Theory of Deep Learning

Gitta Kutyniok

(Technische Universität Berlin)

Spring School Series “Models and Data”
DASIV SmartState Center, USC, March 17–20, 2019

The 21st Century

Various technological advances in the 21st century are only possible through *integrated mathematical modeling, simulation, and optimization*.

The 21st Century

Various technological advances in the 21st century are only possible through *integrated mathematical modeling, simulation, and optimization*.

Further Examples:

- Turbines

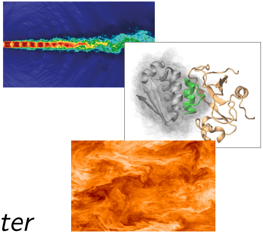
  $\rightsquigarrow$ *Adjoint based jet-noise minimization*

- Atomistic molecular dynamics

  $\rightsquigarrow$ *Simulations with ultralong timescales*

- Star formation

  $\rightsquigarrow$ *Understanding of turbulent accretion of matter*

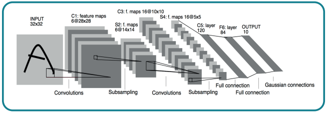


There is a pressing need to go beyond
pure modeling, simulation, and optimization approaches!

The Data Science Side: Impact of Deep Learning

Self-Driving Cars

Surveillance



Legal Issues

Health Care

Very few theoretical results explaining their success!

From Data-Driven to Model-Based Approaches

Problems, Viewpoints and Solution Strategies:

Data Science
Modeling, Simulation, Optimization

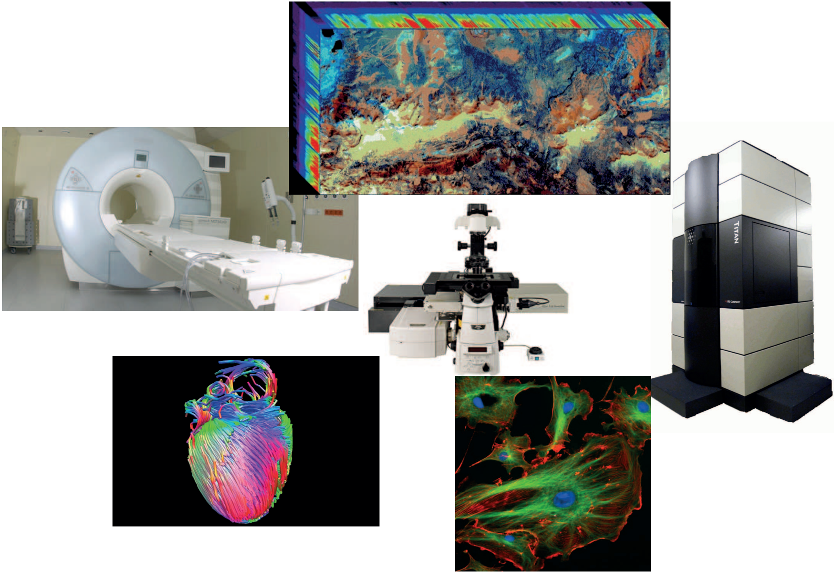
- Pure data-driven approaches.
  Detect structural components in data sets!
- Machine learning with physical constraints.
  Insert physical information in machine learning algorithm!
- Parametric differential equations.
  Learn parameters from given data sets!
- Data assimilation.
  Combine sparse data with physical model to generate a general model!
- Data analysis on simulation data.
  Study simulation generated data in search of underlying laws!

Optimal balancing of
data-driven and model-based approaches!

Outline

1. Inverse Problems
    - Examples from Imaging Sciences
    - Sparse Regularization via Shearlets
2. Solving Inverse Problems with Deep Neural Networks
    - A Bit of History
3. Limited-Angle Computer Tomography
    - Learning (only) the Invisible: A Hybrid Approach
    - Numerical Experiments
4. Conclusions

Modern Imaging Science



Tikhonov Regularization

Standard Tikhonov Regularization:

Given an ill-posed inverse problem $Kx = y$, where $K : X \to Y$, an approximate solution $x^{\alpha} \in X$, $\alpha > 0$, can be determined by minimizing

$$J_{\alpha}(x) := \|Kx - y\|^2 + \alpha \|x\|^2, \quad x \in X.$$

Tikhonov Regularization

Standard Tikhonov Regularization:

Given an ill-posed inverse problem $Kx = y$, where $K : X \to Y$, an approximate solution $x^\alpha \in X$, $\alpha > 0$, can be determined by minimizing

$$J_\alpha(x) := \|Kx - y\|^2 + \alpha\|x\|^2, \quad x \in X.$$

Generalization:

$$\tilde{J}_\alpha(x) := \|Kx - y\|^2 + \alpha\mathcal{P}(x), \quad x \in X.$$

The penalty term $\mathcal{P}$

- ensures continuous dependence on the data,
- incorporates properties of the solution.

Some Examples for $\mathcal{P}$:

$$\|x\|_{TV}, \;\; \|x\|_{H^s}, \;\; \|(\langle x, \psi_\lambda \rangle)_\lambda\|_1, \ldots$$

Tikhonov Regularization

Standard Tikhonov Regularization:

Given an ill-posed inverse problem $Kx = y$, where $K : X \to Y$, an approximate solution $x^\alpha \in X$, $\alpha > 0$, can be determined by minimizing

$$J_\alpha(x) := \|Kx - y\|^2 + \alpha\|x\|^2, \quad x \in X.$$

Generalization:

$$\tilde{J}_\alpha(x) := \|Kx - y\|^2 + \alpha\mathcal{P}(x), \quad x \in X.$$

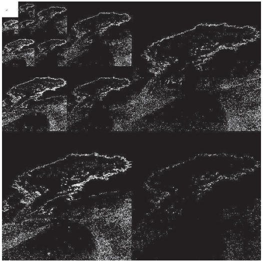
The penalty term $\mathcal{P}$

- ensures continuous dependence on the data,
- incorporates properties of the solution.

Some Examples for $\mathcal{P}$:

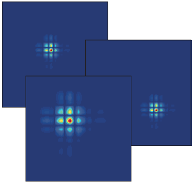
$$\|x\|_{TV}, \;\; \|x\|_{H^s}, \;\; \|(\langle x, \psi_\lambda \rangle)_\lambda\|_1, \ldots$$

The World is Compressible!



Wavelet Transform (JPEG2000):

$$f \mapsto (\langle f, \psi_{j,m} \rangle)_{j,m}.$$



Definition: For a wavelet $\psi \in L^2(\mathbb{R}^2)$, a wavelet system is defined by

$$\{\psi_{j,m} : j \in \mathbb{Z}, m \in \mathbb{Z}^2\}, \quad \text{where } \psi_{j,m}(x) := 2^j \psi(2^j x - m).$$

Sparsity

Novel Paradigm:

For each class of data, there exists a sparsifying system!

Sparsity

Novel Paradigm:

For each class of data, there exists a sparsifying system!

Two Viewpoints of 'Sparsifying System':

Let $\mathcal{C} \subseteq \mathcal{H}$ and $(\psi_\lambda)_\lambda \subseteq \mathcal{H}$.

- Decay of Coefficients. Consider the decay for $n \to \infty$ of the sorted sequence of coefficients

$$(|\langle x, \psi_{\lambda_n} \rangle|)_n \quad \text{for all } x \in \mathcal{C}.$$

- Approximation Properties. Consider the decay for $N \to \infty$ of the error of best N-term approximation, i.e.,

$$\inf_{\#\Lambda_N = N, (c_\lambda)_\lambda} \left\| x - \sum_{\lambda \in \Lambda_N} c_\lambda \psi_\lambda \right\| \quad \text{for all } x \in \mathcal{C}.$$

Sparsity-Based Approaches to Inverse Problems

Compressed Sensing (Candès, Romberg, Tao and Donoho; 2006) :

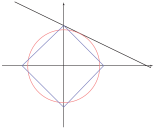
- Goal: Solve an underdetermined linear problem

$$y = Ax, \quad A \text{ an } n \times N\text{-matrix with } n \ll N,$$

for a solution $x \in \mathbb{R}^N$ admitting a sparsifying system $(\psi_\lambda)_\lambda$.

