LEARNED IMAGE RECONSTRUCTION FOR HIGH RESOLUTION TOMOGRAPHIC IMAGING

Andreas Hauptmann

University of Oulu

Research Unit of Mathematical Sciences

&

University College London

Centre for Medical Image Computing

Department of Computer Science

2019 AAPM Annual Meeting, San Antonio, 18 July 2019

Finnish Centre of Excellence in
Inverse Modelling and Imaging
2018-2025

CENTRES OF EXCELLENCE
IN RESEARCH

UNIVERSITY OF OULU

UNIVERSITY OF JYVÄSKYLÄ

TAMPERE UNIVERSITY OF TECHNOLOGY

FINNISH METEOROLOGICAL INSTITUTE

UNIVERSITY OF EASTERN FINLAND

LUT
Lappeenranta
University of Technology

UNIVERSITY OF HELSINKI

Aalto University

Simon Arridge, Paul Beard, Marta Betcke, Ben Cox, and Nam Huyhn, Vivek Muthurangu, and Jennifer Steeden (UCL), Felix Lucka (CWI), Jonas Adler and Ozan Öktem (KTH)

PROBLEM FORMULATION

We consider the setting with:

True unknown $x_{true} \in X$
Measured data $y \in Y$
Forward operator $\mathcal{A} : X \to Y$

Then the measurement is given by

$$y = \mathcal{A}(x_{true}) + \delta y.$$

Given the measurement y we aim to find a mapping $\mathcal{F}^\dagger : Y \to X$, such that

$$\mathcal{F}^\dagger(y) \approx x_{true}.$$

We aim to parametrise the mapping $\mathcal{F}_\theta^\dagger : Y \to X$, with some parameter set θ.

$\Rightarrow$ Find an optimal set of parameters, such that

$$\mathcal{F}_\theta^\dagger(y) \approx x_{true}.$$

MODEL ENFORCED RECONSTRUCTION

Direct reconstruction followed by learning based post-processing.

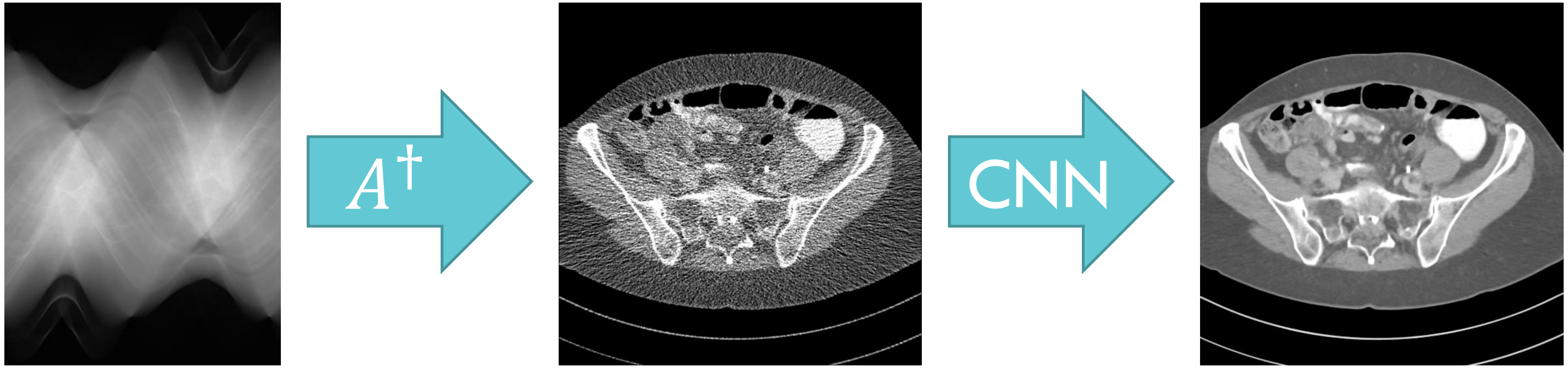
Reconstruction is carried out using a simple/fast inversion step.

Post-processing is used to remove artefacts and noise.

Given a reconstruction operator $\mathcal{A}^\dagger : Y \to X$, then we apply

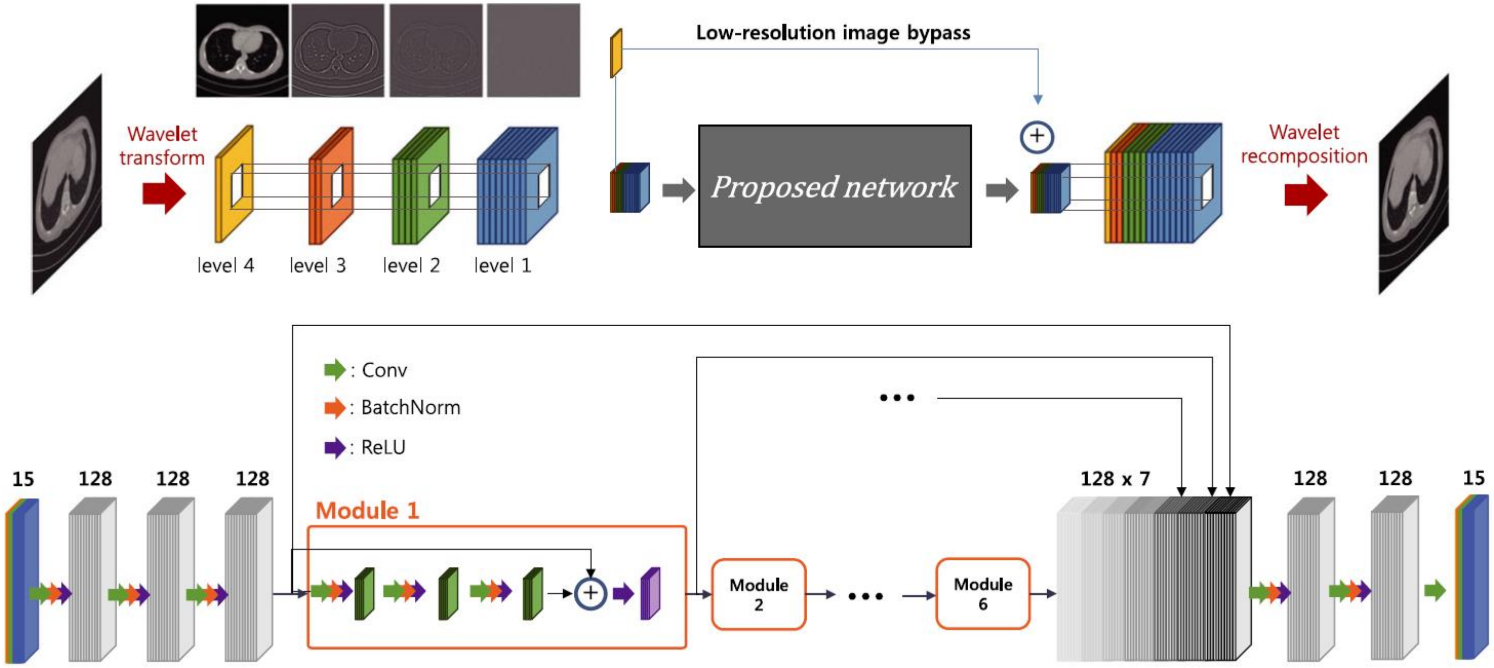
$$\mathcal{F}_\theta^\dagger = \mathcal{G}_\theta \circ \mathcal{A}^\dagger$$

where $\mathcal{G}_\theta : X \to X$ is typically a sophisticated CNN.



A DEEP CNN FOR LOW-DOSE X-RAY CT
[KANG ET AL., *MEDICAL PHYSICS*, 2017]

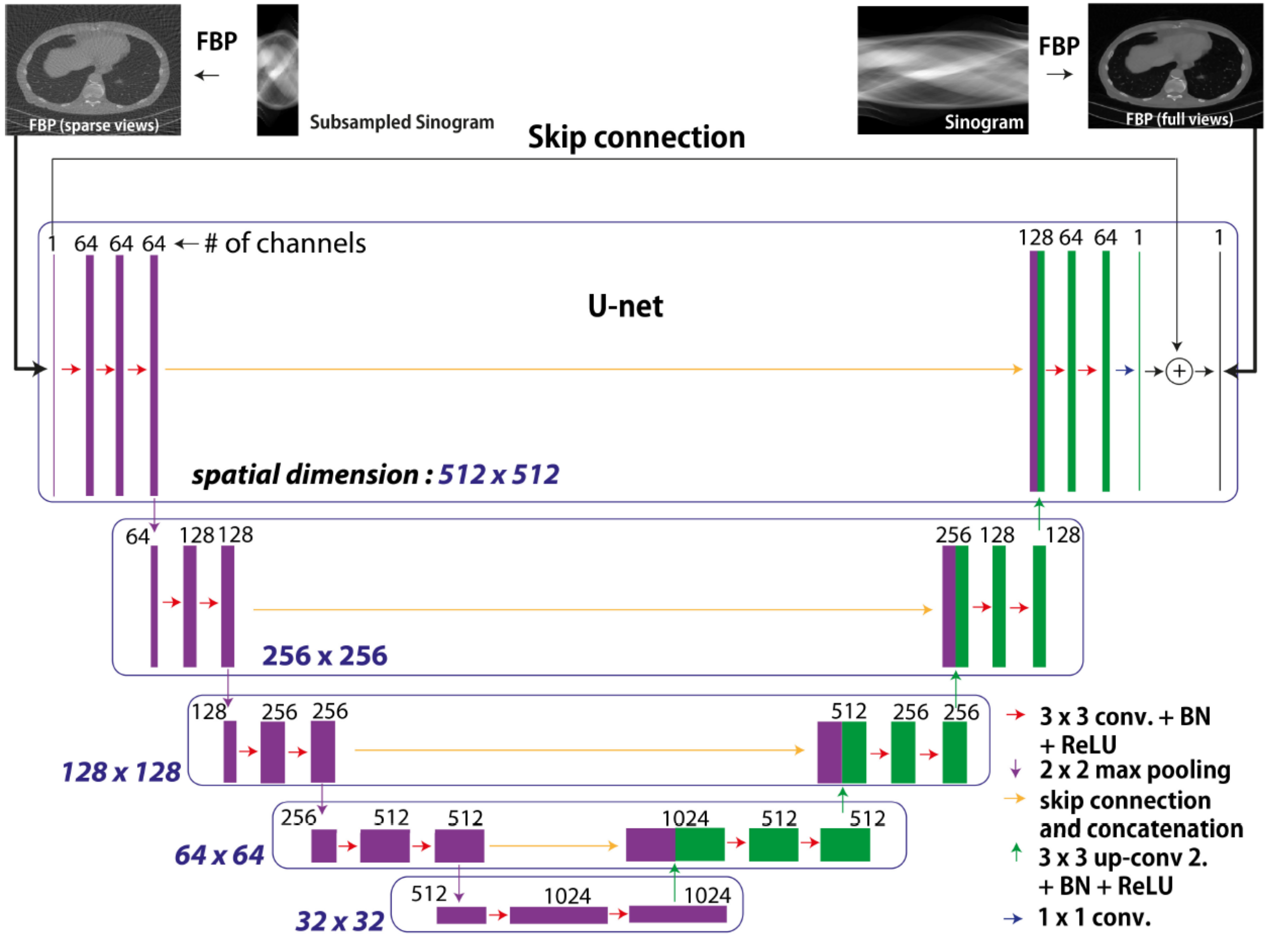
- Reconstruction by FBP
- Decomposition into Wavelet coefficients
- Learned denoising of coefficients
- 2nd place at the 2016 AAPM low-dose CT Grand Challenge



