Machine Learning for CS MRI: From Model-Based Methods to Deep Learning

Bihan Wen

Nanyang Assistant Professor

School of Electrical and Electronic Engineering (EEE)
Nanyang Technological University (NTU)
Artificial Intelligence (AI): Data-Driven Models

Data-Driven Models > Analytical Models
• Artificial Intelligence (AI): Data-Driven Models

• Machine Learning

Deep Learning, Optimization, Sparse Coding, etc.
• Artificial Intelligence (AI): Data-Driven Models

• Machine Learning

• Solutions with Mathematical Analysis

Mathematical Analysis, Convergence Guarantee, etc.
• Artificial Intelligence (AI): Data-Driven Models

• Machine Learning

• Solutions with Mathematical Analysis

• Applications with State-of-the-art Results
Computer Vision vs. Image Reconstruction

Computer Vision
- Image Analysis

Image Reconstruction
- Image Restoration
- Computational Imaging
- Video Analysis
Computer Vision vs. Image Formation

Computer Vision

Image Reconstruction

Classification + Localization

understanding

Sensing
Outline

• Compressed Sensing MRI

• Why Do We Need Data Models?

• Tutorial on Transform Learning (TL) for MRI

• From Model-Based Method to Deep Learning
Magnetic Resonance Imaging (MRI)

MR sampling → MR image reconstruction
CS MRI

MR sampling

MR image reconstruction

Image Space

K-Space
Why MRI?

- Non-invasive
Why MRI?

- Non-invasive
- Non-ionizing
Why MRI?

- Non-invasive
- Non-ionizing
- Variety of Contrast and Visualization
Why Compressed Sensing (CS)?

- Scan time is too long
Why Compressed Sensing (CS)?

- Scan time is too long
- Image Resolution
CS MRI

$x$: MR image

$F_u x = y$

$y$: k-Space measurement

MR sampling
CS MRI

MR sampling

MR image reconstruction
CS MRI

Naïve Reconstruction

under-sampling

CS: better reconstruction via image modeling
Compressed Sensing MRI

Transform-domain Sparsity

\[
\hat{x} = \arg\min_x \| \Psi x \|_0 \quad \text{s.t.} \quad F_u x = y
\]

Sparsity as the regularizer

\[
\hat{x} = \arg\min_x \left\| \Psi x \right\|_0 + \nu \| F_u x - y \|_2^2
\]
What makes images look like images?

How to distinguish desired pattern from others?
Why do we need Data Model?

Data Models
Why do we need Data Model?

Machine Learning

Simple Neural Network

Deep Learning Neural Network

Input Layer

Hidden Layer

Output Layer

Shallow Methods

Transform Learning

Dictionary Learning

PCA

GMM

Data Models

Deep Methods

CNN

RNN

FCN

U-Net

GAN

Transfer Learning

Machine Learning
A vector $x \in \mathbb{R}^n$ is **sparse** $\iff$ Most of its coefficients are equal to zero.

**Define:** $\|x\|_0 = \text{number of non-zero coefficients in } x$.

Dense signal may be sparse in certain transform domain.

**Example:** $x(n) = \sin\left(\frac{2\pi n}{128}\right)$,
- Sinusoids are sparse in DFT domain.
Sparsity

Natural Image

2D Discrete Cosine Transform (2D DCT)

Highly sparse DCT coefficients
Sparsity

i.i.d. White Gaussian Noise

2D Discrete Cosine Transform (2D DCT)

i.i.d. White Gaussian
Sparsity and Beyond

Level-set-based CS
Ye et al.

Sparse MRI
Lustig et al.

TL-MRI
Ravishankar & Bresler

PANO
Qu et al.

Deep MRI
Wang et al.

STROLLR-MRI
Wen et al.

V-Net
Hammerink et al.

PSF Imaging
Liang

DL-MRI
Ravishankar & Bresler

LOST
Akgäne et al.

PBDW
Qu et al.

LORAKS
Haldar

LASSI
Ravishankar et al.

ADMM-Net
Yang et al.

DeepRes-Net MRI
Lee et al.

