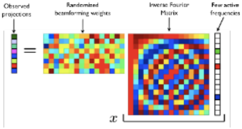
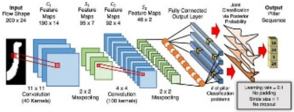
Machine Learning for CS MRI:
From Model-Based Methods to Deep Learning

Bihan Wen

Nanyang Assistant Professor

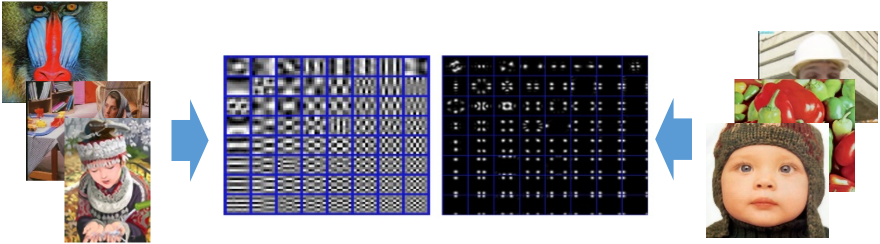
School of Electrical and Electronic Engineering (EEE)
Nanyang Technological University (NTU)





About me

- Artificial Intelligence (AI): Data-Driven Models

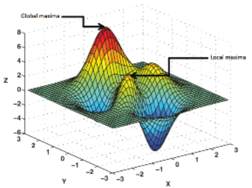


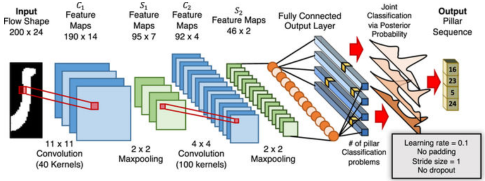
Data-Driven Models > Analytical Models

About me

- Artificial Intelligence (AI): Data-Driven Models

- Machine Learning

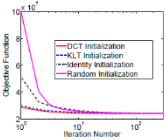




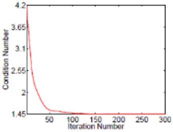
Deep Learning, Optimization, Sparse Coding, etc.

About me

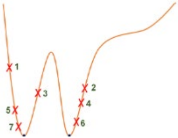
- Artificial Intelligence (AI): Data-Driven Models
- Machine Learning
- Solutions with Mathematical Analysis

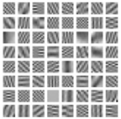


Objective Function



$\kappa(W)$ - Random Init

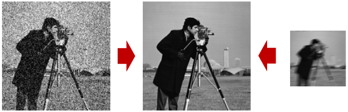



Learnt W - Random Init

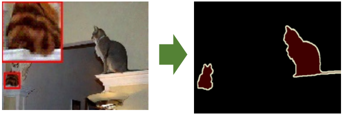
Mathematical Analysis, Convergence Guarantee, etc.

About me

- Artificial Intelligence (AI): Data-Driven Models
- Machine Learning
- Solutions with Mathematical Analysis
- Applications with State-of-the-art Results

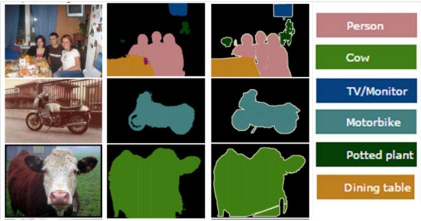


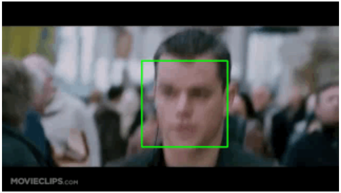
Restoration



Analysis

Computer Vision vs. Image Reconstruction

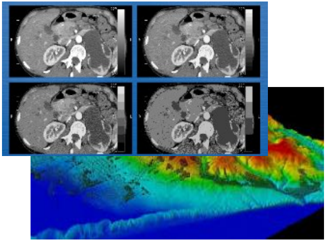




Computer Vision

- Image Analysis
- Video Analysis

Image Reconstruction

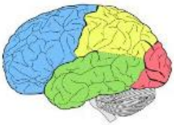


- Image Restoration
- Computational Imaging

Computer Vision vs. Image Formation

“Cat”