FBPConvNet

[JIN ET AL., *IEEE TRANSACTIONS ON IMAGE PROCESSING*, 2017]

- Reconstruction by FBP
- Learned denoising of reconstructed image
- Residual U-Net architecture



MODEL BASED RECONSTRUCTION

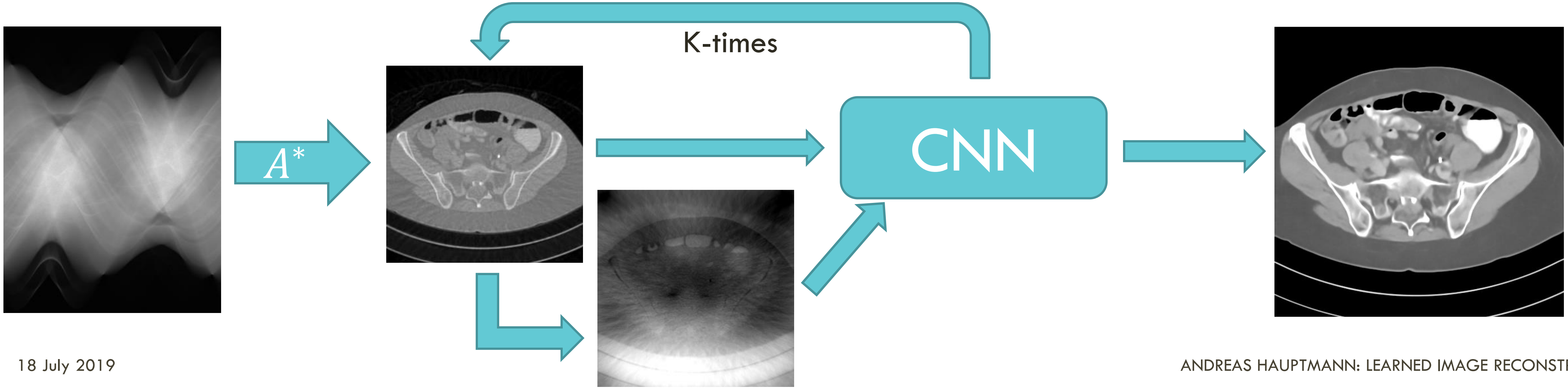
Forward and adjoint operator are used directly

Typically done in an iterative approach

In this case we learn an iterative update

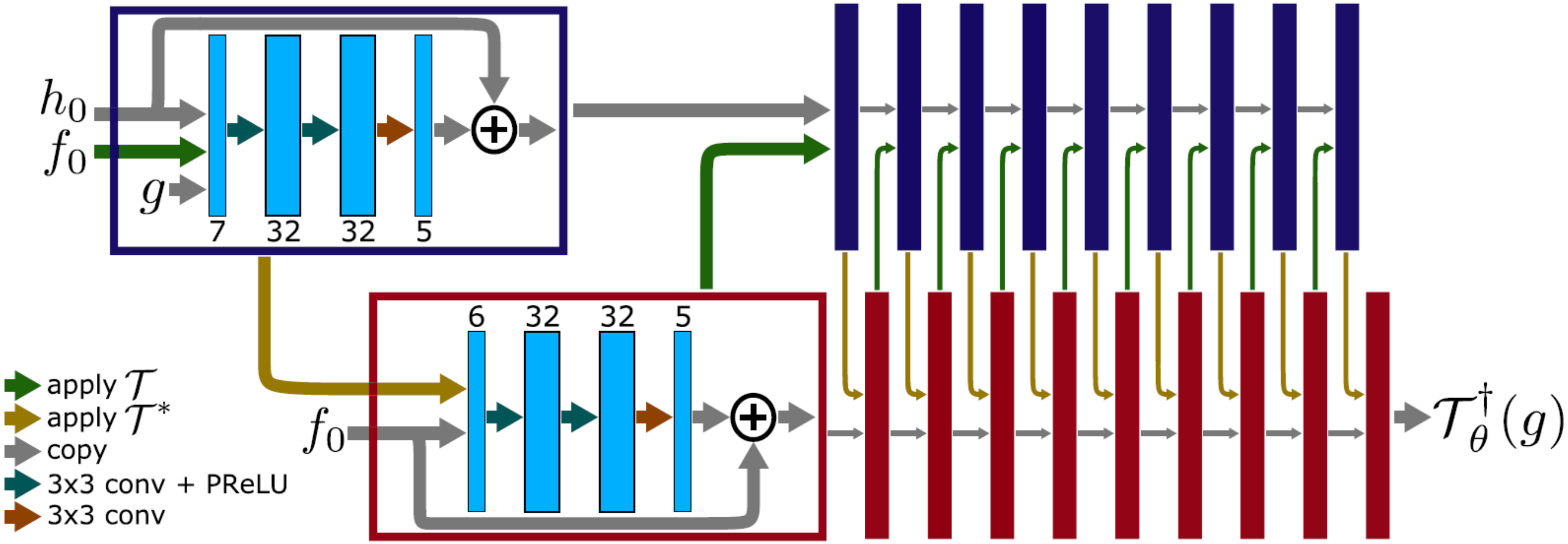
$$x_{k+1} = \mathcal{G}_\theta(\nabla d(y, Ax_k), x_k),$$

where $d(y, Ax_k)$ denotes the data-fit and $\mathcal{G}_\theta : X \times X \to X$ is typically a simple CNN.



LEARNED PRIMAL-DUAL RECONSTRUCTION
[ADLER & ÖKTEM, *IEEE TRANSACTIONS ON MEDICAL IMAGING*, 2018]

- Learned iterative reconstruction
- Learning in image (primal) and data (dual) domain



CARDIOVASCULAR MAGNETIC RESONANCE IMAGING

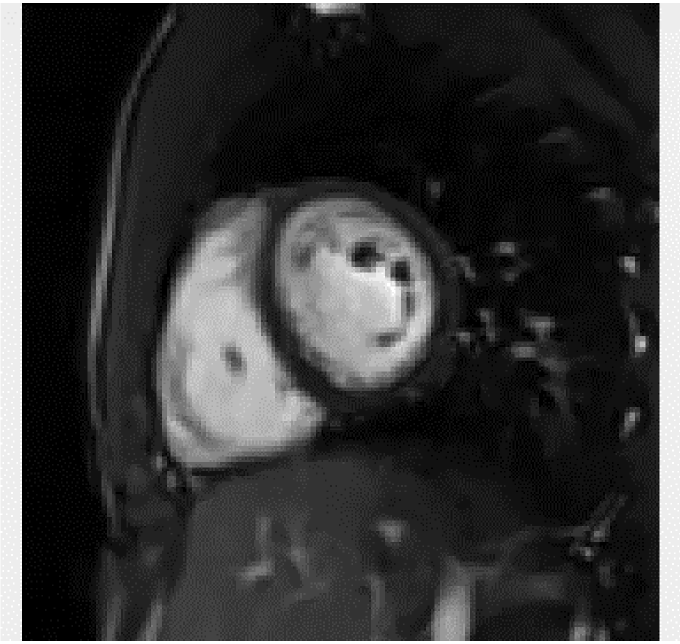
Forward model: Fourier transform $\mathcal{F}_k$

Data given in k-space: $y = \mathcal{F}_k x$

Reconstruction by inverse Fourier transform: $\mathcal{F}_k^{-1} y$

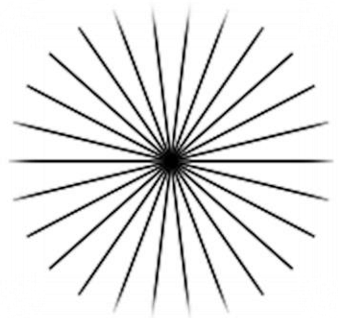
- Gold-standard taken under breath-hold, $\sim$10 seconds
- Need for real-time imaging for pediatric imaging

Gold-standard breath-hold

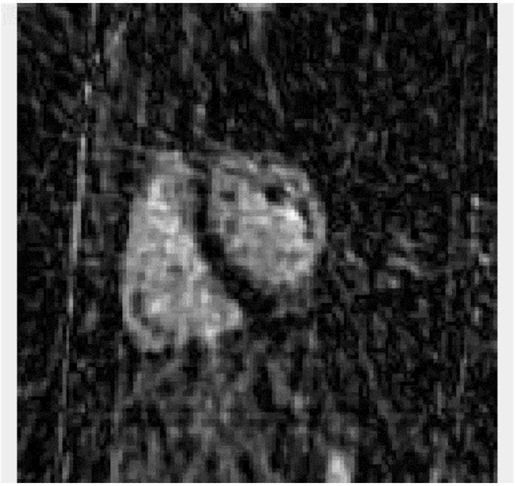


REAL-TIME IMAGING IN CMR AND DEEP LEARNING

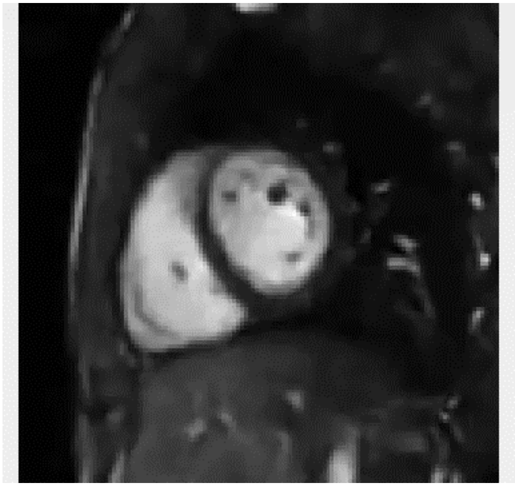
K-space data



Direct reconstruction
F_k^{-1}



CNN



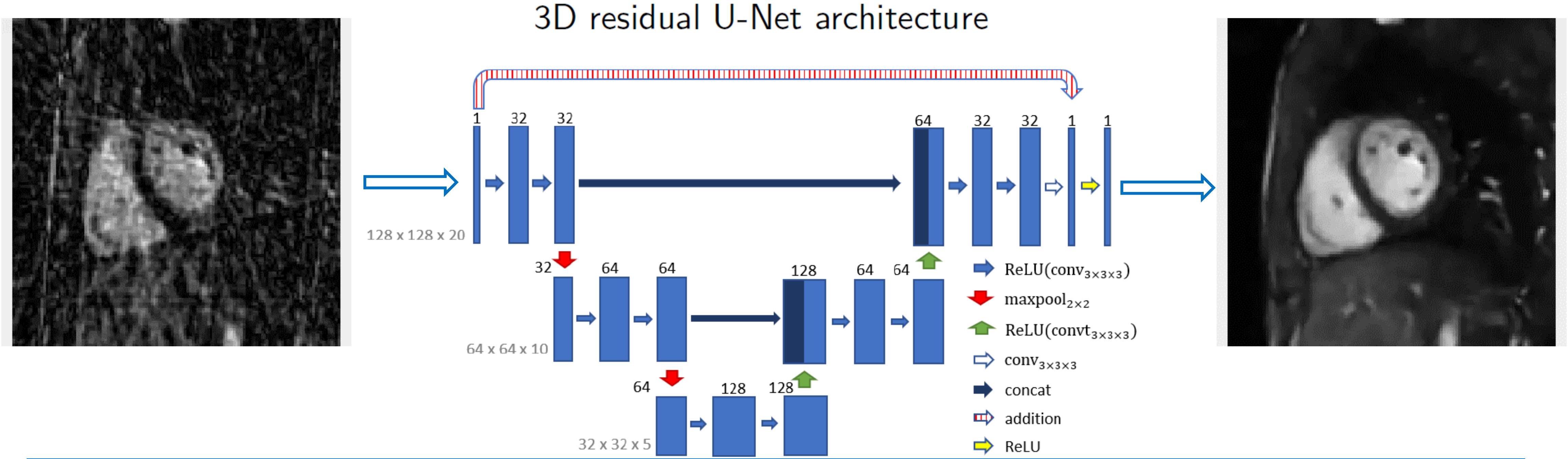
Pro:

- Fast reconstruction and post-processing
- Training with magnitude images possible

Contra:

- Many samples needed for training
- Artefacts ideally to be noise-like

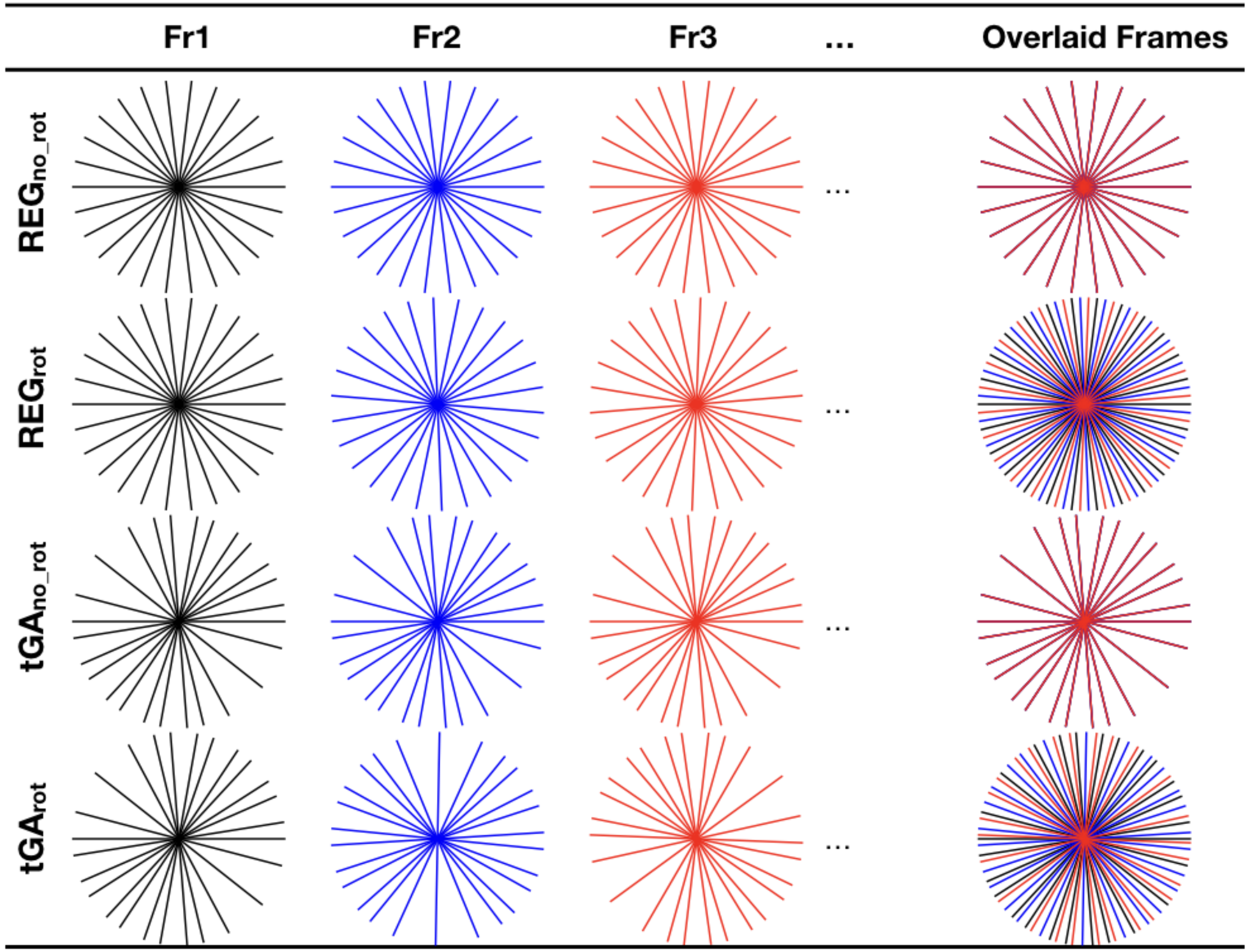
NETWORK ARCHITECTURE AND TRAINING



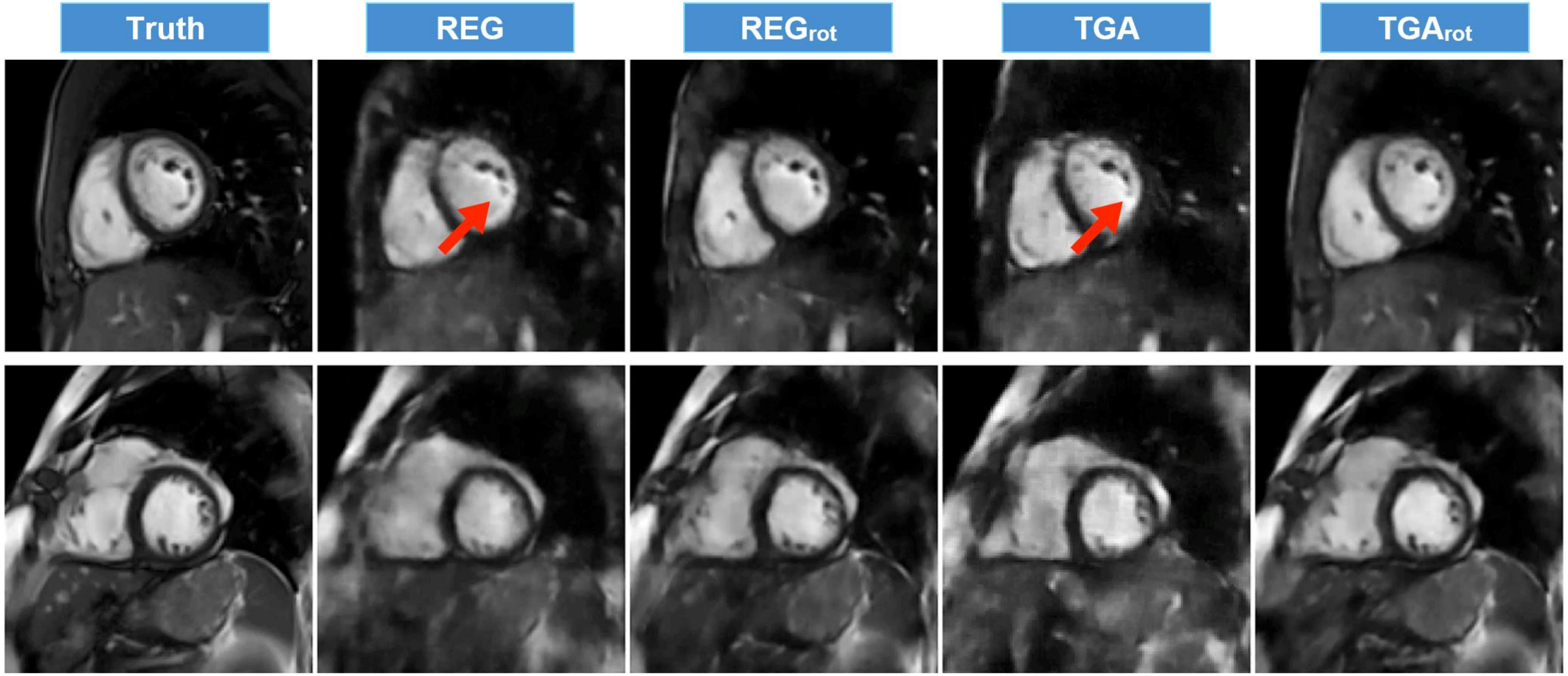
- Trained to minimise the ℓ^2-loss of output to the desired ground truth (or possible ℓ^1-loss)
- 2276 (2D+time) data sets from ~250 patients
- Real-time data 13x accelerated, simulated from magnitude images

COMPARISON OF RADIAL SAMPLING PATTERNS

- Nature of artefacts is crucial for learning task
- Test of radial rotating/non-rotating sampling patterns



COMPARISON OF RADIAL SAMPLING PATTERNS



	REG	REGrot	TGA	TGArot
RMSE ($\times 10^{-2}$)	8.0 ± 1.5 *	4.9 ± 1.0 *	8.4 ± 1.6 *	4.2 ± 1.3
SSIM	0.64 ± 0.04 *	0.83 ± 0.03 *	0.63 ± 0.05 *	0.87 ± 0.03