- Approach: Recover x by the ℓ_1-analysis minimization problem

$$\min_{\tilde{x}} \|(\langle \tilde{x}, \psi_\lambda \rangle)_\lambda\|_1 \text{ subject to } y = A\tilde{x}$$



Sparsity-Based Approaches to Inverse Problems

Compressed Sensing (Candès, Romberg, Tao and Donoho; 2006) :

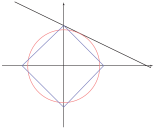
- Goal: Solve an underdetermined linear problem

$$y = Ax, \quad A \text{ an } n \times N\text{-matrix with } n \ll N,$$

  for a solution $x \in \mathbb{R}^N$ admitting a sparsifying system $(\psi_\lambda)_\lambda$.

- Approach: Recover x by the ℓ_1-analysis minimization problem

$$\min_{\tilde{x}} \|(\langle \tilde{x}, \psi_\lambda \rangle)_\lambda\|_1 \text{ subject to } y = A\tilde{x}$$



Some Earlier Footprints in Inverse Problems:

- Donoho (1995): Wavelet-Vaguelette decomposition.
- Chambolle, DeVore, Lee, Lucier (1998): Penalty on the Besov norm.
- Daubechies, Defries, De Mol (2004): General sparsity constraints.
- ...

Regularization by Sparsity

Functional with ℓ_1-Penalty Term:

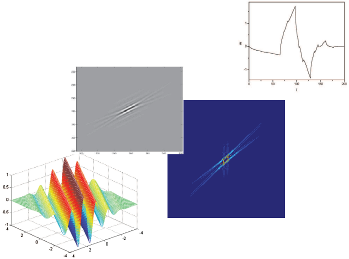
$$J_\alpha(x) := \|Kx - y\|^2 + \alpha \|(\langle x, \psi_\lambda \rangle)_\lambda\|_1, \quad x \in X.$$

Regularization by Sparsity

Functional with ℓ_1-Penalty Term:

$$J_\alpha(x) := \|Kx - y\|^2 + \alpha\|(\langle x, \psi_\lambda\rangle)_\lambda\|_1, \quad x \in X.$$

Applied Harmonic Analysis Approach:
Wavelets, Ridgelets, Curvelets, Shearlets,...



Desiderata:

- Multiscale representation system.
- Partition of Fourier domain.
- Fast algorithms: $x \mapsto (\langle x, \psi_\lambda\rangle)_\lambda \rightsquigarrow x$.
- Optimality for the considered class.
  $\rightsquigarrow$ Here: Functions governed by anisotropic features.

Shearlets come into Play

Mathematical Model for Images

Key Observation:

> Images are governed by edge-like structures!

Mathematical Model for Images

Key Observation:

Images are governed by edge-like structures!

Definition (Donoho; 2001):

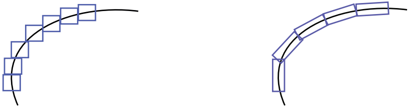
Let $\nu > 0$. We then define the class of *cartoon-like functions* by

$$\mathcal{E}^2(\mathbb{R}^2) = \{f \in L^2(\mathbb{R}^2) : f = f_1 + \chi_B f_2\},$$

where $B \subset [0,1]^2$ with $\partial B \in C^2$, and the functions f_1 and f_2 satisfy $f_1, f_2 \in C_0^2([0,1]^2)$, $\|f_1\|_{C^2}, \|f_2\|_{C^2}, \|\partial B\|_{C^2} < \nu$.

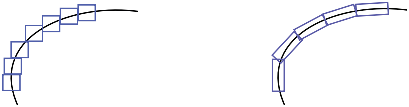
Key Ideas of the Shearlet Construction

Wavelet versus Shearlet Approximation:



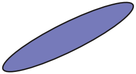
Key Ideas of the Shearlet Construction

Wavelet versus Shearlet Approximation:



Parabolic scaling ('width ≈ length^2'):

$$A_{2^j} = \begin{pmatrix} 2^j & 0 \\ 0 & 2^{j/2} \end{pmatrix}, \quad j \in \mathbb{Z}.$$



Orientation via shearing:

$$S_k = \begin{pmatrix} 1 & k \\ 0 & 1 \end{pmatrix}, \quad k \in \mathbb{Z}.$$

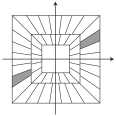
Advantage:

- Shearing leaves the digital grid $\mathbb{Z}^2$ invariant.
- Uniform theory for the continuum and digital situation.

(Cone-adapted) Discrete Shearlet Systems

Definition (K, Labate; 2006):

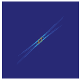
The (cone-adapted) discrete shearlet system $\mathcal{SH}(c; \phi, \psi, \tilde{\psi})$, $c > 0$, generated by $\phi \in L^2(\mathbb{R}^2)$ and $\psi, \tilde{\psi} \in L^2(\mathbb{R}^2)$ is the union of



$$\{\phi(\cdot - cm) : m \in \mathbb{Z}^2\},$$

$$\{2^{3j/4}\psi(S_k A_{2^j} \cdot - cm) : j \geq 0, |k| \leq \lceil 2^{j/2} \rceil, m \in \mathbb{Z}^2\},$$

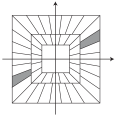
$$\{2^{3j/4}\tilde{\psi}(\tilde{S}_k \tilde{A}_{2^j} \cdot - cm) : j \geq 0, |k| \leq \lceil 2^{j/2} \rceil, m \in \mathbb{Z}^2\}.$$



(Cone-adapted) Discrete Shearlet Systems

Definition (K, Labate; 2006):

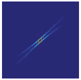
The (cone-adapted) discrete shearlet system $\mathcal{SH}(c;\phi,\psi,\tilde{\psi})$, $c > 0$, generated by $\phi \in L^2(\mathbb{R}^2)$ and $\psi, \tilde{\psi} \in L^2(\mathbb{R}^2)$ is the union of



$$\{\phi(\cdot - cm) : m \in \mathbb{Z}^2\},$$

$$\{2^{3j/4}\psi(S_k A_{2^j} \cdot - cm) : j \geq 0, |k| \leq \lceil 2^{j/2} \rceil, m \in \mathbb{Z}^2\},$$

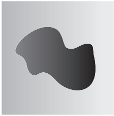
$$\{2^{3j/4}\tilde{\psi}(\tilde{S}_k \tilde{A}_{2^j} \cdot - cm) : j \geq 0, |k| \leq \lceil 2^{j/2} \rceil, m \in \mathbb{Z}^2\}.$$



Theorem (K, Lim; 2011):

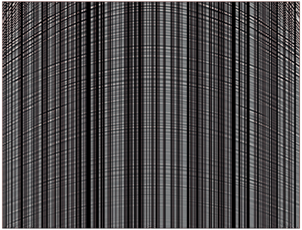
Let $\phi, \psi, \tilde{\psi} \in L^2(\mathbb{R}^2)$ be compactly supported, and let $\hat{\psi}$, $\hat{\tilde{\psi}}$ satisfy certain decay condition. Then $\mathcal{SH}(\phi,\psi,\tilde{\psi})$ provides an optimally sparse approximation of $f \in \mathcal{E}^2(\mathbb{R}^2)$, i.e.,

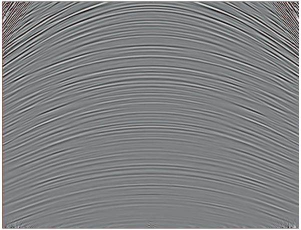
$$\|f - f_N\|_2^2 \lesssim N^{-2}(\log N)^3 \text{ and } |\langle f, \sigma_{\eta_n}\rangle| \lesssim n^{-\frac{3}{2}}(\log n)^{\frac{3}{2}}.$$



Applications

Inpainting:

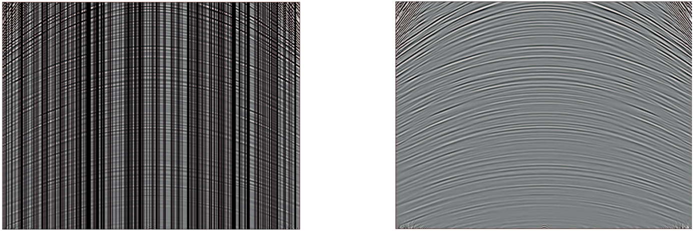




(Source: K, Lim; 2012)

Applications

Inpainting:



(Source: K, Lim; 2012)

2D&3D (parallelized) Fast Shearlet Transform (`www.ShearLab.org`):

- Matlab *(K, Lim, Reisenhofer; 2013)*
- Julia *(Loarca; 2017)*
- Python *(Look; 2018)*
- Tensorflow *(Loarca; 2019)*

Edge Analysis

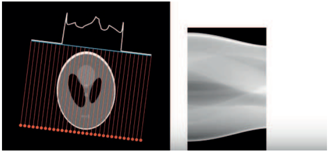
The location of edges in images and their geometric properties can contain valuable information about the depicted object. Several algorithms are available that use the anisotropic nature of shearlets to obtain this information.

