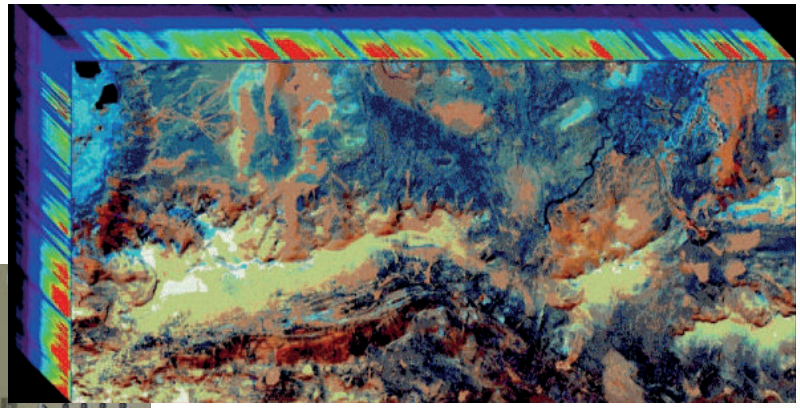
Lecture 2: Deep Learning meets Inverse Problems

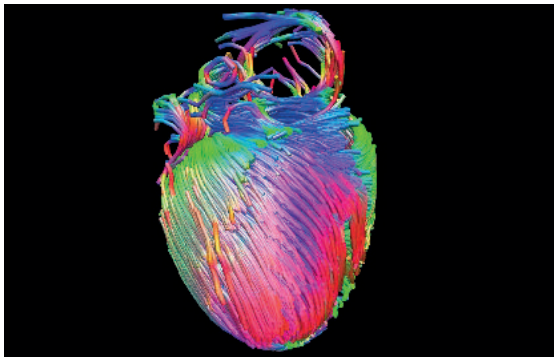
Gitta Kutyniok

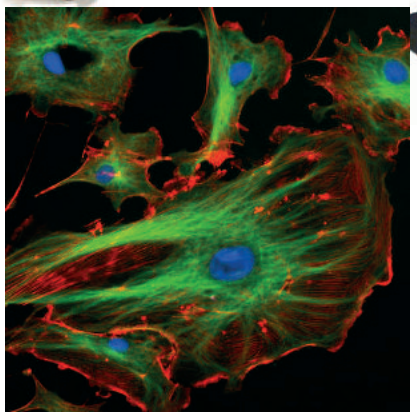
Technische Universität Berlin: Institute for Mathematics
Faculty of Computer Science & Engineering
University of Tromsø: Machine Learning Group

BMS Summer School "Mathematics of Deep Learning"
Zuse Institute Berlin, August 19–30, 2019

Modern Imaging Science







Inverse Problems

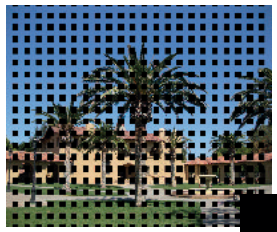
Recovering the original data from a transformed version!

Inverse Problems

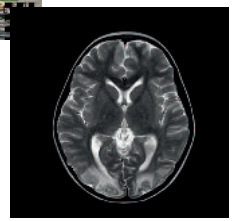
Recovering the original data from a transformed version!

Some Examples from Imaging:

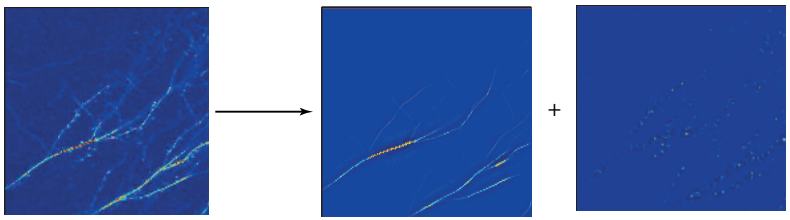
- Inpainting.
  $\rightsquigarrow$ *Recovery from incomplete data.*



- Magnetic Resonance Imaging.
  $\rightsquigarrow$ *Recovery from point samples of the Fourier transform.*



- Feature Extraction.
  $\rightsquigarrow$ *Separating the image into different features.*



Solving Inverse Problems

Regularization of Inverse Problems

General Setting:

Given $K : X \to Y$ and $y \in Y$, compute $x \in X$ with $Kx = y$.

Well-Posedness Conditions (Hadamard):

- Existence: For each $y \in Y$, there exists some $x \in X$ with $Kx = y$.
- Uniqueness: Such an $x \in X$ is unique.
- Stability: $\lim_{n\to\infty} Kx_n \to Kx$ implies $\lim_{n\to\infty} x_n \to x$.

Regularization of Inverse Problems

General Setting:

Given $K : X \to Y$ and $y \in Y$, compute $x \in X$ with $Kx = y$.

Well-Posedness Conditions (Hadamard):

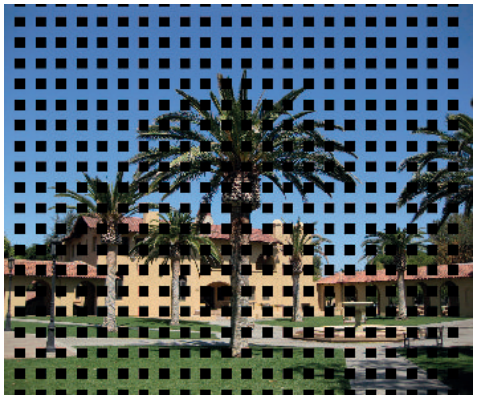
- Existence: For each $y \in Y$, there exists some $x \in X$ with $Kx = y$.
- Uniqueness: Such an $x \in X$ is unique.
- Stability: $\lim_{n\to\infty} Kx_n \to Kx$ implies $\lim_{n\to\infty} x_n \to x$.

Most problems are unfortunately ill-posed!

Examples of Ill-Posed Inverse Problems

Inpainting:

$$K : L^2([0,1]^2) \mapsto L^2([0,1]^2), \quad Kf = f \cdot \chi_\Omega, \ \Omega \subset [0,1]^2$$

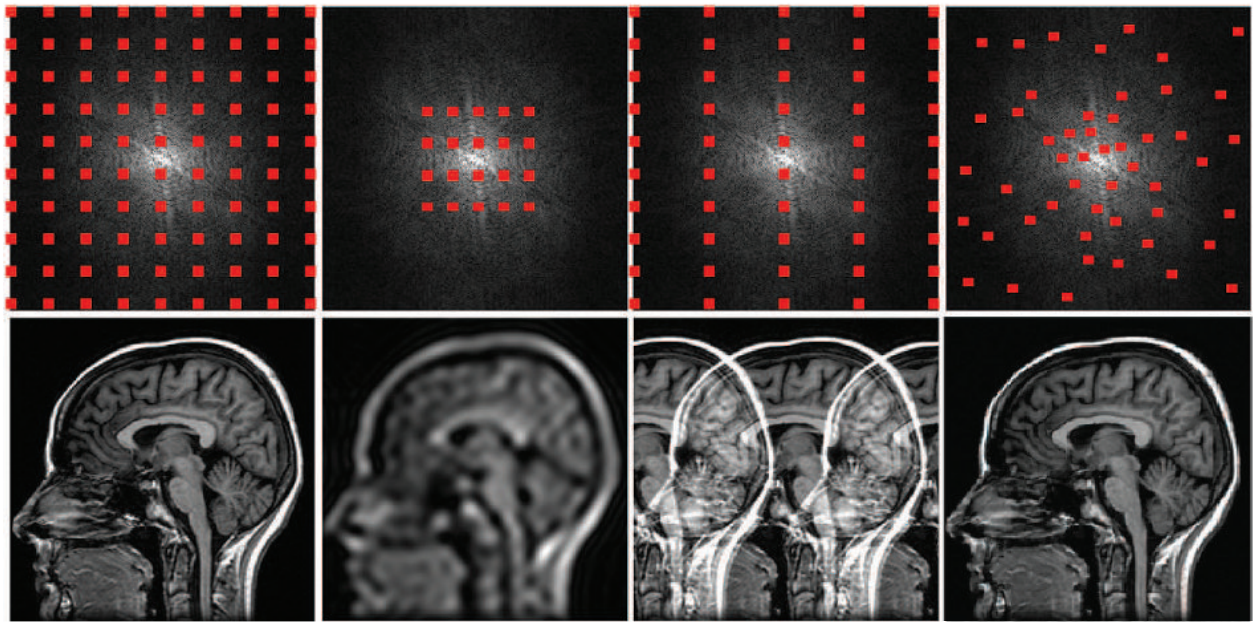


Examples of Ill-Posed Inverse Problems

Magnetic Resonance Imaging:

$$K : L^2([0,1]^2) \cap L^1([0,1]^2) \mapsto L^2([0,1]^2)$$

$$Kf = (\hat{f}(\lambda))_{\lambda \subset \Lambda},\ \Lambda \subset [0,1]^2 \text{discrete}$$

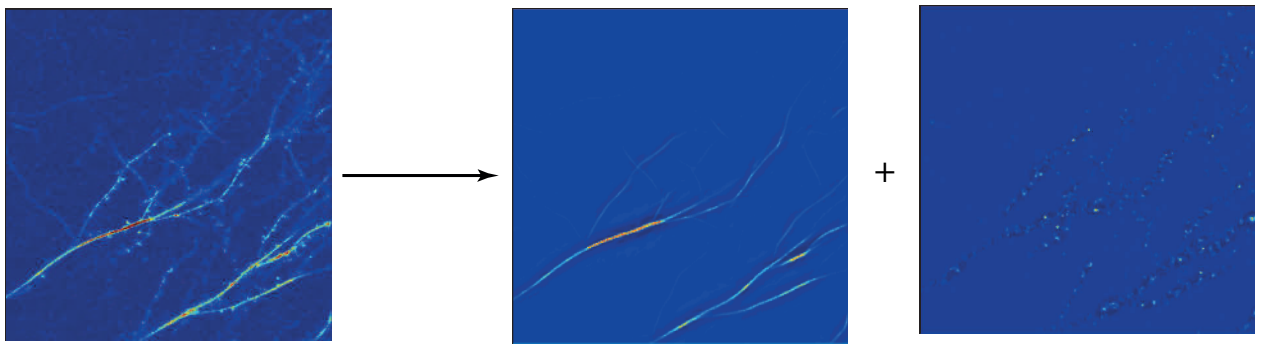


Source: Lustig, Donoho, Pauly; 2007

Examples of Ill-Posed Inverse Problems

Feature Extraction:

$$K : L^2([0,1]^2) \times L^2([0,1]^2) \mapsto L^2([0,1]^2), \quad K(f,g) = f + g$$

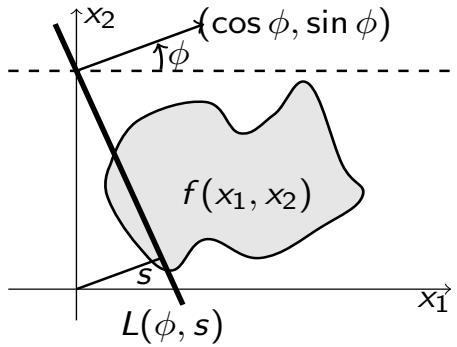


Source: K, Lim; 2011

Limited Angle-(Computed) Tomography

A CT scanner samples the *Radon transform*

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

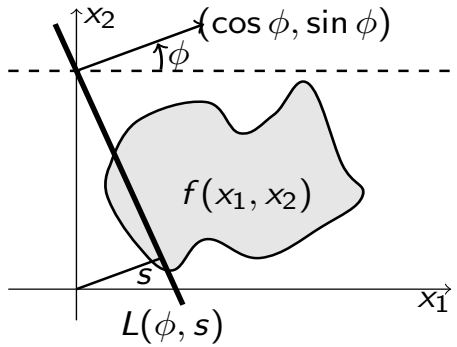


for $L(\phi, s) = \left\{x \in \mathbb{R}^2 : x_1 \cos(\phi) + x_2 \sin(\phi) = s\right\}$,
$\phi \in [-\pi/2, \pi/2)$, and $s \in \mathbb{R}$.

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,...

Ill-Posed Inverse Problems

General Setting:

Given $K : X \to Y$ and $y \in Y$, compute $x \in X$ with $Kx = y$.

Well-Posedness Conditions (Hadamard):

- Existence: For each $y \in Y$, there exists some $x \in X$ with $Kx = y$.
- Uniqueness: Such an $x \in X$ is unique.
- Stability: $\lim_{n\to\infty} Kx_n \to Kx$ implies $\lim_{n\to\infty} x_n \to x$.

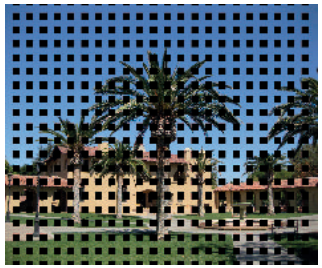
Ill-Posed Inverse Problems:

Need for regularization!

Regularization

Goal:

- Bring in prior information on the seeked solution.
- Ensure continuous dependence on the data.



Problem:

- How do we obtain prior information?
- How do we incorporate it in the solver?

Approaches:

- Take from shelf
- Handcraft
- Learn

Tikhonov Regularization

Standard Tikhonov Regularization:

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

$$J_\alpha(x) := \underbrace{\|Kx - y\|^2}_{\text{Data fidelity term}} + \alpha \cdot \underbrace{\|x\|^2}_{\text{Regularization Term}}, \quad x \in X.$$

Tikhonov Regularization

Standard Tikhonov Regularization:

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

$$J_\alpha(x) := \underbrace{\|Kx - y\|^2}_{\text{Data fidelity term}} + \alpha \cdot \underbrace{\|x\|^2}_{\text{Regularization Term}}, \quad x \in X.$$

Generalization:

$$J_\alpha(x) := \underbrace{\|Kx - y\|^2}_{\text{Data fidelity term}} + \alpha \cdot \underbrace{\mathcal{R}(x)}_{\text{Regularization Term}}, \quad x \in X.$$

The Regularization Term $\mathcal{R}$

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

Regularization Term

Some Examples for $\mathcal{R}$:

- ℓ_2 norm of $x \in \mathbb{R}^d$:

$$\|x\|_{\ell_2}^2 = \sum_{i=1}^{d} |x_i|^2$$

Regularization Term

Some Examples for $\mathcal{R}$:

- ℓ_2 norm of $x \in \mathbb{R}^d$:

$$\|x\|_{\ell_2}^2 = \sum_{i=1}^{d} |x_i|^2$$

- TV-norm of $f : [0,1] \to \mathbb{R}$:

$$\|f\|_{TV} = \sup_k \sum_{i=0}^{n_k - 1} |f(x_{i+1}) - f(x_i)|$$

Regularization Term

Some Examples for $\mathcal{R}$:

- ℓ_2 norm of $x \in \mathbb{R}^d$:

$$\|x\|_{\ell_2}^2 = \sum_{i=1}^{d} |x_i|^2$$

- TV-norm of $f : [0,1] \to \mathbb{R}$:

$$\|f\|_{TV} = \sup_k \sum_{i=0}^{n_k - 1} |f(x_{i+1}) - f(x_i)|$$

- Sobolev norm of $f : \Omega \to \mathbb{R}$:

$$\|f\|_{W^{k,p}(\Omega)} = \left(\sum_{|\alpha| \leq k} \|\partial^\alpha f\|_{L^p(\Omega)}^p \right)^{\frac{1}{p}}$$

Regularization Term

Some Examples for $\mathcal{R}$:

- ℓ_2 norm of $x \in \mathbb{R}^d$:

$$\|x\|_{\ell_2}^2 = \sum_{i=1}^{d} |x_i|^2$$

- TV-norm of $f : [0,1] \to \mathbb{R}$:

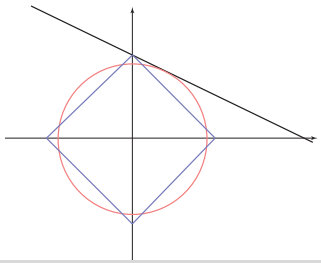
$$\|f\|_{TV} = \sup_k \sum_{i=0}^{n_k - 1} |f(x_{i+1}) - f(x_i)|$$

- Sobolev norm of $f : \Omega \to \mathbb{R}$:

$$\|f\|_{W^{k,p}(\Omega)} = \left(\sum_{|\alpha| \leq k} \|\partial^\alpha f\|_{L^p(\Omega)}^p \right)^{\frac{1}{p}}$$

- Sparsity of $x \in \mathbb{R}^d$:

$$\|x\|_{\ell_1} = \sum_{i=1}^{d} |x_i|$$



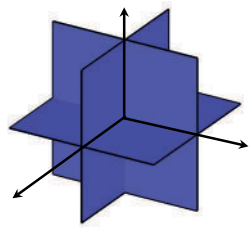
Sparse Regularization

Paradigm for Data Processing: Sparsity!

Sparse Signals:

A signal $x \in \mathbb{R}^n$ is k-sparse, if

$$\|x\|_0 = \#\text{non-zero coefficients} \leq k.$$

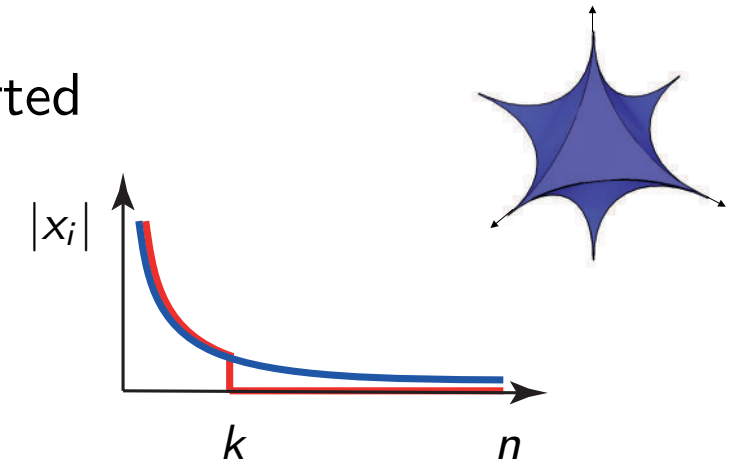


$\rightsquigarrow$ Model Σ_k: Union of k-dimensional subspaces

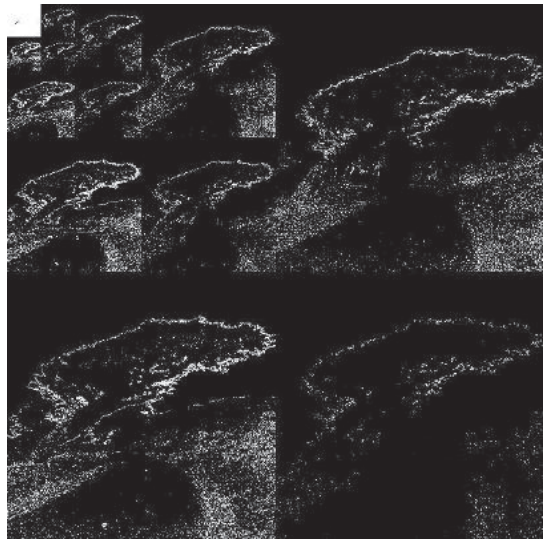
Compressible Signals:

A signal $x \in \mathbb{R}^n$ is compressible, if the sorted coefficients have rapid (power law) decay.

$\rightsquigarrow$ Model: ℓ_p ball with $p \leq 1$

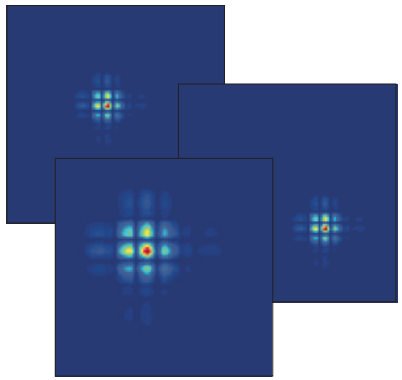


The World is Compressible!



