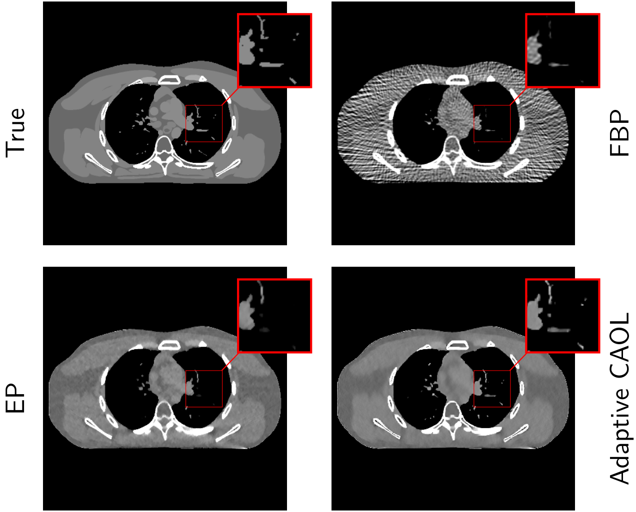
$\left\{ \hat{\boldsymbol{h}}_k \right\}$ adapted to population training data

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 [58]

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



123 views
(out of usual 984)
$\Longrightarrow 8\times$ dose reduction
25 filters 5×5

RMSE (in HU):

FBP	82.8
EP	40.8
Adaptive filters	**35.2**

- Physics / statistics provides dramatic improvement
- Data-adaptive regularization further reduces RMSE, improves fine details

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 [58]:

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

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 [58]

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

Outline

MR with adapted patch 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

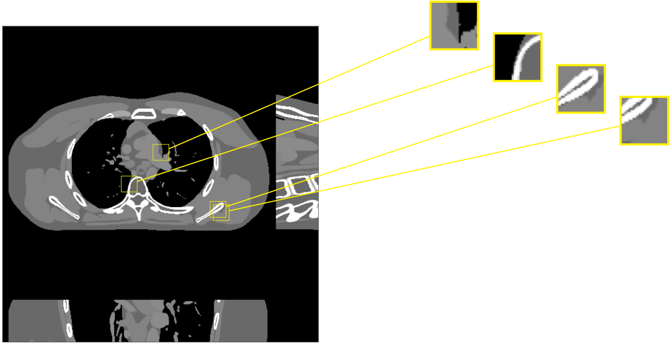
J. Fessler

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



MR reconstruction using adaptive dictionary regularizer

Dictionary-blind MR image reconstruction:

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

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

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

In words: of the many images...

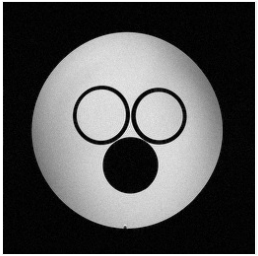
Alternating (nested) minimization:

- 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 [59].
- 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 $\mathcal{D}$, $\boldsymbol{D}\boldsymbol{z}_m$, 0-norm, but monotone decreasing and some convergence theory [59].

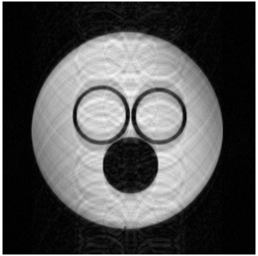
2D CS MRI results I

J. Fessler

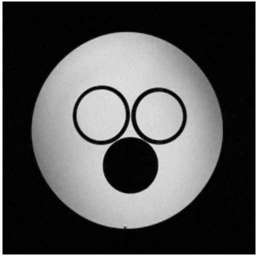
Fully Sampled



Zero-Filled



SOUP-DILLO-MRI



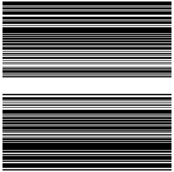
6×6 patches

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

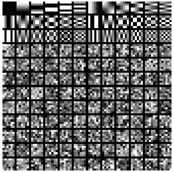
$\boldsymbol{D}_0$: [DCT | random]

[59]