CLINICAL STUDY

tGA radial bSSFP

R = 13 - 14 spokes

1.7mm - 36ms

10 CHD patients

Age = 33.6±16.8 years

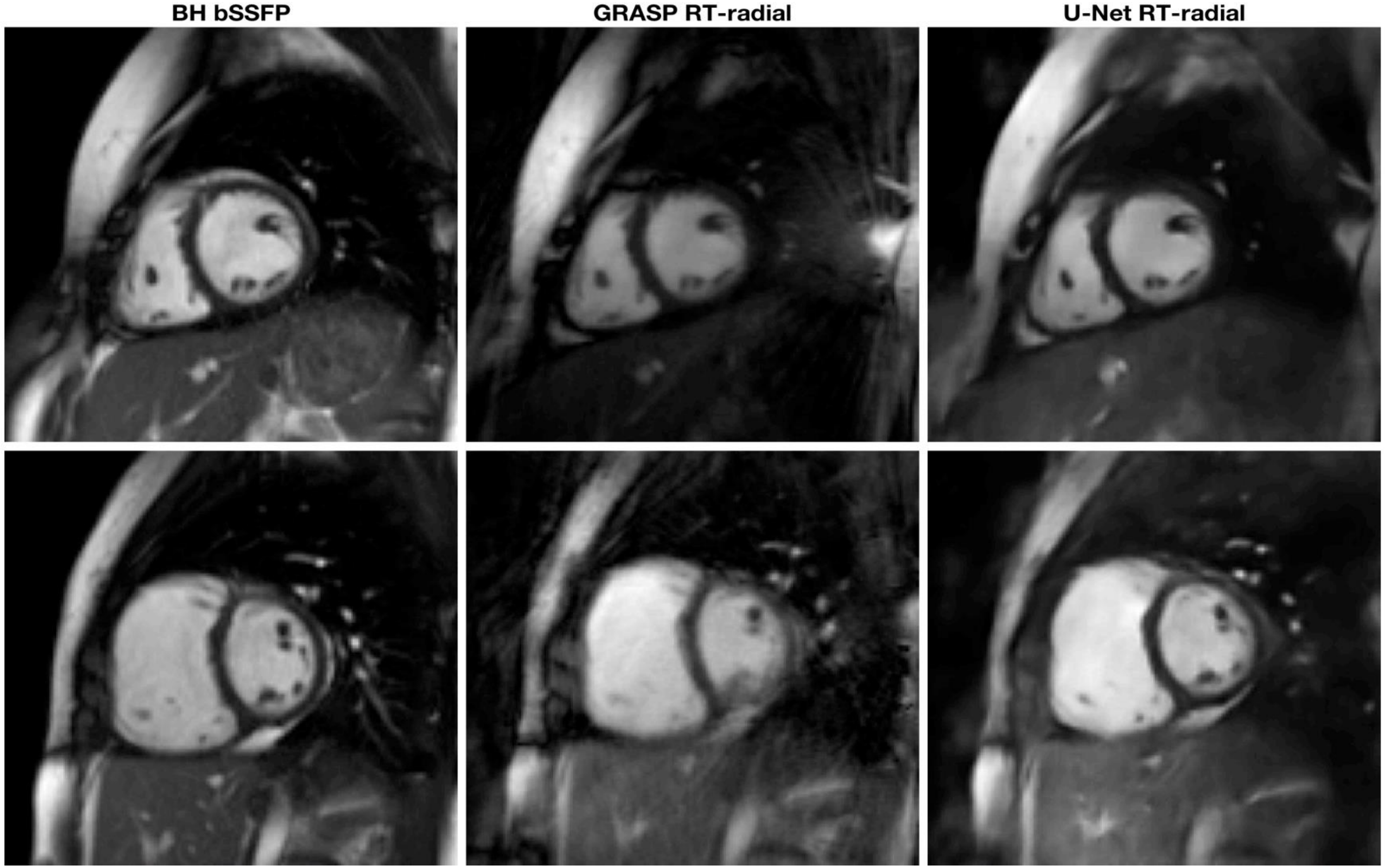
Same Dx as training

Learned reconstruction

GRASP reconstruction

Compare to conventional BH

CLINICAL STUDY: RECONSTRUCTIONS



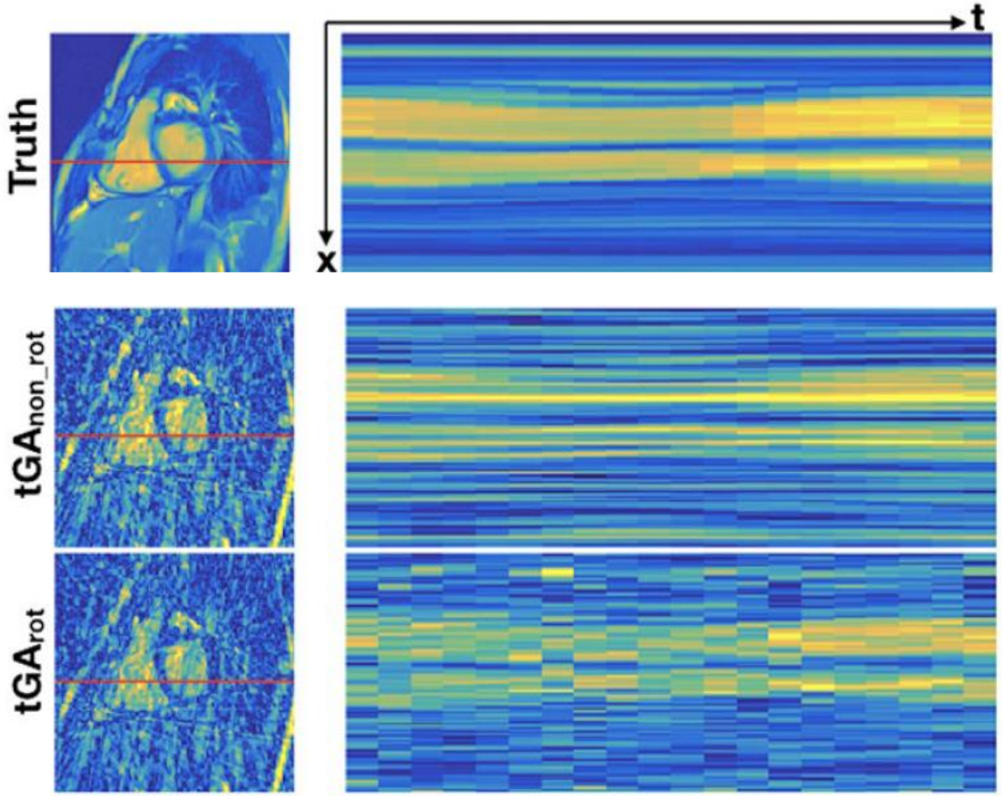
CLINICAL STUDY: VENTRICULAR VOLUMES

	Mean ± Standard deviation		
	BH-bSSFP	RT GRASP	RT U-Net
LV EDV (mL)	148 ± 44	143 ± 44	151 ± 46
LV ESV (mL)	56 ± 27	**60 ± 29***	58 ± 29
LV EF (%)	64 ± 10	**60 ± 11*†**	63 ± 11
RV EDV (mL)	213 ± 97	**198 ± 89*†**	204 ± 92
RV ESV (mL)	92 ± 49	89 ± 48	**91 ± 47***
RV EF (%)	58 ± 7	57 ± 6	57 ± 6

* *Values are significantly different from BH-bSSFP ($p<0.05$)*
† *Values are significantly different from RT U-Net ($p<0.05$)*

[Hauptmann et al., *Magnetic Resonance in Medicine*, 2019]

WHEN TO CONSIDER POST-PROCESSING?



- If artefacts are incoherent in time: CNN learns interpolation in time
- If artefacts are **not** noise-like: CNN needs to rely on features learned from the training data
- Model knowledge needed to design a robust learning task!

MODEL BASED RECONSTRUCTIONS

- Model-based iterative reconstructions are shown to outperform post-processing approaches:
  - Especially if artefacts are not noise-like/spatially correlated:
    - Streaking artefacts/limited-view in CT
- Typically these iterative methods are trained end-to-end, i.e. data to final iterate.

Method	PSNR (dB)	SSIM	Runtime (ms)	Parameters
FBP	19.75	0.597	4	1
TV	28.06	0.928	5 166	1
Learned U-Net	29.20	0.943	9	10^7
Learned Gradient Descent	32.02	-	-	-
Learned Primal-Dual	38.28	0.988	49	$2.4 \cdot 10^5$

[Adler&Öktem, 2017&2018]

PROBLEM WITH END-TO-END TRAINING?

- End-to-end training is not feasible for large image size:
  - ➢ Either due to memory limitation, with 12GB GPU we can train 1024x1024 in 2D or 256x256x16 in 3D.
  - ➢ Or Computational cost during training, due to multiple application of forward operator

Possible solution: Greedy training of each iterate separately.

- ➢ That way we can separate evaluation of forward operator from the training task.

A GREEDY APPROACH

Classical variational approach: Find x from measurement y as a minimiser of:

$$x \in \arg\min_{x'} \left\{ J(x') \right\} = \arg\min_{x'} \left\{ \mathcal{D}(x; g) + \lambda R(x') \right\}.$$

$$\mathcal{D}(x_k; g) = \frac{1}{2} \| A x_k - y \|_2^2$$

and

$$\nabla \mathcal{D}(x_k; g) = A^*(A x_k - y).$$

Instead of solving explicitly by proximal gradient descent

$$x_{k+1} = \operatorname{prox}_{R,(\lambda \gamma_{k+1})} \left(x_k - \gamma_{k+1} \nabla \mathcal{D}(x_k; g) \right),$$

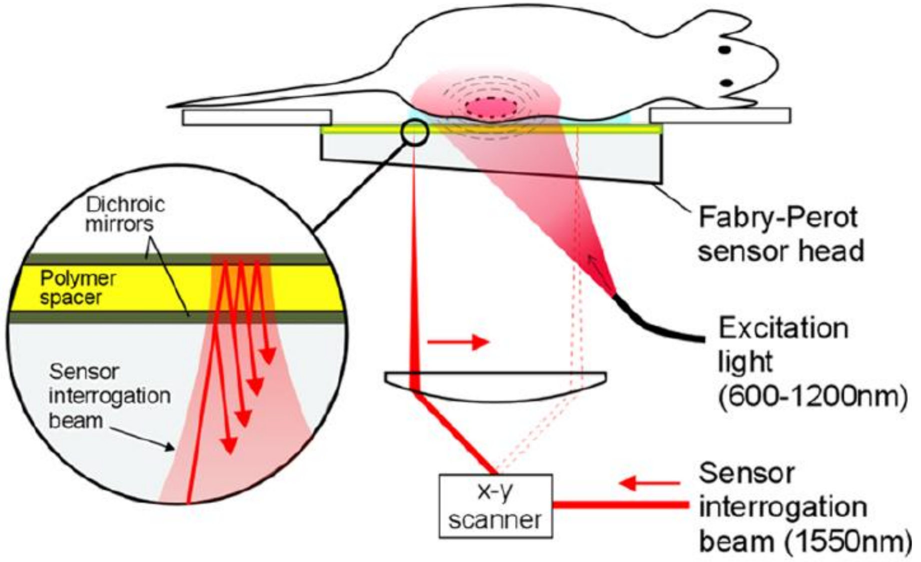
we propose to find a function (formulated as CNN) such that

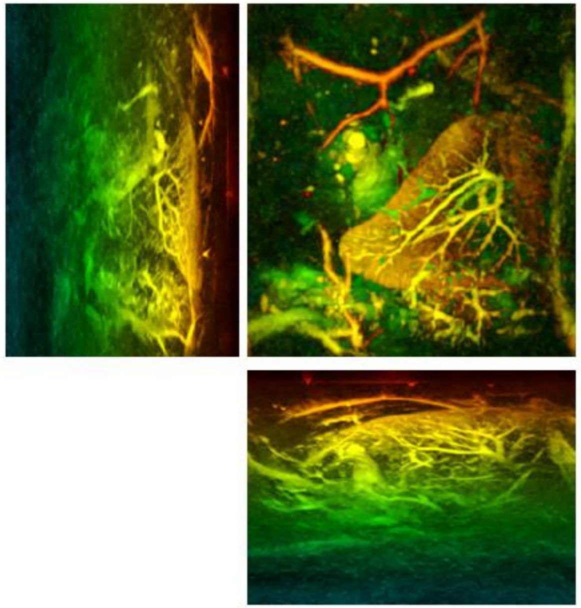
$$x_{k+1} = G_{\theta_k}(\nabla \mathcal{D}(x_k; g), x_k).$$

EXAMPLE: PHOTOACOUSTIC TOMOGRAPHY

- Fabry Perot polymer film ultrasound sensor is a planar interferometer

- Acoustically-induced thickness changes are detected optically





- Standard method is a raster scan: sequentially interrogate all pixels

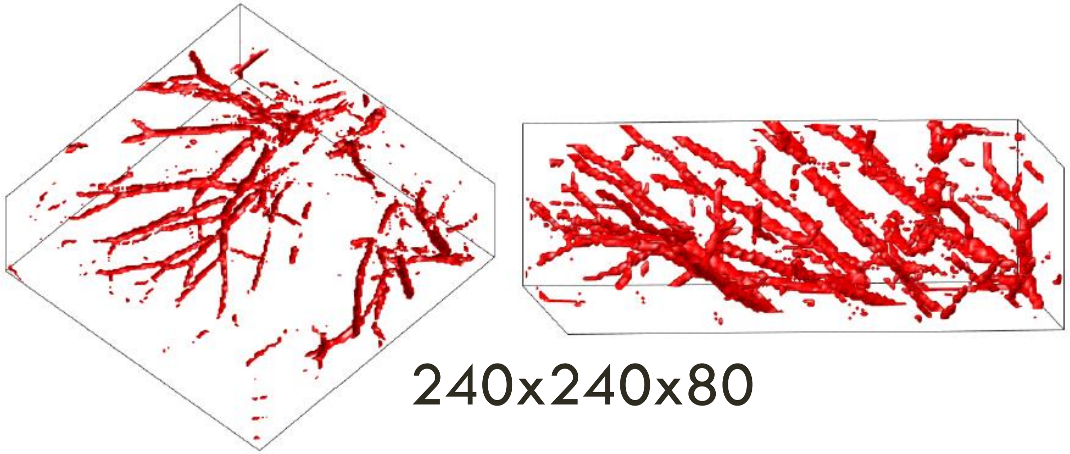
- Interrogate sensor with patterns instead
  - ➢ Compressed Sensing

