King's College London

Engineering and Physical Sciences Research Council

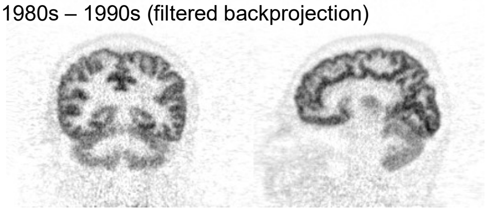
Advancing PET image reconstruction: from MAP to generative AI

ANDREW J. READER

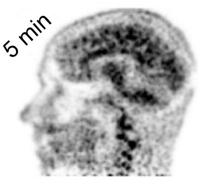
1

Model-based reconstruction: maths, stats, physics, images

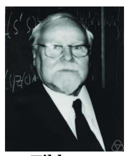
1980s – 1990s (filtered backprojection)



Radon



Regularise (MAPEM)

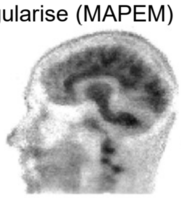


Tikhonov

By Konrad Jacobs, Erlangen - Mathematisches Institut Oberwolfach (MFO), https://opc.mfo.de/detail?photoID=4215, CC BY-SA 2.0 de, https://commons.wikimedia.org/w/index.php?curid=4534096

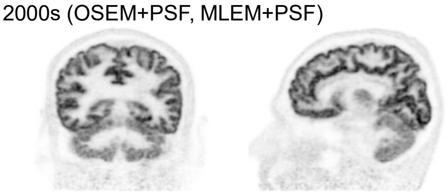
1990s (iterative reconstruction, OSEM, MLEM)

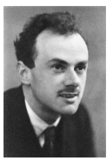
Poisson



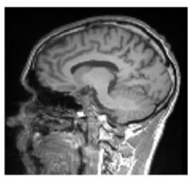
MRI guidance

2000s (OSEM+PSF, MLEM+PSF)





Dirac

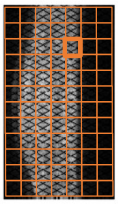


Colsher 1980, Kinahan & Rogers 1989, Shepp & Vardi 1982, Hudson & Larkin 1994

2

Poisson log-likelihood: accurate noise modelling

$$\Pr(m_i|q_i(\boldsymbol{x})) = \exp(-q_i(\boldsymbol{x}))\,\frac{q_i(\boldsymbol{x})^{m_i}}{m_i!}$$



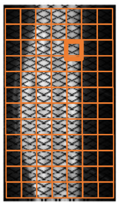
$\boldsymbol{m}$

$$\Pr(\boldsymbol{m}|\boldsymbol{q}(\boldsymbol{x})) = \prod_{i=1}^{I} \exp(-q_i(\boldsymbol{x}))\,\frac{q_i(\boldsymbol{x})^{m_i}}{m_i!} = L(\boldsymbol{x}|\boldsymbol{m})$$

LIKELIHOOD

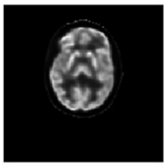
LOG-LIKELIHOOD

$$\ln L(\boldsymbol{x}|\boldsymbol{m}) = \sum_{i=1}^{I}(-q_i(\boldsymbol{x}) + m_i \ln q_i(\boldsymbol{x})) + const.$$



$\boldsymbol{q}(\boldsymbol{x})$

FORWARD MODEL $\boldsymbol{A}$



$\boldsymbol{x}$

USING THE GRADIENT (or SCORE) of the LOG-LIKELIHOOD

$$\hat{\boldsymbol{x}} = \arg\max_{\boldsymbol{x}} \sum_{i=1}^{I}(m_i \ln(\boldsymbol{A}\boldsymbol{x})_i - (\boldsymbol{A}\boldsymbol{x})_i)$$

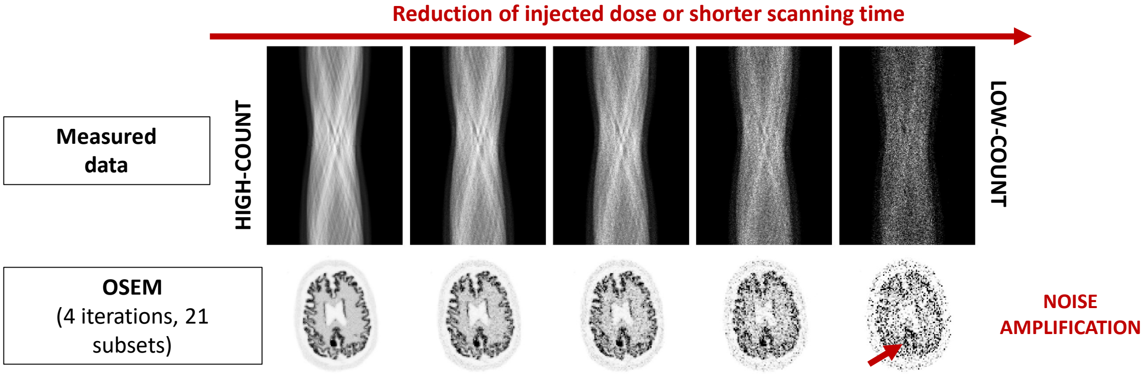
$$x_j^{EM} = \frac{x_j^k}{\sum_{i=1}^{I} a_{ij}} \sum_{i=1}^{I} a_{ij} \frac{m_i}{(\boldsymbol{A}\boldsymbol{x}^k)_i}$$

Shepp & Vardi 1982 Lange & Carson 1984

3

The problem of noise in PET Image Reconstruction

Images courtesy G. Corda –D'Incan



4

The problem of noise

PARAMETERS (often pixel values) — NOISY

$$L(\boldsymbol{x}|\boldsymbol{m})$$

- ML estimates can be highly noisy
- Prior knowledge about the parameter values can compensate for noise, using Bayes' theorem

POSTERIOR — LIKELIHOOD — PRIOR

$$\mathrm{p}(\boldsymbol{x}|\boldsymbol{m}) \propto \mathrm{p}(\boldsymbol{m}|\boldsymbol{x})\mathrm{p}(\boldsymbol{x})$$

