Deep Learning meets Physics: Taking the Best out of Both Worlds in Imaging Science

Gitta Kutyniok

(Ludwig-Maximilians-Universität München)

Zoom webinar on
Mathematics, Physics and Machine Learning
Instituto Superior Técnico, Lisbon, December 2, 2020

LUDWIG-
MAXIMILIANS-
UNIVERSITÄT
MÜNCHEN

The 21st Century

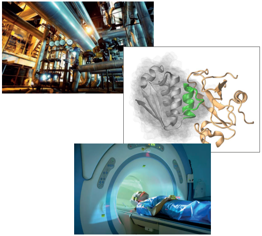
Various technological advances in the 21st century are only possible through *integrated mathematical modeling, simulation, and optimization*.

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*

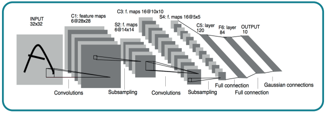


There is a pressing need to go beyond pure modeling, simulation, and optimization approaches!

Impact of Deep Learning (Artificial Intelligence)

Self-Driving Cars

Surveillance



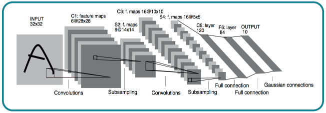
Legal Issues

Health Care

Impact of Deep Learning (Artificial Intelligence)

Self-Driving Cars

Surveillance



Legal Issues

Health Care

Very few theoretical results explaining their success!

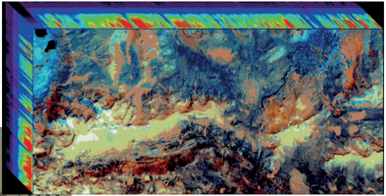
From Data-Driven to Model-Based Approaches

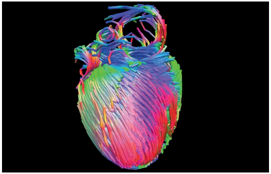
Problems, Viewpoints and Solution Strategies:

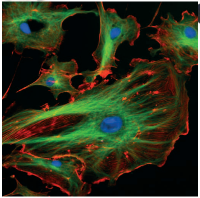
- Pure data-driven approaches.
  Detect structural components in data sets!
- Machine learning with physical constraints.
  Insert physical information in machine learning algorithm!
- Parametric differential equations.
  Learn parameters from given data sets!
- Data assimilation.
  Combine sparse data 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!

Modern Imaging Science







Inverse Problems

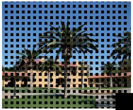
Recovering the original data from a transformed version!

Inverse Problems

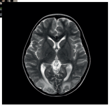
Recovering the original data from a transformed version!

Some Examples from Imaging:

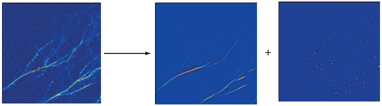
- Inpainting.
  ⇝ *Recovery from incomplete data.*



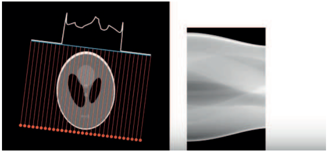
- Magnetic Resonance Imaging.
  ⇝ *Recovery from point samples of the Fourier transform.*



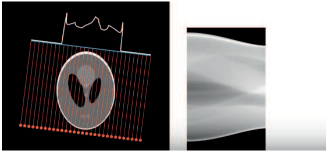
- Feature Extraction.
  ⇝ *Separating the image into different features.*



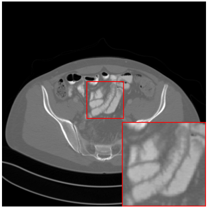
Computed Tomography (CT)

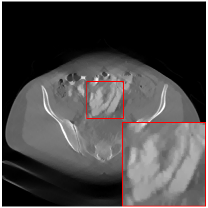


Computed Tomography (CT)

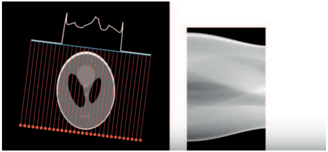


Problem with Limited-Angle Tomography:

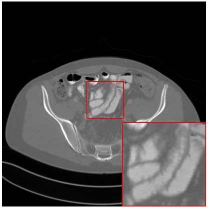


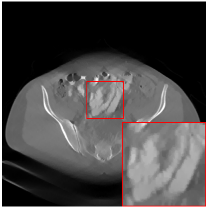


Computed Tomography (CT)



Problem with Limited-Angle Tomography:





The data is too complex for mathematical modeling!

Outline

1. Model-Based Side
   - Sparse Regularization of Inverse Problems
   - Shearlet Systems come into Play
2. Data-Based Side
   - Deep Neural Networks
   - Deep Learning and Inverse Problems
3. Taking the Best out of Both Worlds
   - General Conceptual Approach
   - LtI: Learning the Invisible
   - DeNSE: Deep Network Shearlet Edge Extractor
4. Conclusions

Model-Based Approach to Inverse Problems:

Sparse Regularization

Solving Inverse Problems

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

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

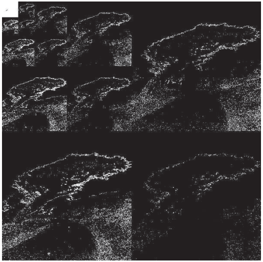
an approximate solution $f^\alpha \in X$, $\alpha > 0$, can be determined by

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

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

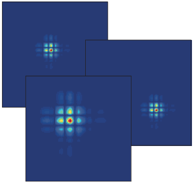
- ensures continuous dependence on the data,
- incorporates properties of the solution.

The World is Compressible!



Wavelet Transform (JPEG2000):

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

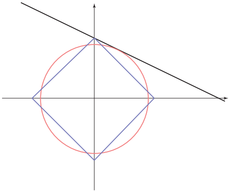


Definition: For a wavelet $\psi \in L^2(\mathbb{R}^2)$, a wavelet system is defined by

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

How to Penalize Non-Sparsity?