Classification
+ Localization



understanding

Computer
Vision

Image
Reconstruction

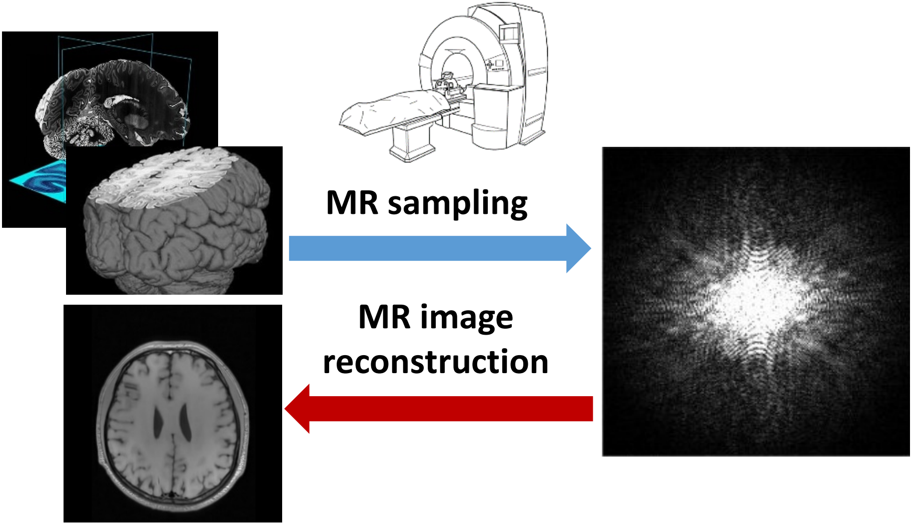
90	91	92	93	94	95	96	97	98	99
80	81	82	83	84	85	86	87	88	89
70	71	72	73	74	75	76	77	78	79
60	61	62	63	64	65	66	67	68	69
50	51	52	53	54	55	56	57	58	59
40	41	42	43	44	45	46	47	48	49
30	31	32	33	34	35	36	37	38	39
20	21	22	23	24	25	26	27	28	29
10	11	12	13	14	15	16	17	18	19
0	1	2	3	4	5	6	7	8	9

Sensing

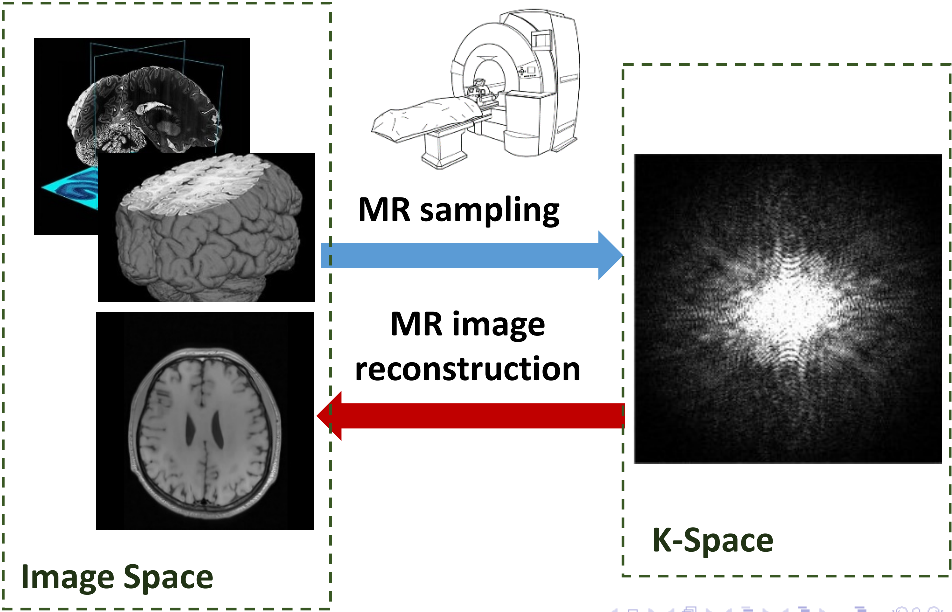
Outline

- 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)



CS MRI



Why MRI?

- Non-invasive

Why MRI?

- Non-invasive

- Non-ionizing

Why MRI?

- Non-invasive
- Non-ionizing
- Variety of Contrast and Visualization

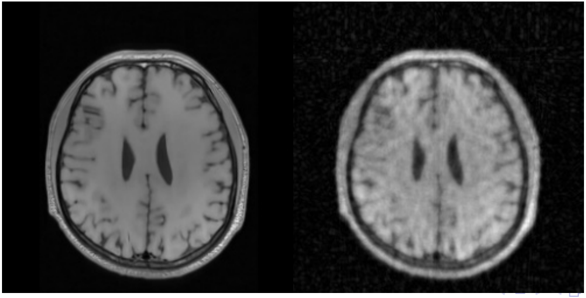
Why Compressed Sensing (CS)?

- **Scan time is too long**

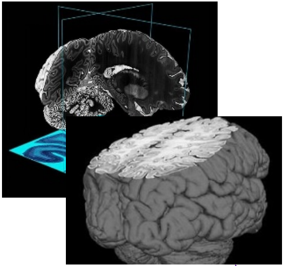
Why Compressed Sensing (CS)?