- The posterior: likelihood function × $\mathrm{p}(\boldsymbol{x})$
- Can use the notion of "energy" for modelling the prior:

$$\mathrm{p}(\boldsymbol{x}) \propto \exp[-\beta U(\boldsymbol{x})]$$

- Images considered improbable (low $\mathrm{p}(\boldsymbol{x})$), assigned high energy $U(\boldsymbol{x})$
- Maximum *a posteriori* (MAP) reconstruction

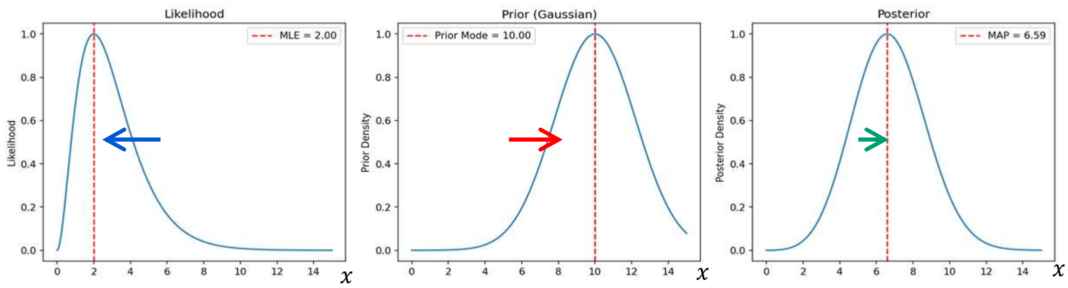
The MAP objective

- Maximum *a posteriori* (MAP) reconstruction

$$\hat{\boldsymbol{x}} = \arg\max_{\boldsymbol{x}} \sum_{i=1}^{I} (m_i \ln[\boldsymbol{A}\boldsymbol{x}]_i - [\boldsymbol{A}\boldsymbol{x}]_i) - \beta U(\mathbf{x})$$

LOG LIKELIHOOD — LOG PRIOR

LOG POSTERIOR

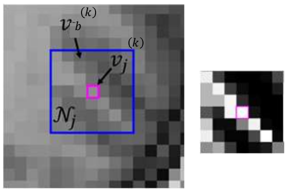


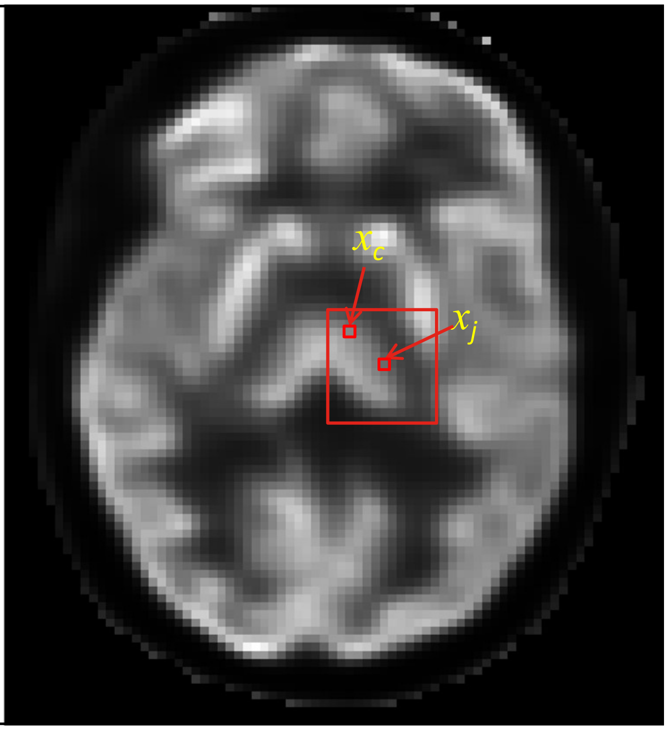
Examples of the energy function $U(x)$ for the prior

Can use guidance weights, w_{cj}, based on similarity of values in pixels c and j in a guide image (e.g. MRI, or current PET estimate).

$$U(\boldsymbol{x}) = \sum_{j=1}^{J} \sum_{c \in N_j} \xi_{cj} w_{cj} \psi(x_c - x_j)$$

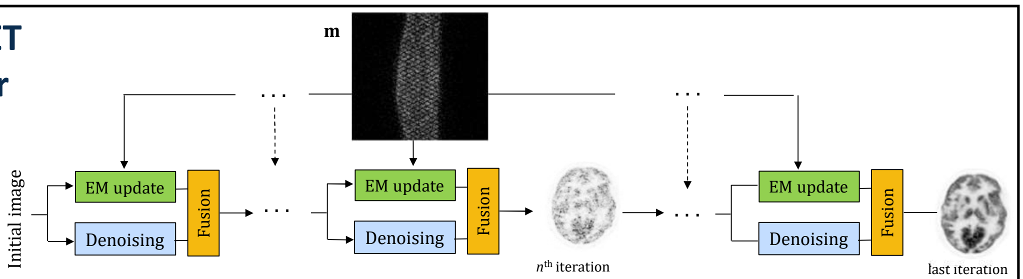
If binary weights used, called "Bowsher" method.





7

Example for PET Quadratic prior



MAP objective function

$$\hat{\boldsymbol{x}} = \arg\max_{\boldsymbol{x}} \sum_{i=1}^{I} (m_i \ln(\boldsymbol{Ax})_i - (\boldsymbol{Ax})_i) - \beta U(\boldsymbol{x})$$

$$U(\boldsymbol{x}) = \frac{1}{4} \sum_{j=1}^{J} \sum_{l=1}^{J} w_{jl} (x_j - x_l)^2$$

UPDATE ~ GRADIENT (or SCORE) of the LOG-LIKELIHOOD

$$x_j^{EM} = \frac{x_j^k}{\sum_{i=1}^{I} a_{ij}} \sum_{i=1}^{I} a_{ij} \frac{m_i}{(\boldsymbol{Ax}^k)_i}$$

UPDATE ~ GRADIENT OF the LOG PRIOR

$$x_j^{SM} = \frac{1}{2\sum_{l=1}^{J} w_{jl}} \sum_{l=1}^{J} w_{jl} (x_l^k + x_j^k)$$

$$x_j^{k+1} = \frac{2x_j^{EM}}{\left(1 - \beta v_j x_j^{SM}\right) + \sqrt{\left(1 - \beta v_j x_j^{SM}\right)^2 + 4\beta v_j x_j^{EM}}}$$