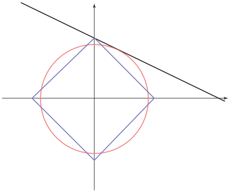
Intuition:



$\rightsquigarrow$ *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

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

Shearlets come into Play

Mathematical Model for Images

Key Observation:

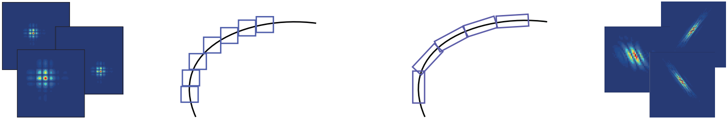
Images are governed by edge-like structures!

Mathematical Model for Images

Key Observation:

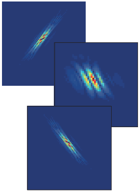
Images are governed by edge-like structures!

Problem with Wavelets:



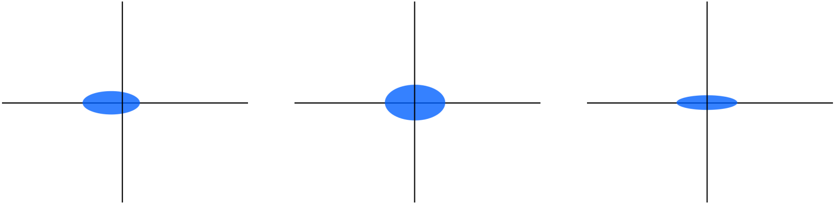
Shearlets (Kutyniok, Labate; 2006):

$$A_j := \begin{pmatrix} 2^j & 0 \\ 0 & 2^{j/2} \end{pmatrix}, \qquad S_k := \begin{pmatrix} 1 & k \\ 0 & 1 \end{pmatrix}, \quad j, k \in \mathbb{Z}.$$



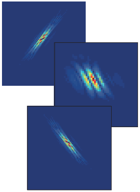
Then

$$\psi_{j,k,m} := 2^{\frac{3j}{4}} \psi(S_k A_j \cdot - m).$$



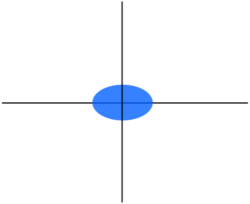
Shearlets (Kutyniok, Labate; 2006):

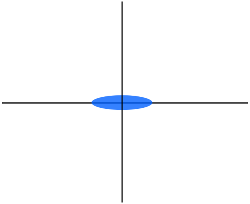
$$A_j := \begin{pmatrix} 2^j & 0 \\ 0 & 2^{j/2} \end{pmatrix}, \qquad S_k := \begin{pmatrix} 1 & k \\ 0 & 1 \end{pmatrix}, \quad j,k \in \mathbb{Z}.$$



Then

$$\psi_{j,k,m} := 2^{\frac{3j}{4}} \psi(S_k A_j \cdot -m).$$

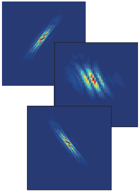




Shearlets

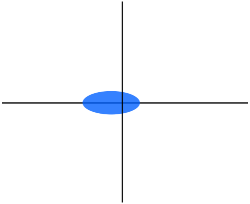
Shearlets (Kutyniok, Labate; 2006):

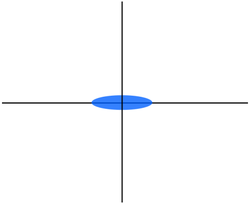
$$A_j := \begin{pmatrix} 2^j & 0 \\ 0 & 2^{j/2} \end{pmatrix}, \qquad S_k := \begin{pmatrix} 1 & k \\ 0 & 1 \end{pmatrix}, \quad j, k \in \mathbb{Z}.$$



Then

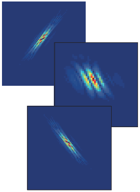
$$\psi_{j,k,m} := 2^{\frac{3j}{4}} \psi(S_k A_j \cdot - m).$$





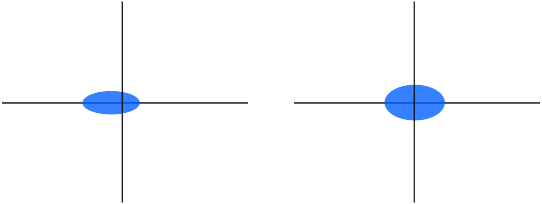
Shearlets (Kutyniok, Labate; 2006):

$$A_j := \begin{pmatrix} 2^j & 0 \\ 0 & 2^{j/2} \end{pmatrix}, \qquad S_k := \begin{pmatrix} 1 & k \\ 0 & 1 \end{pmatrix}, \quad j,k \in \mathbb{Z}.$$



Then

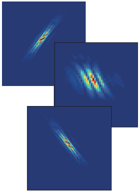
$$\psi_{j,k,m} := 2^{\frac{3j}{4}} \psi(S_k A_j \cdot - m).$$



Shearlets

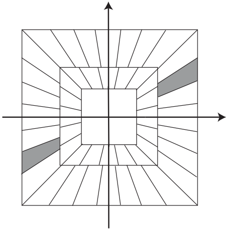
Shearlets (Kutyniok, Labate; 2006):

$$A_j := \begin{pmatrix} 2^j & 0 \\ 0 & 2^{j/2} \end{pmatrix}, \qquad S_k := \begin{pmatrix} 1 & k \\ 0 & 1 \end{pmatrix}, \quad j,k \in \mathbb{Z}.$$



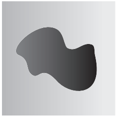
Then

$$\psi_{j,k,m} := 2^{\frac{3j}{4}} \psi(S_k A_j \cdot - m).$$



Shearlets are Optimal

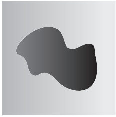
Model of Images (Donoho; 2001):
"Cartoon-functions are functions governed by a discontinuity curve."