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) :

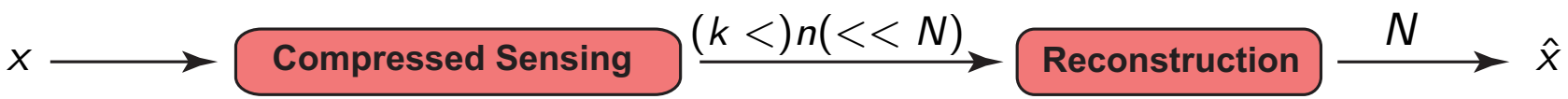
- Goal: Solve an underdetermined linear problem

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

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

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



Sparsifying System

Functional Analytic Properties:

- $(\psi_\lambda)_\lambda$ can be an orthonormal basis.
- $(\psi_\lambda)_\lambda$ can form a frame, i.e., there exist $0 < A \leq B < \infty$ with

$$A\|x\|^2 \leq \sum_\lambda |\langle x, \psi_\lambda \rangle|^2 \leq B\|x\|^2 \quad \text{for all } x \in \mathcal{H}.$$

Sparsifying System

Functional Analytic Properties:

- $(\psi_\lambda)_\lambda$ can be an orthonormal basis.
- $(\psi_\lambda)_\lambda$ can form a frame, i.e., there exist $0 < A \leq B < \infty$ with

$$A\|x\|^2 \leq \sum_\lambda |\langle x, \psi_\lambda \rangle|^2 \leq B\|x\|^2 \quad \text{for all } x \in \mathcal{H}.$$

Basic Facts about Frames:

- The frame operator $S : \mathcal{H} \to \mathcal{H}$, $Sx = \sum_\lambda \langle x, \psi_\lambda \rangle \psi_\lambda$ is invertible.
- The dual frame $(\tilde{\psi}_\lambda)_\lambda := (S^{-1}\psi_\lambda)_\lambda$ yields

$$x = \sum_\lambda \langle x, \psi_\lambda \rangle \tilde{\psi}_\lambda = \sum_\lambda \langle x, \tilde{\psi}_\lambda \rangle \psi_\lambda \quad \text{for all } x \in \mathcal{H}.$$

Sparsifying System

Functional Analytic Properties:

- $(\psi_\lambda)_\lambda$ can be an orthonormal basis.
- $(\psi_\lambda)_\lambda$ can form a frame, i.e., there exist $0 < A \leq B < \infty$ with

$$A\|x\|^2 \leq \sum_\lambda |\langle x, \psi_\lambda \rangle|^2 \leq B\|x\|^2 \quad \text{for all } x \in \mathcal{H}.$$

Basic Facts about Frames:

- The frame operator $S : \mathcal{H} \to \mathcal{H}$, $Sx = \sum_\lambda \langle x, \psi_\lambda \rangle \psi_\lambda$ is invertible.
- The dual frame $(\tilde{\psi}_\lambda)_\lambda := (S^{-1}\psi_\lambda)_\lambda$ yields

$$x = \sum_\lambda \langle x, \psi_\lambda \rangle \tilde{\psi}_\lambda = \sum_\lambda \langle x, \tilde{\psi}_\lambda \rangle \psi_\lambda \quad \text{for all } x \in \mathcal{H}.$$

Some Advantages of Redundancy:

- Flexibility in expansions $x = \sum_\lambda c_\lambda \psi_\lambda$.
- Robustness against loss of coefficients $\langle x, \psi_\lambda \rangle$.

Notion of Optimality

Two Viewpoints of Optimality of $(\psi_\lambda)_\lambda$: Let $\mathcal{C} \subseteq \mathcal{H}$.

- Decay of Coefficients. $\beta > 0$ is largest (for all systems) with

$$|\langle x, \psi_{\lambda_n} \rangle| \lesssim n^{-\beta} \text{ as } n \to \infty, \quad \text{for all } x \in \mathcal{C}.$$

- Approximation Properties. $\gamma > 0$ is largest (for all systems) with

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

Notion of Optimality

Two Viewpoints of Optimality of $(\psi_\lambda)_\lambda$: Let $\mathcal{C} \subseteq \mathcal{H}$.

- Decay of Coefficients. $\beta > 0$ is largest (for all systems) with

$$|\langle x, \psi_{\lambda_n} \rangle| \lesssim n^{-\beta} \text{ as } n \to \infty, \quad \text{for all } x \in \mathcal{C}.$$

- Approximation Properties. $\gamma > 0$ is largest (for all systems) with

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

Situation of an ONB: For the best N-term approximation x_N of x, we have

$$\|x - x_N\|^2 = \sum_{\lambda \notin \Lambda_N} |c_\lambda|^2 = \sum_{n > N} |\langle x, \psi_{\lambda_n} \rangle|^2$$

These viewpoints coincide!

Notion of Optimality

Two Viewpoints of Optimality of $(\psi_\lambda)_\lambda$: Let $\mathcal{C} \subseteq \mathcal{H}$.

- Decay of Coefficients. $\beta > 0$ is largest (for all systems) with

$$|\langle x, \psi_{\lambda_n} \rangle| \lesssim n^{-\beta} \text{ as } n \to \infty, \quad \text{for all } x \in \mathcal{C}.$$

- Approximation Properties. $\gamma > 0$ is largest (for all systems) with

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

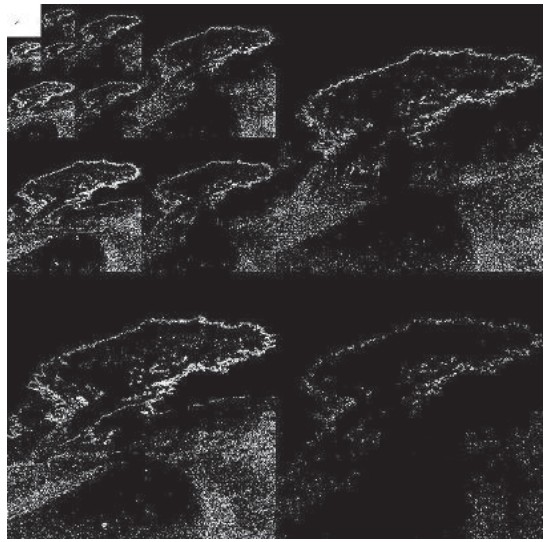
Situation of an ONB: For the best N-term approximation x_N of x, we have

$$\|x - x_N\|^2 = \sum_{\lambda \notin \Lambda_N} |c_\lambda|^2 = \sum_{n > N} |\langle x, \psi_{\lambda_n} \rangle|^2$$

These viewpoints coincide!

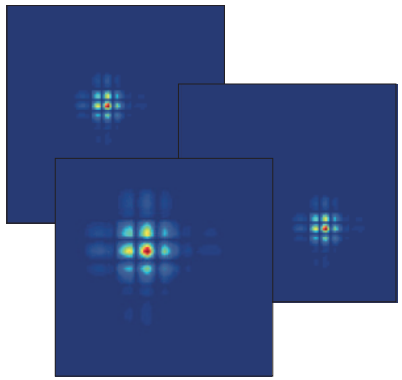
Situation of a Frame: These viewpoints differ significantly!

The World is Compressible!



Wavelet Transform (JPEG2000):

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

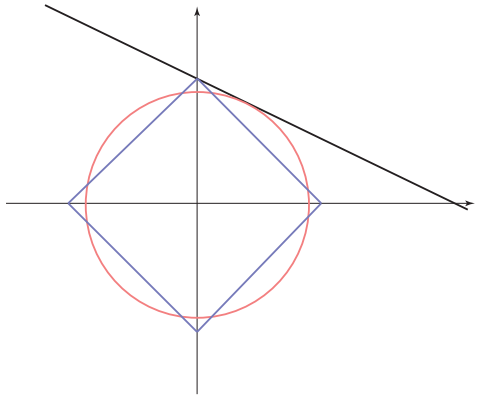


Definition: For a wavelet $\psi \in L^2(\mathbb{R}^2)$, a wavelet system is defined by

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

How to Penalize Non-Sparsity?

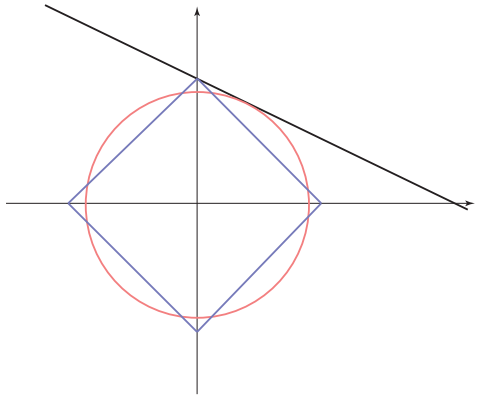
Intuition:



$\rightsquigarrow$ *Use the ℓ_1 norm!*

How to Penalize Non-Sparsity?

Intuition:



$\rightsquigarrow$ *Use the ℓ_1 norm!*

Sparse Regularization:

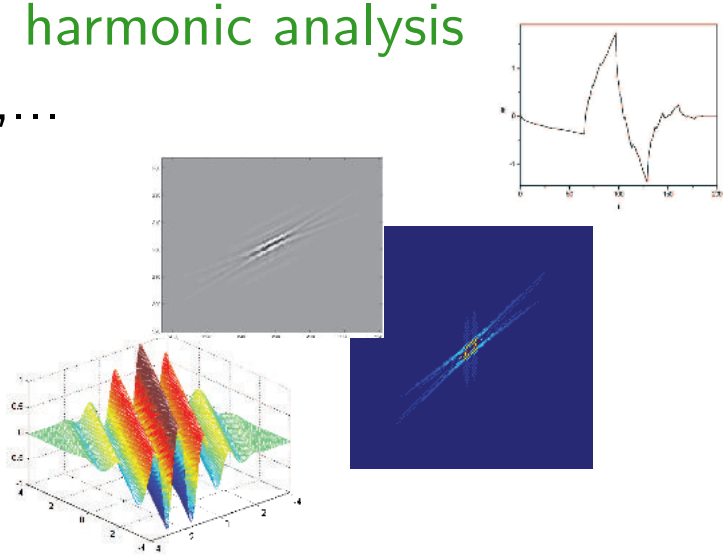
Solve an ill-posed inverse problem $Kf = g$ by

$$f^\alpha := \operatorname{argmin}_f \Big[\underbrace{\|Kf - g\|^2}_{\text{Data fidelity term}} + \alpha \cdot \underbrace{\|(\langle f, \psi_{j,m} \rangle)_{j,m}\|_1}_{\text{Regularization Term}} \Big].$$

Sparsifying ONB/Frames

How to find such a system:

- Use an existing one such as from applied harmonic analysis
  Wavelets, Ridgelets, Curvelets, Shearlets,...

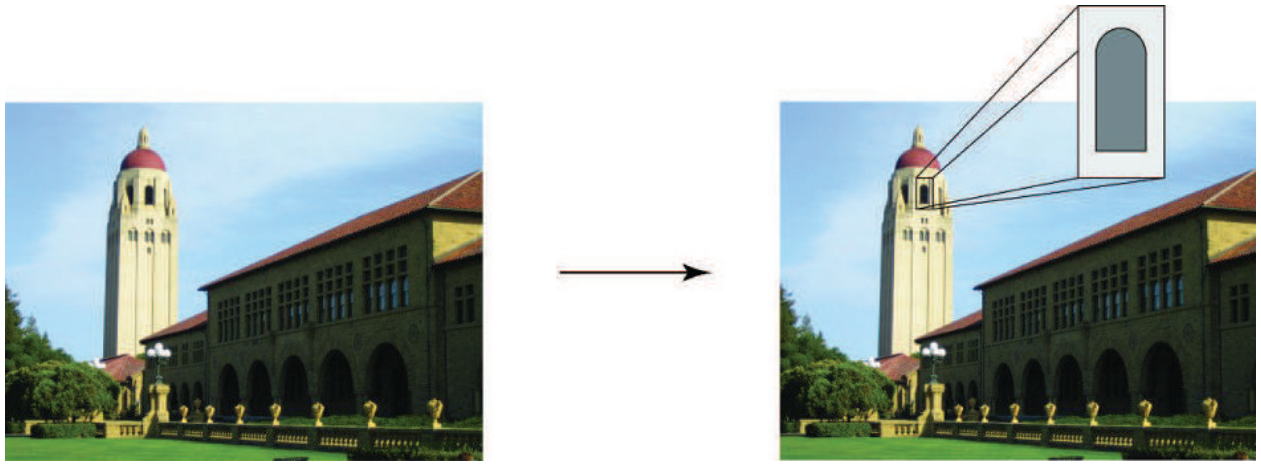


- Construct a novel system based on information about the data.
- Learn a system by dictionary learning algorithms.
  - K-SVD (Aharon, Elad, and Bruckstein; 2006)
  - ...

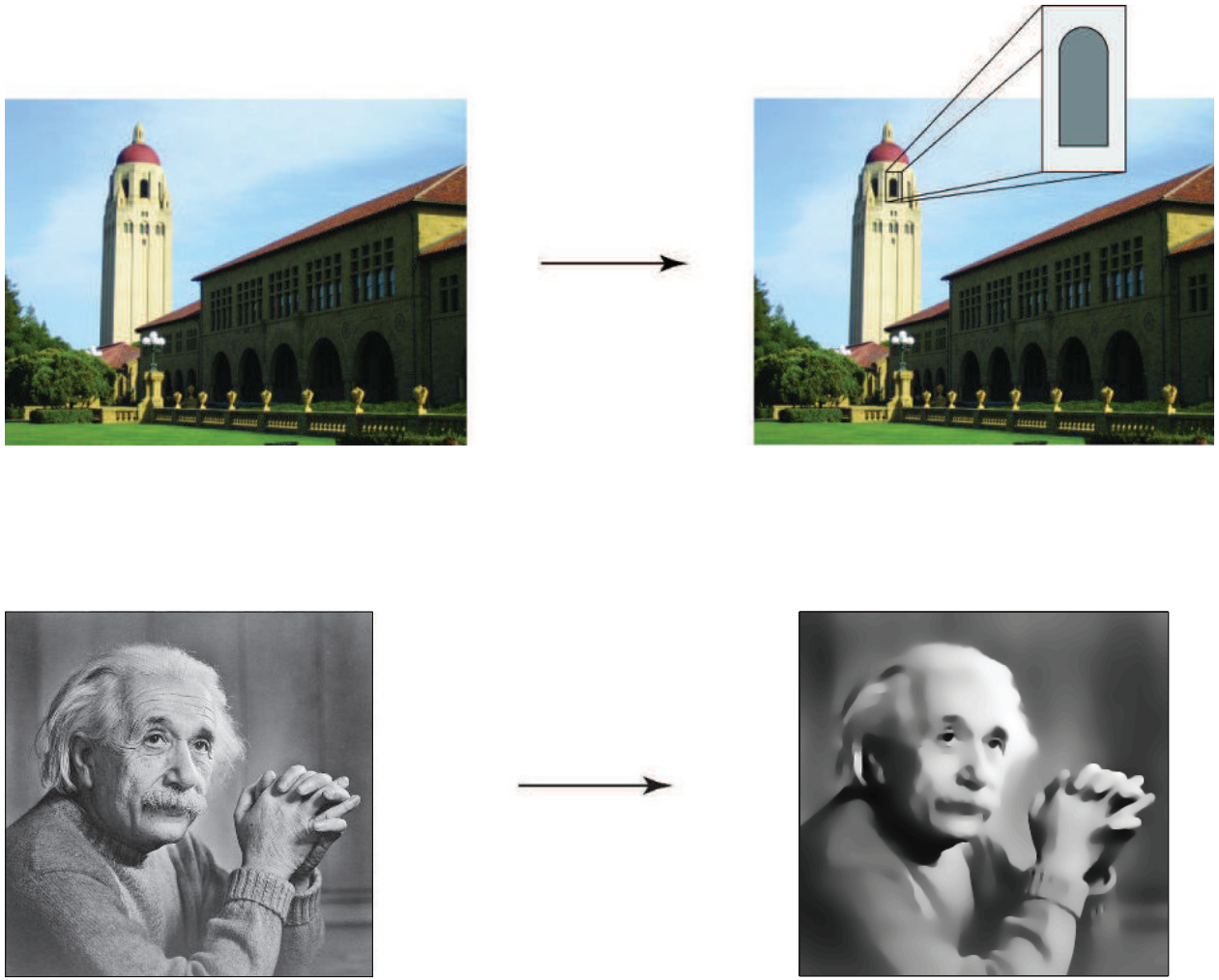
Regularization Term in Imaging

What is an Image?

What is an Image?



What is an Image?



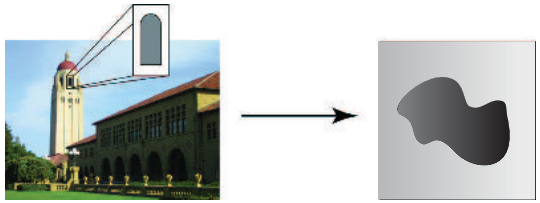
Fitting Model

Definition (Donoho; 2001):

The set of cartoon-like functions $\mathcal{E}^2(\mathbb{R}^2)$ is defined by

$$\mathcal{E}^2(\mathbb{R}^2) = \{f \in L^2(\mathbb{R}^2) : f = f_0 + f_1 \cdot \chi_B\},$$

where $\emptyset \neq B \subset [0,1]^2$ simply connected with C^2-boundary and bounded curvature, and $f_i \in C^2(\mathbb{R}^2)$ with supp $f_i \subseteq [0,1]^2$ and $\|f_i\|_{C^2} \leq 1$, $i = 0, 1$.



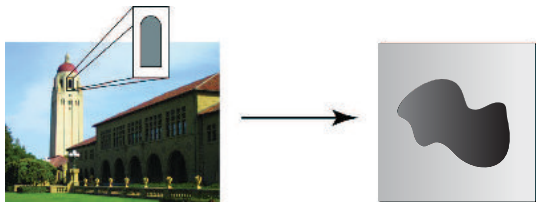
Fitting Model

Definition (Donoho; 2001):

The set of cartoon-like functions $\mathcal{E}^2(\mathbb{R}^2)$ is defined by

$$\mathcal{E}^2(\mathbb{R}^2) = \{f \in L^2(\mathbb{R}^2) : f = f_0 + f_1 \cdot \chi_B\},$$

where $\emptyset \neq B \subset [0,1]^2$ simply connected with C^2-boundary and bounded curvature, and $f_i \in C^2(\mathbb{R}^2)$ with supp $f_i \subseteq [0,1]^2$ and $\|f_i\|_{C^2} \leq 1$, $i = 0, 1$.



Theorem (Donoho; 2001):

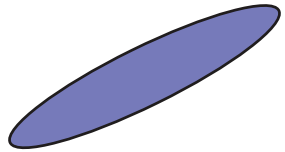
Let $(\psi_\lambda)_\lambda \subseteq L^2(\mathbb{R}^2)$. Allowing only polynomial depth search, we have the following optimal behavior for $f \in \mathcal{E}^2(\mathbb{R}^2)$:

$$\|f - f_N\|_2 \asymp N^{-1} \quad \text{and} \quad |\langle f, \psi_{\lambda_n}\rangle| \lesssim n^{-\frac{3}{2}} \quad \text{as } N, n \to \infty.$$

Scaling and Orientation

Parabolic scaling ('width $\approx$ length^2'):

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



Orientation via shearing:

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

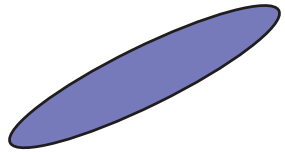
Definition (K, Labate; 2006):
For $\psi \in L^2(\mathbb{R}^2)$, the associated shearlet system is defined by

$$\{2^{\frac{3j}{4}} \psi(S_k A_{2^j} \cdot - m) \, : \, j, k \in \mathbb{Z}, m \in \mathbb{Z}^2\}.$$

Scaling and Orientation

Parabolic scaling ('width $\approx$ length^2'):

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



Orientation via shearing:

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

Definition (K, Labate; 2006):
For $\psi \in L^2(\mathbb{R}^2)$, the associated shearlet system is defined by

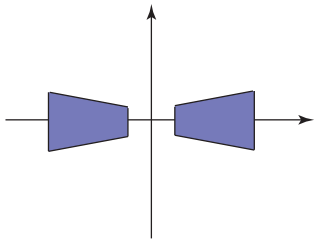
$$\{2^{\frac{3j}{4}} \psi(S_k A_{2^j} \cdot - m) \, : \, j, k \in \mathbb{Z}, m \in \mathbb{Z}^2\}.$$

Disadvantage: Non-uniform treatment of directions!