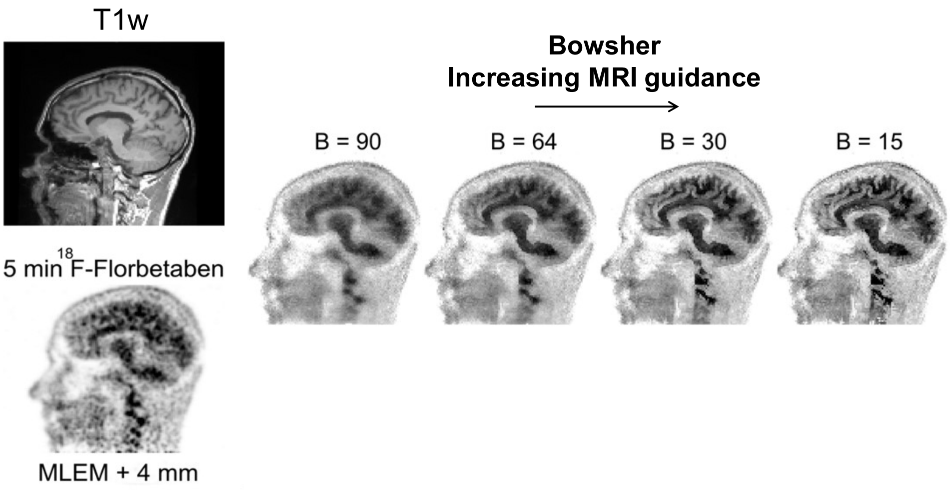
$$v_j = \frac{\sum_{l=1}^{J} w_{jl}}{s_j}$$

De Pierro TMI 1995

8

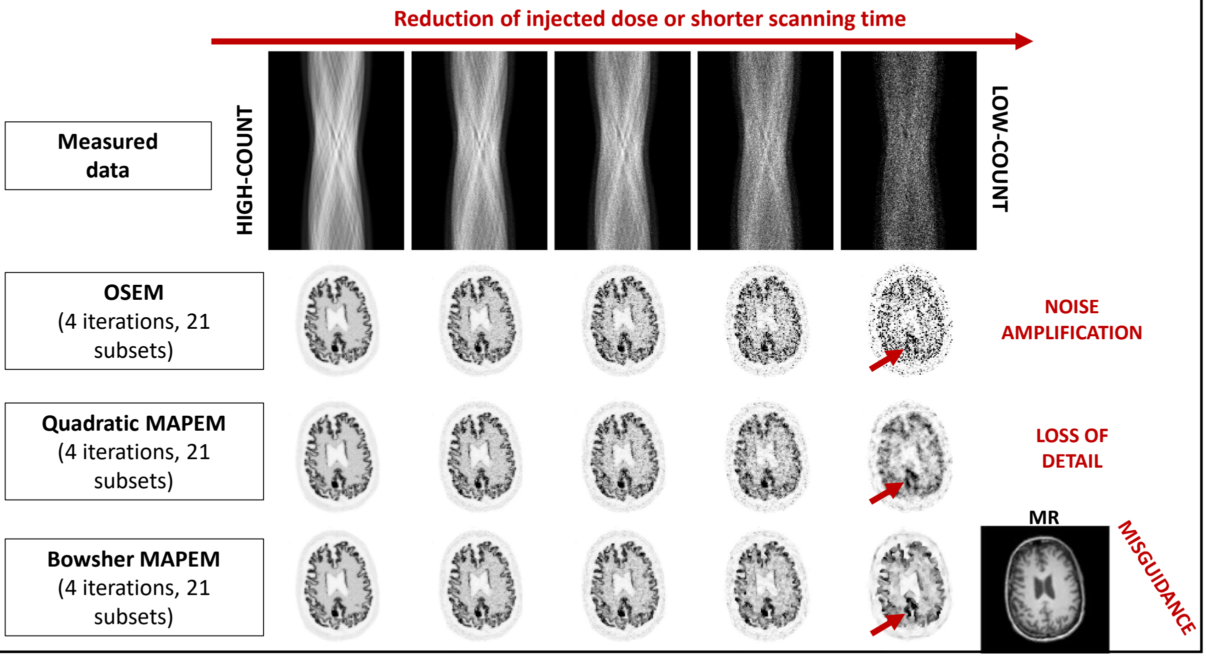
Example results

$$U(\boldsymbol{x}) = \sum_{j=1}^{J} \sum_{c \in N_j} w_{cj} \left(x_c - x_j \right)^2$$



Limitations of ML and MAP Image Reconstruction

Images courtesy
G. Corda –D'Incan



The motivation for deep learning

- Conventional MLEM and MAPEM
  - Noisy, low-resolution data → noisy images with Gibbs ringing
  - Noise compensation (*regularisation*) can be too simple (quadratic, TV, RDP, …)
    … or too imposing (e.g. MRI guidance can be wrong!)
- Assumes
  - Imaging system model
  - Data noise distribution
  - How to regularise (i.e. a model of how images should appear)

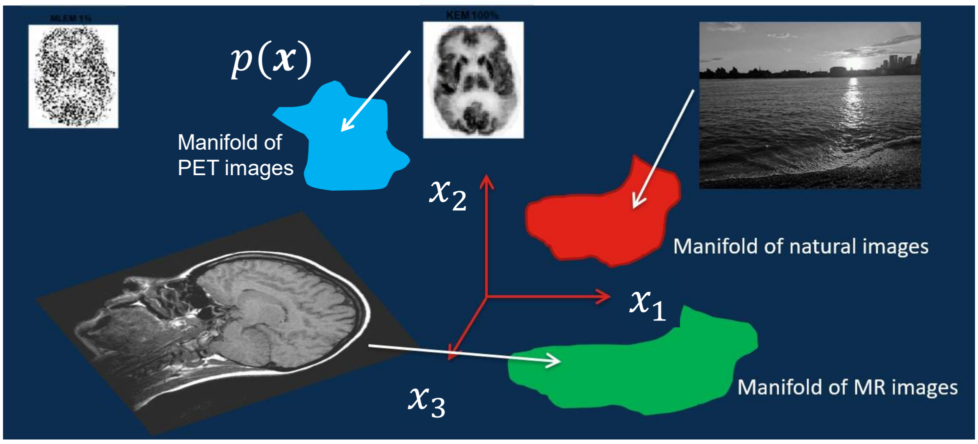
… but do we really know all these things?

