Deep Learning and Modeling: Taking the Best out of Both Worlds

Gitta Kutyniok
(Technische Universität Berlin)

IPAM Workshop "Deep Learning and Medical Applications"
IPAM, Los Angeles, January 27 – 31, 2020

The 21st Century

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

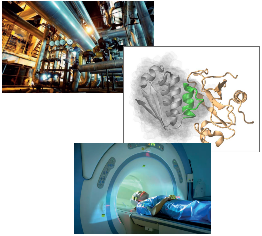
The 21st Century

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



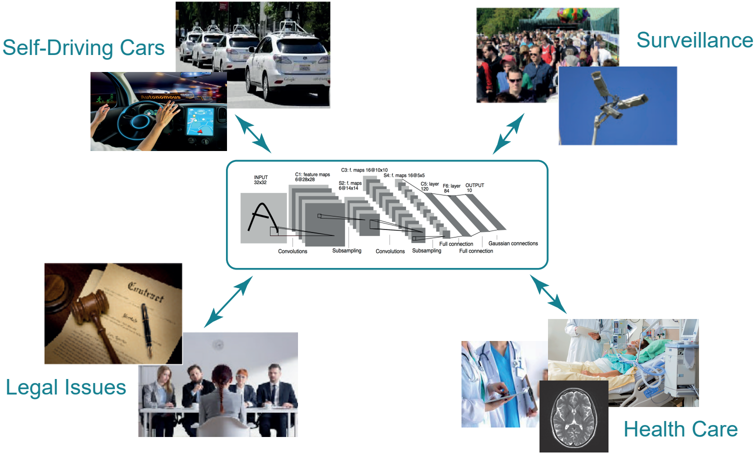
Further Examples:

- Gas networks
  ⇝ *Modeling of gigantic control systems*
- Atomistic molecular dynamics
  ⇝ *Simulations with ultralong timescales*
- Medical imaging
  ⇝ *Recovery from distorted data sets*



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

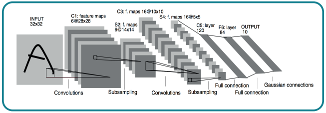
Impact of Deep Learning (Artificial Intelligence)



Impact of Deep Learning (Artificial Intelligence)

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:

- 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. Imaging Science
   - Sparse Regularization of Inverse Problems
   - Optimality of Shearlet Systems
   - A Microlocal Viewpoint: Wavefront Sets
2. Deep Neural Networks and Inverse Problems
   - A Mathematical Viewpoint
   - Conceptual Approaches
3. Taking the Best out of Both Worlds
   - (Limited-Angle) Computed Tomography
   - LtI: Learning the Invisible
   - DeNSE: Deep Network Shearlet Edge Extractor
4. Conclusions

Solving Inverse Problems

Tikhonov Regularization:
Given an (ill-posed) inverse problem

$$Kf = g, \qquad \text{where} \quad K : X \to Y,$$

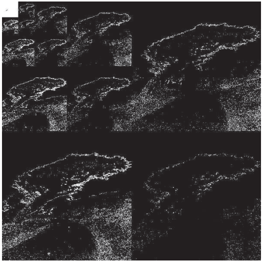
an approximate solution $f^\alpha \in X$, $\alpha > 0$, can be determined by

$$f^\alpha := \operatorname{argmin}_{f \in X} \Big[\underbrace{\|Kf - g\|^2}_{\text{Data fidelity term}} + \alpha \cdot \underbrace{\mathcal{P}(f)}_{\text{Penalty term}} \Big].$$

Penalty Term: The penalty term $\mathcal{P}$

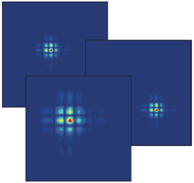
- ensures continuous dependence on the data,
- incorporates properties of the solution.

The World is Compressible!



Wavelet Transform (JPEG2000):

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

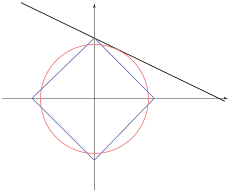


Definition: For a wavelet $\psi \in L^2(\mathbb{R}^2)$, a wavelet system is defined by

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

How to Penalize Non-Sparsity?

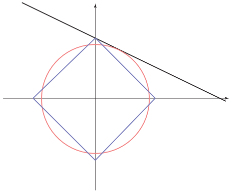
Intuition:



$\rightsquigarrow$ *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

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

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

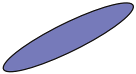
Wavelet versus Shearlet Approximation:

Key Ideas of the Shearlet Construction

Wavelet versus Shearlet Approximation:

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

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



Orientation via shearing:

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

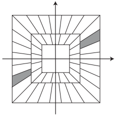
Advantage:

- 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):

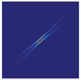
The (cone-adapted) discrete shearlet system $\mathcal{SH}(c;\phi,\psi,\tilde{\psi})$, $c>0$, generated by $\phi \in L^2(\mathbb{R}^2)$ and $\psi,\tilde{\psi} \in L^2(\mathbb{R}^2)$ is the union of



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

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

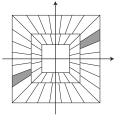
$$\{2^{3j/4}\tilde{\psi}(\tilde{S}_k \tilde{A}_{2^j}\,\cdot\, - cm) : j \geq 0, |k| \leq \lceil 2^{j/2} \rceil, m \in \mathbb{Z}^2\}.$$



(Cone-adapted) Discrete Shearlet Systems

Definition (K, Labate; 2006):

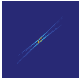
The (cone-adapted) discrete shearlet system $\mathcal{SH}(c;\phi,\psi,\tilde{\psi})$, $c > 0$, generated by $\phi \in L^2(\mathbb{R}^2)$ and $\psi, \tilde{\psi} \in L^2(\mathbb{R}^2)$ is the union of



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

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

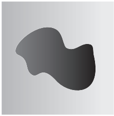
$$\{2^{3j/4}\tilde{\psi}(\tilde{S}_k \tilde{A}_{2^j} \cdot - cm) : j \geq 0, |k| \leq \lceil 2^{j/2} \rceil, m \in \mathbb{Z}^2\}.$$



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.,

$$\|f - f_N\|_2 \leq C \cdot N^{-1} \cdot (\log N)^{3/2}, \quad N \to \infty.$$



A Microlocal Viewpoint

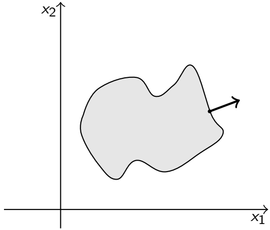
Considering edge-structures together with their direction!

A Microlocal Viewpoint

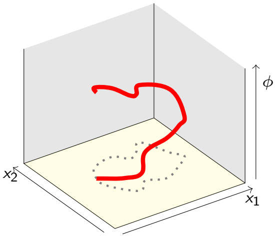
Considering edge-structures together with their direction!

Wavefront Sets:

- Notion for singularities and their direction.
- The direction indicates the propagation of the singularity.