- **Scan time is too long**

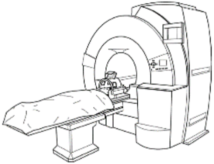
- **Image Resolution**



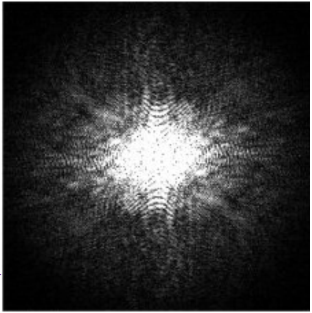
CS MRI



x: MR image



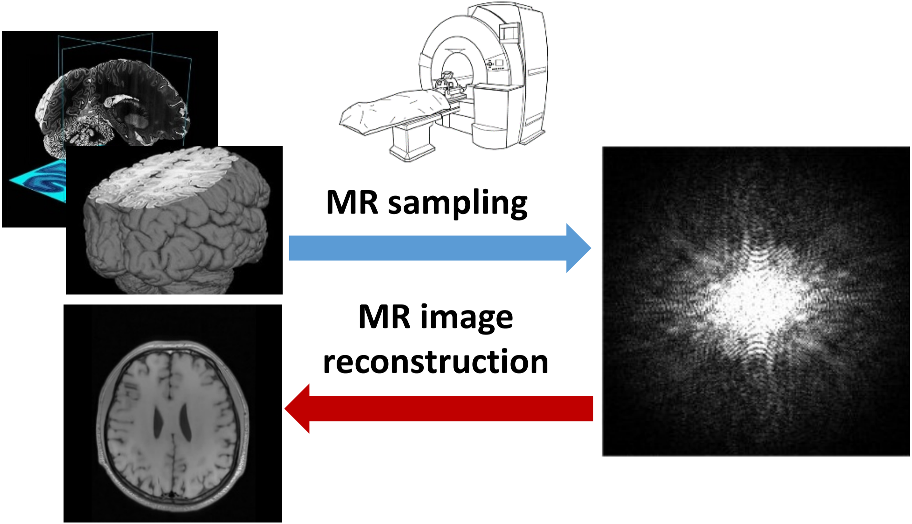
MR sampling



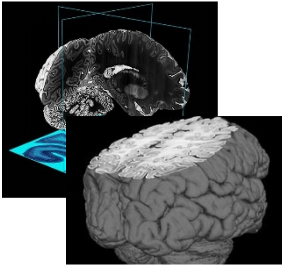
y: k-Space measurement

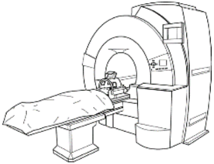
$$\boldsymbol{F}_u \boldsymbol{x} = \boldsymbol{y}$$

CS MRI

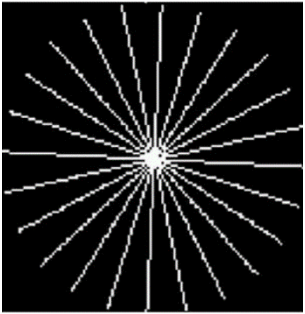


CS MRI

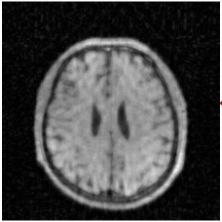




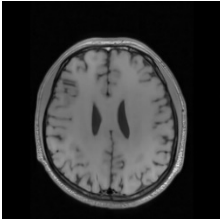
under-sampling



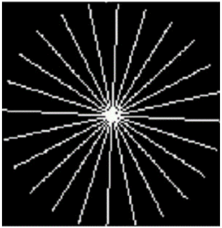
Naïve Reconstruction



CS: better reconstruction via image modeling



Compressed Sensing MRI



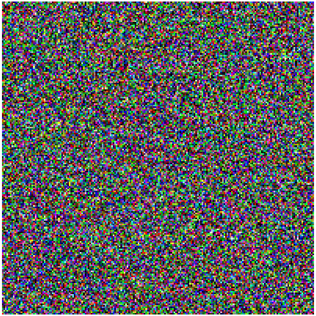
Transform-domain Sparsity

$$\hat{\boldsymbol{x}} = \underset{\boldsymbol{x}}{\operatorname{argmin}} \| \Psi \boldsymbol{x} \|_0 \quad \text{s.t.} \quad \boldsymbol{F}_u \boldsymbol{x} = \boldsymbol{y}$$

$$\hat{\boldsymbol{x}} = \underset{\boldsymbol{x}}{\operatorname{argmin}} \| \Psi \boldsymbol{x} \|_0 + v \| \boldsymbol{F}_u \boldsymbol{x} - \boldsymbol{y} \|_2^2$$

Sparsity as the regularizer

Why do we need Data Model?



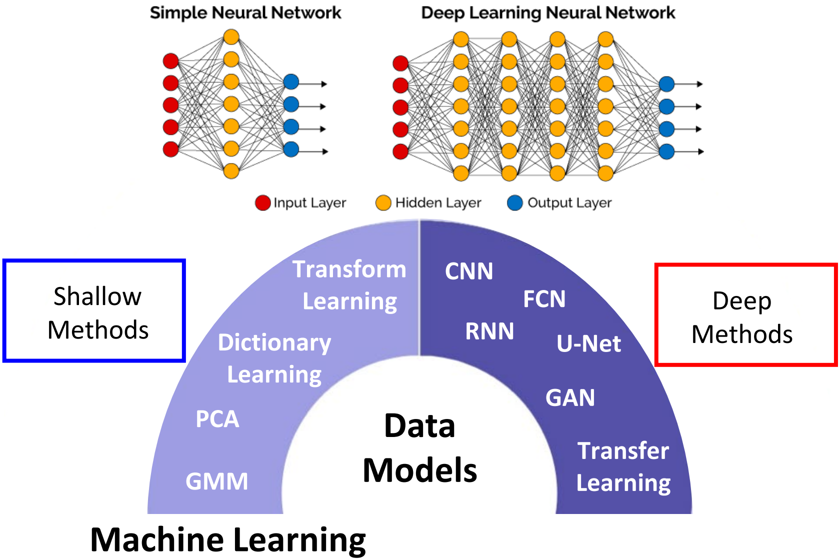
What makes images look like images?