- Deep learning can use
  - Real-data examples to learn
    - more accurate imaging and noise models (and their 'inverse')
  - Ground truth or high-quality reference data
    - to learn the probability distribution of high-quality images

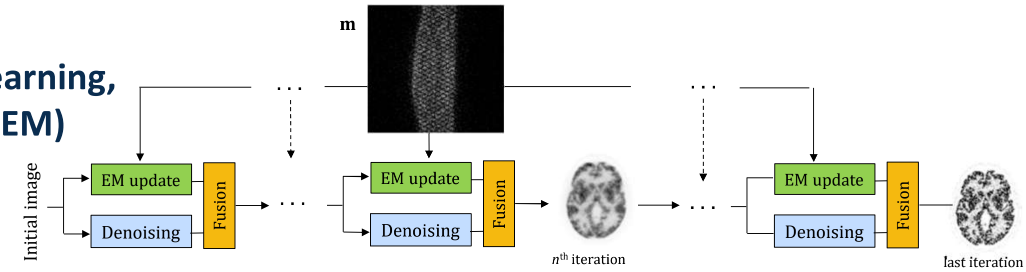
Data

- Ground truth or high-quality reference data
  - to learn the probability distribution of high-quality images

...with or without paired measurement data



FBSEM-Net (supervised learning, unrolled MAPEM)



MAP objective function

$$\hat{\boldsymbol{x}} = \arg\max_{\boldsymbol{x}} \sum_{i=1}^{I} \left(m_i \ln(\boldsymbol{Ax})_i - (\boldsymbol{Ax})_i\right) - \beta U(\boldsymbol{x})$$

$$U(\boldsymbol{x}) = \frac{1}{4}\sum_{j=1}^{J}\sum_{l=1}^{J} w_{jl}\left(x_j - x_l\right)^2$$

$$x_j^{EM} = \frac{x_j^k}{\sum_{i=1}^{I} a_{ij}} \sum_{i=1}^{I} a_{ij} \frac{m_i}{(\boldsymbol{Ax}^k)_i}$$

$$x_j^{SM} = CNN(x^k)$$

$$x_j^{k+1} = \frac{2x_j^{EM}}{\left(1 - \beta v_j x_j^{SM}\right) + \sqrt{\left(1 - \beta v_j x_j^{SM}\right)^2 + 4\beta v_j x_j^{EM}}}$$

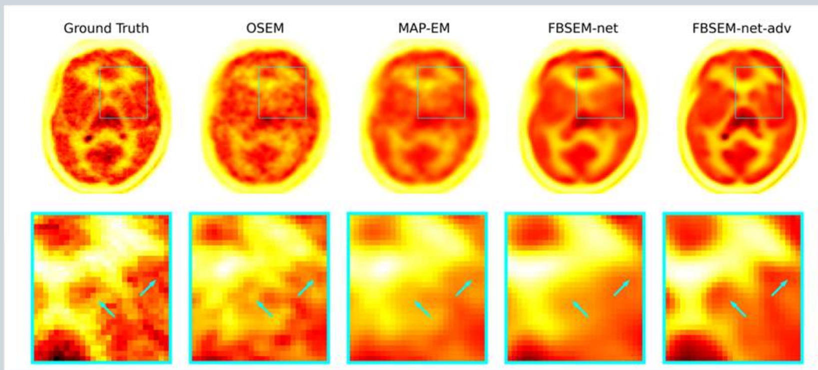
$$v_j = \frac{\sum_{l=1}^{J} w_{jl}}{s_j}$$

De Pierro TMI 1995

13

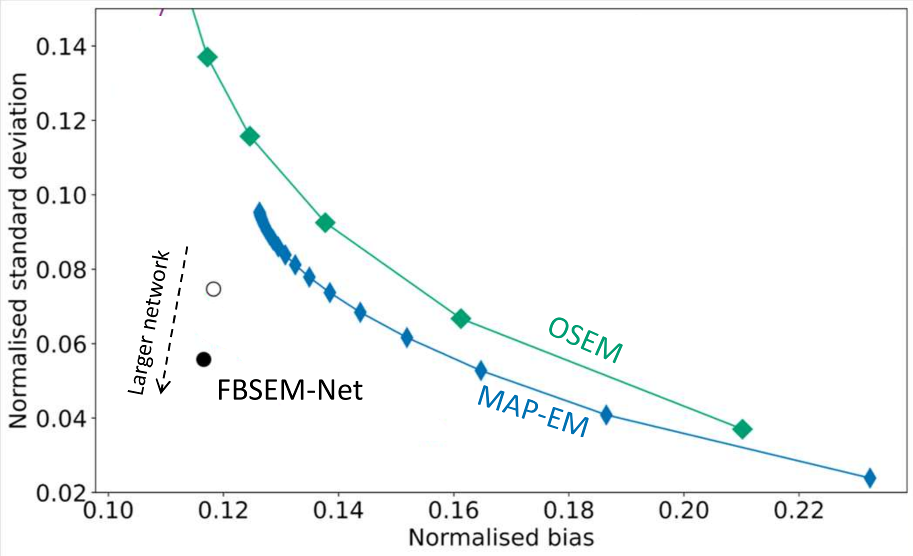
Example reconstructions – 2D realistic simulation data (low count)

Larger CNN



[6] G. Wang, J. Qi. Penalized likelihood PET image reconstruction using patch-based edge-preserving regularization. IEEE Trans Med Imaging. 2012 Dec;31(12):2194-204.
[7] A. Mehranian, A.J. Reader. Model-Based Deep Learning PET Image Reconstruction Using Forward-Backward Splitting Expectation-Maximization. *IEEE Trans Radiat Plasma Med Sci.* 2020 Jun 23;5(1):54-64. 14