Theorem (Kutyniok, Lim; 2011):
"Shearlets fulfill the optimal compression rate for cartoon-functions."

Shearlets are Optimal

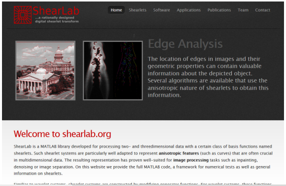
Model of Images (Donoho; 2001):
"Cartoon-functions are functions governed by a discontinuity curve."



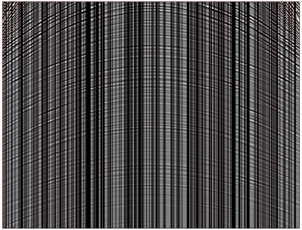
> **Theorem (Kutyniok, Lim; 2011):**
> "Shearlets fulfill the optimal compression rate for cartoon-functions."

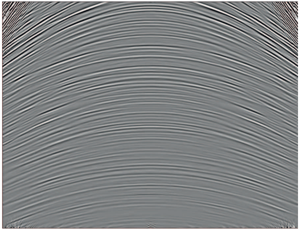
2D&3D (parallelized) Fast Shearlet Transform (`www.ShearLab.org`):

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



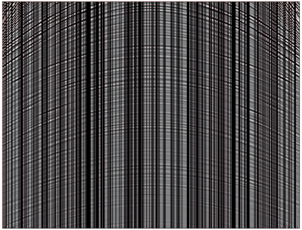
Inpainting

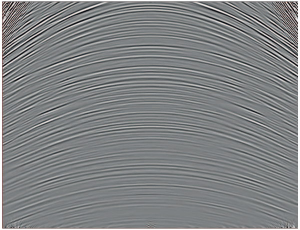




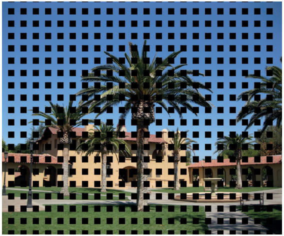
(Source: Kutyniok, Lim; 2012)

Inpainting





(Source: Kutyniok, Lim; 2012)



(Source: Kutyniok, Lim; 2015)

A Microlocal Viewpoint

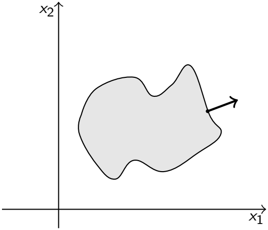
Considering edge-structures and their direction!

A Microlocal Viewpoint

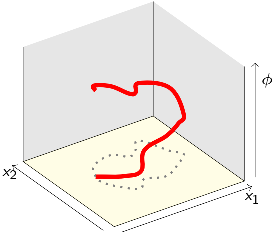
Considering edge-structures and their direction!

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

Why is the Wavefront Set Important?

Analysis of Image:

- Information about core structures of image.
- Detecting directional information often important.

Analysis of Inverse Problem:

- Analysis of available data.
- Example: Information about missing parts.

Image Reconstruction:

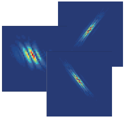
- Knowledge of wavefront set can be used for regularization.
- Example: Canonical relation between wavefront set of Radon transform and reconstructed image.

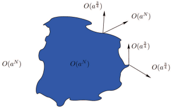
Shearlets and Wavefront Sets

Theorem (K, Labate; 2006)(Grohs; 2011):
We have

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

Shearlets can identify the wavefront set at fine scales!



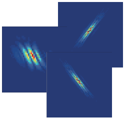


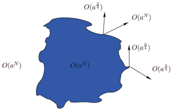
Shearlets and Wavefront Sets

Theorem (K, Labate; 2006)(Grohs; 2011):
We have

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

Shearlets can identify the wavefront set at fine scales!





Shearlets are a representation system which...

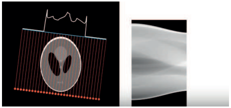
- ...is generated by one or a few 'mother functions',
- ...precisely resolves the wavefront set,
- ...provides optimally sparse approximation of cartoons,
- ...allows for compactly supported analyzing elements,
- ...is associated with fast decomposition algorithms,
- ...treats the continuum and digital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

Mathematical Modeling Reaches a Barrier

Limited Angle-(Computed) Tomography

A CT scanner samples the *Radon transform*

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

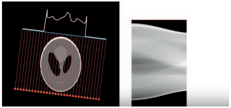


for $L(\phi, s) = \{x \in \mathbb{R}^2 : x_1 \cos(\phi) + x_2 \sin(\phi) = s\}$,
$\phi \in [-\pi/2, \pi/2)$, and $s \in \mathbb{R}$.

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

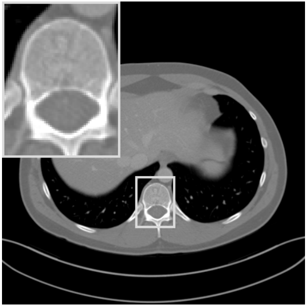
Applications: Dental CT, breast tomosynthesis, electron tomography,...