Welcome to shearlab.org

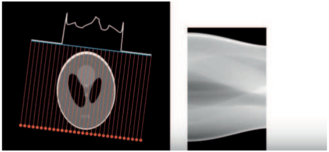
ShearLab is a MATLAB library developed for processing two- and threedimensional data with a certain class of basis functions named shearlets. Such shearlet systems are particularly well adapted to represent **anisotropic features** (such as curves) that are often crucial

Mathematical Modeling Reaches a Barrier

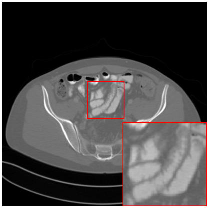
Computed Tomography (CT)

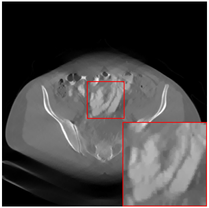


Computed Tomography (CT)



Problem with Limited-Angle Tomography:



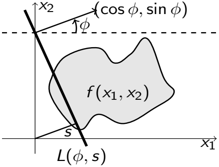


The data is too complex for mathematical modeling!

Limited Angle-(Computed) Tomography

A CT scanner samples the *Radon transform*

$$\mathcal{R}f(\phi, s) = \int_{L(\phi,s)} f(x)dS(x),$$

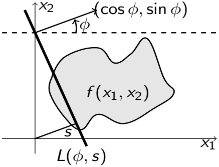


for $L(\phi, s) = \{x \in \mathbb{R}^2 : x_1 \cos(\phi) + x_2 \sin(\phi) = s\}$,
$\phi \in [-\pi/2, \pi/2)$, and $s \in \mathbb{R}$.

Limited Angle-(Computed) Tomography

A CT scanner samples the *Radon transform*

$$\mathcal{R}f(\phi, s) = \int_{L(\phi,s)} f(x)dS(x),$$



for $L(\phi, s) = \{x \in \mathbb{R}^2 : x_1 \cos(\phi) + x_2 \sin(\phi) = s\}$, $\phi \in [-\pi/2, \pi/2)$, and $s \in \mathbb{R}$.

Challenging inverse problem if $\mathcal{R}f(\cdot, s)$ is only sampled on $[-\phi, \phi] \subset [-\pi/2, \pi/2)$.

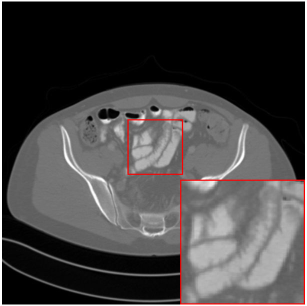
Applications: Dental CT, breast tomosynthesis, electron tomography,...

Model-Based Approaches Fail

Sparse Regularization:

$$\operatorname{argmin}_f \left[\underbrace{\|\mathcal{R}f - g\|^2}_{\text{Data fidelity term}} + \alpha \cdot \underbrace{\|(\langle f, \psi_{j,k,m}\rangle)_{j,k,m}\|_1}_{\text{Penalty term}} \right].$$

Clinical Data:



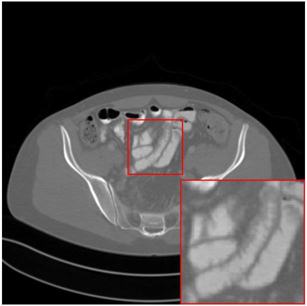
Original Image

Model-Based Approaches Fail

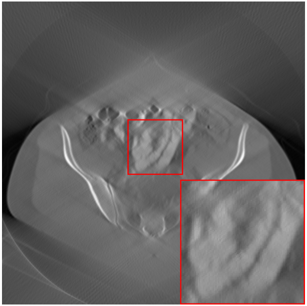
Sparse Regularization:

$$\operatorname{argmin}_f \Big[\underbrace{\|\mathcal{R}f - g\|^2}_{\text{Data fidelity term}} + \alpha \cdot \underbrace{\|(\langle f, \psi_{j,k,m} \rangle)_{j,k,m}\|_1}_{\text{Penalty term}} \Big].$$

Clinical Data:



Original Image



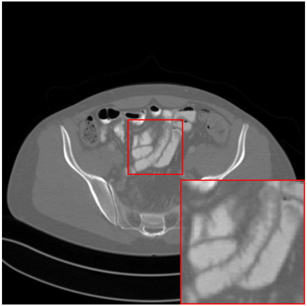
Filtered Backprojection

Model-Based Approaches Fail

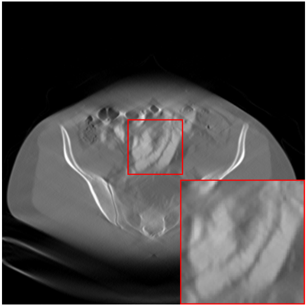
Sparse Regularization:

$$\operatorname{argmin}_f \left[\underbrace{\|\mathcal{R}f - g\|^2}_{\text{Data fidelity term}} + \alpha \cdot \underbrace{\|(\langle f, \psi_{j,k,m} \rangle)_{j,k,m}\|_1}_{\text{Penalty term}} \right].$$

Clinical Data:



Original Image



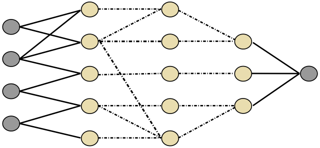
Sparse Regularization with Shearlets

Let's bring Deep Learning into the Game

Neural Networks from a Mathematical Perspective

Definition:

Assume the following notions:



- $d \in \mathbb{N}$: Dimension of input layer.
- L: Number of layers.
- N: Number of neurons.
- $\sigma : \mathbb{R} \to \mathbb{R}$: (Non-linear) function called *rectifier*.
- $W_\ell : \mathbb{R}^{N_{\ell-1}} \to \mathbb{R}^{N_\ell}$, $\ell = 1, \ldots, L$: Affine linear maps ($x \mapsto Ax + b$)

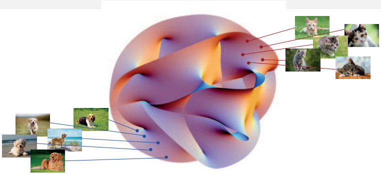
Then $\Phi : \mathbb{R}^d \to \mathbb{R}^{N_L}$ given by

$$\Phi(x) = W_L \sigma(W_{L-1} \sigma(\ldots \sigma(W_1(x))), \quad x \in \mathbb{R}^d,$$

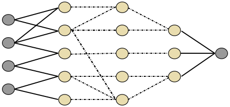
is called a *(deep) neural network (DNN)*. A DNN with only few non-zero weights is called *sparsely connected*.

Training of Deep Neural Networks

High-Level Set Up:



- Samples $(x_i, f(x_i))_{i=1}^m$ of a function such as $f : \mathcal{M} \to \{1, 2, \dots, K\}$.



- Select an architecture of a deep neural network, i.e., a choice of d, L, $(N_\ell)_{\ell=1}^L$, and σ.
  Sometimes selected entries of the matrices $(A_\ell)_{\ell=1}^L$, i.e., weights, are set to zero at this point.
- Learn the affine-linear functions $(W_\ell)_{\ell=1}^L = (A_\ell \cdot + b_\ell)_{\ell=1}^L$ by

$$\min_{A_\ell, b_\ell} \sum_{i=1}^m \mathcal{L}(\Phi_{A_\ell, b_\ell}(x_i), f(x_i)) + \lambda \mathcal{R}(A_\ell, b_\ell)$$

yielding the network $\Phi_{A_\ell, b_\ell} : \mathbb{R}^d \to \mathbb{R}^{N_L}$,

$$\Phi_{A_\ell, b_\ell}(x) = W_L \sigma(W_{L-1} \sigma(\dots \sigma(W_1(x))).$$

This is often done by stochastic gradient descent.