Bias-variance assessment



	Model-based (MLEM, MAPEM…)		Unrolled DL methods (FBSEM-Net, …)	Generative AI methods (e.g. diffusion)
Deep learned parts of model	None	X	Images with Meas. Data	Images only
Handcrafted parts of model	Maths, physics, stats, images		Maths, physics, stats	Maths, physics, stats
Generalisation (for domain shift, OOD)	Excellent	X	Moderate	Good
Fully 3D recon?	Yes		Yes	Yes
Training data needs	N/A		~10 - 100 pairs (3D)	~10 - 100 (3D)
Network parameters	N/A		100k – 30 million	>30 million
Reconstruction time	Fast		Slow	Slow
Image quality (within training distribution)	**Moderate**	→	Good	Quite good
Strengths	**Simple, established, trusted, reliable(?)**		**Scanner-specific supervised learning**	**No training pairs needed. Generated samples: uncertainty**

Generative-Model-Based Fully 3D PET Image Reconstruction by Conditional Diffusion Sampling

George Webber[1], Yuya Mizuno[2], Oliver D Howes[2], Alexander Hammers[3], Andrew P King[1], Andrew J Reader[1]

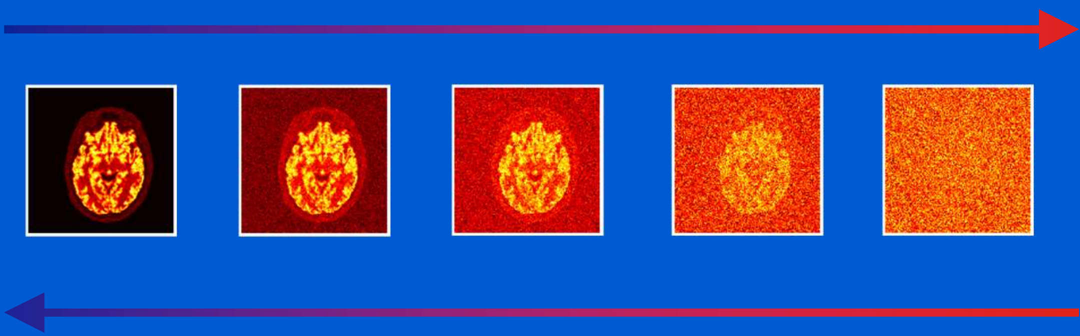
1. School of Biomedical Engineering & Imaging Sciences, King's College London, UK
2. Institute of Psychiatry, Psychology & Neuroscience, King's College London, UK
3. Guy's and St Thomas' PET Centre & King's College London, UK

EPSRC Centre for Doctoral Training
Smart Medical Imaging

KING'S College LONDON

17

Forward diffusion process (random)



Reverse diffusion process (generative)

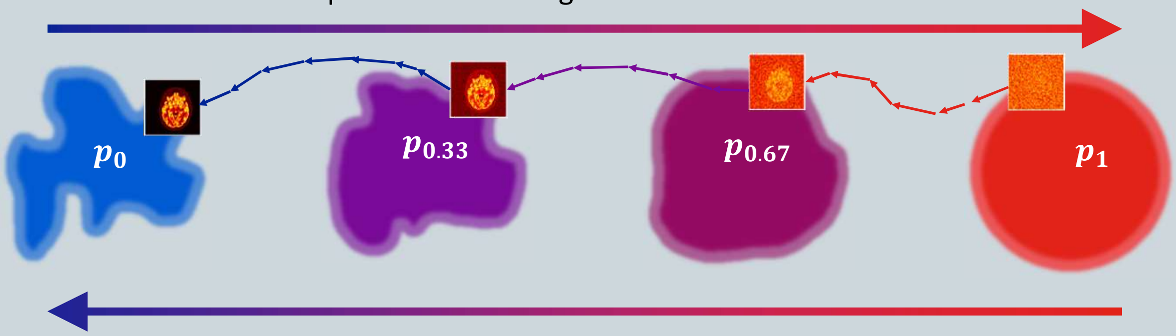
Score-based generative models (SGMs)

[3] J. Ho *et al*. "*Denoising diffusion probabilistic models*". In: Advances in Neural Information Processing Systems. Vol. 33. 2020:6840-6851.

18

Score-based generative models

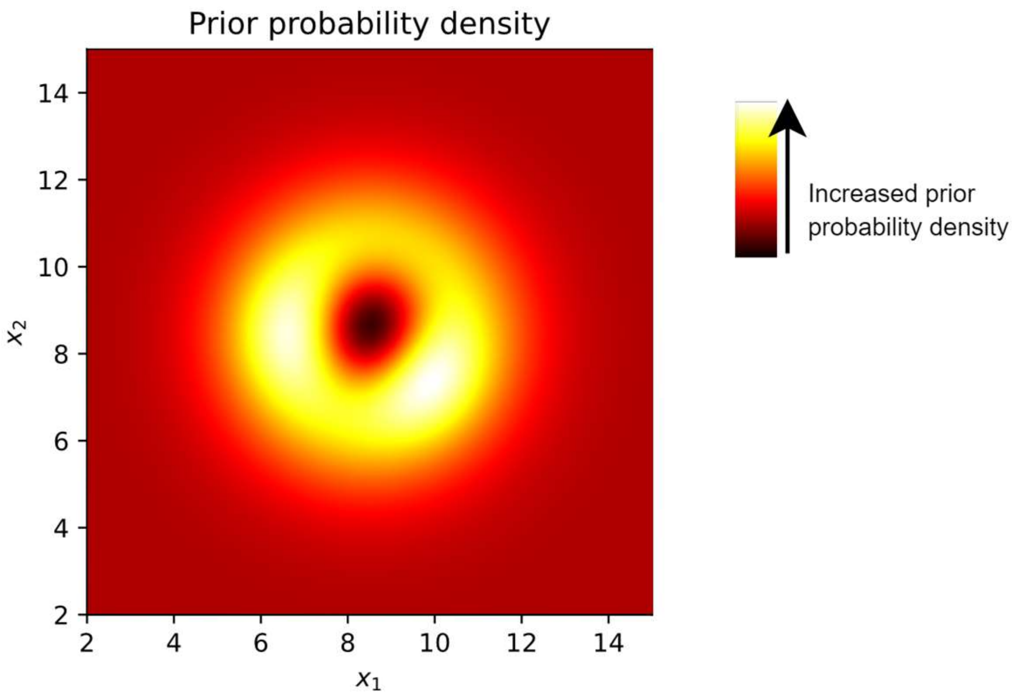
Forward diffusion process: increasing levels of additive Gaussian noise