Classic CS MRI

Semi-Adaptive CS MRI

Learning-Based CS MRI

Deep Learning CS MRI

<table>
<thead>
<tr>
<th>Methods</th>
<th>Sparse Model</th>
<th>Block Matching</th>
<th>Supervised Learning</th>
<th>Low-Rank Modeling</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sparse MRI [5]</td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PBDW [19]</td>
<td>✓</td>
<td>✓</td>
<td></td>
<td></td>
</tr>
<tr>
<td>LORAKS [23]</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PANO [20]</td>
<td>✓</td>
<td>✓</td>
<td></td>
<td></td>
</tr>
<tr>
<td>DLMRI [6]</td>
<td></td>
<td></td>
<td></td>
<td>✓</td>
</tr>
<tr>
<td>SOUPDIL-MRI [28]</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>LASSI [22]</td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>STL-MRI [24]</td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Frist-MRI [27]</td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>STROLLR-MRI [12]</td>
<td>✓</td>
<td>✓</td>
<td></td>
<td>✓</td>
</tr>
<tr>
<td>ADMM-Net [8]</td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>BCD-Net [9, 33]</td>
<td>✓</td>
<td></td>
<td></td>
<td>✓</td>
</tr>
</tbody>
</table>

[Wen et al., SPM 2020]
Transform-Learning (TL) based regularizer

\[
\hat{x} = \arg\min_x \| \Psi x \|_0 + \nu \| F_u x - y \|_2^2
\]

\[
\hat{x} = \arg\min_x \| F_u x - y \|_2^2 + R_{TL}(x)
\]

Transform-Learning (TL) based regularizer

[Wen et al., SPM 2020]
TL-based MRI

\( \mathcal{R}_{TL}(x) \)

1. Sparsifying Transform Learning (STL)
2. Unitary Transform Learning (UT)
3. Learning a UNION of Transforms (UNITE)
4. Flipping and Rotation Invariant Sparsifying Transform (FRIST)
5. Sparsifying TRansfOrm Learning and Low-Rankness (STROLLLR)

[Wen et al., SPM 2020]
Transform Learning

Octobos: Learning a Union of Transforms

Online Transform Learning

Sparsifying Transform Learning

FRIST: flipping and rotation invariant Transform Learning

Parent transform

Child transforms

STROLLR: Transform Learning with Low-Rank Regularization

[IJCV 15’, IP 17’, JSTSP 15’, TIP 20’]
1. Sparsifying Transform Learning (STL)

\[
\hat{x} = \arg\min_x \| F_w x - y \|_2^2 + \mathcal{R}_{TL}(x)
\]

\[
\mathcal{R}_{STL}(x) \triangleq \arg\min_W \sum_{i=1}^{N} \left\{ \| WP_i x - b_i \|_2^2 + \tau^2 \| b_i \|_0 \right\} + \frac{\lambda}{2} \| W \|_F^2 - \lambda \log(\det W)
\]

Well-conditioning Regularizer for $W$

1. Coefficient $\lambda$ controls the condition number
2. Prevents trivial solution, i.e., $W = 0$
2. Unitary Transform Learning (UT)

\[
\hat{x} = \arg\min_x \| F_w x - y \|_2^2 + \mathcal{R}_{TL}(x)
\]

\[
\mathcal{R}_{UT}(x) \triangleq \arg\min_W \sum_{i=1}^{N} \left\{ \| WP_i x - b_i \|_2^2 + \tau^2 \| b_i \|_0 \right\} \quad \text{s.t.} \quad W^H W = I_n
\]

1. When \( \lambda \to \infty \), it is equivalent to unitary condition.

2. Wavelets, DCT, Discrete Fourier Transforms are all unitary.

3. Closed-form solution
TL-based MRI

Sensing and Reconstruction

Input

K-space

Under-sampling

Output

MR image

\( x \)

Reconstructed

\( \hat{x} \)

[Wen et al., SPM 2020]

K-space

\( y \)

\( \rightarrow \)

Patch extraction

\( W \)

Image Update

Initialization

Transform Learning Schemes
3. Learning a UNION of Transforms (UNITE)

\[
\hat{x} = \arg\min_x \| F_w x - y \|^2_2 + \mathcal{R}_{TL}(x)
\]

\[
\mathcal{R}_{UNITE}(x) \triangleq \arg\min_{\{b_i\}, \{W_k, C_k\}} \sum_{k=1}^{K} \sum_{i \in C_k} \{ \| W_k P_i x - b_i \|^2_2 + \tau^2 \| b_i \|_0 \}
\]

s.t. \[ W_k^H W_k = I_n, \quad \{C_k\} \in G \quad \forall k. \]