Goal: $\Phi_{A_\ell, b_\ell} \approx f$

Deep Neural Networks and Inverse Problems

Solving Inverse Problems by Deep Learning

Setup:

Given N training samples $(f_i, g_i)_{i=1}^N$ following the forward model

$$g_i = Kf_i + \eta.$$

Goal:

- Determine a reconstruction operator $\mathcal{T}_\theta$ such that

$$g = Kf + \eta \quad \Longrightarrow \quad \mathcal{T}_\theta(g) \approx f.$$

- $\mathcal{T}_\theta$ is parametrized by $\theta \in \mathbb{R}^p$ and learned from training data.

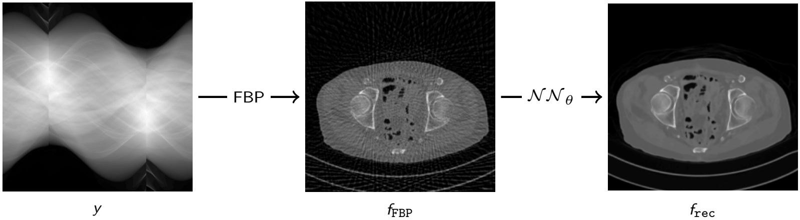
Evaluation:

Evaluate the quality of $\mathcal{T}_\theta$ by testing on the test data $(f_i, g_i)_{i=N+1}^K$ following the forward model.

Typical Deep Learning Approaches to Inverse Problems

Denoising Direct Inversion [Ye et al.,2016] [Unser et. al.,2017], ...:

- Idea: Direct inversion with filtered backprojection, train CNN to remove noise.
- Illustration:

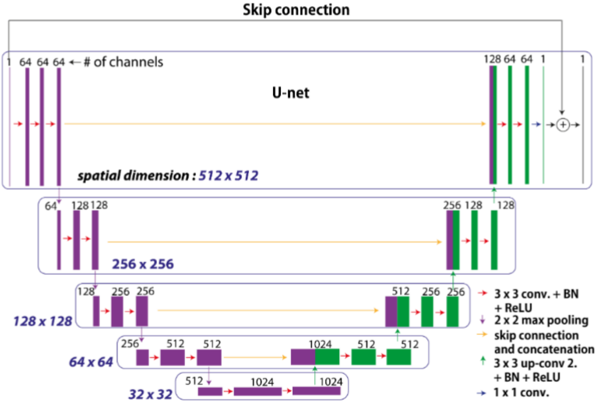

y
— FBP ⟶
f_{FBP}
$\longrightarrow \mathcal{N}\mathcal{N}_\theta \longrightarrow$
$f_{\mathtt{rec}}$

Inversion & denoising ⇝ Simple, ad-hoc approach to inverse problems

- Intuition:
  - CNN learns structured noise/artifacts.
  - Rationale: Without taking FBP, CNN needs to learn physics of CT.

Denoising Direct Inversions - The CNN Architecture

- U-Net architecture, originally used for segmentation [Ronneberger et al.,2015]
- Based on fully-convolutional networks [Long et al.,2014]
- Encoder-Decoder CNN with skip-connections



Solvers for Generalized Tikhonov Regularization

Generalized Tikhonov Regularization:

$$\operatorname{argmin}_f \left[\|Kf - g\|^2 + \alpha \cdot \mathcal{P}(f) \right]$$

Douglas-Rachford (or ADMM or ...) results in the iterations:

(1) $f_{k+1} := \operatorname{prox}_{\gamma\alpha\mathcal{P}}(h_k)$;

(2) $h_{k+1} := h_k + \operatorname{prox}_{\gamma J}(2f_{k+1} - h_k) - f_{k+1}$;

where $\gamma > 0$, $J := \|K \cdot - g\|^2$, and $\operatorname{prox}_J(h) := \operatorname{argmin}_u J(u) + \frac{1}{2}\|u - h\|_2^2$.

Observations:

- For $\mathcal{P} = \|\cdot\|_1$, (1) in soft-thresholding ⇝ denoising;
- (2) amounts in solving a linear system ⇝ Tikhonov-regularization;

Other Deep Learning Approaches to Inverse Problems

Plug-and-play with CNN-denoising [Bouman et al.,2013], [Elad et al.,2016], . . .

- Iterative solvers such as Douglas-Rachford or ADMM contain a denoising step.
- Replace this step by a trained CNN.

Learned Iterative Schemes [Pock et. al.,2017], [Adler et al.,2017], . . .

- Iterative solvers such as ADMM or Primal-Dual contain proximal steps.
- Replace these steps by parameterized operators (not necessarily prox), where the parameters are learned.

The “Best” Deep Learning Approach to Limited-Angle CT

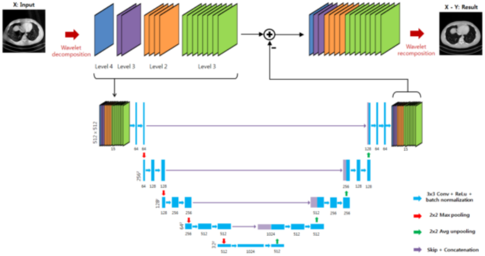


Image source: [Gu & Ye, 2017]:

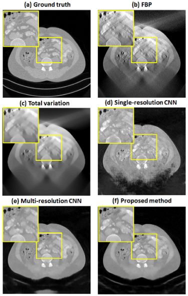


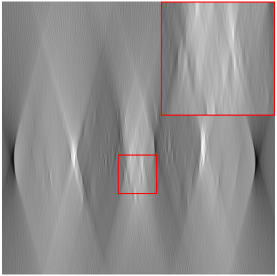
Image source: [Gu & Ye, 2017]:

- Missing theory, unclear what the neural network really does:
  - Entire image is processed!
  - Which features are modified?
  - Lack of a clear interpretation!
- The neural network needs to learn a lot of streaking artifacts (+noise)

[J. Gu and J. C. Ye. Multi-scale wavelet domain residual learning for limited-angle CT reconstruction. In: Procs Fully3D (2017), pp. 443447.]

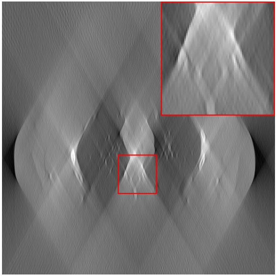
A True Hybrid Approach

Zooming in on the Recovery Problem



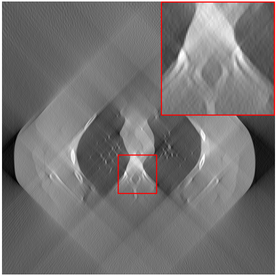
$\phi = 15^\circ$, filtered backprojection (FBP)

Zooming in on the Recovery Problem



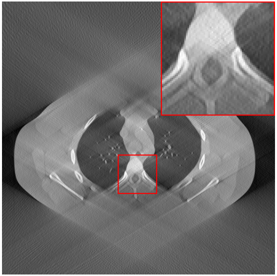
$\phi = 30^\circ$, filtered backprojection (FBP)

Zooming in on the Recovery Problem



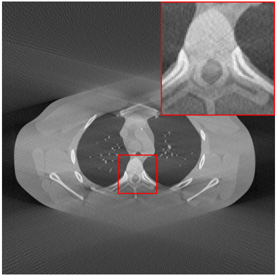
$\phi = 45^\circ$, filtered backprojection (FBP)

Zooming in on the Recovery Problem



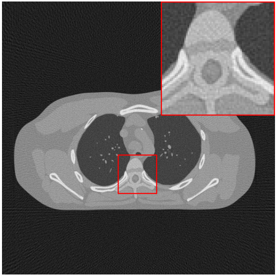
$\phi = 60^\circ$, filtered backprojection (FBP)

Zooming in on the Recovery Problem



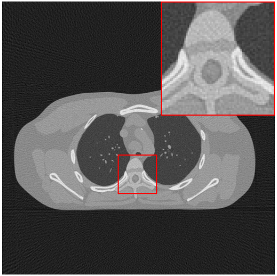
$\phi = 75^\circ$, filtered backprojection (FBP)

Zooming in on the Recovery Problem



$\phi = 90^\circ$, filtered backprojection (FBP)

Zooming in on the Recovery Problem



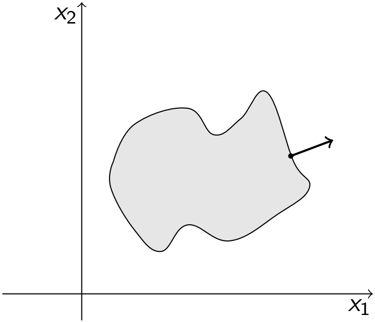
$\phi = 90^\circ$, filtered backprojection (FBP)

Some Observations:

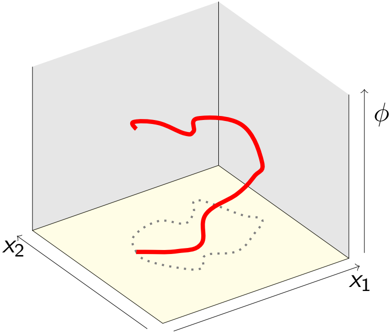
- Only certain boundaries/features seem to be *"visible"*!
- Missing wedge creates artifacts!
- Highly ill-posed inverse problem!

