Learning data-driven priors for image reconstruction: From bilevel optimisation to neural network-based unrolled schemes

Kostas Papafitsoros

School of Mathematical Sciences
Queen Mary University of London

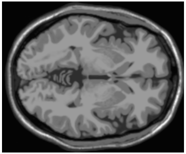
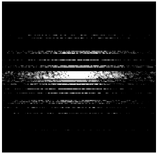
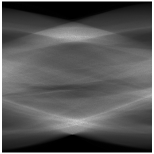
Workshop on Recent Advances in Iterative Reconstruction
UCL
22 May 2023

Inverse problems in imaging – General framework

We are given data $\mathbf{z}$ that satisfy

$$\mathbf{z} = \mathbf{A}\mathbf{x}_{true} + \mathbf{e}$$

$\mathbf{x}_{true}$	ground truth image
$\mathbf{e}$	random noise
$\mathbf{A} : V^n \to V^m$	bounded linear operator, $V = \{\mathbb{R}, \mathbb{C}\}$

Identity	Convolution	Fourier coefficients	Radon transform
			

Inverse problems in imaging – General framework

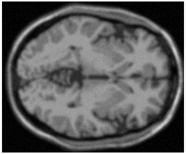
We are given data $\mathbf{z}$ that satisfy

$$\mathbf{z} = \mathbf{A}\mathbf{x}_{true} + \mathbf{e}$$

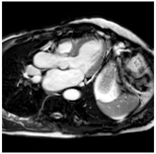
$\mathbf{x}_{true}$	ground truth image
$\mathbf{e}$	random noise
$\mathbf{A} : V^n \to V^m$	bounded linear operator, $V = \{\mathbb{R}, \mathbb{C}\}$

Identity

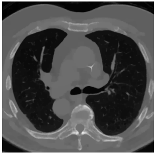
Convolution



Fourier coefficients



Radon transform



Variational regularisation approach

$$\min_{\mathbf{x}\in X} \underbrace{d(\mathbf{Ax}, \mathbf{z})}_{\text{data discrepancy}} + \underbrace{\mathcal{R}_\lambda(\mathbf{x})}_{\text{regularisation with weight } \lambda}$$

$\mathcal{R}_\lambda$: regularisation, imposes regularity, makes problem well-posed

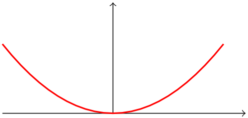
d : ensures reconstruction close to the data
(its choice is dictated by the statistics of the noise, e.g. L^2 for Gaussian)

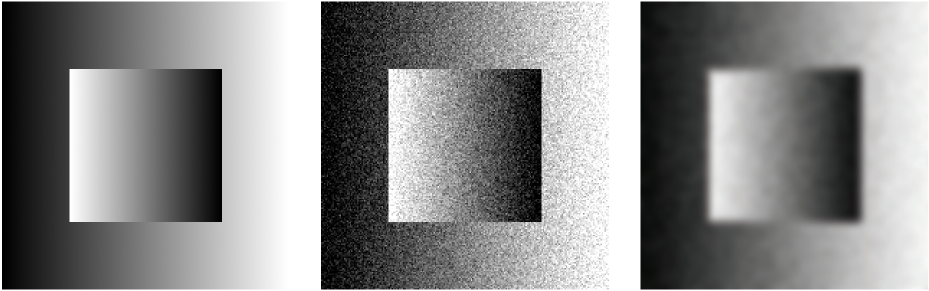
λ : balances the effect of data discrepancy and regularisation

Regularisation functionals

Tikhonov:
[Tikhonov (1963)]

$$\mathcal{R}_\lambda(\mathbf{x}) = \frac{\lambda}{2}\|\nabla \mathbf{x}\|_2^2$$

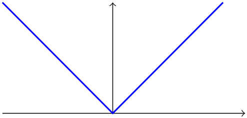


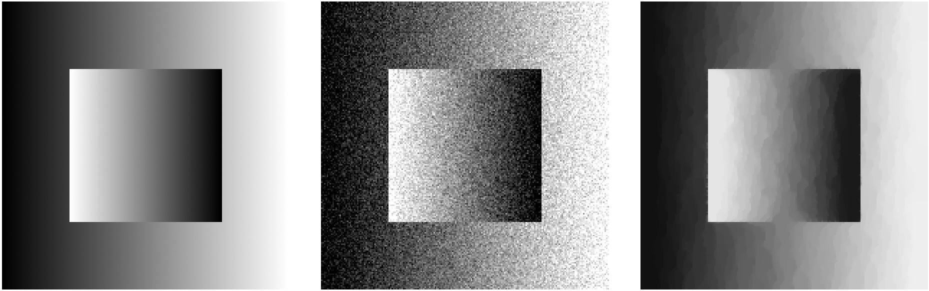


Regularisation functionals

Total Variation:
[Rudin, Osher, Fatemi (1992)]

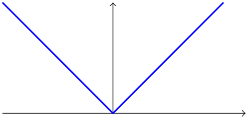
$$\mathcal{R}_\lambda(\mathbf{x}) = \mathrm{TV}(\mathbf{x}) = \lambda \|\nabla \mathbf{x}\|_1$$





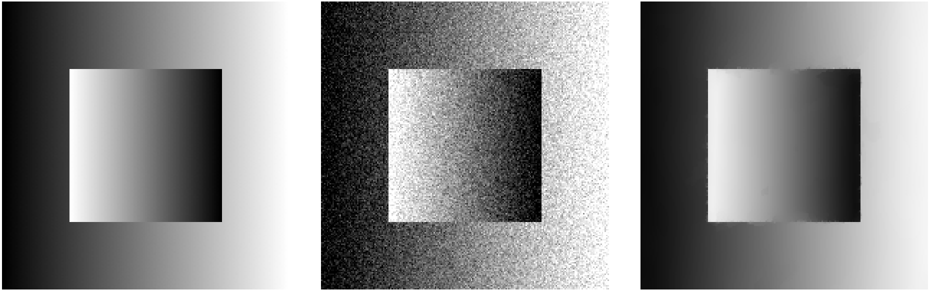
Total Generalised Variation:

[Bredies, Kunisch, Pock (2010)]



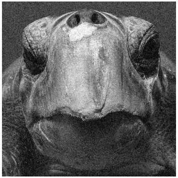
$$\mathrm{TGV}(\mathbf{x}) = \min_{\mathbf{w}} \ \lambda_1 \|\nabla \mathbf{x} - \mathbf{w}\|_1 + \lambda_2 \|\mathcal{E}\mathbf{w}\|_1$$

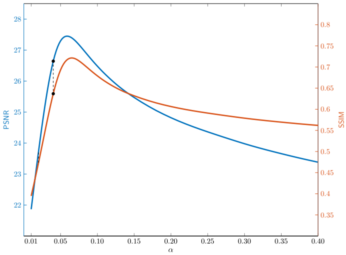
$\mathcal{E}\mathbf{w}$: symmetrised gradient, $\left(\frac{\nabla \mathbf{w} + (\nabla \mathbf{w})^T}{2}\right)$