[Jathoul et al., *Nature Photonics*, 2015]

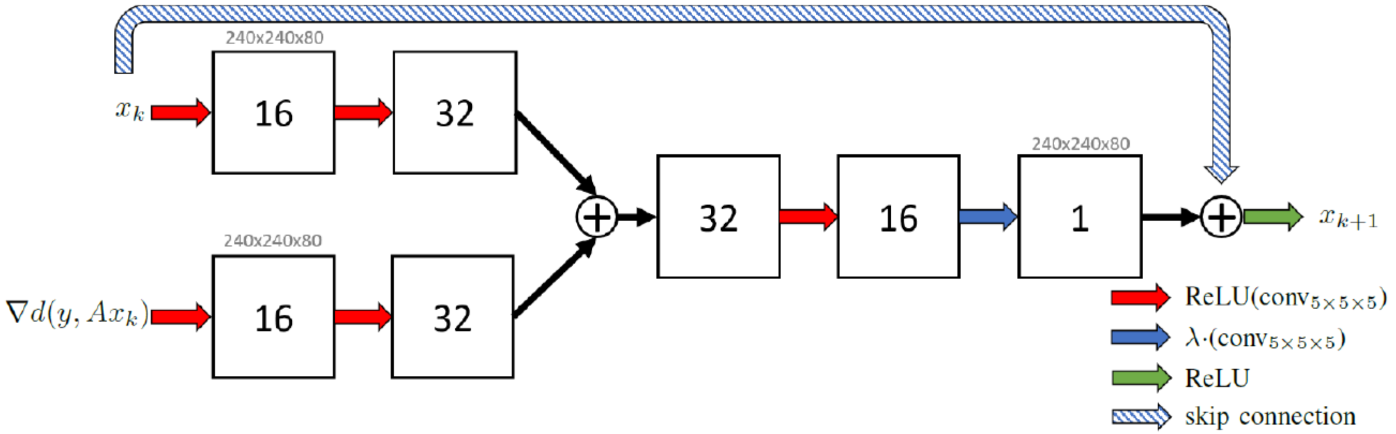
TRAINING ON VESSEL PHANTOMS

- Trained to minimise the ℓ^2-distance of new iterate x_{k+1} to the true solution
- With the computation of the gradient, total training time for 5 iterations takes 7 days
- Compare: End-to-end training would take about $\sim$140 days

$$x_{k+1} = G_{\theta_k}(\nabla \mathcal{D}(x_k; g), x_k).$$



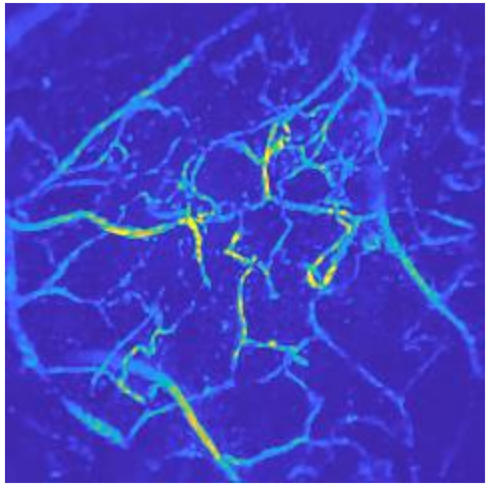
240x240x80



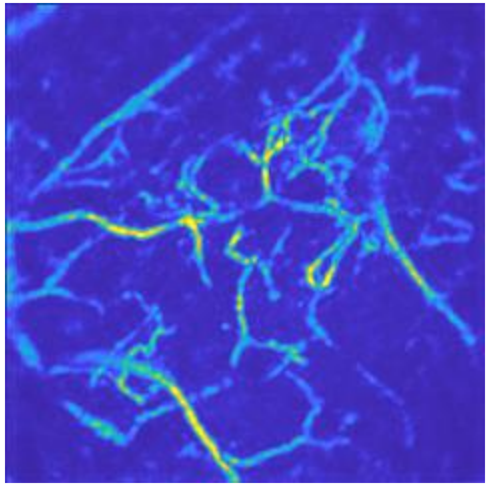
APPLICATION TO HUMAN IN-VIVO MEASUREMENTS

- Reduces reonstruction time by a factor 4 (by reduction of iterations)
- Considerably improves reconstruction quality

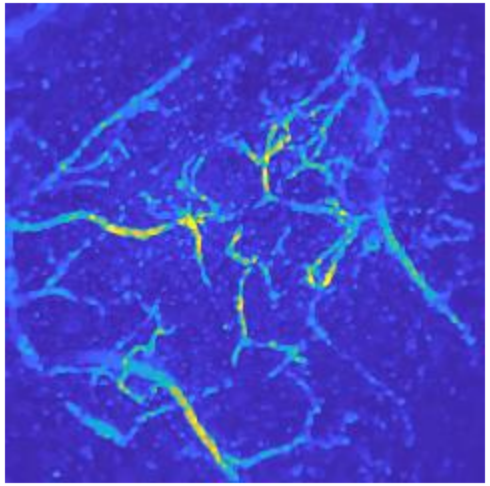
Reference
Fully-sampled data



Learned Reconstruction
4x sub-sampled, 5 Iterations,
Time: 2.5 min., PSNR: 41.40



Total Variation Reconstruction
4x sub-sampled, 20 Iterations,
Time: 10 min., PSNR: 38.05

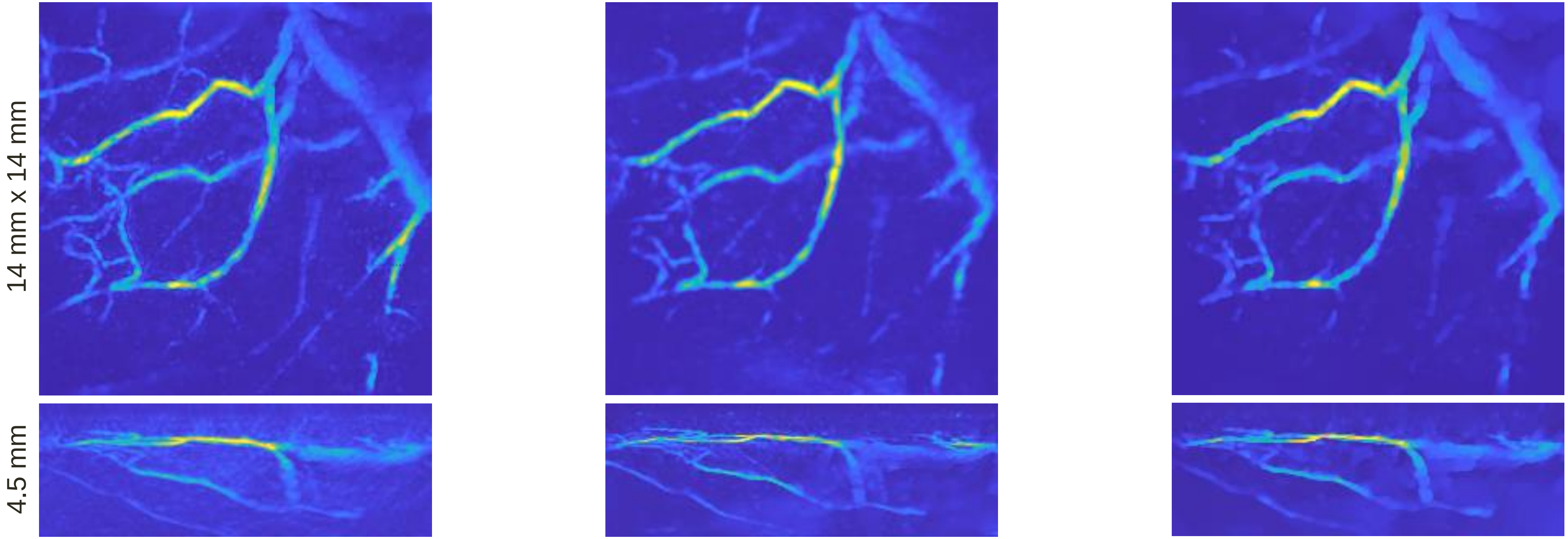


[Hauptmann et al., *IEEE Transactions on Medical Imaging*, 2018]

ACCELERATION BY USING AN APPROXIMATE MODEL

- Reduces reonstruction time by another factor of ~8 (→ 32x compared to TV)

Reference Fully-sampled data	Learned Reconstruction 4x sub-sampled, 5 Iterations, Time: 20 sec., PSNR: 42.18	Total Variation Reconstruction 4x sub-sampled, 20 Iterations, Time: 10 min., PSNR: 41.16



[Hauptmann et al., *Machine Learning for Medical Image Reconstruction*, 2018]

ANOTHER SOLUTION: A MULTISCALE APPROACH

To save memory and computation time: Compute only final iterate in full resolution

Discretisation space: $S_i = \{X_i, Y_i\}, \ i \in \{0, \ldots, N\}$

Reconstructed image and data: $\{x_i, y_i\} \in S_i$

The desired finest resolution is $S_N = \{X_N, Y_N\} = \{X, Y\}$

Corresponding forward and adjoint operators:
$\mathcal{A}_i : X_i \to Y_i$ and $\mathcal{A}_i^* : Y_i \to X_i$

Projection and upsampling operator:
$\pi_i : S_N \to S_i \quad \text{and} \quad \tau : S_{i-1} \to S_i \quad \text{for } i \in \{1, \ldots, N\}.$

MULTI SCALE LEARNED GRADIENT DESCENT

We compute the gradient for each iterate by

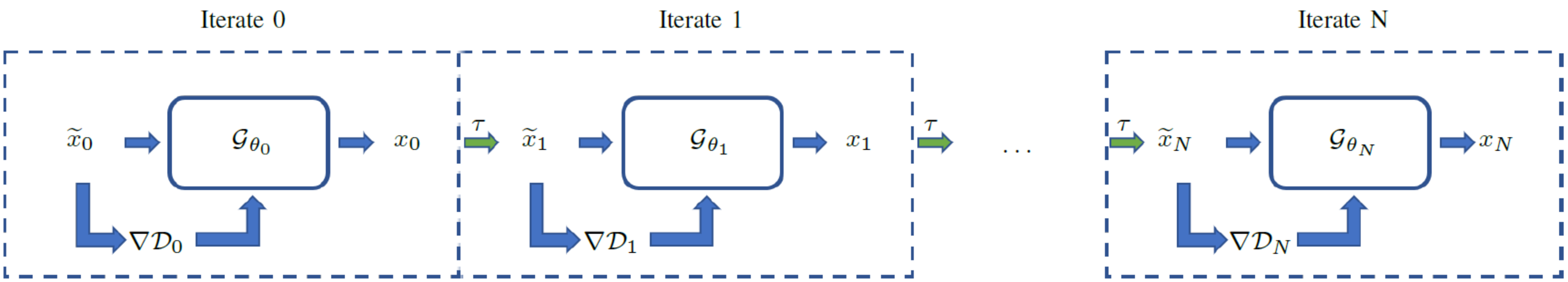
$$\nabla \mathcal{D}_i(x_i; g) := \mathcal{A}_i^* \left(\mathcal{A}_i(x_i) - \pi_i(y)\right).$$

The update is then performed by

$$\begin{cases} x_i = \mathcal{G}_{\theta_i}\left(\widetilde{x}_i, \nabla \mathcal{D}_i(\widetilde{x}_i; y)\right) \\ \widetilde{x}_{i+1} = \tau(x_i). \end{cases}$$

Algorithm 1 Multi-scale learned gradient descent (MS-LGD)

1: **for** $i = 0, \dots, N$ **do**
2:   **if** $i = 0$ **then**
3:     $\widetilde{x}_0 \leftarrow \mathcal{A}_0^\dagger \pi_0(y)$
4:   **else**
5:     $\widetilde{x}_i \leftarrow \tau(x_{i-1})$
6:   **end if**
7:   $x_i \leftarrow \mathcal{G}_{\theta_i}\left(\widetilde{x}_i, \nabla \mathcal{D}_i(\widetilde{x}_i; y)\right)$
8: **end for**
9: $x^* \leftarrow x_N$



MULTI SCALE LEARNED FILTERED GRADIENT DESCENT

Due to upsampling we lose high-frequency components, especially in the last iterate.