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



Visualization in phase space

A Microlocal Viewpoint

Considering edge-structures independent of resolution!

Defnition:

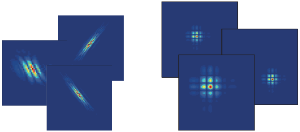
- Let $f \in L^2(\mathbb{R}^2)$ and $k \in \mathbb{N}$. A point $(x, \lambda) \in \mathbb{R}^2 \times \mathbb{S}^1$ is called k-regular directed point of f if there exist open neighbourhoods U_x, U_λ of x and λ and a smooth function $\phi \in C^\infty(\mathbb{R}^2)$ with supp $\phi \subset U_x$, $\phi(x) = 1$ such that there is $C_k > 0$ and

$$|\widehat{\phi f}(\xi)| \leq C_k(1 + |\xi|)^{-k}, \text{ for all } \xi \in \mathbb{R}^2 \setminus \{0\} \text{ such that } \xi/|\xi| \in V_\lambda.$$

- The complement of the set of all k-regular directed points is called the *k-wavefront set* denoted by $WF_k(f)$.
- $WF(f)$ is defined as the complement of the set of all points (x, λ), which are k-regular directed points for all k.

Shearlets and Wavefront Sets

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])

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

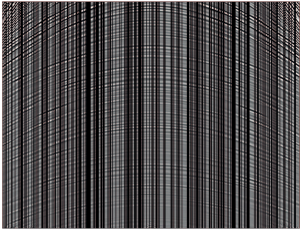
Summary

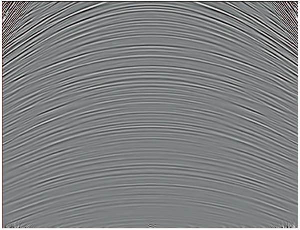
Shearlets are a representation system which...

- ...is generated by one or a few 'mother functions',
- ...provides optimally sparse approximation of cartoons,
- ...precisely resolves the wavefront set,
- ...allows for compactly supported analyzing elements,
- ...is associated with fast decomposition algorithms,
- ...treats the continuum and digital 'world' uniformly.

Applications

Inpainting:

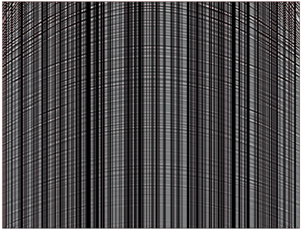


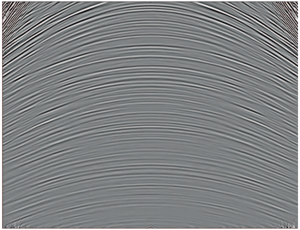


(Source: K, Lim; 2012)

Applications

Inpainting:





(Source: K, Lim; 2012)

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

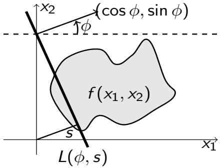
- Matlab *(K, Lim, Reisenhofer; 2013)*
- Julia *(Loarca; 2017)*
- Python *(Look; 2018)*
- Tensorflow *(Loarca; 2019)*

Mathematical Modeling Reaches a Barrier

Limited Angle-(Computed) Tomography

A CT scanner samples the *Radon transform*

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

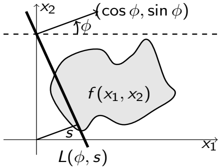


for $L(\phi, s) = \{x \in \mathbb{R}^2 : x_1 \cos(\phi) + x_2 \sin(\phi) = s\}$, $\phi \in [-\pi/2, \pi/2)$, and $s \in \mathbb{R}$.

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

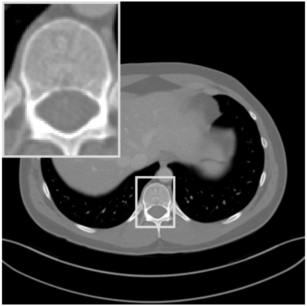
Applications: Dental CT, breast tomosynthesis, electron tomography,...

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:



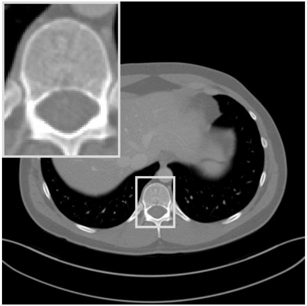
Original Image

Model-Based Approaches Fail

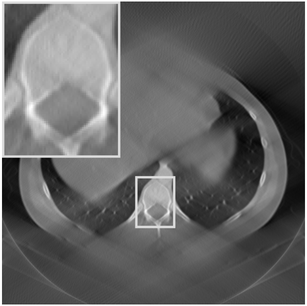
Sparse Regularization:

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

Clinical Data:



Original Image



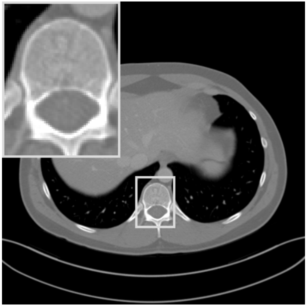
Filtered Backprojection

Model-Based Approaches Fail

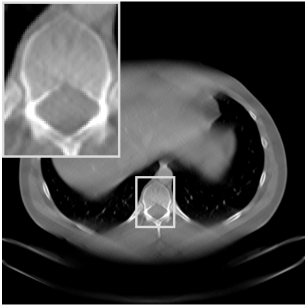
Sparse Regularization:

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

Clinical Data:



Original Image



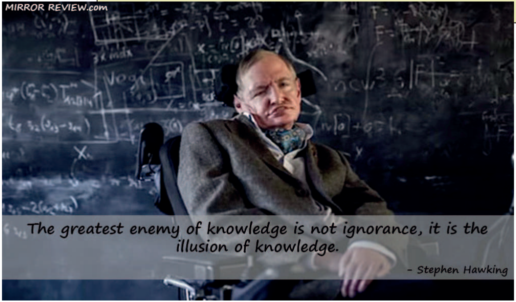
Sparse Regularization with Shearlets

Let's bring Deep Learning into the Game

Deep Learning = Alchemy?

„Ali Rahimi, a researcher in artificial intelligence (AI) at Google in San Francisco, California, took a swipe at his field last December—and received a 40-second ovation for it. Speaking at an AI conference, Rahimi charged that machine learning algorithms, in which computers learn through trial and error, have become a form of „alchemy." Researchers, he said, do not know why some algorithms work and others don't, nor do they have rigorous criteria for choosing one AI architecture over another....“

Science, May 2018



Fundamental Questions concerning Deep Neural Networks

- *Expressivity:*
  - How powerful is the network architecture?
  - Can it indeed represent the correct functions?

$\rightsquigarrow$ *Applied Harmonic Analysis, Approximation Theory, ...*

Fundamental Questions concerning Deep Neural Networks

- *Expressivity:*
  - How powerful is the network architecture?
  - Can it indeed represent the correct functions?

$\rightsquigarrow$ *Applied Harmonic Analysis, Approximation Theory, ...*

- *Learning:*
  - Why does the current learning algorithm produce anything reasonable?
  - What are good starting values?

