Wasserstein for learning image regularisers

Carola-Bibiane Schönlieb

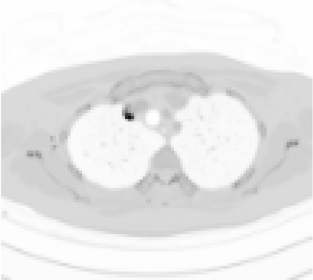
Department for Applied Mathematics and Theoretical Physics
Cantab Capital Institute for the Mathematics of Information
EPSRC Centre for Mathematical Imaging in Healthcare
Alan Turing Institute
University of Cambridge, UK

Banff, 11 December 2018

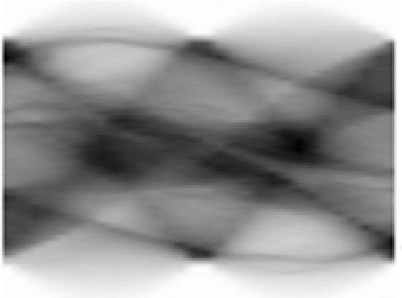
UNIVERSITY OF CAMBRIDGE

Forward problem: knowing physical quantity compute the measurements
Inverse problem: measuring the datum compute the physical quantity.

Physical quantity
(x-ray attentuation)



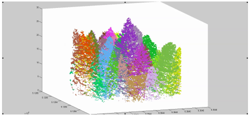
Measurements
(sinogram)



Mathematically: compute u from $f = T(u) + n$.

CT images from LUNA dataset `https://luna16.grand-challenge.org`.

Inverse problems in imaging?

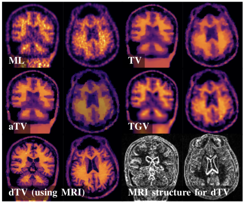


Multi-sensor image analysis for forest conservation

J. Lee, CBS, D. Coomes, et al. '15–

Unveiling the invisible for art conservation

Weickert '13; D'Autume, Panayotova, Ricciardi, CBS '17



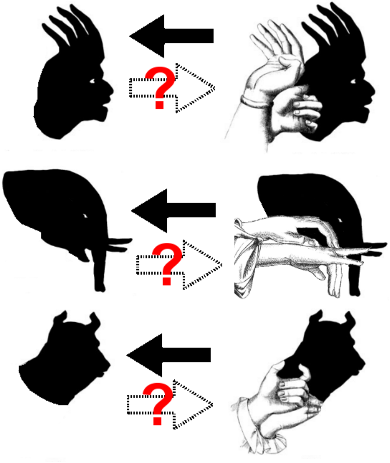
Positron emission tomography

Burger et al. '08; Ehrhardt, Markiewicz, CBS '17, '18

3D conversion

Bertalmio et al. '98; Morel, Masnou '98; Chan, Kang, Shen '00; Burger, He, CBS, SIAM Imaging Science '09; CBS, CUP '15J. Hocking, CBS et al. Heritage Sciences '18

Ill-posedness



The variational approach

General task: **restore u** from an **observed datum f** where

$$f = \underbrace{T(u)}_{\text{forward model}} + \underbrace{n}_{\text{noise}}.$$

Variational approach: Compute u as a minimizer of

$$\mathcal{J}(u) = \alpha \underbrace{R(u)}_{\text{regularization}} + \underbrace{D(T(u), f)}_{\text{data fidelity}} \to \min_{u \in B},$$