The influence of regularisation parameters

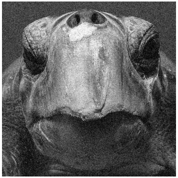
$$\min_{\mathbf{x}} \; \frac{1}{2}\|\mathbf{A}\mathbf{x}-\mathbf{z}\|_2^2 + \lambda\|\nabla\mathbf{x}\|_1, \quad \lambda > 0$$

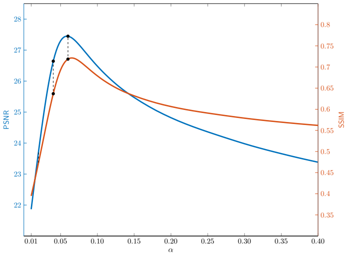




The influence of regularisation parameters

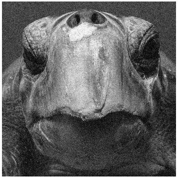
$$\min_{\mathbf{x}} \ \frac{1}{2}\|\mathbf{Ax}-\mathbf{z}\|_2^2 + \lambda\|\nabla\mathbf{x}\|_1, \quad \lambda > 0$$

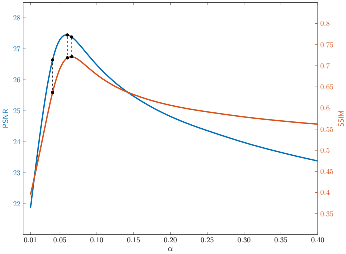




The influence of regularisation parameters

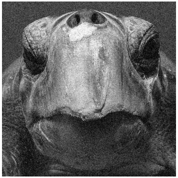
$$\min_{\mathbf{x}} \; \frac{1}{2}\|\mathbf{A}\mathbf{x} - \mathbf{z}\|_2^2 + \lambda\|\nabla\mathbf{x}\|_1, \quad {\color{red}\lambda > 0}$$

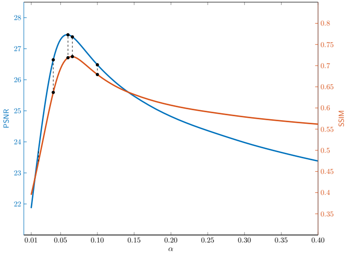




The influence of regularisation parameters

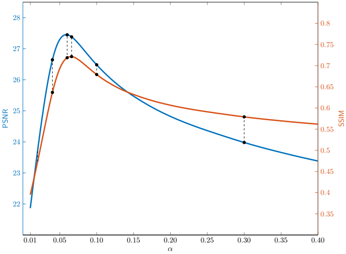
$$\min_{\mathbf{x}} \frac{1}{2}\|\mathbf{A}\mathbf{x} - \mathbf{z}\|_2^2 + \lambda\|\nabla\mathbf{x}\|_1, \quad \lambda > 0$$





The influence of regularisation parameters

$$\min_{\mathbf{x}} \ \frac{1}{2}\|\mathbf{A}\mathbf{x} - \mathbf{z}\|_2^2 + \lambda\|\nabla\mathbf{x}\|_1, \quad {\color{red}\lambda > 0}$$



Regularisation functionals with spatially varying parameters

Scalar Total Variation

$\lambda \in \mathbb{R}$, $\lambda > 0$

$\mathrm{TV}(\mathbf{x}) = \ \lambda \|\mathbf{x}\|_1$

Spatially varying Total Variation:

$\mathbf{\Lambda} \in \mathbb{R}^{qn}$, $\mathbf{\Lambda} > 0$

$\mathrm{TV}(\mathbf{x}) = \ \|\mathbf{\Lambda}\nabla\mathbf{x}\|_1$

Regularisation functionals with spatially varying parameters

Spatially varying Total Variation:

$\mathbf{\Lambda} \in \mathbb{R}^{qn}$, $\mathbf{\Lambda} > 0$

$$\mathrm{TV}(\mathbf{x}) = \|\mathbf{\Lambda}\nabla\mathbf{x}\|_1$$

[Strong, Chan, 2005]

[Dong, Hintermüller, Rincon-Camacho, 2011]

[Jalalzai, 2014]

[Calatroni, Cao, De los Reyes, Schönlieb, Valkonen, 2017]

[Hintermüller, Rautenberg, Wu, Langer, 2017]

[Hintermüller, Papafitsoros, Rautenberg, 2017]

[Dong, Schönlieb, 2020]

[De los Reyes, Villacís, 2022]

[Pragliola, Calatroni, Lanza, Sgallari, 2021-22]

$\rightarrow$ Mostly for denoising and static 2D inverse problems...

How to choose the parameters λ, $\mathbf{\Lambda}$?

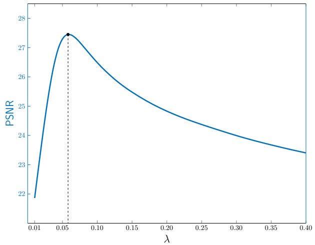
Bilevel optimization – supervised learning

"Supervised learning" approach using ground truth-degraded data pairs:

[De los Reyes, Schönlieb, Pock, Kunisch, Calatroni, Valkonen, Villacis...(2013-)]

$$\min_{\mathbf{x},\, \lambda,\, (\text{or } \mathbf{\Lambda})} \frac{1}{2}\|\mathbf{x} - \mathbf{x}_{true}\|_2^2$$

s.t. $\mathbf{x}$ solves the TV (or TGV) problem with parameters λ $\left(\text{or } \mathbf{\Lambda}\right)$

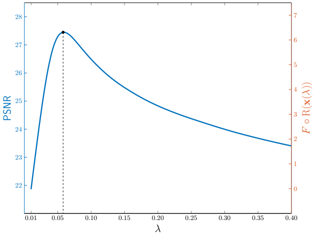


How to choose the parameters without using u_{true}?

Bilevel optimisation – supervised learning

$$\min_{\mathbf{x},\, \lambda,\, (\text{or } \boldsymbol{\Lambda})} \frac{1}{2}\|\mathbf{x} - \mathbf{x}_{true}\|^2_{L^2(\Omega)}$$

s.t. $\mathbf{x}$ solves the TV (or TGV) problem with parameters λ (or $\boldsymbol{\Lambda}$)



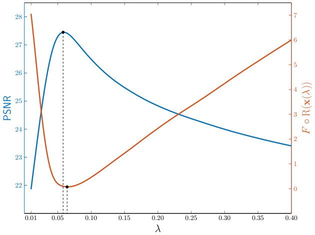
Bilevel optimisation – unsupervised learning

"Unsupervised learning" without using ground truth-degraded data pairs:

[Hintermüller, Papafitsoros, Rautenberg, Wu, Langer, Sun,...(2017-)]

$$\min_{\mathbf{x},\,\lambda,\,(\text{or }\boldsymbol{\Lambda})} F(R\mathbf{x}) := \frac{1}{2}\|\max(R\mathbf{x}-\overline{\sigma}^2,0)\|_2^2+\frac{1}{2}\|\min(R\mathbf{x}-\underline{\sigma}^2,0)\|_2^2$$