How to distinguish desired pattern from others?

Why do we need Data Model?

Data Models

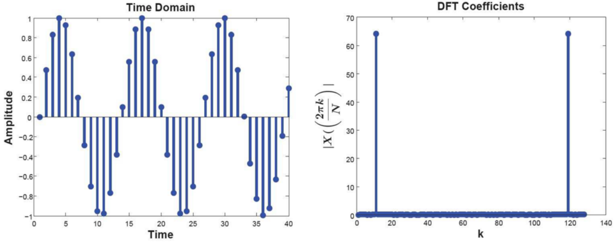
Why do we need Data Model?



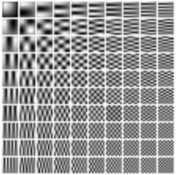
Sparsity

- A vector $x \in \mathbb{R}^n$ is **sparse** $\Leftrightarrow$ Most of its coefficients are equal to zero.
- **Define:** $\|x\|_0$ = number of non-zero coefficients in x.
- Dense signal may be sparse in certain transform domain.

- **Example:** $x(n) = \sin(\frac{2\pi 10}{128} n)$,
  - **Sinusoids** are sparse in DFT domain.

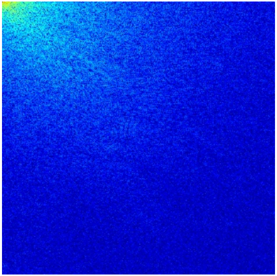


Sparsity



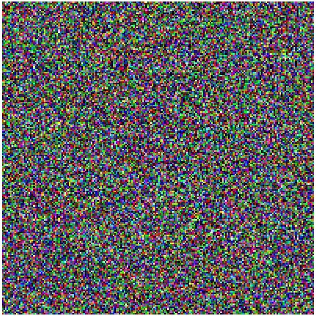
2D Discrete Cosine Transform (2D DCT)

Natural Image

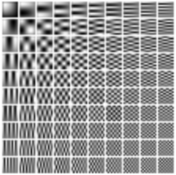


Highly sparse DCT coefficients

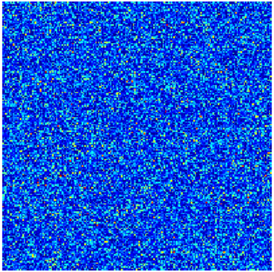
Sparsity



i.i.d. White Gaussian Noise

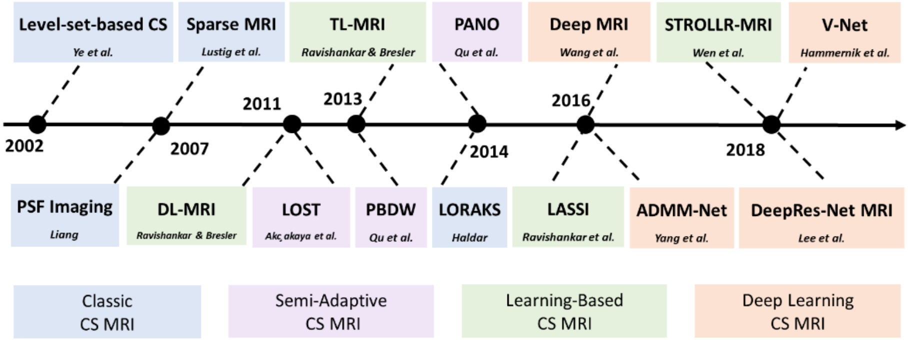


2D Discrete Cosine Transform (2D DCT)



i.i.d. White Gaussian

Sparsity and Beyond



Methods	Sparse Model				Block Matching	Supervised Learning	Low-Rank Modeling
	Fixed	Directional	DL	TL			
Sparse MRI [5]	✓						
PBDW [19]	✓	✓					
LORAKS [23]							✓
PANO [20]	✓				✓		
DLMRI [6]			✓				
SOUPDIL-MRI [28]			✓				
LASSI [22]			✓				✓
STL-MRI [24]				✓			
FRIST-MRI [27]		✓		✓			
STROLLR-MRI [12]				✓	✓		✓
ADMM-Net [8]						✓	
BCD-Net [9, 33]				✓		✓	

[Wen et al., SPM 2020]

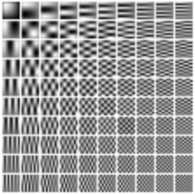
Transform Learning for Better Sparsity

Fixed Transform

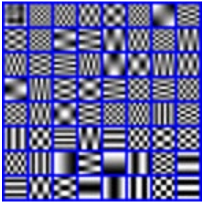
$$\hat{\boldsymbol{x}} = \underset{\boldsymbol{x}}{\operatorname{argmin}} \; \| \Psi \boldsymbol{x} \|_0 + \upsilon \, \| \boldsymbol{F}_u \boldsymbol{x} - \boldsymbol{y} \|_2^2$$

$$\hat{\boldsymbol{x}} = \underset{\boldsymbol{x}}{\operatorname{argmin}} \; \| \boldsymbol{F}_u \boldsymbol{x} - \boldsymbol{y} \|_2^2 + \boxed{\mathfrak{R}_{TL}(\boldsymbol{x})}$$

