Learning Deep Denoisers for Low-Field MRI with Noisy Data

Nikola Janjušević

Presenting work done in collaboration with: Prof. Li Feng, Prof. Yao Wang

Center for Advanced Imaging Innovation and Research (CAI^2R)
Department of Electrical and Computer Engineering
New York University

February 21st 2024

Outline

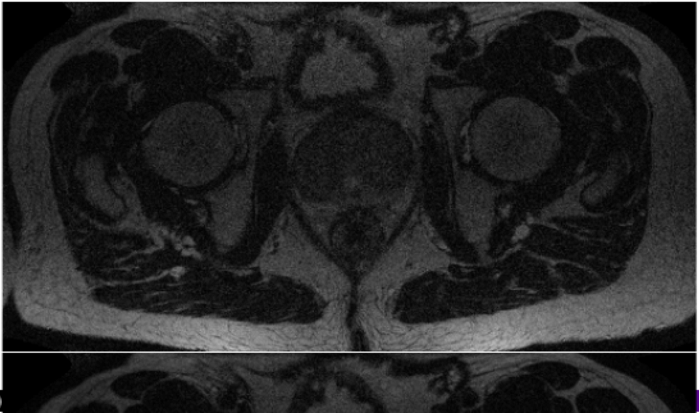
1. Motivation
2. Noise and Observation Model
3. Deep Denoiser Architecture
4. Learning Deep MRI Denoisers
5. Summary and Future Work

Outline

1. Motivation
   - Low-Field MRI
   - Existing Tools
2. Noise and Observation Model
3. Deep Denoiser Architecture
4. Learning Deep MRI Denoisers
5. Summary and Future Work

Low-Field MRI Acquisition

- Low-cost construction
- Slower T2*-decay
  - Promising for new applications (lung-imaging)
- Low-Field strength $\Rightarrow$ Low-SNR
- Scan Averaging $\Rightarrow$ High(er)-SNR
- K-space undersampling~~K-space undersampling~~ Scan Denoising



Goals

To develop a technique for Low-Field MRI Denoising: SNAC-DL

- Data-driven denoiser
- Coil-Number Agnostic
- No ground-truth data necessary (Unsupervised / Self-Supervised)

Self-supervised Noise-Adaptive Convolutional Dictionary-Learning (SNAC-DL) for LFMRI

Deep Neural Network Training

Parameterized Denoiser: $f(\boldsymbol{y}) = f(\boldsymbol{y},\, \Theta)$.

Supervised Loss Function

Over distribution of clean ($\boldsymbol{x}$) and noisy ($\boldsymbol{y}$) images ($p(\boldsymbol{x},\boldsymbol{y})$):

$$\begin{aligned}
\mathcal{L} &= \mathbb{E}_{p(\boldsymbol{x},\boldsymbol{y})}[\text{distance}(\boldsymbol{x},\, f(\boldsymbol{y}))] \\
&= \iint \text{distance}(\boldsymbol{x},\, f(\boldsymbol{y}))\, p(\boldsymbol{x},\boldsymbol{y})\, d\boldsymbol{x}\, d\boldsymbol{y} \\
&\approx \frac{1}{|\mathcal{D}|} \sum_{(\boldsymbol{x},\boldsymbol{y})\sim\mathcal{D}} \text{distance}(\boldsymbol{x}, f(\boldsymbol{y}))
\end{aligned}$$

Optimize Θ via SGD on loss function: $\Theta \leftarrow \Theta - \text{lr} * \nabla_\Theta \mathcal{L}$.

MSE Loss

$$\mathcal{L}_{\text{MSE}} = \mathbb{E}[\|\boldsymbol{x} - f(\boldsymbol{y})\|_2^2]$$

Noise2Noise Loss

Self-Supervised Loss

Consider two noisy observations:

$$\boldsymbol{y} = \boldsymbol{x} + \boldsymbol{\nu}_y$$
$$\boldsymbol{z} = \boldsymbol{x} + \boldsymbol{\nu}_z$$

with noise independent to signal ($\boldsymbol{\nu} \perp\!\!\!\perp \boldsymbol{x}$).

Noise2Noise Loss function

$$\mathcal{L}_{\text{N2N}} = \mathbb{E}[\|\boldsymbol{z} - f(\boldsymbol{y})\|_2^2]$$

$$\begin{aligned}
\mathcal{L}_{\text{MSE}} &= \mathbb{E}[\|\boldsymbol{x} - f(\boldsymbol{y})\|_2^2] \\
&= \mathbb{E}[\|\boldsymbol{x} - \boldsymbol{z} + \boldsymbol{z} - f(\boldsymbol{y})\|_2^2] \\
&= \mathcal{L}_{\text{N2N}} - N\sigma_z^2 + 2\,\mathbb{E}[\langle \boldsymbol{\nu}_z, f(\boldsymbol{y})\rangle] \\
&= \mathcal{L}_{\text{N2N}} - N\sigma_z^2 \qquad (\boldsymbol{\nu}_z \perp\!\!\!\perp \boldsymbol{\nu}_y,\, \mathbb{E}[\boldsymbol{\nu}_z] = 0) \\
\nabla_\Theta \mathcal{L}_{\text{MSE}} &= \nabla_\Theta \mathcal{L}_{\text{N2N}}
\end{aligned}$$

SURE Loss

Self-Supervised Loss

Consider *one* noisy observation:

$$\boldsymbol{y} = \boldsymbol{x} + \boldsymbol{\nu}, \quad \boldsymbol{\nu} \sim \mathcal{N}(0,\, \sigma^2 \boldsymbol{I})$$

with noise independent to signal ($\boldsymbol{\nu} \perp\!\!\!\perp \boldsymbol{x}$).