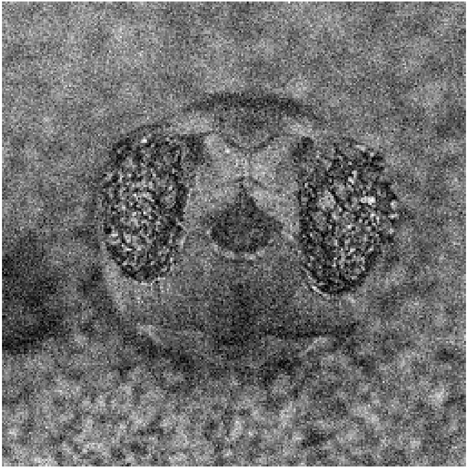
s.t. $\mathbf{x}$ solves the TV (or TGV) problem with parameters λ (or $\boldsymbol{\Lambda}$)



Numerical results: bilevel TV unsupervised learning

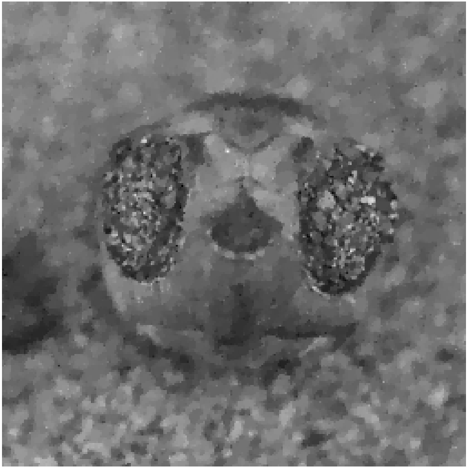
scalar TV, PSNR=27.75, SSIM=0.7701
bilevel TV, PSNR=27.52, SSIM=0.7722

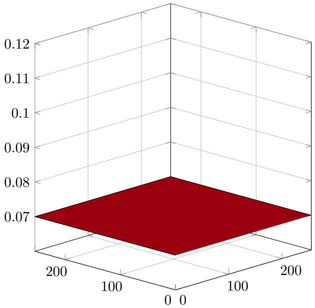
Numerical results: bilevel TV unsupervised learning



scalar TV, PSNR=27.75, SSIM=0.7701
bilevel TV, PSNR=27.52, SSIM=0.7722

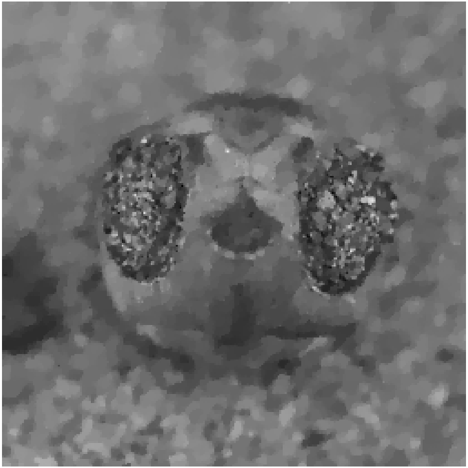
Numerical results: bilevel TV unsupervised learning

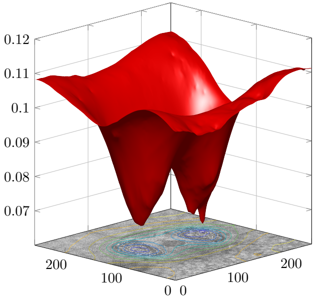




scalar TV, PSNR=27.75, SSIM=0.7701
bilevel TV, PSNR=27.52, SSIM=0.7722

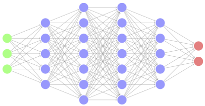
Numerical results: bilevel TV unsupervised learning





scalar TV, PSNR=27.75, SSIM=0.7701
bilevel TV, PSNR=27.52, SSIM=0.7722

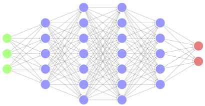
Deep learning-based methods



$$\underset{\Theta}{\text{minimize}} \quad \frac{1}{M}\sum_{i=1}^{M} l(\mathcal{N}_\Theta(\mathbf{z}^i), \mathbf{x}^i_{true}) + r(\Theta)$$

- Application of deep learning to inverse problems and mathematical imaging has seen a tremendous development over the last years [Arridge, Maass, Öktem, Schönlieb, 2019]
- Fewer theoretical underpinnings; instabilities; challenging interpretation [Antun, Renna, Poon, Adcock, Hansen, 2020]

Deep learning-based methods



$$\underset{\Theta}{\text{minimize}} \quad \frac{1}{M}\sum_{i=1}^{M} l(\mathcal{N}_\Theta(\mathbf{z}^i), \mathbf{x}^i_{true}) + r(\Theta)$$

- Application of deep learning to inverse problems and mathematical imaging has seen a tremendous development over the last years [Arridge, Maass, Öktem, Schönlieb, 2019]
- Fewer theoretical underpinnings; instabilities; challenging interpretation [Antun, Renna, Poon, Adcock, Hansen, 2020]

Combine

Interepretability of model-based methods

Flexibility and adaptability of deep learning

Combining deep learning and model-based methods

- Learning the regularization functional $\mathcal{R}$ from data and embed it into a variational data fitting/regularisation functional scheme
  [Lunz, Öktem, Schönlieb, 2018]
  [Li, Schwab, Antholzer, Haltmeier, 2020]
- Enforcing the reconstruction to be close to an output of a network via $\mathcal{R}$, e.g. $\mathcal{R}(\cdot) = \| \cdot - u_\Theta(\mathbf{s})\|_2^2$ for some network with trainable parameters Θ and input $\mathbf{s}$
  [Duff, Campbell, Ehrhardt, 2021]
  [Kofler, Haltmeier, Schaeffter, Kachelriess, Dewey, Wald, Kolbitsch, 2020]
- Substituting proximal operators in classical iterativeschemes by learned NN denoisers (in a "plug-and-play" fashion)
  [Meinhardt, Moller, Hazirbas, Cremers, 2017]
  [Romano, Elad, and Milanfar, 2017]
- Using learned iterative schemes
  [Adler, Öktem, 2017-18]
  [Hammernik, Klatzer, Kobler, Recht, Sodickson, Pock, Knoll, 2018]

Reviews:

[Arridge, Maass, Öktem, Schönlieb, 2019], [Monga, Li, Eldar, 2021], [Mukherjee, Hauptmann, Öktem, Pereyra, Schönlieb, 2023], [Kamilov, Bouman, Buzzard, Wohlberg, 2023]

Deep learning-based parameter learning

[Afkham, Chung, Chung, 2019], [Nekhili, Descombes, Calatroni, 2022]

Offline phase

- Given $(\mathbf{z}_i, \mathbf{x}^i_{true})_{i=1}^M$ employ, e.g., a bilevel scheme to compute optimal $(\lambda_i)_{i=1}^M$

- Use training data $(\lambda_i, \mathbf{z}_i)_{i=1}^M$ to train a neural network $\mathcal{N}_\Theta$

$$\min_\Theta \mathcal{L}(\Theta) := \frac{1}{M} \sum_{i=1}^M l(\mathcal{N}_\Theta(\mathbf{z}_i), \lambda_i)$$

