Geometric Understanding of Supervised and Unsupervised Deep Learning for Biomedical Image Reconstruction

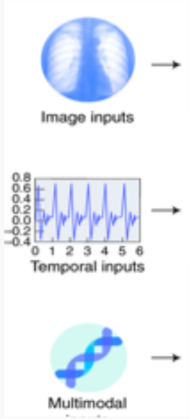
Jong Chul Ye, Ph.D

Professor
BISPL - BioImaging, Signal Processing, and Learning lab.
KAIST, Korea

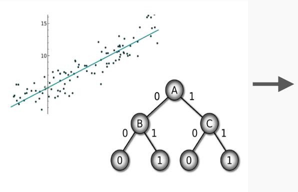
Deep Learning and Medical Applications
JANUARY 27 - 31, 2020

Classical Learning vs Deep Learning

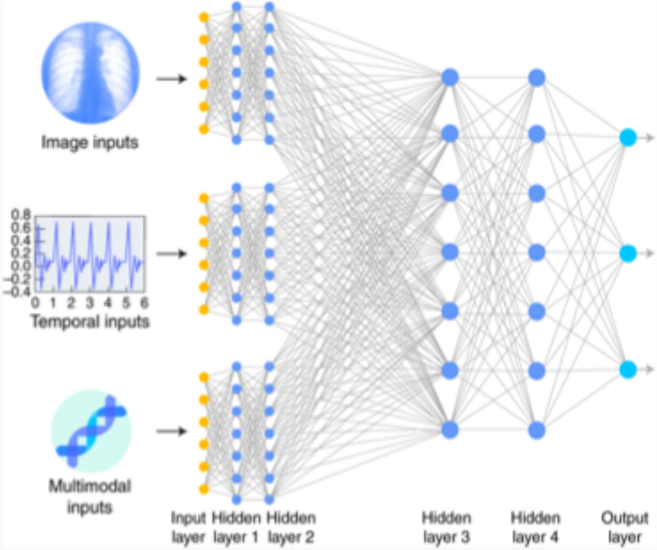
Classical machine learning



Feature Engineering



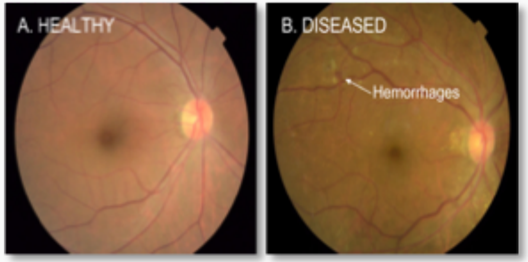
Deep learning
(no feature engineering)



Esteva et al, Nature Medicine, (2019)

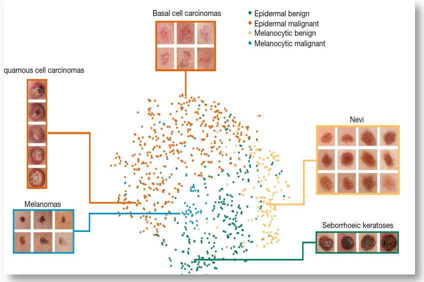
Deep Learning Era in Medical Imaging

Diabetic eye diagnosis



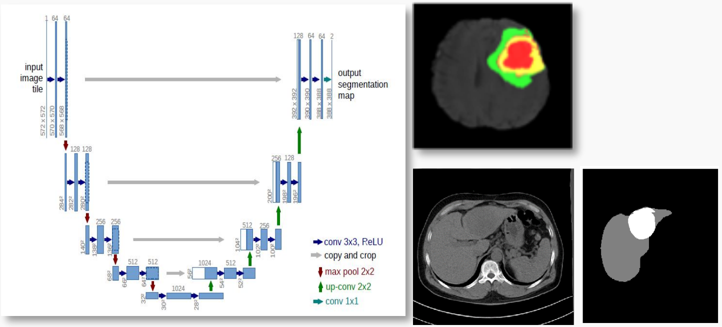
Gulshan, V. et al. JAMA (2016)

Skin Cancer diagnosis



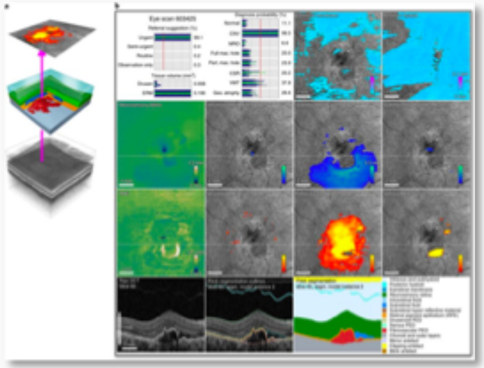
Esteva et al, Nature (2017)

Image segmentation



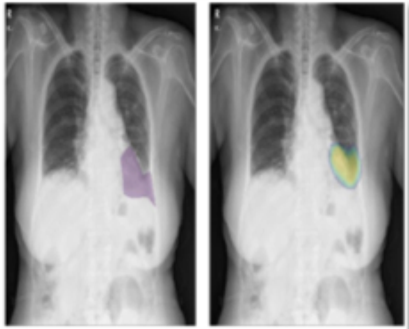
Ronneberger et al, MICCAI, 2015

OCT diagnosis



De Fauw et al, Nature Medicine (2018)

Chest X-ray



Courtesy of Kyu Hwan Jung @Vuno

Image registration

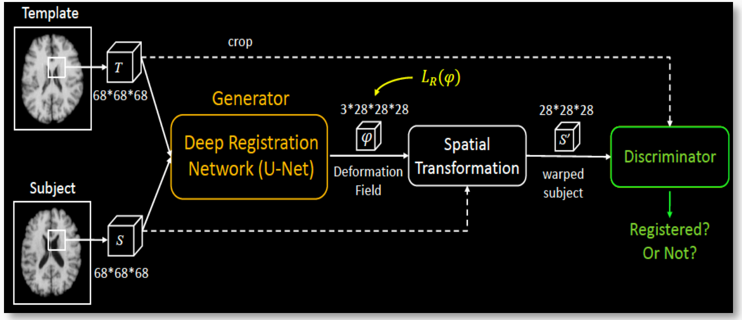
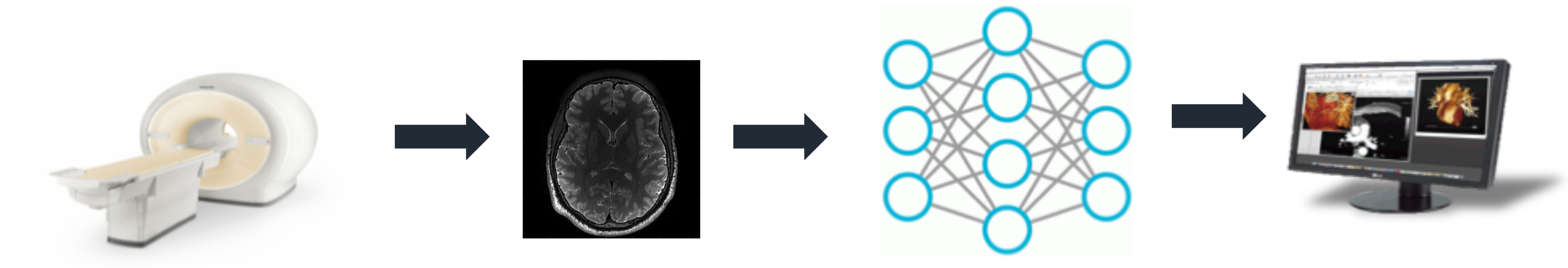


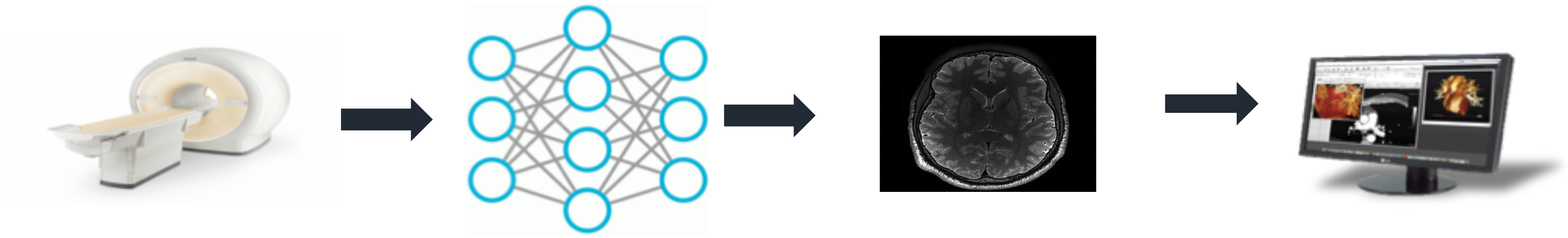
Figure courtesy of X. Cao & D. Shen

Deep Learning for Image Reconstruction

Diagnosis & analysis



Focus of this talk: Image reconstruction



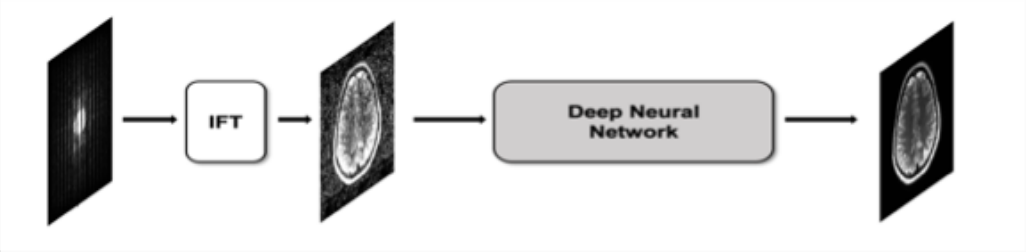
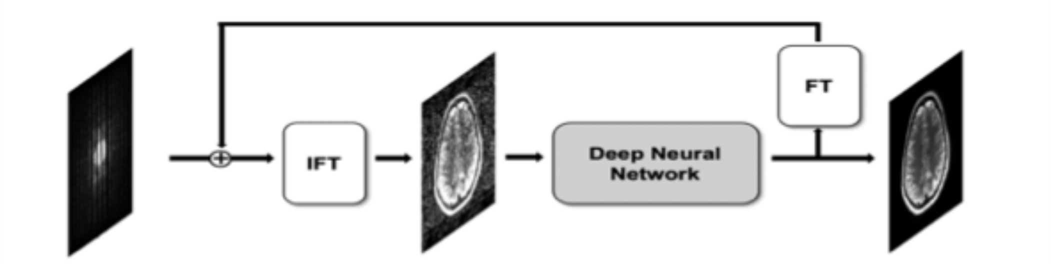
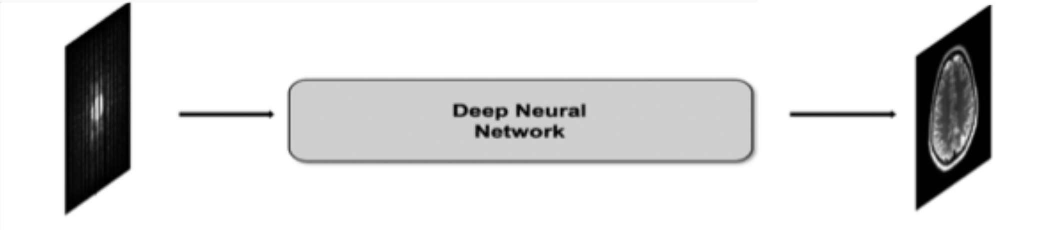


Image-domain learning

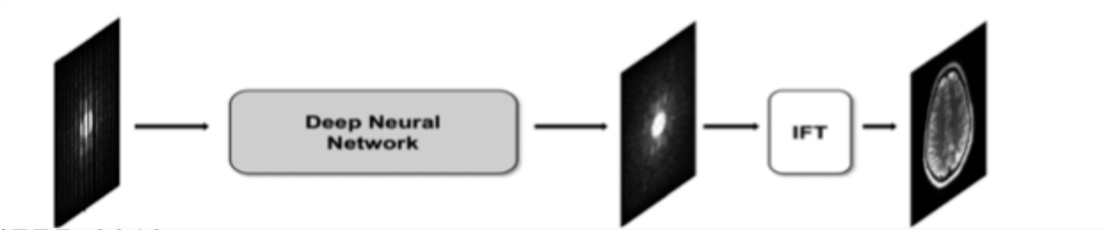
Hybrid-domain learning





Domain-Transform learning

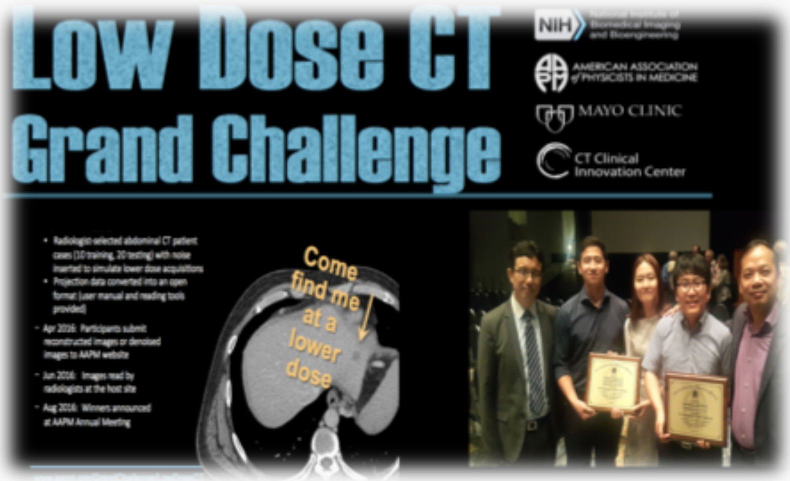
Sensor-domain learning



Han et al, IEEE TMI, 2019, Ravishankar, Proc IEEE, 2019

Deep Learning Revolution for Inverse Problem

- ❑ **Accuracy**: high quality recon > CS
- ❑ **Fast** reconstruction time
- ❑ **Business model:** vendor-driven training
- ❑ **Interpretable** models

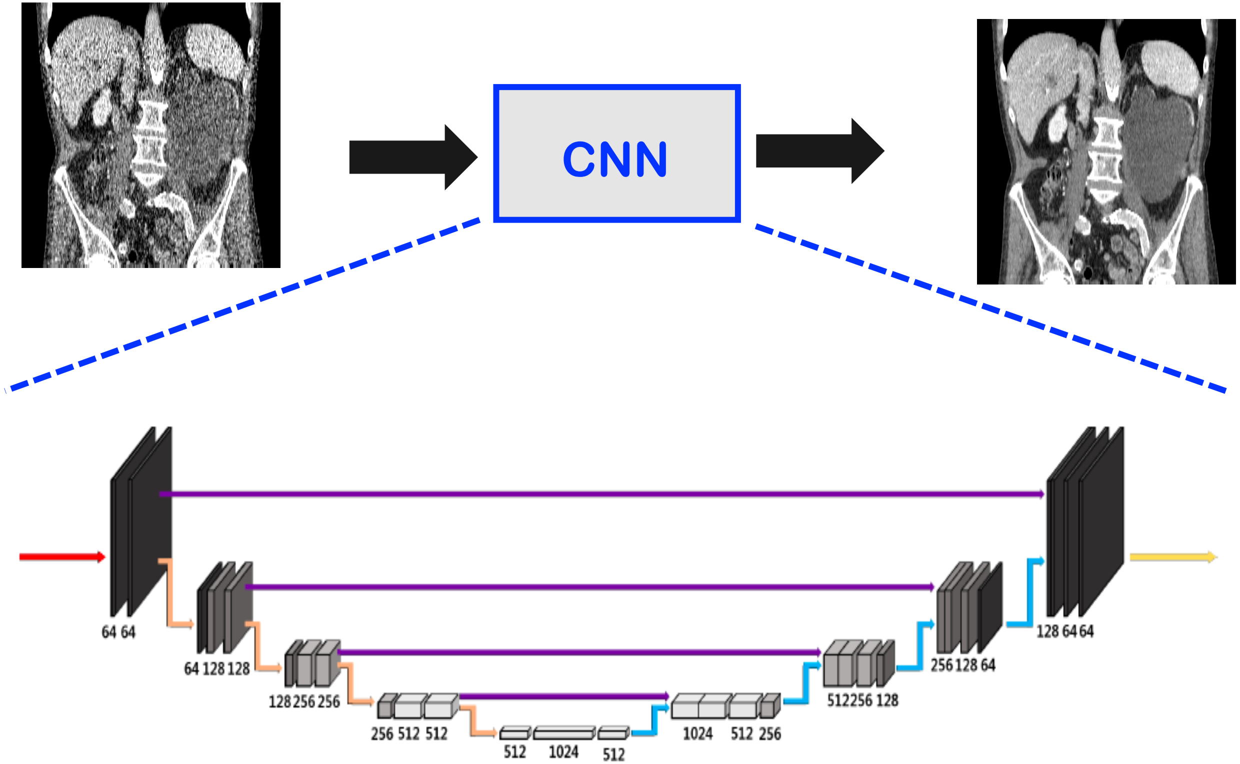


DOES IT CREATE ANY ARTIFICIAL FEATURES ?

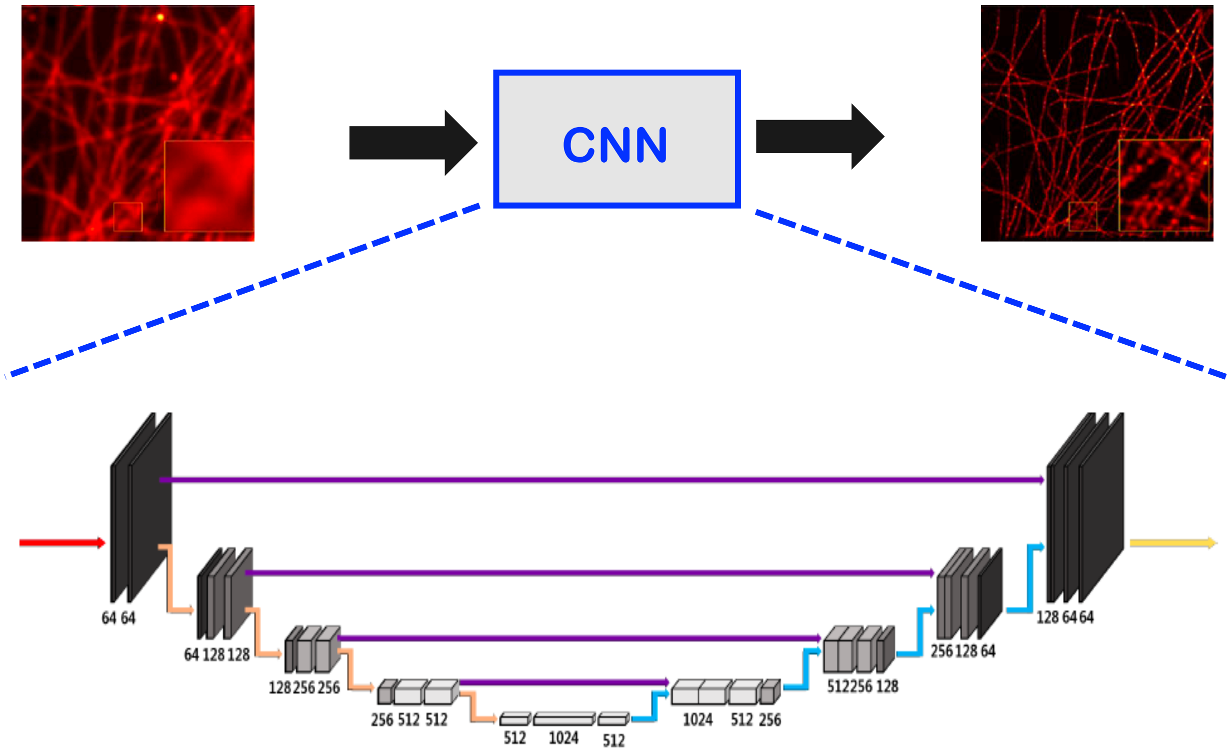
Geometry of Supervised Learning

Ye et al, SIAM J. Imaging Sciences, 2018; Ye et al, ICML, 2019

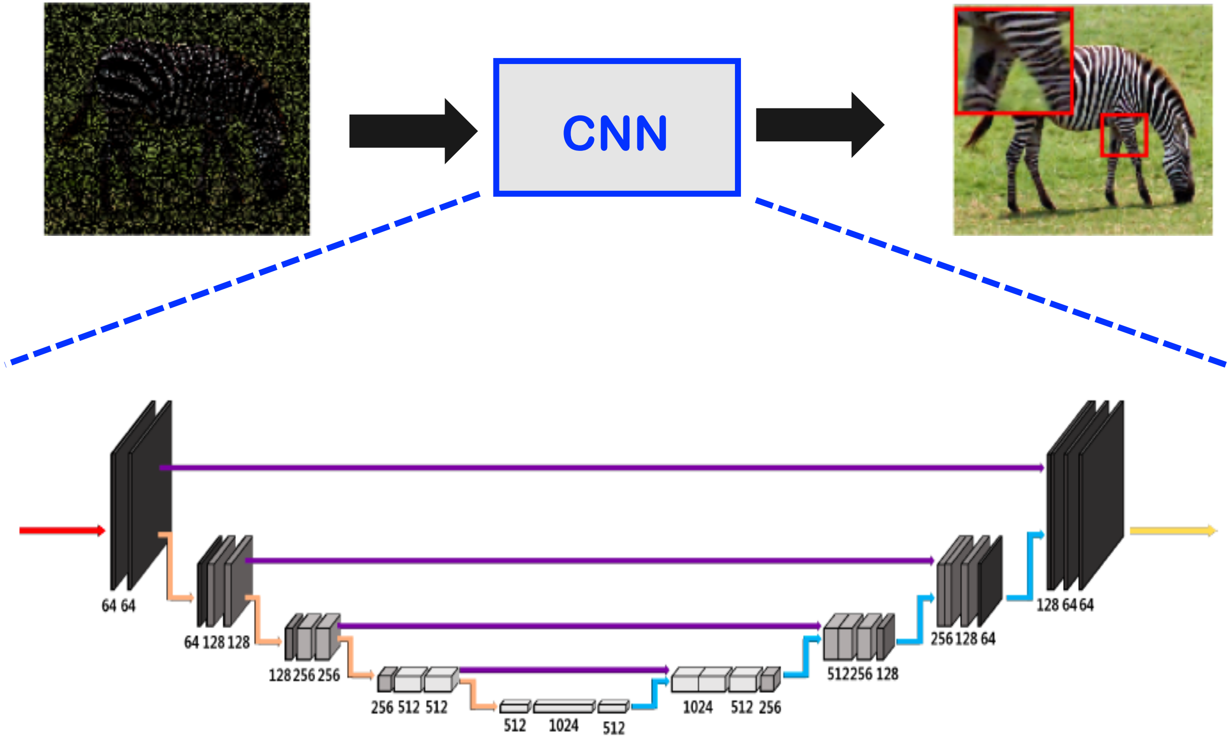
Encoder-Decoder CNN for Inverse Problems



Encoder-Decoder CNN for Inverse Problems



Encoder-Decoder CNN for Inverse Problems



Successful applications to various inverse problems

Why Same Architecture Works for Different Inverse Problems ?