Example of Classical (Band-Limited) Shearlet

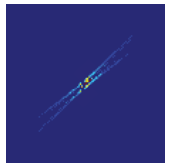
Let $\psi \in L^2(\mathbb{R}^2)$ be defined by

$$\hat{\psi}(\xi) = \hat{\psi}(\xi_1, \xi_2) = \hat{\psi}_1(\xi_1)\, \hat{\psi}_2(\tfrac{\xi_2}{\xi_1}),$$



where

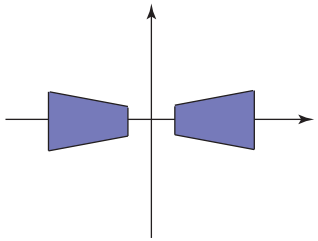
- ψ_1 wavelet, $\mathrm{supp}(\hat{\psi}_1) \subseteq [-2, -\frac{1}{2}] \cup [\frac{1}{2}, 2]$ and $\hat{\psi}_1 \in C^\infty(\mathbb{R})$.
- $\mathrm{supp}(\hat{\psi}_2) \subseteq [-1, 1]$ and $\hat{\psi}_2 \in C^\infty(\mathbb{R})$.



Example of Classical (Band-Limited) Shearlet

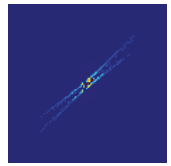
Let $\psi \in L^2(\mathbb{R}^2)$ be defined by

$$\hat{\psi}(\xi) = \hat{\psi}(\xi_1, \xi_2) = \hat{\psi}_1(\xi_1)\, \hat{\psi}_2(\tfrac{\xi_2}{\xi_1}),$$

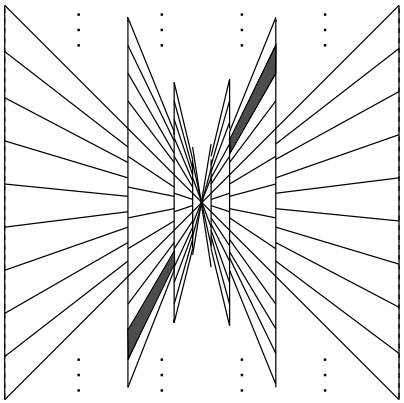


where

- ψ_1 wavelet, $\mathrm{supp}(\hat{\psi}_1) \subseteq [-2, -\frac{1}{2}] \cup [\frac{1}{2}, 2]$ and $\hat{\psi}_1 \in C^\infty(\mathbb{R})$.
- $\mathrm{supp}(\hat{\psi}_2) \subseteq [-1, 1]$ and $\hat{\psi}_2 \in C^\infty(\mathbb{R})$.



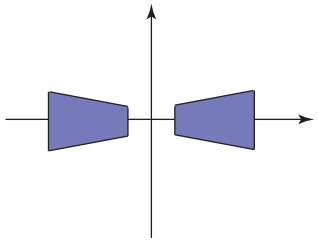
Induced tiling of Fourier domain:



Example of Classical (Band-Limited) Shearlet

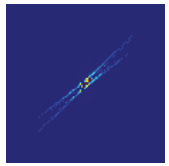
Let $\psi \in L^2(\mathbb{R}^2)$ be defined by

$$\hat{\psi}(\xi) = \hat{\psi}(\xi_1, \xi_2) = \hat{\psi}_1(\xi_1)\, \hat{\psi}_2(\tfrac{\xi_2}{\xi_1}),$$

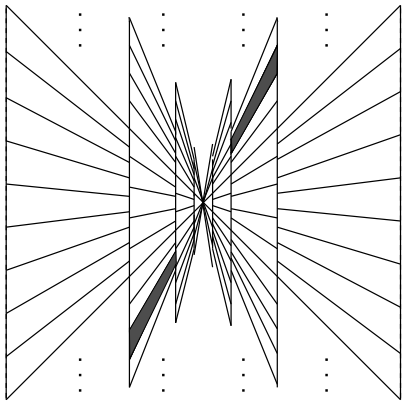


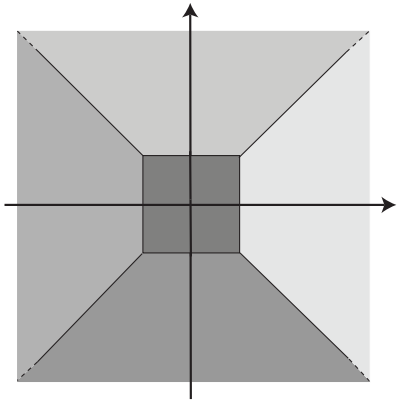
where

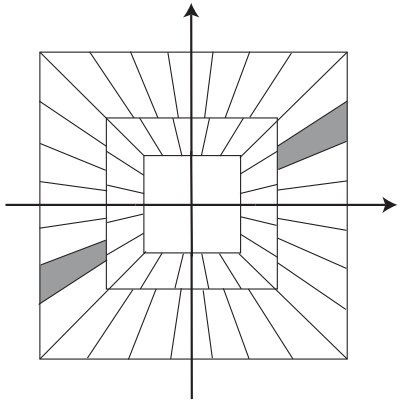
- ψ_1 wavelet, $\mathrm{supp}(\hat{\psi}_1) \subseteq [-2, -\frac{1}{2}] \cup [\frac{1}{2}, 2]$ and $\hat{\psi}_1 \in C^\infty(\mathbb{R})$.
- $\mathrm{supp}(\hat{\psi}_2) \subseteq [-1, 1]$ and $\hat{\psi}_2 \in C^\infty(\mathbb{R})$.



Induced tiling of Fourier domain:







(Cone-adapted) Shearlet Systems

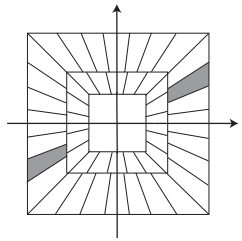
Definition (K, Labate; 2006):

The (cone-adapted) 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) Shearlet Systems

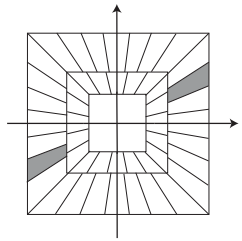
Definition (K, Labate; 2006):

The (cone-adapted) 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

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

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

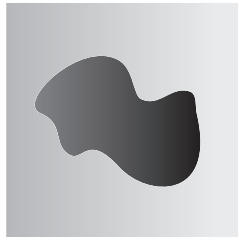
$$\{2^{3j/4}\tilde{\psi}(\tilde{S}_k \tilde{A}_{2^j}\ \cdot\ - cm) : j \geq 0, |k| \leq \lceil 2^{j/2} \rceil, m \in \mathbb{Z}^2\}.$$



Theorem (K, Lim; 2011):

Let $\phi,\psi,\tilde{\psi} \in L^2(\mathbb{R}^2)$ be compactly supported, and let $\hat{\phi}$, $\hat{\psi}$, $\hat{\tilde{\psi}}$ satisfy certain decay conditions. Then $\mathcal{SH}(c;\phi,\psi,\tilde{\psi}) = (\sigma_\eta)_\eta$ provides an optimally sparsifying system for $f \in \mathcal{E}^2(\mathbb{R}^2)$, i.e., for $N,n \to \infty$,

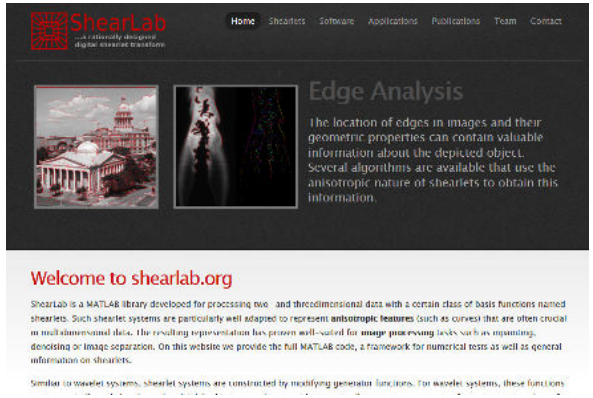
$$\|f - f_N\|_2 \lesssim N^{-1}(\log N)^3 \text{ and } |\langle f, \sigma_{\eta_n}\rangle| \lesssim n^{-\frac{3}{2}}(\log n)^{\frac{3}{2}}.$$



Implementation

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

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



Deep Neural Networks and Inverse Problems

~~Deep~~ Learning Enters the Game

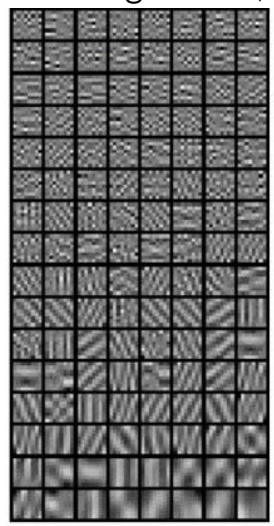
Idea: Instead of "handcrafting" a sparsity prior, learn it from data.

Some Examples: Dictionary learning (Elad et al.; 2006), parameter selection, e.g. via bilevel optimization (Kunisch et al.; 2013), analysis operator learning, e.g. (Pock et al.; 2014), blind compressed sensing, e.g. (Bressler et al.; 2015), ...

Discussion:

- Higher (initial) computational burden in learning stage
- Depending on application: marginal to substantial gain
- Do we simply re-learn wavelets, shearlets etc.?!
- No automatic correction for errors in modelling the operator
- Drawback: Still need to solve a variational formulation

Source: Eksioglu et al.; 2014

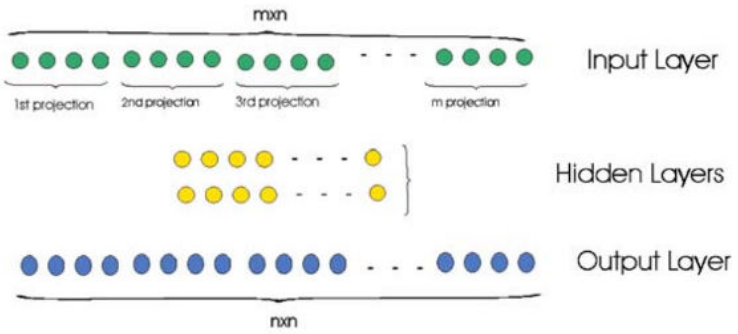


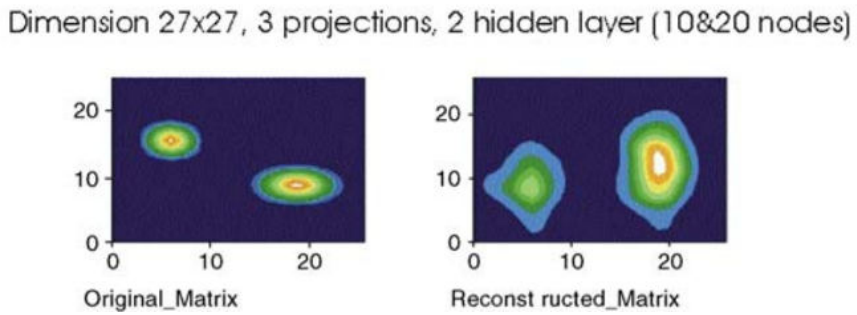
Digging Deeper I

An early approach based on neural networks:

Paschalis et al.; 2004:

- Train feedforward neural network $\mathcal{NN}_\theta : \mathbb{R}^m \to \mathbb{R}^n$ such that $\mathcal{NN}_\theta(\vec{y}) \approx \vec{x}$
- On CPU, no CNNs, ...
- $\rightsquigarrow$ very small toy-examples & more a proof of concept





Digging Deeper II

Another early approach based on neural networks:

LeCun et al.; 2010:

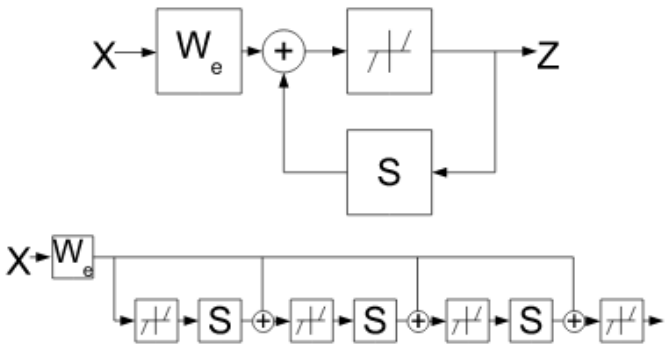
- Based on ISTA (Daubechies et al.; 2004):

$$x_{k+1} := \mathcal{S}_{\tau/\alpha}\left(x_k - 1/\alpha \cdot W_e^T(W_e x_k - y)\right),$$

  which is an algorithm for solving

$$\text{argmin}_{\mathbf{x}} \|W_e x - y\|_2^2 + \tau \cdot \|x\|_1.$$

- Unroll/Unfold iterations and cast as "CNN-like" neural network.

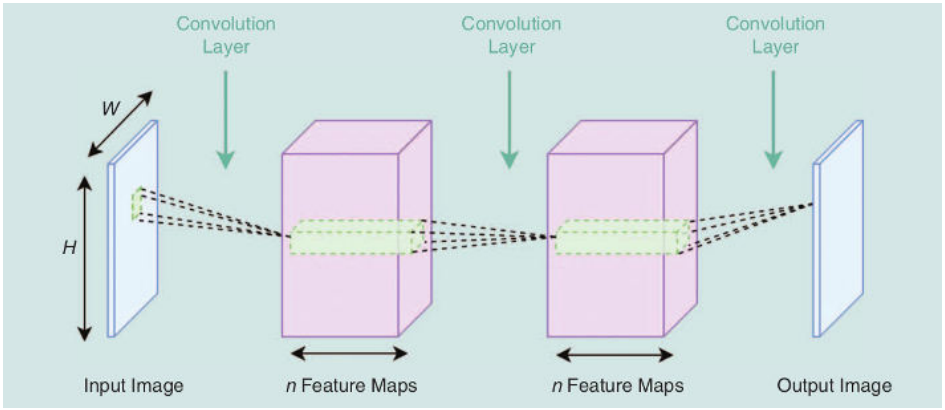


Getting Really Deep: CNNs come into Play

- In 2012 first CNN `AlexNet` wins classification challenge (Hinton et al.; 2012)
- Since then: All winners are CNN-based by huge margin.
- Convolutional-layer: i-th output channel given by

$$o_i = \rho\left(\sum_{j=1}^{c_1} I_j *_2 w_j^i + b_i\right),$$

where $I = [I_1, \ldots, I_{c_1}] \in \mathbb{R}^{k\times k\times c_1}$ input array, $w^i = [w_1^i, \ldots, w_{c1}^i] \in \mathbb{R}^{s\times s\times c_1}$ filter, $b_i \in \mathbb{R}$ bias, non-linearity ρ applied elementwise.



3-layer CNN, [Lucas et al.,2018]

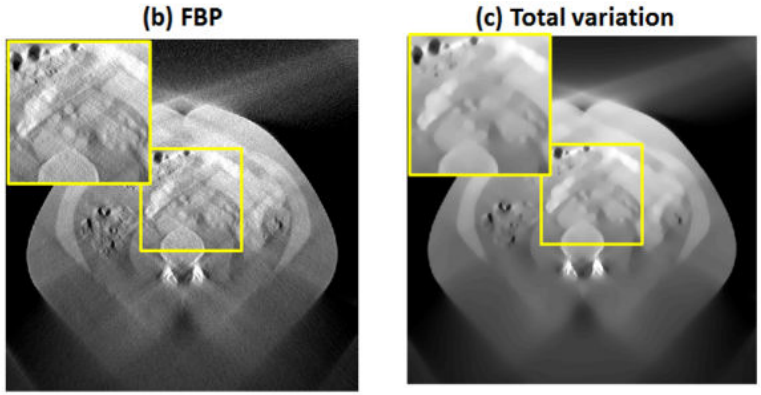
Similar Improvement for Inverse Problems?

Denoising: (Advantage: results are relatively easy to compare)

- 2009: CNN less than 1dB PSNR over wavelet approach (Jain et al.; 2009).
- 2017: Deep CNN-ResNet less than 0.7dB over BM3D (Zhang et al.; 2017).
- "Denoising white additive Gaussian noise is solved"?! (Elad et al.; 2016)

Medical imaging:

- Almost impossible to compare
  $\rightsquigarrow$ too many design choices
- Often no fair comparisons are done
- Effort to reproduce results is immense
- . . . let's look at examples now



(Ye et al.; 2018)

Overview of Current Research Results

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.

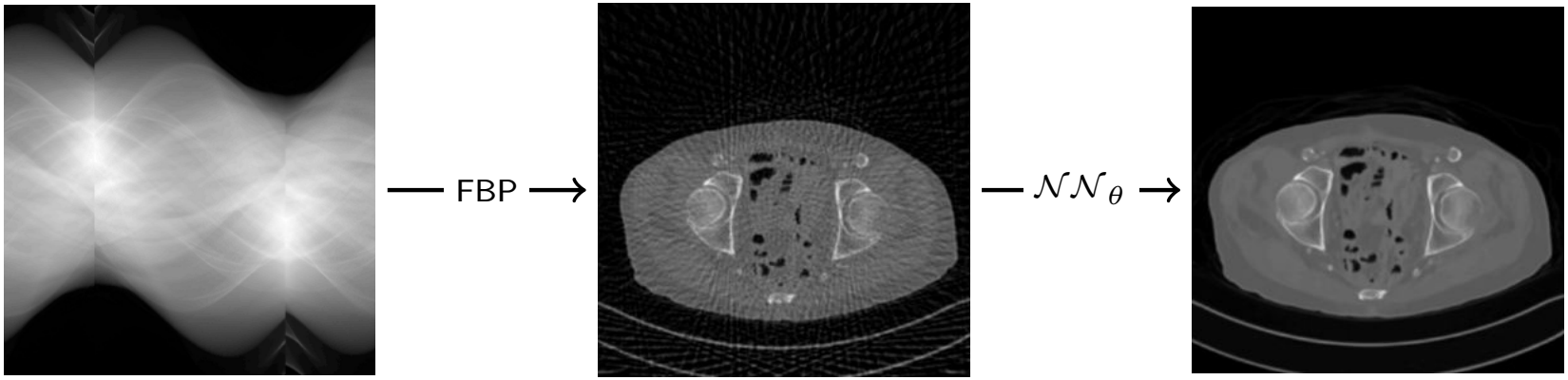
Evaluation:

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:

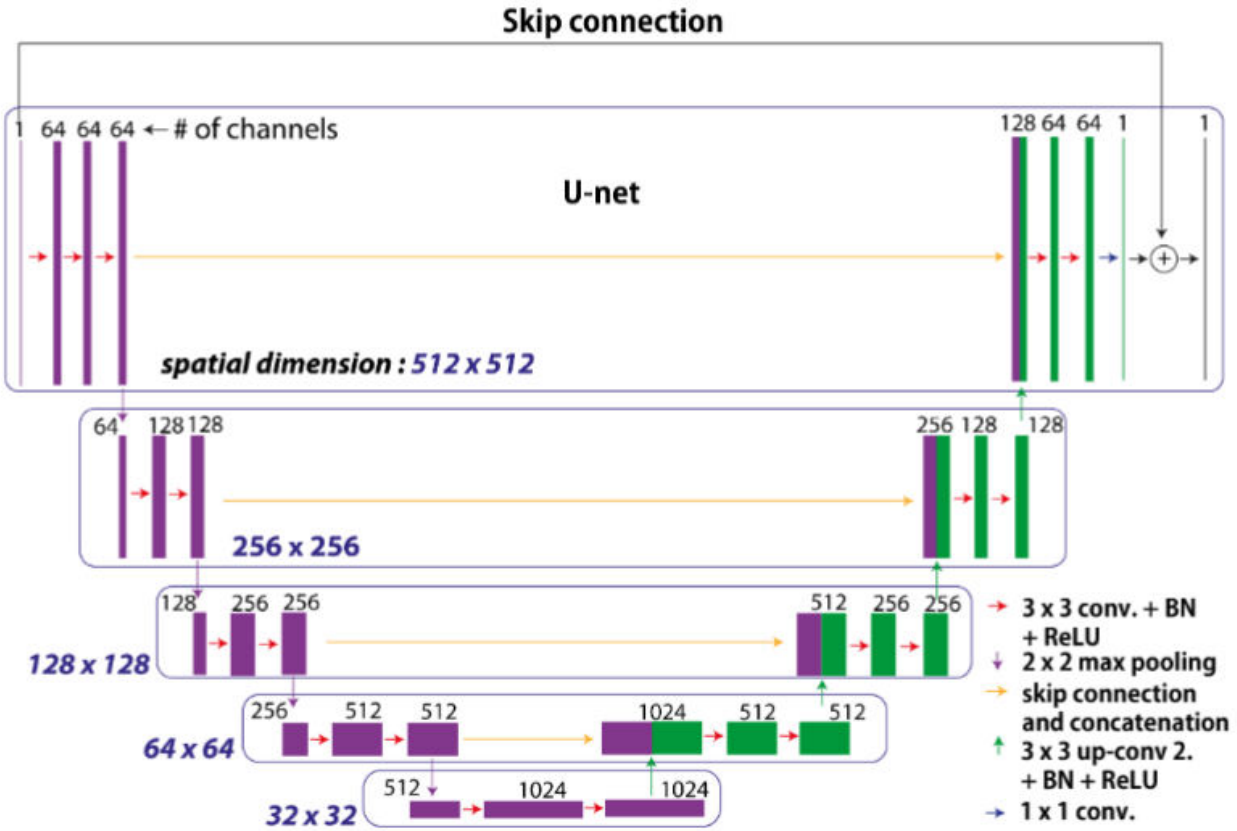


Inversion & denoising $\rightsquigarrow$ Simple, ad-hoc approach to inverse problems

- 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



Other Deep Learning Approaches to Inverse Problems

Tikhonov Regularization:

$$\text{argmin}_f \left[\|Ax - y\|^2 + \ \alpha \cdot \mathcal{R}(x) \right]$$

Solving the problem by Douglas-Rachford (or ADMM ...) results in the iterations

- $x_{k+1} := \text{prox}_{\gamma\lambda\mathcal{R}}(v_k)$
- $v_{k+1} := v_k + \text{prox}_{\gamma f}(2x_{k+1} - v_k) - x_{k+1}$

where $\gamma > 0$, $f = \|A \cdot -y\|_2^2$ and $\text{prox}_f(v) := \text{argmin}_z f(z) + \frac{1}{2}\|z - v\|_2^2$.