Online phase

- Given some new data $\mathbf{z}_{test}$ compute $\lambda_\Theta = \mathcal{N}_\Theta(\mathbf{z}_{test})$ and solve

$$\min_{\mathbf{x}} \frac{1}{2} \|\mathbf{A}\mathbf{x} - \mathbf{z}_{test}\|_2^2 + \lambda_\Theta \|\nabla \mathbf{x}\|_1$$

Deep learning-based parameter learning

[Afkham, Chung, Chung, 2019], [Nekhili, Descombes, Calatroni, 2022]

Offline phase

- Given $(\mathbf{z}_i, \mathbf{x}^i_{true})_{i=1}^M$ employ, e.g., a bilevel scheme to compute optimal $(\lambda_i)_{i=1}^M$
  Computationally intensive, esp. 3D and dynamic, not generalisable...
- Use training data $(\lambda_i, \mathbf{z}_i)_{i=1}^M$ to train a neural network $\mathcal{N}_\Theta$

$$\min_\Theta \mathcal{L}(\Theta) := \frac{1}{M} \sum_{i=1}^{M} l(\mathcal{N}_\Theta(\mathbf{z}_i), \lambda_i)$$

Online phase

- Given some new data $\mathbf{z}_{test}$ compute $\lambda_\Theta = \mathcal{N}_\Theta(\mathbf{z}_{test})$ and solve

$$\min_{\mathbf{x}} \frac{1}{2} \|\mathbf{A}\mathbf{x} - \mathbf{z}_{test}\|_2^2 + \lambda_\Theta \|\nabla \mathbf{x}\|_1$$

A deep unrolled scheme

$$\mathbf{z} = A\mathbf{x}_{true} + \mathbf{e}$$

$$\min_{\mathbf{x}} \frac{1}{2}\|\mathbf{A}\mathbf{x} - \mathbf{z}\|_2^2 + \|\mathbf{\Lambda}\nabla\mathbf{x}\|_1, \quad (P)$$

NET_Θ : overparameterised CNN with parameters Θ, e.g. U-net

S : an iterative scheme that solves (P), e.g. PDHG [Chambolle, Pock, 2011]

$$\mathbf{x}_T = S^T(\mathbf{x}_0, \mathbf{z}, \mathbf{\Lambda}, \mathbf{A}), \quad \mathbf{x}_T \to S^*(\mathbf{z}, \mathbf{\Lambda}, \mathbf{A}) \text{ as } T \to \infty$$

Fix $T \in \mathbb{N}$

$$\begin{cases} \mathbf{\Lambda}_\Theta = \mathrm{NET}_\Theta(\mathbf{x}_0) \\ \mathbf{x}_1 = S^1(\mathbf{x}_0, \mathbf{z}, \mathbf{\Lambda}_\Theta, \mathbf{A}) \\ \mathbf{x}_2 = S^2(\mathbf{x}_0, \mathbf{z}, \mathbf{\Lambda}_\Theta, \mathbf{A}) \\ \quad\vdots \\ \mathbf{x}_T = S^T(\mathbf{x}_0, \mathbf{z}, \mathbf{\Lambda}_\Theta, \mathbf{A}) \end{cases}$$

$$\mathbf{x}_T = \mathcal{N}_{\boldsymbol{\Theta}}^{\boldsymbol{T}}(\mathbf{x}_0)$$

A deep unrolled scheme

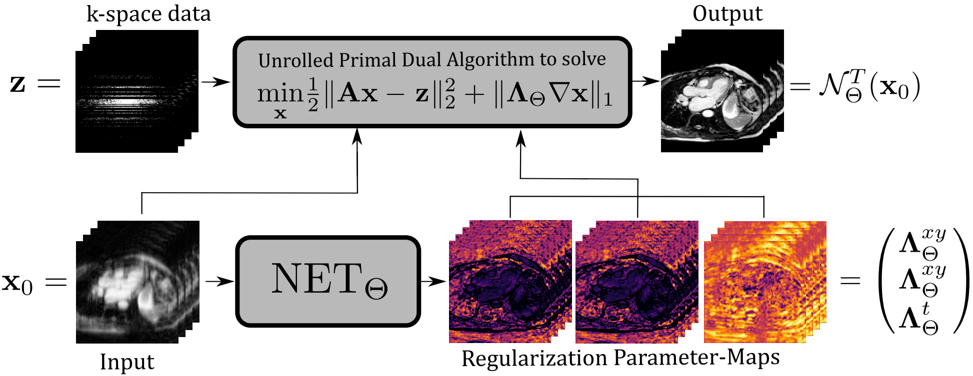
$$\begin{aligned}\mathcal{N}_\Theta^T(\mathbf{x}_0) &= S^T(\mathbf{x}_0, \mathbf{z}, \boldsymbol{\Lambda}_\Theta, \mathbf{A}) \\ &= S^T(\mathbf{x}_0, \mathbf{z}, \mathrm{NET}_\Theta(\mathbf{x}_0), \mathbf{A})\end{aligned}$$

End-to-end training

Set of training pairs $\mathcal{D} = \{(\mathbf{x}_0^i, \mathbf{x}_{\mathrm{true}}^i)_{i=1}^M : \mathbf{x}_0^i = A^* \mathbf{z}_i, \mathbf{z}_i := \mathbf{A}\mathbf{z}_{true}^i + \mathbf{e}_i\}$

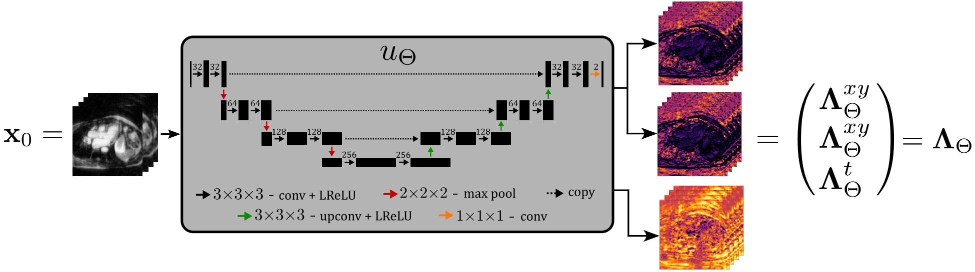
$$\min_\Theta \mathcal{L}(\Theta) := \frac{1}{M} \sum_{i=1}^M l\big(\mathcal{N}_\Theta^T(\mathbf{x}_0^i)\,,\, \mathbf{x}_{\mathrm{true}}^i\big) + r(\Theta)$$

A deep unrolled scheme



$$\mathrm{TV}_{\mathbf{\Lambda}^{xy,t}}(\mathbf{x}) := \sum_z \mathbf{\Lambda}^{xy}(z)|\nabla_x \mathbf{x}(z)| + \mathbf{\Lambda}^{xy}(z)|\nabla_y \mathbf{x}(z)| + \mathbf{\Lambda}^{t}(z)|\nabla_t \mathbf{x}(z)|$$