Transform-Learning (TL) based regularizer



2D DCT



Learned Transform

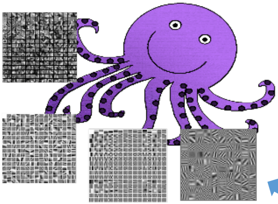
[Wen et al., SPM 2020]

$$\mathfrak{R}_{TL}(\boldsymbol{x})$$

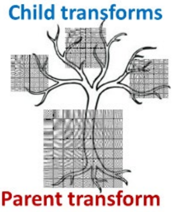
1. Sparsifying Transform Learning (STL)
2. Unitary Transform Learning (UT)
3. Learning a UNIon of Transforms (UNITE)
4. Flipping and Rotation Invariant Sparsifying Transform (FRIST)
5. Sparsifying TRansfOrm Learning and Low-Rankness (STROLLR)

[Wen et al., SPM 2020]

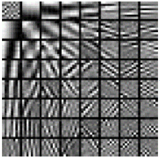
Transform Learning



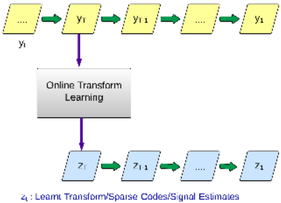
Octobos: Learning a Union of Transforms



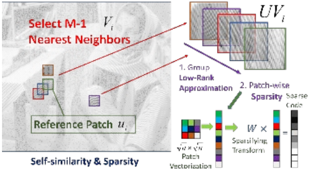
FRIST: flipping and rotation invariant Transform Learning



Sparsifying Transform Learning



Online Transform Learning



STROLLR: Transform Learning with Low-Rank Regularization

[IJCV 15', IP 17', JSTSP 15', TIP 20']

1. Sparsifying Transform Learning (STL)

$$\hat{\boldsymbol{x}} = \underset{\boldsymbol{x}}{\operatorname{argmin}} \ \|\boldsymbol{F}_u\boldsymbol{x} - \boldsymbol{y}\|_2^2 + \boxed{\Re_{TL}(\boldsymbol{x})}$$

$$\Re_{STL}(\boldsymbol{x}) \triangleq \underset{\boldsymbol{W},\{\boldsymbol{b}_i\}}{\operatorname{argmin}} \sum_{i=1}^{N} \{\|\boldsymbol{W}\mathbf{P}_i\boldsymbol{x} - \boldsymbol{b}_i\|_2^2 + \tau^2 \|\boldsymbol{b}_i\|_0\} + \boxed{\frac{\lambda}{2}\|\boldsymbol{W}\|_F^2 - \lambda \log(\det \boldsymbol{W})}$$

Well-conditioning Regularizer for W

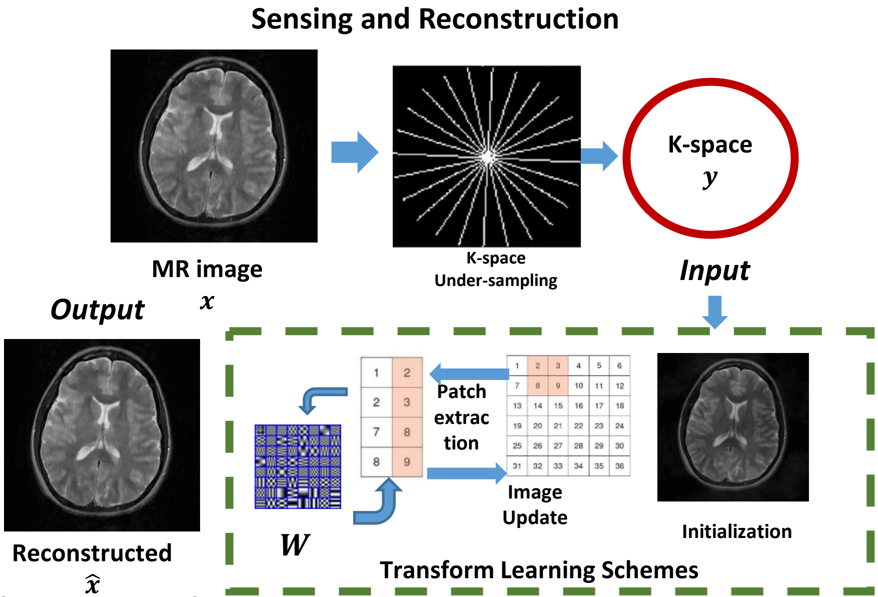
1. Coefficient λ controls the condition number
2. Prevents trivial solution, i.e., $W = 0$

2. Unitary Transform Learning (UT)

$$\hat{\boldsymbol{x}} = \underset{\boldsymbol{x}}{\operatorname{argmin}} \; \|\boldsymbol{F}_u \boldsymbol{x} - \boldsymbol{y}\|_2^2 + \boxed{\mathfrak{R}_{TL}(\boldsymbol{x})}$$

$$\mathfrak{R}_{UT}(\boldsymbol{x}) \triangleq \underset{\boldsymbol{W},\{\boldsymbol{b}_i\}}{\operatorname{argmin}} \sum_{i=1}^{N} \{\|\boldsymbol{W}\mathbf{P}_i\boldsymbol{x} - \boldsymbol{b}_i\|_2^2 + \tau^2 \|\boldsymbol{b}_i\|_0\} \quad \boxed{\text{s.t.} \quad \boldsymbol{W}^H\boldsymbol{W} = \boldsymbol{I}_n}$$