$\rightsquigarrow$ *Differential Geometry, Optimal Control, Optimization, ...*

Fundamental Questions concerning Deep Neural Networks

- *Expressivity:*
  - How powerful is the network architecture?
  - Can it indeed represent the correct functions?

$\rightsquigarrow$ *Applied Harmonic Analysis, Approximation Theory, ...*

- *Learning:*
  - Why does the current learning algorithm produce anything reasonable?
  - What are good starting values?

$\rightsquigarrow$ *Differential Geometry, Optimal Control, Optimization, ...*

- *Generalization:*
  - Why do deep neural networks perform that well on data sets, which do not belong to the input-output pairs from a training set?
  - What impact has the depth of the network?

$\rightsquigarrow$ *Learning Theory, Optimization, Statistics, ...*

Fundamental Questions concerning Deep Neural Networks

- *Expressivity:*
  - How powerful is the network architecture?
  - Can it indeed represent the correct functions?

$\rightsquigarrow$ *Applied Harmonic Analysis, Approximation Theory, ...*

- *Learning:*
  - Why does the current learning algorithm produce anything reasonable?
  - What are good starting values?

$\rightsquigarrow$ *Differential Geometry, Optimal Control, Optimization, ...*

- *Generalization:*
  - Why do deep neural networks perform that well on data sets, which do not belong to the input-output pairs from a training set?
  - What impact has the depth of the network?

$\rightsquigarrow$ *Learning Theory, Optimization, Statistics, ...*

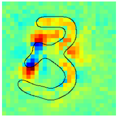
- *Interpretability:*
  - Why did a trained deep neural network reach a certain decision?
  - Which components of the input do contribute most?

$\rightsquigarrow$ *Information Theory, Uncertainty Quantification, ...*

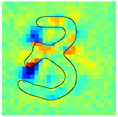
Interpretability of Deep Neural Networks

Main Goal: We aim to understand decisions of "black-box" predictors!

map for digit 3

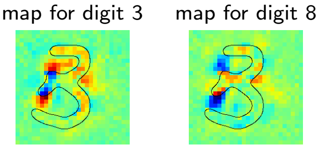


map for digit 8



Interpretability of Deep Neural Networks

Main Goal: We aim to understand decisions of “black-box” predictors!

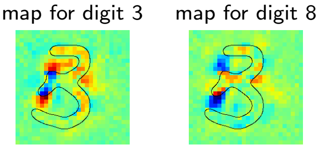


Some Previous Relevance Mapping Methods:

- Gradient based methods:
  - Sensitivity Analysis (Simonyan, Vedaldi, Zisserman, 2013)
  - SmoothGrad (Smilkov, Thorat, Kim, Viégas, Wattenberg, 2017)
- Backwards propagation based methods:
  - Guided Backprop (Springenberg, Dosovitskiy, Brox, Riedmiller, 2015)
  - Layer-wise Relevance Propagation (Bach, Binder, Montavon, Klauschen, Müller, Samek, 2015)
  - Deep Taylor (Montavon, Samek, Müller, 2018)

Interpretability of Deep Neural Networks

Main Goal: We aim to understand decisions of "black-box" predictors!



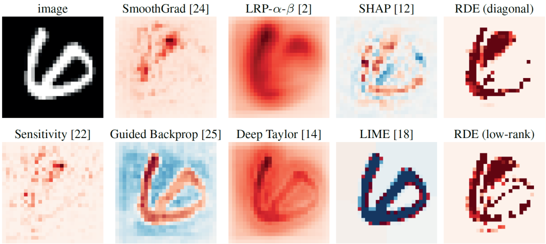
Rate-Distortion Explanation (MacDonald, Wäldchen, Hauch, K; 2019):

- Rigorous definition of relevance by information theory.
  $\rightsquigarrow$ *Regarding relevant pixels as key information to transmit.*
- Formulation of interpretability as optimization problem.
- Theoretical analysis of complexity.

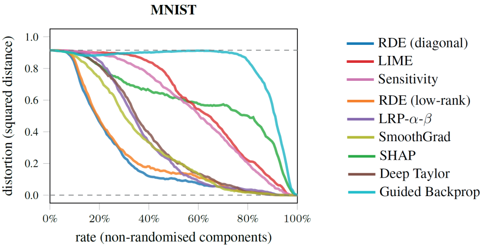
For details see the poster outside!

Numerical Experiments on MNIST

Classification of the Digit 6:



Quality Measure:



Deep Neural Networks and Inverse Problems

Typical Deep Learning Approaches to Inverse Problems

Denoising Direct Inversion [Kang,Min,Ye,2017], [Unser et. al.,2017], [Antholzer et al.,2019]

- Idea: Direct inversion, e.g. with filtered backprojection, then train CNN to remove (structured) noise and artefacts.

Typical Deep Learning Approaches to Inverse Problems

Denoising Direct Inversion [Kang,Min,Ye,2017], [Unser et. al.,2017], [Antholzer et al.,2019]

- Idea: Direct inversion, e.g. with filtered backprojection, then train CNN to remove (structured) noise and artefacts.

Plug-and-play with CNN-denoising [Venkatakrishnan,Bouman,Wohlberg,2013], [Romano,Elad,Milanfar,2016], [Meinhardt et al.,2017], [Reehorst,Schniter,2019]

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

Typical Deep Learning Approaches to Inverse Problems

Denoising Direct Inversion [Kang,Min,Ye,2017], [Unser et. al.,2017], [Antholzer et al.,2019]

- Idea: Direct inversion, e.g. with filtered backprojection, then train CNN to remove (structured) noise and artefacts.

Plug-and-play with CNN-denoising [Venkatakrishnan,Bouman,Wohlberg,2013], [Romano,Elad,Milanfar,2016], [Meinhardt et al.,2017], [Reehorst,Schniter,2019]

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

Learned Iterative Schemes [Gregor,LeCun,2010], [Yang et al.,2016], [Hammernick et al.,2016] [Adler,Öktem,2017], [Hammernick et al.,2018], [Hauptmann et al.,2018]

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

Typical Deep Learning Approaches to Inverse Problems

Denoising Direct Inversion [Kang,Min,Ye,2017], [Unser et. al.,2017], [Antholzer et al.,2019]

- Idea: Direct inversion, e.g. with filtered backprojection, then train CNN to remove (structured) noise and artefacts.

Plug-and-play with CNN-denoising [Venkatakrishnan,Bouman,Wohlberg,2013], [Romano,Elad,Milanfar,2016], [Meinhardt et al.,2017], [Reehorst,Schniter,2019]

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

Learned Iterative Schemes [Gregor,LeCun,2010], [Yang et al.,2016], [Hammernick et al.,2016] [Adler,Öktem,2017], [Hammernick et al.,2018], [Hauptmann et al.,2018]

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

Generative Models Priors [Bora et al.,2017], [Mixon,Villar,2018], [Hand,Voroninski,2018], [Wei,Yang,Wang,2019], [Shah,Hegde,2019], [Xu,Zeng,Romberg,2019]

- Solve $\min_{z\in\mathbb{R}^k} \|AG(z) - y\|_2^2$, where G is a generative model (e.g. GAN).