Model-Based Approaches Fail

Sparse Regularization:

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

Clinical Data:



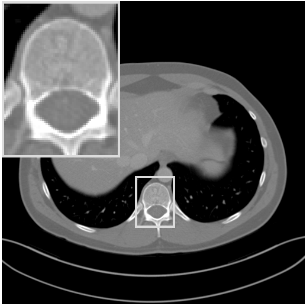
Original Image

Model-Based Approaches Fail

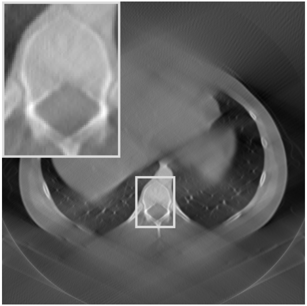
Sparse Regularization:

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

Clinical Data:



Original Image



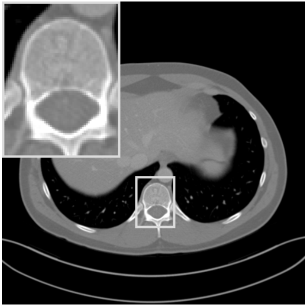
Filtered Backprojection

Model-Based Approaches Fail

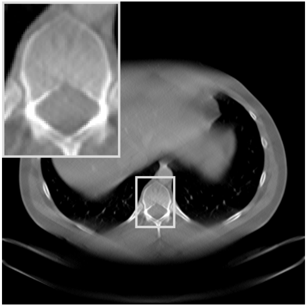
Sparse Regularization:

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

Clinical Data:



Original Image



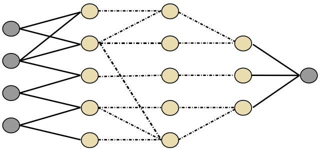
Sparse Regularization with Shearlets

Let's bring Deep Learning into the Game

The Mathematics of Deep Neural Networks

Definition:

Assume the following notions:



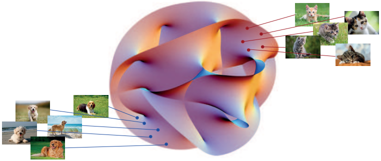
- $d \in \mathbb{N}$: Dimension of input layer.
- L: Number of layers.
- N: Number of neurons.
- $\rho : \mathbb{R} \to \mathbb{R}$: (Non-linear) function called *activation function*.
- $T_\ell : \mathbb{R}^{N_{\ell-1}} \to \mathbb{R}^{N_\ell}$, $\ell = 1, \ldots, L$: Affine linear maps.

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

$$\Phi(x) = T_L \rho(T_{L-1}\rho(\ldots\rho(T_1(x)))), \quad x \in \mathbb{R}^d,$$

is called *(deep) neural network (DNN)*.

Training of Deep Neural Networks



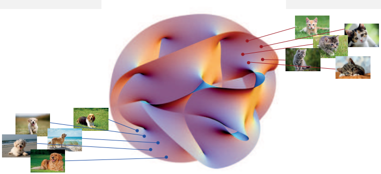
High-Level Set Up:

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

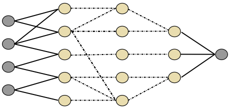
Training of Deep Neural Networks

High-Level Set Up:

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



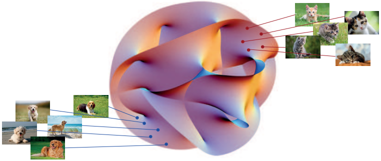
- Select an architecture of a deep neural network, i.e., a choice of d, L, $(N_\ell)_{\ell=1}^L$, and ρ.
  Sometimes selected entries of the matrices $(A_\ell)_{\ell=1}^L$, i.e., weights, are set to zero at this point.



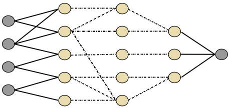
Training of Deep Neural Networks

High-Level Set Up:

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



- Select an architecture of a deep neural network, i.e., a choice of d, L, $(N_\ell)_{\ell=1}^L$, and ρ.
  Sometimes selected entries of the matrices $(A_\ell)_{\ell=1}^L$, i.e., weights, are set to zero at this point.



- Learn the affine-linear functions $(T_\ell)_{\ell=1}^L = (A_\ell \cdot + b_\ell)_{\ell=1}^L$ by

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

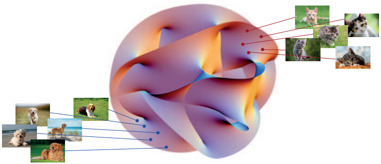
  yielding the network $\Phi_{(A_\ell, b_\ell)_\ell} : \mathbb{R}^d \to \mathbb{R}^{N_L}$,

$$\Phi_{(A_\ell, b_\ell)_\ell}(x) = T_L \rho(T_{L-1} \rho(\ldots \rho(T_1(x)))).$$

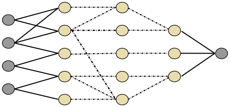
  This is often done by stochastic gradient descent.

Training of Deep Neural Networks

High-Level Set Up:



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



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

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

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

$$\Phi_{(A_\ell, b_\ell)_\ell}(x) = T_L \rho(T_{L-1} \rho(\ldots \rho(T_1(x)))).$$

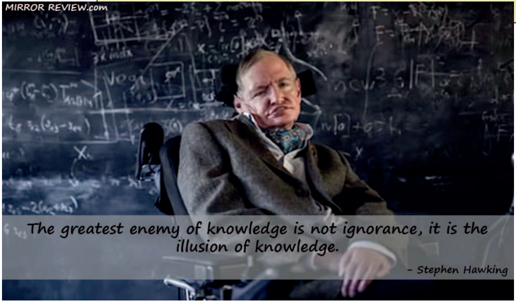
This is often done by stochastic gradient descent.

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

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, 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, 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, ...*

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).

Models and Data: A Microlocal Perspective

General Mission Statement:

- Employ model-based approaches as far as they are reliable.
- Apply deep learning only when it is necessary.

Models and Data: A Microlocal Perspective

General Mission Statement:

- Employ model-based approaches as far as they are reliable.
- Apply deep learning only when it is necessary.