Classical Methods for Inverse Problems

Step 1: Signal Representation

coefficients

$$x = \sum_i \langle b_i, x \rangle \tilde{b}_i$$

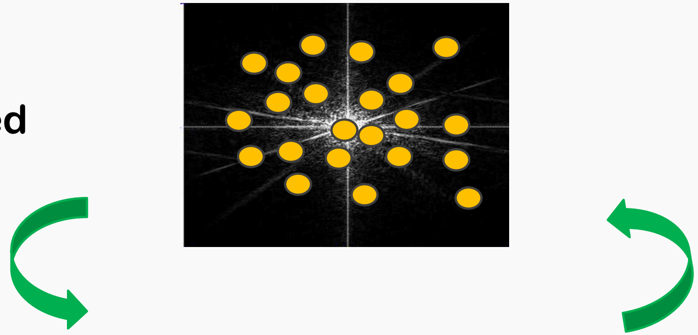
Analysis frame

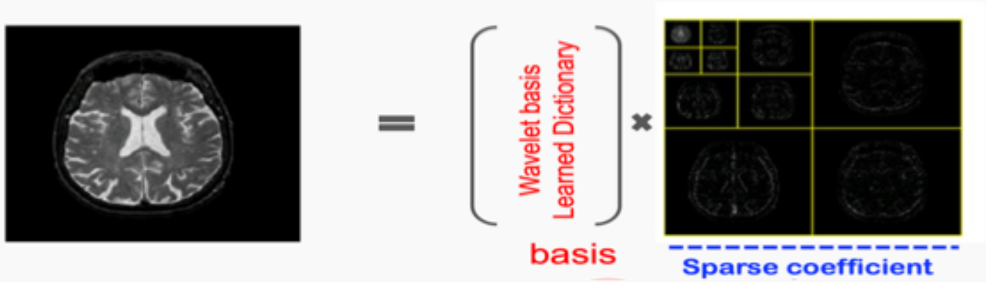
Synthesis frame

Classical Methods for Inverse Problems

Step 2:   Basis Pursuit by Optimization

Eg. Compressed Sensing



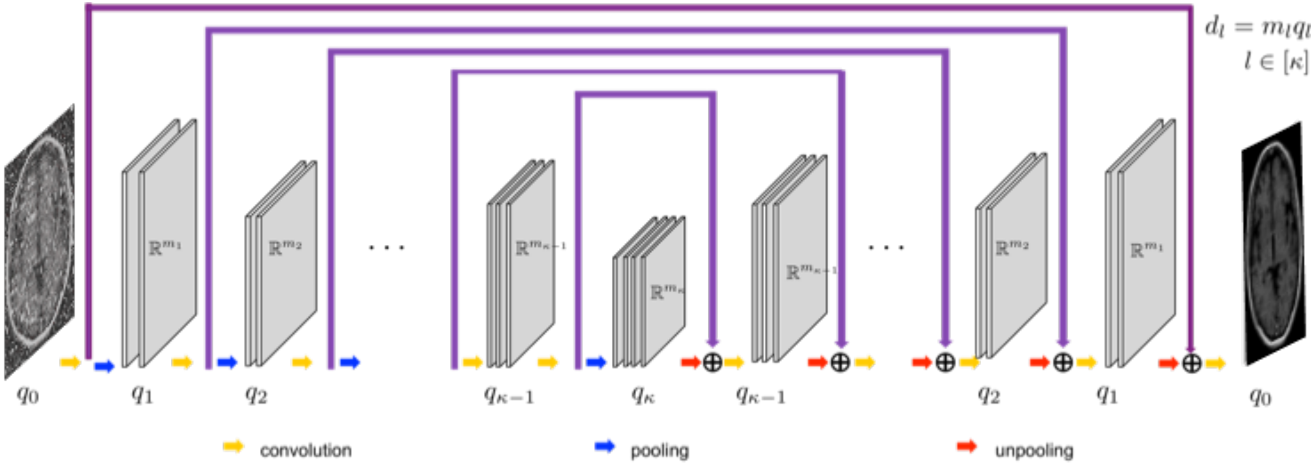


$$x = \sum_i \tilde{b}_i \langle b_i , x \rangle$$

Why do They Look so Different ?
Any Link between Them ?

Our Theoretical Findings

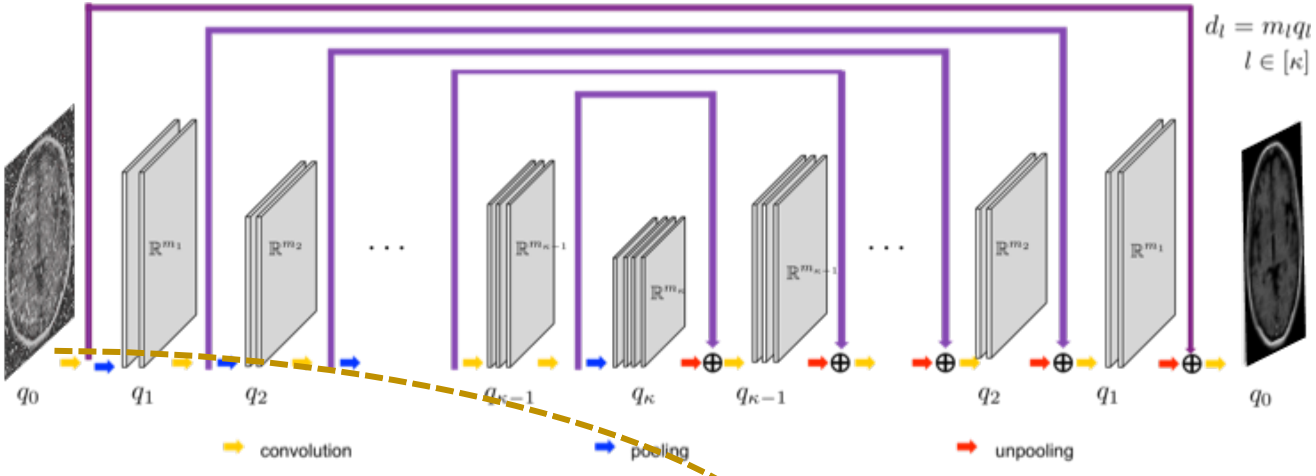
Ye et al, SIIMS, 2018; Ye et al, ICML, 2019



$$y = \sum_i \langle b_i(x), x \rangle \tilde{b}_i(x)$$

Our Theoretical Findings

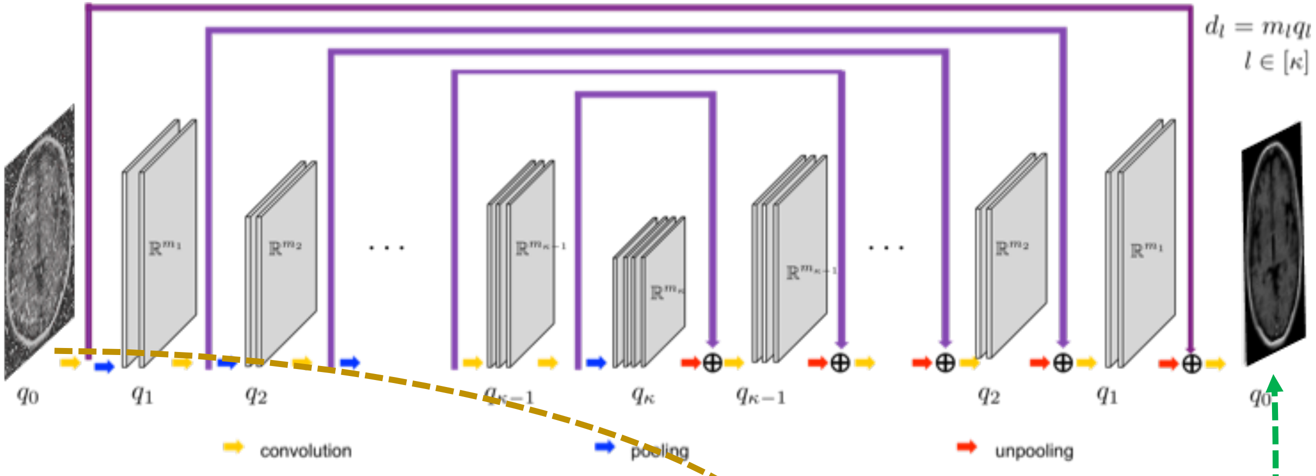
Ye et al, SIIMS, 2018; Ye et al, ICML, 2019



$$y = \sum_i \langle b_i(x), x \rangle \tilde{b}_i(x)$$

Our Theoretical Findings

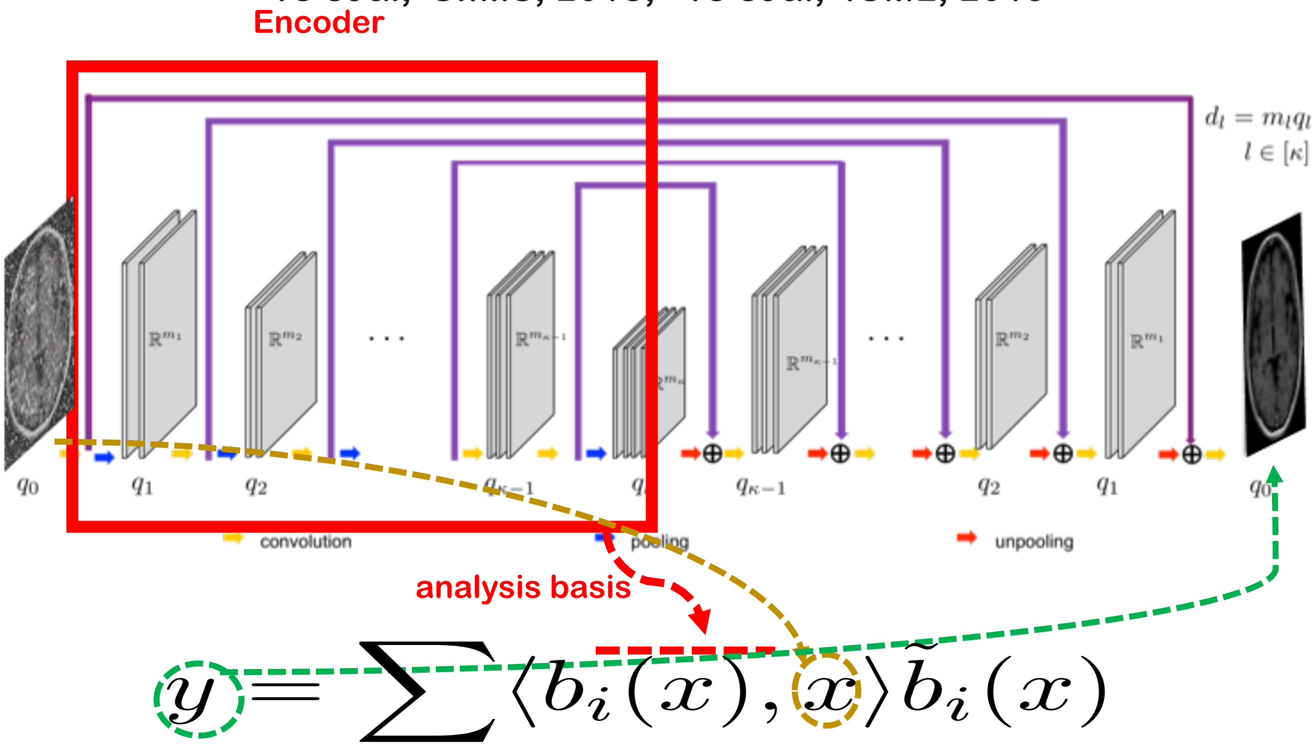
Ye et al, SIIMS, 2018; Ye et al, ICML, 2019



$$y = \sum_i \langle b_i(x), x \rangle \tilde{b}_i(x)$$

Our Theoretical Findings

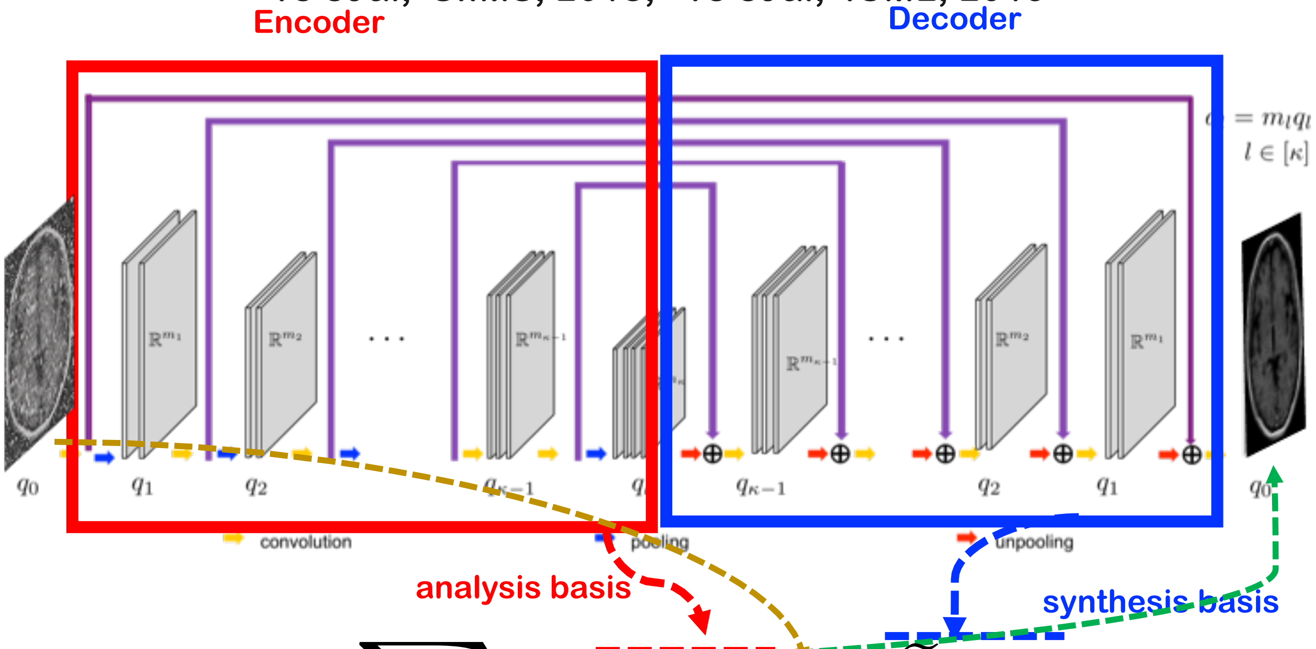
Ye et al, SIIMS, 2018; Ye et al, ICML, 2019



$$y = \sum_i \langle b_i(x), x \rangle \tilde{b}_i(x)$$

Our Theoretical Findings

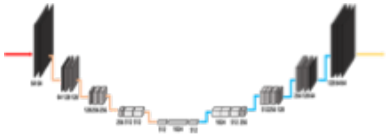
Ye et al, SIIMS, 2018; Ye et al, ICML, 2019



$$y = \sum_i \langle b_i(x), x \rangle \tilde{b}_i(x)$$

Linear Encoder-Decoder (ED) CNN

$$y = \tilde{B} B^\top x = \sum_i \langle x, b_i \rangle \tilde{b}_i$$



$$\begin{aligned} B &= E^1 E^2 \cdots E^\kappa, \\ \tilde{B} &= D^1 D^2 \cdots D^\kappa \end{aligned}$$

pooling

un-pooling

$$E^l = \begin{bmatrix} \Phi^l \circledast \psi^l_{1,1} & \cdots & \Phi^l \circledast \psi^l_{q_l,1} \\ \vdots & \ddots & \vdots \\ \Phi^l \circledast \psi^l_{1,q_{l-1}} & \cdots & \Phi^l \circledast \psi^l_{q_l,q_{l-1}} \end{bmatrix} \qquad D^l = \begin{bmatrix} \tilde{\Phi}^l \circledast \tilde{\psi}^l_{1,1} & \cdots & \tilde{\Phi}^l \circledast \tilde{\psi}^l_{1,q_l} \\ \vdots & \ddots & \vdots \\ \tilde{\Phi}^l \circledast \tilde{\psi}^l_{q_{l-1},1} & \cdots & \tilde{\Phi}^l \circledast \tilde{\psi}^l_{q_{l-1},q_l} \end{bmatrix}$$

Learned filters

Linear E-D CNN w/ Skipped Connection

$$y = \tilde{B}B^\top x = \sum_i \langle x, b_i \rangle \tilde{b}_i$$

$$B = [E^1 \cdots E^\kappa \quad E^1 \cdots E^{\kappa-1} S^\kappa \quad \cdots \quad E^1 S^2 \quad S^1]$$

$$\tilde{B} = [D^1 \cdots D^\kappa \quad D^1 \cdots D^{\kappa-1} \tilde{S}^\kappa \quad \cdots \quad D^1 \tilde{S}^2 \quad \tilde{S}^1]$$

more redundant expression

$$S^l = \begin{bmatrix} I_{m_{l-1}} \circledast \psi^l_{1,1} & \cdots & I_{m_{l-1}} \circledast \psi^l_{q_l,1} \\ \vdots & \ddots & \vdots \\ I_{m_{l-1}} \circledast \psi^l_{1,q_{l-1}} & \cdots & I_{m_{l-1}} \circledast \psi^l_{q_l,q_{l-1}} \end{bmatrix}$$

$$\tilde{S}^l = \begin{bmatrix} I_{m_{l-1}} \circledast \tilde{\psi}^l_{1,1} & \cdots & I_{m_{l-1}} \circledast \tilde{\psi}^l_{1,q_l} \\ \vdots & \ddots & \vdots \\ I_{m_{l-1}} \circledast \tilde{\psi}^l_{q_{l-1},1} & \cdots & I_{m_{l-1}} \circledast \tilde{\psi}^l_{q_{l-1},q_l} \end{bmatrix}$$

Learned filters

Deep Convolutional Framelets

Perfect reconstruction

$$x = \tilde{B}B^\top x = \sum_i \langle x, b_i \rangle \tilde{b}_i$$

Frame conditions

w/o skipped connection

$$\tilde{\Phi}^l \Phi^{l\top} = \alpha I_{m_{l-1}}, \; \Psi^l \tilde{\Psi}^{l\top} = \frac{1}{r\alpha} I_{rq_{l-1}}$$

w skipped connection

$$\tilde{\Phi}^l \Phi^{l\top} = \alpha I_{m_{l-1}}, \; \Psi^l \tilde{\Psi}^{l\top} = \frac{1}{r(\alpha+1)} I_{rq_{l-1}}$$

Ye et al, SIIMS 2018; Ye et al, ICML 2019

Deep Convolutional Framelets

Perfect reconstruction

$$x = \tilde{B}B^\top x = \sum_i \langle x, b_i \rangle \tilde{b}_i$$

Frame conditions

Frame Conditions for Pooling layers

w/o skipped connection

$$\tilde{\Phi}^l \Phi^{l\top} = \alpha I_{m_{l-1}}, \quad \Psi^l \tilde{\Psi}^{l\top} = \frac{1}{r\alpha} I_{rq_{l-1}}$$

w skipped connection

$$\tilde{\Phi}^l \Phi^{l\top} = \alpha I_{m_{l-1}}, \quad \Psi^l \tilde{\Psi}^{l\top} = \frac{1}{r(\alpha+1)} I_{rq_{l-1}}$$

Ye et al, SIAM J. Imaging Science, 2018

Role of ReLUs?

Generator for Multiple Expressions

$$y = \tilde{B}(x)B(x)^\top x = \sum_i \langle x, b_i(x)\rangle \tilde{b}_i(x)$$

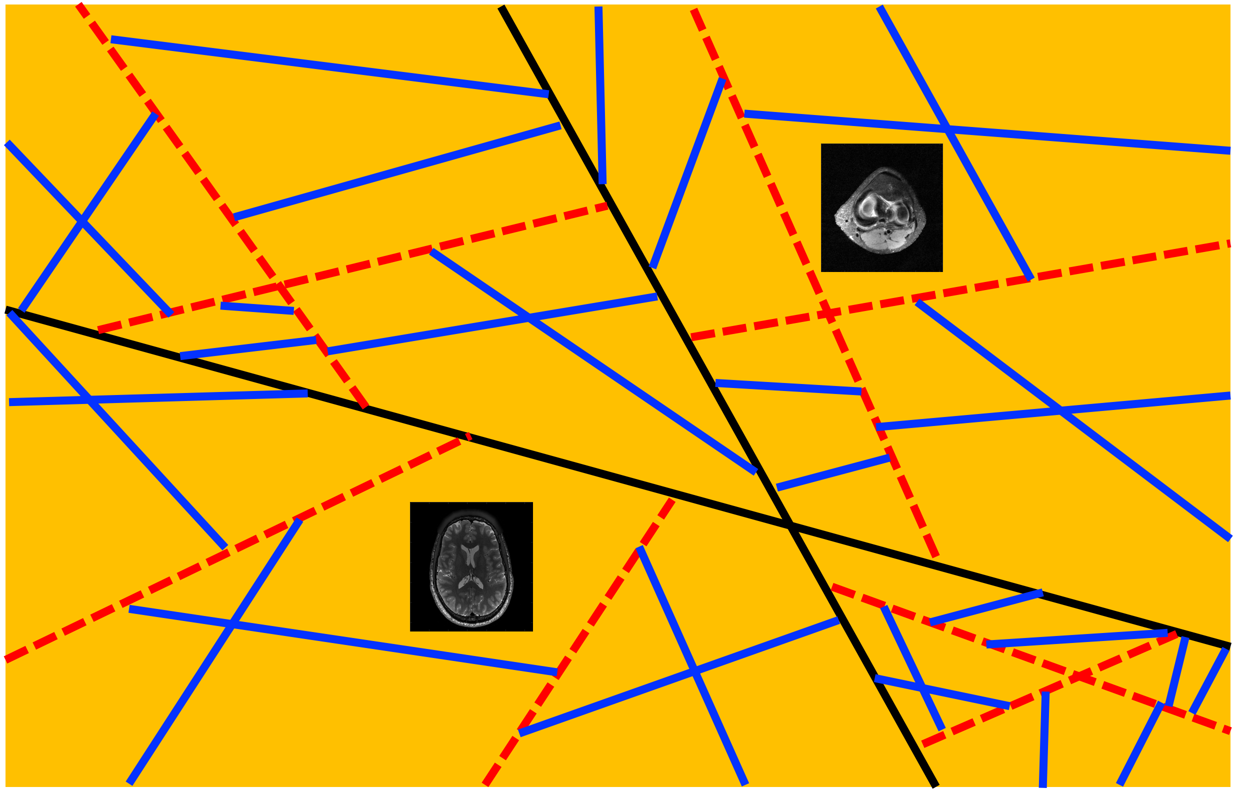
$$B(x) = E^1\Sigma^1(x)E^2 \cdots \Sigma^{\kappa-1}(x)E^\kappa,$$
$$\tilde{B}(x) = D^1\tilde{\Sigma}^1(x)D^2 \cdots \tilde{\Sigma}^{\kappa-1}(x)D^\kappa$$

$$\Sigma^l(x) = \begin{bmatrix} \sigma_1 & 0 & \cdots & 0 \\ 0 & \sigma_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma_{m_l} \end{bmatrix}$$