Taking a Microlocal Viewpoint

General Mission Statement:

- Employ model-based approaches as far as they are reliable.
- Apply deep learning only where model-based methods fail.

Taking a Microlocal Viewpoint

General Mission Statement:

- Employ model-based approaches as far as they are reliable.
- Apply deep learning only where model-based methods fail.

Guiding Principle:

- Edges are key features of each image.
- Recovery of the wavefront set is crucial:
  - Use as prior [Davison; 1983],....
  - Reveal missing parts.
  - ...

- Apply the shearlet transform to “sense” the wavefront set.

Learning the Invisible (LtI)

joint with

Maximilian März (TU Berlin)

Wojciech Samek and Vignesh Srinivan (Fraunhofer HHI Berlin)

Tatiana Bubba, Matti Lassas, and Samuli Siltanen (University of Helsinki)

A Related Deep Learning Approach to Limited-Angle CT

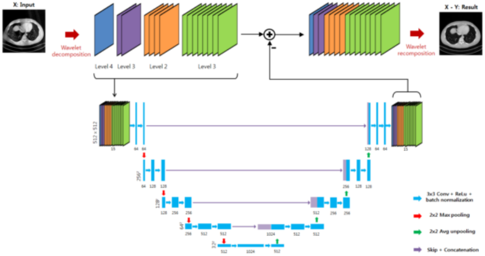


Image source: [Gu & Ye, 2017]:

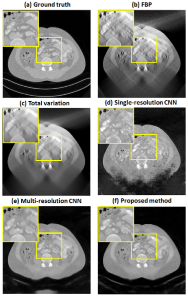
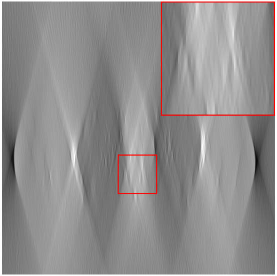


Image source: [Gu & Ye, 2017]:

- 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)

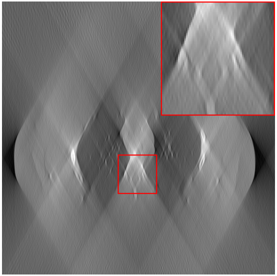
[J. Gu and J. C. Ye. Multi-scale wavelet domain residual learning for limited-angle CT reconstruction. In: Procs Fully3D (2017), pp. 443447.]

Zooming in on the Recovery Problem



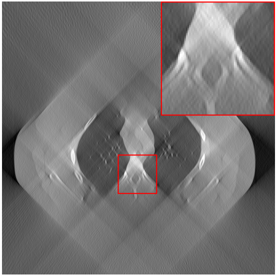
$\phi = 15^\circ$, filtered backprojection (FBP)

Zooming in on the Recovery Problem



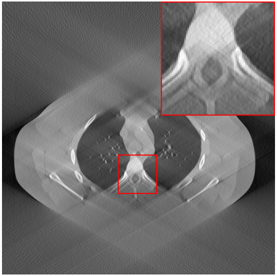
$\phi = 30^\circ$, filtered backprojection (FBP)

Zooming in on the Recovery Problem



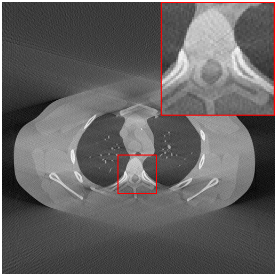
$\phi = 45^\circ$, filtered backprojection (FBP)

Zooming in on the Recovery Problem



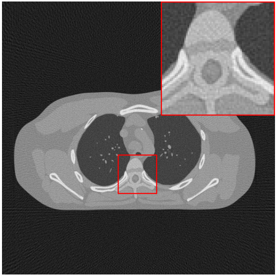
$\phi = 60^\circ$, filtered backprojection (FBP)

Zooming in on the Recovery Problem



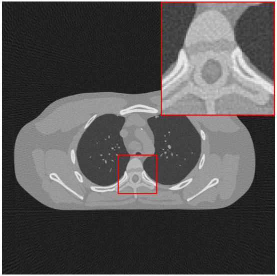
$\phi = 75^\circ$, filtered backprojection (FBP)

Zooming in on the Recovery Problem



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

Zooming in on the Recovery Problem



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

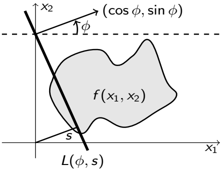
Some Observations:

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

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).

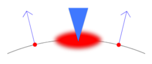


"visible": singularities tangent to sampled lines

"invisible": singularities not tangent to sampled lines

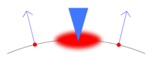
Shearlets can Help

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



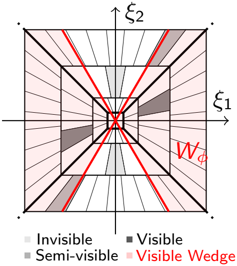
Shearlets can Help

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



The Shearlet Transform:

- Shearlets can identify the wavefront set at fine scales.
- Shearlets can separate the visible and invisible part.



Models versus Data

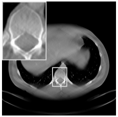
The High-level Idea:

- 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)"

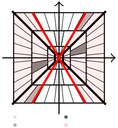
(Bubba, K, Lassas, März, Samek, Siltanen, Srinivan; 2018)

Step 1: *Reconstruct the visible*



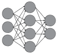
$$f^* := \mathrm{argmin}_{f\geq 0} \| \mathcal{R}_\phi f - g \|_2^2 + \| \mathrm{SH}_\psi(f) \|_{1,w}$$

- 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*

$$\mathcal{N}\mathcal{N}_\theta : \ \mathrm{SH}_\psi(f^*)_{\mathcal{I}_{\mathtt{vis}}} \ \text{———}$$

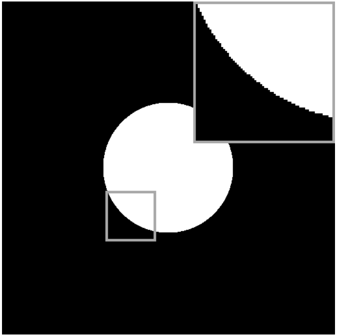


$$\longrightarrow F \ \left(\overset{!}{\approx} \mathrm{SH}_\psi(f_{\mathrm{gt}})_{\mathcal{I}_{\mathtt{inv}}} \right)$$