Other Deep Learning Approaches to Inverse Problems

Tikhonov Regularization:

$$\operatorname{argmin}_f \left[\|Ax - y\|^2 + \ \alpha \cdot \mathcal{R}(x) \right]$$

Solving the problem by Douglas-Rachford (or ADMM ...) results in the iterations

- $x_{k+1} := \operatorname{prox}_{\gamma\lambda\mathcal{R}}(v_k)$
- $v_{k+1} := v_k + \operatorname{prox}_{\gamma f}(2x_{k+1} - v_k) - x_{k+1}$

where $\gamma > 0$, $f = \|A \cdot -y\|_2^2$ and $\operatorname{prox}_f(v) := \operatorname{argmin}_z f(z) + \frac{1}{2}\|z - v\|_2^2$.

Observations:

- v-update amounts in solving a linear system $\rightsquigarrow$ Tikhonov-regularization
- For $\mathcal{R} = \|\cdot\|_1$, x-update results in soft-thresholding $\rightsquigarrow$ denoising

Other Deep Learning Approaches to Inverse Problems

Tikhonov Regularization:

$$\operatorname{argmin}_f \left[\|Ax - y\|^2 + \ \alpha \cdot \mathcal{R}(x) \right]$$

Solving the problem by Douglas-Rachford (or ADMM ...) results in the iterations

- $x_{k+1} := \operatorname{prox}_{\gamma\lambda\mathcal{R}}(v_k)$
- $v_{k+1} := v_k + \operatorname{prox}_{\gamma f}(2x_{k+1} - v_k) - x_{k+1}$

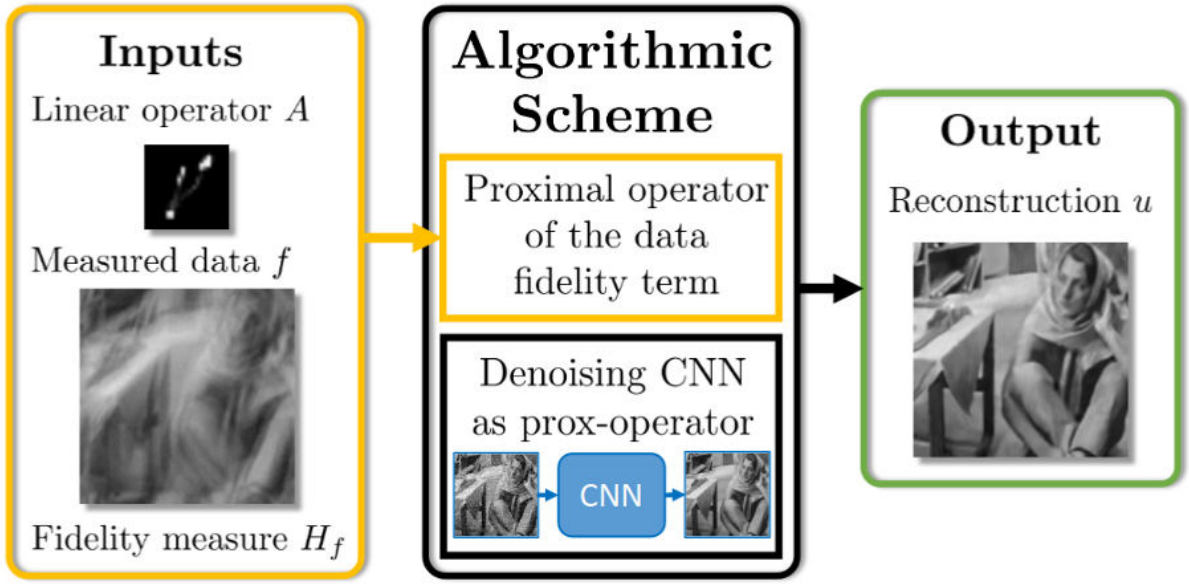
where $\gamma > 0$, $f = \|A \cdot -y\|_2^2$ and $\operatorname{prox}_f(v) := \operatorname{argmin}_z f(z) + \frac{1}{2}\|z - v\|_2^2$.

Observations:

- v-update amounts in solving a linear system $\rightsquigarrow$ Tikhonov-regularization
- For $\mathcal{R} = \|\cdot\|_1$, x-update results in soft-thresholding $\rightsquigarrow$ denoising

Plug-and-play with CNN-denoising (Bouman et al.; 2013), (Elad et al.; 2016), ...:
Replace the denoising step by a trained CNN.

Plug-and-play with CNN-denoising II



(Cremers et al.; 2017)

Advantage:

- Train very sophisticated denoising-CNN on many images
  - $\rightsquigarrow$ reuse for any other inverse problem
  - $\rightsquigarrow$ separate training and inverse problem

Drawback:

- Potentially "slow", due to many applications of A

Learned Iterative Schemes

General idea: Combine mathematical structure of variational methods with deep learning!

General procedure (Yang et al; 2016), (Pock et al.; 2017), (Adler et al.; 2017),...:

1) Pick an algorithm for solving the problem

$$\min_x \|Ax - y\|_2^2 + \mathcal{R}(x)$$

for a general (convex) regularizer $\mathcal{R}$ (e.g. ADMM, Primal-Dual, ...).

2) Replace the proximal-steps by parametrized operators (not necessarily prox), where the parameters are sought to be learned.

Compressed Sensing using Generative Models

Generative Models:

- Examples are variational auto-encoder or generative adversarial networks (GANs)
- General: Neural networks

$$\mathbb{R}^k \ni z \mapsto G(z) \in \mathbb{R}^n,$$

where $k \ll n$ and G is trained to produce elements similar to training data.

Task:

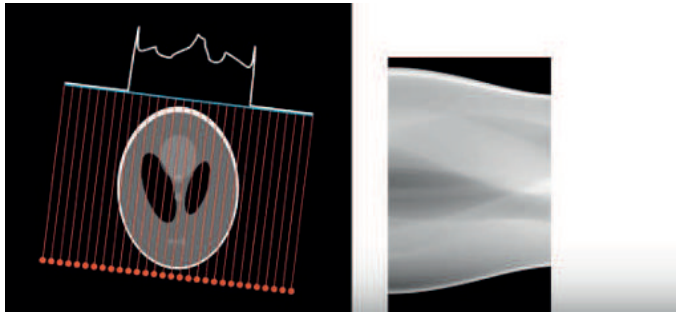
- Let $A \in \mathbb{R}^{m \times n}$ Gaussian matrix, measurements $y = Ax_0 + \eta$ be given.
- Solve $z_0 \in \operatorname{argmin}_{z \in \mathbb{R}^k} \|AG(z) - y\|_2^2$ (non-convex) to within additive ε of optimum.

Theorem (Bora et al.; 2017): G generative model from d-layer ReLU neural network and $m \in \mathcal{O}(kd \log(n))$. Then with overwhelming probability

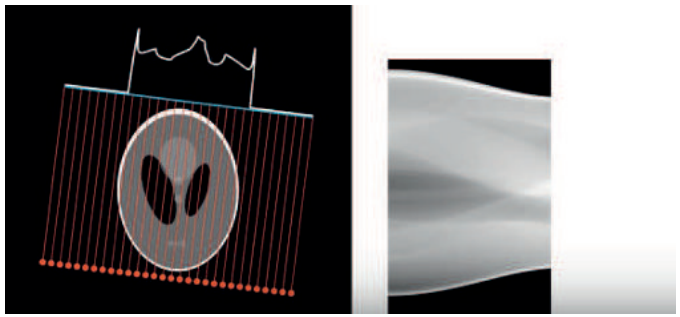
$$\|G(z_0) - x_0\|_2 \leq 6 \min_{z \in \mathbb{R}^k} \|G(z) - x_0\|_2 + 3\|\eta\|_2 + 2\varepsilon.$$

Mathematical Modeling Reaches a Barrier

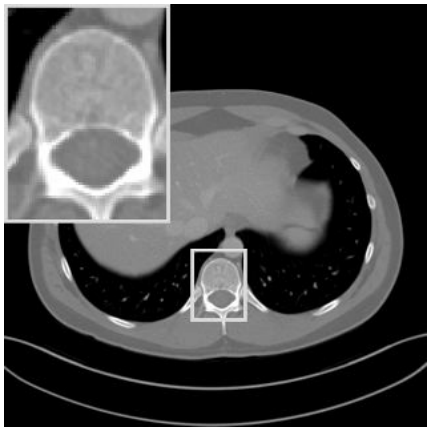
Computed Tomography (CT)

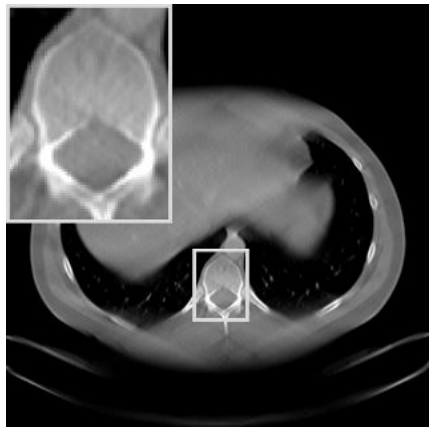


Computed Tomography (CT)



Problem with Limited-Angle Tomography:





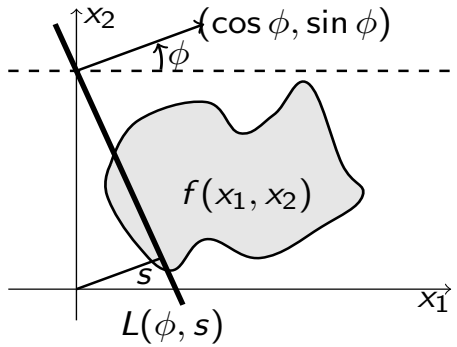
The data is too complex for mathematical modeling!

Limited Angle-(Computed) Tomography

A CT scanner samples the *Radon transform*

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

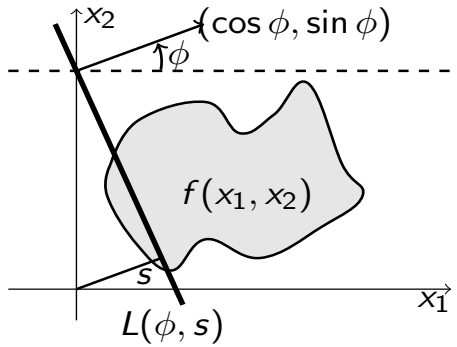
for $L(\phi, s) = \left\{x \in \mathbb{R}^2 : x_1 \cos(\phi) + x_2 \sin(\phi) = s\right\}$,
$\phi \in [-\pi/2, \pi/2)$, and $s \in \mathbb{R}$.



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 $[-\phi, \phi] \subset [-\pi/2, \pi/2)$.

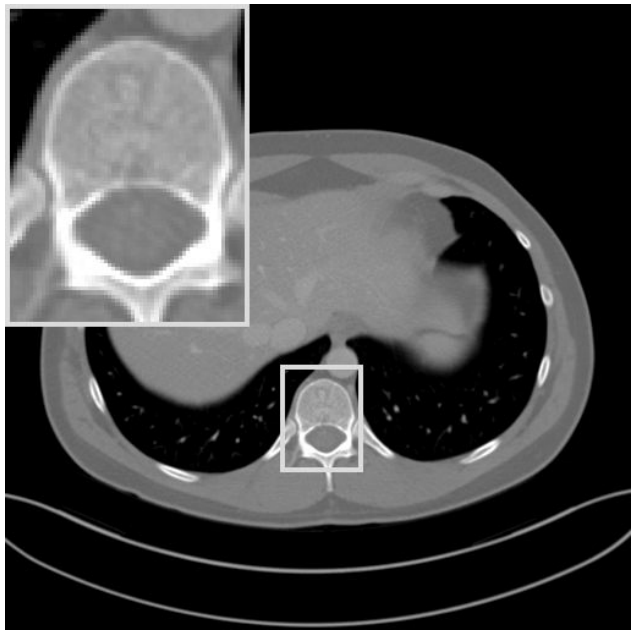
Applications: Dental CT, breast tomosynthesis, electron tomography,...

Model-Based Approaches Fail

Sparse Regularization:

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

Clinical Data:



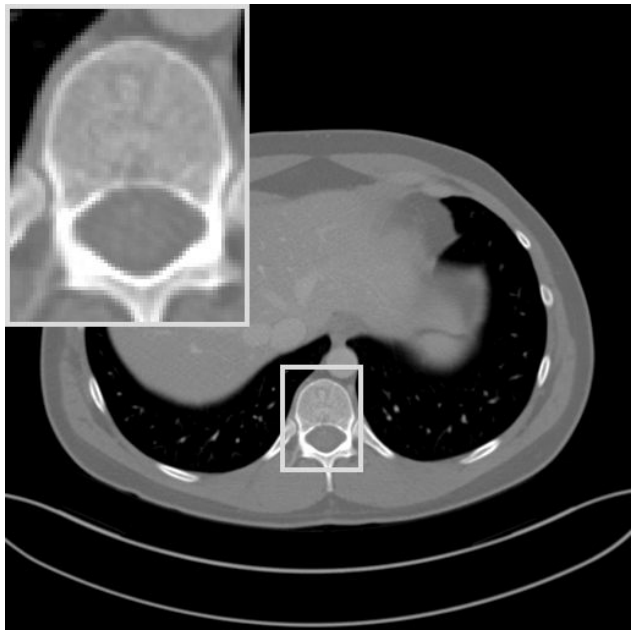
Original Image

Model-Based Approaches Fail

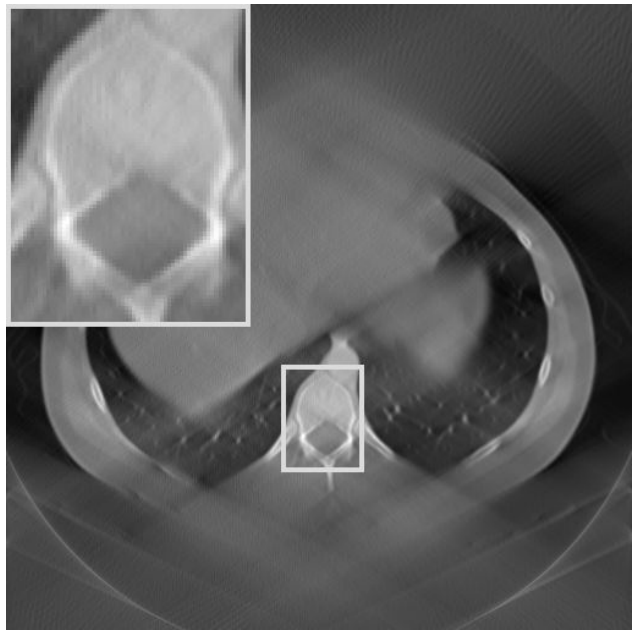
Sparse Regularization:

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

Clinical Data:



Original Image



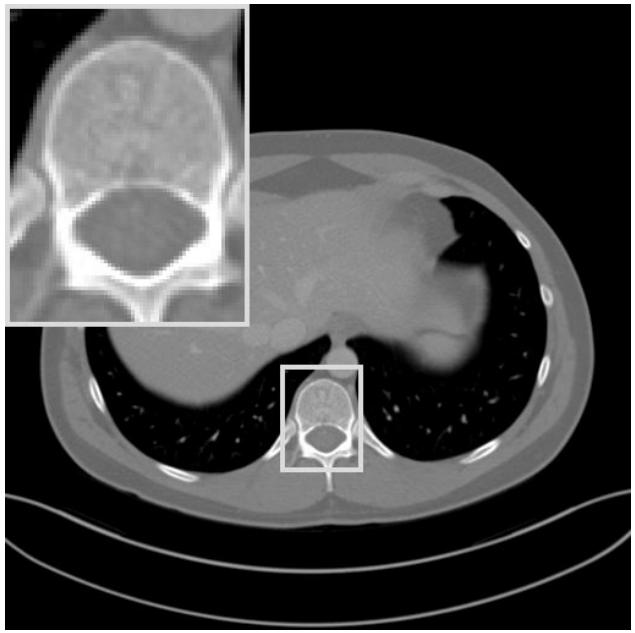
Filtered Backprojection

Model-Based Approaches Fail

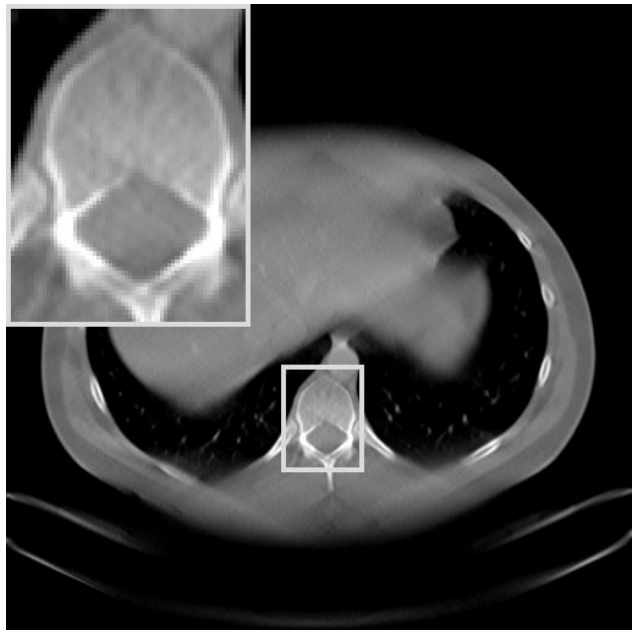
Sparse Regularization:

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

Clinical Data:



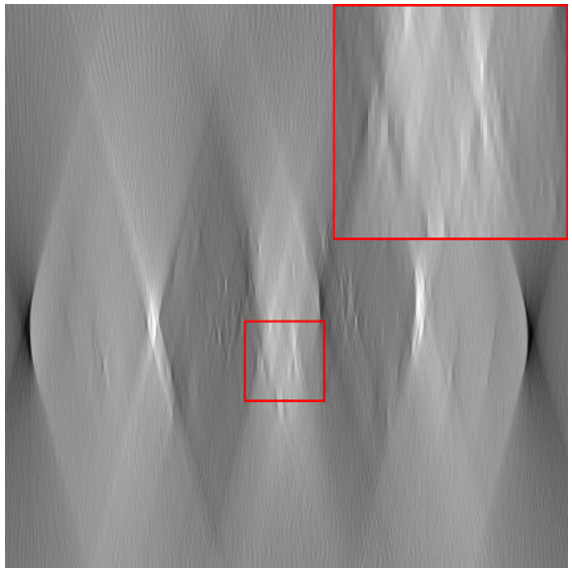
Original Image



Sparse Regularization with Shearlets

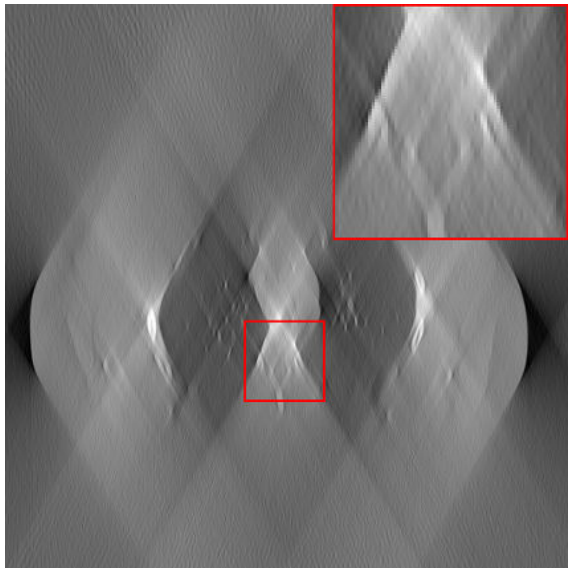
A True Hybrid Approach

Zooming in on the Recovery Problem



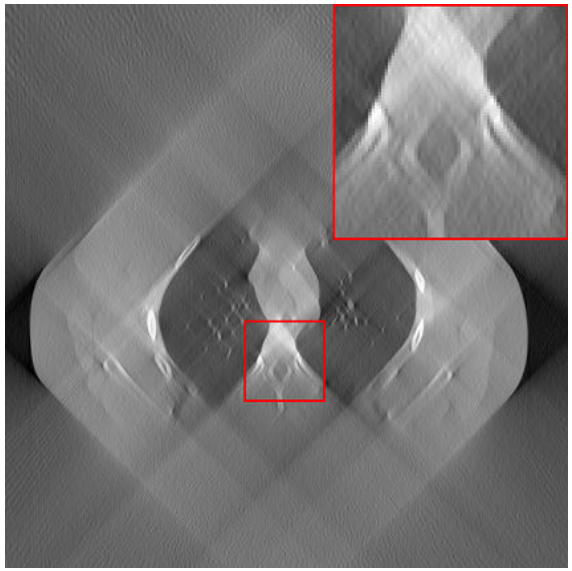
$\phi = 15°$, filtered backprojection (FBP)

Zooming in on the Recovery Problem



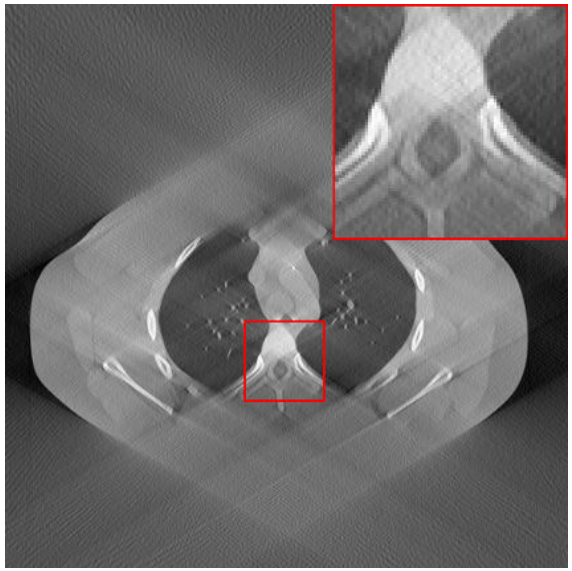
$\phi = 30°$, filtered backprojection (FBP)

Zooming in on the Recovery Problem



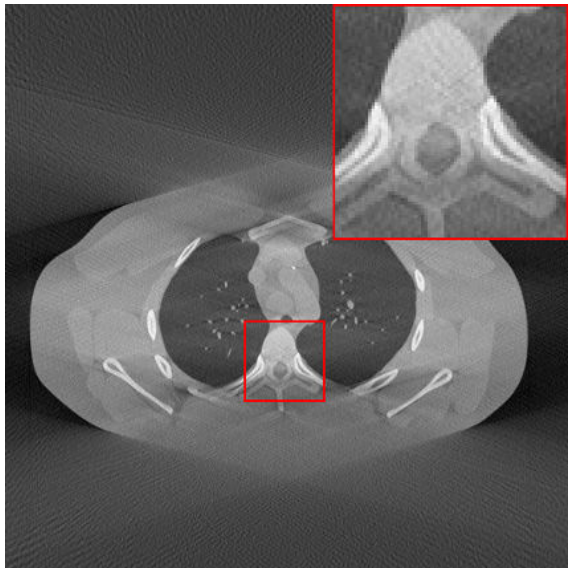
$\phi = 45°$, filtered backprojection (FBP)

Zooming in on the Recovery Problem



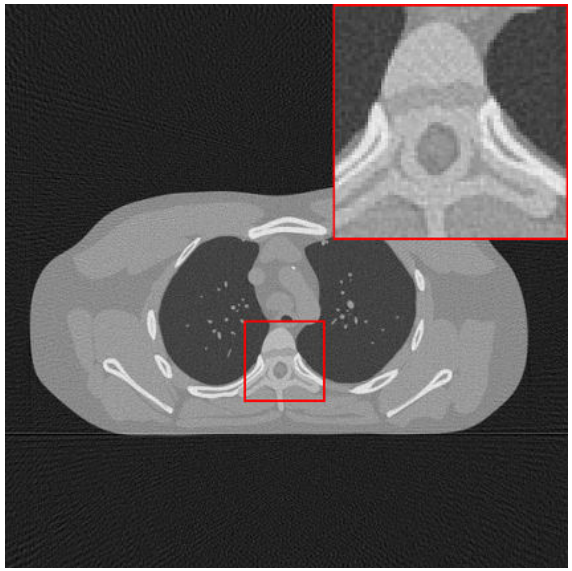
$\phi = 60°$, filtered backprojection (FBP)

Zooming in on the Recovery Problem



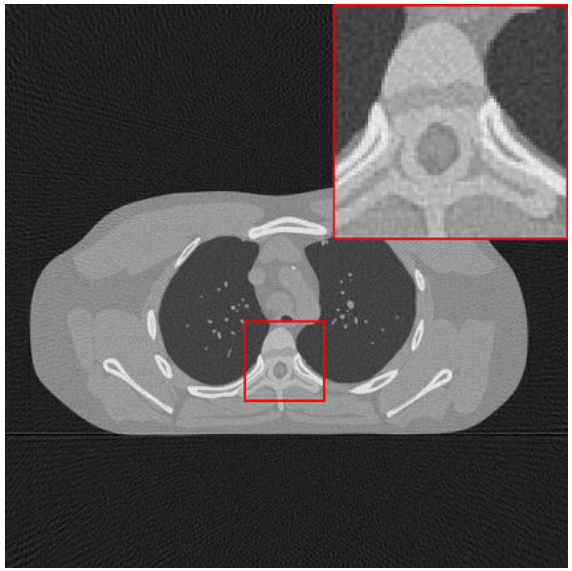
$\phi = 75°$, filtered backprojection (FBP)

Zooming in on the Recovery Problem



$\phi = 90°$, filtered backprojection (FBP)

Zooming in on the Recovery Problem



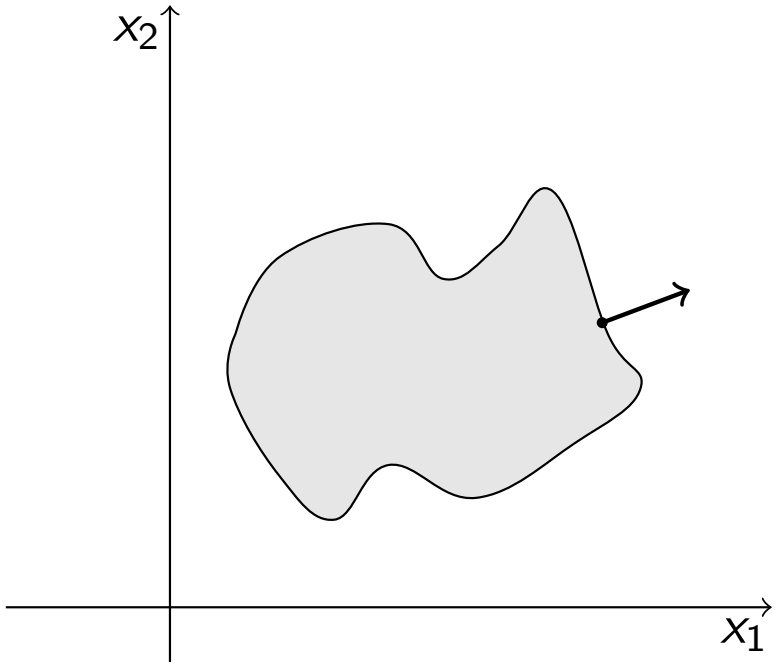
$\phi = 90°$, filtered backprojection (FBP)

Some Observations:

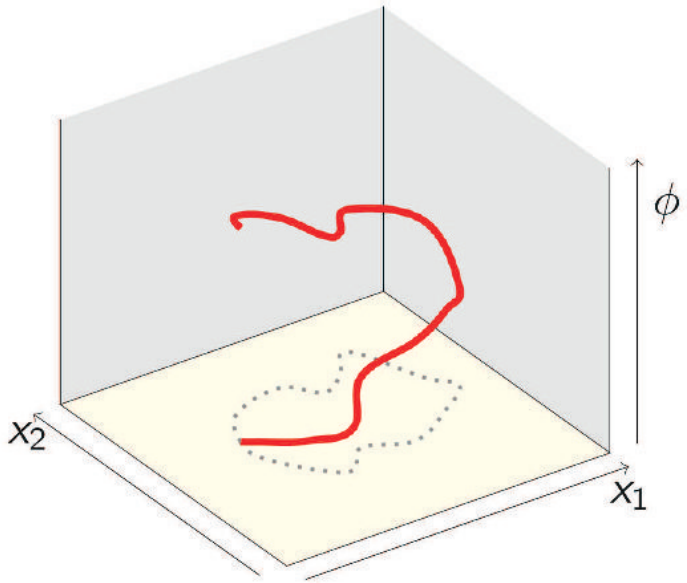
- Only certain boundaries/features seem to be *"visible"*!
- Missing wedge creates artifacts!
- Highly ill-posed inverse problem!

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

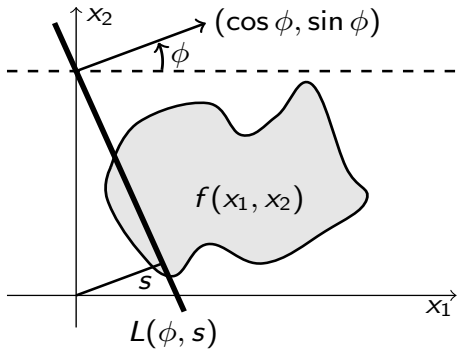


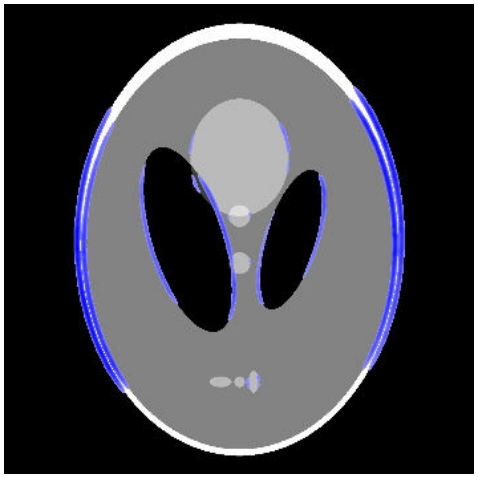
Visualization in phase space

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 $W(\mathcal{R}f)$ 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).



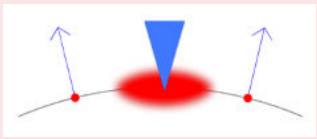


"visible": singularities tangent to sampled lines

"invisible": singularities not tangent to sampled lines

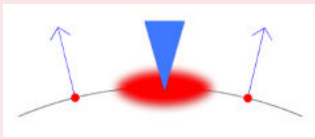
Shearlets can Help

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

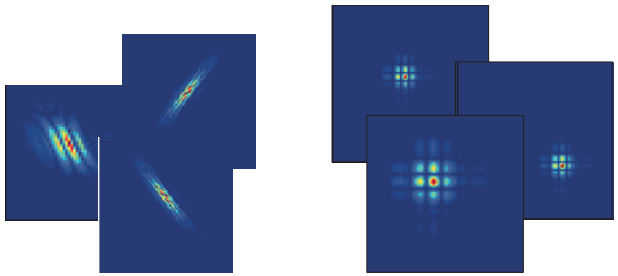


Shearlets can Help

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:

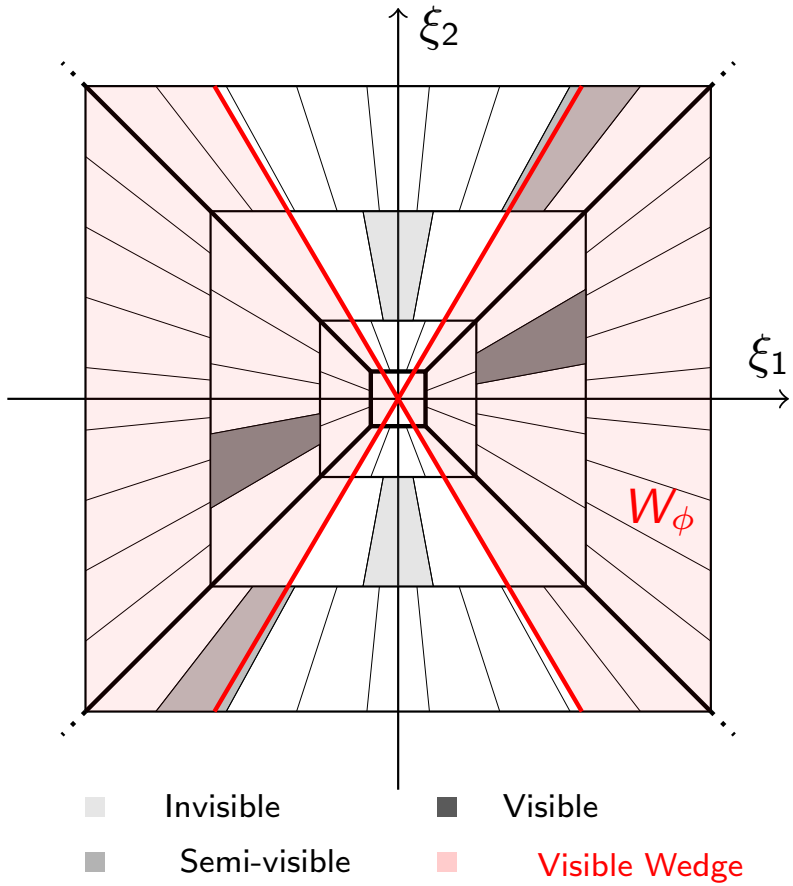
- Continuous Shearlet Transform:

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

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

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

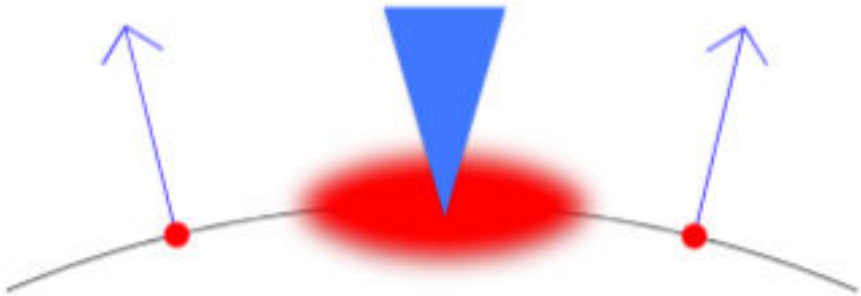
Shearlets can Separate the Visible and Invisible Part



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.



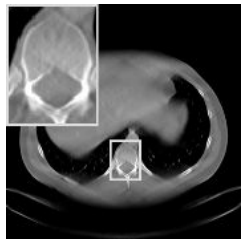
Practical Questions:

- 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)"

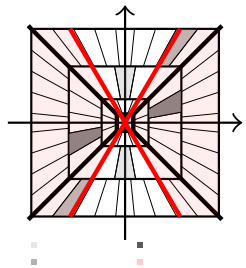
(Bubba, K, Lassas, März, Samek, Siltanen, Srinivan; 2018)

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*

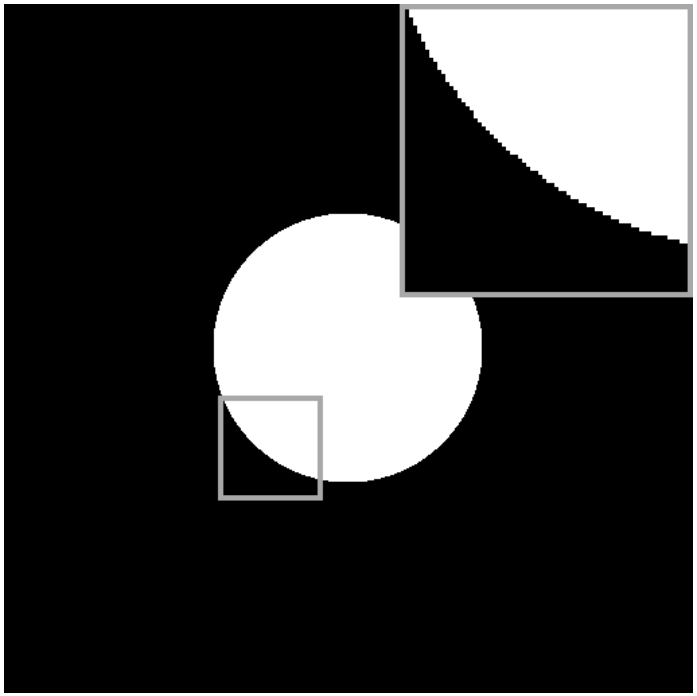
$$\mathcal{NN}_\theta : \mathrm{SH}_\psi(f^*)_{\mathcal{I}_{\mathtt{vis}}} \longrightarrow F \left(\overset{!}{\approx} \mathrm{SH}_\psi(f_{\mathtt{gt}})_{\mathcal{I}_{\mathtt{inv}}} \right)$$

Step 3: *Combine*

$$f_{\mathtt{LtI}} = \mathrm{SH}_\psi^T \left(\mathrm{SH}_\psi(f^*)_{\mathcal{I}_{\mathtt{vis}}} + F \right) \tag{1}$$