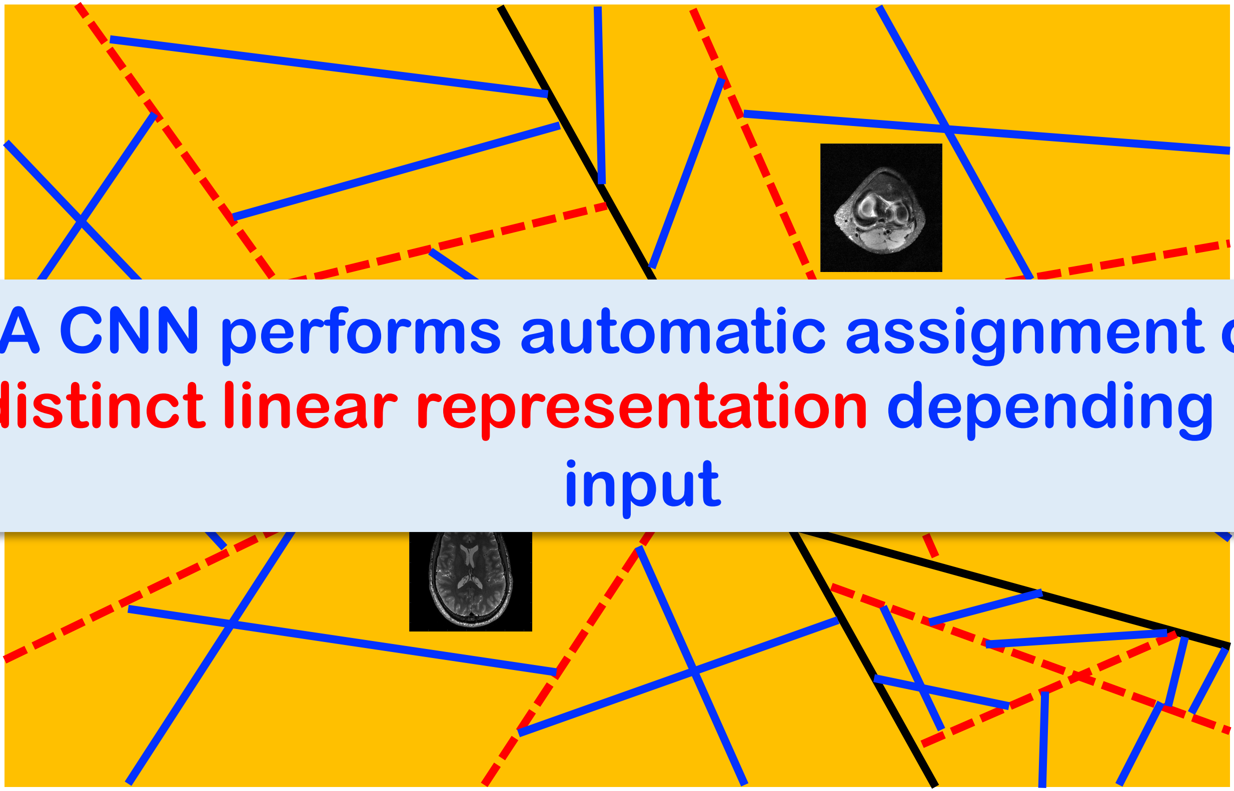
Input dependent {0,1} matrix

--> Input adaptivity

Input Space Partitioning for Multiple Expressions

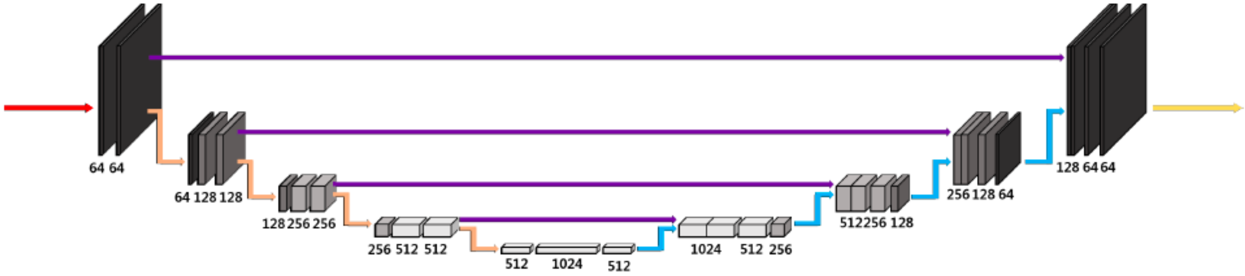


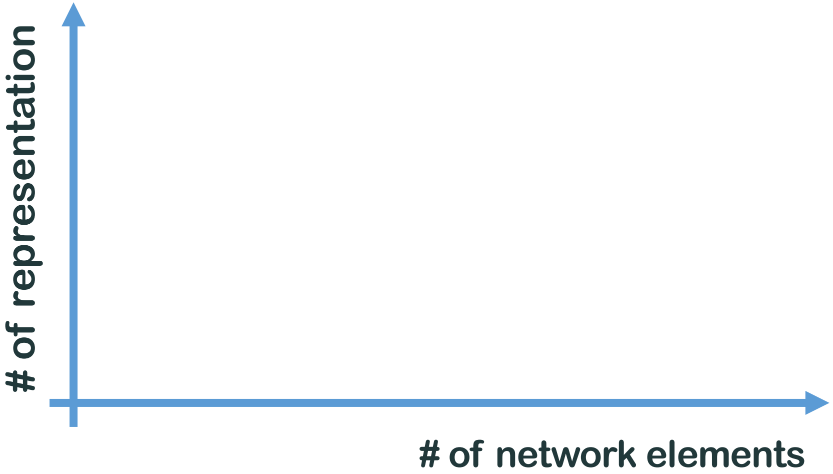
Input Space Partitioning for Multiple Expressions



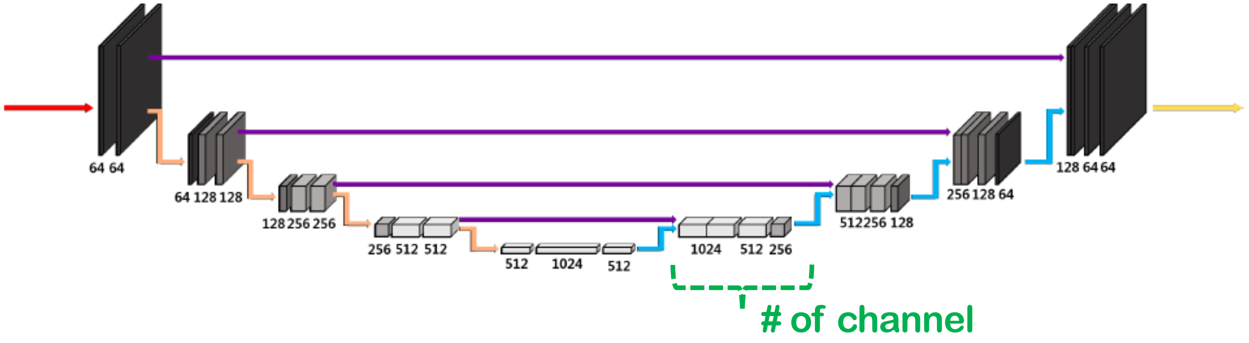
A CNN performs automatic assignment of distinct linear representation depending on input

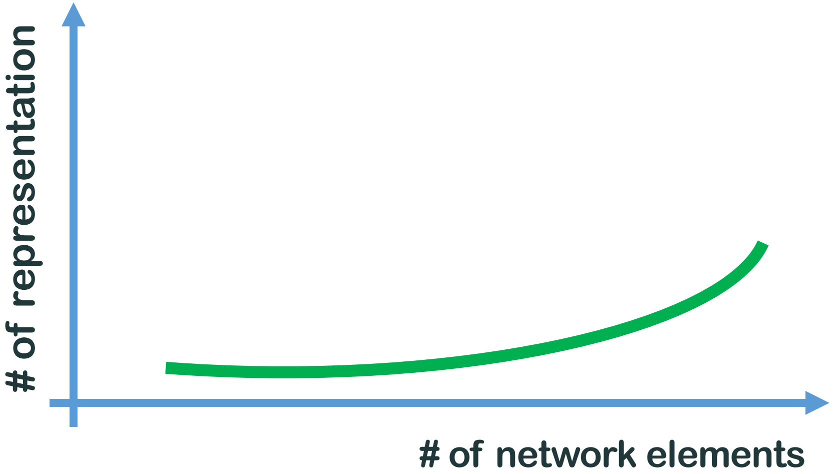
Expressivity of E-D CNN



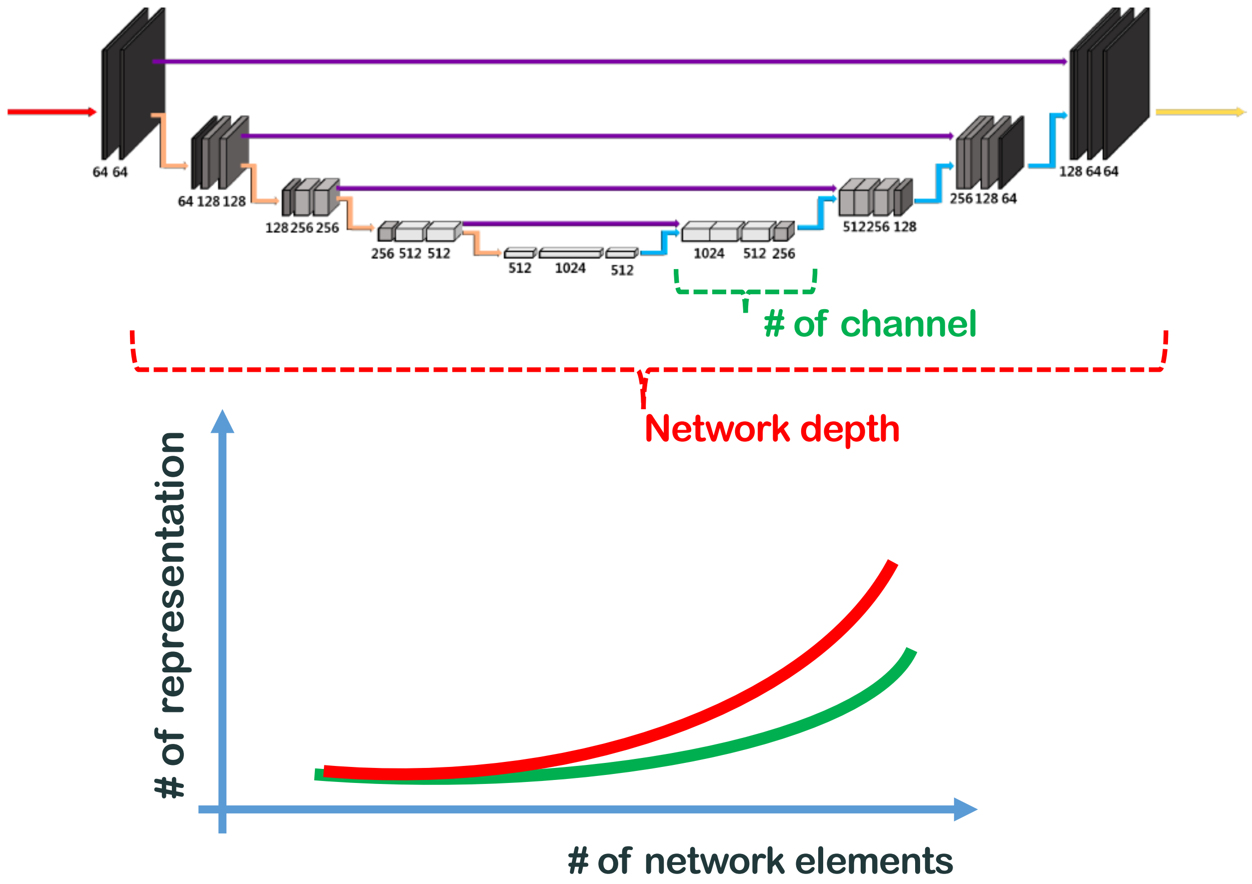


Expressivity of E-D CNN

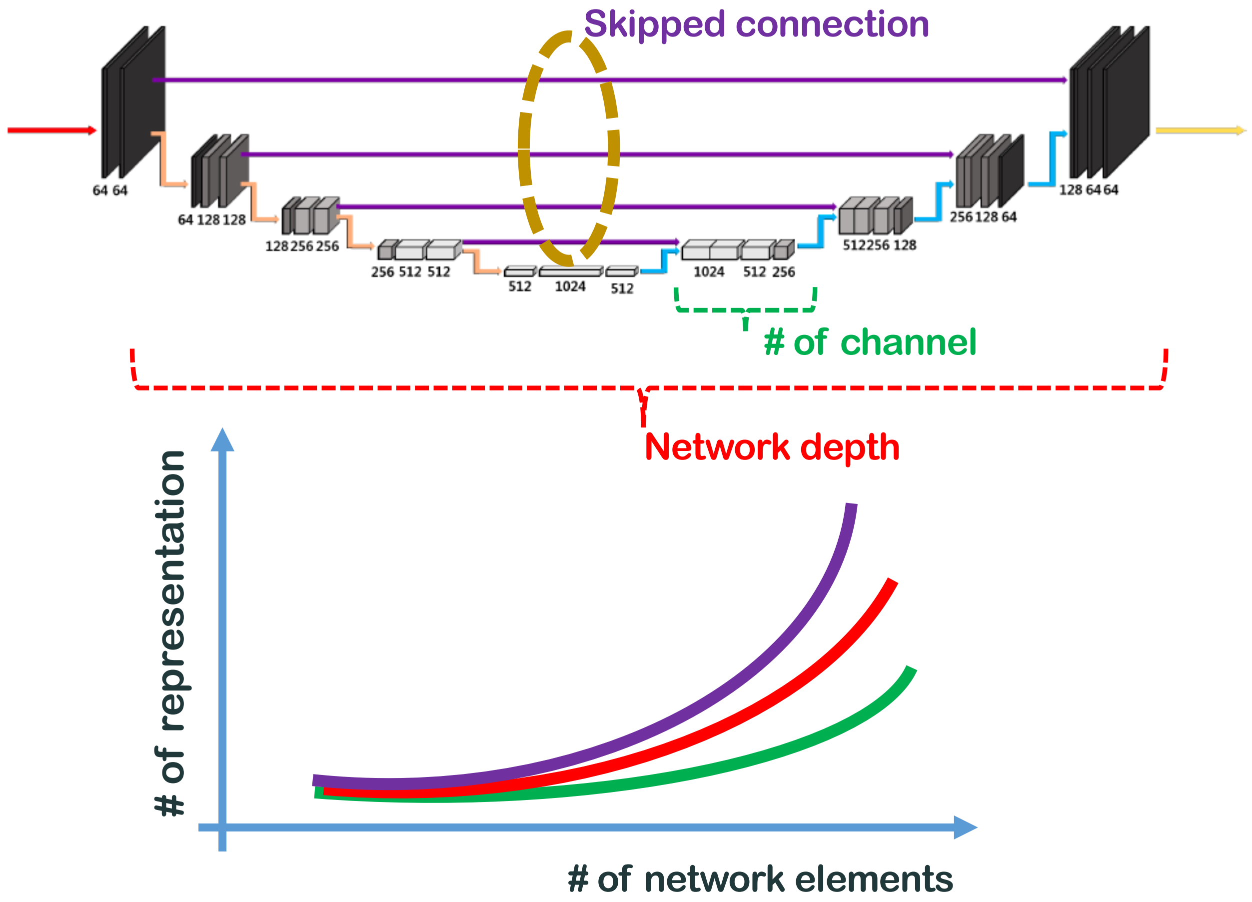




Expressivity of E-D CNN



Expressivity of E-D CNN



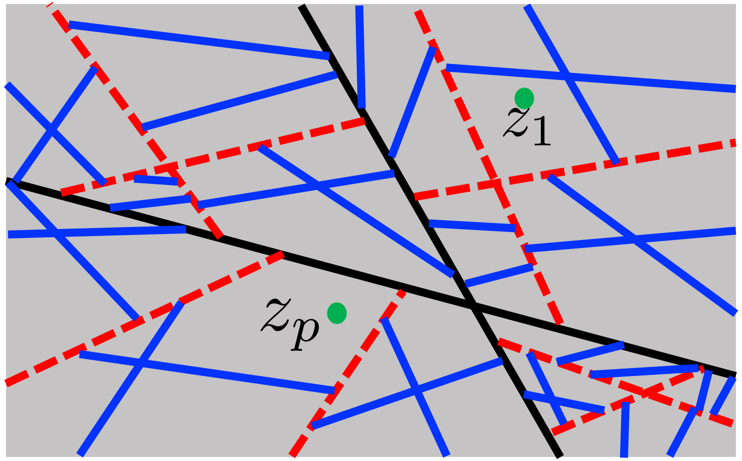
Lipschitz Continuity

Related to the generalizability

$$\|F(\mathbf{W}, x^{(1)}) - F(\mathbf{W}, x^{(2)})\|_2 \leq K\|x^{(1)} - x^{(2)}\|_2$$

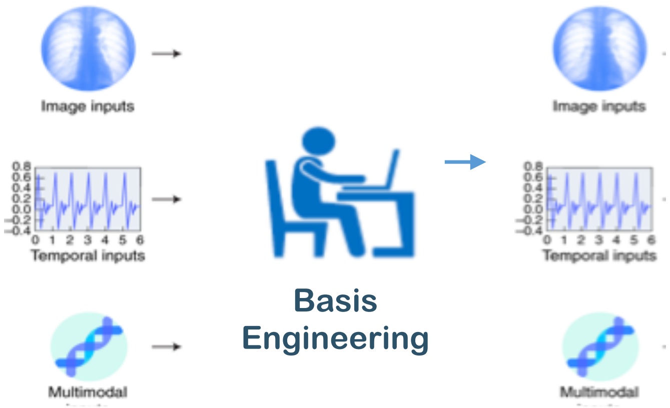
$$K = \max_p K_p, \quad K_p = \|\tilde{B}(z_p)B(z_p)^\top\|_2$$

Dependent on the Local Lipschitz

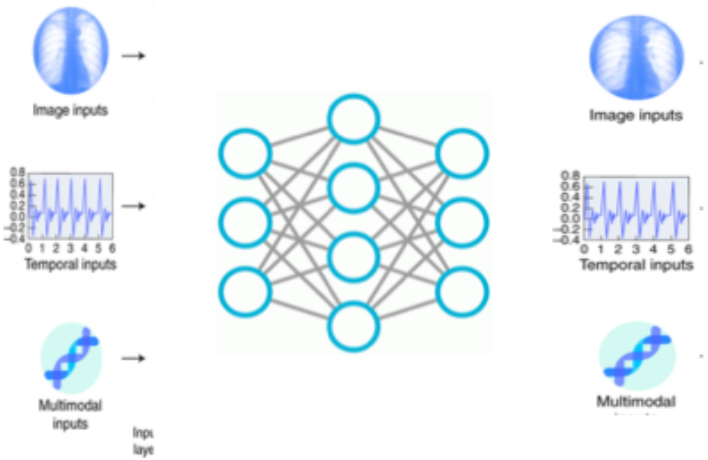


Regularized Recon vs. Deep Recon

Classical Regularized Recon
(basis engineering)



Deep Recon
(no basis engineering)



THEORY-DRIVEN CNN DESIGN

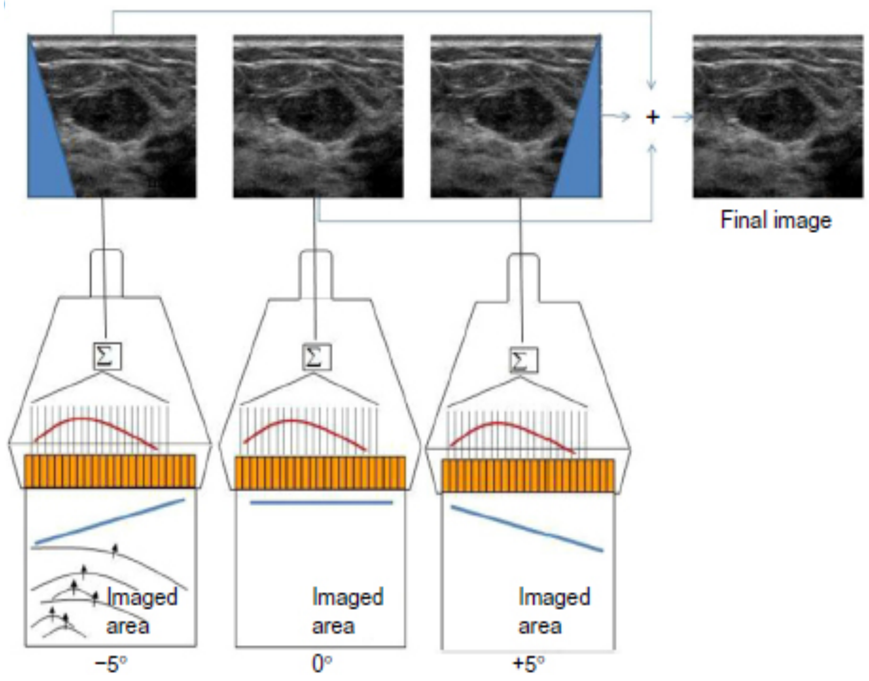
:some snapshots

Universal Deep Beamformer

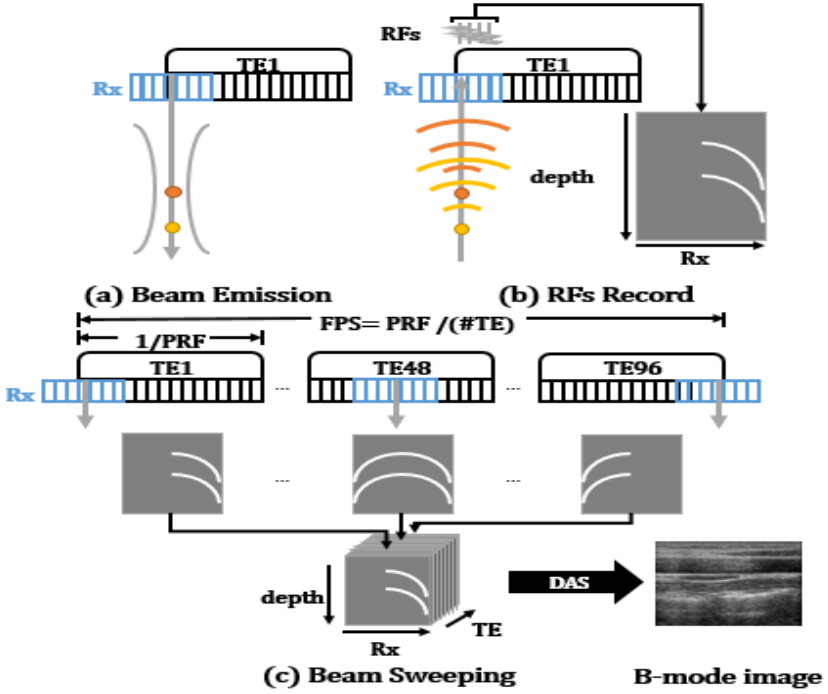
Khan et al, MICCAI, 2019 (oral presentation)