Step 3: *Combine*

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

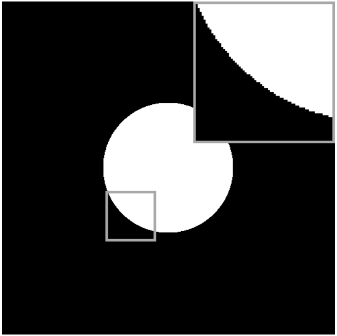
Verify the concept of (in-)visibility



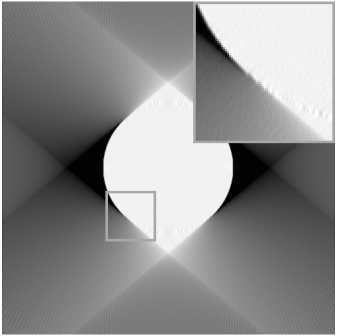
f_{gt}

Numerical Simulation

Verify the concept of (in-)visibility



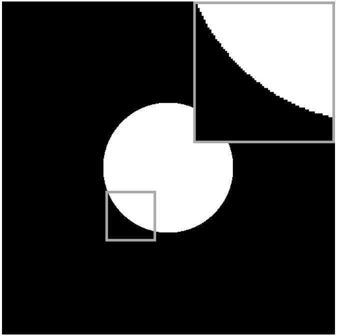
f_{gt}



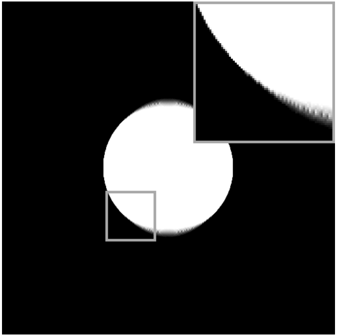
FBP

Numerical Simulation

Verify the concept of (in-)visibility



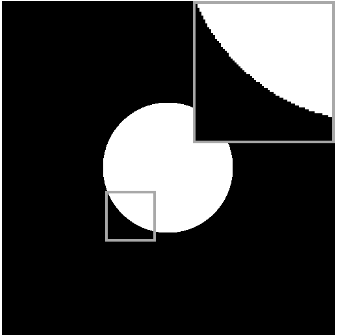
f_{gt}



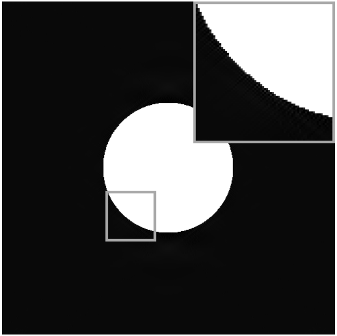
ℓ_1-analysis shearlet solution f^*

Numerical Simulation

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



$f_{\mathtt{gt}}$

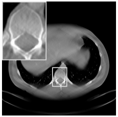


$\mathrm{SH}_\psi^T\left(\mathrm{SH}_\psi(f^*)_{\mathcal{I}_{\mathtt{vis}}} + \mathrm{SH}_\psi(f_{\mathtt{gt}})_{\mathcal{I}_{\mathtt{inv}}}\right)$

Our Approach "Learn the Invisible (LtI)"

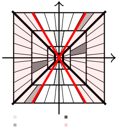
(Bubba, K, Lassas, März, Samek, Siltanen, Srinivan; 2018)

Step 1: *Reconstruct the visible*

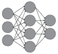


$$f^* := \mathrm{argmin}_{f \geq 0} \| \mathcal{R}_\phi f - g \|_2^2 + \| \mathrm{SH}_\psi(f) \|_{1,w}$$

- 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*

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

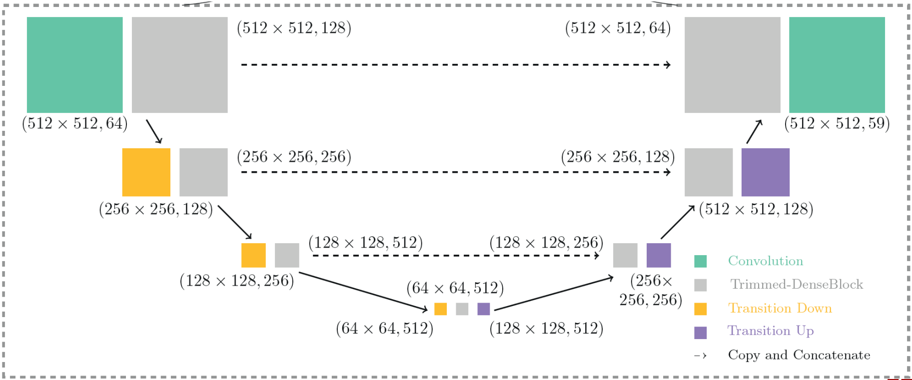
Step 3: *Combine*

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

Our Approach – Step 2: PhantomNet

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

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



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

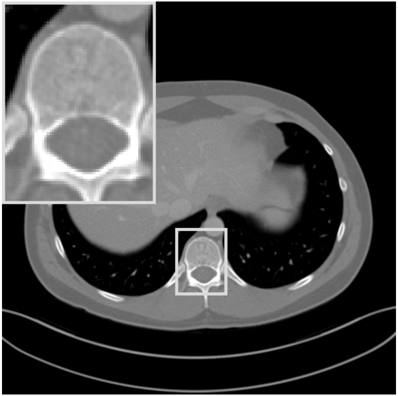
Setup

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
- ...

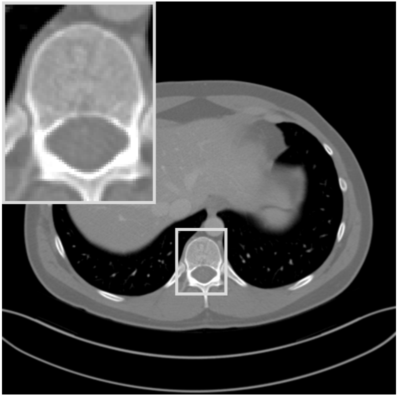
[1]We would like to thank Dr. Cynthia McCollough, the Mayo Clinic, the American Association of Physicists in Medicine (AAPM), and grant EB01705 and EB01785 from the National Institute of Biomedical Imaging and Bioengineering for providing the Low-Dose CT Grand Challenge data set.