Additionally we compute a filtered version of the gradient by substituting the adjoint with the filtered backprojection:

$$\nabla^\dagger \mathcal{D}_i(x_i; y) := \mathcal{A}_i^\dagger\left(\mathcal{A}_i(x_i) - \pi_i(y)\right)$$

Algorithm 2 Multi-scale learned filtered gradient descent (MS-LFGD)

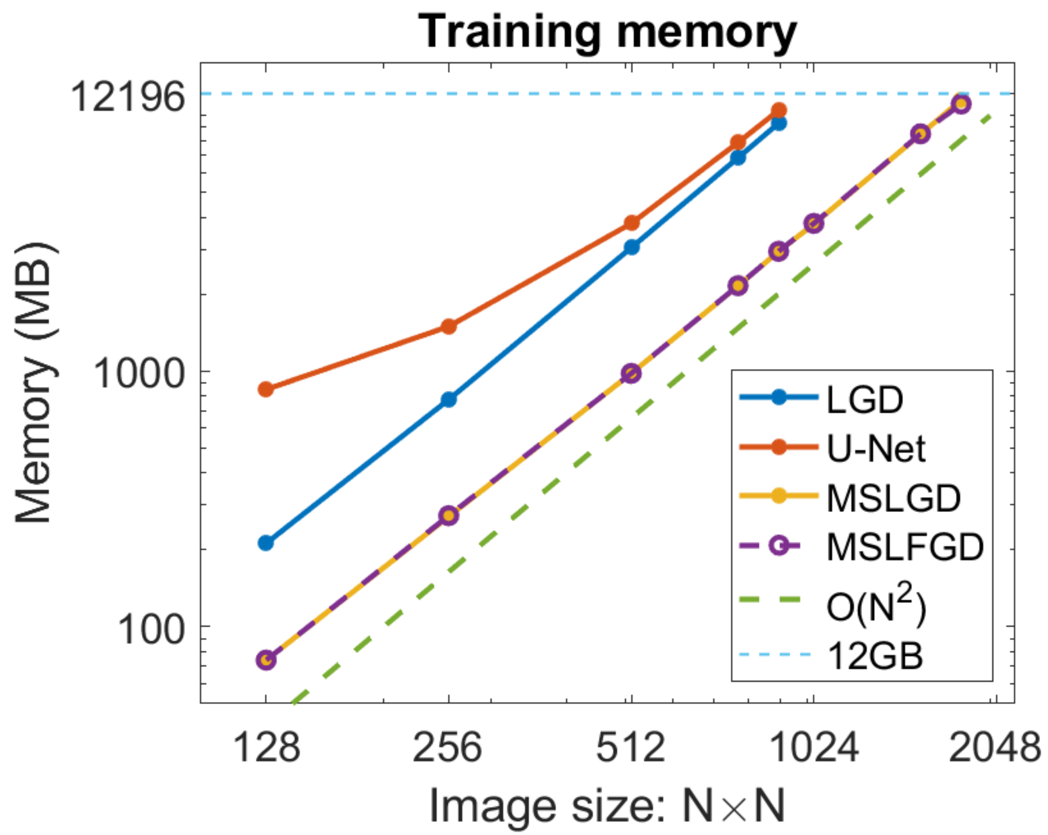
1: **for** $i = 0, \dots, N$ **do**
2:     **if** $i = 0$ **then**
3:         $\widetilde{x}_0 \leftarrow \mathcal{A}_0^\dagger \pi_0(y)$
4:     **else**
5:         $\widetilde{x}_i \leftarrow \tau(f_{i-1})$
6:     **end if**
7:     $\widetilde{x}_{res} \leftarrow \mathcal{A}_i(\widetilde{x}_i) - \pi_i(y)$
8:     $\nabla \mathcal{D}_i(\widetilde{x}_i; y) \leftarrow \mathcal{A}_i^*(x_{res})$
9:     $\nabla^\dagger \mathcal{D}_i(\widetilde{x}_i; y) \leftarrow \mathcal{A}_i^\dagger(x_{res})$
10:     $x_i \leftarrow \mathcal{G}_{\theta_i}\left(\widetilde{x}_i, \nabla \mathcal{D}_i(\widetilde{x}_i; y), \nabla^\dagger \mathcal{D}_i(\widetilde{x}_i; y)\right)$
11: **end for**
12: $f^* \leftarrow f_N$

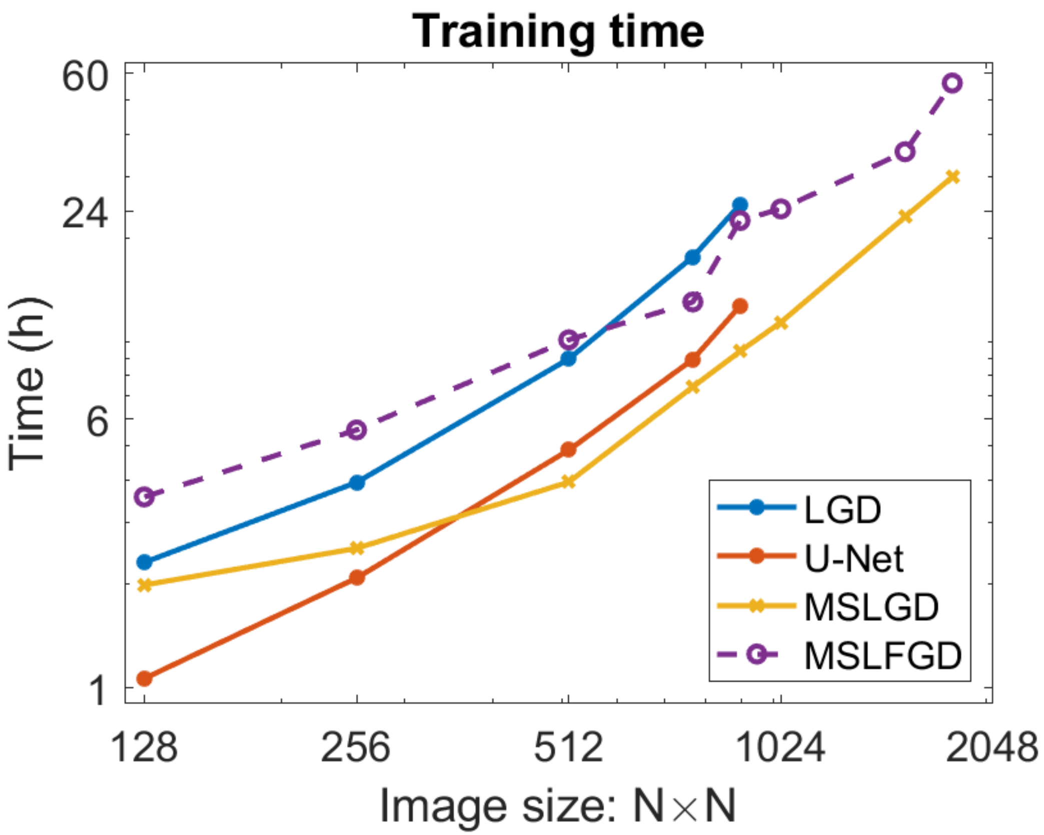
A SCALABILITY STUDY ON SIMULATED DATA

- We create phantoms of random ellipses
- Test algorithms for increasing image size
- Fan-beam geometry, 512 angles, 0.05% Gaussian noise
- Test following algorithms:
  - Learned Gradient Descent (LGD)
  - Post-processing with U-Net
  - Multi-scale Learned Gradient Descent (MS-LGD)
  - Multi-scale Learned filtered Gradient Descent (MS-LFGD)

A SCALABILITY STUDY: RESULTS

Test on a single NVIDIA Titan XP GPU (12GB)

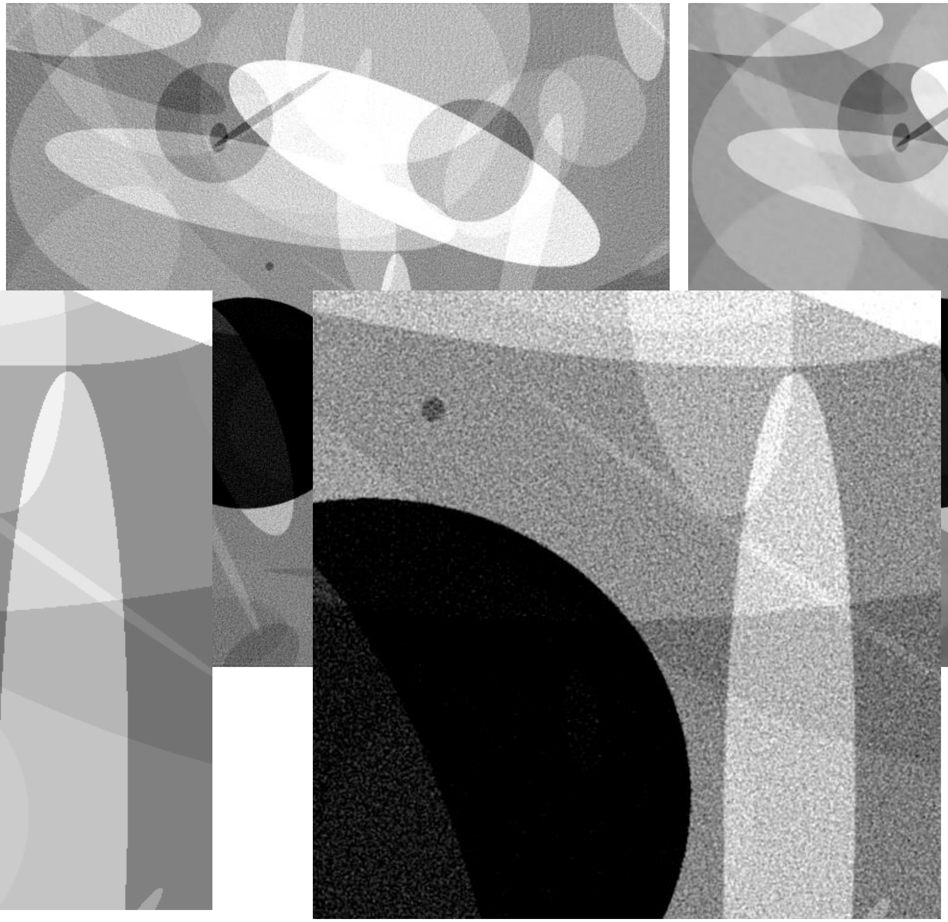




RECONSTRUCTION WITH MS-LFGD (1536X1536)