Ultrasound Acquisition Modes

Planewave Imaging

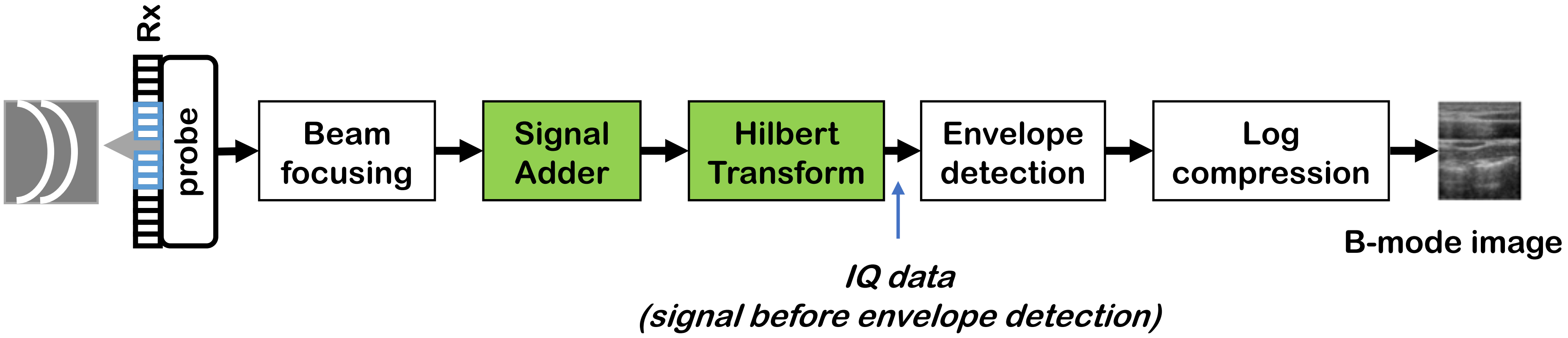


Focused Imaging



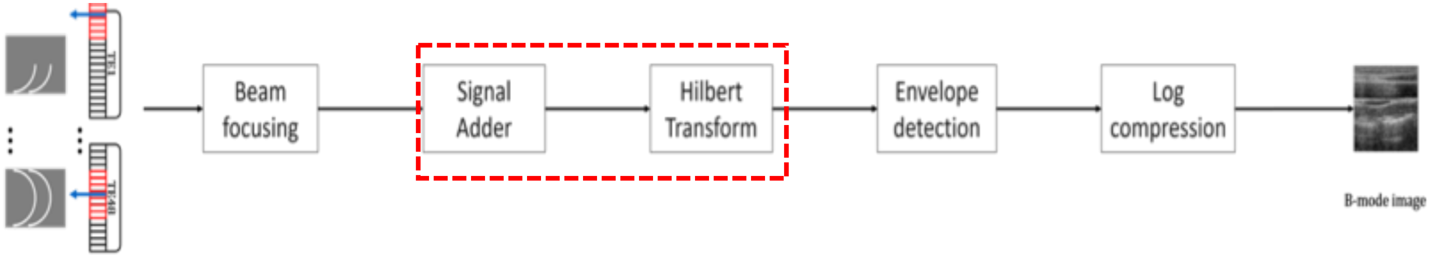
Conventional approach

Conventional *delay-and-sum* (DAS) Beamforming Pipeline



Adaptive Beamformer

Conventional Beamforming Pipeline



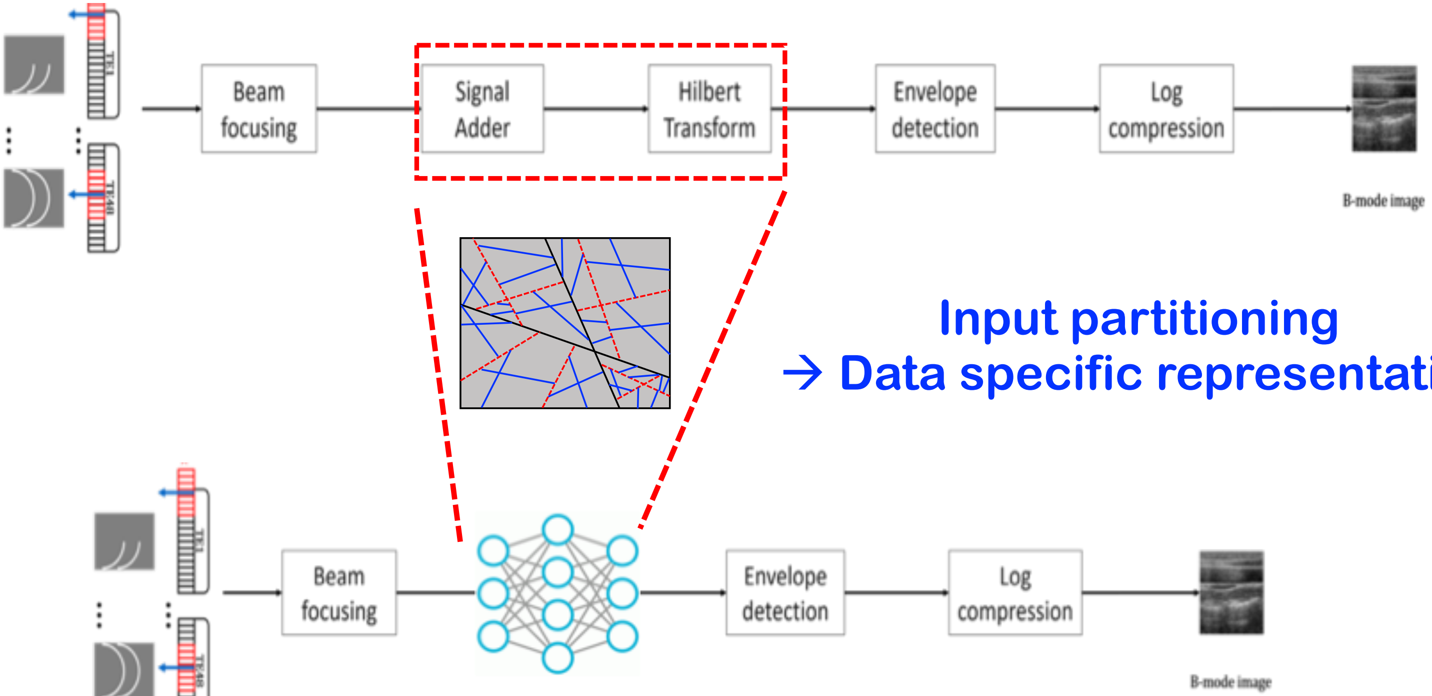
$$\begin{bmatrix} I \\ Q \end{bmatrix} = \begin{bmatrix} z_l[n] \\ \mathcal{H}(z_l)[n] \end{bmatrix}$$

Data specific weights

$$z_l[n] = \boldsymbol{w}_l[n]^\top \boldsymbol{y}_l[n] \qquad \boldsymbol{w}_l[n] = \frac{\boldsymbol{R}_l[n]^{-1}\mathbf{1}}{\mathbf{1}^H \boldsymbol{R}_l[n]^{-1}\mathbf{1}}$$

Adaptive and Compressive Deep Beamformer

Conventional Beamforming Pipeline

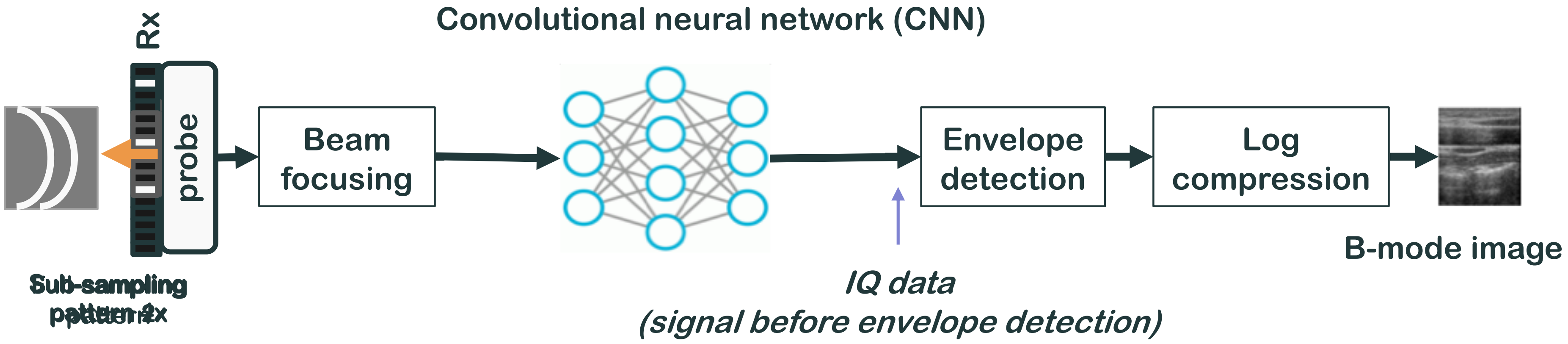


Input partitioning
→ Data specific representation

Deep Beamformer

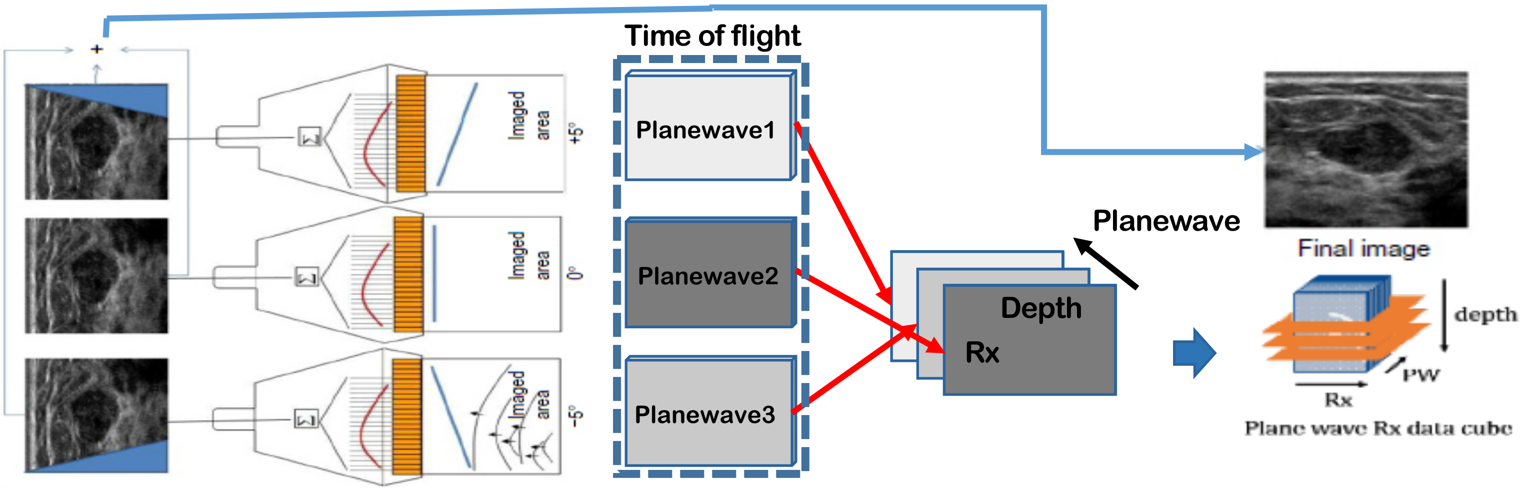
Khan et al, MICCAI 2019

Universal Deep Beamformer

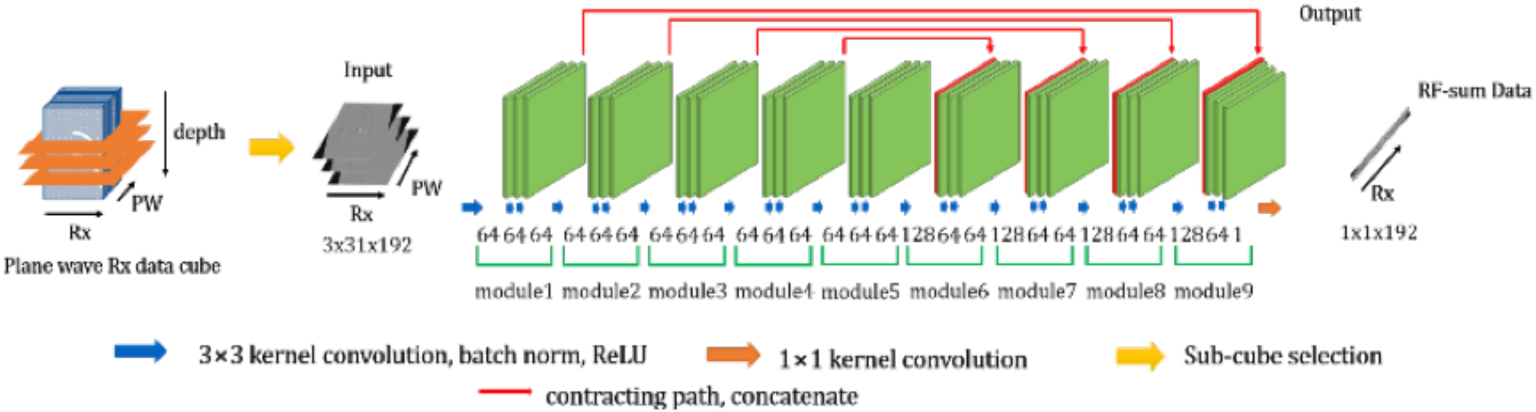


Application to Planar Wave Imaging

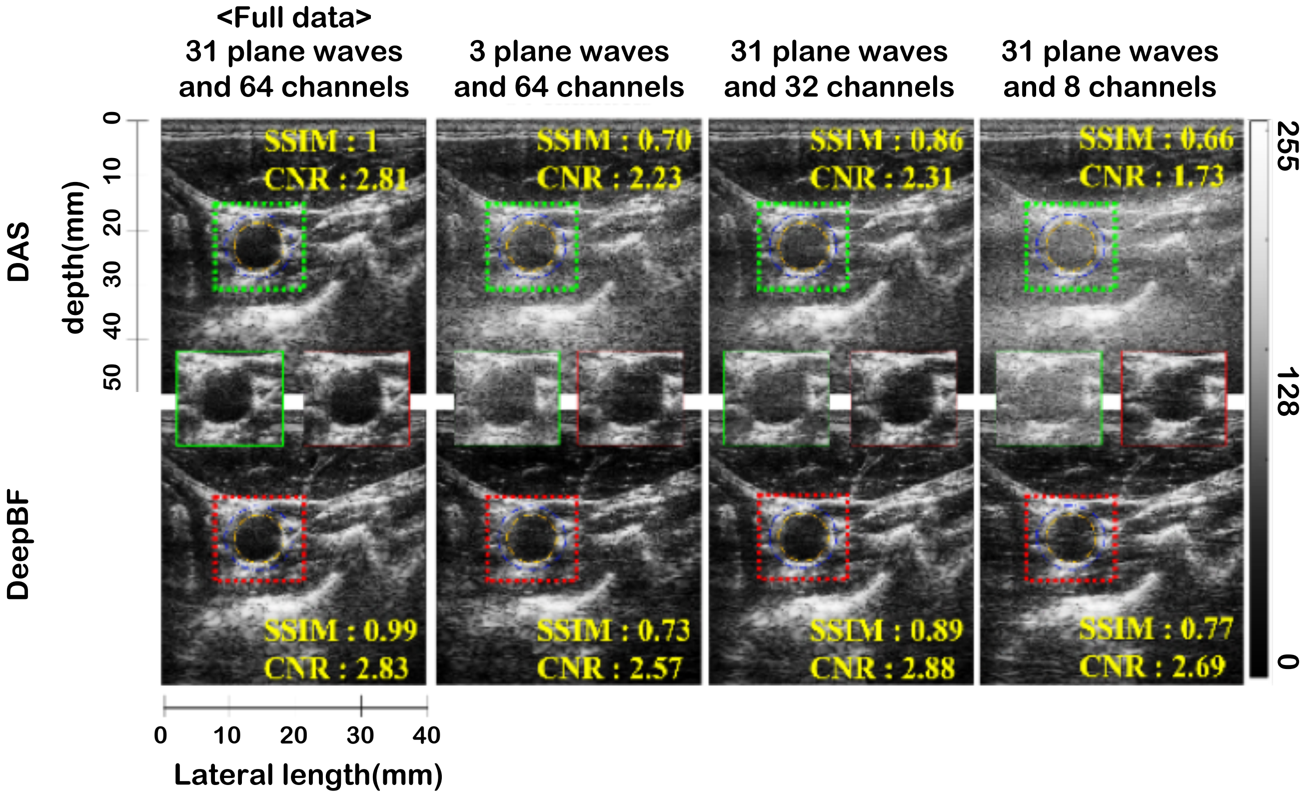
Data Generation

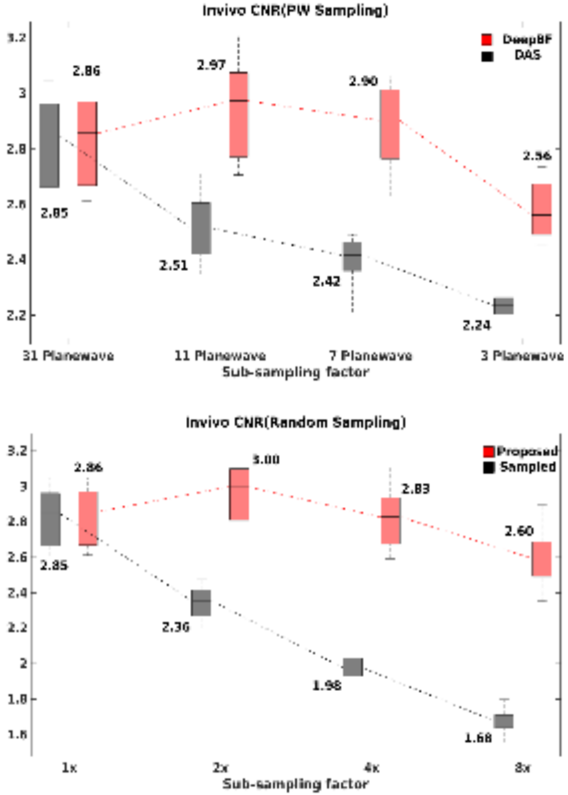


Proposed Method



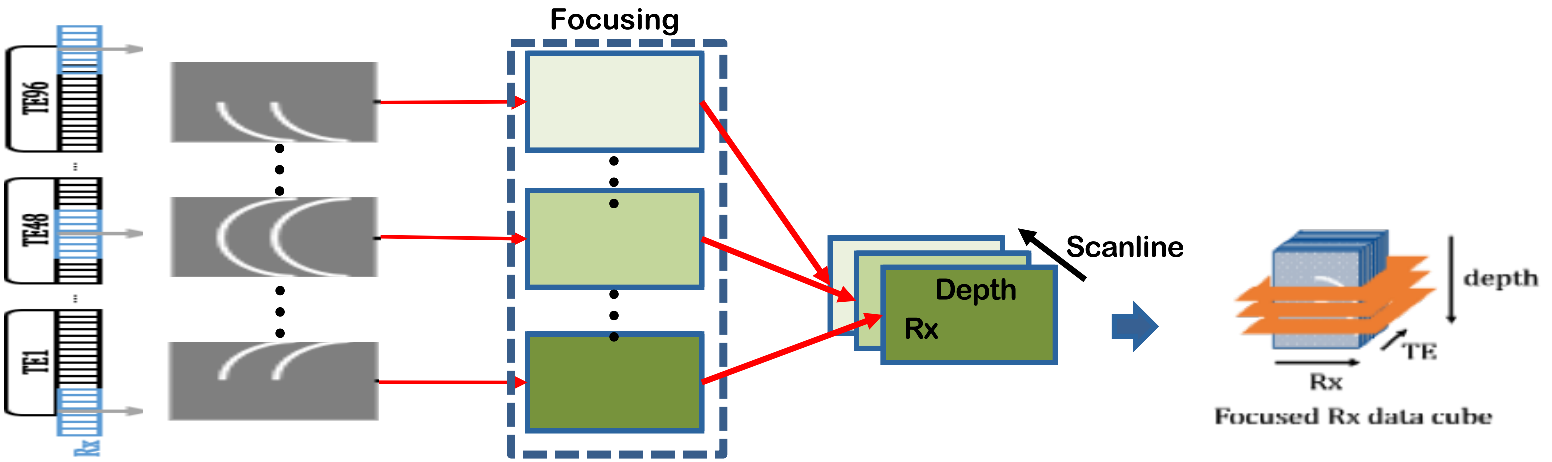
Results (Planewave / in-vivo)



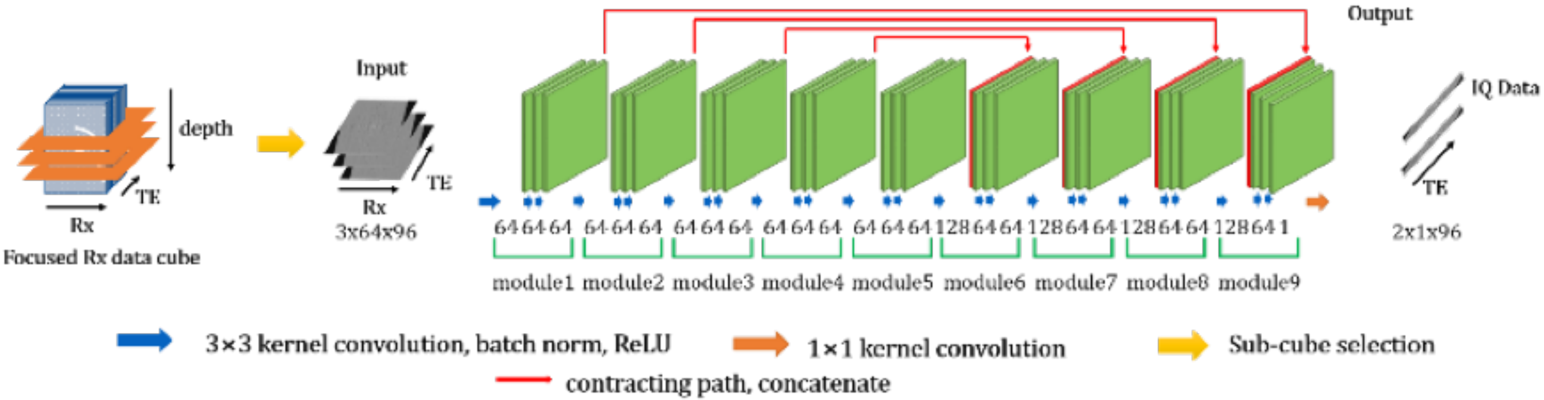


Application to Focused B-Mode Imaging

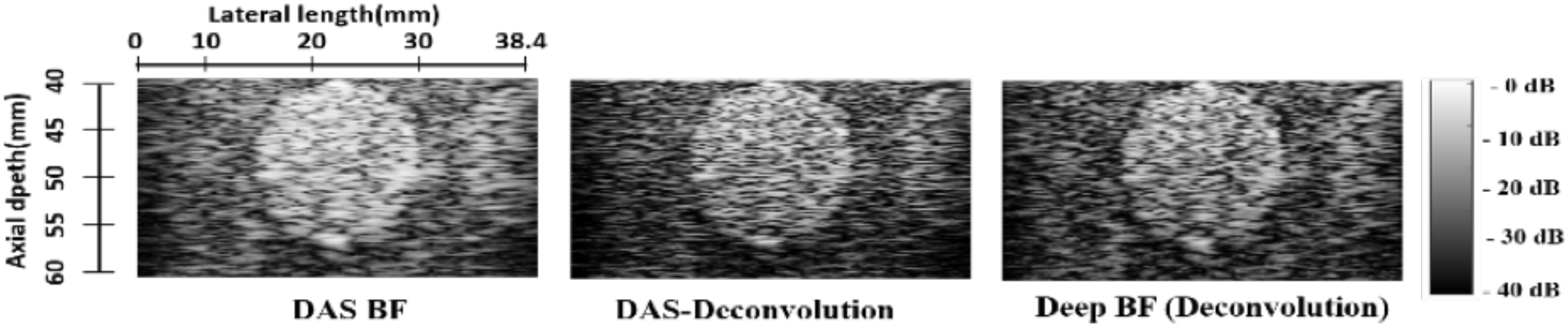
Data Generation

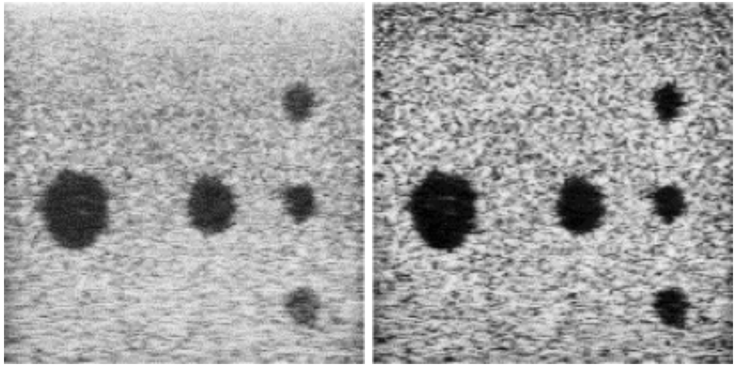


Proposed Method

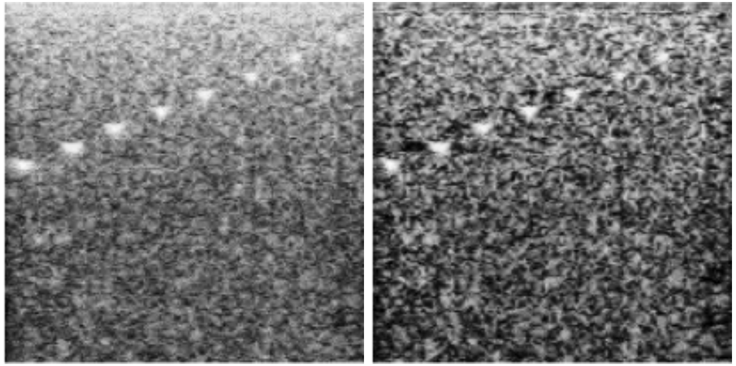


Results of Deconvolution BF (phantom)