Evaluation on Test Patient

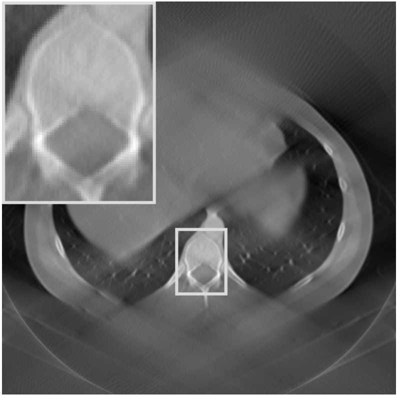


f_{gt}

Evaluation on Test Patient

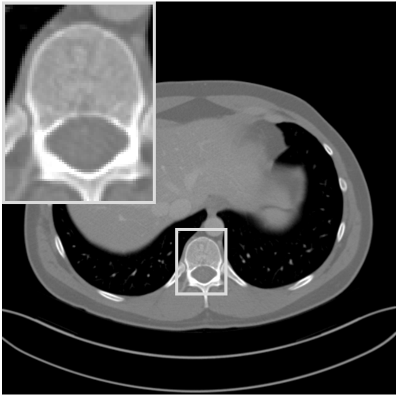


f_{gt}

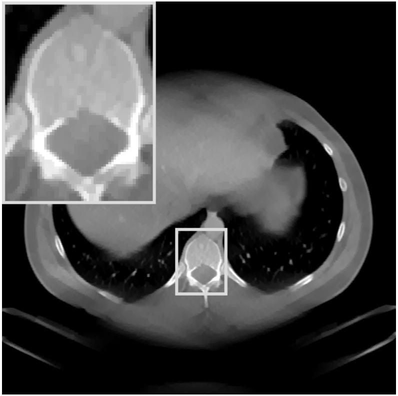


f_{FBP}: RE = 0.50, HaarPSI=0.35

Evaluation on Test Patient

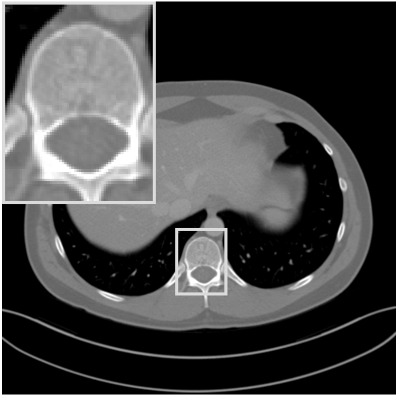


f_{gt}

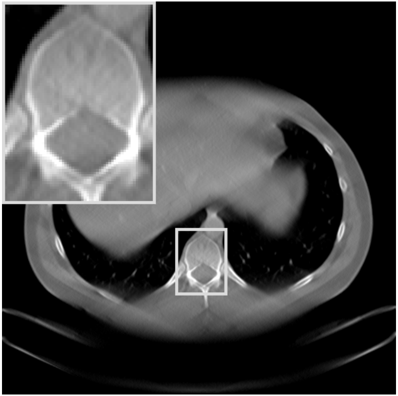


f_{TV}: RE $= 0.21$, HaarPSI$=0.41$

Evaluation on Test Patient

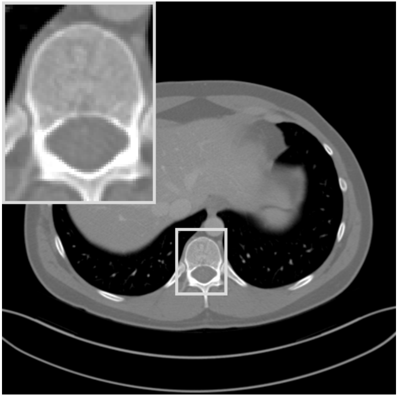


f_{gt}



f^*: RE = 0.19, HaarPSI=0.43

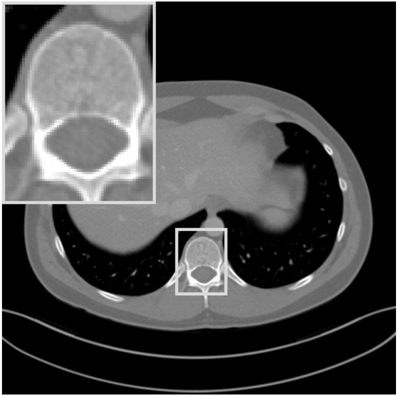
Evaluation on Test Patient



f_{gt}

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

Evaluation on Test Patient



$f_{\mathtt{gt}}$

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

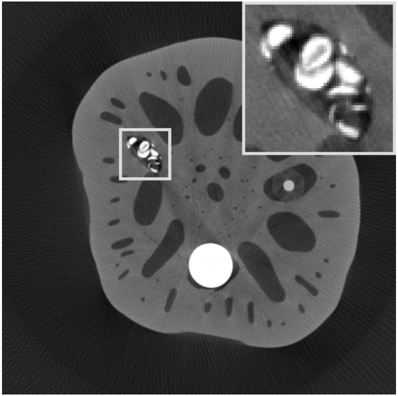
HaarPSI (Reisenhofer, Bosse, K, and Wiegand; 2018)