Phantom

FBP



MS-LFGD

APPLICATION TO HUMAN PHANTOMS

We use the database supplied by Mayo Clinic for the AAPM Low Dose CT Grand Challenge

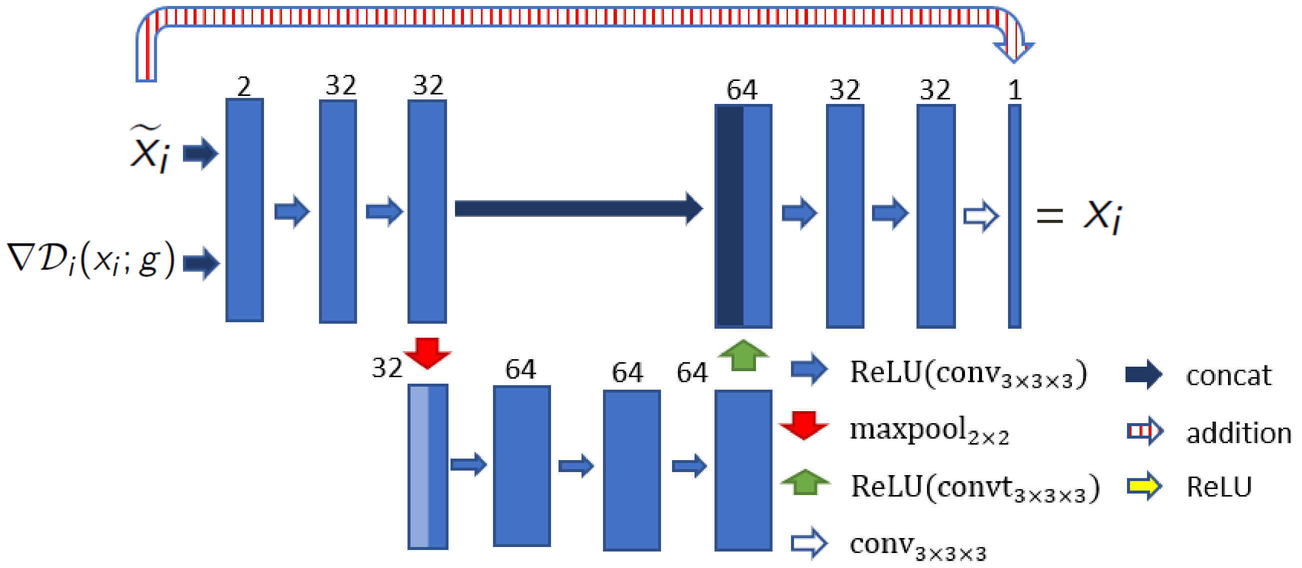
- Data is simulated following the Beer-Lamberts law with Poisson noise
- Training on 9 patients with 2168 slices of size 512x512
- Testing on 1 patient with 210 slices

We consider three setups of increasing difficulty:

	Angles	Photon count	Scales	Angles per Scale
Case 1	600	8000	5	600, 300, 150, 75, 37
Case 2	240	6000	5	240, 120, 60, 30, 30
Case 3	120	5000	5	120, 60, 30, 30, 30

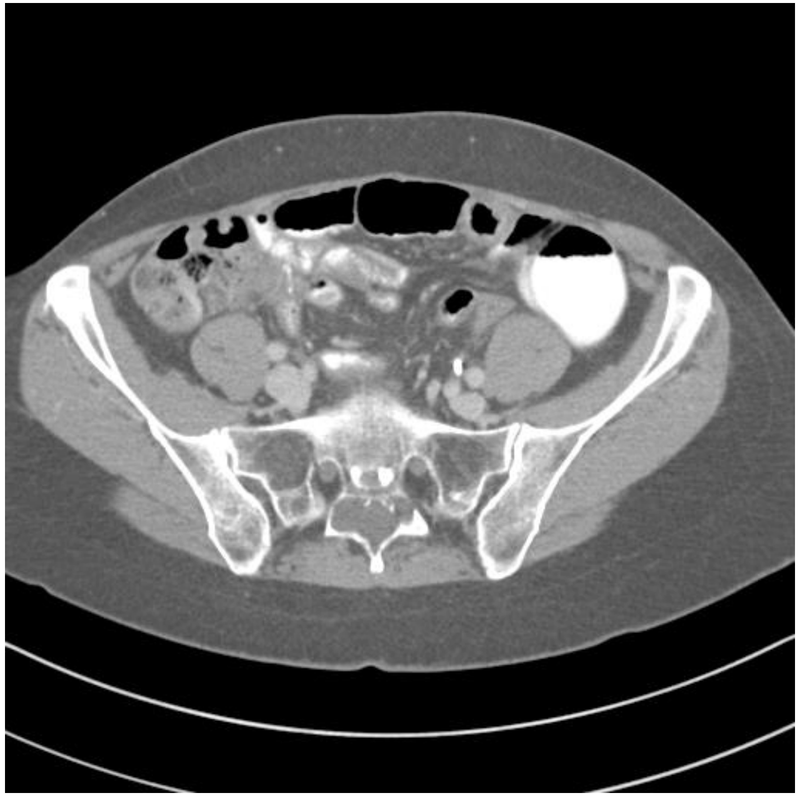
NETWORK AND TRAINING DETAILS

- Networks are trained for 50,000 iterations (end-to-end), 5 iterates/scales
- Minimise ℓ^2-distance to full-dose scan
- Sub-networks consist of a residual mini U-Net