Fundamental Understanding of the Problem

This Phenomenon is well understood and mathematically analyzed via the concept of *microlocal analysis*, in particular, *wavefront sets*.



$f = \mathbb{I}_D$ for a set $D \subseteq \mathbb{R}^2$ with smooth boundary



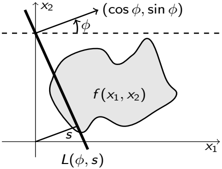
Visualization in phase space

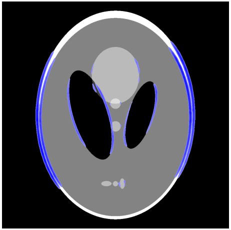
Definition: The wavefront set of a distribution f is the completement of all such location/direction pairs (t, s), where for a local window ϕ the function $\widehat{\phi f}$ decays rapidly in direction s.

Visibility in CT

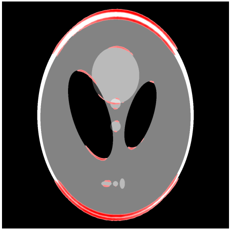
Theorem ([Quinto, 1993]): Let $L_0 = L(\phi_0, s_0)$ be a line in the plane. Let $(x_0, \xi_0) \in \mathrm{WF}(f)$ such that $x_0 \in L_0$ and ξ_0 is a normal vector to L_0.

- The singularity of f at (x_0, ξ_0) causes a unique singularity in $\mathcal{R} f$ at (ϕ_0, s_0).
- Singularities of f not tangent to $L(\phi_0, s_0)$ do not cause singularities in $\mathcal{R} f$ at (ϕ_0, s_0).





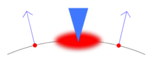
"visible": singularities tangent to sampled lines



"invisible": singularities not tangent to sampled lines

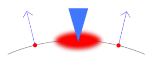
Shearlets can Help

Key Idea: Filling the missing angle is an inpainting problem of the wavefront set!

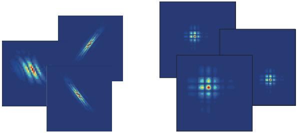


Shearlets can Help

Key Idea: Filling the missing angle is an inpainting problem of the wavefront set!



Theorem (K, Labate, 2006): “Shearlets can identify the wavefront set at fine scales.”



More Precisely:

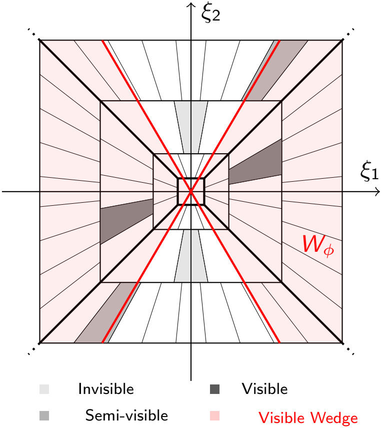
- Continuous Shearlet Transform:

$$L^2(\mathbb{R}^2) \ni f \mapsto \mathcal{SH}_\psi f(a,s,t) = \langle f, \psi_{a,s,t} \rangle, \quad (a,s,t) \in \mathbb{R}_+ \times \mathbb{R} \times \mathbb{R}^2.$$

- Resolution of Wavefront Sets (simplified from [K & Labate, 2006], [Grohs, 2011])

$$\begin{aligned} \mathrm{WF}(f)^c \quad = \quad & \Big\{ (t_0, s_0) \in \mathbb{R}^2 \times [-1,1] : \text{for } (t,s) \text{ in neighborhood } U \text{ of } (t_0,s_0): \\ & |\mathcal{SH}_\psi f(a,s,t)| = \mathcal{O}(a^k) \text{ as } a \longrightarrow 0, \forall k \in \mathbb{N}, \text{ unif. over } U \Big\} \end{aligned}$$

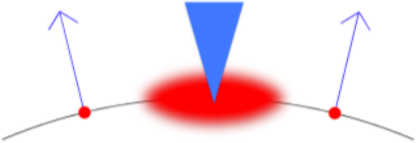
Shearlets can Separate the Visible and Invisible Part



The High-level Idea

Avenue of Research

- Shearlets are proven to resolve the wavefront set.
- Use them in sparse/limited angle tomography for filling in missing parts of the wavefront set.

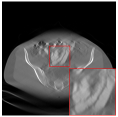


Practical Questions:

- How can we access the visible parts with shearlets?
  $\rightsquigarrow$ *Sparse Regularization!*
- How can we inpaint the missing parts?
  $\rightsquigarrow$ *Deep Learning!*

Our Approach "Learn the Invisible (LtI)"

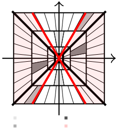
(Bubba, K, Lassas, März, Samek, Siltanen, Srinivan; 2018)



Step 1: *Reconstruct the visible*

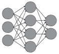
$$f^* := \operatorname{argmin}_{f \geq 0} \| \mathcal{R}_\phi f - g \|_2^2 + \| \mathrm{SH}_\psi(f) \|_{1,w}$$

- Best available classical solution (little artifacts, denoised)
- Access "wavefront set" via sparsity prior on shearlets:
  - For $(j,k,l) \in \mathcal{I}_{\mathtt{inv}}$: $\mathrm{SH}_\psi(f^*)_{(j,k,l)} \approx 0$
  - For $(j,k,l) \in \mathcal{I}_{\mathtt{vis}}$: $\mathrm{SH}_\psi(f^*)_{(j,k,l)}$ reliable and near perfect



Step 2: *Learn the invisible*

$$\mathcal{NN}_\theta : \ \mathrm{SH}_\psi(f^*)_{\mathcal{I}_{\mathtt{vis}}} \ \longrightarrow$$



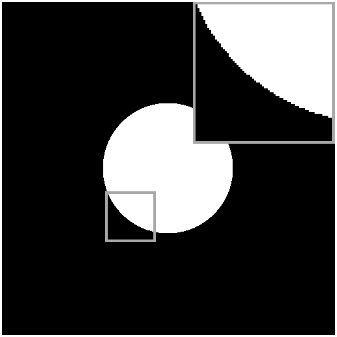
$$\longrightarrow F \ \left(\overset{!}{\approx} \mathrm{SH}_\psi(f_{\mathtt{gt}})_{\mathcal{I}_{\mathtt{inv}}} \right)$$

Step 3: *Combine*

$$f_{\mathtt{LtI}} = \mathrm{SH}_\psi^T \left(\mathrm{SH}_\psi(f^*)_{\mathcal{I}_{\mathtt{vis}}} + F \right)$$