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

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

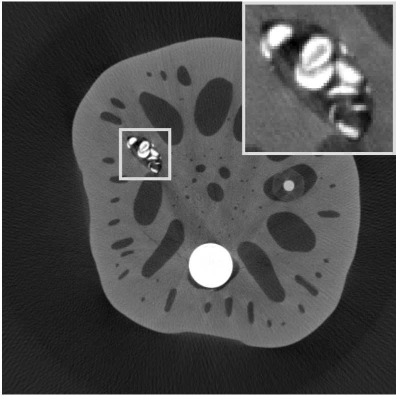
www.haarpsi.org

Generalization to Lotus Root

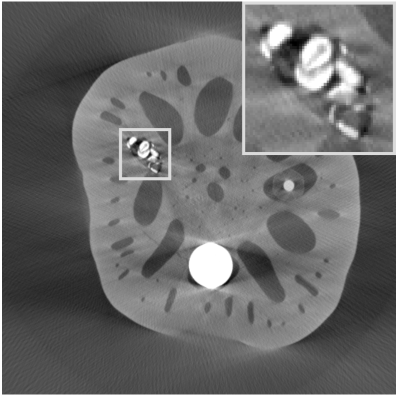


f_{gt}

Generalization to Lotus Root

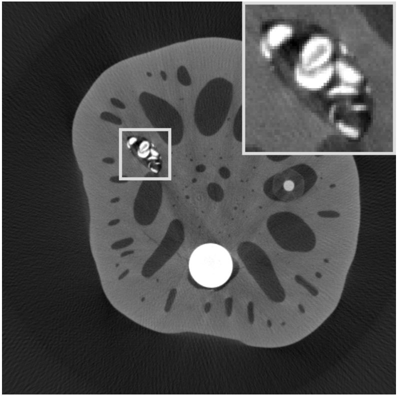


f_{gt}

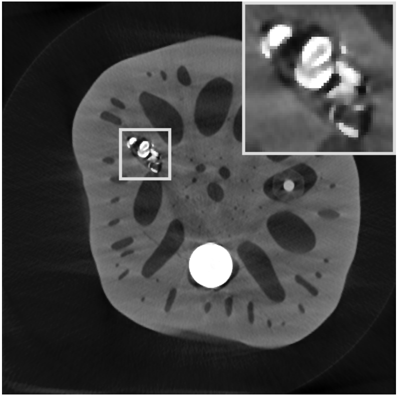


f_{FBP}: RE $= 0.31$, HaarPSI$=0.61$

Generalization to Lotus Root

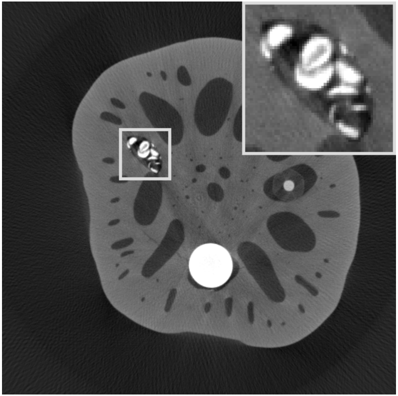


f_{gt}

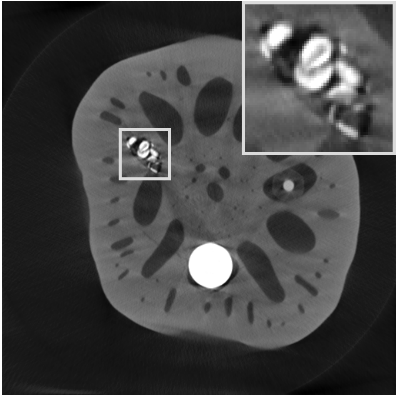


f_{TV}: RE $= 0.12$, HaarPSI$=0.74$

Generalization to Lotus Root

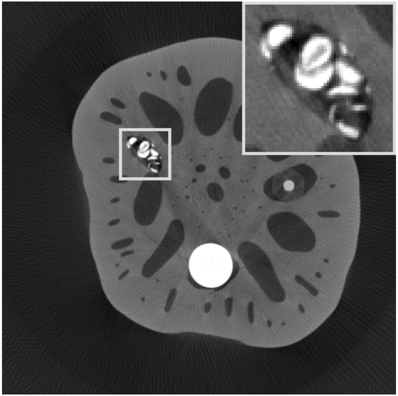


f_{gt}

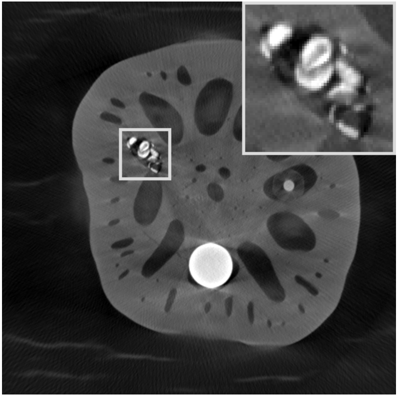


f^*: RE $= 0.11$, HaarPSI$=0.75$

Generalization to Lotus Root

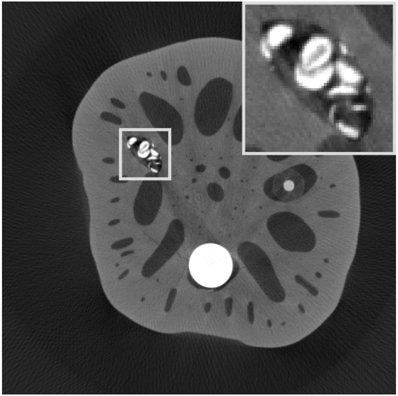


f_{gt}

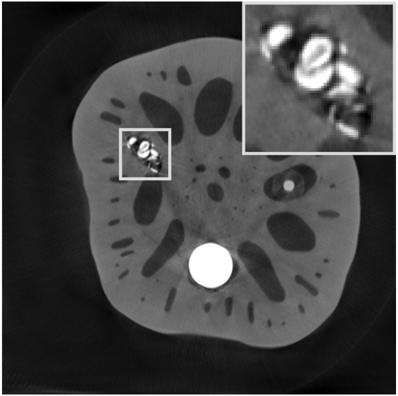


$f_{\text{[Gu \& Ye, 2017]}}$: RE $= 0.25$, HaarPSI$=0.62$

Generalization to Lotus Root



f_{gt}



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

Deep Network Shearlet Edge Extractor (DeNSE)

joint with

Hector Andrade-Loarca (TU Berlin)

Ozan Öktem (KTH Royal Institute of Technology)

Philipp Petersen (University of Vienna)

Computed Tomography and Wavefront Sets

Course of Action:

1. Detect the wavefront set of the sinogram.
2. Apply the inverse canonical relation.
3. Use the wavefront set of the reconstructed image as prior.

Canonical Relation:

The canonical relation C satisfies

$$\mathrm{WF}\big(R(f)\big) = C \circ \mathrm{WF}(f) \quad \text{whenever } f \in \mathcal{D}'(\mathbb{R}^2),$$

where

$$\begin{aligned} C \;=\; & \Big\{\big(\theta, p, s(-x \cdot \omega(\theta)^{\perp}\mathrm{d}\theta + \mathrm{d}p); x, s\omega(\theta)\mathrm{d}x\big) \in T^*(\mathbb{M}) : \\ & (\theta, p) \in \mathbb{M}, x \in \mathbb{R}^2, s \neq 0, x \cdot \omega(\theta) = p\Big\}. \end{aligned}$$

Deep Network Shearlet Edge Extractor (DeNSE)

(Andrade-Loarca, K, Öktem, Petersen; 2019)

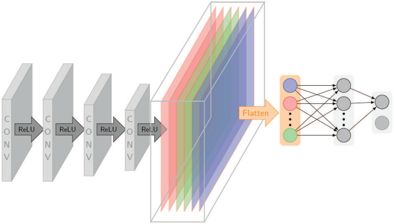
Key Steps:

(1) Apply the shearlet transform to an image.
   $\rightsquigarrow$ Extract the correct features.
   $\rightsquigarrow$ Derive a good data representation.

(2) Consider patches of shearlet coefficients.
   $\rightsquigarrow$ Localize to each position.

(3) Apply a convolutional neural network.
   $\rightsquigarrow$ Predict the direction (180 directions) in each patch.

Network Architecture:

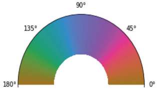


Training of the ConvNet

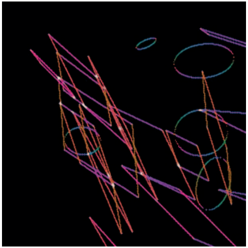
Training Data: Patches of the shearlet transform of images made of...

1. ...random sums of ellipses and parallelograms of different contrasts, sizes, and orientations.
2. ...random sums of ellipses and parallelograms, but convolves these images with a kernel to generate a function with a higher-order wavefront set.
3. ...the BSDS500 (Berkeley Segmentation data set) provided by the Computer Vision Group of UC Berkeley. It comprises 503 natural images of different types.
4. ...the Semantic Boundaries data set (SBD) with 11355 natural images, again provided by the Computer Vision Group of UC Berkeley.