Numerical Simulation

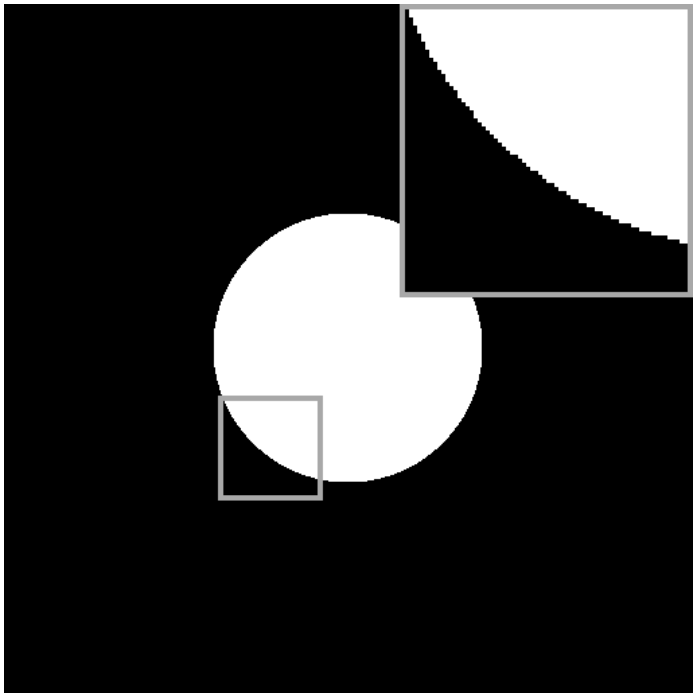
Verify the concept of (in-)visibility



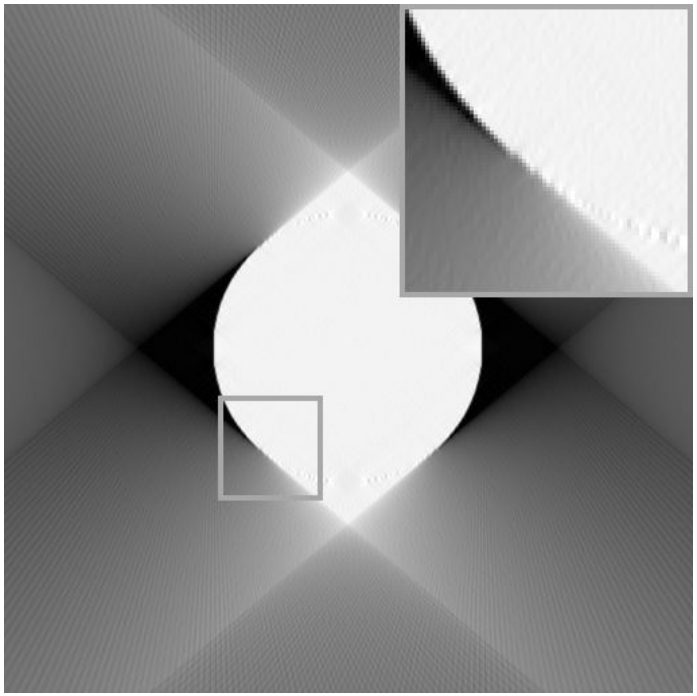
f_{gt}

Numerical Simulation

Verify the concept of (in-)visibility



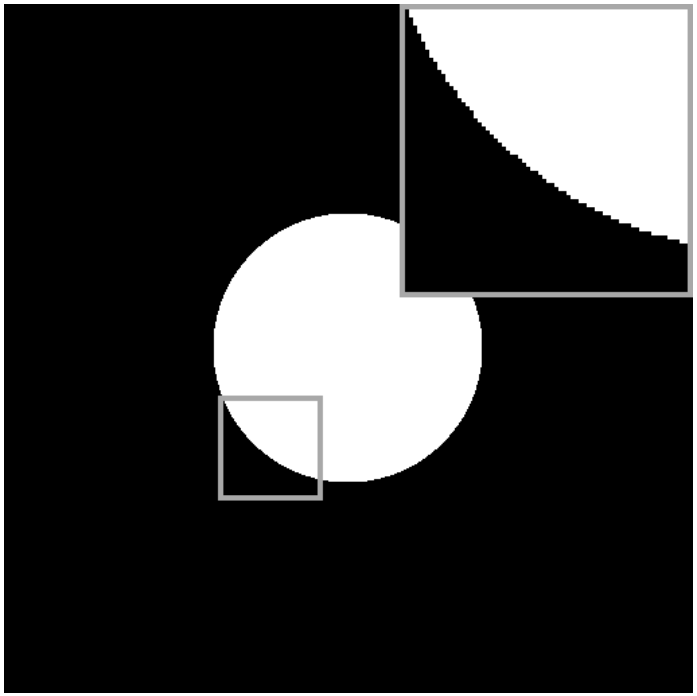
f_{gt}



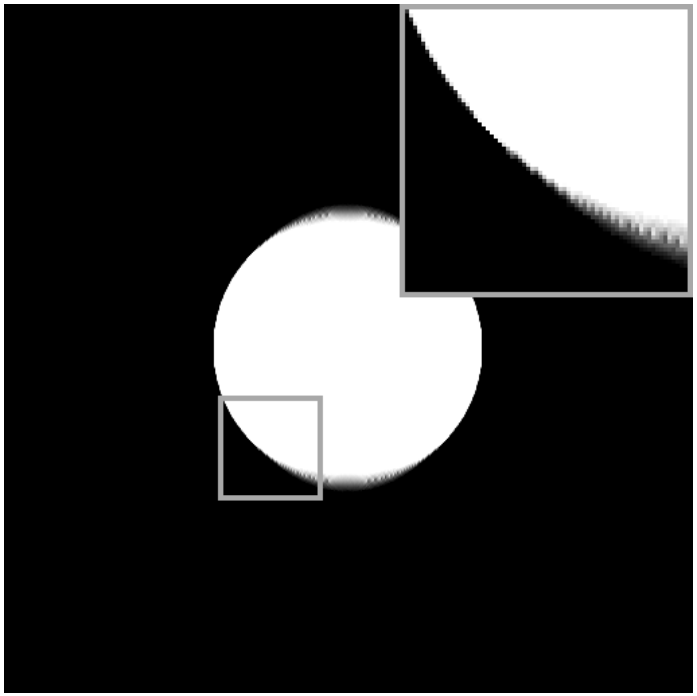
FBP

Numerical Simulation

Verify the concept of (in-)visibility



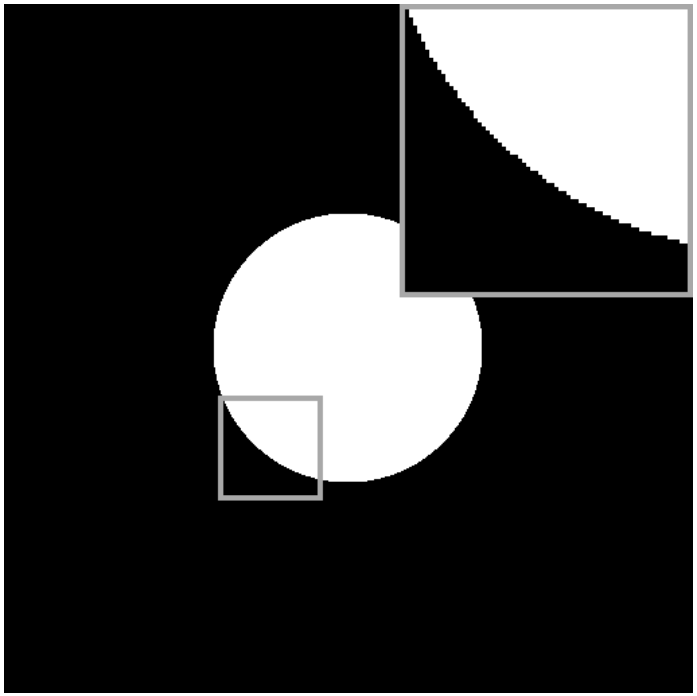
f_{gt}



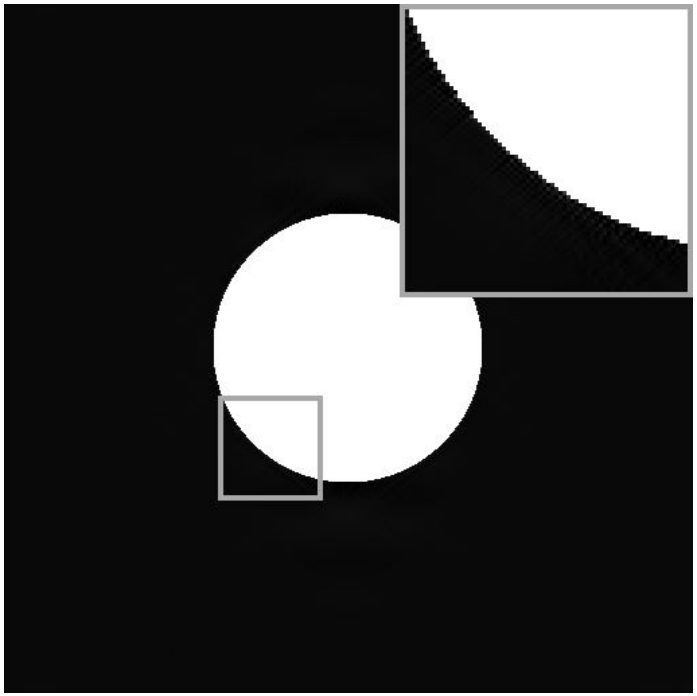
ℓ_1-analysis shearlet solution f^*

Numerical Simulation

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



$f_{\mathtt{gt}}$

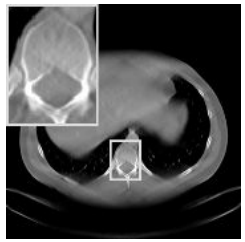


$$\mathsf{SH}_\psi^T \left(\mathsf{SH}_\psi(f^*)_{\mathcal{I}_{\mathtt{vis}}} + \mathsf{SH}_\psi(f_{\mathtt{gt}})_{\mathcal{I}_{\mathtt{inv}}} \right)$$

Our Approach "Learn the Invisible (LtI)"

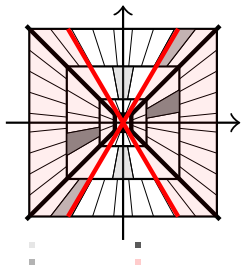
(Bubba, K, Lassas, März, Samek, Siltanen, Srinivan; 2018)

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*

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

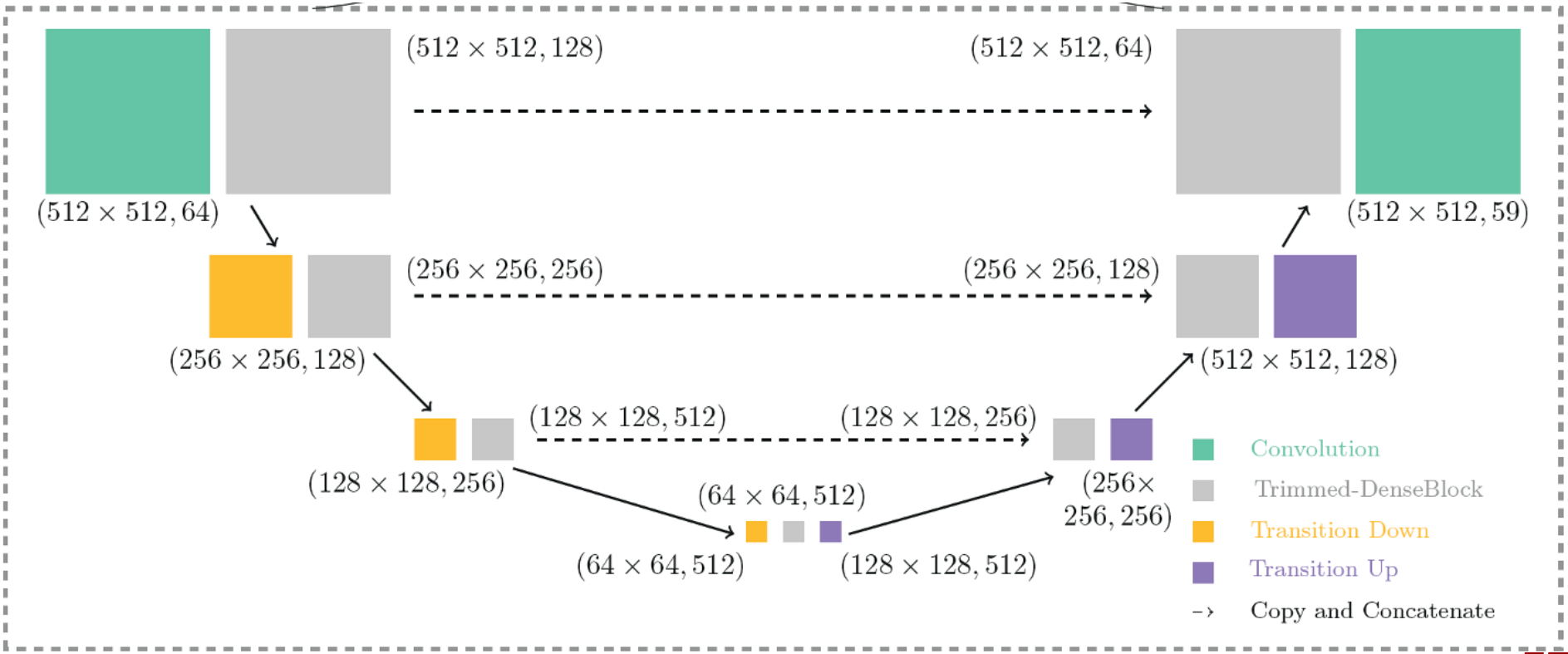
Step 3: *Combine*

$$f_{\mathtt{LtI}} = \mathrm{SH}_\psi^T \left(\mathrm{SH}_\psi(f^*)_{\mathcal{I}_{\mathtt{vis}}} + F \right) \tag{2}$$

Our Approach – Step 2: PhantomNet

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

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



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

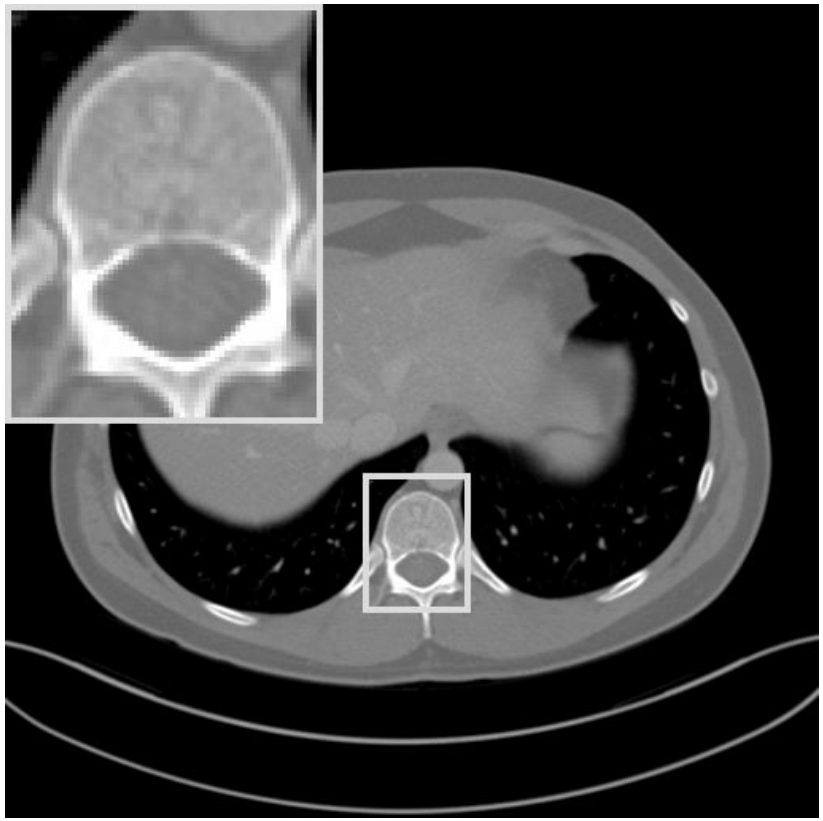
Setup

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
- ...

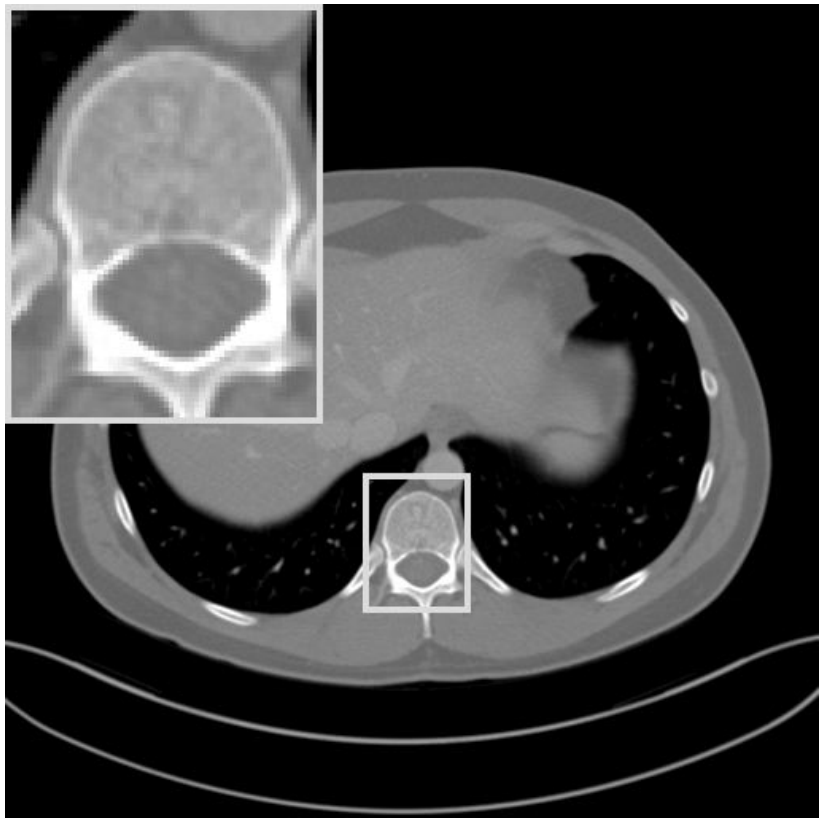
[1] We would like to thank Dr. Cynthia McCollough, the Mayo Clinic, the American Association of Physicists in Medicine (AAPM), and grant EB01705 and EB01785 from the National Institute of Biomedical Imaging and Bioengineering for providing the Low-Dose CT Grand Challenge data set.