A deep unrolled scheme



$$\mathrm{TV}_{\mathbf{\Lambda}^{xy,t}}(\mathbf{x}) := \sum_z \mathbf{\Lambda}^{xy}(z)|\nabla_x \mathbf{x}(z)| + \mathbf{\Lambda}^{xy}(z)|\nabla_y \mathbf{x}(z)| + \mathbf{\Lambda}^{t}(z)|\nabla_t \mathbf{x}(z)|$$

A deep unrolled scheme: Consistency analysis

$$\min_{\Theta \in \mathbb{R}^\ell} \mathcal{L}^T(\Theta) := \frac{1}{M} \sum_{i=1}^{M} l(\mathcal{N}_\Theta^T(\mathbf{x}_0^i), \mathbf{x}_{true}^i) + r(\Theta)$$

$$\mathcal{N}_\Theta^T(\mathbf{x}_0) := S^T(\mathbf{x}_0, \mathbf{z}, \mathbf{\Lambda}_\Theta(\mathbf{x}_0))$$

$$\min_{\Theta \in \mathbb{R}^\ell} \mathcal{L}^*(\Theta) := \frac{1}{M} \sum_{i=1}^{M} l(\mathcal{N}_\Theta^*(\mathbf{x}_0^i), \mathbf{x}_{true}^i) + r(\Theta)$$

$$\mathcal{N}_\Theta^*(\mathbf{x}_0) := S^*(\mathbf{z}, \mathbf{\Lambda}_\Theta(\mathbf{x}_0))$$

Unrolled PDHG algorithm [Chambolle, Pock, 2011]

Input: $L = \|\mathbf{K}\|$, $\tau\sigma \leq 1/L^2$, $\theta = 1$, initial guess $\mathbf{x}_0$
Parameters: number of iterations $T > 0$
Output: reconstructed image $\mathbf{x}_T$

1: $\bar{\mathbf{x}}_0 = \mathbf{x}_0$
2: $\mathbf{y}_0 = \mathbf{0}$
3: **for** $k < T$ **do**
4: $\quad \mathbf{y}_{k+1} = \mathrm{prox}_{\sigma f^*}(\mathbf{y}_k + \sigma \mathbf{K}\bar{\mathbf{x}}_k)$
5: $\quad \mathbf{x}_{k+1} = \mathrm{prox}_{\tau g}(\mathbf{x}_k - \tau \mathbf{K}^{\mathsf{T}} \mathbf{y}_{k+1})$
6: $\quad \bar{\mathbf{x}}_{k+1} = \mathbf{x}_{k+1} + \theta(\mathbf{x}_{k+1} - \mathbf{x}_k)$
7: **end for**

$$\min_{\mathbf{x} \in X} f(\mathbf{K}\mathbf{x}) + g(\mathbf{x})$$

$$f(\mathbf{y}) = f(\mathbf{p}, \mathbf{q}) := f_1(\mathbf{p}) + f_2(\mathbf{q}) = \frac{1}{2}\|\mathbf{p} - \mathbf{z}\|_2^2 + \|\boldsymbol{\Lambda}_\Theta\, \mathbf{q}\|_1, \ \mathbf{K} := \begin{bmatrix} \mathbf{A} \\ \nabla \end{bmatrix}, \ g(\mathbf{x}) := \mathbf{0}$$

A deep unrolled scheme: Consistency analysis

Assumptions:

(i) The operator $\mathbf{A} : X \to Z$ is injective

(ii) The fidelity term $d(\cdot, \mathbf{z})$ is $\mu_{\mathbf{z}}$-strongly convex and Lipschitz continuously differentiable for every $\mathbf{z} \in Z$

(iii) The parameters $\sigma, \tau > 0$ in PDHG are small enough such that the matrix

$$\mathbf{M} = \begin{pmatrix} \frac{1}{\tau}\mathbf{I} & -\mathbf{K}^{\mathsf{T}} \\ -\mathbf{K} & \frac{1}{\sigma}\mathbf{I} \end{pmatrix}$$

is symmetric, positive definite, defining a norm $\|\cdot\|_{\mathbf{M}}$

(iv) The regularization function $r : \mathbb{R}^{\ell} \to \overline{\mathbb{R}} := \mathbb{R} \cup \{+\infty\}$ is proper and lower semicontinuous.

(v) The loss function $l : X \times X \to \mathbb{R}$ is continuous.

(vi) The activation functions in the NET_{Θ} are continuous.

Proposition. The following statements hold:

(i) The solution map $\mathbf{\Lambda} \mapsto S^*(\mathbf{z}, \mathbf{\Lambda})$ is Lipschitz for every $\mathbf{z} \in Z$. In particular the following bound holds for every $\mathbf{\Lambda}_1, \mathbf{\Lambda}_2 \in \mathbb{R}^{qn}_+$,

$$\|S^*(\mathbf{\Lambda}_1, \mathbf{z}) - S^*(\mathbf{\Lambda}_2, \mathbf{z})\|_2 \leq \frac{2\|\nabla\|}{\lambda_{\min}(\mathbf{A}^\intercal \mathbf{A})\mu_{\mathbf{z}}} \|\mathbf{\Lambda}_1 - \mathbf{\Lambda}_2\|_2.$$

(ii) The map $\mathbf{\Lambda} \mapsto S^T(\mathbf{x}_0, \mathbf{z}, \mathbf{\Lambda})$ is Lipschitz for every $\mathbf{z} \in Z$, $\mathbf{x}_0 \in X$ and $T \in \mathbb{N}$.

(iii) For $\|\mathbf{\Lambda}\|_2 \leq \overline{\mathbf{\Lambda}}$ we obtain the following sub-linear rate, for $\mathbf{v}_0 := (\mathbf{x}_0, \mathbf{y}_0)$

$$\|S^T(\mathbf{x}_0, \mathbf{z}, \mathbf{\Lambda}) - S^*(\mathbf{z}, \mathbf{\Lambda})\|_2 \leq \frac{3C_{\mathbf{z},\mathbf{A}}}{T^{1/4}} \left(1 + \|\mathbf{v}_0 - \mathbf{v}^*(\mathbf{\Lambda}, \mathbf{z})\|_{\mathbf{M}}\right),$$

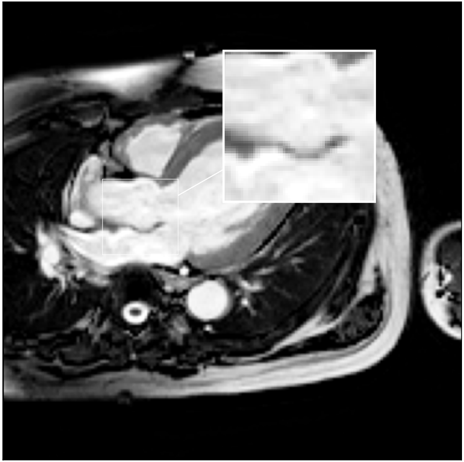
(iv) Whenever $\mathbf{\Lambda}_T \to \mathbf{\Lambda}$ as $T \to \infty$, it holds $S^T(\mathbf{x}_0, \mathbf{z}, \mathbf{\Lambda}_T) \to S^*(\mathbf{z}, \mathbf{\Lambda})$ for every $\mathbf{x}_0 \in X$, $\mathbf{z} \in Z$.