SURE Loss function

$$\mathcal{L}_{\text{SURE}} = \mathbb{E}[\|\boldsymbol{y} - f(\boldsymbol{y})\|_2^2 - N\sigma^2 + 2\sigma^2 \nabla_{\boldsymbol{y}} \cdot f(\boldsymbol{y})]$$

$$\begin{aligned}
\mathcal{L}_{\text{MSE}} &= \mathbb{E}[\|\boldsymbol{x} - f(\boldsymbol{y})\|_2^2] \\
&= \mathbb{E}[\|\boldsymbol{x} - \boldsymbol{y} + \boldsymbol{y} - f(\boldsymbol{y})\|_2^2] \\
&= \mathbb{E}[\|\boldsymbol{y} - f(\boldsymbol{y})\|_2^2 - N\sigma^2 + 2\langle \boldsymbol{\nu}, f(\boldsymbol{y})\rangle] \\
&= \mathbb{E}[\|\boldsymbol{y} - f(\boldsymbol{y})\|_2^2 - N\sigma^2 + 2\sigma^2 \nabla_{\boldsymbol{y}} \cdot f(\boldsymbol{y})] \\
&= \mathcal{L}_{\text{SURE}} \\
\nabla_\Theta \mathcal{L}_{\text{MSE}} &= \nabla_\Theta \mathcal{L}_{\text{SURE}}
\end{aligned}$$

Approximate divergence term with finite-difference:

$$\nabla_{\boldsymbol{y}} \cdot f(\boldsymbol{y}) \approx \boldsymbol{b}^T \left(\frac{f(\boldsymbol{y} + \epsilon \boldsymbol{b}) - f(\boldsymbol{y})}{\epsilon} \right), \quad \boldsymbol{b} \sim \mathcal{N}(0,\, \boldsymbol{I}).$$

Losses

Loss	Pro	Con
Supervised	- desired objective	- ground-truth needed
Noise2Noise	- no ground-truth needed - equivalent to $\mathcal{L}_{\text{MSE}}$	- 2 noisy samples needed
SURE	- only 1 noisy sample needed	- must estimate σ - only approximately equivalent to $\mathcal{L}_{\text{MSE}}$

Outline

1 Motivation
2 Noise and Observation Model
3 Deep Denoiser Architecture
4 Learning Deep MRI Denoisers
5 Summary and Future Work

Image-Space Observation Model

multi-coil

Multi-coil image (y)

$$y = Sx + \nu$$

$$\nu[i] \sim \mathcal{CN}(0, \Sigma) \quad \forall\ i$$

Sensitivity maps (s)

$$\sum_{i=1}^{C} |s_i|^2 = 1 \quad \text{for all pixels}$$

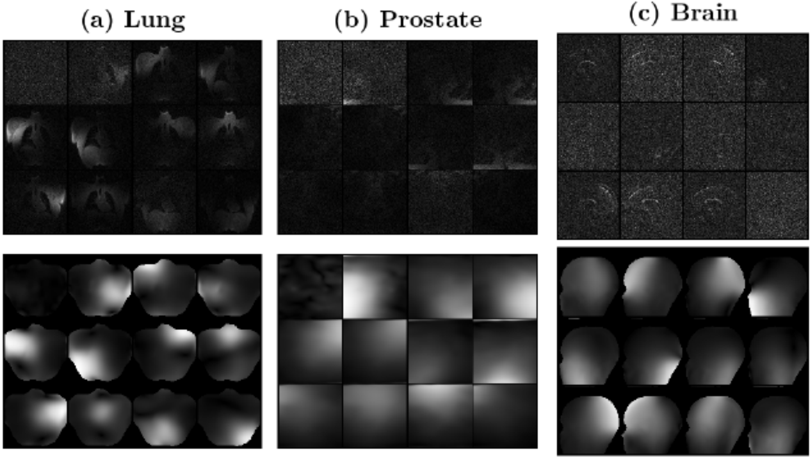


Figure: **Top**: Multi-coil image-space (y). **Bottom**: Sensitivity maps (s).

Image-Space Observation Model

coil-combined

Coil-combined image ($\tilde{y}$)

$$\tilde{y} = S^H y = x + \tilde{\nu}$$
$$\tilde{\nu} \sim \mathcal{CN}(0, \mathbf{diag}(\boldsymbol{\sigma}^2))$$

Effective noise-level (σ)

$$\boldsymbol{\sigma}[i]^2 = (\boldsymbol{s}[i])^H \boldsymbol{\Sigma} (\boldsymbol{s}[i]) \quad \forall\, i$$

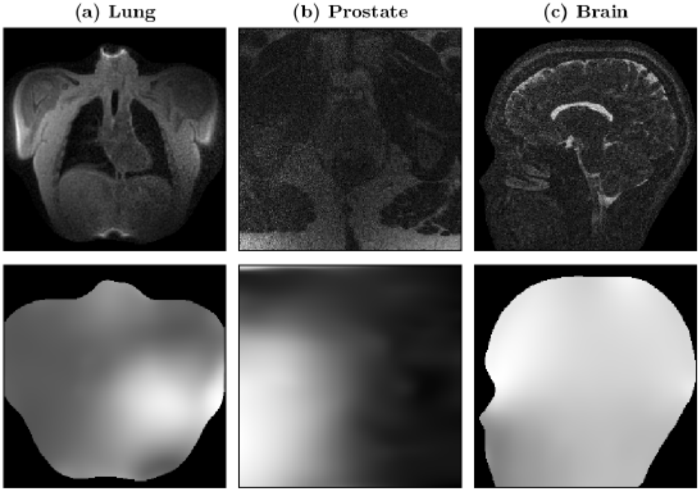


Figure: **Top**: Coil-combined image-space ($\tilde{y}$). **Bottom**: Noise-level maps (σ).