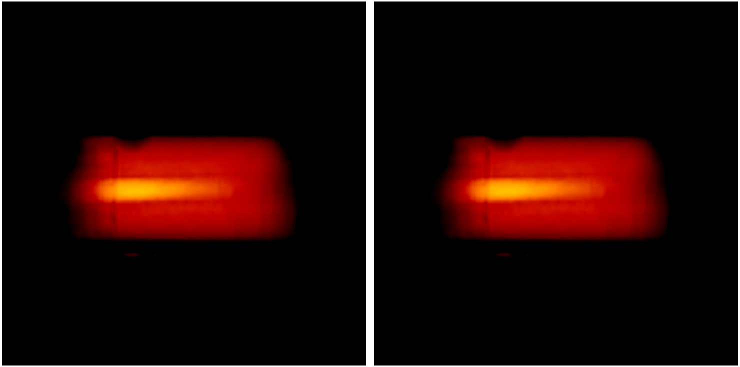
where

- $R(u)$ is a prior/regularizer that models a-priori information on u weighted by positive α, e.g., $R(u) = \|\nabla u\|_1$ (in infinite dimensions $|Du|(\Omega)$)
- $D(\cdot, \cdot)$ is a distance function, e.g. $D(Tu, g) = \|Tu - f\|_2^2$ and B suitable Banach space, e.g., $B = BV(\Omega)$.

Engl, Hanke, Neubauer '96; Rudin, Osher, Fatemi, Physica D '92; Natterer, Wübbeling '01; Candes, Romberg, Tao, IEEE Trans Inf Theory '06; Kaltenbacher, Neubauer, Scherzer '08; Schuster, Kaltenbacher, Hofmann, Kazimierski '12

UNIVERSITY OF CAMBRIDGE

4D Bregman-TV Reconstruction from 20% line-sampling

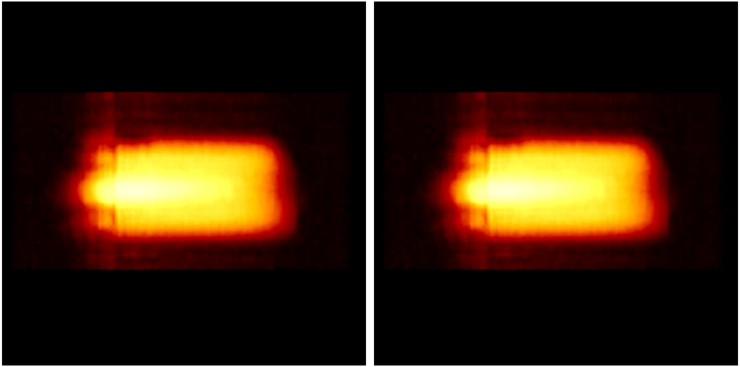


$$\text{Acquire } \mathcal{SF}(u) \Rightarrow \text{ reconstruct } u(x,t) = \text{argmin}_v \alpha \; \|\nabla v\|_1 + \|\mathcal{SF}(v) - f\|_2^2$$

M. Benning, A. Sederman, CBS, L. Gladden, et al.; M. Benning, L. Gladden, D. Holland, CBS, and T. Valkonen, J. Magnetic Resonance 238, pp. 26 - 43, 2014; Osher, Burger et al. '05; M. Lustig et al. '07

UNIVERSITY OF CAMBRIDGE

Fourier inversion from 20% line-sampling



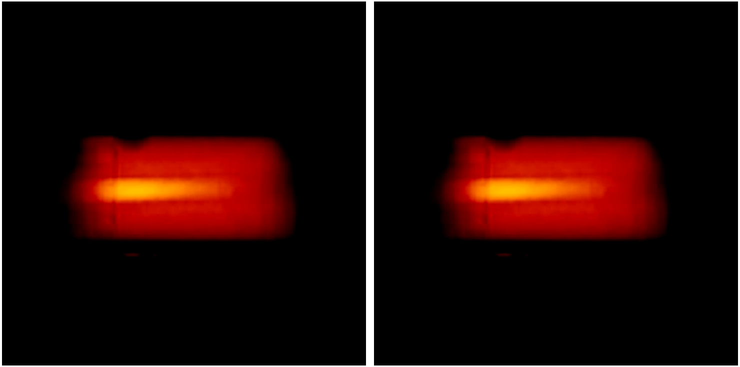
Acquire $\mathcal{SF}(u) \Rightarrow$ reconstruct $u(x,t)$ by zero-filling.