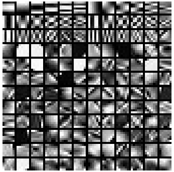
Sampling ($2.5\times$)



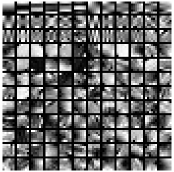
Initial $\boldsymbol{D}$



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

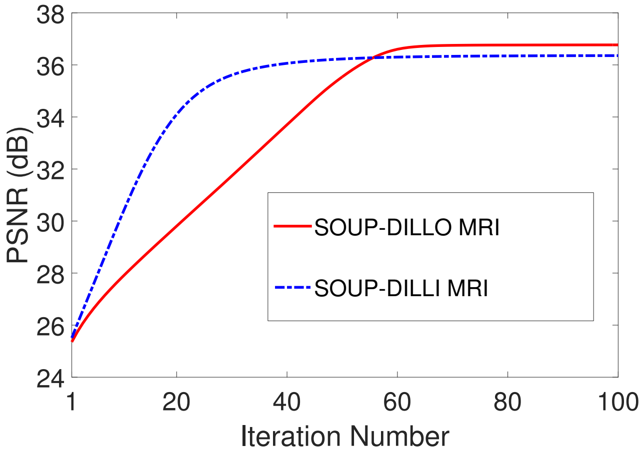


imag$\{\boldsymbol{D}\}$



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

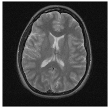
2D CS MRI results II

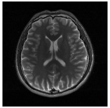


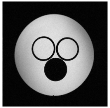
(SNR vs fully sampled image.)
Using $\|z_m\|_0$ leads to higher SNR than $\|z_m\|_1$.
Adaptive case is non-convex anyway...

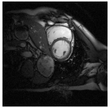
Matlab code: `http://web.eecs.umich.edu/~fessler/irt/reproduce/`
`https://gitlab.eecs.umich.edu/fessler/soupdil_dinokat`

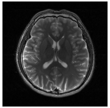
2D CS MRI results III

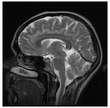

(a)

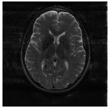

(b)


(c)


(d)


(e)

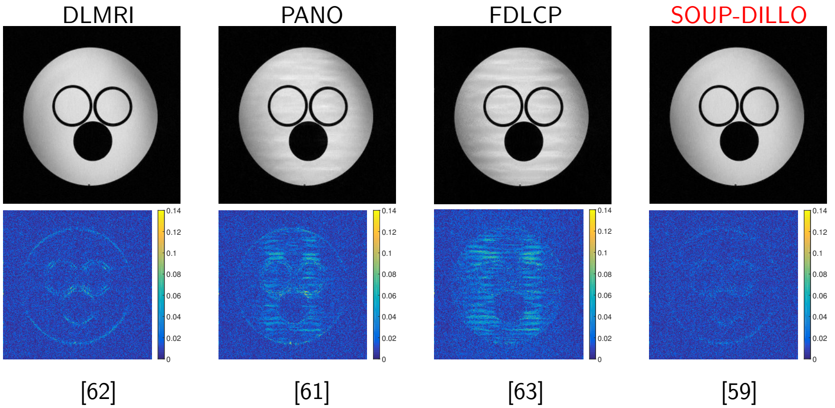

(f)


(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.				[60]	[61]	[62]	[59]	[59]

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) $\boldsymbol{T}$ (for patches)
- or convolutional versions thereof [57, 64]

ML-based regularized optimization problem using M image patches:

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

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

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

Alternative: blind adaptive learned dictionary [62] or learned sparsifying transform [65].
Double minimization (so very “deep?”) More interpretable than CNNs?