Image-Space Observation Model

whitened multi-coil

Whitened MC ($\boldsymbol{y}_w$)

$$\boldsymbol{y}_w[i] = \boldsymbol{\Sigma}^{-1/2}(\boldsymbol{y}_w[i]) \quad \forall\, i$$
$$\boldsymbol{y}_w = \boldsymbol{S}\boldsymbol{x}_w + \boldsymbol{\nu}_w$$
$$\boldsymbol{\nu}_w \sim \mathcal{CN}(0,\, \sigma_w^2 \boldsymbol{I})$$

Whitened s-maps ($\boldsymbol{S}_w$)

$$\boldsymbol{s}_w[i] = \frac{\boldsymbol{\Sigma}^{-1/2}(\boldsymbol{s}[i])}{z[i]} \quad \forall\, i$$
$$z[i] = \sqrt{(\boldsymbol{s}[i])^H \boldsymbol{\Sigma}^{-1}(\boldsymbol{s}[i])} \quad \forall\, i$$

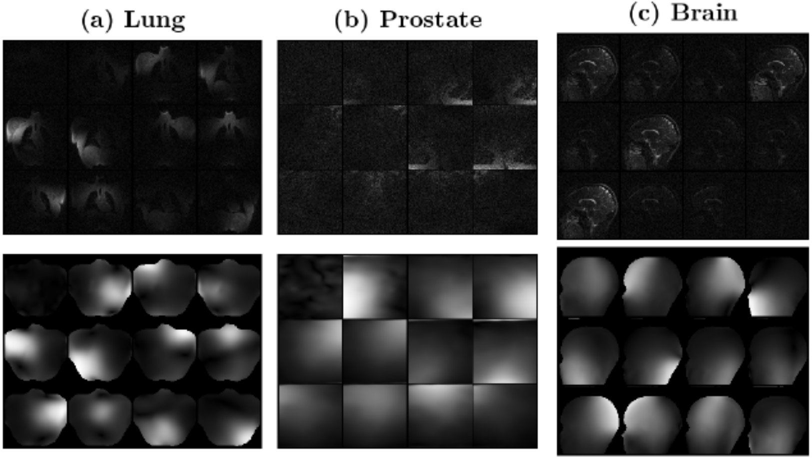


Figure: **Top**: Whitened MC image ($\boldsymbol{y}_w$). **Bottom**: Whitened s-maps ($\boldsymbol{s}_w$).

Image-Space Observation Model

whitened coil-combined

Whitened CC ($\tilde{y}$)

$$\tilde{\boldsymbol{y}}_w = \boldsymbol{S}_w^H \boldsymbol{y}_w = \boldsymbol{x}_w + \tilde{\boldsymbol{\nu}}_w$$
$$\tilde{\boldsymbol{\nu}}_w \sim \mathcal{CN}(0,\ \sigma_w^2 \boldsymbol{I})$$

Renormalized Whitened CC

$$\tilde{\boldsymbol{y}}_w / z$$

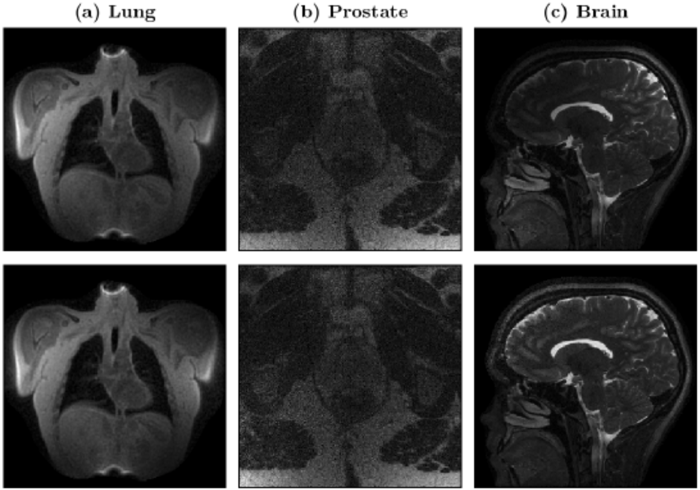


Figure: **Top**: Whitened CC image ($\tilde{y}_w$). **Bottom**: Renormalized CC image ($\tilde{y}_w / z$).