M. Benning, A. Sederman, CBS, L. Gladden, et al.

4D MRI from sub-sampled data

UNIVERSITY OF CAMBRIDGE

4D Bregman-TV Reconstruction from 20% line-sampling



$$\text{Acquire } \mathcal{SF}(u) \Rightarrow \text{ reconstruct } u(x,t) = \text{argmin}_v \alpha \; \|\nabla v\|_1 + \|\mathcal{SF}(v) - f\|_2^2$$

M. Benning, A. Sederman, CBS, L. Gladden, et al.; M. Benning, L. Gladden, D. Holland, CBS, and T. Valkonen, J. Magnetic Resonance 238, pp. 26 - 43, 2014; Osher, Burger et al. '05; M. Lustig et al. '07

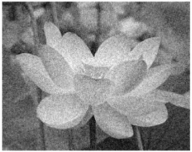
Connection to nonlinear PDEs

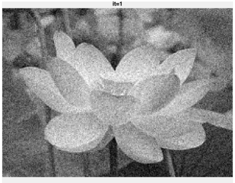
Nonlinear image smoothing with total variation regularisation

$$\alpha \|\nabla u\|_1 + \frac{1}{2}\|u - f\|^2 \to \min_u$$

with steepest descent

$$u_t = \alpha\, p + (u - f), \quad p \in \partial \|\nabla u\|_1, \quad \text{in } \Omega,$$





TV scale space

Perona, Malik ’90; Rudin, Osher, Fatemi, Physica D ’92; and various contributions from Ambrosio, Caselles, Chambolle, Lions, Morel, Novaga, . . .

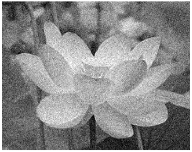
Connection to nonlinear PDEs

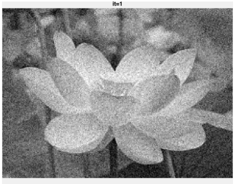
Nonlinear image smoothing with total variation regularisation

$$\alpha {\color{red}\|\nabla u\|_1} + \frac{1}{2}\|u - f\|^2 \to \min_u$$

with steepest descent for $|\nabla u| \neq 0$

$$u_t = \alpha\, {\color{red}\mathrm{div}\left(\frac{\nabla u}{|\nabla u|}\right)} + (u - f), \quad \text{in } \Omega,$$





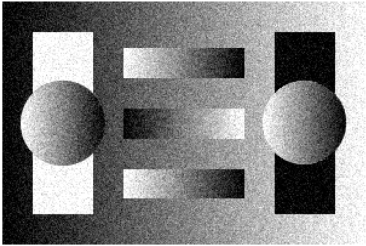
TV scale space

Perona, Malik ’90; Rudin, Osher, Fatemi, Physica D ’92; and various contributions from Ambrosio, Caselles, Chambolle, Lions, Morel, Novaga, . . .

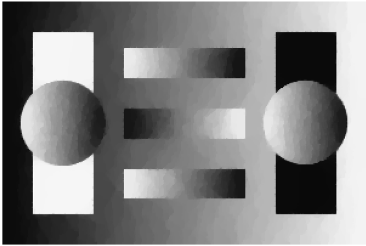
What is the right sparsity?

UNIVERSITY OF CAMBRIDGE

$\min_u \{\alpha\|\nabla u\|_1 + \|u - f\|_2^2\}$



Noisy image



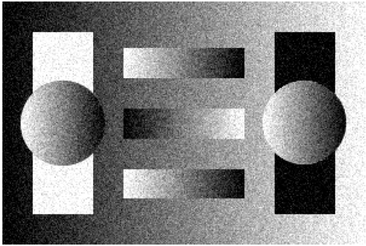
TV denoised image

Image courtesy of K. Papafitsoros

References: Rudin, Osher, Fatemi '92; Hinterberger, Scherzer, Computing '06; Bredies, Kunisch, Pock, SIAM Imaging '10; Papafitsoros, CBS, J. Math. Imaging & Vision, '13 ...