Theorem

Let the training set $\mathcal{D}$ be fixed and consider the sequence of functionals $\mathcal{L}^T : \mathbb{R}^\ell \to \overline{\mathbb{R}}$, $T \in \mathbb{N}$, as well as $\mathcal{L}^* : \mathbb{R}^\ell \to \overline{\mathbb{R}}$ defined as before. Then we have that $\mathcal{L}^T$ Γ-converges to $\mathcal{L}^*$ as $T \to \infty$.

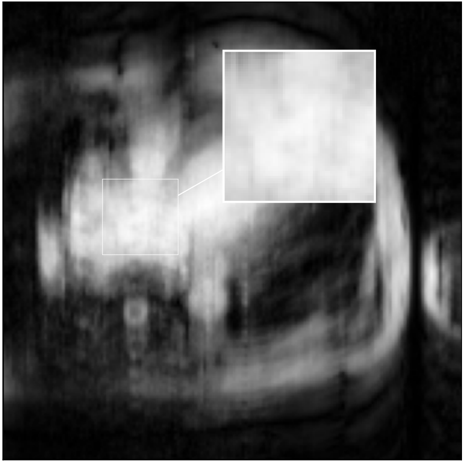
Let $\epsilon_T \to 0$. Suppose that Θ_T is an ϵ_T-minimizer of $\mathcal{L}^T$ i.e. $\mathcal{L}^T(\Theta_T) \le \inf_{\Theta \in \mathbb{R}^\ell} \mathcal{L}^T(\Theta) + \epsilon_T$. Then, if Θ is an accumulation point of $(\Theta_T)_{T \in \mathbb{N}}$ it is a minimizer of $\mathcal{L}^*$ and $\mathcal{L}^*(\Theta) = \limsup_{T \to \infty} \mathcal{L}^T(\Theta_T)$.

Application to dynamic MRI



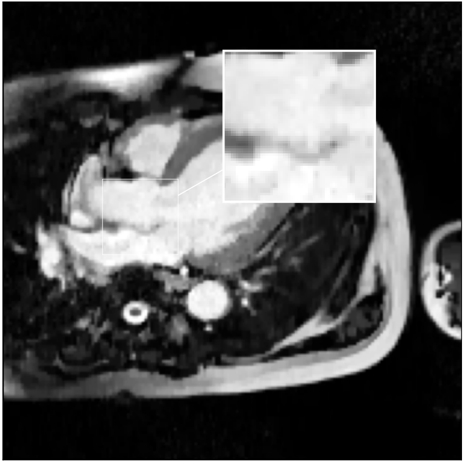
Ground truth

Application to dynamic MRI

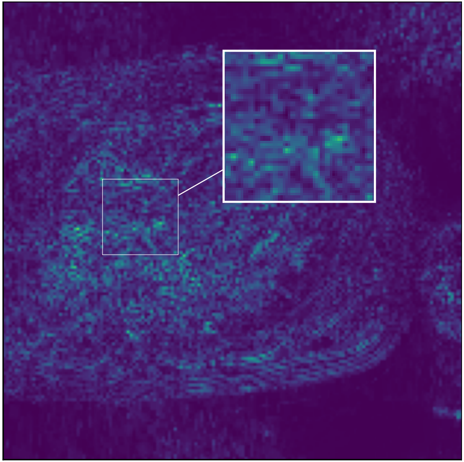


zero filled

Application to dynamic MRI

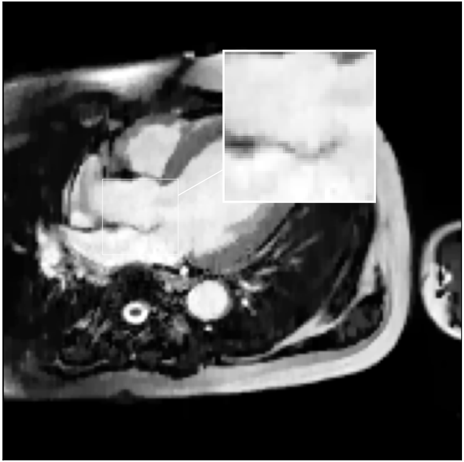


Reconstruction, best scalar λ w.r.t. PSNR

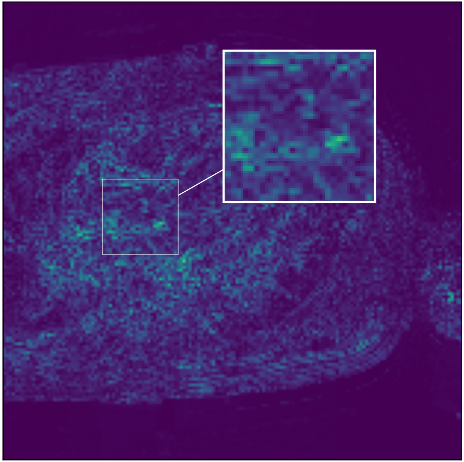


Difference to the ground truth

Application to dynamic MRI

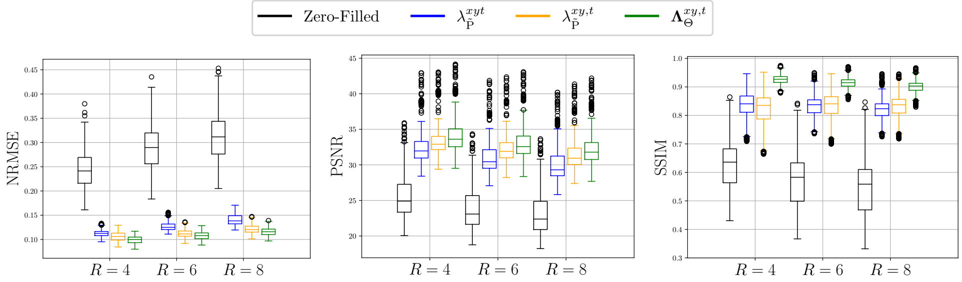


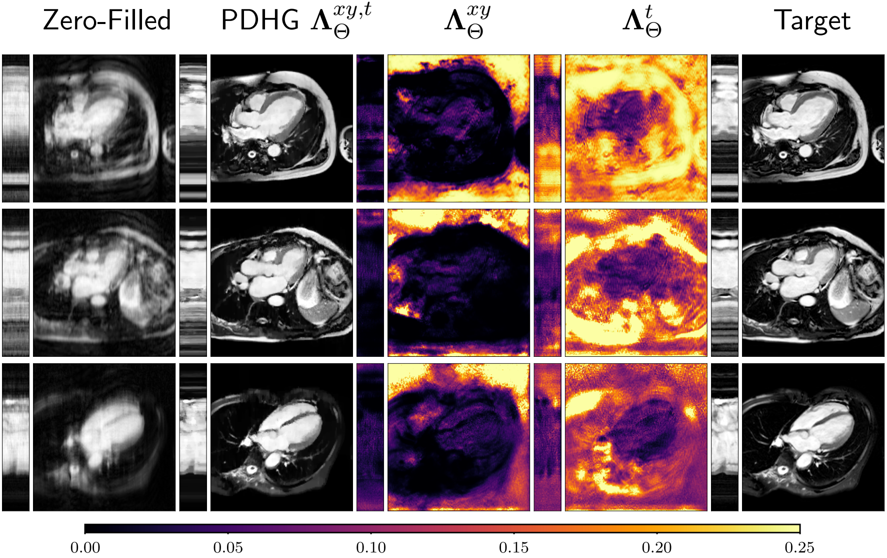
Reconstruction with $\mathbf{\Lambda}_\Theta$