Image-Space Observation Model

Summary

- **Spaces**
  - CC:

$$\tilde{\boldsymbol{y}} \sim \mathcal{CN}(\boldsymbol{x},\ \mathbf{diag}(\boldsymbol{\sigma}^2))$$

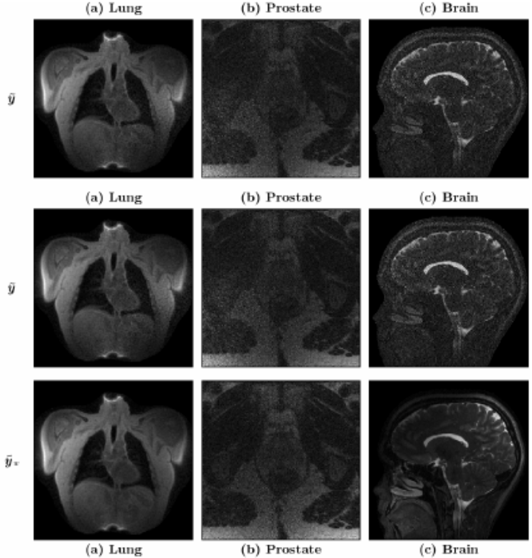
  - Whitened-CC:

$$\tilde{\boldsymbol{y}}_w \sim \mathcal{CN}(\boldsymbol{x}_w,\ \sigma_w^2 \boldsymbol{I})$$

  - Renorm-Whitened-CC:

$$\tilde{\boldsymbol{y}}_w / z \sim \mathcal{CN}(\boldsymbol{x},\ \sigma_w^2\, \mathbf{diag}(1/z))$$

- **Proposed Use**
  - Coil-combined $\Rightarrow$ coil-number agnostic
  - Denoise $\tilde{\boldsymbol{y}}_w$ $\Rightarrow$ constant noise-level
  - Display / evaluate with renormalization



Outline

1. Motivation
2. Noise and Observation Model
3. Deep Denoiser Architecture
4. Learning Deep MRI Denoisers
5. Summary and Future Work

Sparse Representation

Observation model: $\boldsymbol{y} = \boldsymbol{x} + \boldsymbol{\nu}, \quad \boldsymbol{\nu} \sim \mathcal{N}(0, \sigma^2 I)$

Given dictionary $\boldsymbol{D}$, assume: $\exists\, z \text{ s.t. } \boldsymbol{x} = \boldsymbol{D}z$, and z is sparse

Basis Pursuit DeNoising (BPDN)

$$\underset{z}{\text{minimize}} \quad \frac{1}{2}\|\boldsymbol{y} - \boldsymbol{D}\boldsymbol{z}\|_2^2 + \lambda\|\boldsymbol{z}\|_1$$

(Note: $\lambda = \lambda(\sigma)$)

Iterative Soft-Thresholding Algorithm (ISTA)

Let $\boldsymbol{z}^{(0)} = \boldsymbol{0}$, step-size $\eta \in (0, 1/L]$, $L = \|\boldsymbol{D}\|_2$.

$$\boldsymbol{z}^{(k+1)} = \mathrm{ST}_{\eta\lambda}\left(\boldsymbol{z}^{(k)} - \eta\boldsymbol{D}^T(\boldsymbol{D}\boldsymbol{z}^{(k)} - \boldsymbol{y})\right), \quad k = 0, 1, \ldots \infty,$$

$\mathrm{ST}_\tau(\boldsymbol{z}) = \mathrm{sign}(\boldsymbol{z}) \circ \mathrm{ReLU}\left(|\boldsymbol{z}| - \tau\right),$

Obtain $\hat{x} = \boldsymbol{D}\boldsymbol{z}^{(\infty)}$.

CDLNet: Unrolling Convolutional Dictionary Learning

ISTA:

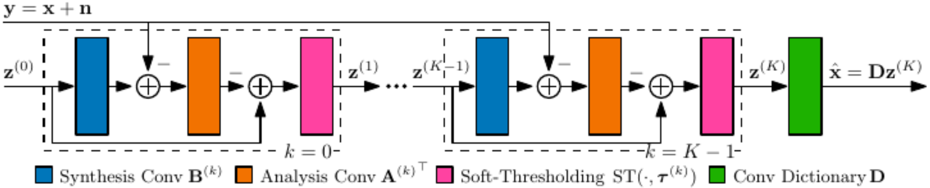
$$z^{(k+1)} = \mathrm{ST}\left(z^{(k)} - \eta D^T (D z^{(k)} - y),\ \eta\lambda\right), \quad k = 0, 1, \ldots, \infty$$

CDLNet: LISTA $+ D$

$$z^{(k+1)} = \mathrm{ST}\left(z^{(k)} - {A^{(k)}}^T (B^{(k)} z^{(k)} - y),\ \tau^{(k)}\right), \quad k = 0, 1, \ldots, K-1,$$

$$\hat{x} := f(y, \Theta) = D z^{(K)} \hat{x} := f(y, \hat{\sigma}, \Theta) = D z^{(K)}$$

- ${A^{(k)}}^T, B^{(k)}$ conv. analysis, synthesis respectively
- $\tau^{(k)} \in \mathbb{R}_+^M$ $\tau^{(k)} = \tau_0^{(k)} + \hat{\sigma}\tau_1^{(k)} \in \mathbb{R}^M$
- D conv. synthesis "*dictionary*"
- $\Theta = \{[A^{(k)}, B^{(k)}, \tau^{(k)}]_{k=0}^{K-1}, D\}$