1. When $\lambda \to \infty$, it is equivalent to unitary condition.
2. Wavelets, DCT, Discrete Fourier Transforms are all unitary.
3. Closed-form solution



[Wen et al., SPM 2020]

3. Learning a UNIon of Transforms (UNITE)

$$\hat{\boldsymbol{x}} = \underset{\boldsymbol{x}}{\operatorname{argmin}} \; \|\boldsymbol{F}_u \boldsymbol{x} - \boldsymbol{y}\|_2^2 + \boxed{\Re_{TL}(\boldsymbol{x})}$$

$$\Re_{UNITE}(\boldsymbol{x}) \triangleq \underset{\{\boldsymbol{b}_i\},\{\boldsymbol{W}_k, C_k\}}{\operatorname{argmin}} \sum_{k=1}^{K} \sum_{i \in C_k} \{\|\boldsymbol{W}_k \mathbf{P}_i \boldsymbol{x} - \boldsymbol{b}_i\|_2^2 + \tau^2 \|\boldsymbol{b}_i\|_0\}$$

$$\text{s.t.} \quad \boldsymbol{W}_k^H \boldsymbol{W}_k = \boldsymbol{I}_n \,, \quad \{C_k\} \in G \quad \forall k.$$

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)

$$\hat{\boldsymbol{x}} = \underset{\boldsymbol{x}}{\operatorname{argmin}} \ \|\boldsymbol{F}_u \boldsymbol{x} - \boldsymbol{y}\|_2^2 + \boxed{\Re_{TL}(\boldsymbol{x})}$$

$$\Re_{FRIST}(\boldsymbol{x}) \triangleq \underset{\boldsymbol{W},\{\boldsymbol{b}_i\},\{C_k\}}{\operatorname{argmin}} \sum_{k=1}^{K} \sum_{i \in C_k} \{\|\boldsymbol{W} \boldsymbol{\Phi}_k \, \mathbf{P}_i \, \boldsymbol{x} - \boldsymbol{b}_i\|_2^2 + \tau^2 \, \| \, \boldsymbol{b}_i \, \|_0\}$$

$$\text{s.t.} \quad \boldsymbol{W}^H \boldsymbol{W} = \boldsymbol{I}_n \,, \ \{C_k\} \in G \quad \forall k,$$

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)

$$\hat{\boldsymbol{x}} = \underset{\boldsymbol{x}}{\operatorname{argmin}} \; \|\boldsymbol{F}_u \boldsymbol{x} - \boldsymbol{y}\|_2^2 + \boxed{\mathfrak{R}_{TL}(\boldsymbol{x})}$$

$$\mathfrak{R}_{TL}(\boldsymbol{x}) = \mathfrak{R}_{STROLLR}(\boldsymbol{x}) \triangleq \gamma^{LR}\, \mathfrak{R}_{LR}(\boldsymbol{x}) + \gamma^{S}\, \mathfrak{R}_{S}(\boldsymbol{x})$$

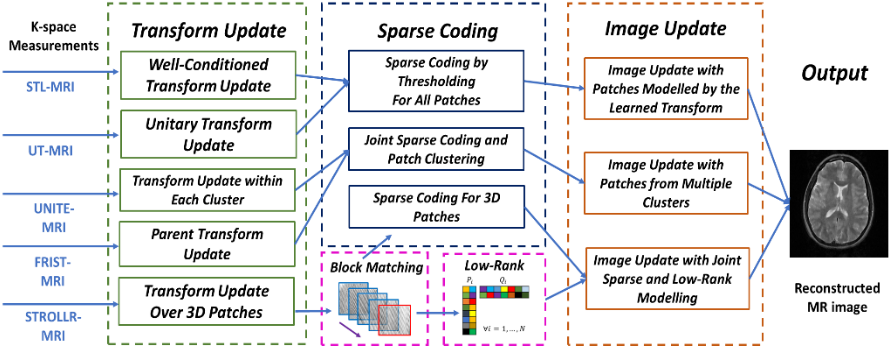
$$\mathfrak{R}_S(\boldsymbol{x}) = \min_{\{\tilde{\boldsymbol{b}}_i\}, \boldsymbol{W}} \sum_{i=1}^{N} \left\{ \left\| \boldsymbol{W}\, \mathbf{C}_i \boldsymbol{x} - \tilde{\boldsymbol{b}}_i \right\|_2^2 + \tau^2 \left\| \tilde{\boldsymbol{b}} \right\|_0 \right\} \;\; \text{s.t.} \;\; \boldsymbol{W}^H \boldsymbol{W} = \boldsymbol{I}_{nl}$$

Transform Learning

$$\mathfrak{R}_{LR}(\boldsymbol{x}) = \min_{\{\boldsymbol{D}_i\}} \sum_{i=1}^{N} \left\{ \|\mathbf{V}_i\, \boldsymbol{x} - \boldsymbol{D}_i\|_F^2 + \theta^2 \operatorname{rank}(\boldsymbol{D}_i) \right\}$$