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)

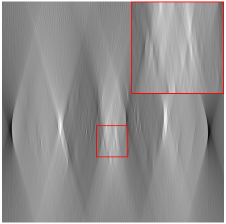
joint with

Maximilian März (TU Berlin)

Wojciech Samek and Vignesh Srinivan (Fraunhofer HHI Berlin)

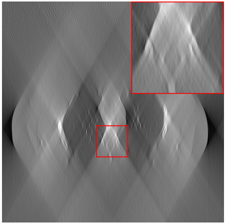
Tatiana Bubba, Matti Lassas, and Samuli Siltanen (University of Helsinki)

Zooming in on the Recovery Problem



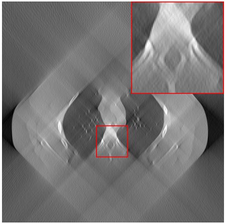
$\phi = 15^{\circ}$, filtered backprojection (FBP)

Zooming in on the Recovery Problem



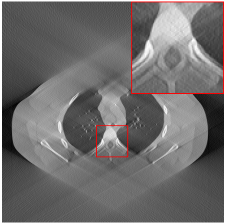
$\phi = 30^\circ$, filtered backprojection (FBP)

Zooming in on the Recovery Problem



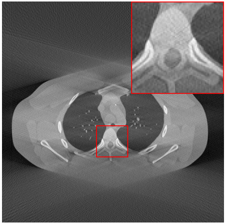
$\phi = 45^\circ$, filtered backprojection (FBP)

Zooming in on the Recovery Problem



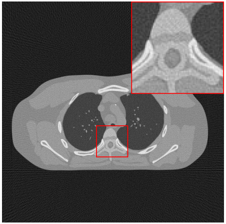
$\phi = 60^\circ$, filtered backprojection (FBP)

Zooming in on the Recovery Problem



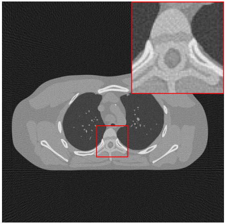
$\phi = 75^\circ$, filtered backprojection (FBP)

Zooming in on the Recovery Problem



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

Zooming in on the Recovery Problem



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

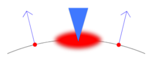
Illustration of Theorem ([Quinto, 1993]):

"visible": singularities tangent to sampled lines

"invisible": singularities not tangent to sampled lines

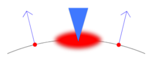
Shearlets can Help

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



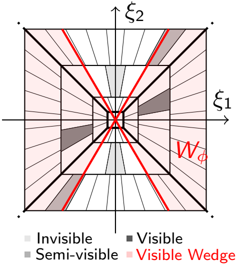
Shearlets can Help

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



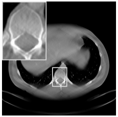
The Shearlet Transform:

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



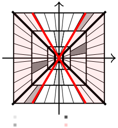
Our Approach "Learn the Invisible (LtI)"

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*

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

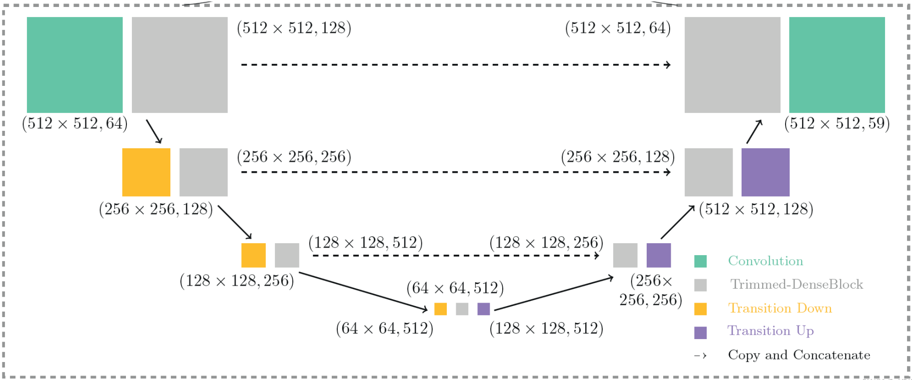
Step 3: *Combine*

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

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

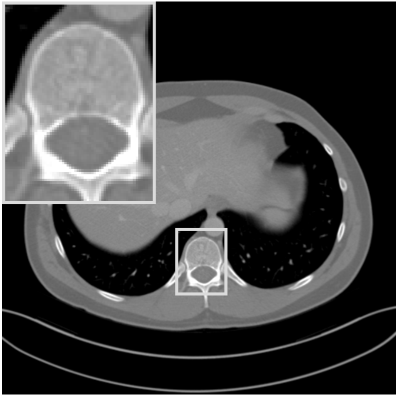
Setup

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
- ...

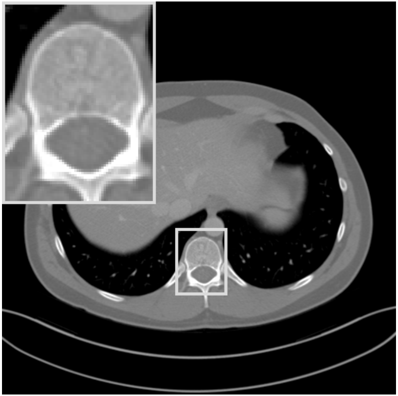
[1]We would like to thank Dr. Cynthia McCollough, the Mayo Clinic, the American Association of Physicists in Medicine (AAPM), and grant EB01705 and EB01785 from the National Institute of Biomedical Imaging and Bioengineering for providing the Low-Dose CT Grand Challenge data set.