UNIVERSITY OF CAMBRIDGE

$\min_u \{\alpha\|\nabla^2 u\|_1 + \|u - f\|_2^2\}$

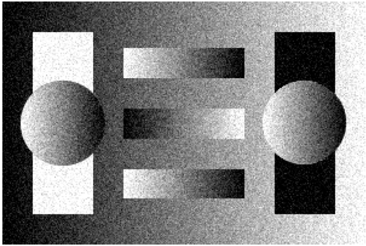


Noisy image   TV^2 denoised image

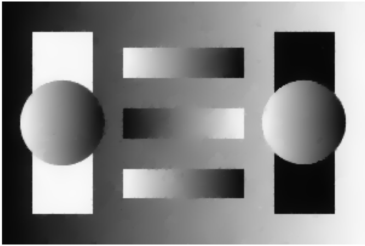
Image courtesy of K. Papafitsoros
References: Rudin, Osher, Fatemi '92; Hinterberger, Scherzer, Computing '06; Bredies, Kunisch, Pock, SIAM Imaging '10; Papafitsoros, CBS, J. Math. Imaging & Vision, '13 ...

What is the right sparsity?

$$\min_u \left\{ \min_w \{\alpha_1 \|\nabla u - w\|_1 + \alpha_2 \|Ew\|\} + \|u - f\|_2^2 \right\}$$



Noisy image



TGV^2 denoised image

Image courtesy of K. Papafitsoros

References: Rudin, Osher, Fatemi '92; Hinterberger, Scherzer, Computing '06; Bredies, Kunisch, Pock, SIAM Imaging '10; Papafitsoros, CBS, J. Math. Imaging & Vision, '13 . . .

Regulariser zoo

No 'one fit all' solution

- Multi-resolution analysis, wavelets (e.g. Daubechies, Mallat, Unser, Kutyniok, Foucart & Rauhut, . . .).
- Other Banach-space norms, e.g. Sobolev norms, Besov norms, etc. (e.g. Lassas, Siltanen 09)
- Higher-order total variation regularisation (Infimal convolution Chambolle, Lions 97; Setzer, Steidl, Teuber 11, Total Generalised Variation Bredies, Kunisch, Pock 10, . . .)
- Non-local regularisation (non-local TV Osher, Gilboa, . . . ; non-local means Morel . . .)
- Anisotropic regularisation Weickert98
- Free-discontinuity problems Mumford, Shah; Tomarelli et al.
- and mixtures of the above . . . and probably more which I have forgotten . . .

Introductory books to variational & PDE imaging Chan & Shen 05; Scherzer 10; Bredies & Lorenz 11 – currently only in German.

Joint work with Sebastian Lunz and Ozan Öktem

S. Lunz, O. Öktem, CBS, Adversarial Regularizers in Inverse Problems, NIPS 2018

Existing methods for deep inversion

UNIVERSITY OF CAMBRIDGE

Existing Approaches

- Fully Learned Models
- Post Processing[1]
- Iterative Schemes[2][3][4]

Variational formulation could resolve some shortcomings of existing algorithms (e.g., provable notion of stability of regularisation)

[1] Kyong Hwan Jin et al. "Deep convolutional neural network for inverse problems in imaging". In: *IEEE Transactions on Image Processing* 26.9 (2017), pp. 4509–4522.

[2] Jonas Adler and Ozan Öktem. "Learned primal-dual reconstruction". In: *arXiv preprint arXiv:1707.06474* (2017).

[3] Tim Meinhardt et al. "Learning Proximal Operators: Using Denoising Networks for Regularizing Inverse Imaging Problems". In: *arXiv preprint arXiv:1704.03488* (2017).

[4] Kerstin Hammernik et al. "Learning a variational network for reconstruction of accelerated MRI data". In: *Magnetic resonance in medicine* (2017).