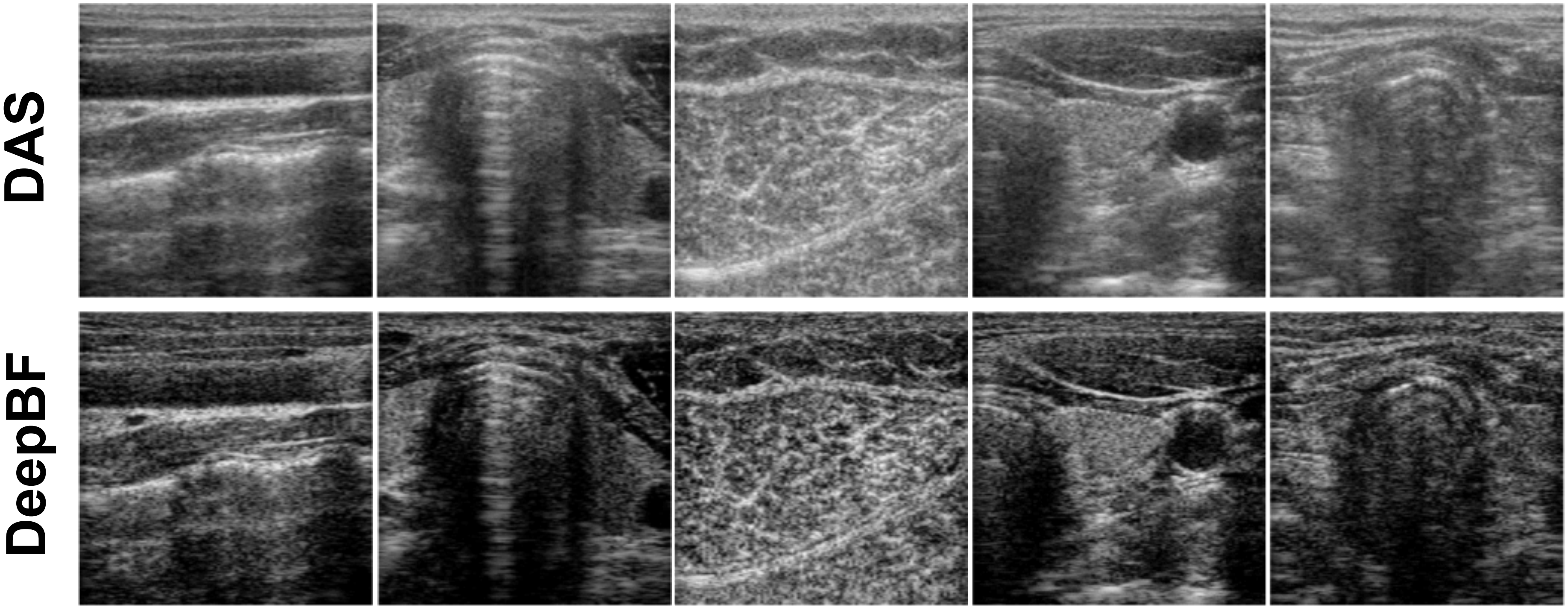


DAS DeepBF



DAS DeepBF

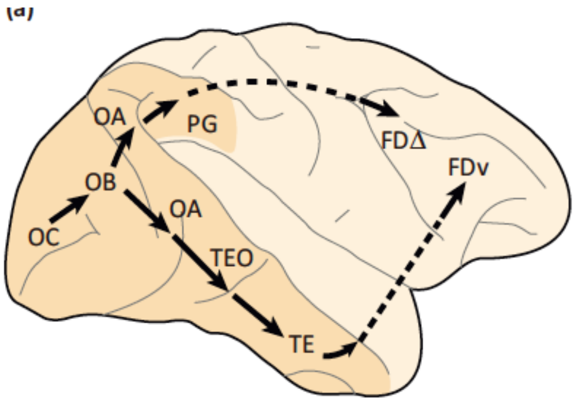
Results of Deconvolution BF (in-vivo)

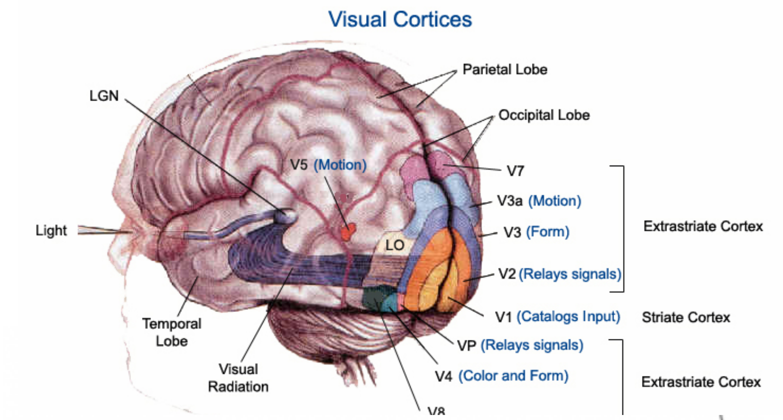


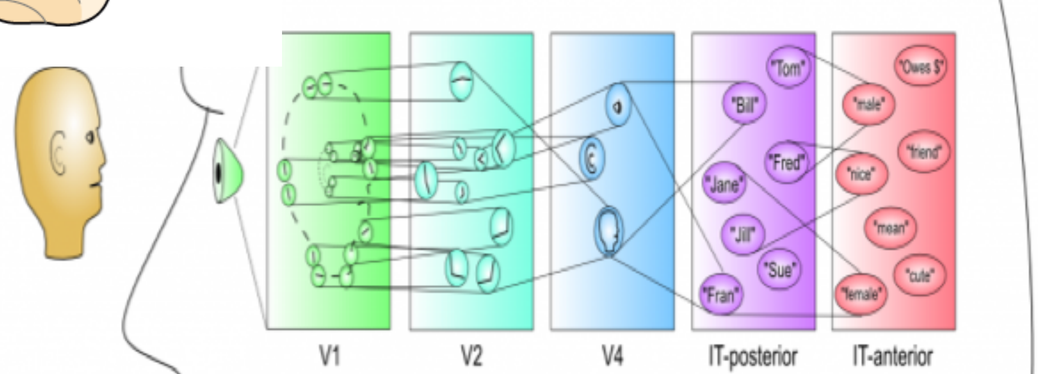
Which domain is good for learning?

Han et al, IEEE Trans. Medical Imaging (in press), 2019
Lee et al, MRM (in press), 2019
Han et al, Medical Physics, 2020

Image Domain Learning is Essential?





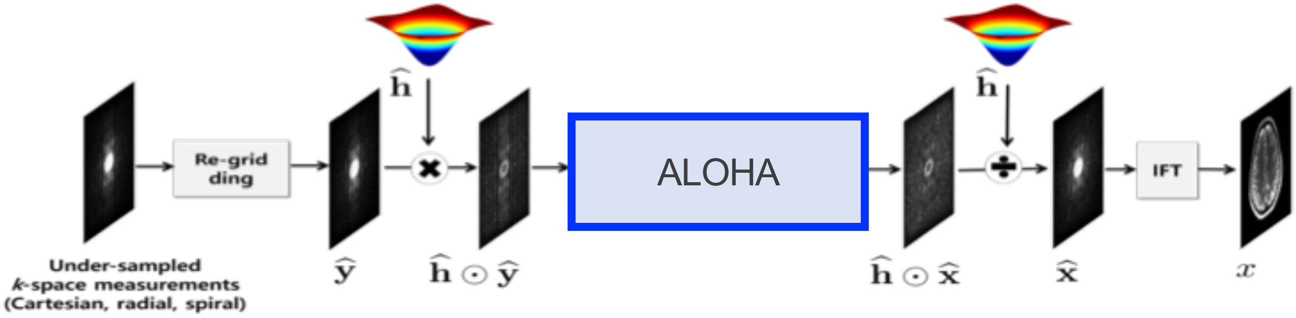


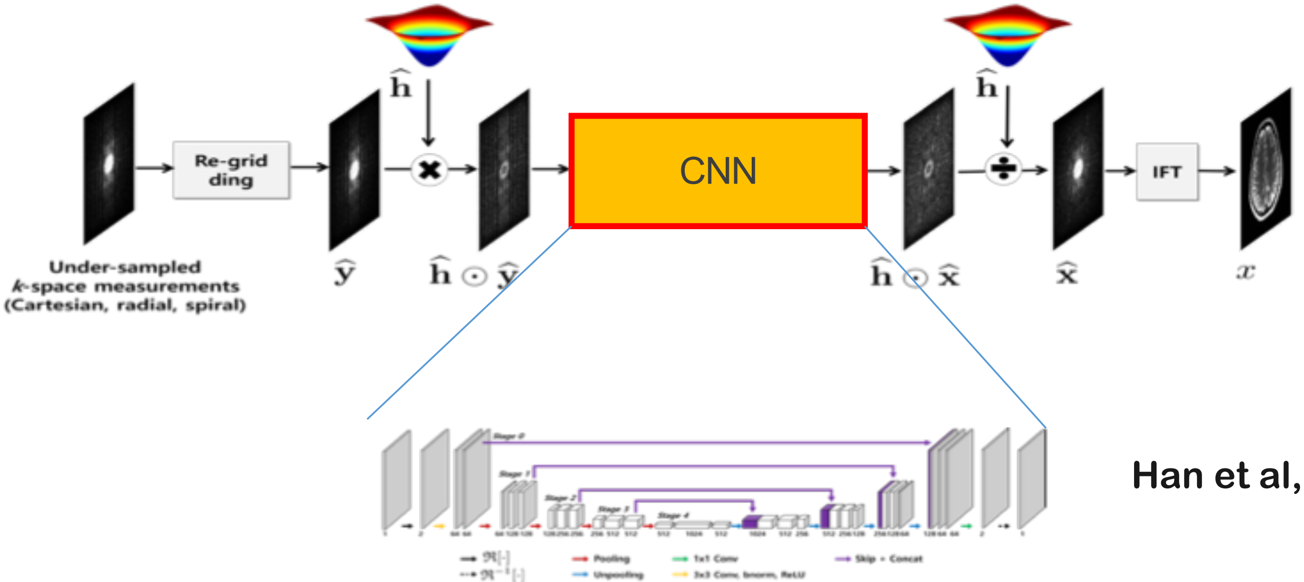
Kravitz et al, Trends in Cognitive Sciences January 2013, Vol. 17, No. 1

k-Space Deep Learning

Han et al, IEEE Trans. Medical Imaging (in press), 2019
Lee et al, MRM (in press), 2019

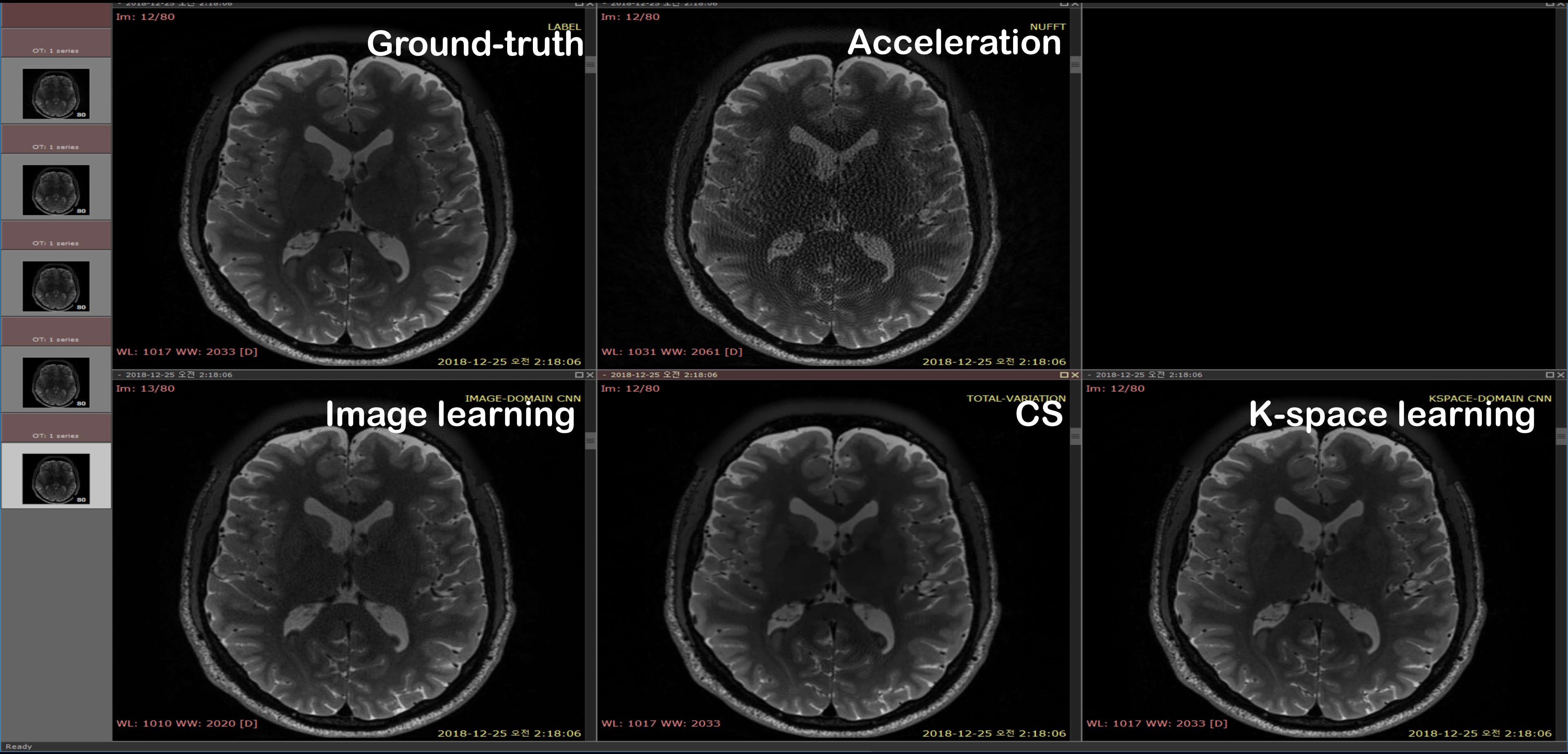
k-Space Deep Learning





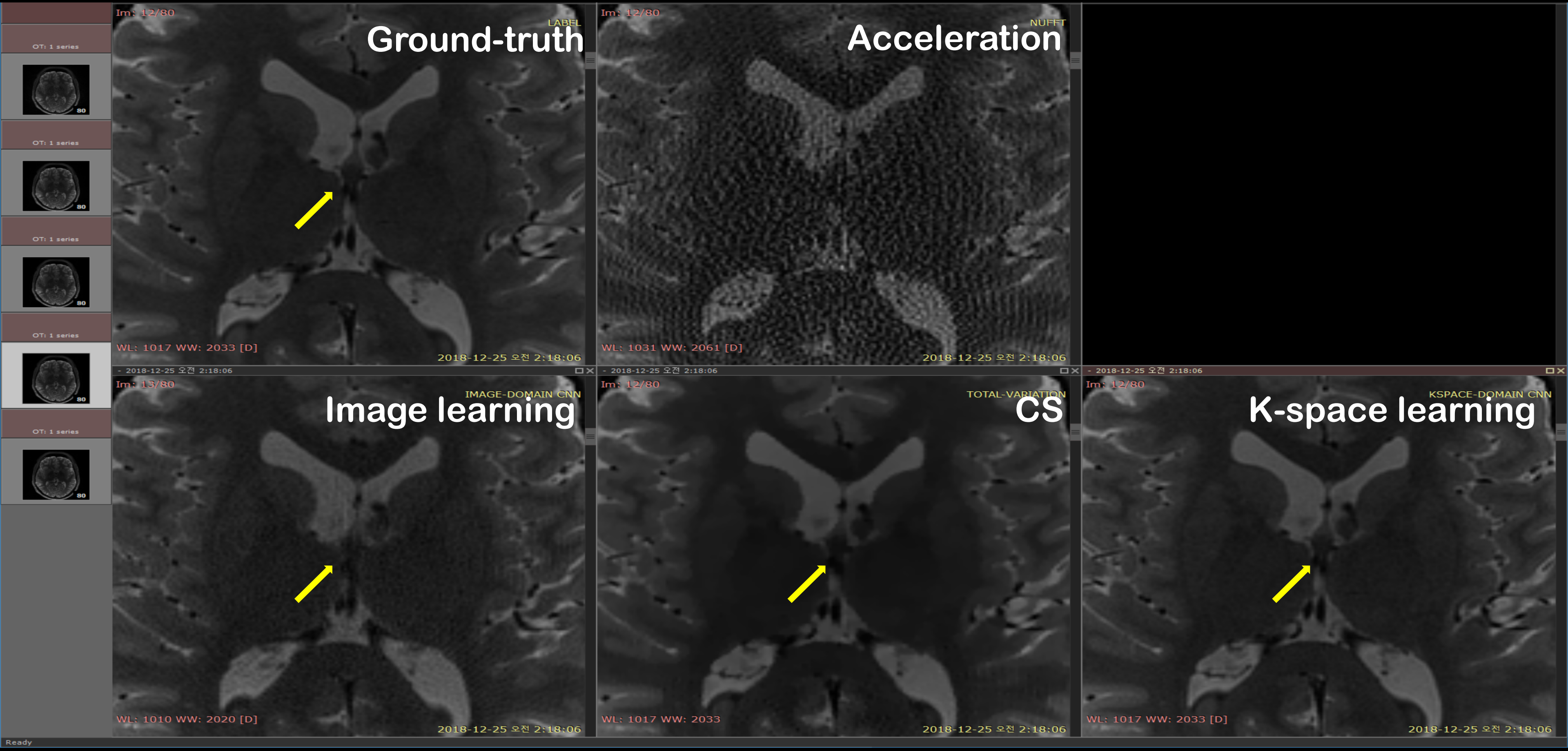
Han et al, IEEE TMI,2019

K-space Deep Learning (Radial R=6)



Han et al, *IEEE TMI* (in press)

K-space Deep Learning (Radial R=6)



Han et al, *IEEE TMI* (in press)

DBP Domain Deep Learning

Differentiated Backprojection

$$g(\mathbf{r}') = -\frac{1}{2\pi}\int_{\lambda_1}^{\lambda_2} \frac{d\lambda}{|\mathbf{r}' - \mathbf{r}_0(\lambda)|}\left[\frac{\partial}{\partial q} p(\mathbf{r}_0(q), \boldsymbol{\beta}(\mathbf{r}', \lambda))\right]_{q=\lambda}$$

Han et al, Medical Physics 46 (12), e855-e872, 2020
Han et al, arXiv preprint arXiv:1906.06854

Two Approaches for CT Reconstruction

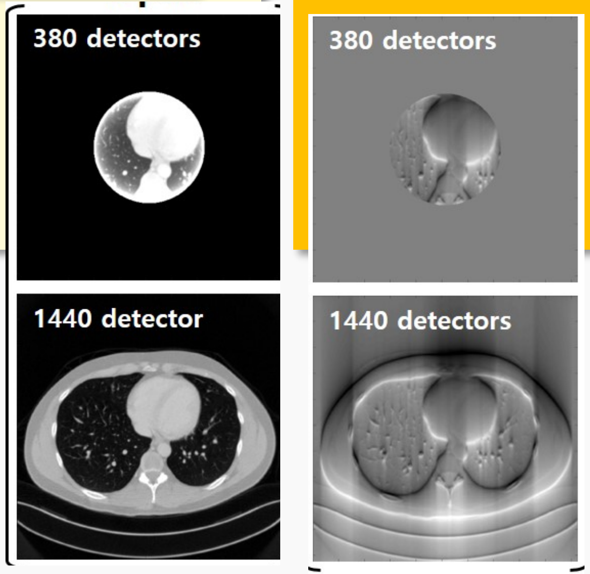
Zou, Y et al, PMB (2004).