Outline

Cost function for convolutional sparsity regularization:

$$\arg\min_{\boldsymbol{x}} \frac{1}{2} \|\boldsymbol{A}\boldsymbol{x} - \boldsymbol{y}\|_{\boldsymbol{W}}^2 + \beta \left(\min_{\zeta} \sum_{k=1}^K \frac{1}{2} \|\boldsymbol{h}_k * \boldsymbol{x} - \zeta_k\|_2^2 + \alpha \|\zeta_k\|_1 \right)$$

Alternating minimization, aka block coordinate descent (BCD), updates:

Sparse code: $\zeta_k^{(n+1)} = \mathrm{soft}\{\boldsymbol{h}_k * \boldsymbol{x}^{(n)}, \alpha\}$

Image: $\boldsymbol{x}^{(n+1)} = \arg\min_{\boldsymbol{x}} F(\boldsymbol{x}; \boldsymbol{y}, \boldsymbol{z}^{(n)})$

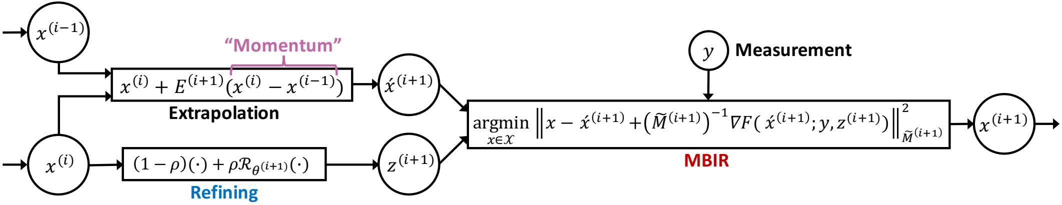
$$F(\boldsymbol{x}; \boldsymbol{y}, \boldsymbol{z}^{(n)}) \triangleq \frac{1}{2} \|\boldsymbol{A}\boldsymbol{x} - \boldsymbol{y}\|_{\boldsymbol{W}}^2 + \beta \left(\sum_{k=1}^K \frac{1}{2} \left\|\boldsymbol{h}_k * \boldsymbol{x} - \zeta_k^{(n+1)}\right\|_2^2 + \alpha \left\|\zeta_k^{(n+1)}\right\|_1 \right)$$

$$= \frac{1}{2} \|\boldsymbol{A}\boldsymbol{x} - \boldsymbol{y}\|_{\boldsymbol{W}}^2 + \beta \frac{1}{2} \left\|\boldsymbol{x} - \boldsymbol{z}^{(n)}\right\|_2^2 \quad \text{(quadratic but } \textit{large} \Longrightarrow \text{majorize)}$$

$$\boldsymbol{z}^{(n)} = \mathcal{R}(\boldsymbol{z}^{(n)}) = \sum_{k=1}^K \mathrm{flip}(\boldsymbol{h}_k) * \mathrm{soft}\{\boldsymbol{h}_k * \boldsymbol{x}^{(n)}\} \quad \text{(denoise} \Longrightarrow \text{learn)}$$

Momentum-Net overview

Unrolled loop network with momentum and quadratic majorizer [68, 69]:



- Diagonal majorizer for CT: $\boldsymbol{M} = \mathrm{Diag}\{\boldsymbol{A}'\boldsymbol{W}\boldsymbol{A}\mathbf{1}\} + \beta \boldsymbol{I} \succeq \boldsymbol{A}'\boldsymbol{W}\boldsymbol{A} + \beta \boldsymbol{I}$
- Learn image mapper ("refiner") $\mathcal{R}$ from training data (supervised).
  cf CNN: filter $\to$ threshold $\to$ filter

Momentum-Net benefits

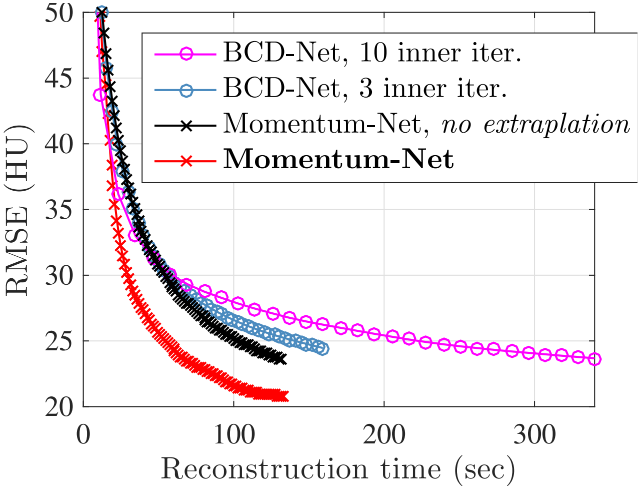
- Image mapper $\mathcal{R}$ is shallow
  $\Longrightarrow$ less risk of over-fitting / hallucination
- Momentum accelerates convergence $\Longrightarrow$ fewer "layers" (outer iterations)
- First unrolled loop approach to have convergence theory
  (under suitable assumptions on $\mathcal{R}$)
- Image update uses original measurements $\boldsymbol{y}$ and imaging physics $\boldsymbol{A}$