Data-driven regulariser

UNIVERSITY OF CAMBRIDGE

A bit of motivation / assumption:

- Design of regularisation functional $R(x)$ does not require solving the variational problem.
- Heuristic: Design regularisation functional to distinguish between artifact-free images (ground truth) and images with artifacts (e.g. noisy images, images with streak artifacts ...).
- How: train regulariser on image distributions, utilising both true image distribution π and distribution of corrupted images ρ.[5]
- If available, can pick ρ distribution of pseudo-inverse

[5] Martin Benning et al. "Learning Filter Functions in Regularisers by Minimising Quotients". In: *Scale Space and Variational Methods in Computer Vision*. Ed. by François Lauze, Yiqiu Dong and Anders Bjorholm Dahl. Cham: Springer International Publishing, 2017, pp. 511–523; Sebastian Lunz, Ozan Öktem and Carola-Bibiane Schönlieb. "Adversarial Regularizers in Inverse Problems". In: *NIPS 2018, arXiv preprint arXiv:1805.11572* (2018); Housen Li et al. "NETT: Solving Inverse Problems with Deep Neural Networks". In: *arXiv preprint arXiv:1803.00092* (2018).

Learned regulariser

Going back to the variational framework we want to replace

$$\arg\min_x \|Tx - y\|_2^2 + \lambda R(x)$$

by

$$\arg\min_x \|Tx - y\|_2^2 + \lambda \Psi_\Theta(x),$$

where $\Psi_\Theta(x)$ is large for 'undesirable' x (coming from ρ) and small for 'desirable' x (coming from π).

Use loss functional of the form

$$\mathbb{E}_{X\sim\pi}\Psi_\Theta(X) - \mathbb{E}_{X\sim\rho}\Psi_\Theta(X)$$

for training Ψ_Θ.

Wasserstein Loss

UNIVERSITY OF CAMBRIDGE

Consider minimising for an appropriately parametrised Ψ_Θ (deep neural network of some kind) the Wasserstein loss functional[6]

$$\mathbb{E}_{X\sim\pi}\Psi_\Theta(X) - \mathbb{E}_{X\sim\rho}\Psi_\Theta(X) + \lambda\cdot\mathbb{E}\left(\|\nabla_x\Psi_\Theta(X)\| - 1\right)_+^2 .$$

Approximation to solution of

$$\sup_{f\in 1-Lip} \mathbb{E}_{X\sim\rho}f(X) - \mathbb{E}_{X\sim\pi}f(X).$$

Motivation: Kantorovich duality for optimal transport.

[6] Martín Arjovsky, Soumith Chintala and Léon Bottou. "Wasserstein Generative Adversarial Networks". In: *Proceedings of the 34th International Conference on Machine Learning, ICML 2017, Sydney, NSW, Australia, 6-11 August 2017*. 2017, pp. 214–223. URL: http://proceedings.mlr.press/v70/arjovsky17a.html.

Distributional Interpretation

UNIVERSITY OF CAMBRIDGE

Assume Ψ_Θ achieves minimal Wasserstein loss.

Question: does gradient descent on Ψ_Θ take us closer to desired image distribution π?

For x sampled from ρ define

$$g_\epsilon(x) = x - \epsilon \cdot \nabla_x \Psi_\Theta(x)$$
$$\rho_\epsilon := (g_\epsilon)_*(\rho)$$

Theorem

Under sufficient regularity assumptions

$$\frac{\partial}{\partial \epsilon} \mathrm{Wass}(\pi, \rho_\epsilon)|_{\epsilon=0} = -\mathbb{E}_{X\sim\rho} \|\nabla_x \Psi_\Theta(X)\|_2^2$$

Data Manifold Distance Theorem

UNIVERSITY OF CAMBRIDGE

Assumption (Data Manifold Assumption)

There is a weakly compact set $\mathcal{M}$ such that $\mathrm{supp}(\pi) \subset \mathcal{M}$

Theorem

Assume DMA. Denote by P the L^2 projection onto the data manifold. Assume

$$P_*(\rho) = \rho \circ P^{-1} = \pi$$

Then, the distance function to the data manifold

$$x \mapsto \min_{y \in \mathcal{M}} \|x - y\|_2$$

is a maximizer to the Wasserstein Loss

$$\sup_{f \in 1-Lip} \mathbb{E}_{X \sim \rho} f(X) - \mathbb{E}_{X \sim \pi} f(X).$$