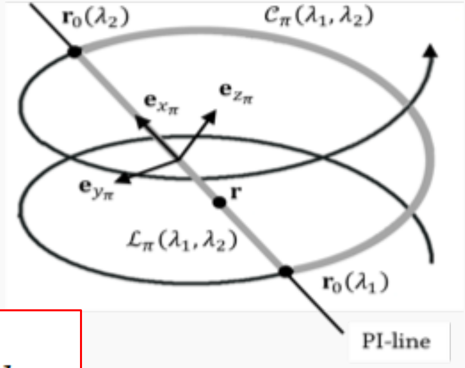
Filtered Backprojection (FBP)

- Ramp filtering
- Back-projection

Backprojection Filtration (BPF)

- Differentiation
- Back-projection
- Hilbert transform

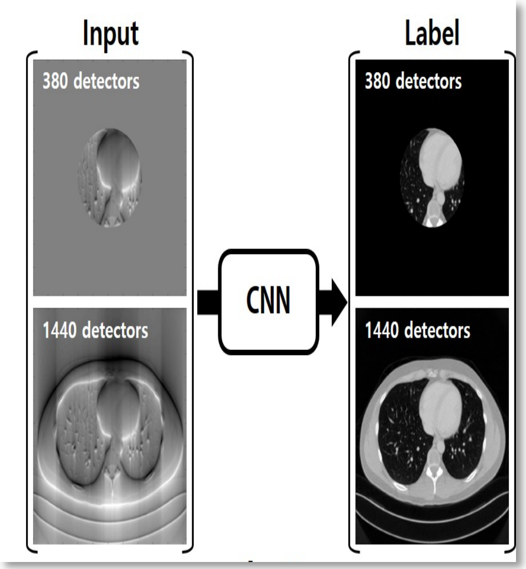




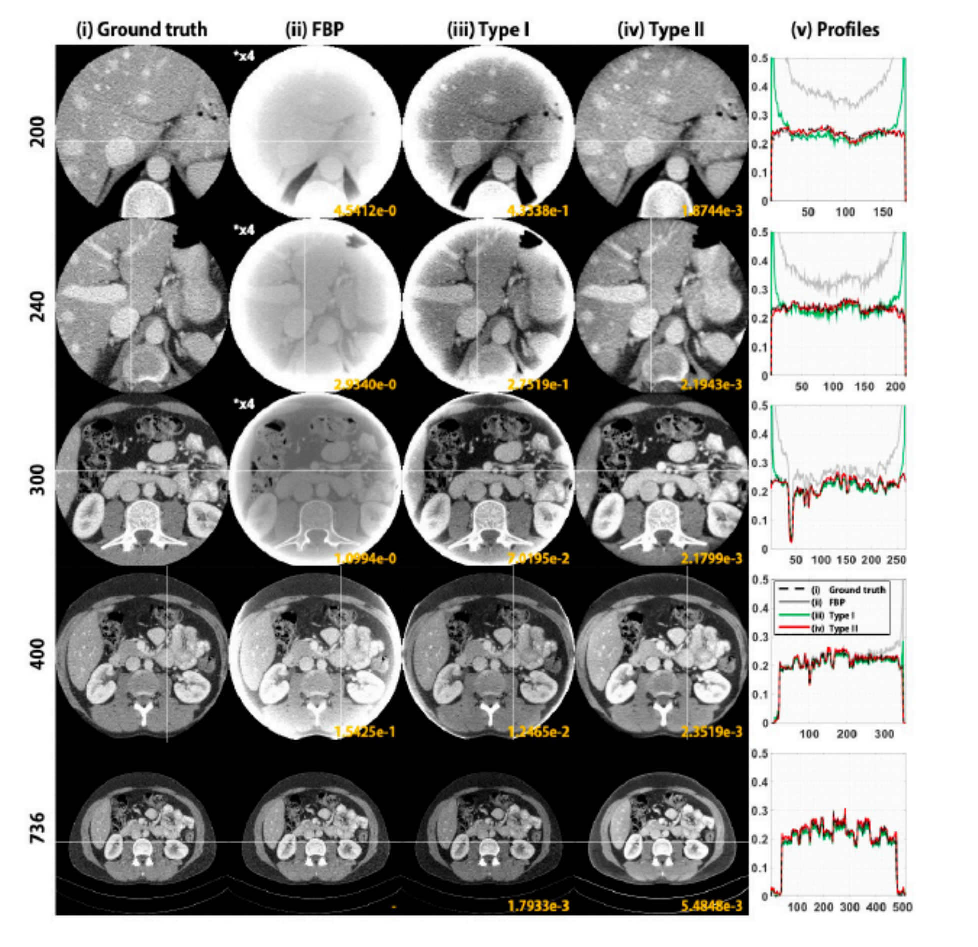
$$f(u) = \frac{c}{\pi\sqrt{1-u^2}} - \frac{1}{\pi\sqrt{1-u^2}} \int_{-1}^{1} \frac{\sqrt{1-s^2}g(s)}{u-s} ds,$$

DBP Domain ROI Tomography

Interior (ROI) Tomography
→ 2-D Deconvolution problem



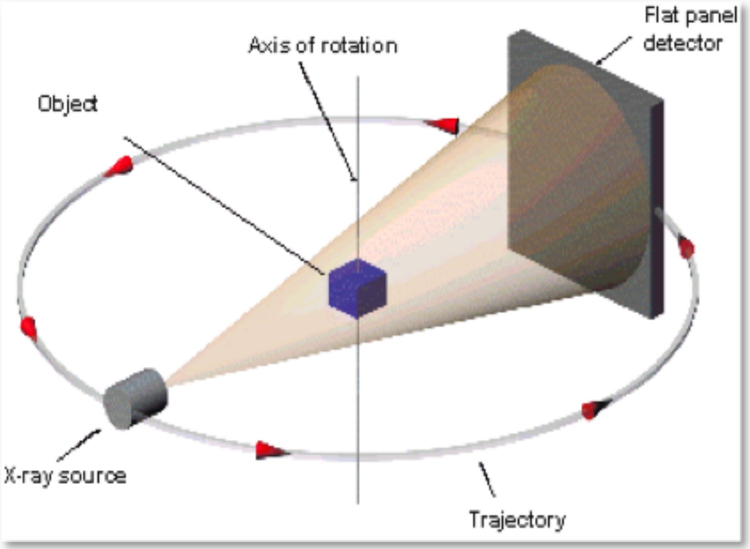
Han et al, Medical Physics, 2019



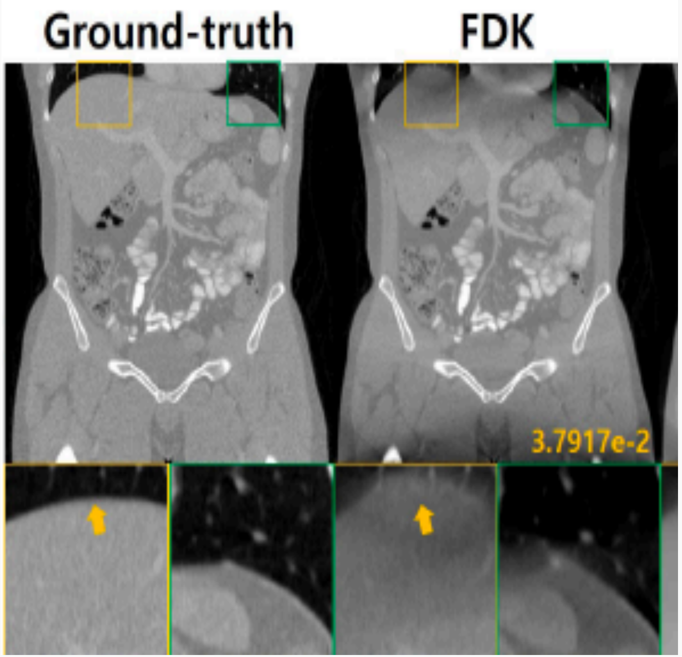
DBP Domain Conebeam Artifact Removal

Han et al, arXiv:1906.06854

Standard Method: FDK Algorithm



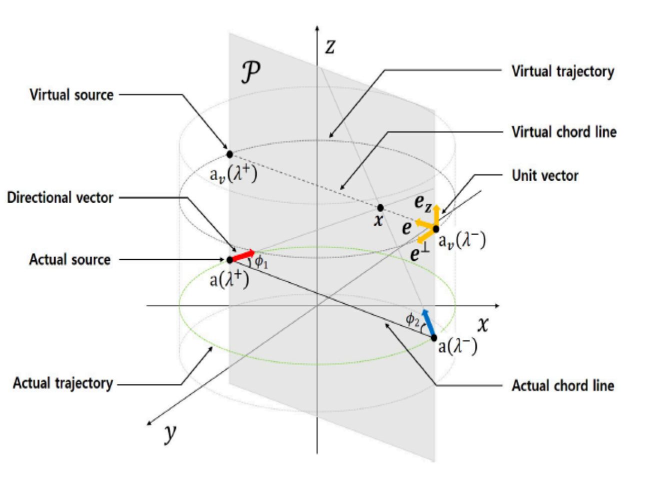
https://www.ndt.net/article/wcndt00/papers/idn730/idn730.htm

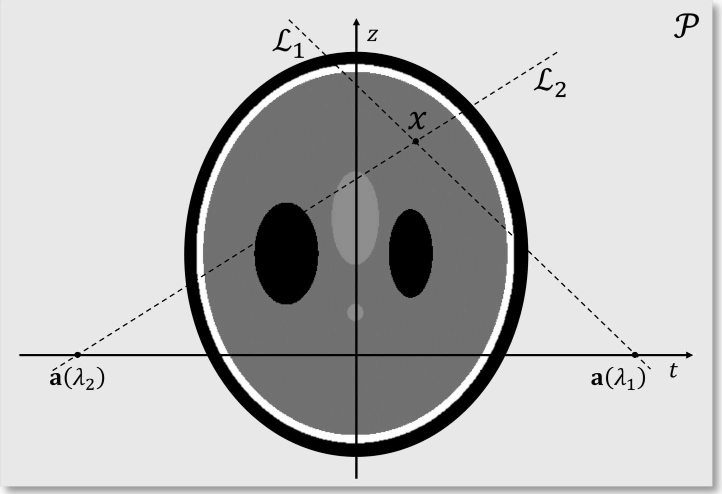


DBP Domain Conebeam Artifact Removal

Han et al, arXiv:1906.06854

Exact Factorization→ 2-D Deconvolution problem

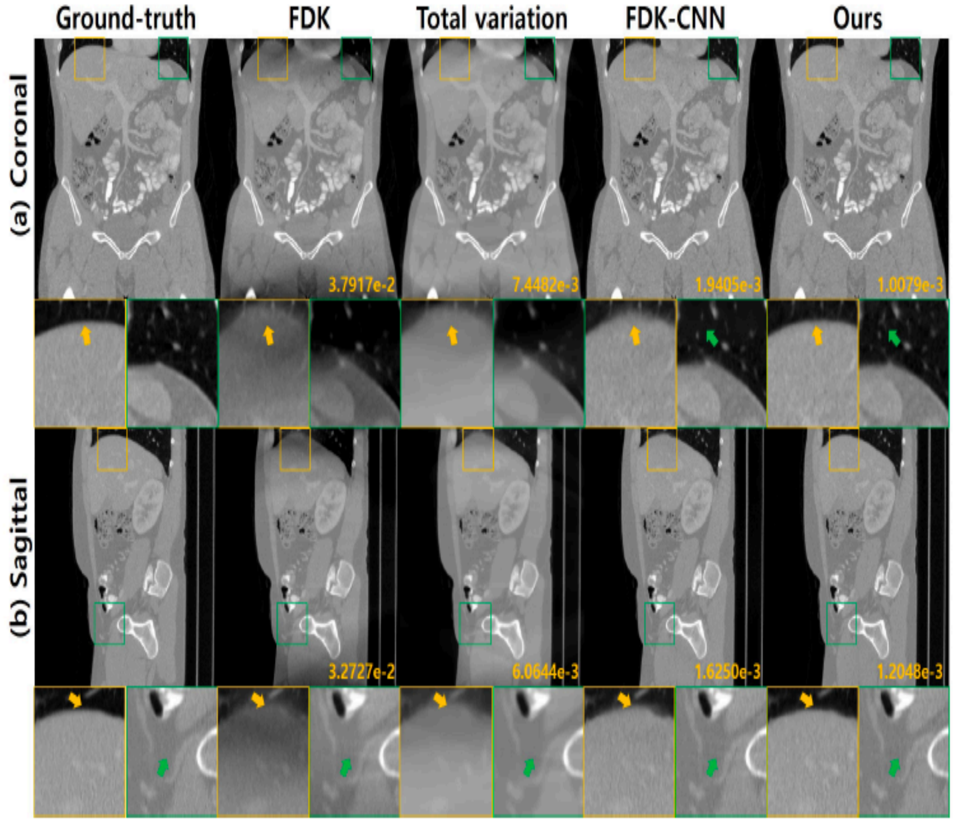


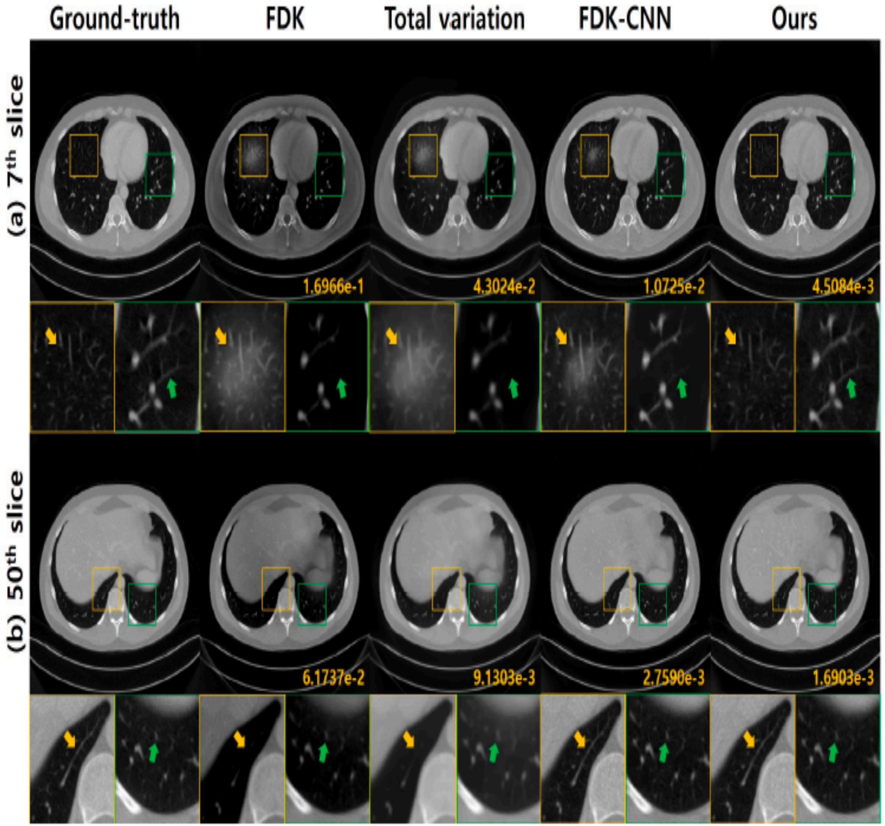


$$g(t,z) = \pi \int_{-\infty}^{\infty} h_H(t-\tau)\left(f(\tau, z_1(\tau)) + f(\tau, z_2(\tau)\right) d\tau$$

DBP Domain Conebeam Artifact Removal

Han et al,   arXiv:1906.06854

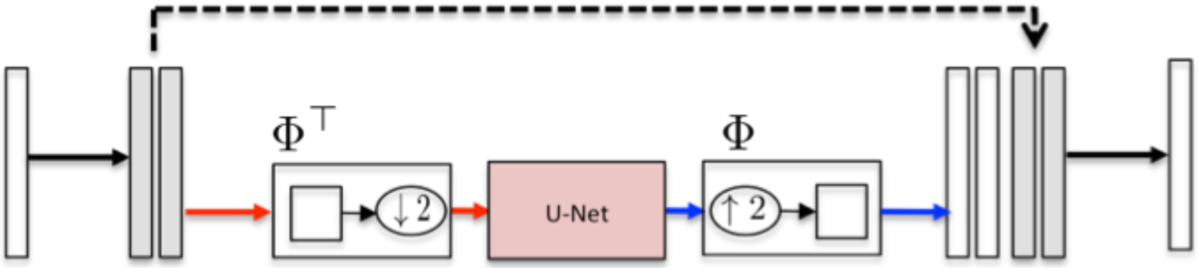




Improving U-Net

Ye et al, SIAM J. Imaging Science, 2018
Han et al, IEEE Trans. Medical Imaging, 2018
Yoo et al, SIAM J. Applied Math, 2019

Limitation of U-Net



$$\Phi^\top$$

$$\Phi$$

$$\Phi = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 1 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \\ 0 & 0 & \vdots & 1 \end{bmatrix}$$

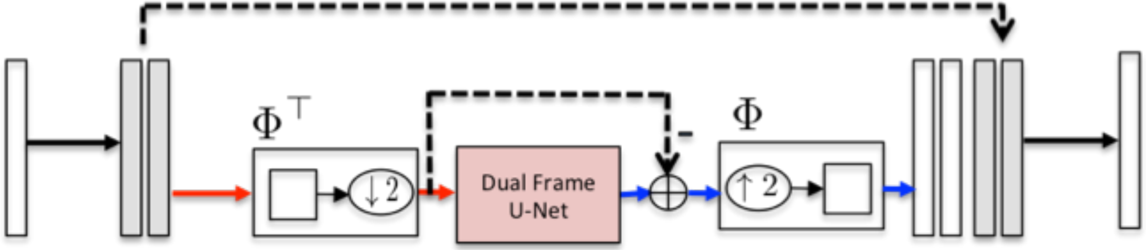
U-Net Pooling does NOT satisfy the frame condition

$$\Phi_{ext}\Phi_{ext}^\top = I + \Phi\Phi^\top \neq I$$

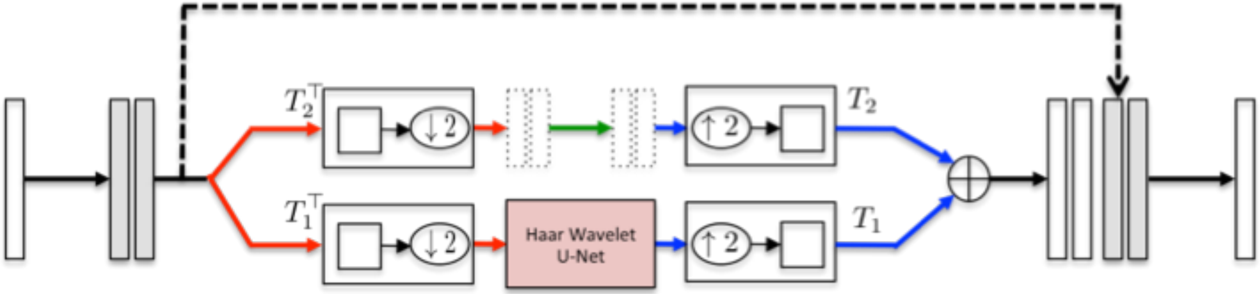
JC Ye et al, SIAM Journal Imaging Sciences, 2018
Y. Han et al, TMI, 2018.

Improving U-net by Frame Conditions

- Dual Frame U-net

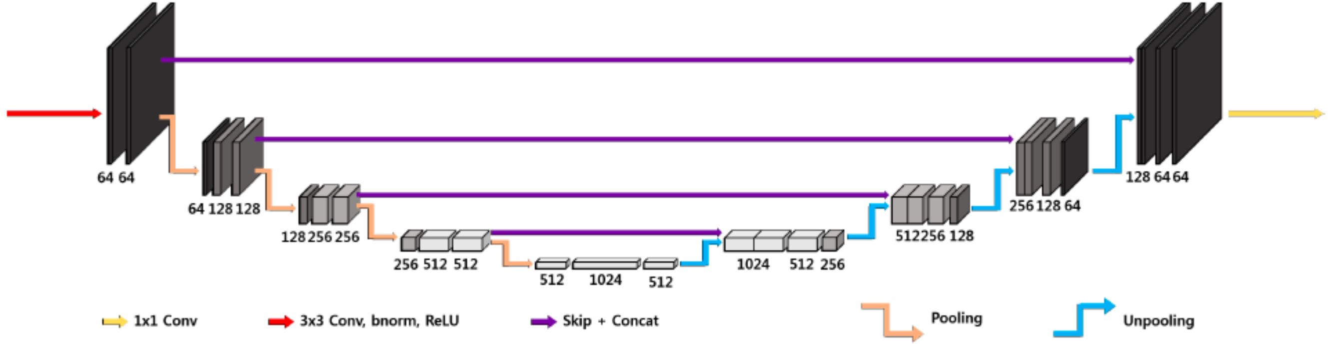


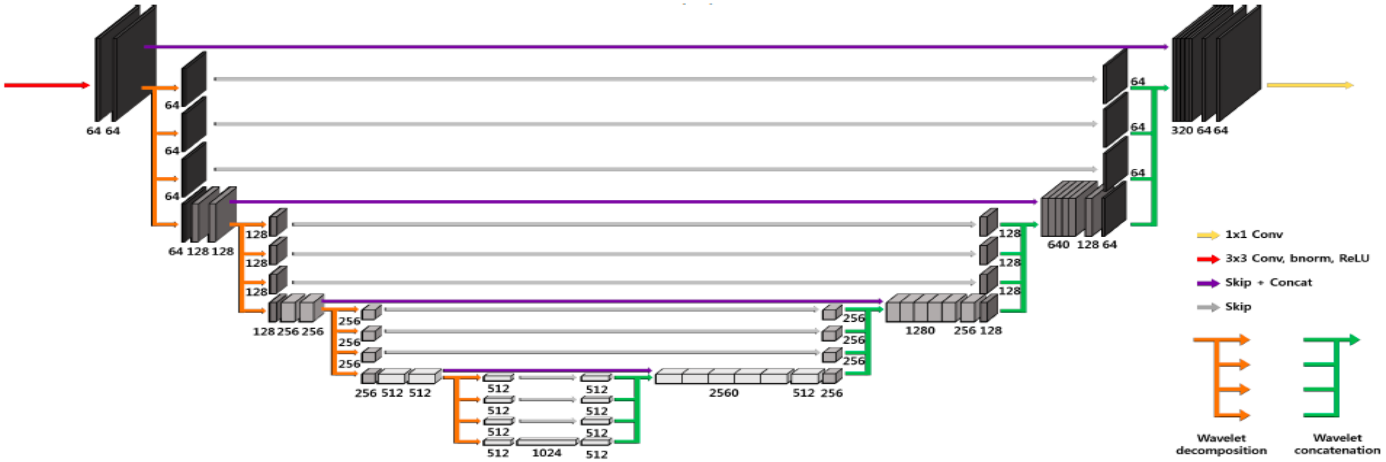
- Tight Frame U-net



JC Ye et al, SIAM Journal Imaging Sciences, 2018
Y. Han and J. C. Ye, TMI, 2018

U-Net vs Tight-Frame U-Net





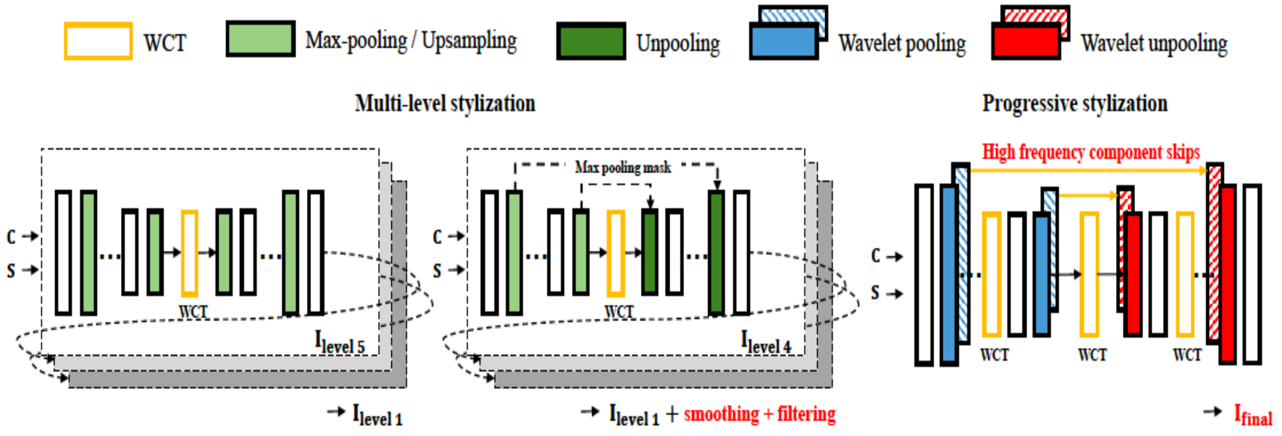
Y. Han and J. C. Ye, TMI, 2018; Yoo et al, SIJAM, 2018

Style Transfer : Power of Tight Frame U-net

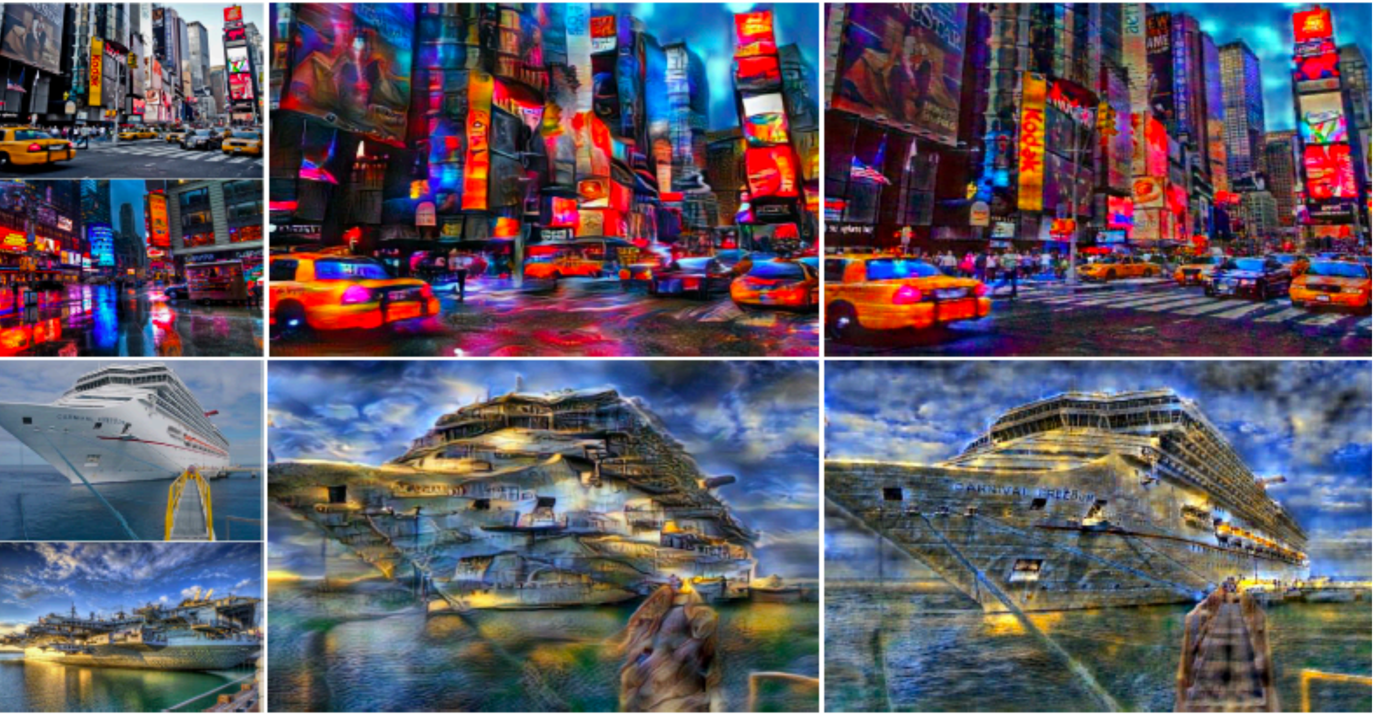
ICCV 2019

Photorealistic Style Transfer via Wavelet Transforms

Jaejun Yoo* Youngjung Uh* Sanghyuk Chun* Byeongkyu Kang Jung-Woo Ha
Clova AI Research, NAVER Corp.