Evaluation on Test Patient

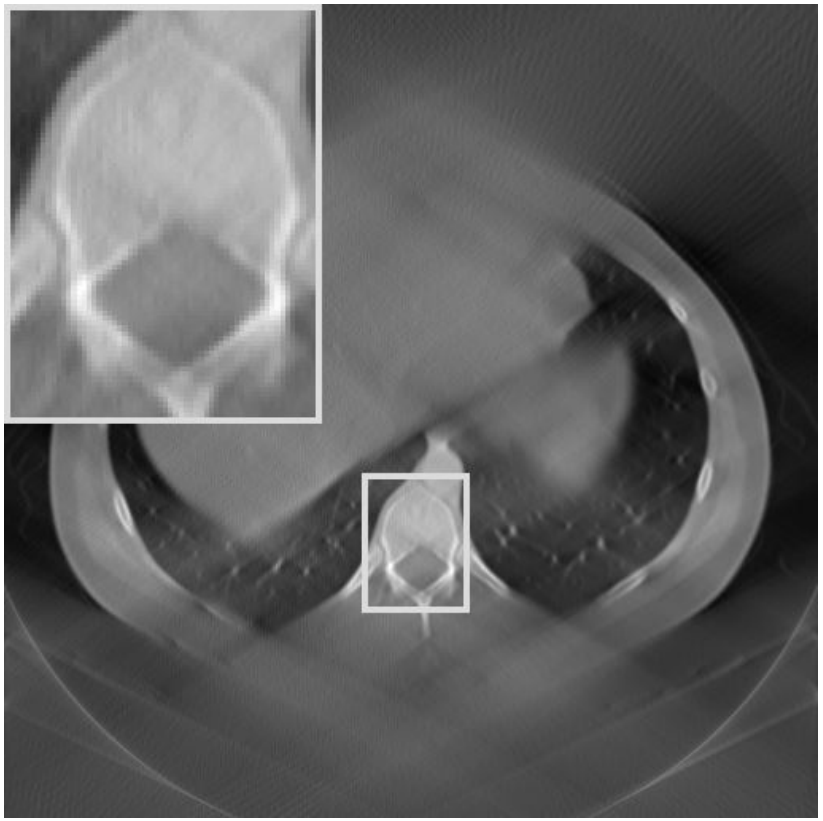


f_{gt}

Evaluation on Test Patient

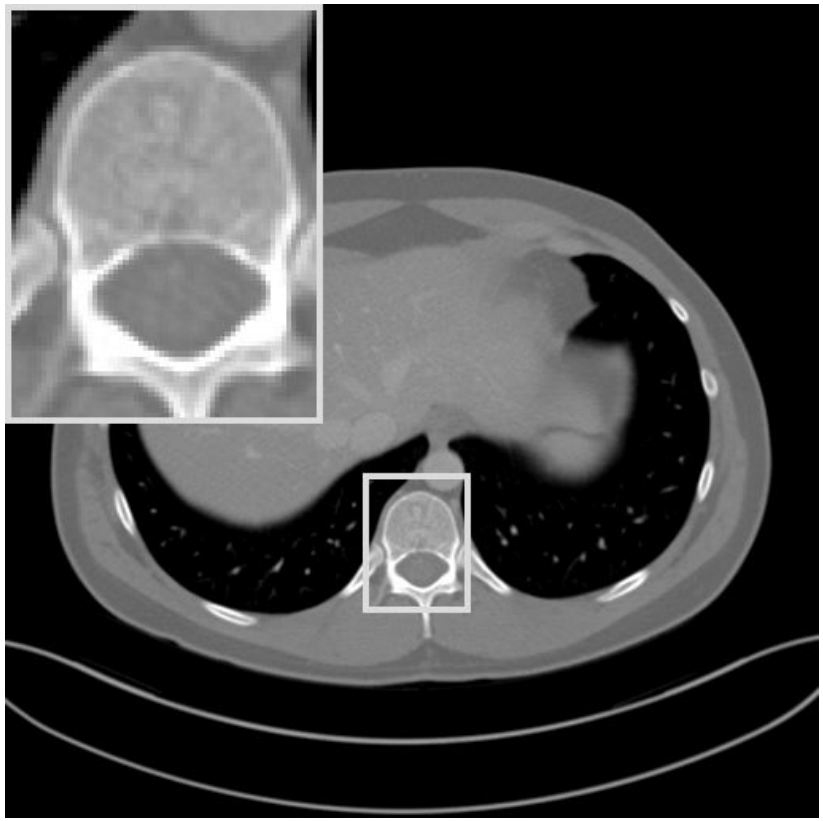


f_{gt}

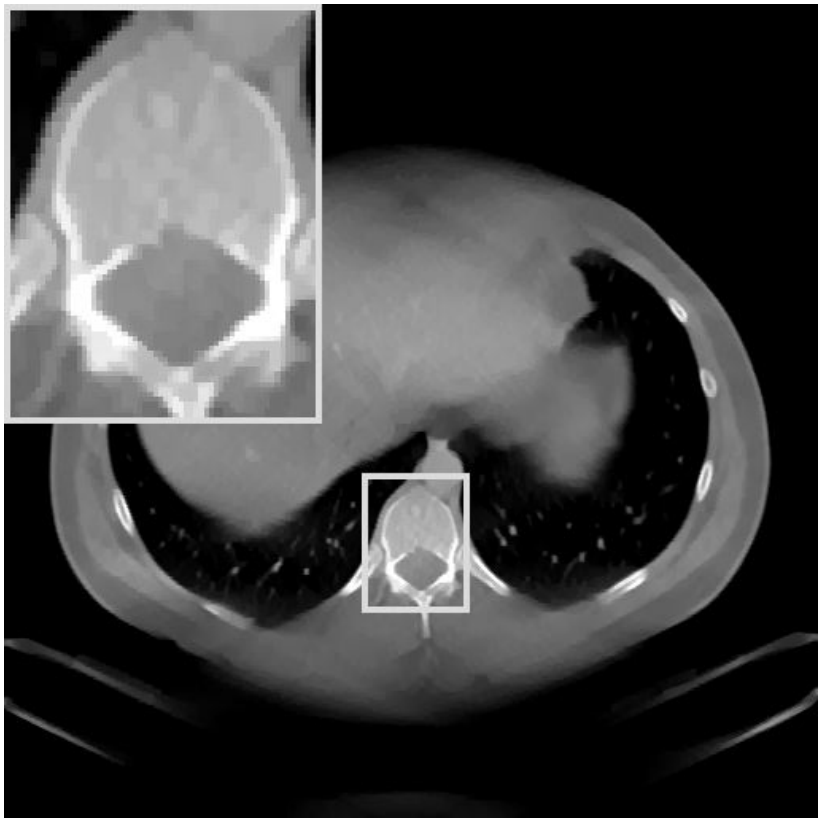


f_{FBP}: RE $= 0.50$, HaarPSI$=0.35$

Evaluation on Test Patient

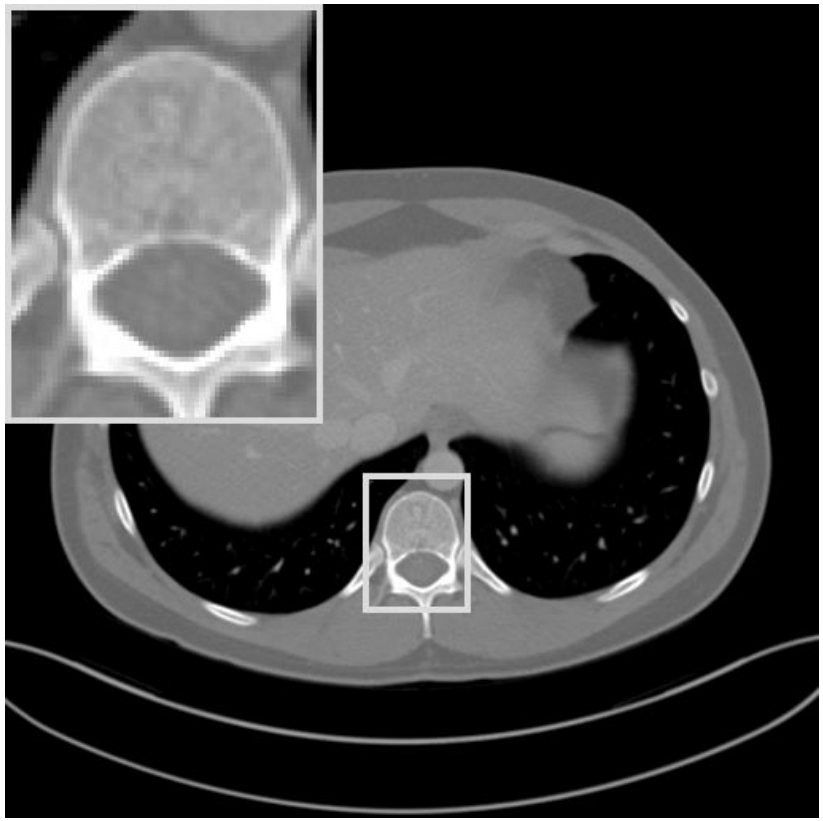


f_{gt}

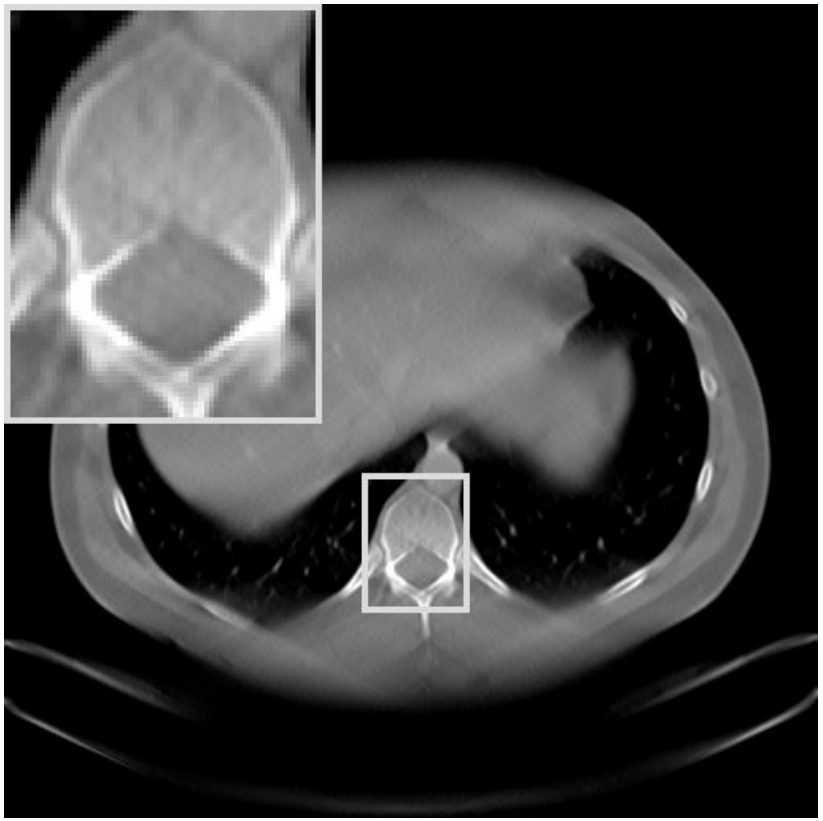


f_{TV}: RE $= 0.21$, HaarPSI$=0.41$

Evaluation on Test Patient

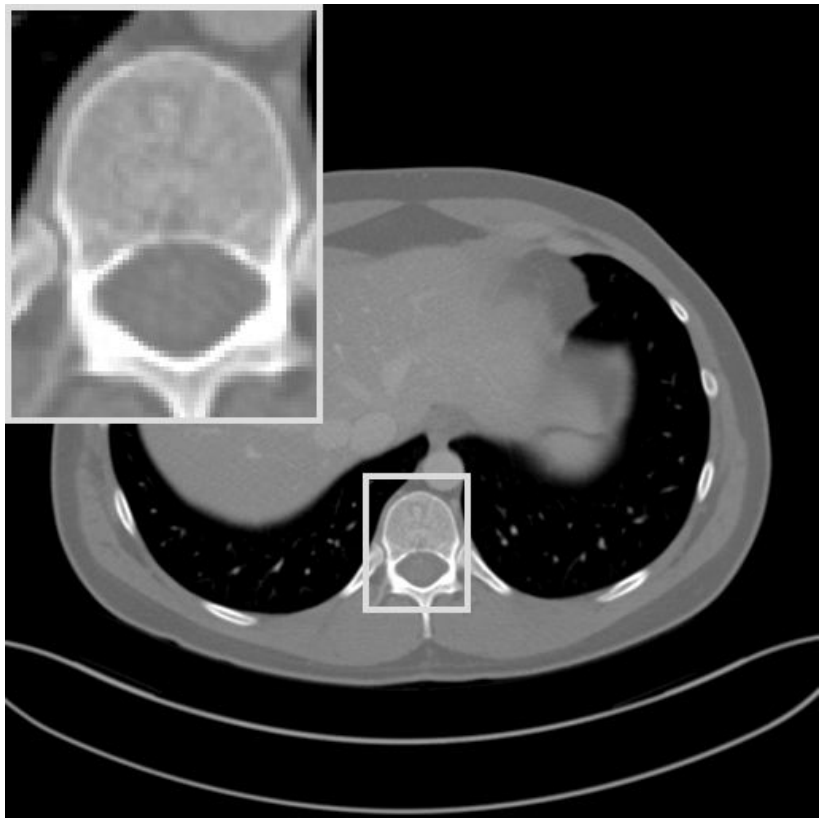


f_{gt}

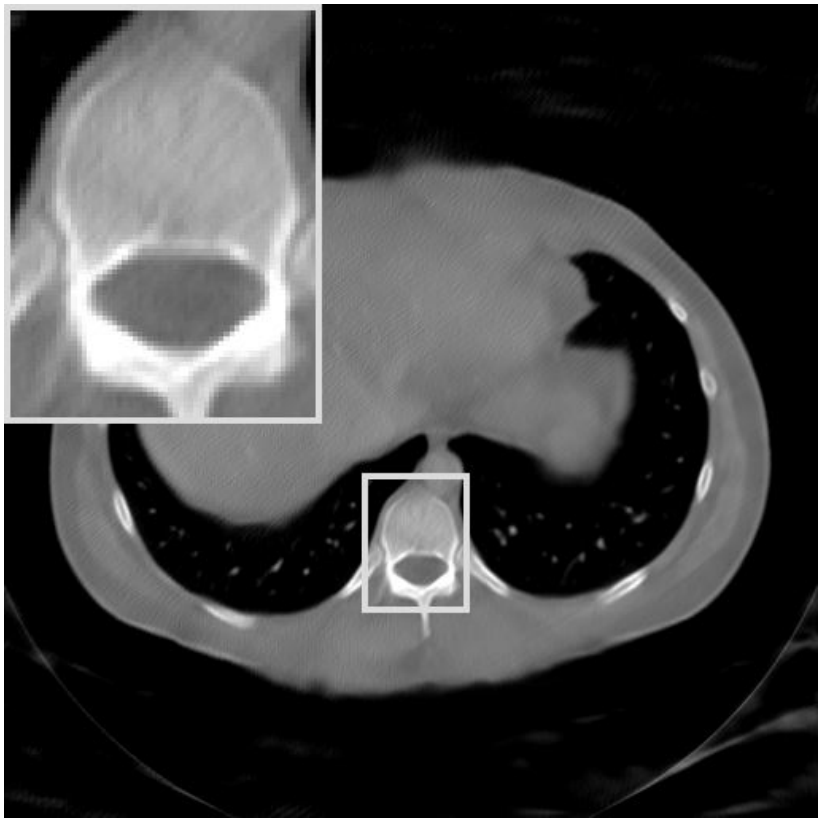


f^*: RE $= 0.19$, HaarPSI$=0.43$

Evaluation on Test Patient

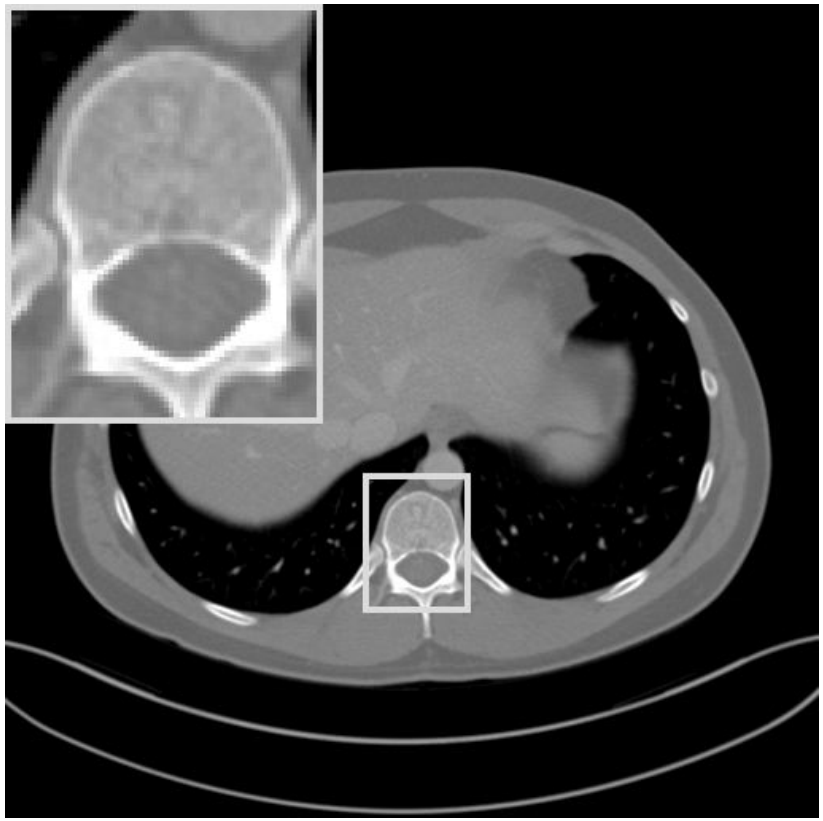


f_{gt}



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

Evaluation on Test Patient



f_{gt}

$f_{\texttt{LtI}}$: 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
$\mathcal{NN}_\theta(f_{\mathtt{FBP}})$	0.15	27.40	0.78	0.52
$\mathcal{NN}_\theta(\mathsf{SH}(f_{\mathtt{FBP}}))$	0.16	26.80	0.74	0.52
$f_{\mathtt{LtI}}$	**0.08**	**32.77**	**0.93**	**0.73**

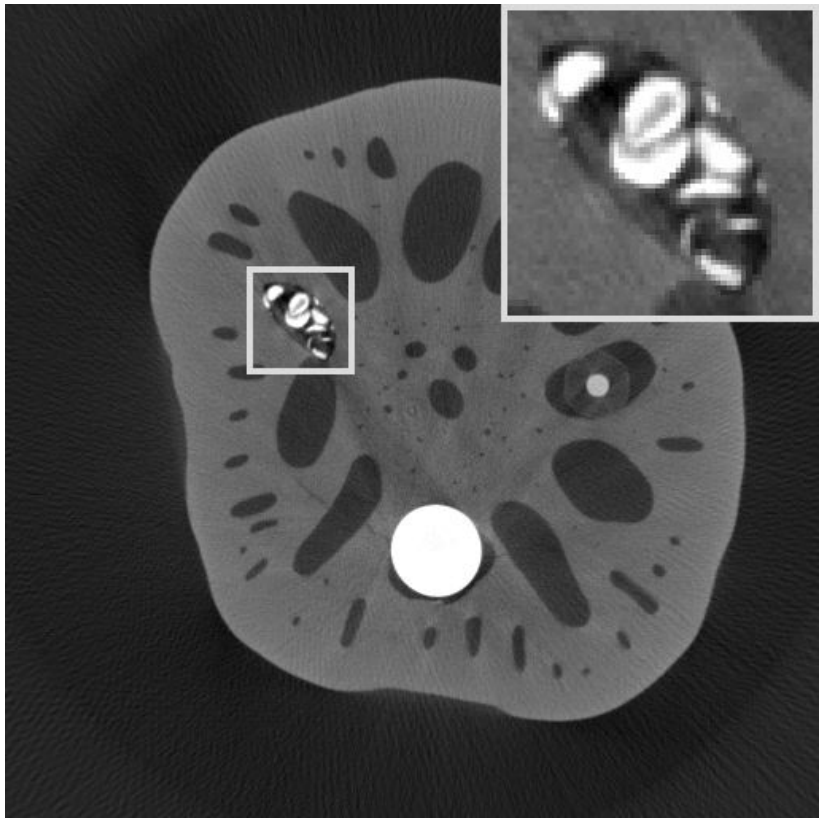
HaarPSI (Reisenhofer, Bosse, K, and Wiegand; 2018)