Existence and weak stability

Under appropriate assumptions we can prove

Theorem

A minimiser of $\|Tx - y\|^2 + \lambda \Psi_\Theta(x)$ *exists.*

Theorem

Let y_n *be a sequence in* Y *with* $y_n \to y$ *in the norm topology and* x_n *sequence of minimizers of the functional*

$$\arg\min_{x \in X} \|Tx - y_n\|^2 + \lambda \Psi_\Theta(x)$$

Then x_n *has a weakly convergent subsequence and its limit* x *is a minimizer of*

$$\|Tx - y\|^2 + \lambda \Psi_\Theta(x).$$

Summary: Learning regulariser

Train Regulariser via:

$$\sum_i \Psi_\Theta(x_i) - \Psi_\Theta(T^\dagger y_i) + \gamma \cdot \left(\|\nabla_x \Psi_\Theta(\epsilon_i x_i + (1-\epsilon_i) T^\dagger y_i)\| - 1 \right)_+^2 .$$

ϵ_i is sampled uniformly in $[0, 1]$.
Reconstruct via:

$$\arg\min_x \|Tx - y\|_2^2 + \lambda \Psi_\Theta(x)$$

for appropriately chosen (estimated) λ. Minimization problem is solved using gradient descent.

CT reconstructions on LIDC dataset

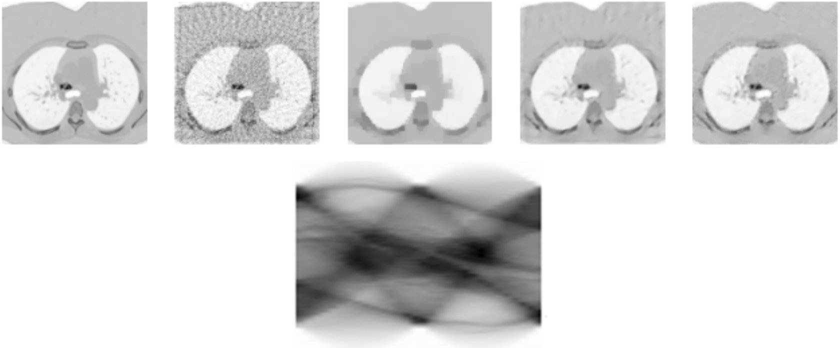
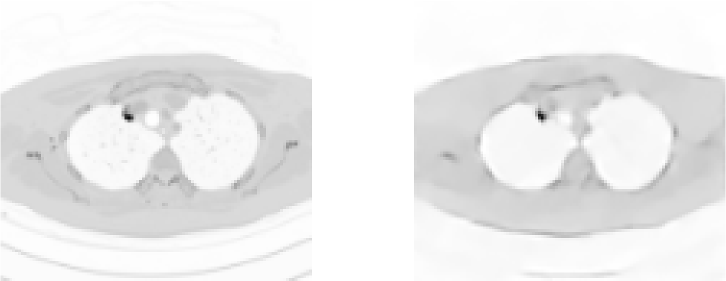


Figure: From left to right: Ground truth, FBP, TV, Post-Processing, Adversarial Reg. Below the Sinogram used for reconstruction.

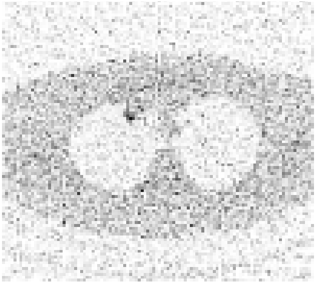
CT reconstruction on LIDC data



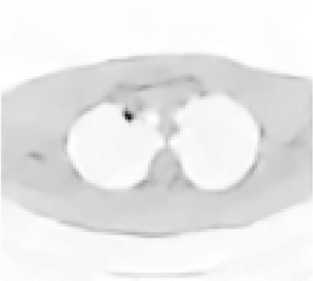
(a) Ground Truth (b) Adversarial Reg.