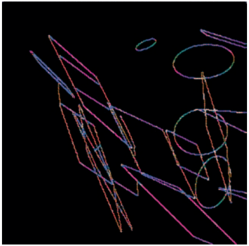
Ellipses and Parallelograms



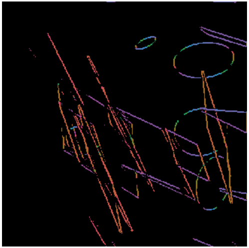
Original



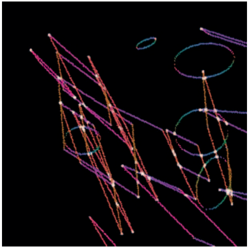
Human Annotation



[Yi, Labate, Easley, Krim; 2009]

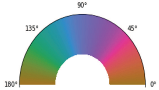


CoShREM [Reisenhofer et al.; 2015]

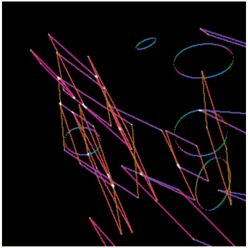


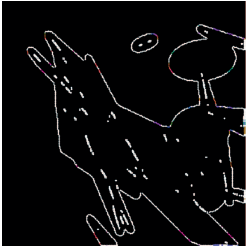
DeNSE

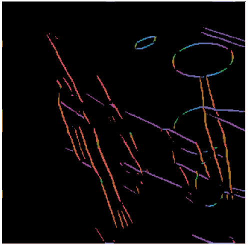
Smoothed Ellipses and Parallelograms

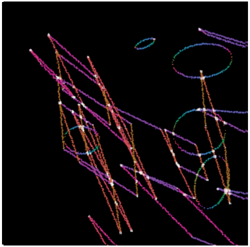



Original


Human Annotation


[Yi, Labate, Easley, Krim; 2009]


CoShREM [Reisenhofer et al.; 2015]


DeNSE

Comparison Results

Comparison for Ellipses/Parallelograms:

Method	MF-score
Canny	49.1
Sobel	40.0
BEL	63.3
Yi-Labate-Easley-Krim	70.3
CoShREM	90.6
DEnSE	**97.5**

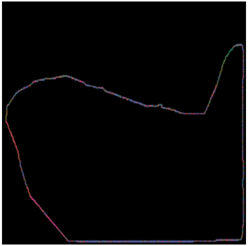
MF-Score:

$$F := \frac{2PR}{R+P}, \quad \text{where}$$

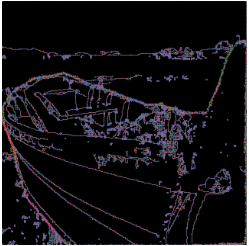
- P is the precision, i.e., the number of true positives divided by the sum of true and false positives,
- R is the recall, i.e., the number of true positives divided by the sum of true positives and false negatives.

BSDS500 Data Set

Original

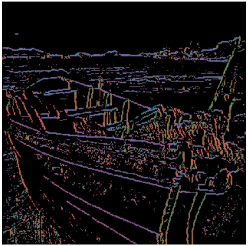


Human Annotation



SEAL [Yu et al; 2018]

CoShREM [Reisenhofer et al.; 2015]



DeNSE

Comparison Results

Comparison for BSDS500 Data Set:

Method	MF-score
gPb-owt-ucm	73.7
gPb	71.5
Mean Shift - Comaniciu, Meer	64.0
Normalized Cuts - Cour, Benezit, Shi	64.2
Fetzenszwalb, Huttenlocher	61.0
Canny	60.3
CoShREM	75.7
DeepEdge	75.3
DEnSE	**95.4**

Computed Tomography and Wavefront Sets

Course of Action:

1. Detect the wavefront set of the sinogram.
2. Apply the inverse canonical relation.
3. Use the wavefront set of the reconstructed image as prior.

Canonical Relation:

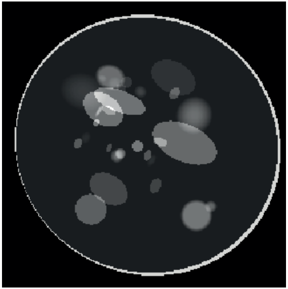
The canonical relation C satisfies

$$\mathrm{WF}\big(R(f)\big) = C \circ \mathrm{WF}(f) \quad \text{whenever } f \in \mathcal{D}'(\mathbb{R}^2),$$

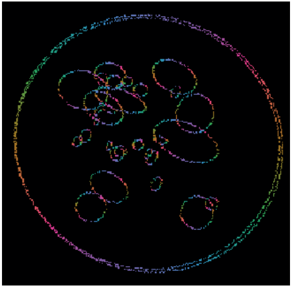
where

$$C = \Big\{\big(\theta, p, s(-x \cdot \omega(\theta)^{\perp}\mathrm{d}\theta + \mathrm{d}p); x, s\omega(\theta)\mathrm{d}x\big) \in T^*(\mathbb{M}) : (\theta, p) \in \mathbb{M}, x \in \mathbb{R}^2, s \neq 0, x \cdot \omega(\theta) = p\Big\}.$$

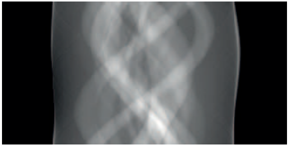
Application of the Canonical Relation



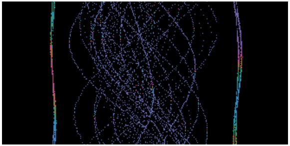
Phantom



Wavefront Set by DeNSE

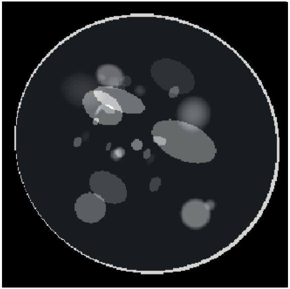


Sinogram

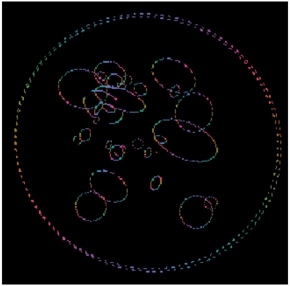


Wavefront Set by Canonical Relation

Application of the Inverse Canonical Relation

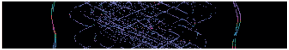


Phantom



Wavefront Set by Inverse Canonical Relation

Low-Dose Sinogram



Wavefront Set by DeNSE

Comparison Results

Comparison for Application of Inverse Canonical Relation:

Inversion technique	**Mean square error**
Tikhonov	443.0
Total variation	380.9
Filtered backprojection	504.3
Canonical relations	168.1

Superior performance to any first-invert-then-extract strategy!

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*.

Deep Learning:

- Impressive performance for *Inverse Problems*.
- *Theoretical foundation* of neural networks *almost entirely missing*: *Expressivity, Learning, Generalization, and Interpretability*.

Combining Both Sides (Limited-Angle Tomography):

- *LtI: Learning the Invisible*
  $\rightsquigarrow$ Accessing the visible part by (sparse regularization) with shearlets.
  $\rightsquigarrow$ Learning only the invisible part.
- *DeNSE: Deep Network Shearlet Edge Extractor*
  $\rightsquigarrow$ Extracting the wavefront set by shearlets and deep learning.
  $\rightsquigarrow$ Applying the canonical relation to use the wavefront set as prior.

THANK YOU!

References available at:

`www.math.tu-berlin.de/∼kutyniok`

Code available at:

`www.ShearLab.org`

Related Books:

- Y. Eldar and G. Kutyniok
  Compressed Sensing: Theory and Applications
  Cambridge University Press, 2012.
- 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)