Evaluation on Test Patient

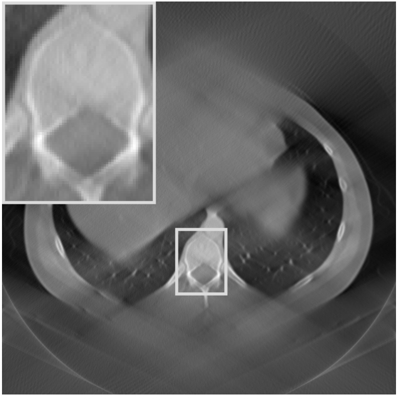


f_{gt}

Evaluation on Test Patient

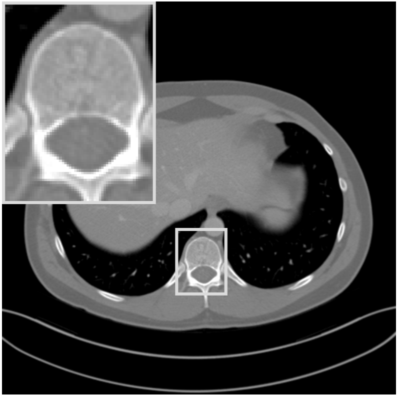


f_{gt}

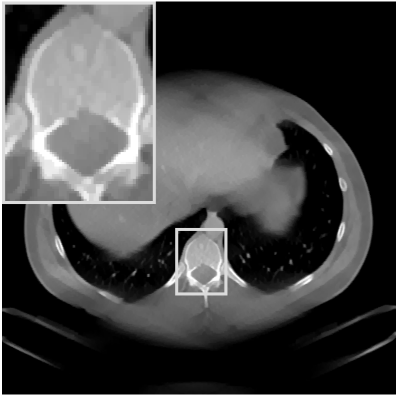


f_{FBP}: RE = 0.50, HaarPSI=0.35

Evaluation on Test Patient

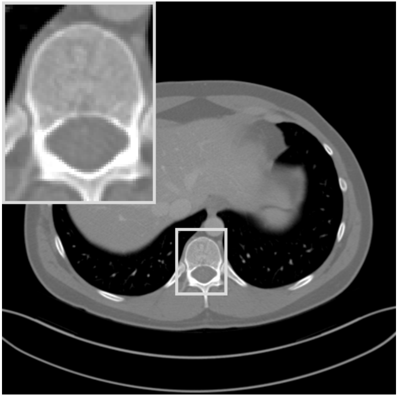


f_{gt}

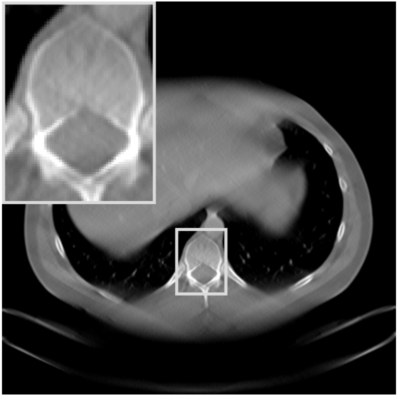


f_{TV}: RE = 0.21, HaarPSI=0.41

Evaluation on Test Patient

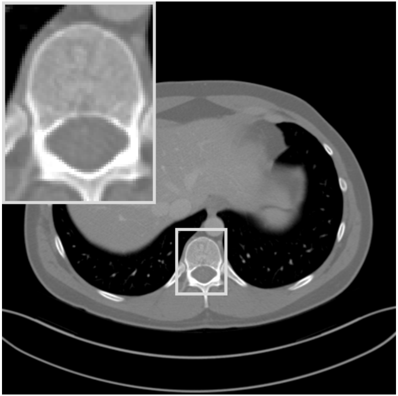


f_{gt}



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

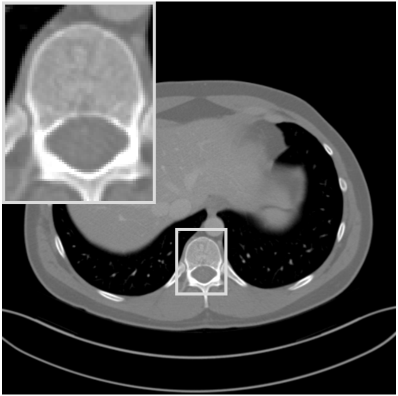
Evaluation on Test Patient



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

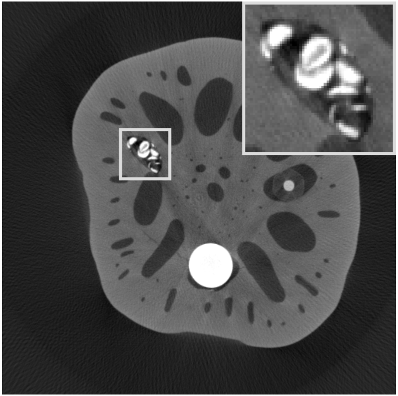
HaarPSI (Reisenhofer, Bosse, K, and Wiegand; 2018)