Figure: Reconstruction from simulated CT measurements LIDC

CT reconstruction on LIDC data



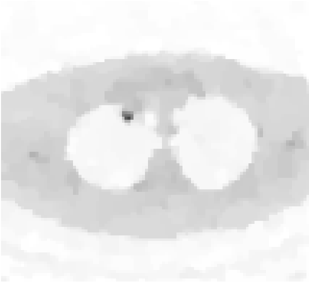
(a) FBP



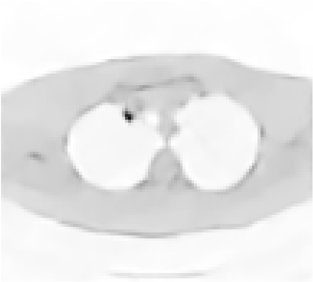
(b) Adversarial Reg.

Figure: Reconstruction from simulated CT measurements LIDC

CT reconstruction on LIDC data



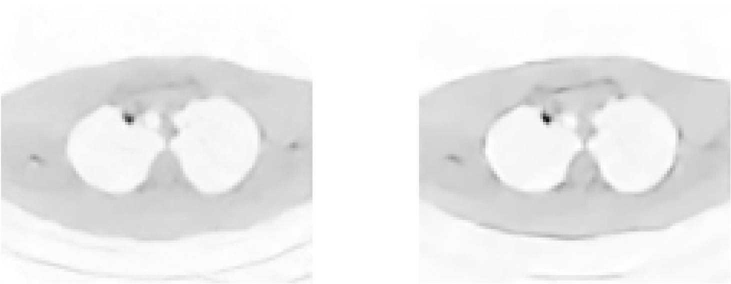
(a) TV



(b) Adversarial Reg.

Figure: Reconstruction from simulated CT measurements LIDC

CT reconstruction on LIDC data



(a) Post-Processing (b) Adversarial Reg.

Figure: Reconstruction from simulated CT measurements on LIDC

CT reconstruction on LIDC data

Method	PSNR (dB)	SSIM
MODEL-BASED		
Filtered Backprojection	14.9	.227
Total Variation	27.7	.890
UNSUPERVISED		
Adversarial Reg. (ours)	30.5	.927
SUPERVISED		
Post-Processing	31.2	.936

Table: CT reconstruction on LIDC data, high noise

Take away messages

Learning of variational models by learning appropriate parametrisation of regulariser

- Wasserstein GAN regulariser parametrising with distributions of ‘desirable’ and ‘undesirable’ solutions (deep)

Take away messages

UNIVERSITY OF CAMBRIDGE

Learning of variational models by learning appropriate parametrisation of regulariser

- Wasserstein GAN regulariser parametrising with distributions of ‘desirable’ and ‘undesirable’ solutions (deep)
- Advantages:
  - Data term allows to insert knowledge about forward operator and noise model very directly
  - Stability theory (even if weak)
  - GAN: Loss on distributions rather than images directly
  - GAN: No paired training data necessary
- Disadvantages: computational complexity. forward

Conclusion

Philosophy: learning structured but adaptive imaging models with guarantees

See also forthcoming Acta Numerica 2019.

Cambridge Image Analysis

UNIVERSITY OF CAMBRIDGE

- Dr Angelica Aviles-Rivero
- Dr Noemie Debroux
- Dr Yury Korolev
- Dr Lukas Lang
- Dr Pan Liu
- Dr Jingwei Liang
- Dr Matt Thorpe
- Thomas Buddenkotte
- Veronica Corona
- Tamara Grossmann
- Sebastian Lunz
- Lisa Kreusser
- Simone Parisotto
- Erlend Riis
- Philip Sellars
- Ferdia Sherry
- Rob Tovey
- Jon Williams

Visit us at `http://www.damtp.cam.ac.uk/research/cia/`
and check out our IMAGES network at
`http://www.images.group.cam.ac.uk`

Sponsored by EPSRC, London Mathematical Society, The Leverhulme Trust, Alan Turing Institute, Horizon2020, and The Royal Society.

Thank you very much for your attention!

More information see:

http://www.ccimi.maths.cam.ac.uk

http://www.cmih.maths.cam.ac.uk

http://www.damtp.cam.ac.uk/research/cia/

Email: cbs31@cam.ac.uk