[68, 69] Il Yong Chun, Zhengyu Huang, Hongki Lim, J A Fessler
Momentum-Net: Fast and convergent iterative neural network for inverse problems
http://arxiv.org/abs/1907.11818,

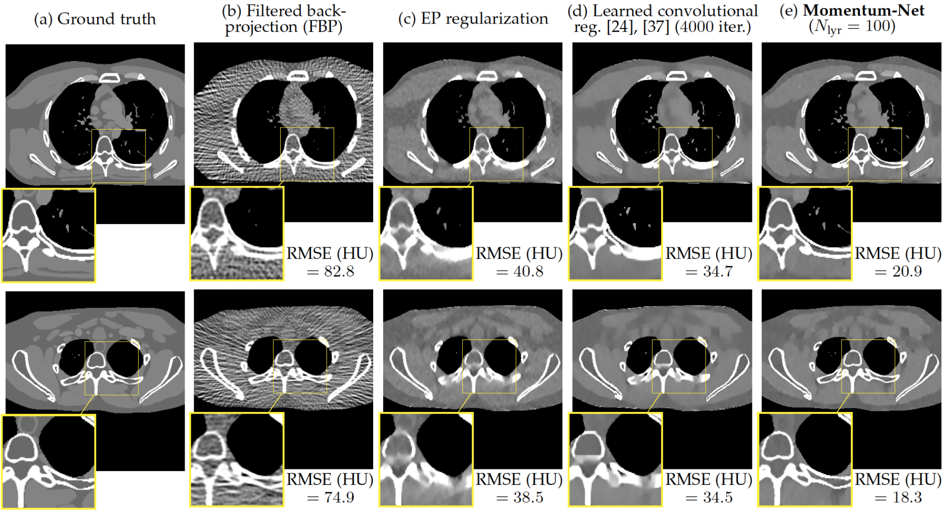
IEEE Tr. on PAMI, 2020 http://doi.org/10.1109/TPAMI.2020.3012955

Momentum-Net preliminary results

Illustration of benefits of momentum:



Momentum-Net preliminary image results



Sparse-view CT with 123/984 views, $I_0 = 10^5$, 800-1200 mod. HU display.

Outline

Summary

- CT image reconstruction has evolved greatly in the 50+ years since Allan Cormack's seminal papers [70, 71]
  - physics
  - statistics
  - regularization and optimization
  - data adaptive methods inspired by machine learning
- 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
- More comparisons between model-based methods with adaptive regularizers and CNN-based methods needed

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. Litjens, T. Kooi, B. E. Bejnordi, A. A. A. Setio, F. Ciompi, M. Ghafoorian, J. A. W. M. . . Laak, B. . Ginneken, and C. I. Sanchez. "A survey on deep learning in medical image analysis." In: *Med. Im. Anal.* 42.C (Dec. 2017), 60–88.

[3] 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.

[4] V. Cheplygina, M. . Bruijne, and J. P. W. Pluim. "Not-so-supervised: A survey of semi-supervised, multi-instance, and transfer learning in medical image analysis." In: *Med. Im. Anal.* 54 (May 2019), 280–96.

[5] A. Esteva, A. Robicquet, B. Ramsundar, V. Kuleshov, M. DePristo, K. Chou, C. Cui, G. Corrado, S. Thrun, and J. Dean. "A guide to deep learning in healthcare." In: *Nature Medicine* 25.1 (Jan. 2019), 24–9.