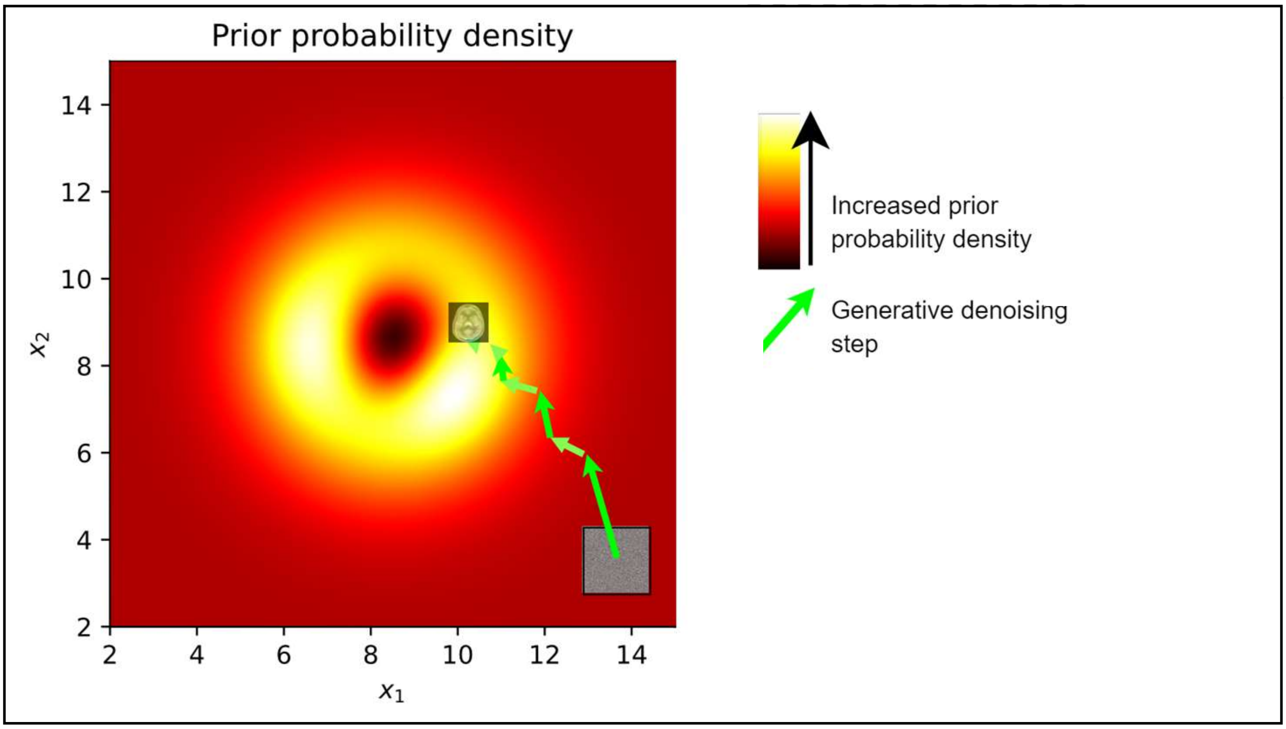


Reverse generative process, uses the gradient of $\log p_t(\boldsymbol{x}_t)$ SCORE VECTOR

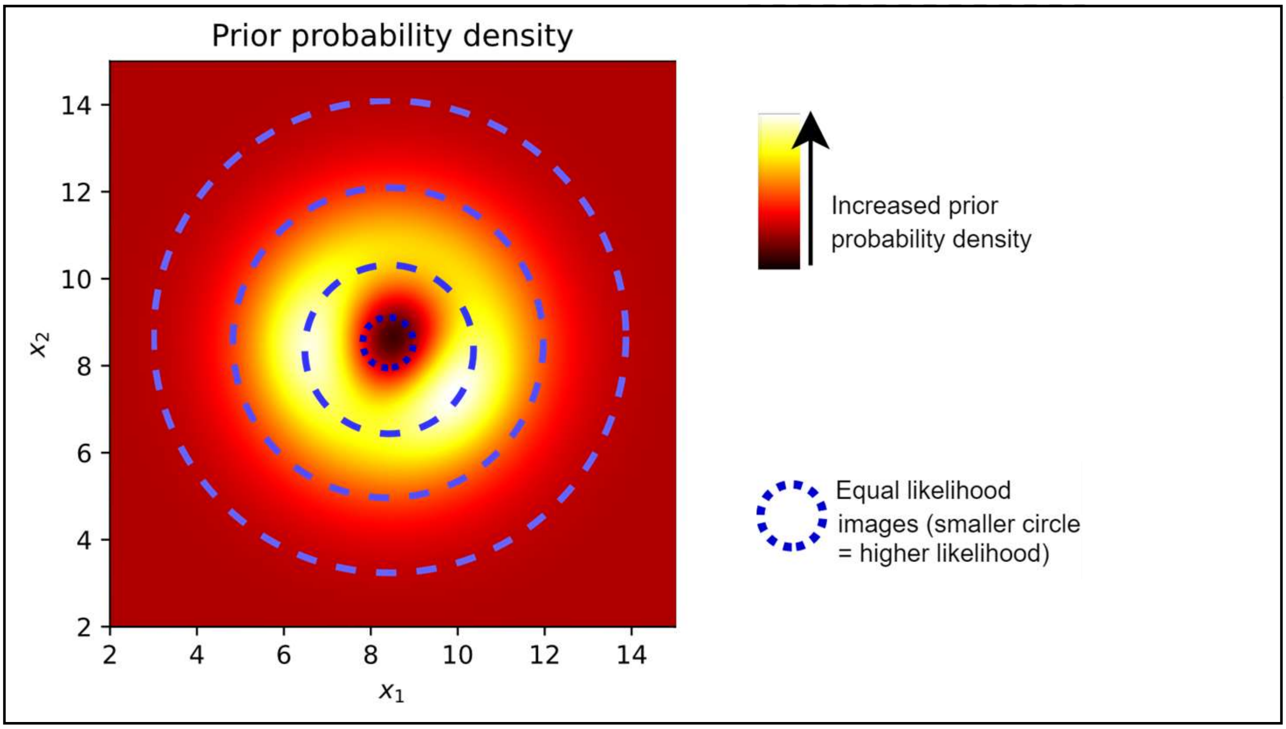
Approximate $\nabla_x \log p_t\,(\boldsymbol{x}_t)$ with $s_{\boldsymbol{\theta}}(\boldsymbol{x}_t, t)$