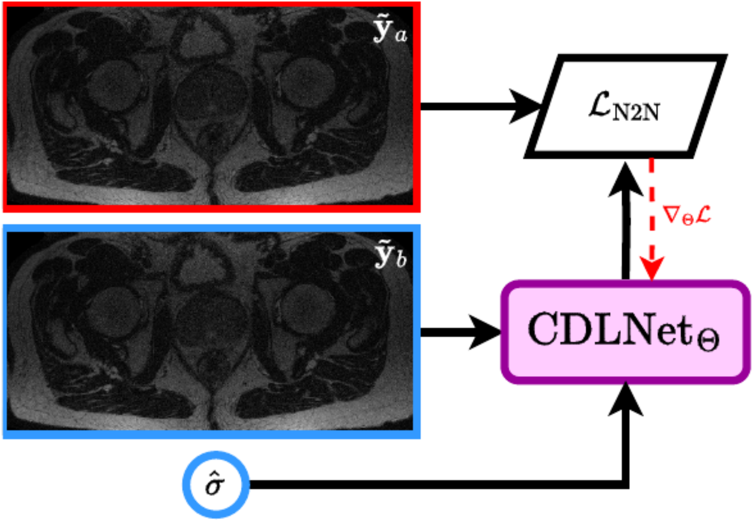


Outline

1. Motivation
2. Noise and Observation Model
3. Deep Denoiser Architecture
4. Learning Deep MRI Denoisers
   - MRI Specific Losses
   - Synthetic Denoising
   - Prostate Image Denoising
   - Lung Image Denoising
5. Summary and Future Work

Average2Average Loss (Avg2Avg)

- coil-combined scans: $\tilde{y}_a$, $\tilde{y}_b$
  - with independent noise
- estimate noise-level in b: $\hat{\sigma}$
  - for CDLNet adaptive-thresholds only
- (**optional**) pre-whiten data
  - to improve coil-combination



Coil2Coil Loss

$$y = Sx + \nu$$

$$\begin{bmatrix} y_A \\ y_B \end{bmatrix} = \begin{bmatrix} S_A \\ S_B \end{bmatrix} x + \begin{bmatrix} \nu_A \\ \nu_B \end{bmatrix}$$

$$\tilde{y}_1 = \frac{1}{\|s_A\|^2} S_A^H y_A = x + \frac{1}{\|s_A\|^2} S_A^H \nu_A$$

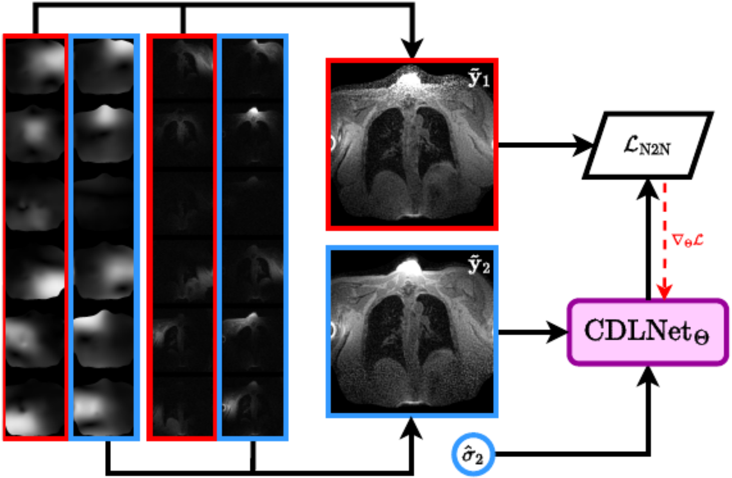
$$\tilde{y}_2 = \frac{1}{\|s_B\|^2} S_B^H y_B = x + \frac{1}{\|s_B\|^2} S_B^H \nu_B$$

If y is coil-whitened with $\nu \sim \mathcal{CN}(0, \sigma_w^2 I)$, then,

$$\tilde{y}_1 \sim \mathcal{CN}\left(x, \frac{\sigma_w^2}{\|s_A\|^2}\right)$$

$$\tilde{y}_2 \sim \mathcal{CN}\left(x, \frac{\sigma_w^2}{\|s_B\|^2}\right)$$

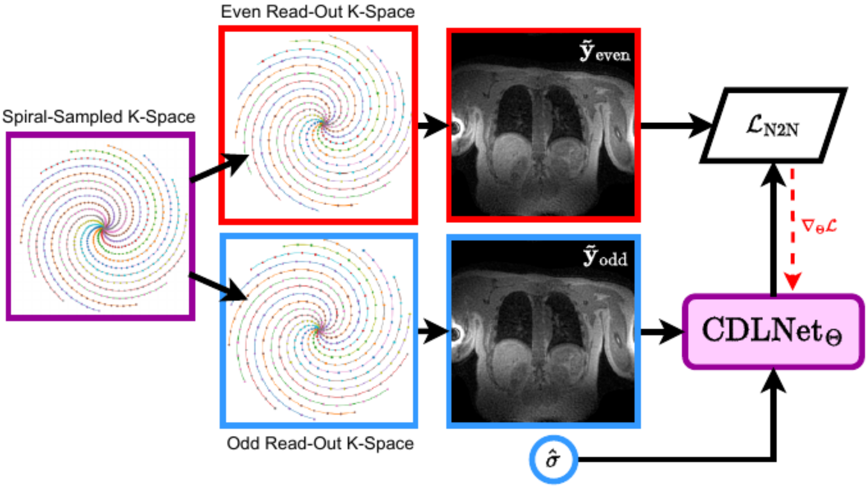
with $\tilde{\nu}_1 \perp \tilde{\nu}_2$ and $\mathbb{E}[\tilde{\nu}] = 0$.