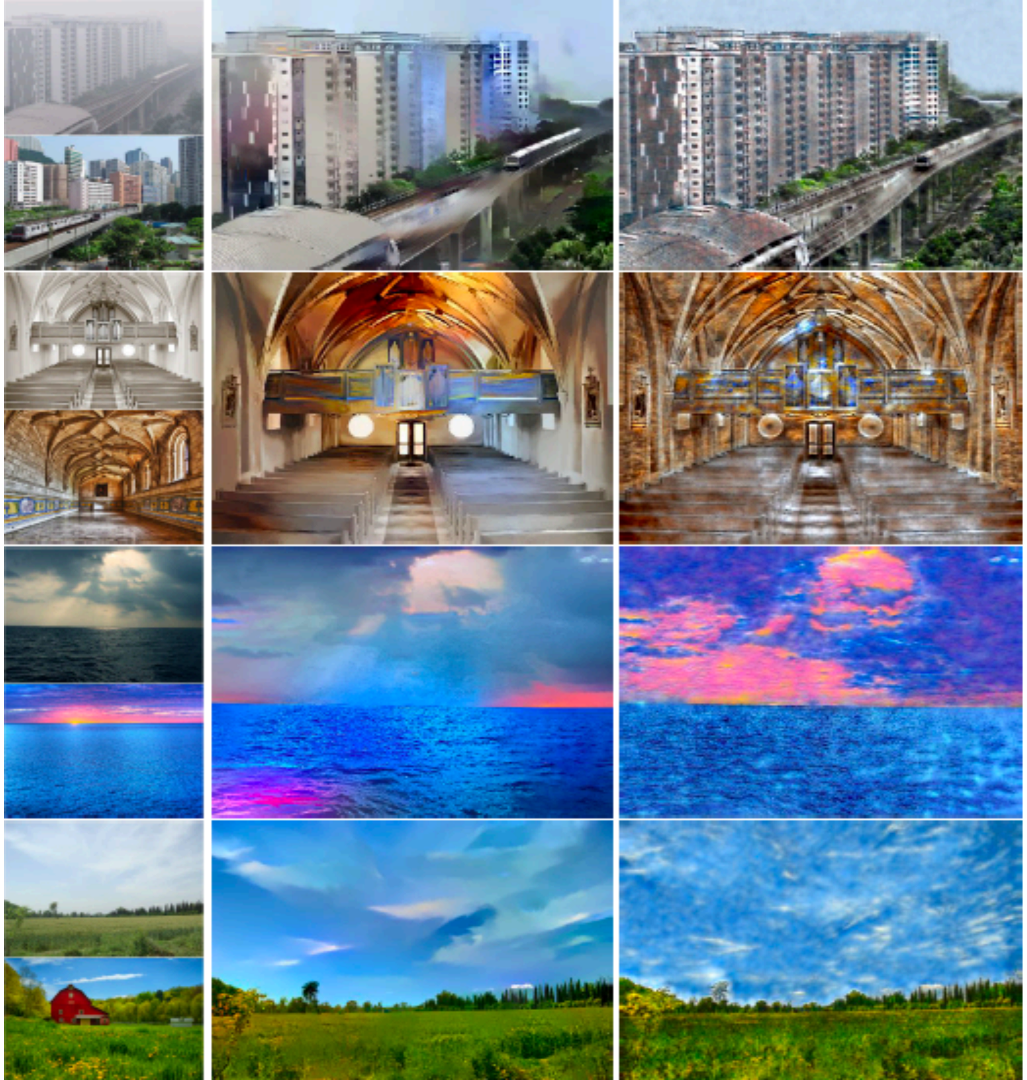


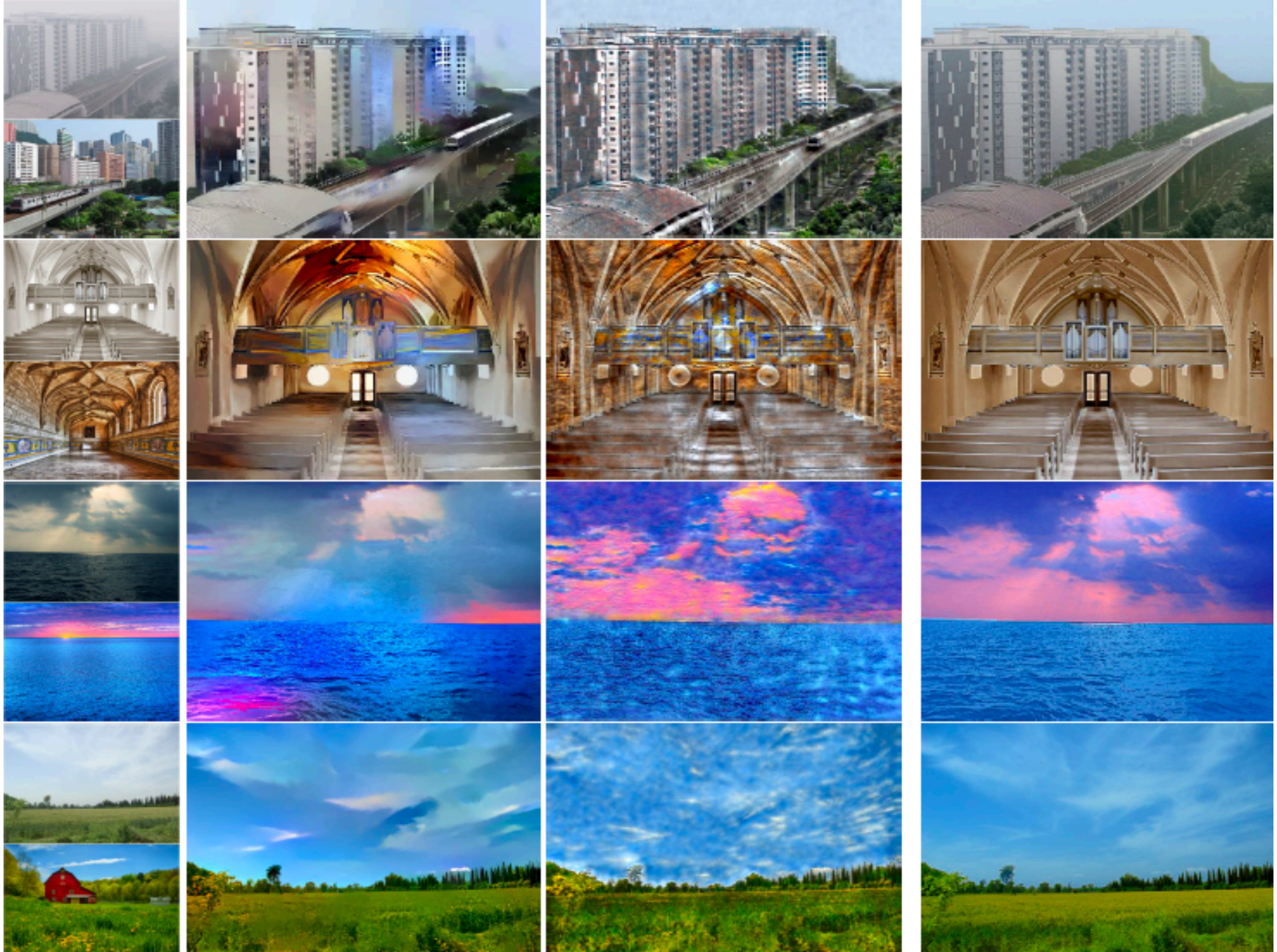
(a) WCT (artistic) (b) PhotoWCT (c) Ours (WCT^2)



(a) Input (b) WCT (c) PhotoWCT

(a) Input (b) WCT (c) PhotoWCT (d) Ours (WCT^2)





What if we do not have **label data** for training ?

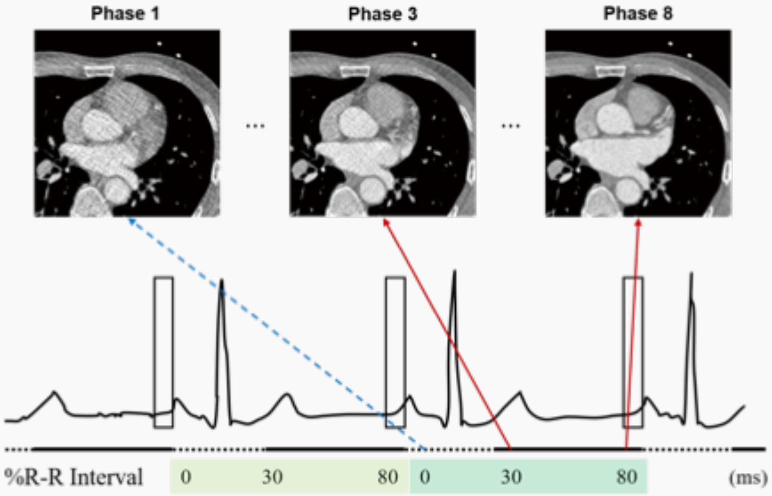
Yann LeCun's Cake Analogy

- "Pure" Reinforcement Learning (cherry)
  - The machine predicts a scalar reward given once in a while.
  - **A few bits for some samples**
- Supervised Learning (icing)
  - The machine predicts a category or a few numbers for each input
  - Predicting human-supplied data
  - **10→10,000 bits per sample**
- Unsupervised/Predictive Learning (cake)
  - The machine predicts any part of its input for any observed part.
  - Predicts future frames in videos
  - **Millions of bits per sample**

- (Yes, I know, this picture is slightly offensive to RL folks. But I'll make it up)

Motivation

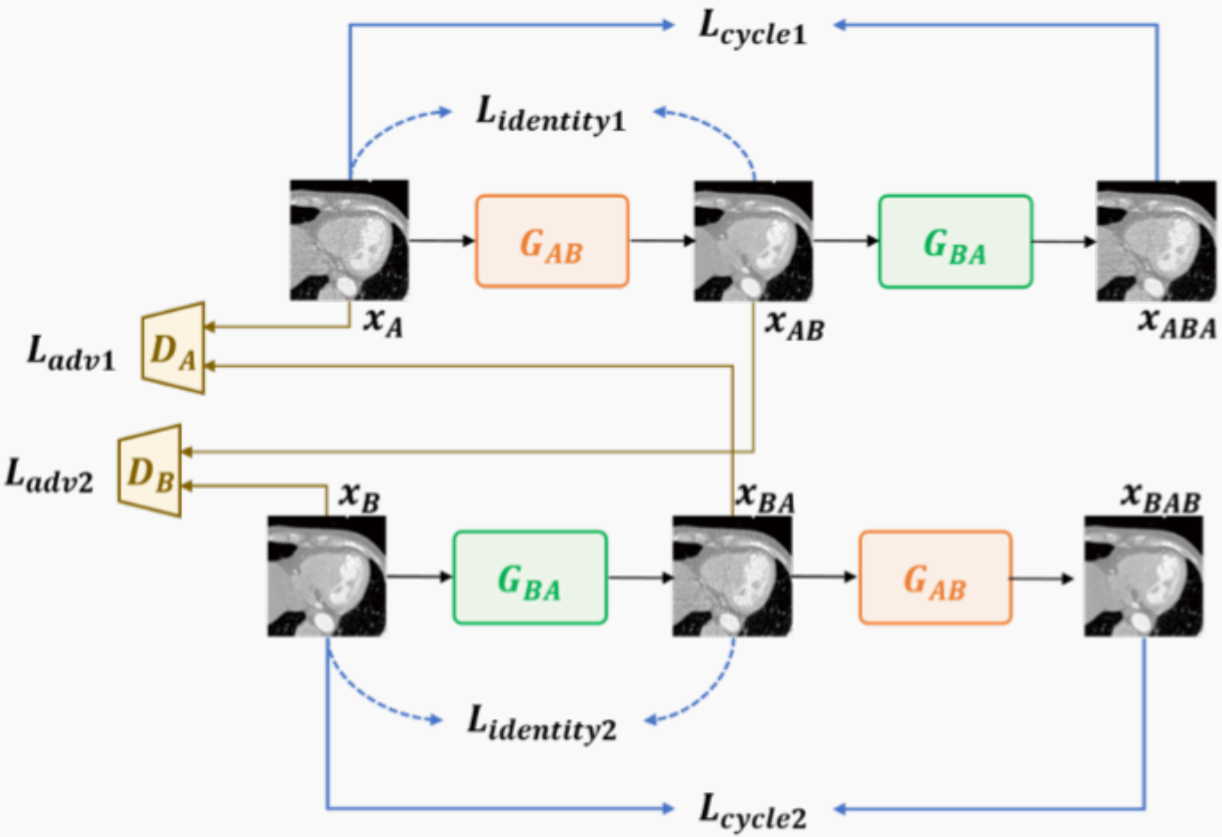
- **Multiphase Cardiac CT denoising**
  - Phase 1, 2: low-dose, Phase 3 ~ 10: normal dose
  - Goal: dynamic changes of heart structure
  - No reference available



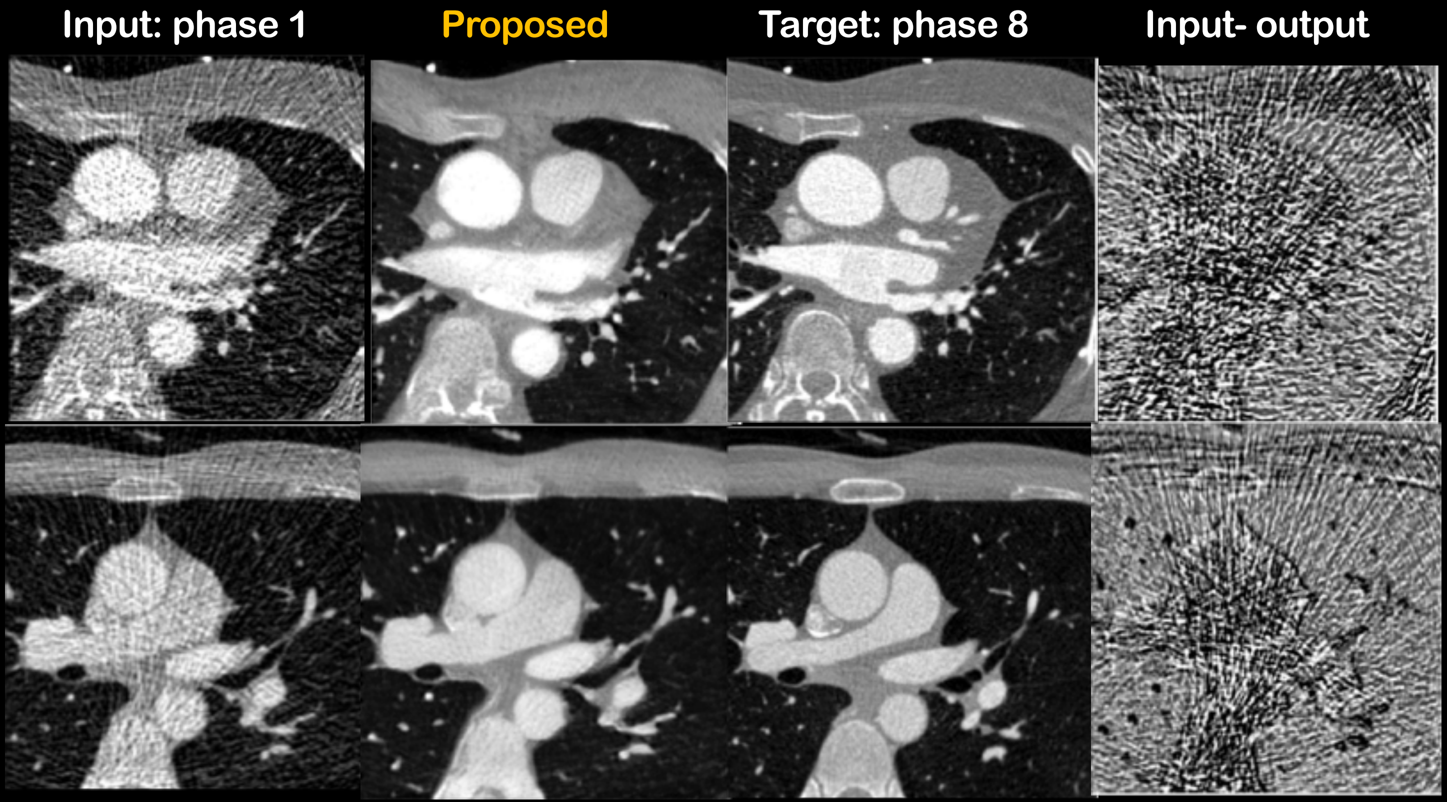
Kang et al, Medical Physics, 2018

Unsupervised Denoising for Low-Dose CT

Kang et al, Medical Physics, 2019

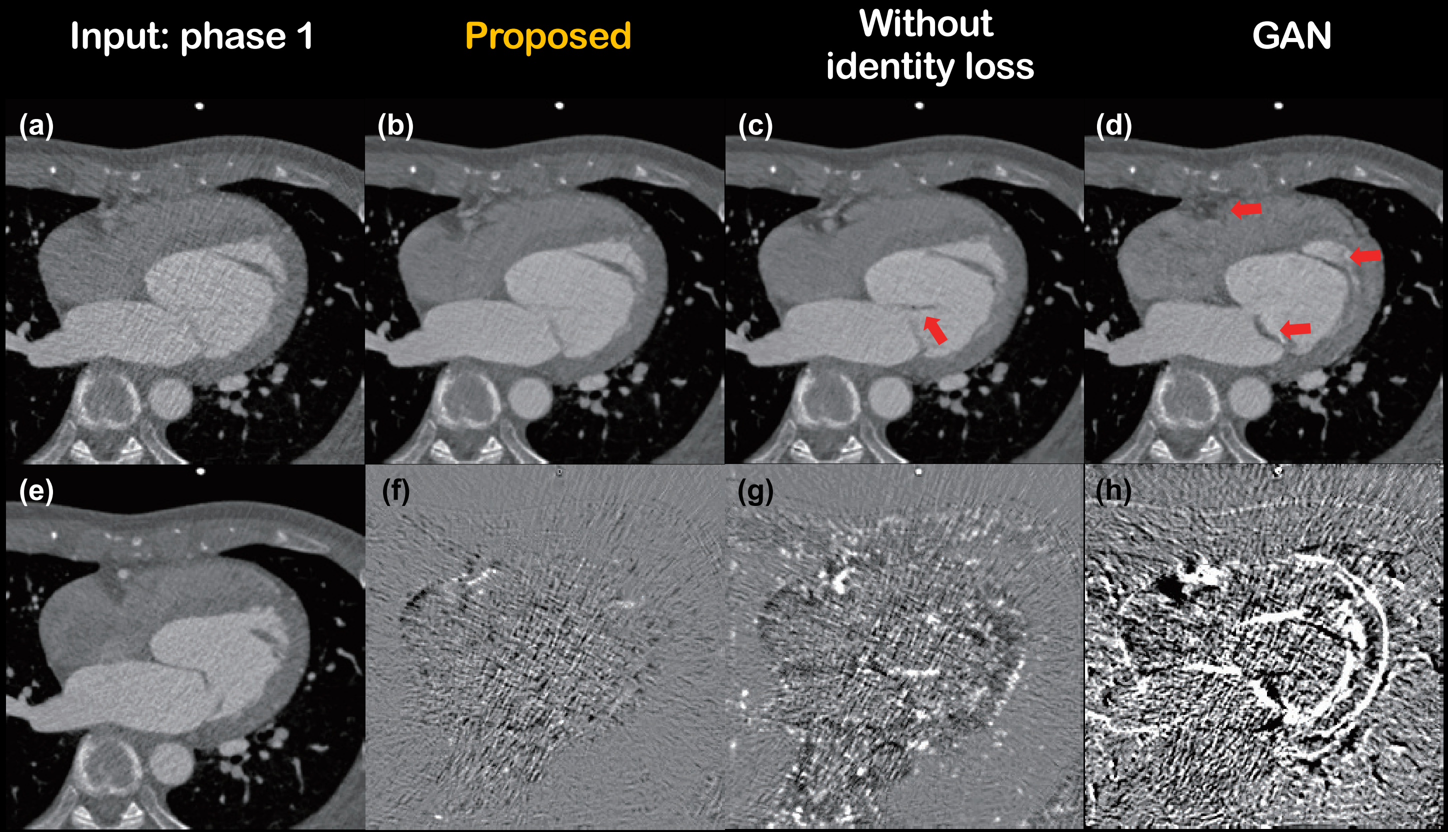


Lose dose (5%) → high dose



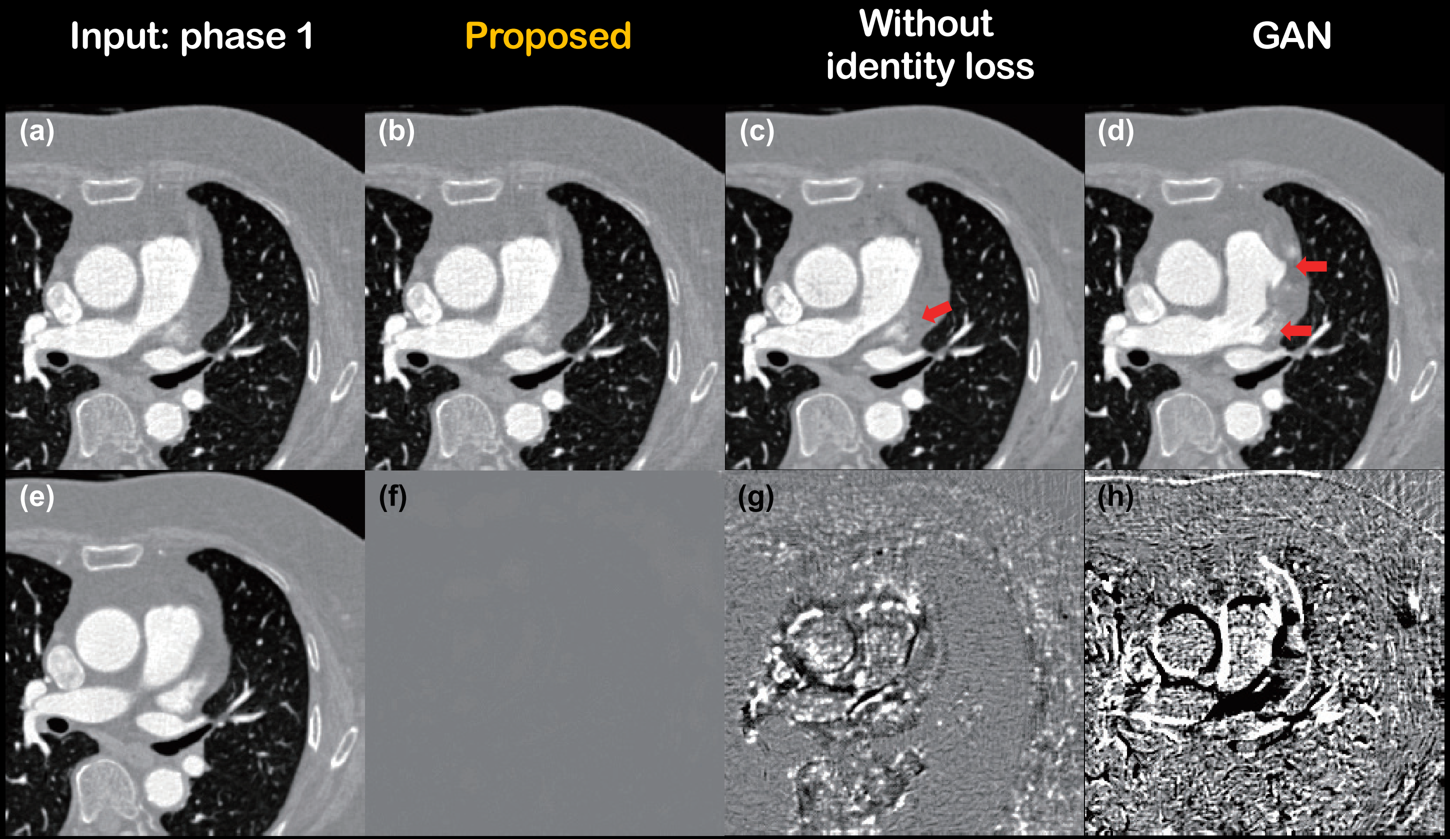
Kang et al, unpublished data

Ablation Study



Kang et al, Medical Physics, 2018

Ablation Study



Kang et al, Medical Physics, 2018

Geometry of CycleGAN

Sim et al, arXiv:1909.12116, 2019
Lim et al, arXiv:1908.09414, 2019
Kim et al, MICCAI, 2019
Kang et al, Medical Physics, 2019