Just a denoising network $s_{\boldsymbol{\theta}}$!
Provide noise level t, and image made noisy at that level x_t : train network to map to the clean image x_0
In the generative process, renoise the denoised image, to be slightly less noisy than before $(t-1)$,
and enter back into the generative process at that new time step (noise level)

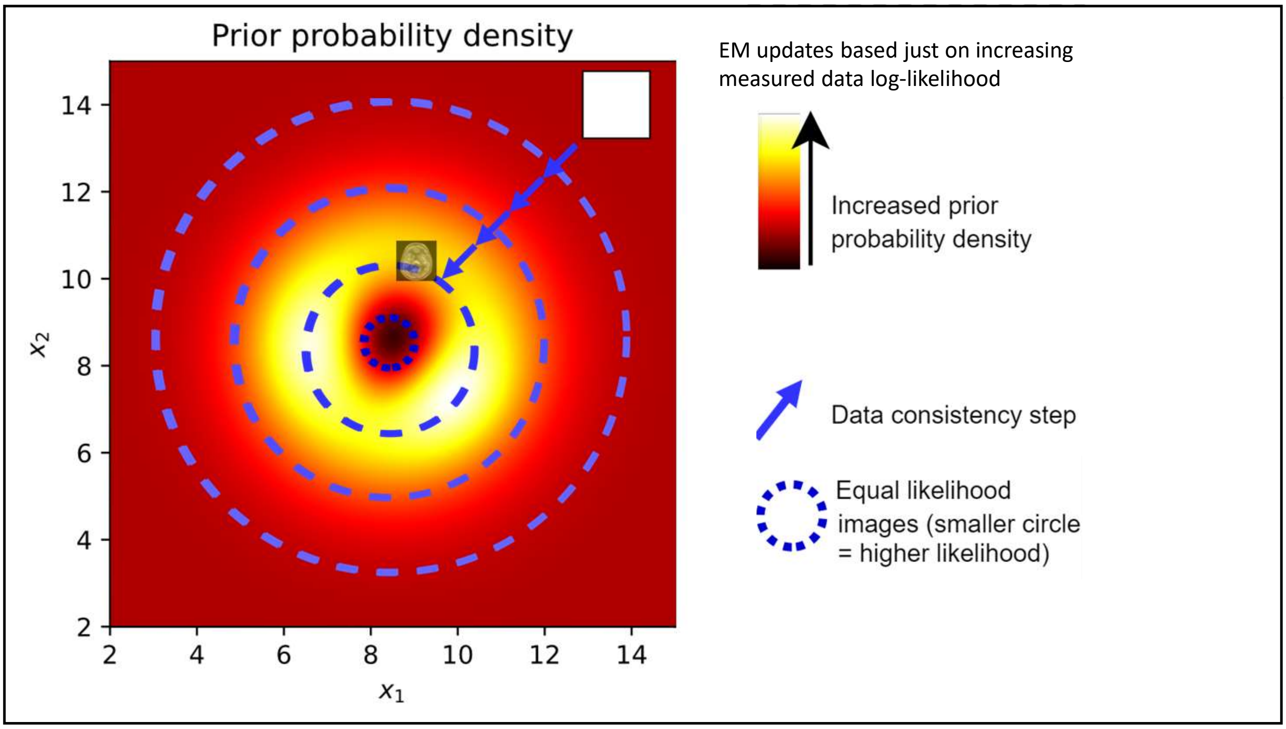



Prior probability density
Increased prior probability density
Generative denoising step
x_1
x_2

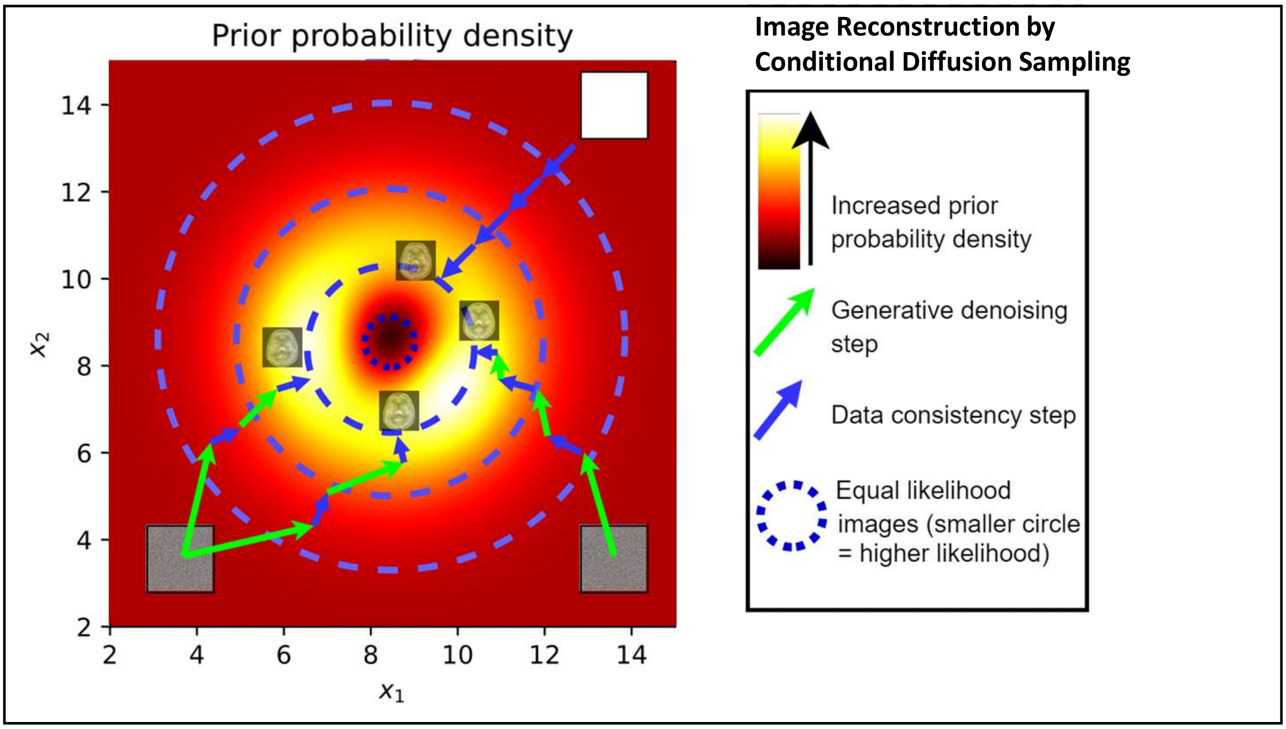
21


Prior probability density
Increased prior probability density
Equal likelihood images (smaller circle = higher likelihood)
x_1
x_2