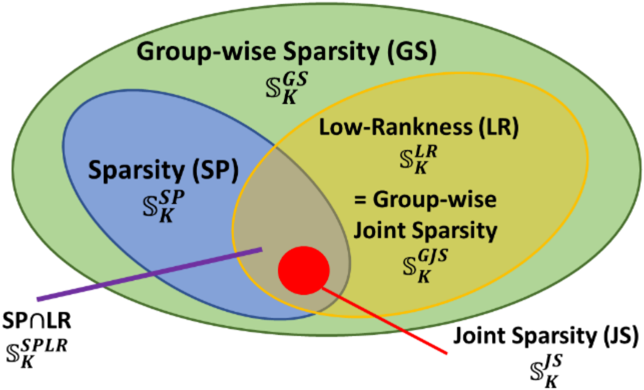
Low-Rank Modelling

A unified framework for TL-based MRI



[Wen et al., SPM 2020]

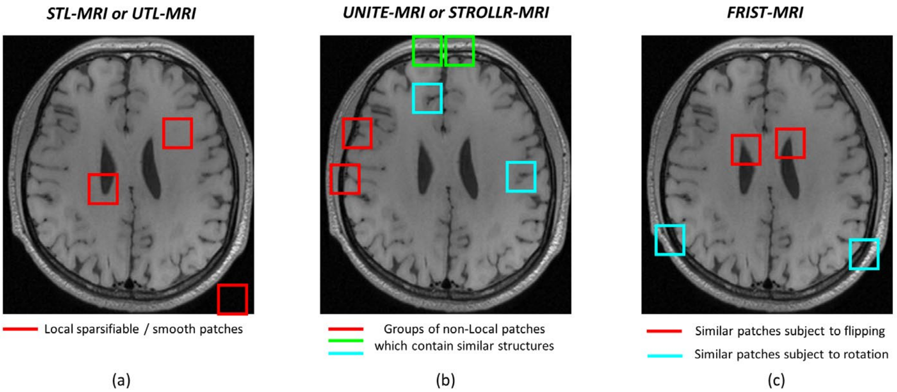
Can you combine any priors, and always gain?



Combine only the complementary priors / image models

[Wen et al., Arxiv 2020]

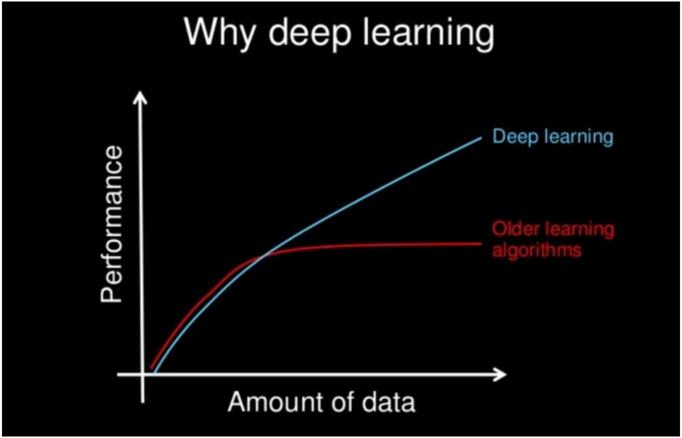
Why Compressed Sensing (CS)?



Combine only the complimentary priors / image models

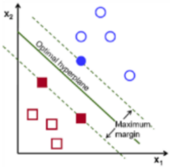
[Wen et al., SPM 2020]

Deep Learning



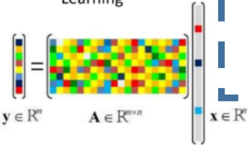
Representation Learning

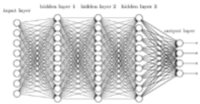
Deep Learning



- SVM
- GMM
- Boosting
- Perceptron
- Decision Tree

- PCA
- Sparse Coding
- Wavelets
- Low-Rankness
- Transform Learning
- Dictionary Learning



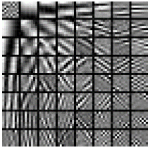


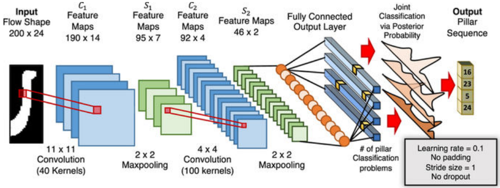
- CNN
- GAN
- RNN

Deep Learning

1990 — 2000 — 2010 — time

Connection to Unrolled Neural Networks

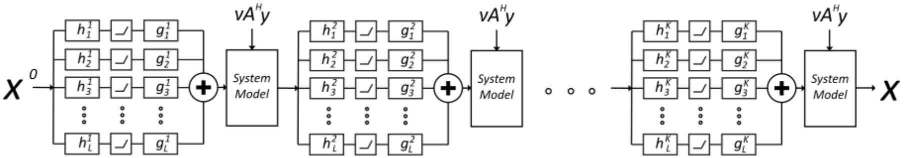




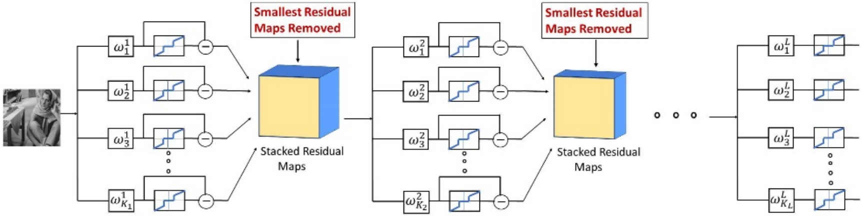
1. Improved Performance
2. Robust to Corruptions