[6] X. Yi, E. Walia, and P. Babyn. "Generative adversarial network in medical imaging: A review." In: *Med. Im. Anal.* 58 (Dec. 2019), p. 101552.

[7] J. Bruna, E. Haber, G. Kutyniok, T. Pock, and Rene Vidal. "Special issue on the mathematical foundations of deep learning in imaging science." In: *J. Math. Im. Vision* 62.3 (2020), 277–8.

[8] D. Rueckert and J. A. Schnabel. "Model-based and data-driven strategies in medical image computing." In: *Proc. IEEE* 108.1 (Jan. 2020), 110–24.

[9] A. Maier, C. Syben, T. Lasser, and C. Riess. "A gentle introduction to deep learning in medical image processing." In: *Zeitschrift für Medizinische Physik* 29.2 (May 2019), 86–101.

[10] S. Ravishankar and Y. Bresler. "Adaptive sampling design for compressed sensing MRI." In: *Proc. Int'l. Conf. IEEE Engr. in Med. and Biol. Soc.* 2011, 3751–5.

[11] 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.

Bibliography II

[12] 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.

[13] G. Godaliyadda, D. H. Ye, M. D. Uchic, M. A. Groeber, G. T. Buzzard, and C. A. Bouman. "A framework for dynamic image sampling based on supervised learning." In: *IEEE Trans. Computational Imaging* 4.1 (Mar. 2018), 1–16.

[14] H. K. Aggarwal and M. Jacob. "J-MoDL: Joint model-based deep learning for optimized sampling and reconstruction." In: *IEEE J. Sel. Top. Sig. Proc.* 14.6 (Oct. 2020), 1151–62.

[15] C. Bahadir, A. Wang, A. Dalca, and M. R. Sabuncu. "Deep-learning-based optimization of the under-sampling pattern in MRI." In: *IEEE Trans. Computational Imaging* (2020).

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

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

[18] 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.

[19] 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.

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

[21] 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.

[22] A. Lucas, M. Iliadis, R. Molina, and A. K. Katsaggelos. "Using deep neural networks for inverse problems in imaging: Beyond analytical methods." In: *IEEE Sig. Proc. Mag.* 35.1 (Jan. 2018), 20–36.

[23] M. T. McCann and M. Unser. "Biomedical image reconstruction: from the foundations to deep neural networks." In: *Found. & Trends in Sig. Pro.* 13.3 (2019), 283–359.

Bibliography III

[24] S. Arridge, P. Maass, O. Oktem, and C-B. Schonlieb. "Solving inverse problems using data-driven models." In: *Acta Numerica* 28 (May 2019), 1–174.

[25] V. Monga, Y. Li, and Y. C. Eldar. *Algorithm unrolling: interpretable, efficient deep learning for signal and image processing*. 2020.

[26] S. Ravishankar, J. C. Ye, and J. A. Fessler. "Image reconstruction: from sparsity to data-adaptive methods and machine learning." In: *Proc. IEEE* 108.1 (Jan. 2020), 86–109.

[27] G. Ongie, A. Jalal, C. A. M. R. G. Baraniuk, A. G. Dimakis, and R. Willett. "Deep learning techniques for inverse problems in imaging." In: *ieee-jsait* (2020).

[28] 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.

[29] Q. De Man, E. Haneda, B. Claus, P. Fitzgerald, B. De Man, G. Qian, H. Shan, J. Min, M. Sabuncu, and G. Wang. "A two-dimensional feasibility study of deep learning-based feature detection and characterization directly from CT sinograms." In: *Med. Phys.* 46.12 (Dec. 2019), e790–800.

[30] P. J. Pickhardt, M. G. Lubner, D. H. Kim, J. Tang, J. A. Ruma, A. Muñoz del Rio, and G-H. Chen. "Abdominal CT with model-based iterative reconstruction (MBIR): Initial results of a prospective Trial comparing ultralow-dose with standard-dose imaging." In: *Am. J. Roentgenol.* 199.6 (Dec. 2012), 1266–74.

[31] FDA. *510k premarket notification of AiCE Deep Learning Reconstruction (Canon)*. 2019.