Numerical Simulation

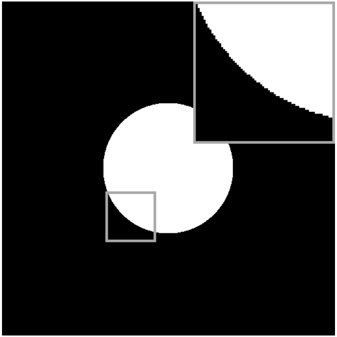
Verify the concept of (in-)visibility



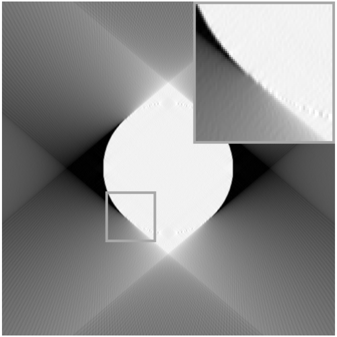
f_{gt}

Numerical Simulation

Verify the concept of (in-)visibility



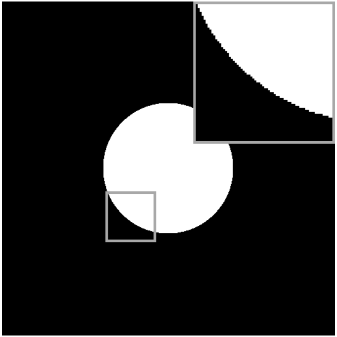
f_{gt}



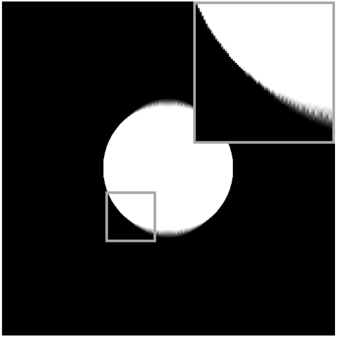
FBP

Numerical Simulation

Verify the concept of (in-)visibility



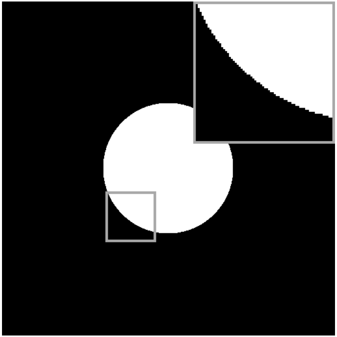
f_{gt}



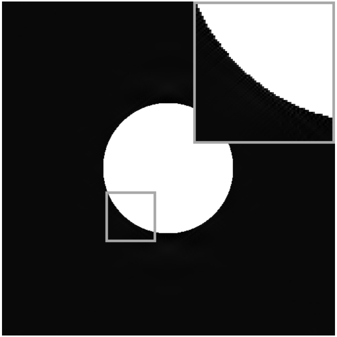
ℓ_1-analysis shearlet solution f^*

Numerical Simulation

Verify the concept of (in-)visibility with the help of an oracle:



$f_{\mathtt{gt}}$

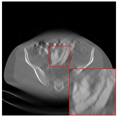


$\mathrm{SH}_\psi^T\left(\mathrm{SH}_\psi(f^*)_{\mathcal{I}_{\mathtt{vis}}} + \mathrm{SH}_\psi(f_{\mathtt{gt}})_{\mathcal{I}_{\mathtt{inv}}}\right)$

Our Approach “Learn the Invisible (LtI)”

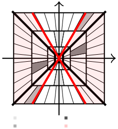
(Bubba, K, Lassas, März, Samek, Siltanen, Srinivan; 2018)

Step 1: *Reconstruct the visible*

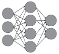


$$f^* := \operatorname{argmin}_{f \geq 0} \| \mathcal{R}_\phi f - g \|_2^2 + \| \mathrm{SH}_\psi(f) \|_{1,w}$$

- Best available classical solution (little artifacts, denoised)
- Access “wavefront set” via sparsity prior on shearlets:
  - For $(j,k,l) \in \mathcal{I}_{\mathtt{inv}}$: $\mathrm{SH}_\psi(f^*)_{(j,k,l)} \approx 0$
  - For $(j,k,l) \in \mathcal{I}_{\mathtt{vis}}$: $\mathrm{SH}_\psi(f^*)_{(j,k,l)}$ reliable and near perfect



Step 2: *Learn the invisible*

$$\mathcal{N}\mathcal{N}_\theta : \; \mathrm{SH}_\psi(f^*)_{\mathcal{I}_{\mathtt{vis}}} \quad \relbar\joinrel\relbar\joinrel\relbar$$  $$\longrightarrow F \; \left(\overset{!}{\approx} \mathrm{SH}_\psi(f_{\mathrm{gt}})_{\mathcal{I}_{\mathtt{inv}}} \right)$$

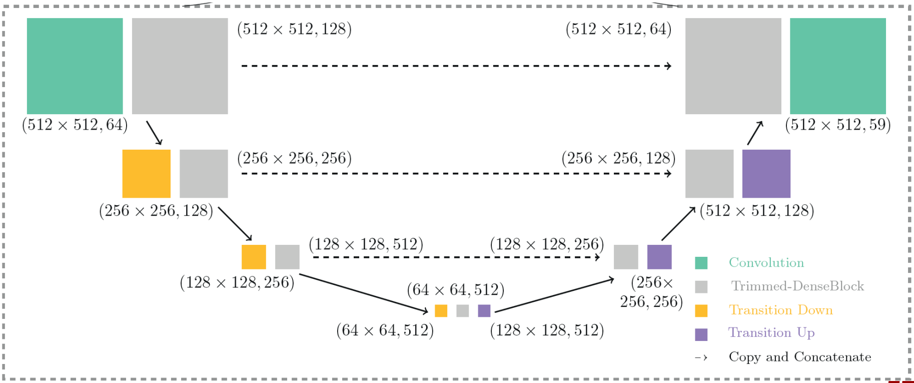
Step 3: *Combine*

$$f_{\mathtt{LtI}} = \mathrm{SH}_\psi^T \left(\mathrm{SH}_\psi(f^*)_{\mathcal{I}_{\mathtt{vis}}} + F \right)$$

Our Approach – Step 2: PhantomNet

U-Net-like CNN architecture $\mathcal{N}\mathcal{N}_\theta$ (40 layers) that is trained by minimizing:

$$\min_{\theta} \frac{1}{N} \sum_{j=1}^{N} \|\mathcal{N}\mathcal{N}_\theta(\mathsf{SH}(f_j^*)) - \mathsf{SH}(f_j^{\mathrm{gt}})_{\mathcal{I}_{\mathtt{inv}}}\|_{w,2}^2.$$



Learning the Invisible

Model Based & Data Driven: Only learn what needs to be learned!

Advantages over Pure Data Based Approach:

- Interpretation of what the CNN does ($\rightsquigarrow$ 3D inpainting)
- Reliability by learning only what is *not visible* in the data
- Better performance due to better input
- The neural network does not process entire image, leading to...
  - ...less blurring by U-net
  - ...fewer unwanted artifacts
- Better generalization

Disadvantage:

- Speed: dominated by ℓ^1-minimization

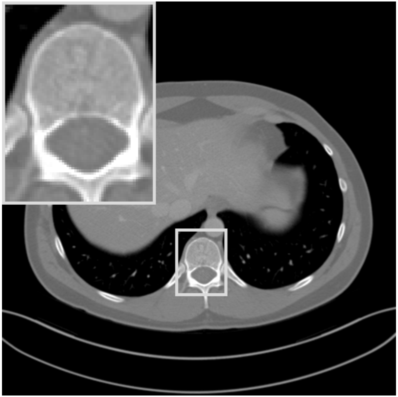
Setup

Experimental Scenarios:

- Mayo Clinic[1]: human abdomen scans provided by the Mayo Clinic for the AAPM Low-Dose CT Grand Challenge.
  - 10 patients (2378 slices of size 512×512 with thickness 3mm)
  - 9 patients for training (2134 slices) and 1 patient for testing (244 slices)
  - simulated noisy *fanbeam* measurements for 60° missing wedge
- Lotus Root: real data measured with the μCT in Helsinki
  - generalization test of our method (training is on Mayo data!)
  - 30° missing wedge
- ...

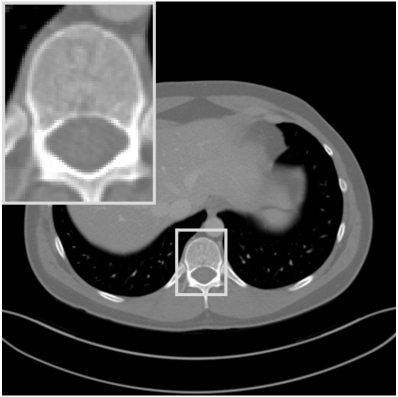
[1] We would like to thank Dr. Cynthia McCollough, the Mayo Clinic, the American Association of Physicists in Medicine (AAPM), and grant EB01705 and EB01785 from the National Institute of Biomedical Imaging and Bioengineering for providing the Low-Dose CT Grand Challenge data set.