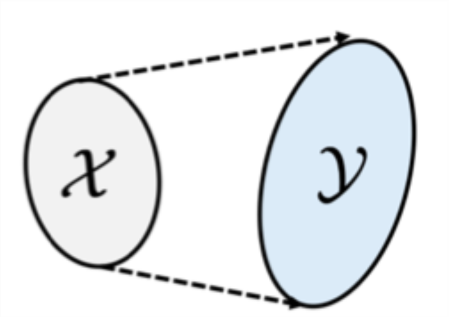
Optimal Transport (OT)

Villani, 2008; Peyre et al, 2019

Kantorovich Formulation: minimizing average transport cost

$$\min_{\pi} \int_{\mathcal{X}\times\mathcal{Y}} \|x - y\| d\pi(x, y)$$

Transport cost



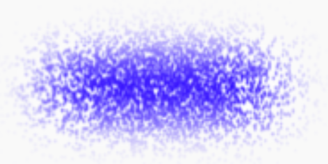
Optimal Transport (OT)

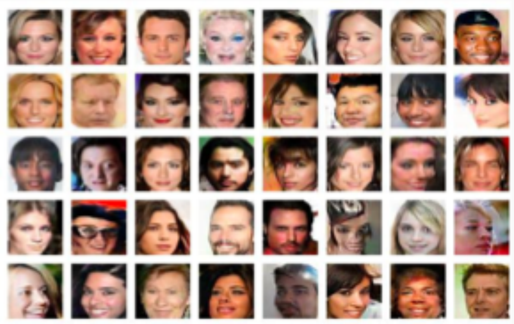
Villani, 2008; Peyre et al, 2019

Wasserstein GAN: if transport cost is given by $c(x,y) = \|x - y\|$

$$\min_{\Theta} \max_{\varphi} \int_{\mathcal{X}} \varphi(x) d\mu(x) - \int_{\mathcal{Y}} \varphi(G_{\Theta}(y)) d\nu(y)$$

1-Lipschitz function





Our Theoretical Findings

Sim et al, arXiv:1909.12116, 2019

Optimal transport: Kantorovich Formulation

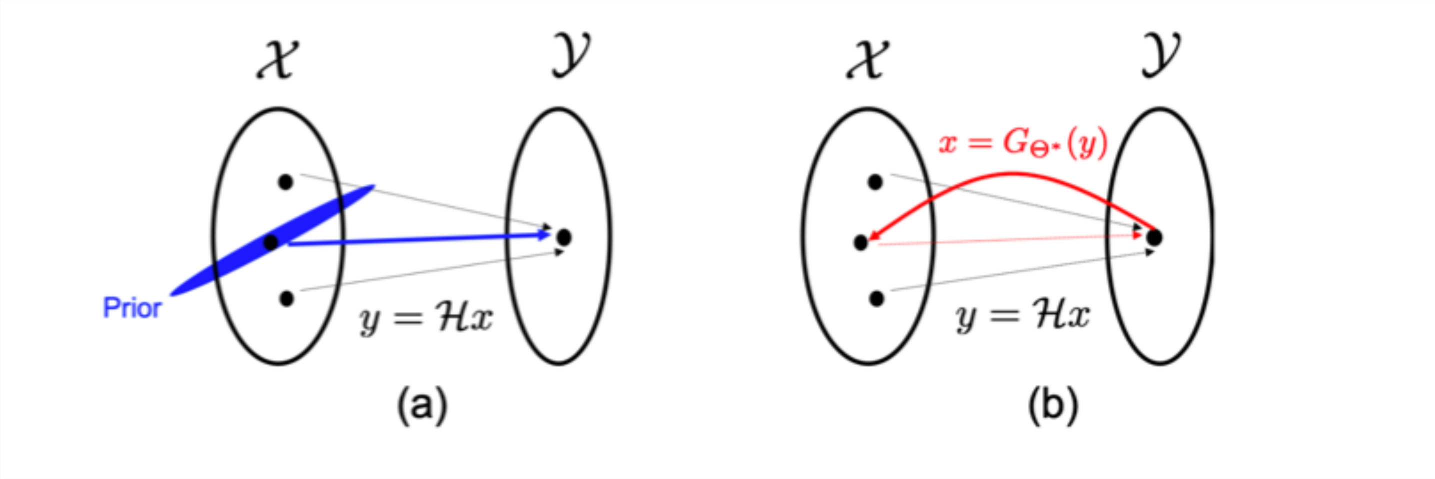
Transport cost

$$\min_{\pi} \int_{\mathcal{X}\times\mathcal{Y}} c(x,y)\, d\pi(x,y)$$

Our Penalized LS Formulation

CNN regularization

$$c(x; y, \Theta, \mathcal{H}) = \|y - \mathcal{H}x\|^q + \|G_\Theta(y) - x\|^p$$



Our Theoretical Findings

Sim et al, arXiv:1909.12116, 2019

Optimal transport with PLS as transport cost

$$\min_{\pi} \int_{\mathcal{X}\times\mathcal{Y}} \boxed{\|y - \mathcal{H}x\|^q + \|G_\Theta(y) - x\|^p} \, d\pi(x, y)$$

Both *x,y* are stochastic: stochastic generalization of PLS

Our Theoretical Findings

Sim et al, arXiv:1909.12116, 2019

Under some regularity conditions,
Kantorovich Dual Formulation → CycleGAN

$$\min_{\Theta,h}\max_{\Phi,\Xi} \ell_{cycle}(\Theta,\mathcal{H}) + \ell_{GAN}(\Theta,\mathcal{H};\Phi,\Xi)$$

Cycle consistency

$$\ell_{cycle}(\Theta,\mathcal{H}) = \rho\int_{\mathcal{X}} \|x - G_\Theta(\mathcal{H}x)\|^p d\mu(x) + \sigma\int_{\mathcal{Y}} \|y - \mathcal{H}G_\Theta(y)\|^q d\nu(y)$$

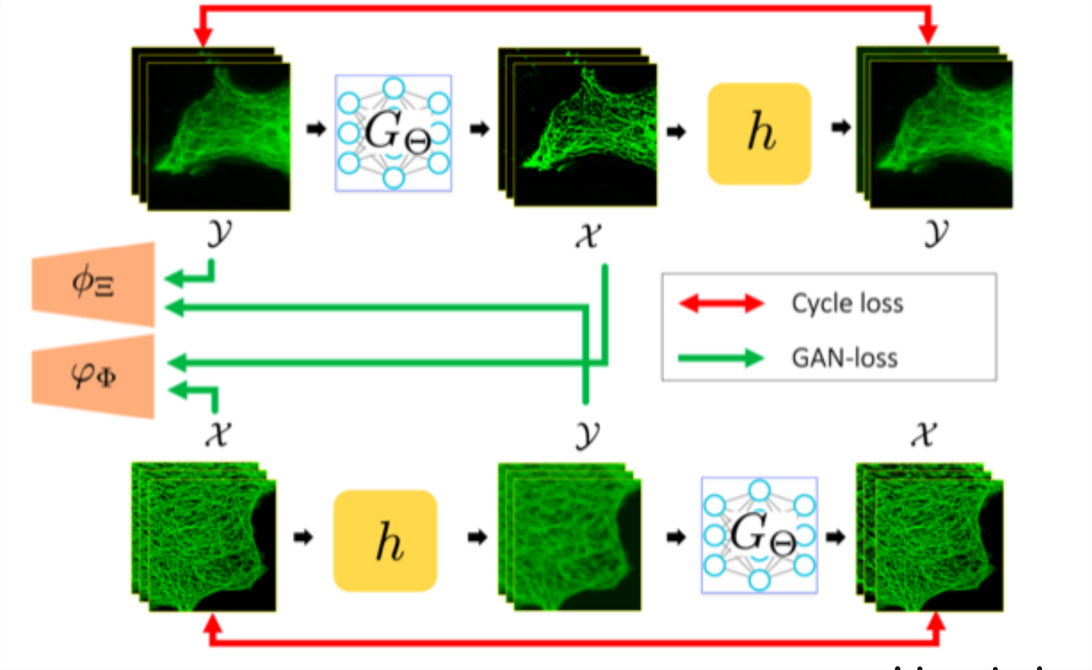
GAN

$$\ell_{GAN}(\Theta,\mathcal{H};\Phi,\Xi) = \int_{\mathcal{X}} \varphi_\Phi(x)d\mu(x) - \int_{\mathcal{Y}} \varphi_\Phi(G_\Theta(y))d\nu(y) + \int_{\mathcal{Y}} \psi_\Xi(y)d\nu(y) - \int_{\mathcal{X}} \psi_\Xi(\mathcal{H}x)d\nu(x)$$

THEORY-DRIVEN CYCLEGAN DESIGN :some snapshots

Unsupervised Blind Deconvolution Microscopy

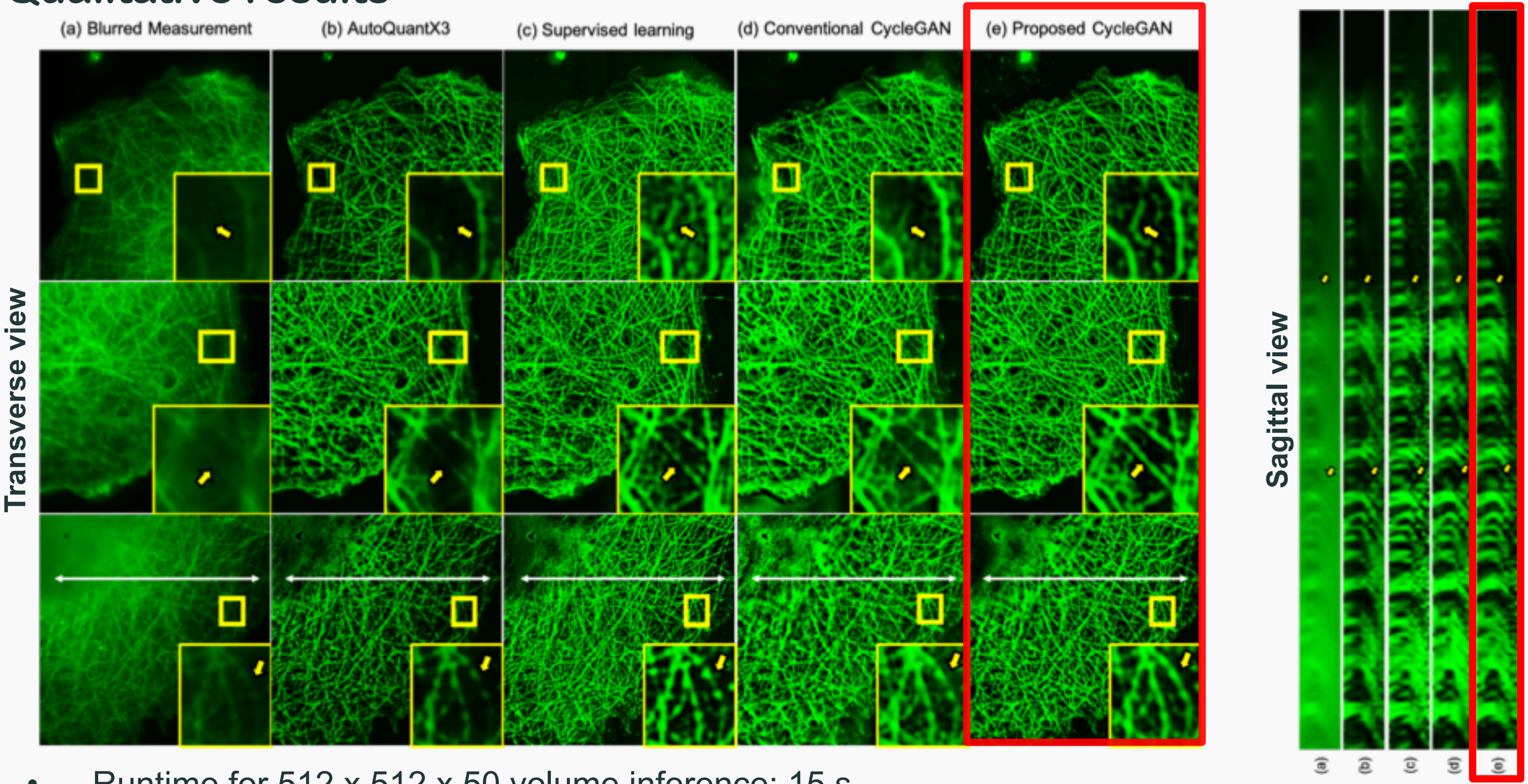
$$c(x, y; \Theta, h) = \|y - h * x\| + \|G_\Theta(y) - x\|$$



Lim et al, arXiv:1908.09414, 2019

Results on Real Microscopy Data

✓ **Qualitative results**

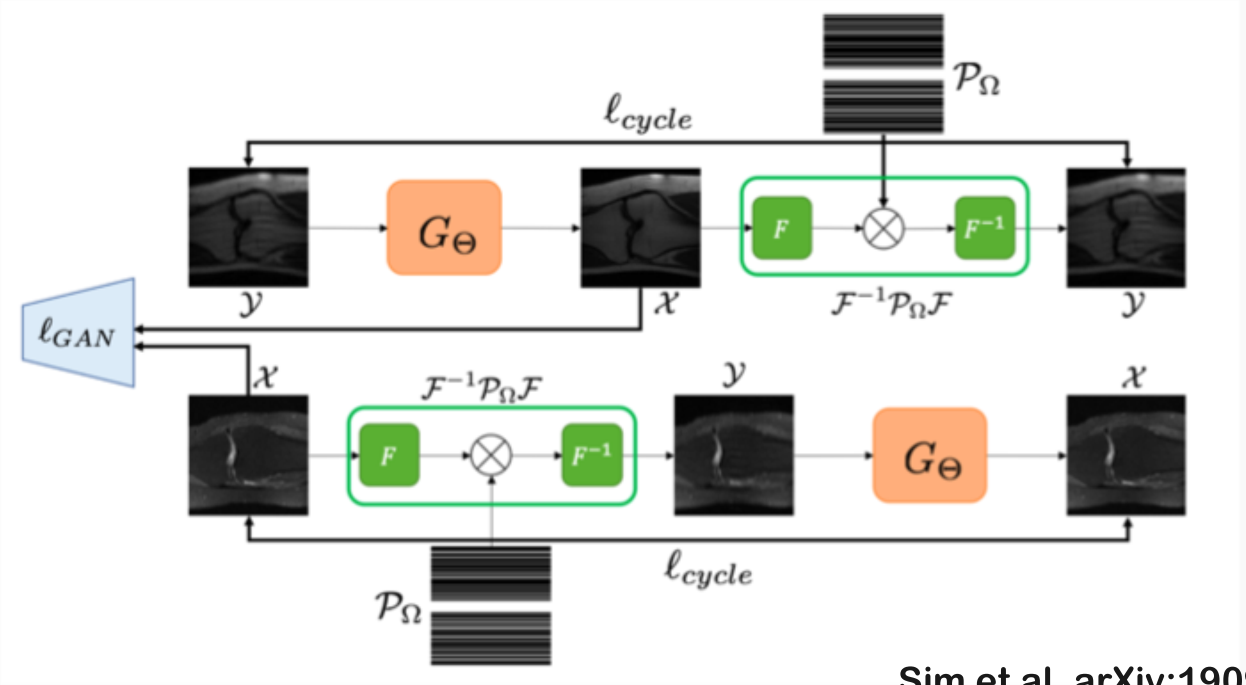


- Runtime for 512 x 512 x 50 volume inference: 15 s

Lim et al, arXiv:1908.09414, 2019

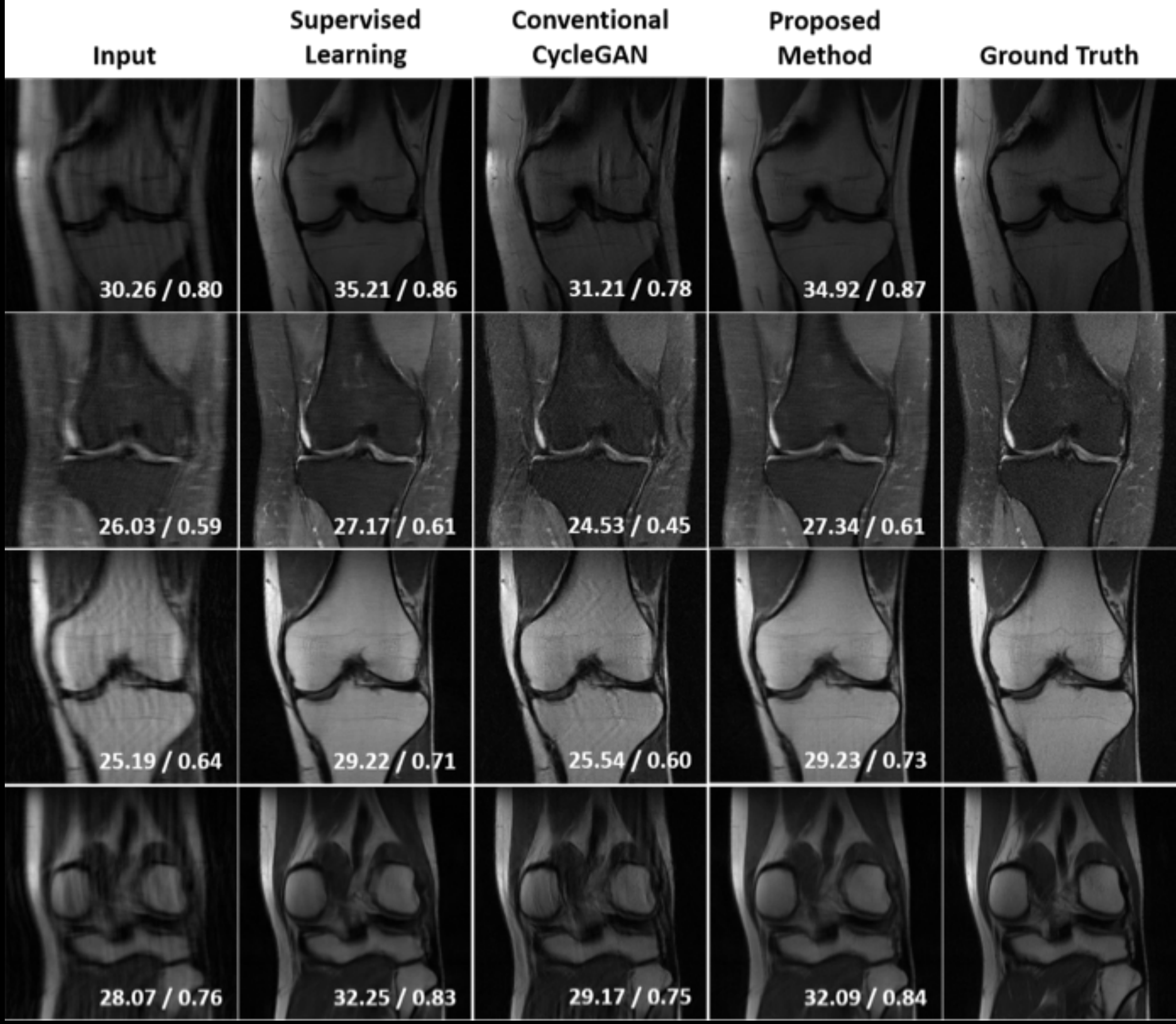
Unsupervised Learning for Accelerated MRI

$$c(x, y; \Theta) = \|y - \mathcal{F}^{-1}\mathcal{P}_\Omega \mathcal{F} x\| + \|G_\Theta(y) - x\|$$



Sim et al, arXiv:1909.12116, 2019

Results on Fast MR Data Set



Sim et al, arXiv:1909.12116, 2019

Summary

- Deep learning has becomes an important platform for medical imaging

- Our theoretical findings
  - Deep neural network with ReLU is a piecewise linear frame representation
  - Unsupervised learning can be solved with optimal transport with PLS cost == cycleGAN

- We can design problem-specific neural networks

Acknowledgement

- Daniel Rueckert (Imperial College)
- Florian Knoll (NYU)
- Fang Liu (Univ. of Wisconsin)
- Mehmet Akcakaya (Univ. of Minnesota)
- Dong Liang (SIAT, China)
- Dinggang Shen (UNC)
- Peder Larson (UCSF)
- Grant
  - NRF of Korea
  - Ministry of Trade Industry and Energy
- Hyunwook Park (KAIST)
- Sung-hong Park (KAIST)
- Jongho Lee (SNU)
- Doshik Hwang (Yonsei Univ)
- Won-Jin Moon (KonkukUniv Medical Center)
- Eungyeop Kim (Gachon Univ. Medical Center)
- Leonard Sunwoo (SNUBH)
- Kyuhwan Jung (Vuno)