Readout2Readout Loss (Ro2Ro)

Spiral Sampling

- split into even/odd along readout dimension
  - may introduce aliasing artifacts
- estimate noise-level in odd: $\hat{\sigma}$
  - for CDLNet adaptive-thresholds only
- (**optional**) pre-whiten data
  - to improve coil-combination



MRI Losses

Loss	Pro	Con
Supervised	- desired objective	- ground-truth needed
Avg2Avg	- no ground-truth needed - equivalent to $\mathcal{L}_{\text{MSE}}$	- 2 noisy samples needed
SURE	- only 1 noisy sample needed	- must estimate σ - only approximately equivalent to $\mathcal{L}_{\text{MSE}}$
Coil2Coil	- only 1 noisy sample needed - equivalent to $\mathcal{L}_{\text{MSE}}$	- must estimate Σ for whitening - must store multi-coil data
Ro2Ro	- only 1 noisy sample needed - equivalent to $\mathcal{L}_{\text{MSE}}$	- May generate readout aliasing artifacts

Synthetic Denoising

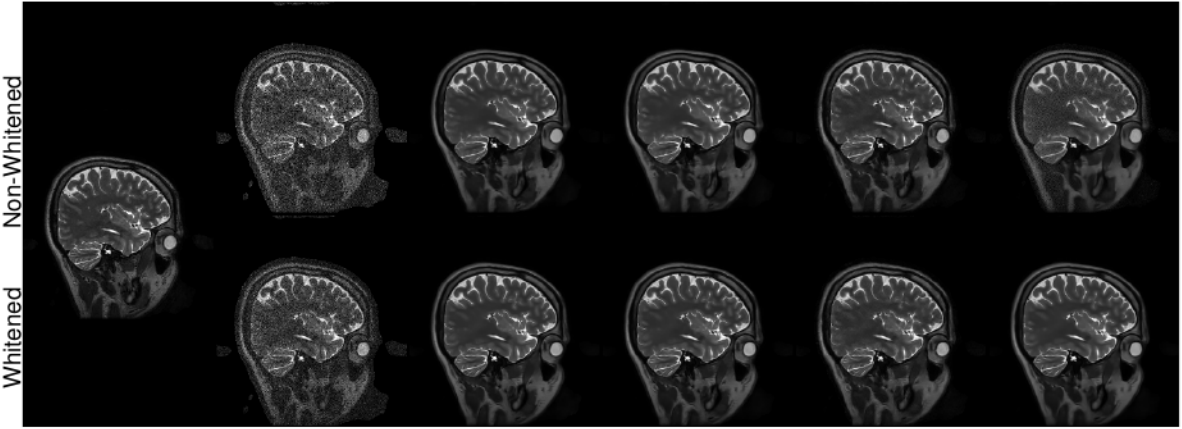
- Generate random (diagonal-dominant) matrix $\boldsymbol{L}$.
- Let $\boldsymbol{\Sigma} = \boldsymbol{L}\boldsymbol{L}^H$.
- Create noisy sample: $y = x + \boldsymbol{L}\nu$ for all pixels, $\nu \sim \mathcal{CN}(0, \boldsymbol{I})$.

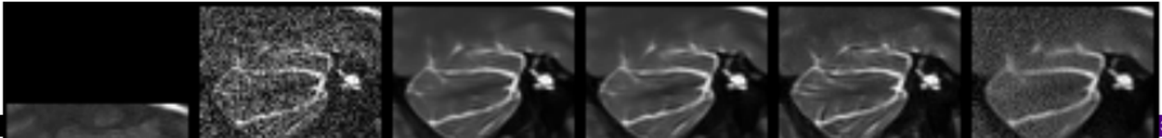
Table: Synthetic Multi-Coil Noise with Coil-Combined Denoising on MoDL Brain Dataset. PSNR / 100×SSIM shown.

Loss	Noise-samples per-image	Non-Whitened	$\hat{\Sigma}$-Whitened
Supervised	n/a	31.1 / 90.3	**32.3** / **92.7**
Avg2Avg	2	31.1 / 89.9	**32.3** / **92.7**
SURE	1	30.4 / 88.9	31.5 / 91.8
Coil2Coil	1	-	32.0 / 92.0

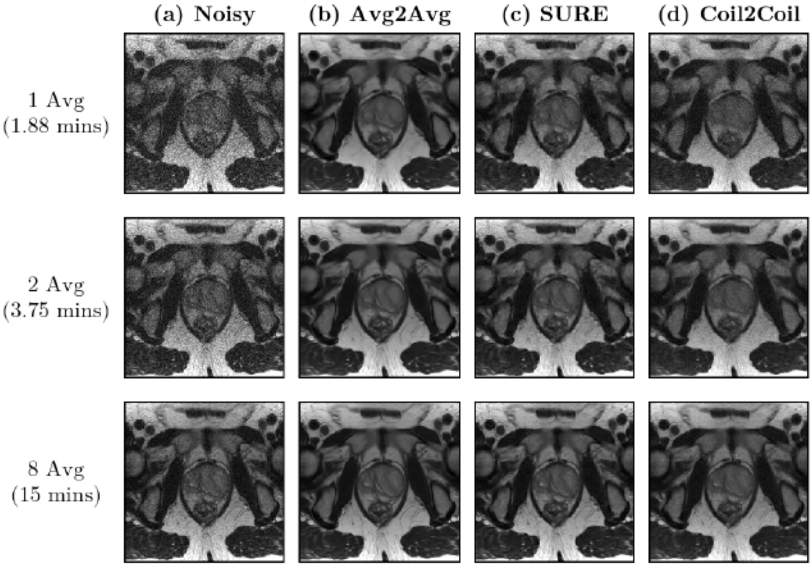
- Whitening improves results
- Avg2Avg equivalent to Supervised