Evaluation on Test Patient

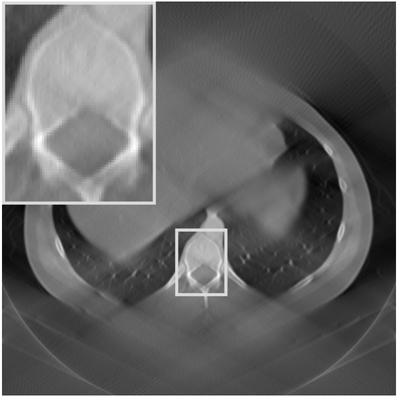


f_{gt}

Evaluation on Test Patient

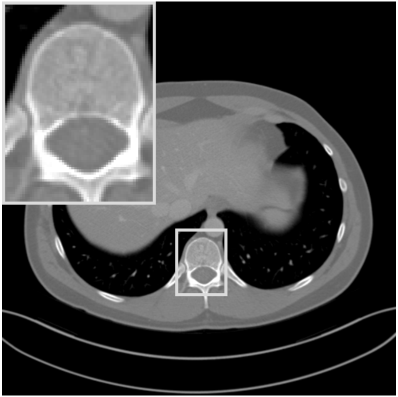


f_{gt}

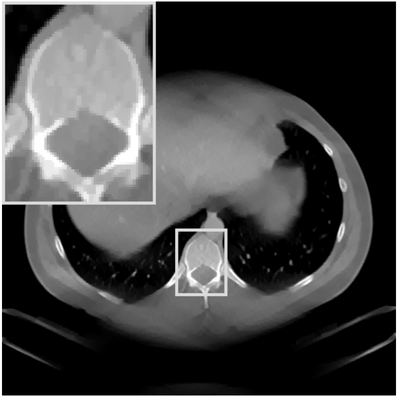


f_{FBP}: RE $=$ 0.50, HaarPSI=0.35

Evaluation on Test Patient

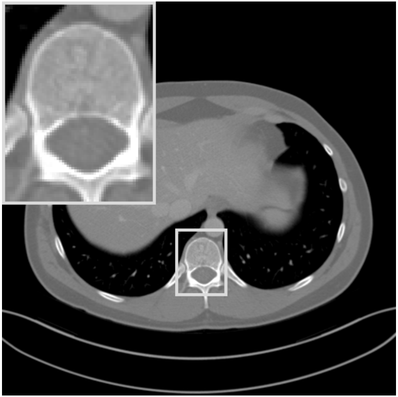


$f_{\mathtt{gt}}$

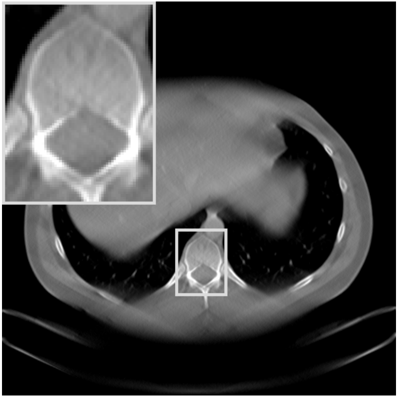


$f_{\mathtt{TV}}$: RE $= 0.21$, HaarPSI$=0.41$

Evaluation on Test Patient

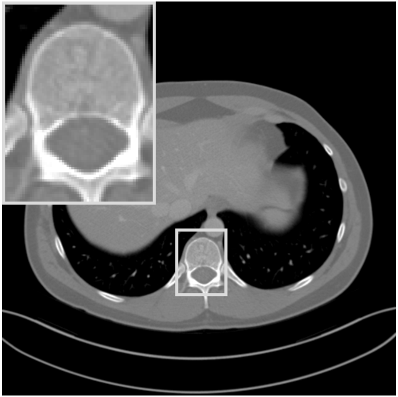


f_{gt}

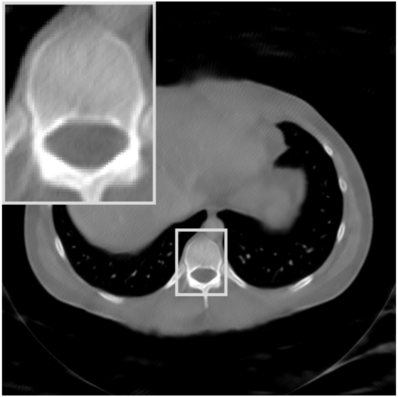


f^*: RE $= 0.19$, HaarPSI$=0.43$

Evaluation on Test Patient

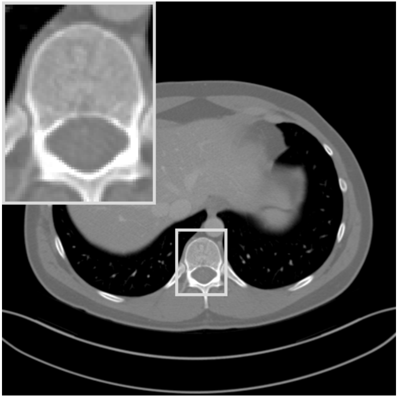


f_{gt}

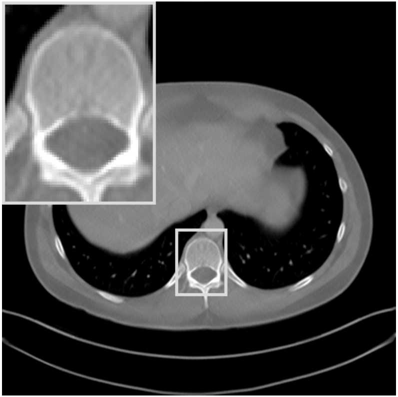


$f_{\text{[Gu \& Ye, 2017]}}$: RE $= 0.22$, HaarPSI$=0.40$

Evaluation on Test Patient



$f_{\mathtt{gt}}$



$f_{\mathtt{LtI}}$: RE $= 0.09$, HaarPSI$=0.76$

Average over Test Patient

Method	RE	PSNR	SSIM	HaarPSI
$f_{\mathtt{FBP}}$	0.47	17.16	0.40	0.32
$f_{\mathtt{TV}}$	0.18	25.88	0.85	0.37
f^*	0.17	26.34	0.85	0.40
$f_{\text{[Gu \& Ye, 2017]}}$	0.25	23.06	0.61	0.34
$\mathcal{N}\mathcal{N}_\theta(f_{\mathtt{FBP}})$	0.15	27.40	0.78	0.52
$\mathcal{N}\mathcal{N}_\theta(\mathrm{SH}(f_{\mathtt{FBP}}))$	0.16	26.80	0.74	0.52
$f_{\mathtt{LtI}}$	**0.08**	**32.77**	**0.93**	**0.73**

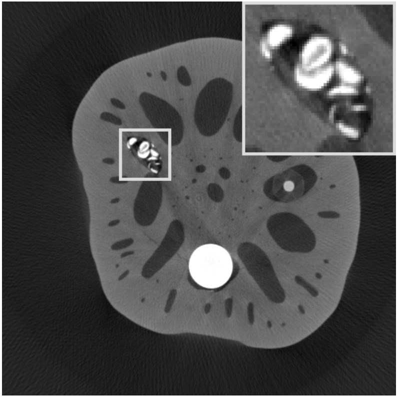
HaarPSI (Reisenhofer, Bosse, K, and Wiegand; 2018)

Advantages over (MS-)SSIM, FSIM, PSNR, GSM, VIF, etc.:

- Achieves higher correlations with human opinion scores.
- Can be computed very efficiently and significantly faster.