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

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

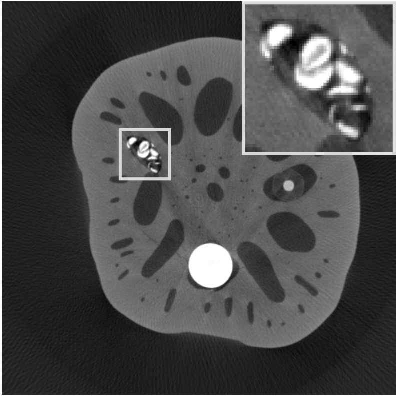
www.haarpsi.org

Generalization to Lotus Root

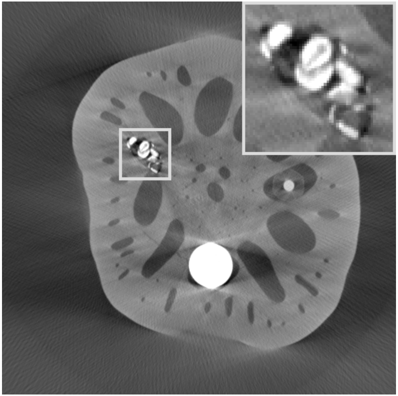


f_{gt}

Generalization to Lotus Root

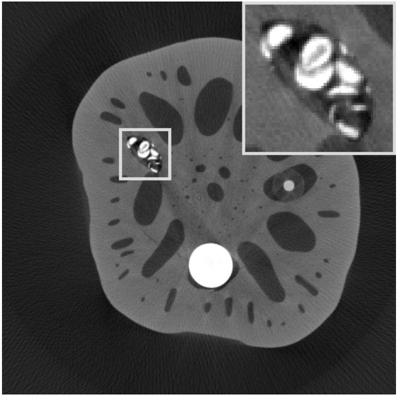


f_{gt}

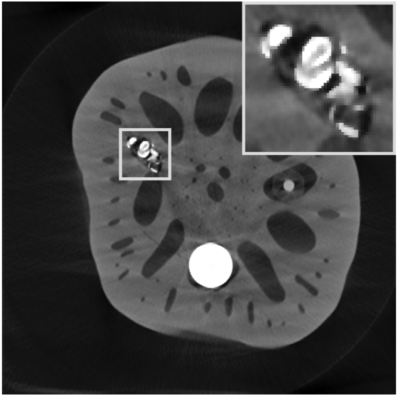


f_{FBP}: RE $= 0.31$, HaarPSI$=0.61$

Generalization to Lotus Root

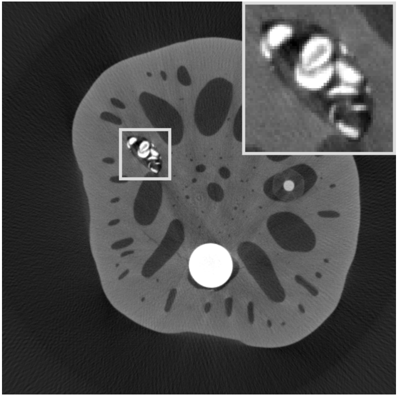


f_{gt}

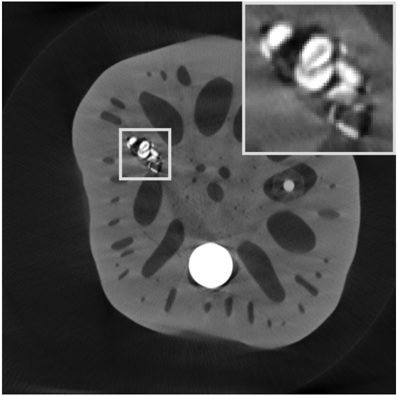


f_{TV}: RE $= 0.12$, HaarPSI$=0.74$

Generalization to Lotus Root

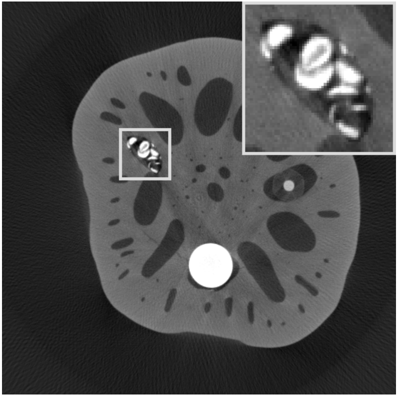


f_{gt}

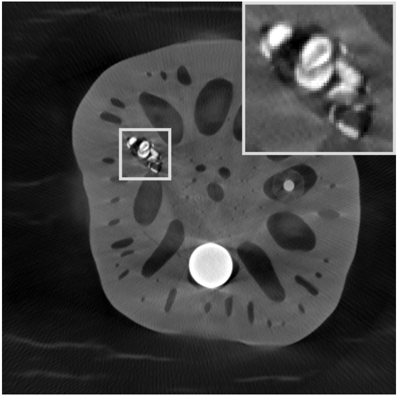


f^*: RE $= 0.11$, HaarPSI$=0.75$

Generalization to Lotus Root

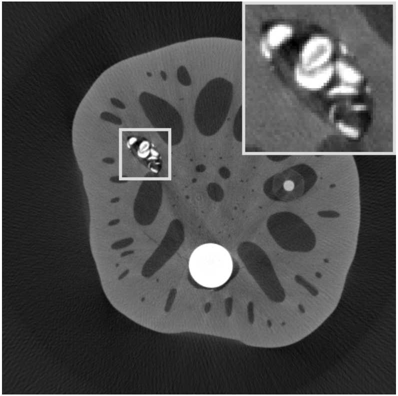


f_{gt}

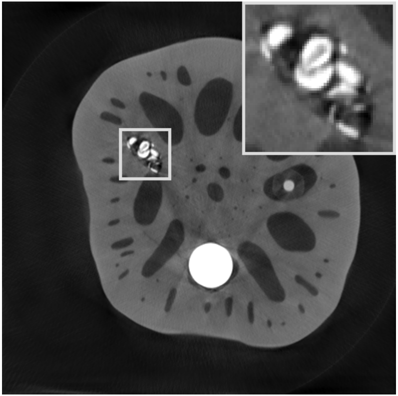


$f_{\text{[Gu \& Ye, 2017]}}$: RE $= 0.25$, HaarPSI$=0.62$

Generalization to Lotus Root



$f_{\mathtt{gt}}$



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

Deep Network Shearlet Edge Extractor (DeNSE)

joint with

Hector Andrade-Loarca (LMU Munich)
Ozan Öktem (KTH Royal Institute of Technology)
Philipp Petersen (University of Vienna)

Key Ideas

Model-Based World:

- Shearlets can precisely resolve the wavefront set.
- Shearlet coefficients provide a significantly improved representation for wavefront set extraction.

$\rightsquigarrow$ *Use shearlets for the heavy lifting in preprocessing!*

Data-Driven World:

- Deep neural networks allow a strong adaptation to a function class.
- Stability can be increased if the data is suitably prepared.

$\rightsquigarrow$ *Apply a neural networks in shearlet domain for classification!*

Key Ideas

Model-Based World:

- Shearlets can precisely resolve the wavefront set.
- Shearlet coefficients provide a significantly improved representation for wavefront set extraction.