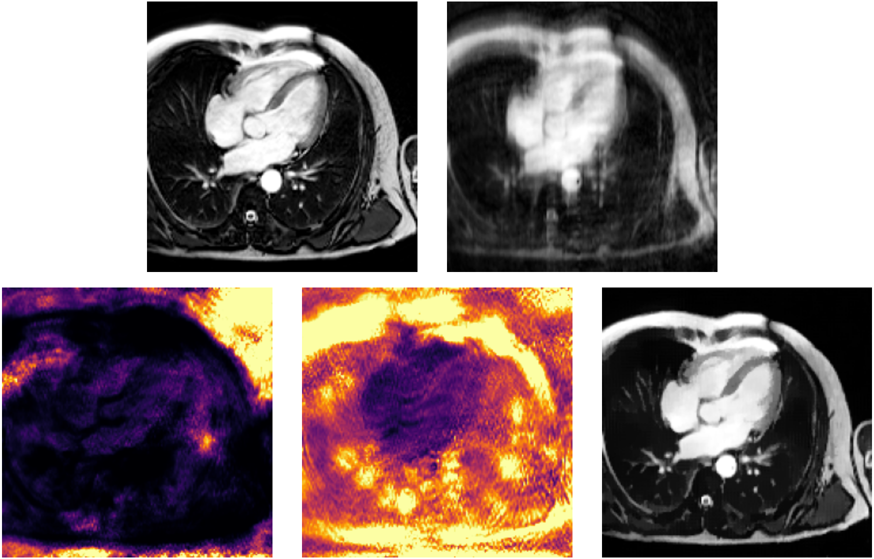
Difference to the ground truth

Application to dynamic MRI

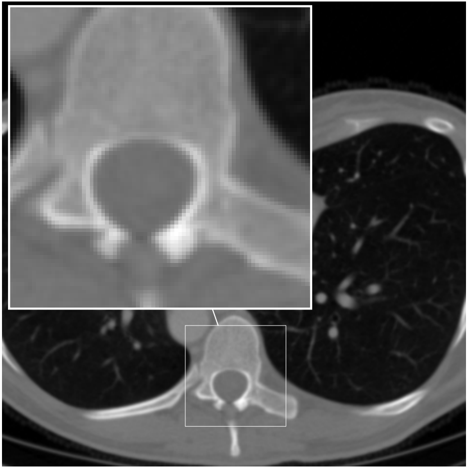



Zero-Filled
PDHG $\mathbf{\Lambda}_{\Theta}^{xy,t}$
$\mathbf{\Lambda}_{\Theta}^{xy}$
$\mathbf{\Lambda}_{\Theta}^{t}$
Target
0.00
0.05
0.10
0.15
0.20
0.25