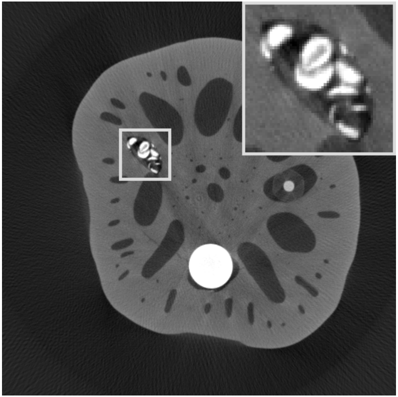
www.haarpsi.org

Generalization to Lotus Root

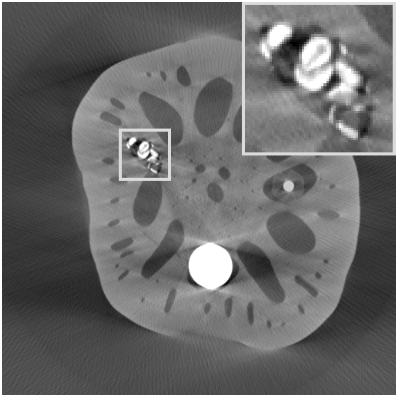


f_{gt}

Generalization to Lotus Root

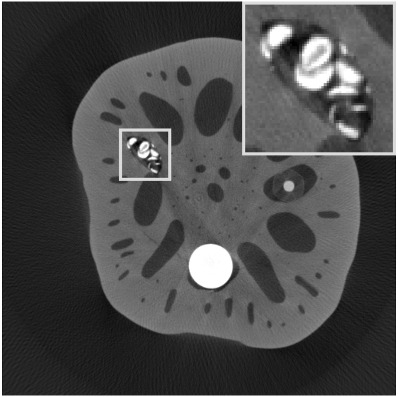


f_{gt}

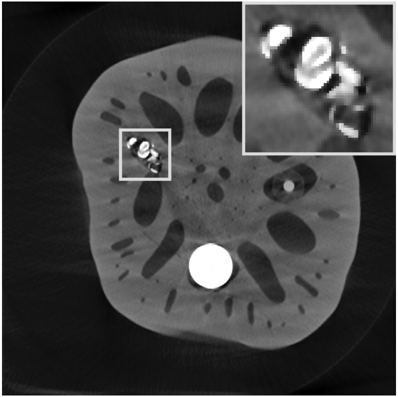


f_{FBP}: RE = 0.31, HaarPSI=0.61

Generalization to Lotus Root

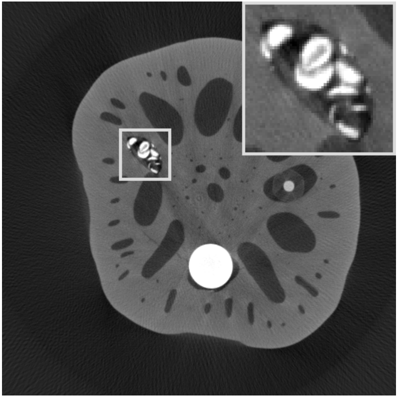


f_{gt}

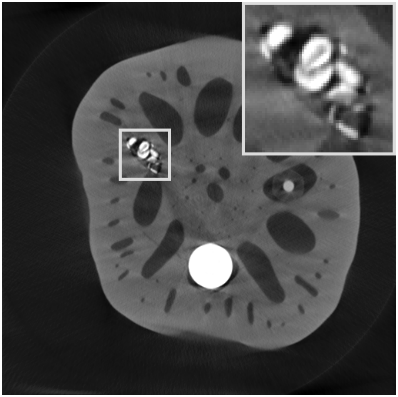


f_{TV}: RE $= 0.12$, HaarPSI$=0.74$

Generalization to Lotus Root

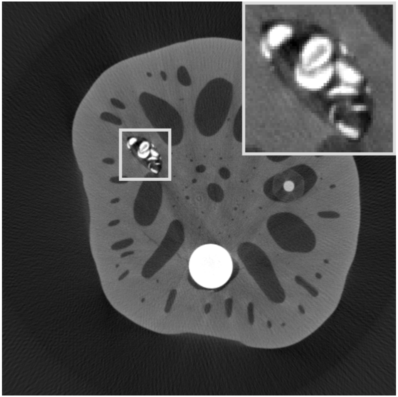


f_{gt}

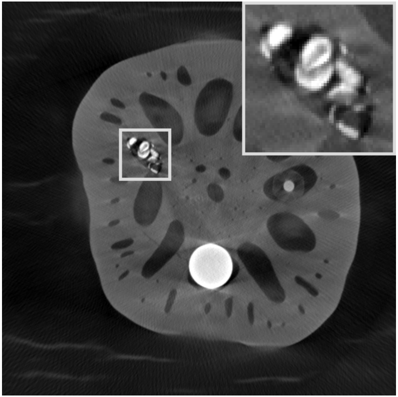


f^*: RE $= 0.11$, HaarPSI$=0.75$

Generalization to Lotus Root

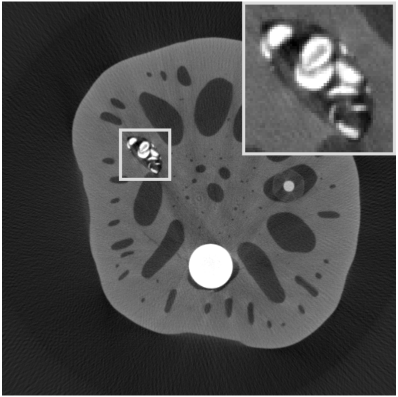


f_{gt}

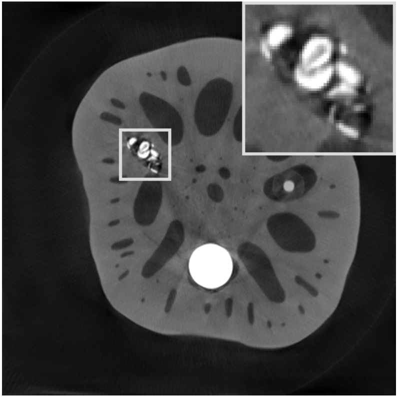


$f_{[\text{Gu \& Ye, 2017}]}$: RE = 0.25, HaarPSI=0.62

Generalization to Lotus Root



$f_{\mathtt{gt}}$



$f_{\mathtt{LtI}}$: RE $=$ 0.11, HaarPSI=0.83

Conclusions

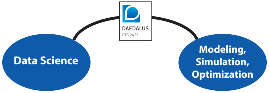
What to take Home...?

Model-Based Side:

- *Inverse problems* can be solved by *sparse regularization*.
- *Shearlets* are optimal for imaging science problems.
- Methods based on *mathematical models* today often *reach a barrier*.

Data-Based Side:

- *Deep neural networks* are nowadays often used for inverse problems.
- A *theoretical foundation* is still largely *missing*.



Combining Both Sides (Limited-Angle Tomography):

- Access and reconstruct the *visible part* using *shearlets*.
- Learn only the *invisible parts* with a *deep neural network*.

$\rightsquigarrow$ *Learning the Invisible (LtI)!*

First BMS Summer School of the new Research Center MATH$^+$:
Mathematics of Deep Learning (August 19–30, 2019)

Speakers:

- Taco Cohen (Qualcomm)
- Francois Fleuret (IDIAP/EPFL)
- Eldad Haber (University of British Columbia)
- Robert Jenssen (Tromso)
- Andreas Krause (ETH Zurich)
- Gitta Kutyniok (TU Berlin)
- Ben Leimkuhler (University of Edinburgh)
- Klaus-Robert Müller (TU Berlin)
- Frank Noé (FU Berlin)
- Christof Schütte (FU Berlin/ZIB)
- Vladimir Spokoiny (HU Berlin/WIAS)
- Rene Vidal (Johns Hopkins Univ.)

Goal of this BMS Summer School at the Zuse Institute Berlin:

This summer school will offer lectures on the theory of deep neural networks, on related questions such as generalization, expressivity, or explainability, as well as on applications of deep neural networks (e.g. to PDEs, inverse problems, or specific real-world problems).

Webpage and Application (Deadline: April 8, 2019):

`http://www.mathplus.de/summer-school-2019/index.html`

THANK YOU!

References available at:

`www.math.tu-berlin.de/∼kutyniok`

Code available at:

`www.ShearLab.org`

Related Books:

- G. Kutyniok and D. Labate
  Shearlets: Multiscale Analysis for Multivariate Data
  Birkhäuser-Springer, 2012.
- P. Grohs and G. Kutyniok
  Theory of Deep Learning
  Cambridge University Press (in preparation)

Applied and Numerical Harmonic Analysis
Gitta Kutyniok
Demetrio Labate
Editors
Shearlets
Multiscale Analysis for Multivariate Data
Birkhäuser