1. A union of transforms \( \{W_k\} \) with the membership \( \{C_k\} \).

2. Patches with similar textures will be grouped together.
4. Flipping and Rotation Invariant Sparsifying Transform (FRISt)

\[
\hat{x} = \arg\min_x \| F_u x - y \|_2^2 + \mathcal{R}_{TL}(x)
\]

\[
\mathcal{R}_{FRISt}(x) \triangleq \arg\min_{W, \{b_i\}, \{C_k\}} \sum_{k=1}^{K} \sum_{i \in C_k} \left\{ \| W \Phi_k P_i x - b_i \|_2^2 + \tau^2 \| b_i \|_0 \right\}
\]

s.t. \( W^H W = I_n, \{C_k\} \in G, \forall k, \)

1. Pre-defined operators \( \{\Phi_k\}; \) One parent transform \( W. \)

2. Prevent overfitting; handles rotation and flipping.
5. Sparsifying TRansfOrm Learning and Low-Rankness (STROLLR)

\[ \hat{x} = \arg\min_x \| F_u x - y \|_2^2 + \mathcal{R}_{TL}(x) \]

\[ \mathcal{R}_{TL}(x) = \mathcal{R}_{STROLLR}(x) \triangleq \gamma^L R \mathcal{R}_{LR}(x) + \gamma^S \mathcal{R}_S(x) \]

\[ \mathcal{R}_S(x) = \min_{\{b_i\}, W} \sum_{i=1}^N \left\{ \| W C_i x - \tilde{b}_i \|_2^2 + \tau^2 \| \tilde{b} \|_0 \right\} \quad \text{s.t.} \quad W^HW = I_{nl} \]

\[ \mathcal{R}_{LR}(x) = \min_{\{D_i\}} \sum_{i=1}^N \left\{ \| V_i x - D_i \|_F^2 + \theta^2 \text{rank}(D_i) \right\} \]

Transform Learning  
Low-Rank Modelling
A unified framework for TL-based MRI

[Wen et al., SPM 2020]
Can you combine any priors, and always gain?

Combine only the complementary priors / image models

[Wen et al., Arxiv 2020]
Why Compressed Sensing (CS)?

Combine only the complimentary priors / image models

[Wen et al., SPM 2020]
Deep Learning

Why deep learning

Performance vs. Amount of data

Deep learning

Older learning algorithms
Deep Learning
Connection to Unrolled Neural Networks

1. Improved Performance
2. Robust to Corruptions
Unrolled Transform Learning for MRI

- **Unrolled TL-MRI**

- **Multi-Layer Transform Residual Learning**

[Ravishankar et al., ISBI 2018]
Some Results

Ground Truth
Example A

Sparse MRI
(39.07 dB)

PANO
(41.61 dB)

DL-MRI
(41.73 dB)

STL-MRI
(41.95 dB)

STROLLR-MRI
(43.27 dB)

[Wen et al., SPM 2020]
Some Results

Ground Truth
Example B

Sparse MRI
(28.03 dB)

PANO
(30.03 dB)

DL-MRI
(29.74 dB)

ADMM-Net
(30.67 dB)

STROLLR-MRI
(32.46 dB)

[Wen et al., SPM 2020]
**Conventional**

- Shallow Model
  - Equivalently one free layer

**Deep Learning**

- Deep Model
  - Multiple free layers
**Conventional**
- Shallow Model
  - One free layer
- Unsupervised
  - No training corpus needed
  - Data efficient

**Deep Learning**
- Deep Model
  - Multiple free layers
- Supervised
  - Training corpus needed
  - Data inefficient

Representation Learning
# Representation Learning

<table>
<thead>
<tr>
<th><strong>Conventional</strong></th>
<th><strong>Deep Learning</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td>Shallow Model</td>
<td>Deep Model</td>
</tr>
<tr>
<td>• One free layer</td>
<td>• Multiple free layers</td>
</tr>
<tr>
<td>Unsupervised</td>
<td>Supervised</td>
</tr>
<tr>
<td>• No training corpus needed</td>
<td>• Training corpus needed</td>
</tr>
<tr>
<td>• Data efficient</td>
<td>• Data inefficient</td>
</tr>
<tr>
<td>Prior-based</td>
<td>Generic</td>
</tr>
<tr>
<td>• Assumption &amp; Understanding of the Data</td>
<td>• Little assumption</td>
</tr>
<tr>
<td>• Regularizer &amp; structures of the Model</td>
<td>• Almost free model</td>
</tr>
</tbody>
</table>
Thank you! Questions??