Application to dynamic MRI

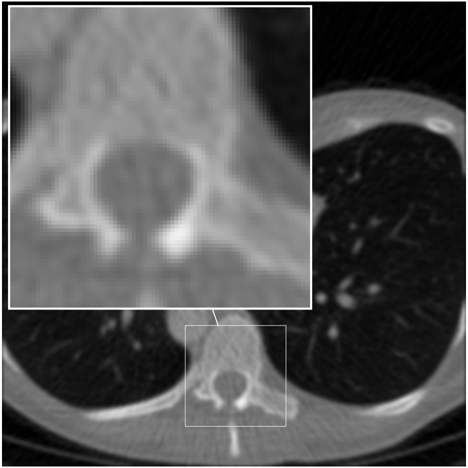


Application to Computerized Tomography



Ground truth

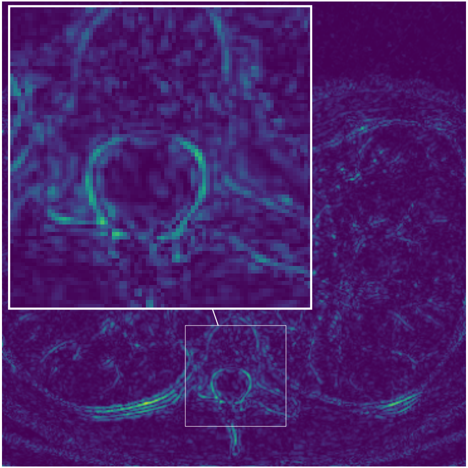
Application to Computerized Tomography



FBP

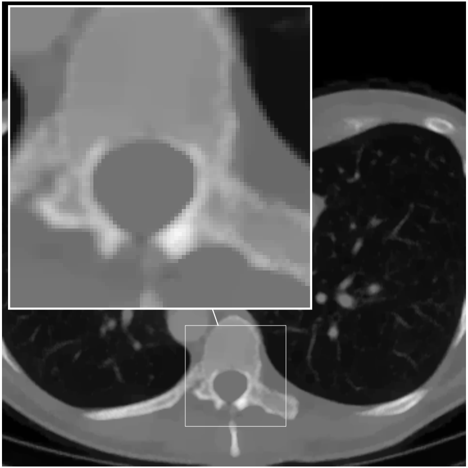
Application to Computerized Tomography

Reconstruction, best scalar λ w.r.t. PSNR

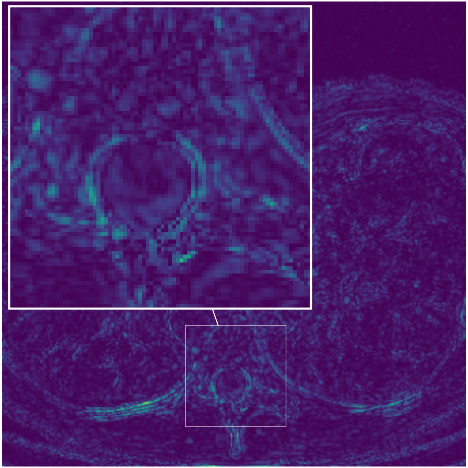


Difference to the ground truth

Application to Computerized Tomography



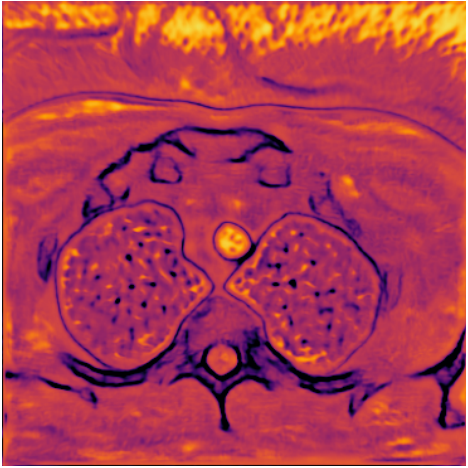
Reconstruction with Λ_Θ



Difference to the ground truth

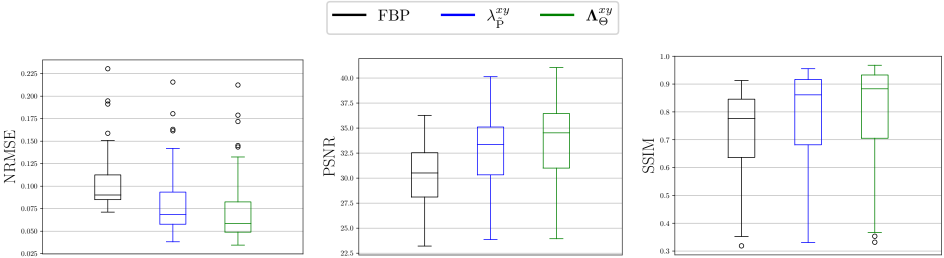
Application to Computerized Tomography

Reconstruction with $\mathbf{\Lambda}_\Theta$

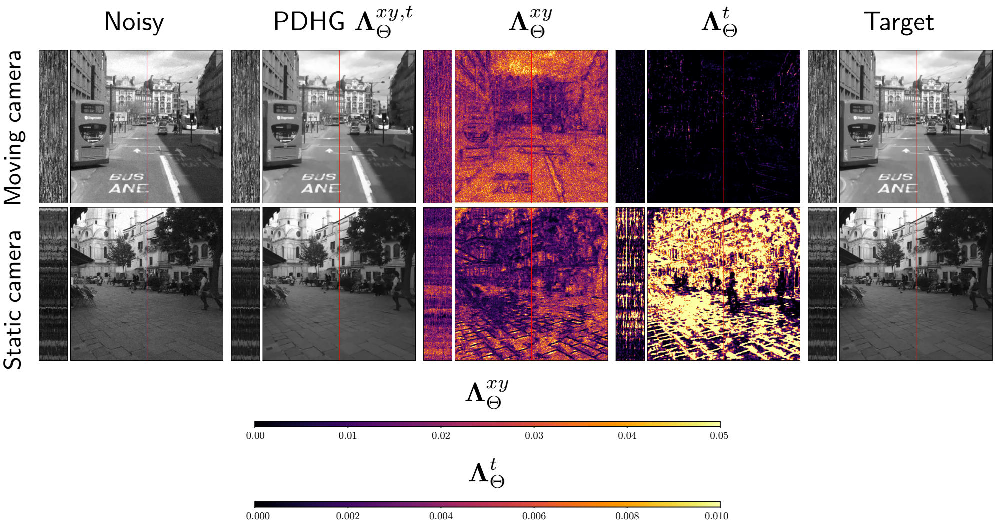


$\mathbf{\Lambda}_\Theta$ produced by unrolled scheme

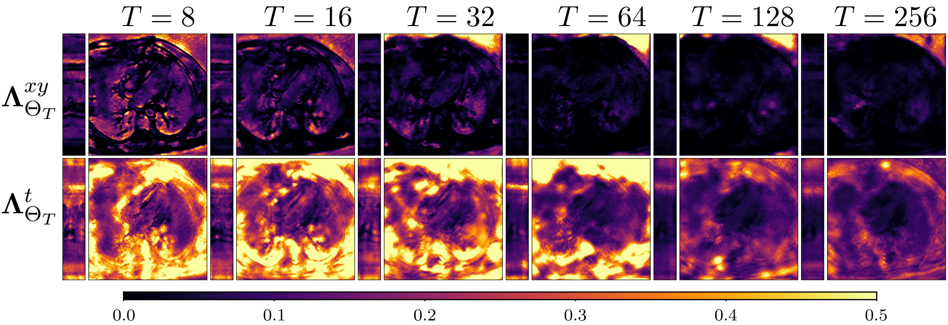
Application to Computerized Tomography



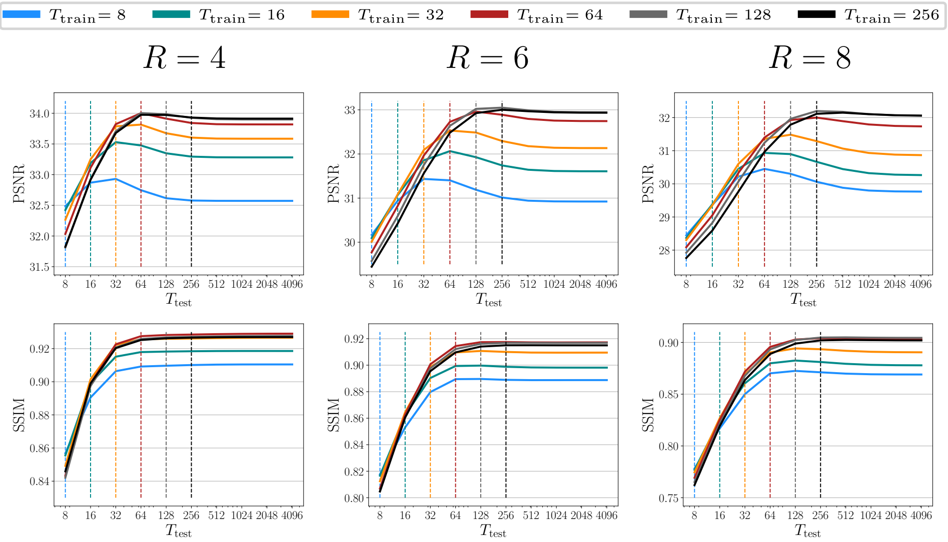
Application to dynamic denoising



Choosing T at training and test times



Choosing T at training and test times



Summary

→ Simple yet efficient data-driven approach to automatically select *spatio-temporal* dependent regularization parameter-maps for the variational image reconstruction

→ Interpretable algorithm since the “black-box” nature of the CNN is placed entirely on the regularization parameter and not to the image itself

→ Consistency analysis in terms of ϵ-minimizers

Open questions:

- Consistency results in terms of stationary points....?
- Investigate theoretically in what degree CNN-produced artefacts in the parameter-maps can affect or create artefacts to the corresponding reconstructions
- Avoid using large T, computing hyper-gradients, deep equilibrium models...

A highly collaborative effort!

- **Learning regularization parameter-maps for variational image reconstruction using deep neural networks and algorithm unrolling**, submitted, arXiv:2301.05888, 2023
- **Unrolled three-operator splitting for parameter-map learning in low dose X-ray CT reconstruction**, Fully3D, 2023
- **CNN-based estimation of spatio-temporal regularization parameter-maps for TV-reconstruction in dynamic cardiac MRI**, ISMRM, 2023

Andreas Kofler, Fabian Altekrüger, Fatima Antarou Ba, Christoph Kolbitsch, Evangelos Papoutsellis, David Schote, Clemens Sirotenko, Felix Zimmermann, Kostas Papafitsoros

MATHS MEETS IMAGE

Hackathon on image reconstruction, segmentation and shape analysis

17 – 19 March 2022 at The Classroom, Berlin

Thank you for your attention

http://kostaspapafitsoros.weebly.com