[32] FDA. *510k premarket notification of Deep Learning Image Reconstruction (GE Medical Systems)*. 2019.

[33] S. Wang, Z. Su, L. Ying, X. Peng, and D. Liang. "Exploiting deep convolutional neural network for fast magnetic resonance imaging." In: *Proc. Intl. Soc. Mag. Res. Med.* 2016, p. 1778.

[34] D. Lee, J. Yoo, and J. C. Ye. *Deep artifact learning for compressed sensing and parallel MRI*. 2017.

Bibliography IV

[35] K. H. Jin, M. T. McCann, E. Froustey, and M. Unser. "Deep convolutional neural network for inverse problems in imaging." In: *IEEE Trans. Im. Proc.* 26.9 (Sept. 2017), 4509–22.

[36] M. Akcakaya, S. Moeller, S. Weingartner, and Kamil Ugurbil. "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.

[37] Y. Han and J. C. Ye. "K-space deep learning for accelerated MRI." In: *IEEE Trans. Med. Imag.* 39.2 (Feb. 2020), 377–86.

[38] M. U. Ghani and W. C. Karl. *Data and image prior integration for image reconstruction using consensus equilibrium*. 2020.

[39] 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.

[40] I. Haggstrom, C. R. Schmidtlein, G. Campanella, and T. J. Fuchs. "DeepPET: A deep encoder-decoder network for directly solving the PET image reconstruction inverse problem." In: *Med. Im. Anal.* 54 (May 2019), 253–62.

[41] 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.

[42] W. Whiteley, V. Panin, C. Zhou, J. Cabello, D. Bharkhada, and J. Gregor. *FastPET: near real-time PET reconstruction from histo-images using a neural network*. 2020.

[43] Y. Yang, J. Sun, H. Li, and Z. Xu. "Deep ADMM-net for compressive sensing MRI." In: *Neural Info. Proc. Sys.* 2016, 10–18.

[44] 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.

[45] 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.

[46] 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.

[47] D. Lee, J. Yoo, S. Tak, and J. C. Ye. "Deep residual learning for accelerated MRI using magnitude and phase networks." In: *IEEE Trans. Biomed. Engin.* 65.9 (Sept. 2018), 1985–95.

[48] 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.

[49] 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.

[50] K. He, X. Zhang, S. Ren, and J. Sun. "Deep residual learning for image recognition." In: *Proc. IEEE Conf. on Comp. Vision and Pattern Recognition*. 2016, 770–8.

[51] K. Zhang, W. Zuo, Y. Chen, D. Meng, and L. Zhang. "Beyond a Gaussian denoiser: residual learning of deep CNN for image denoising." In: *IEEE Trans. Im. Proc.* 26.7 (July 2017), 3142–55.

[52] 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.

[53] 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.

[54] 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.

[55] 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.

[56] 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.

Bibliography VI

[57] I. Y. Chun and J. A. Fessler. *Convolutional analysis operator learning: acceleration and convergence*. 2018.

[58] I. Y. Chun and J. A. Fessler. "Convolutional analysis operator learning: acceleration and convergence." In: *IEEE Trans. Im. Proc.* 29.1 (Jan. 2020), 2108–22.

[59] 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.

[60] 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.

[61] 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.

[62] 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.

[63] 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.

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

[65] 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.

[68] I. Y. Chun, Z. Huang, H. Lim, and J. A. Fessler. *Momentum-Net: Fast and convergent iterative neural network for inverse problems*. 2019.

[69] I. Y. Chun, Z. Huang, H. Lim, and J. A. Fessler. "Momentum-Net: Fast and convergent iterative neural network for inverse problems." In: *IEEE Trans. Patt. Anal. Mach. Int.* (2020). To appear.

Bibliography VII

[70] A. M. Cormack. “Representation of a function by its line integrals, with some radiological applications.” In: *J. Appl. Phys.* 34.9 (Sept. 1963), 2722–7.

[71] A. M. Cormack. “Representation of a function by its line integrals, with some radiological applications II.” In: *japplphy* 35.10 (Oct. 1964), 2908–13.