Synthetic Denoising

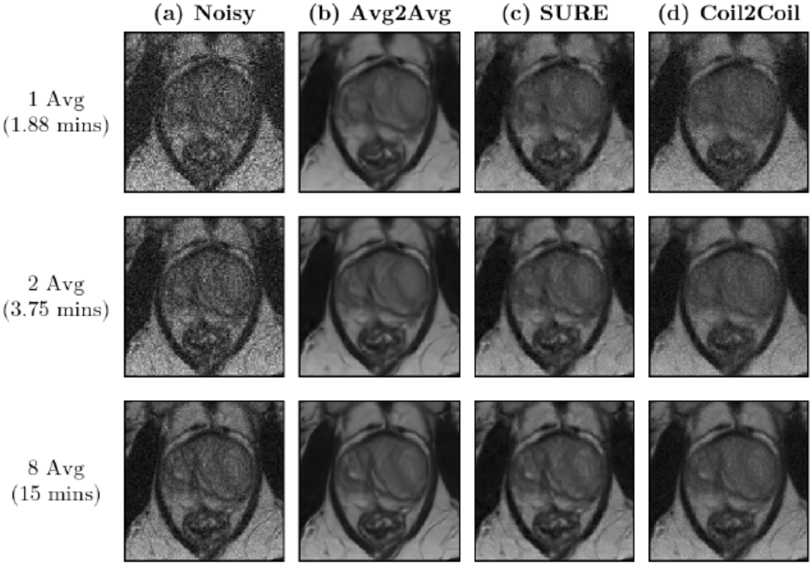




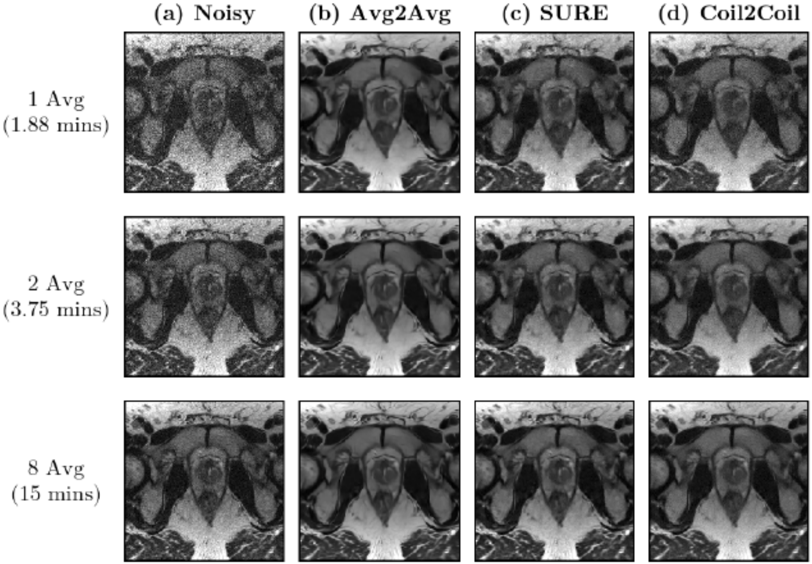
0.55T T2w Prostate Image Denoising



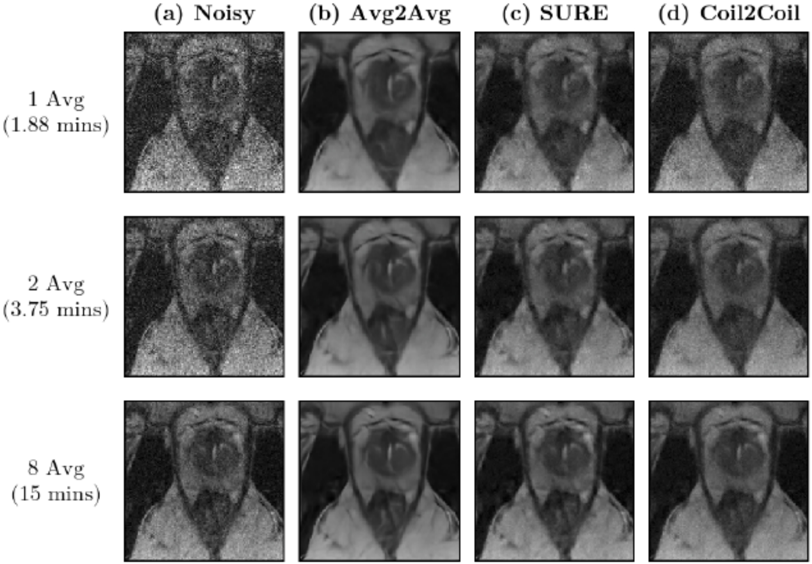
0.55T T2w Prostate Image Denoising



0.55T T2w Prostate Image Denoising



0.55T T2w Prostate Image Denoising

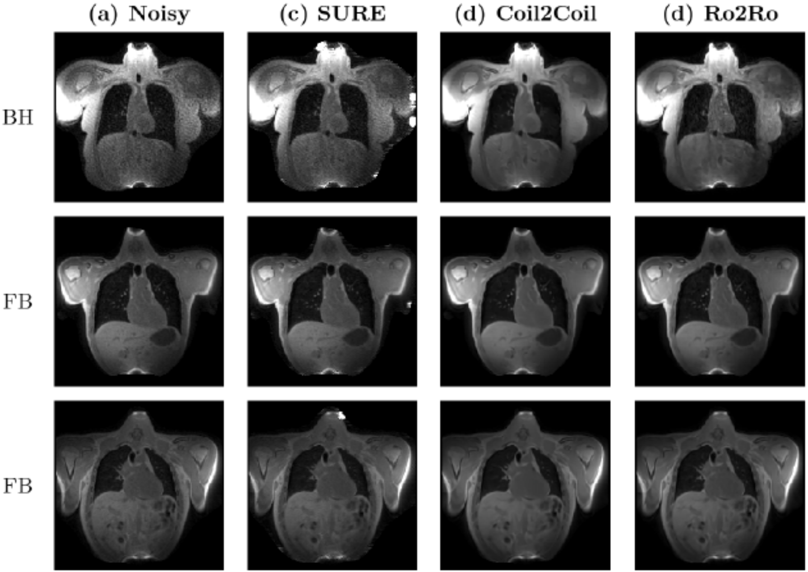


Denoising After GRAPPA Reconstruction

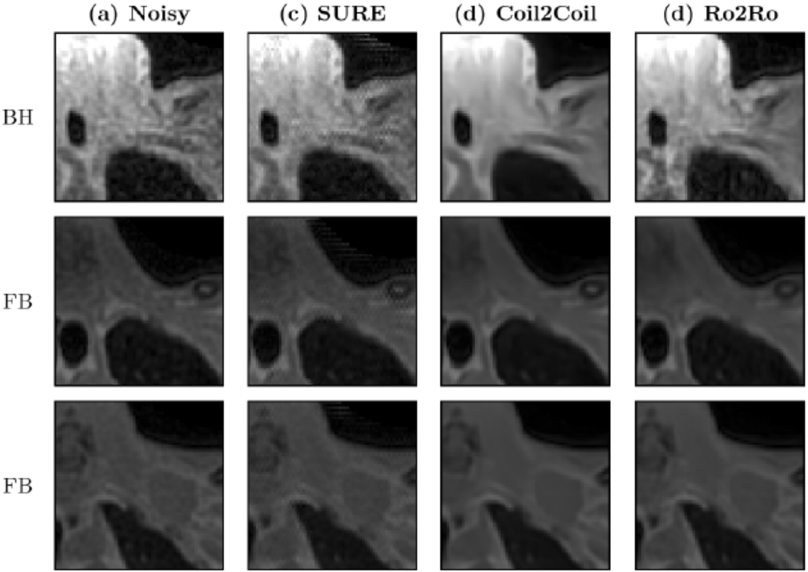
Prostate coil data went through 2x GRAPPA (linear) k-space filling:

- Coil data $(\boldsymbol{k}, \boldsymbol{y})$ is no longer i.i.d. and noise-level estimation is invalid.
- Only Avg2Avg does not require noise-level estimation.

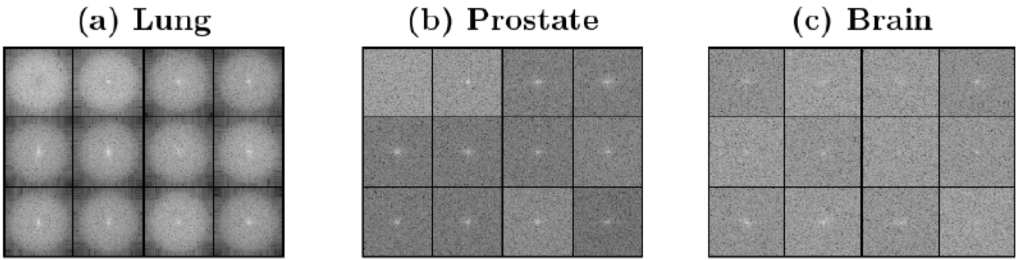
0.55T Lung Image Denoising



0.55T Lung Image Denoising



K-space Analysis



- Source of error: crude noise-level estimation (NLE)
  - Lung: must take into account spiral trajectory (SURE)
  - Prostate: must take into account GRAPPA recon (SURE, Coil2Coil)
- Avg2Avg is the most forgiving method
- Coil2Coil not sensitive to NLE scale-factor
- Ro2Ro preliminary results are promising

Outline

1. Motivation
2. Noise and Observation Model
3. Deep Denoiser Architecture
4. Learning Deep MRI Denoisers
5. Summary and Future Work

Summary and Future Work

- Developed SNAC-DL for Self-Supervised MRI Denoising
- No dynamic (time) dimension required
- Tested in Synthetic Data and Low-Field Lung and Prostate Images
- Proposed different self-supervised losses for MRI
  - Avg2Avg
  - Coil2Coil
  - Ro2Ro

- Sinogram-N2N for Radial Samping, Propeller Sampling Denoising
- Joint Compressed-Sensing and Denoising
- T2w-Guided Diffusion-Weighted MRI Denoising

Acknowledgements

- Hersh Chandarana, MD
- Jingjia Chen, PhD
- Mary Bruno
- Haoyang Pei, MS
- Mahesh Keerthivasan, PhD
- Angela Tong, MD
- Terlika Sood, PhD
- Christoph Maier, MD
- Gloria Huang
- Patricia Johnson, PhD
- Daniel K Sodickson, MD, PhD