$\rightsquigarrow$ *Use shearlets for the heavy lifting in preprocessing!*

Data-Driven World:

- Deep neural networks allow a strong adaptation to a function class.
- Stability can be increased if the data is suitably prepared.

$\rightsquigarrow$ *Apply a neural networks in shearlet domain for classification!*

Compute the wavefront set of the reconstructed CT image!

Key Ideas

Model-Based World:

- Shearlets can precisely resolve the wavefront set.
- Shearlet coefficients provide a significantly improved representation for wavefront set extraction.

$\rightsquigarrow$ *Use shearlets for the heavy lifting in preprocessing!*

Data-Driven World:

- Deep neural networks allow a strong adaptation to a function class.
- Stability can be increased if the data is suitably prepared.

$\rightsquigarrow$ *Apply a neural networks in shearlet domain for classification!*

Compute the wavefront set of the reconstructed CT image!

Use for regularization of inverse problems!

Deep Network Shearlet Edge Extractor (DeNSE)

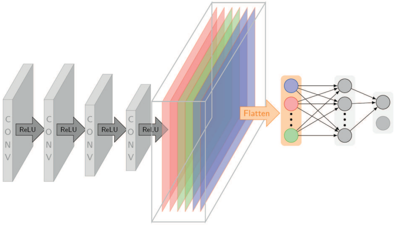
Key Steps:

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

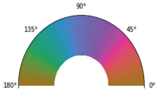
(2) Consider patches of shearlet coefficients.
  ⇝ Localize to each position.

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

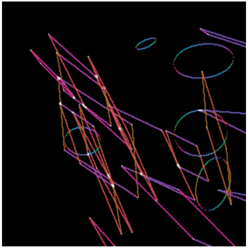
Network Architecture:

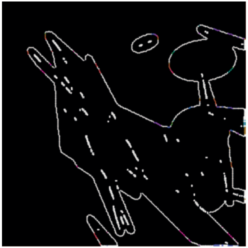


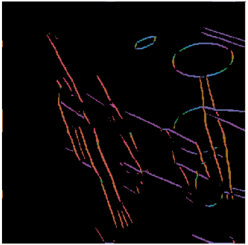
Smoothed Ellipses and Parallelograms

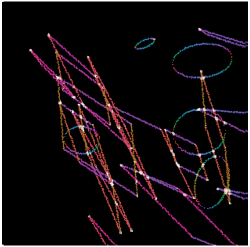



Original


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**

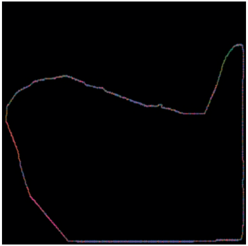
MF-Score:

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

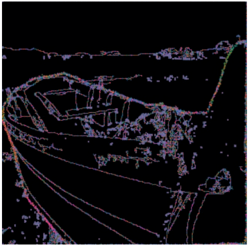
- P is the precision, i.e., the number of true positives divided by the sum of true and false positives,
- R is the recall, i.e., the number of true positives divided by the sum of true positives and false negatives.

BSDS500 Data Set

Original

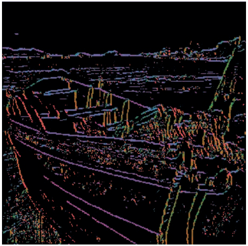


Human Annotation



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. Derive the wavefront set of the reconstructed image.

Canonical Relation:

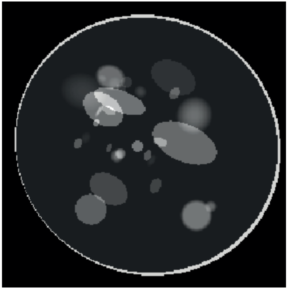
The canonical relation C satisfies

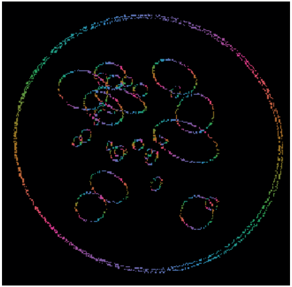
$$\mathrm{WF}\big(R(f)\big) = C \circ \mathrm{WF}(f) \quad \text{whenever } f \in \mathcal{D}'(\mathbb{R}^2),$$

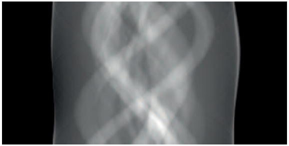
where

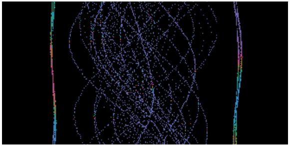
$$C = \Big\{\big(\theta, p, s(-x \cdot \omega(\theta)^{\perp} \mathrm{d}\theta + \mathrm{d}p); x, s\omega(\theta)\mathrm{d}x\big) \in T^*(\mathbb{M}) : (\theta, p) \in \mathbb{M}, x \in \mathbb{R}^2, s \neq 0, x \cdot \omega(\theta) = p\Big\}.$$

Application of the Canonical Relation

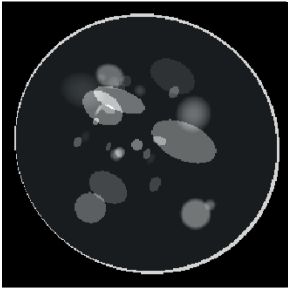

Phantom


Wavefront Set by DeNSE

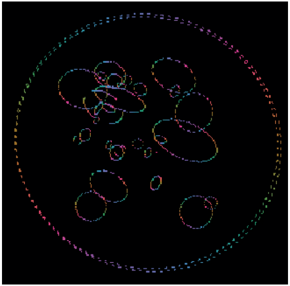

Sinogram


Wavefront Set by Canonical Relation

Application of the Inverse Canonical Relation

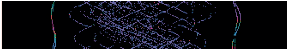


Phantom



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 (soon) available at:

www.math.lmu.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)