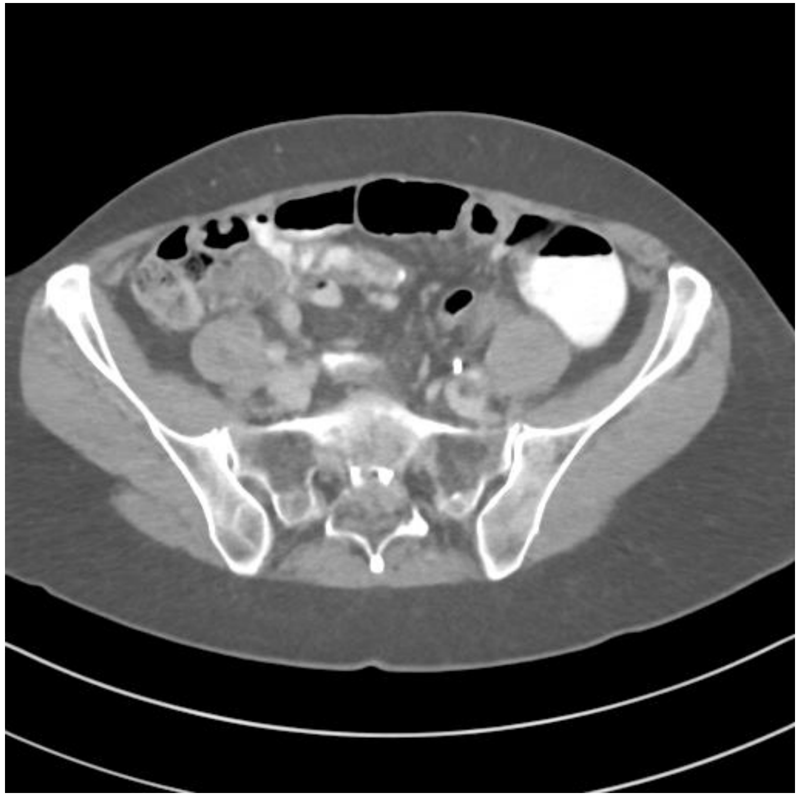
RECONSTRUCTION RESULTS, U-NET

Full-dose scan



FBP

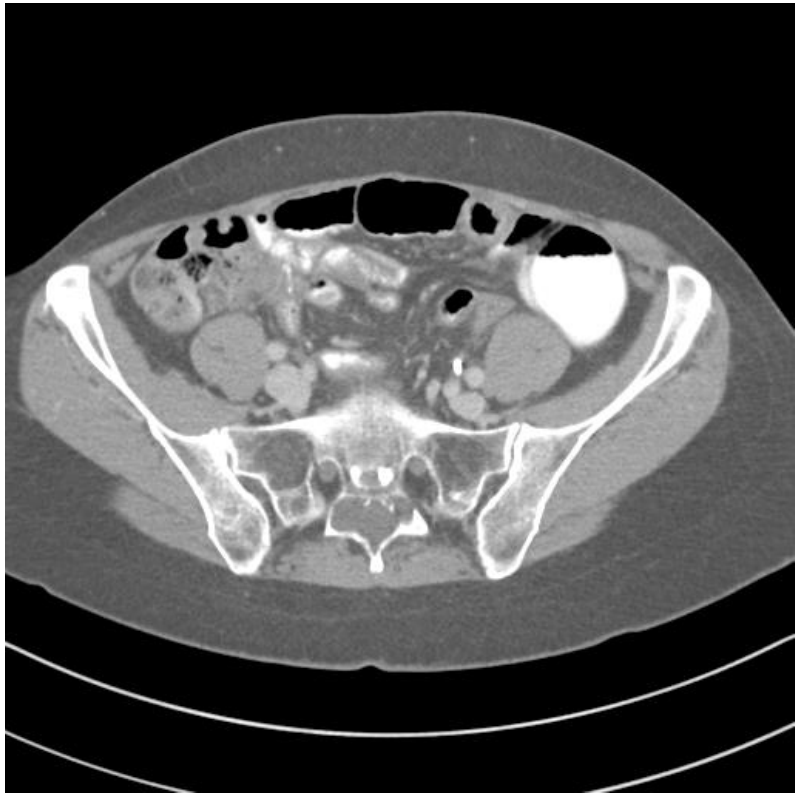
U-Net



[-300, 300]HU

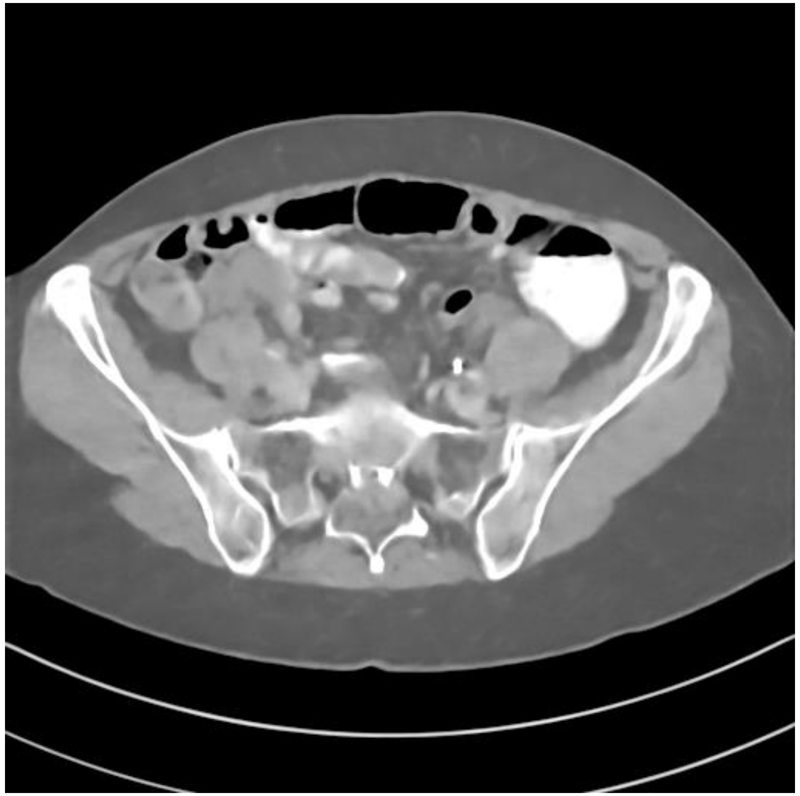
RECONSTRUCTION RESULTS, MS-LGD

Full-dose scan



FBP

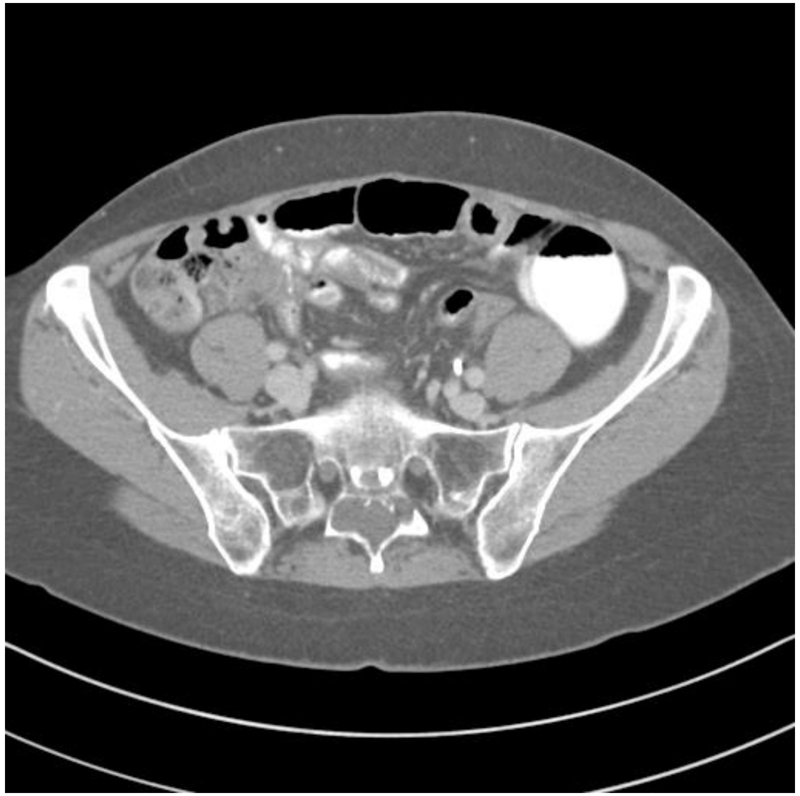
MS-LGD



[-300, 300]HU

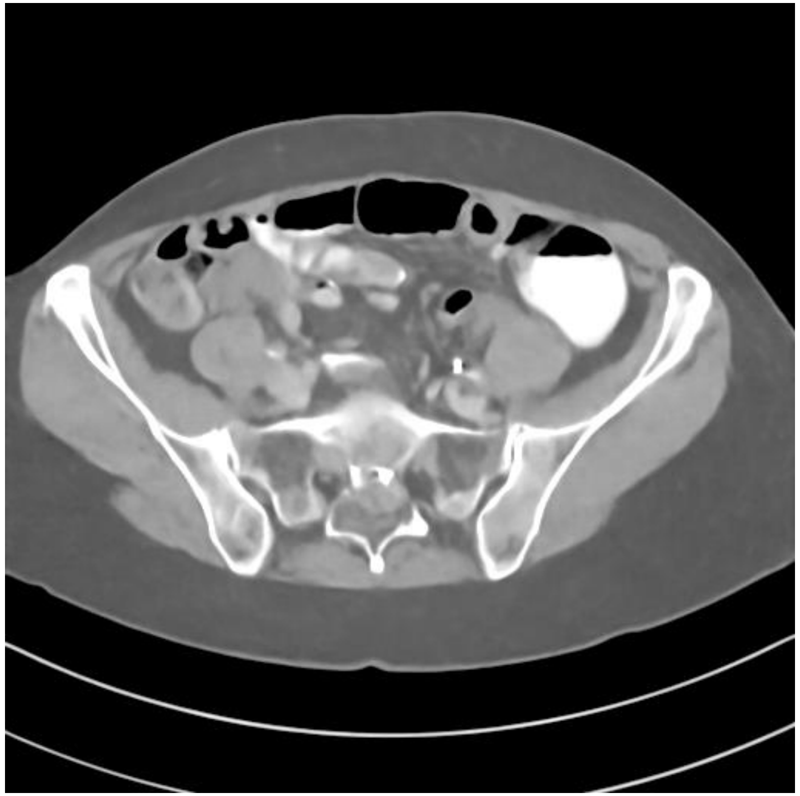
RECONSTRUCTION RESULTS, MS-LFGD

Full-dose scan



FBP

MS-LFGD



[-300, 300]HU

RECONSTRUCTION RESULTS, LGD

Full-dose scan | FBP | LGD

[-300, 300]HU

QUANTITATIVE RESULTS

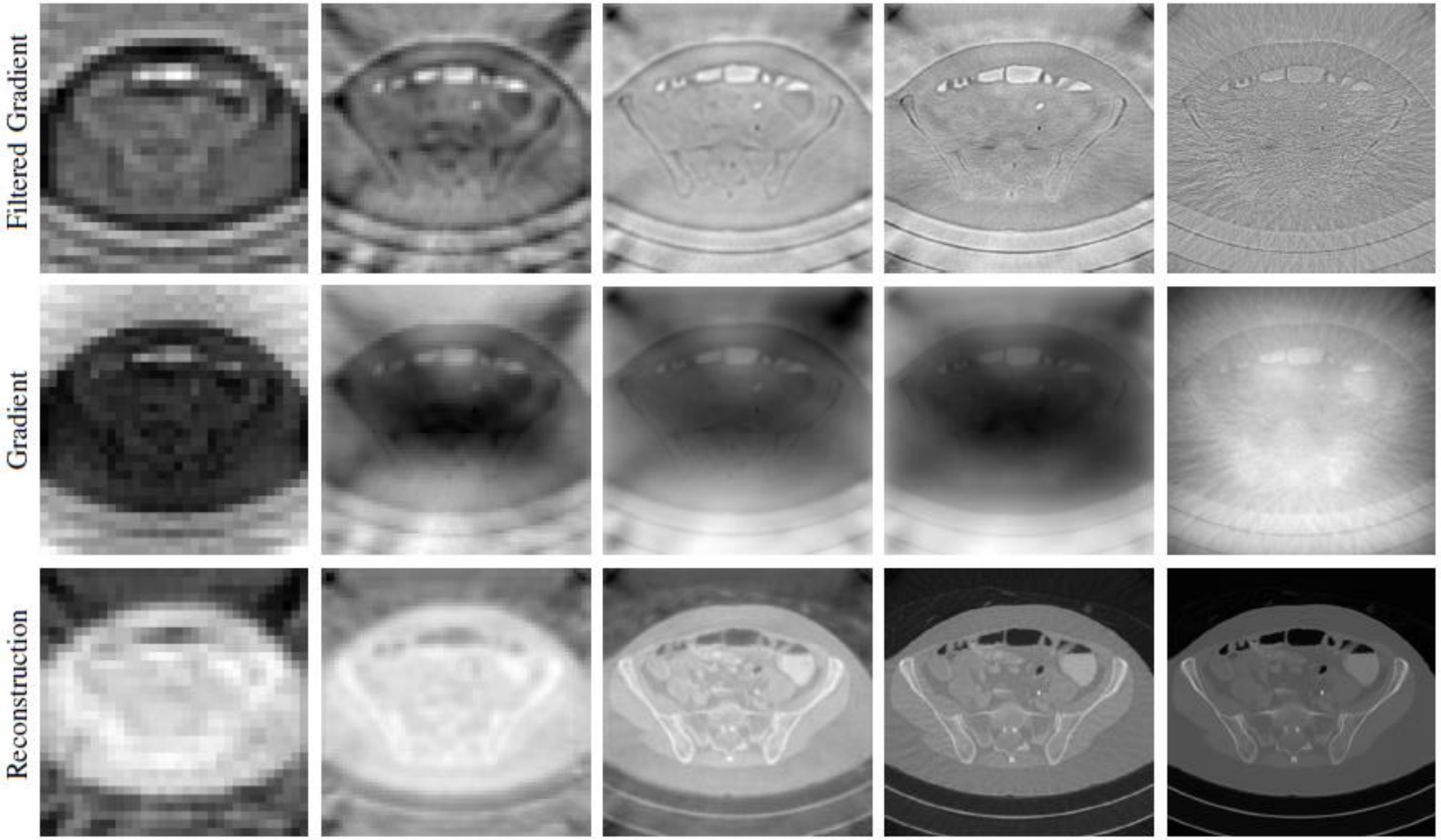
	600 angles		240 angles		120 angles	
	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM
LGD (MINI U-NET)	43.57 ±1.22	0.964 ±0.0033	41.36 ±1.22	0.953 ±0.0039	39.45 ±1.23	0.938 ±0.0054
U-NET	42.24 ±1.50	0.957 ±0.0044	40.41 ±1.38	0.944 ±0.0031	38.03 ±1.32	0.901 ±0.0053
MS-LGD	42.55 ±1.26	0.958 ±0.0033	39.73 ±1.33	0.938 ±0.0044	37.16 ±1.39	0.913 ±0.0065
MS-LFGD	43.34 ±1.22	0.963 ±0.0034	40.80 ±1.24	0.947 ±0.0040	38.30 ±1.26	0.926 ±0.0061

		600 angles		240 angles		120 angles	
	MEMORY	TRAIN.	EXEC.	TRAIN.	EXEC.	TRAIN.	EXEC.
LGD (MINI U-NET)	3042MB	13h40m	0.32s	10h10m	0.21s	5h25m	0.16s
U-NET	3784MB	6h45m	0.19s	4h35m	0.09s	3h50m	0.057s
MS-LGD	980MB	5h35m	0.06s	3h25m	0.05s	2h30m	0.049s
MS-LFGD	980MB	11h15m	0.23s	6h15m	0.12s	4h20m	0.089s

RECONSTRUCTION ON MULTIPLE SCALES

Reduces memory consumption of learned iterative reconstructions

- Possible to train end-to-end
- Faster with competitive results



[Hauptmann, Adler, Arridge, Öktem, *to be submitted (soon)*]

CONCLUDING REMARKS

- Use model knowledge to simplify the training task
- Post-processing can be advised if artefacts are highly incoherent (noise-like)
  $\Rightarrow$ Include temporal domain
- For geometrically restricted problems or correlated artefacts, feeding back model information iteratively is essential
- Speed up possible by using approximate/faster models
- End-to-end training possible by multi-scale approach

Engineering and Physical Sciences Research Council

British Heart Foundation

GIFT Surg

NVIDIA

Thank you for your attention