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

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

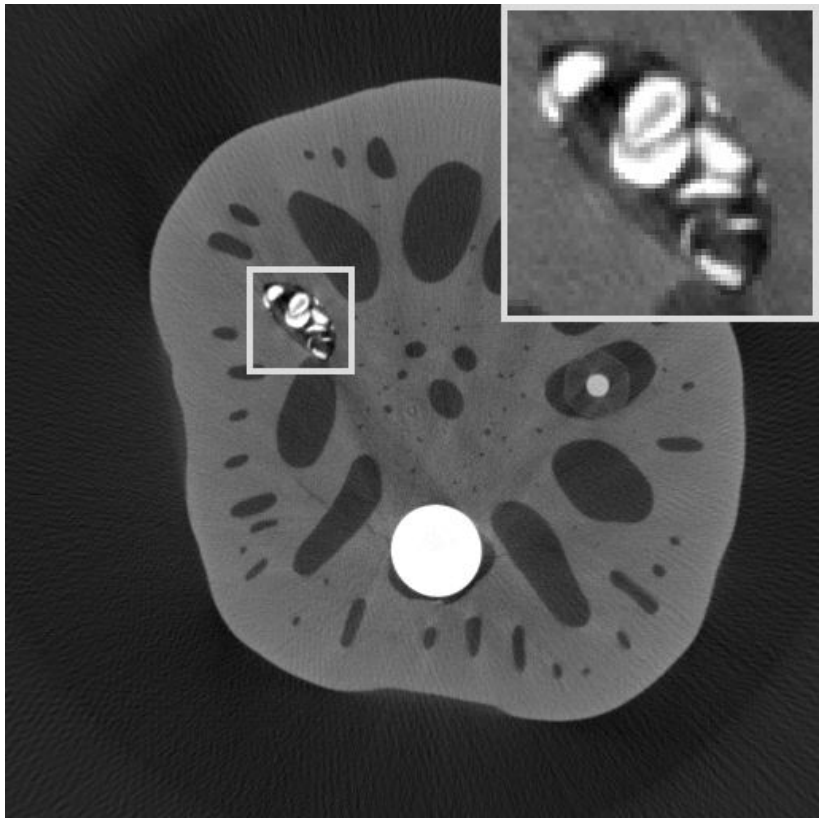
`www.haarpsi.org`

Generalization to Lotus Root

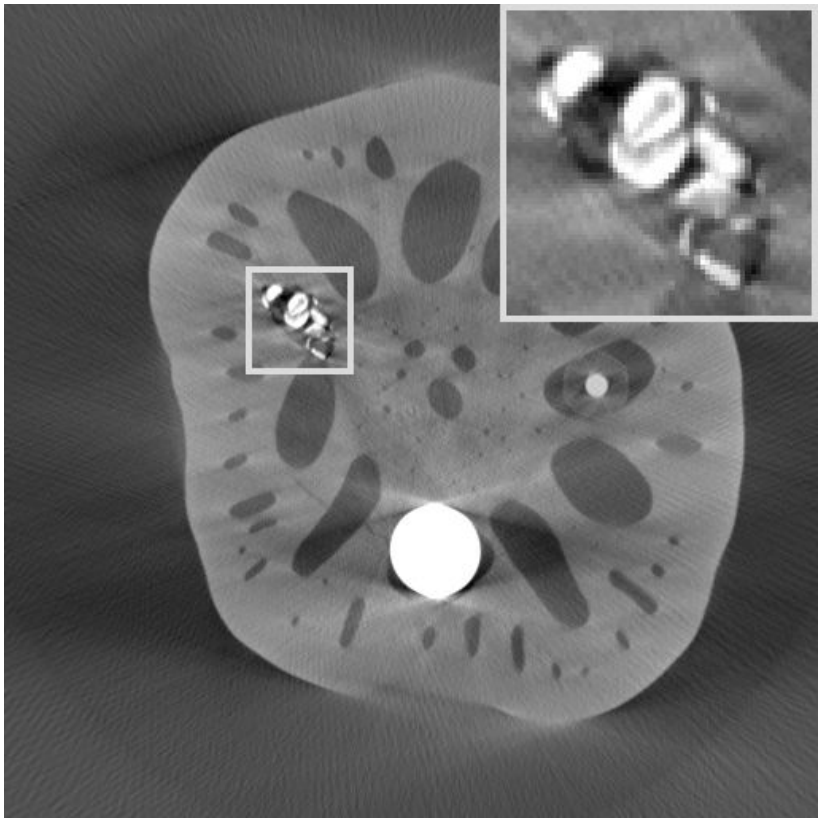


f_{gt}

Generalization to Lotus Root

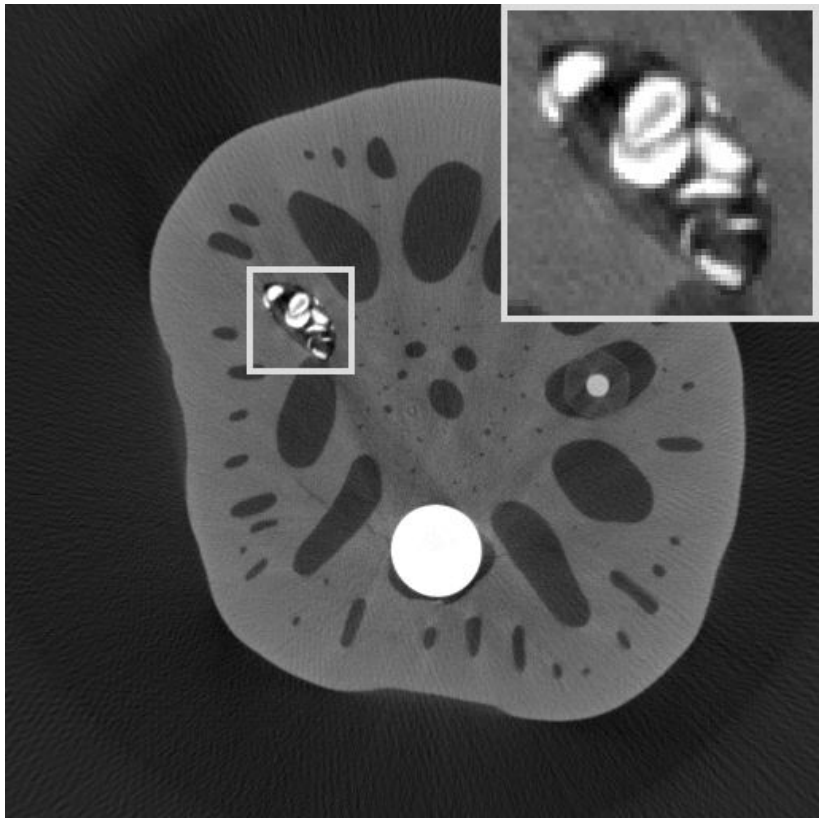


f_{gt}

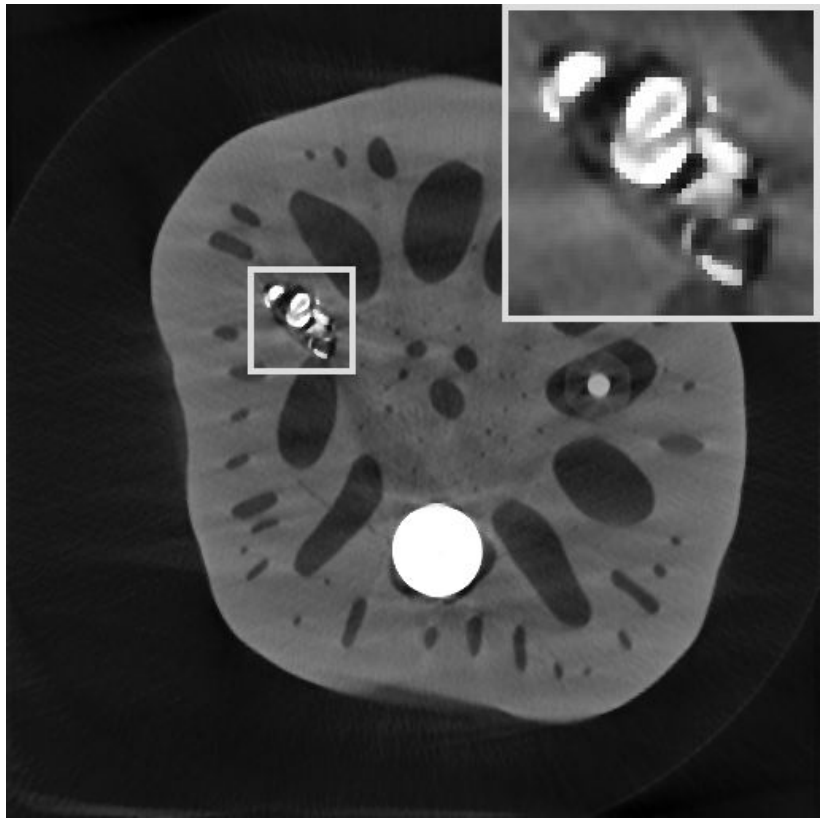


f_{FBP}: RE $= 0.31$, HaarPSI$=0.61$

Generalization to Lotus Root

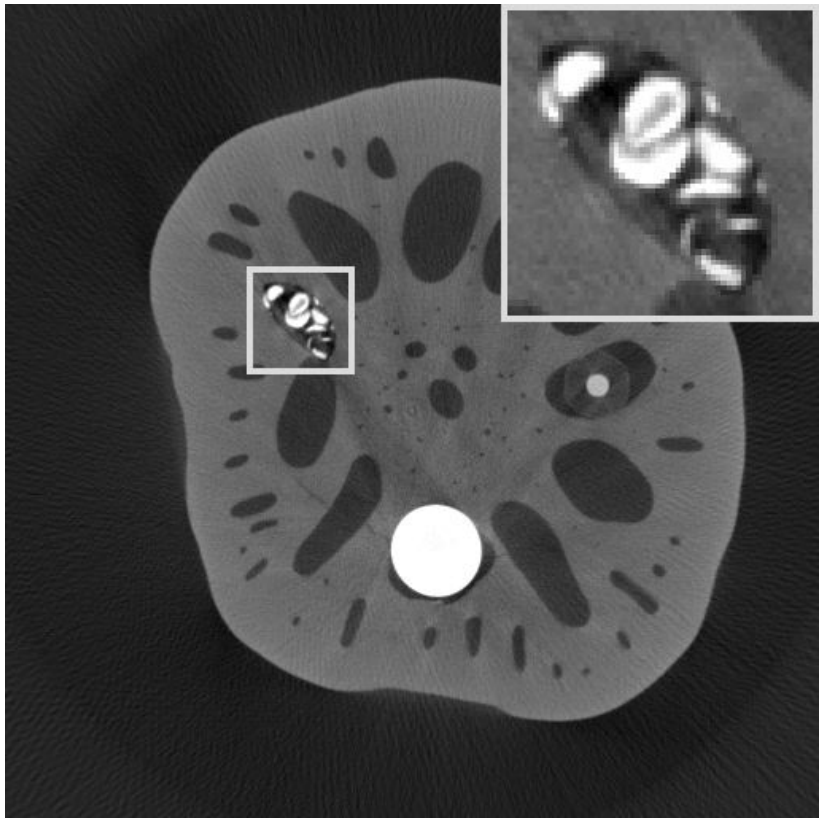


f_{gt}

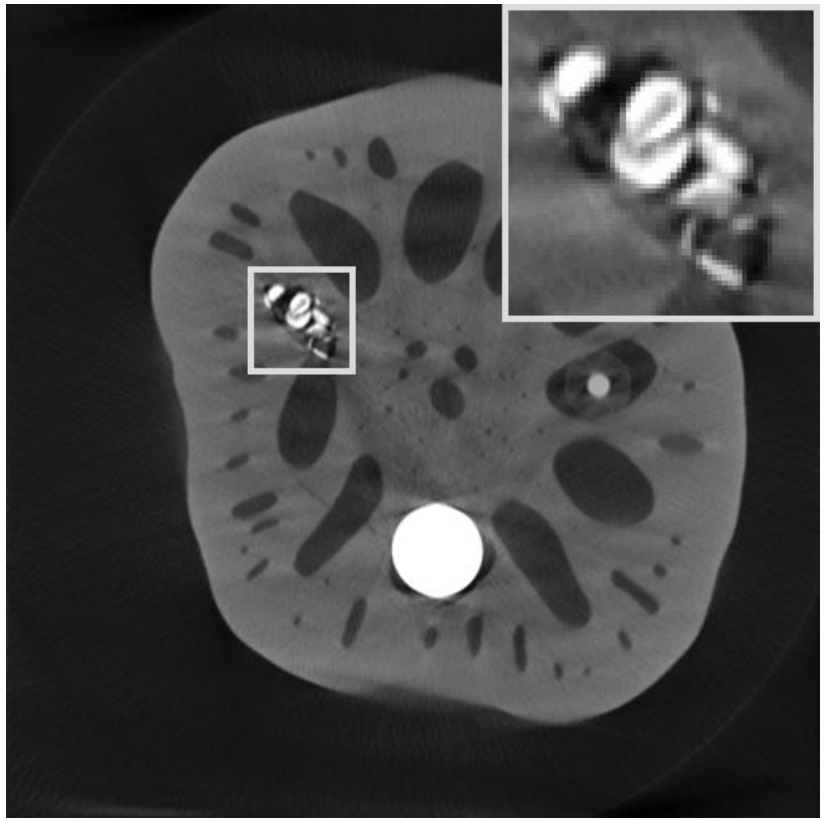


f_{TV}: RE $= 0.12$, HaarPSI$=0.74$

Generalization to Lotus Root

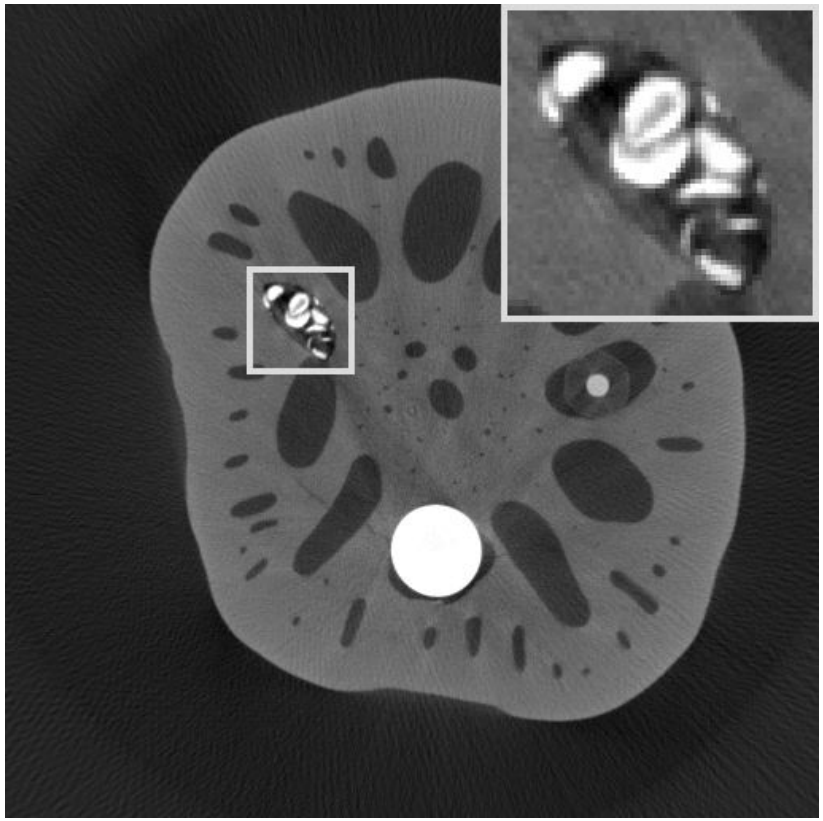


f_{gt}

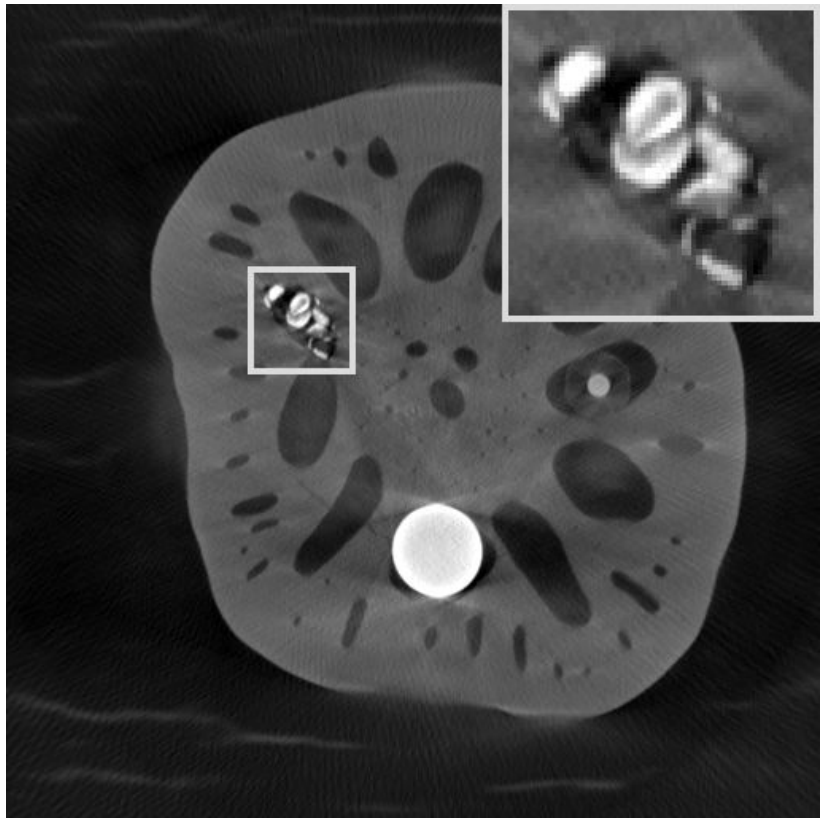


f^*: RE $= 0.11$, HaarPSI$=0.75$

Generalization to Lotus Root

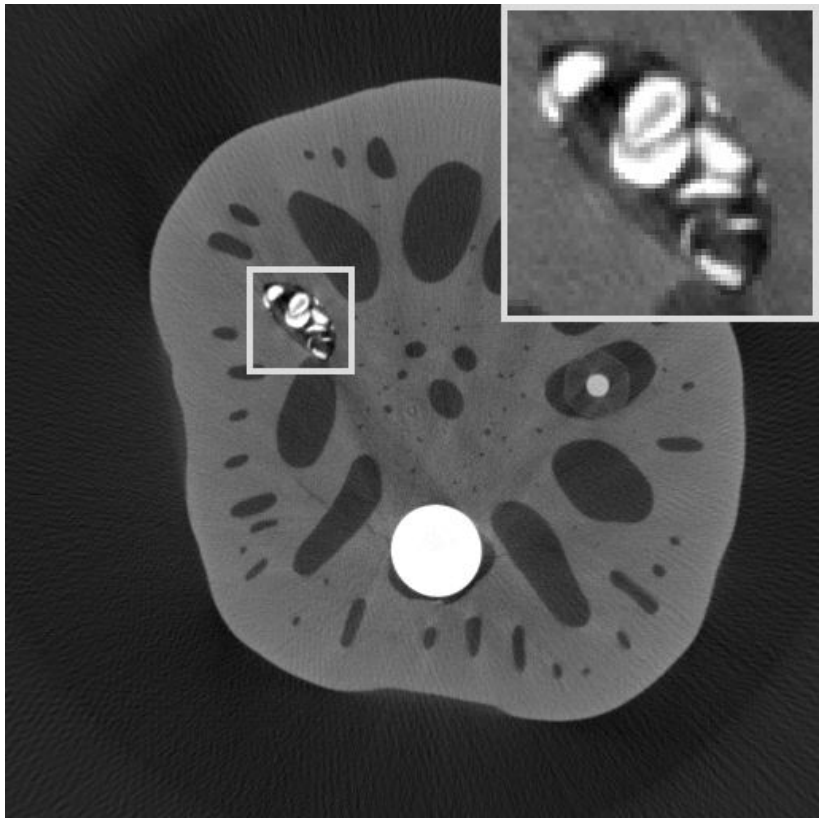


f_{gt}

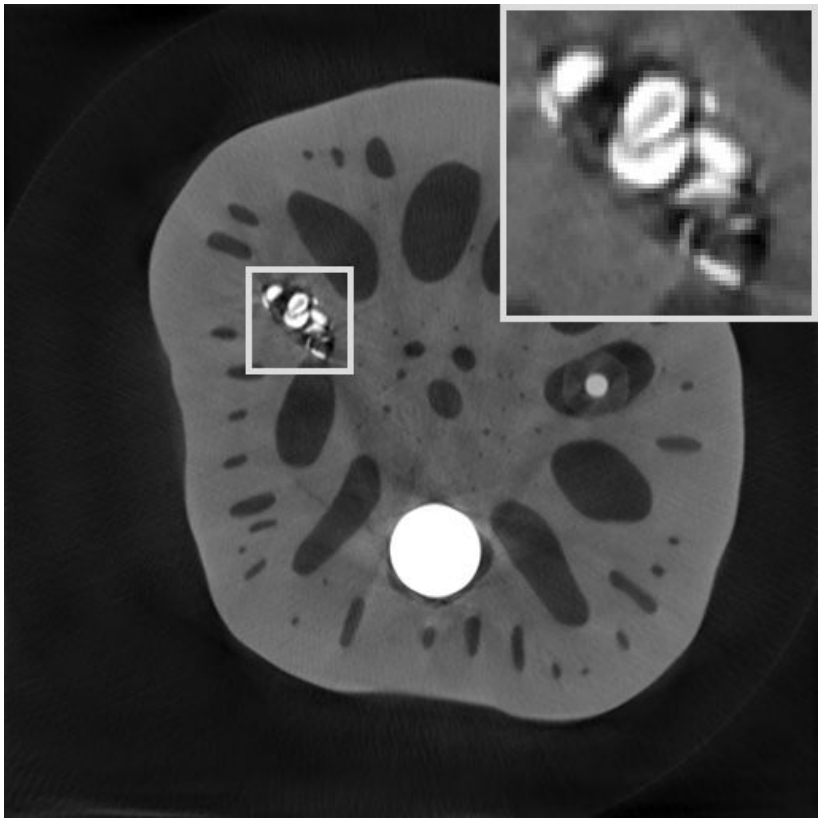


$f_{[\text{Gu \& Ye, 2017}]}$: RE = 0.25, HaarPSI=0.62

Generalization to Lotus Root



f_{gt}



$f_{\texttt{LtI}}$: 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)!*

THANK YOU!

References available at:

www.math.tu-berlin.de/∼kutyniok

Code available at:

www.ShearLab.org

Open positions in mathematics of data science at TU Berlin:

- *MATH$^+$ Junior Research Group Leader (starting January 1, 2020)*
- *Assistant Professor (starting 2020)*

If you are interested, please contact me!