Unrolled Transform Learning for MRI

- Unrolled TL-MRI

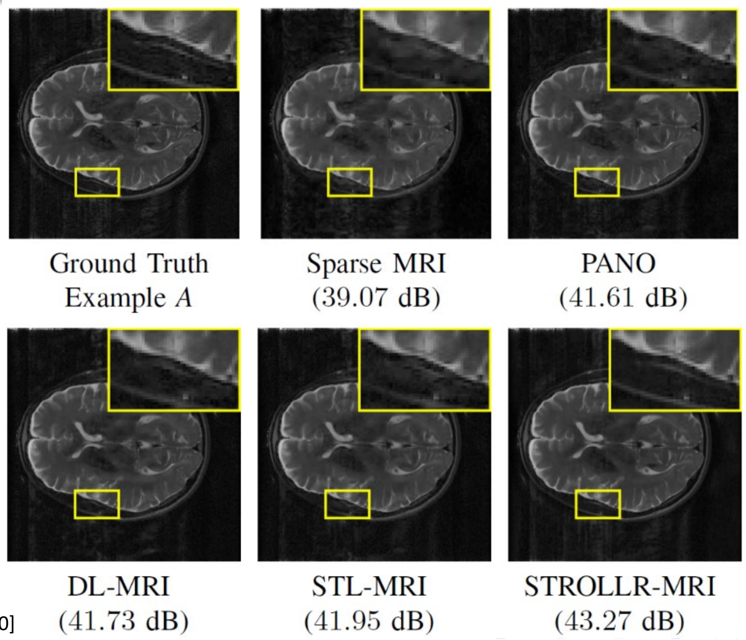


- Multi-Layer Transform Residual Learning



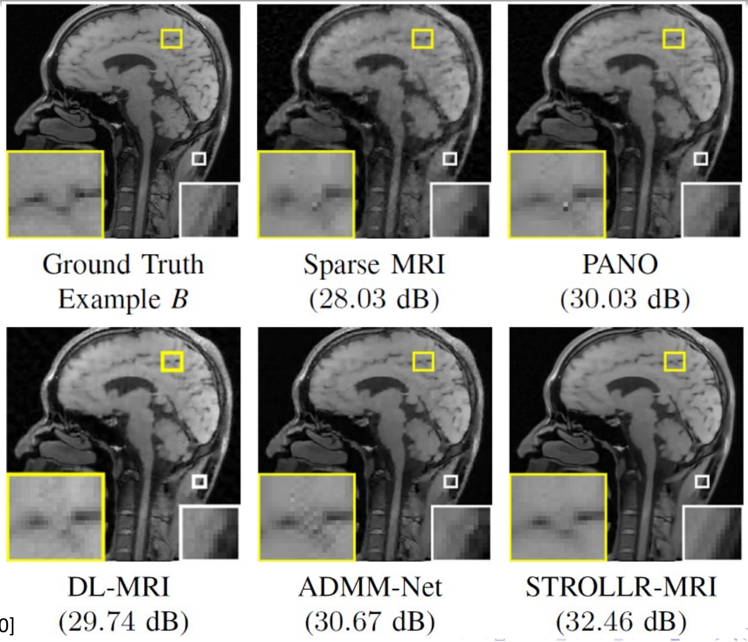
[Ravishankar et al., ISBI 2018]

Some Results



[Wen et al., SPM 2020]

Some Results



[Wen et al., SPM 2020]

Representation Learning

Conventional

- **Shallow Model**
  - **Equivalently one free layer**

Deep Learning

- **Deep Model**
  - **Multiple free layers**

Representation Learning

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

Conventional

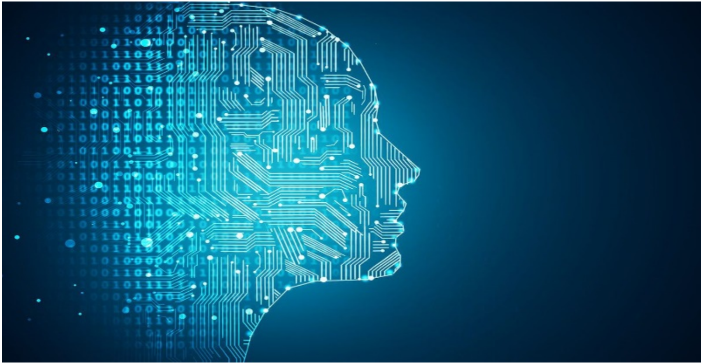
- Shallow Model
  - One free layer
- Unsupervised
  - No training corpus needed
  - Data efficient
- **Prior-based**
  - **Assumption & Understanding of the Data**
  - **Regularizer & structures of the Model**

Deep Learning

- Deep Model
  - Multiple free layers
- Supervised
  - Training corpus needed
  - Data inefficient
- **Generic**
  - **Little assumption**
  - **Almost free model**

?

Thank you! Questions??



Bihan Wen

Homepage: www.bihanwen.com