22

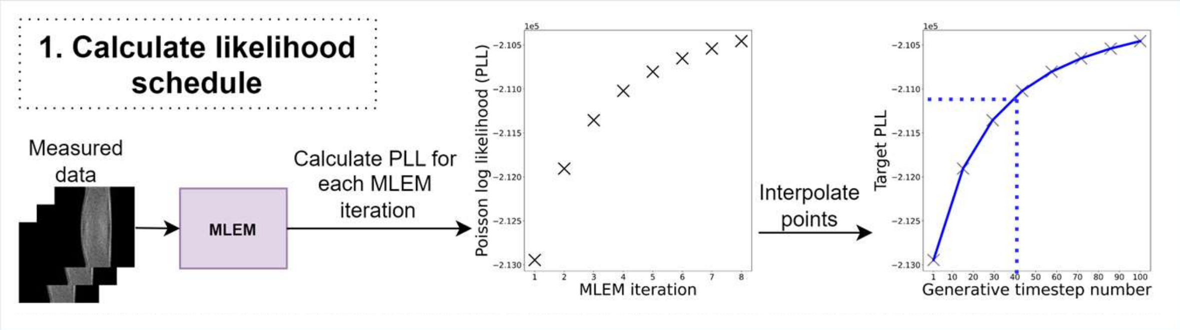

Prior probability density
EM updates based just on increasing measured data log-likelihood
Increased prior probability density
Data consistency step
Equal likelihood images (smaller circle = higher likelihood)
x_1
x_2
2 4 6 8 10 12 14

23

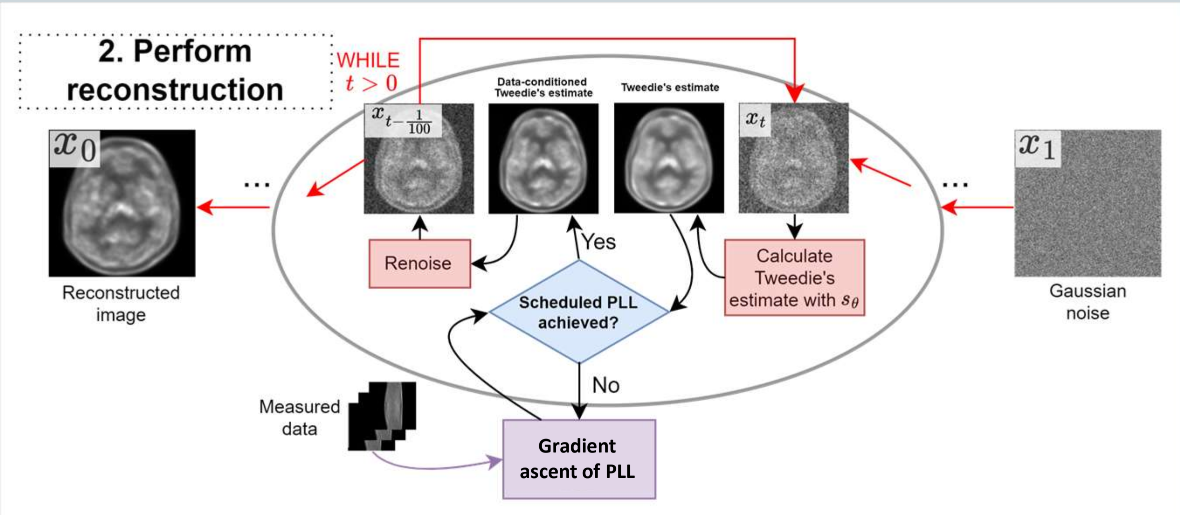

Prior probability density
Image Reconstruction by Conditional Diffusion Sampling
Increased prior probability density
Generative denoising step
Data consistency step
Equal likelihood images (smaller circle = higher likelihood)
x_1
x_2
2 4 6 8 10 12 14

24

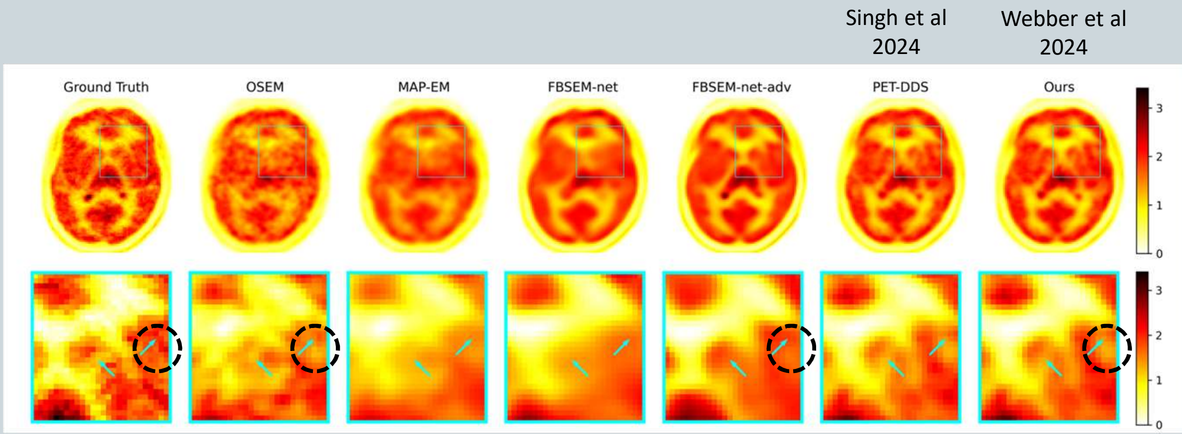
Our *likelihood-scheduled* approach



Our *likelihood-scheduled* approach



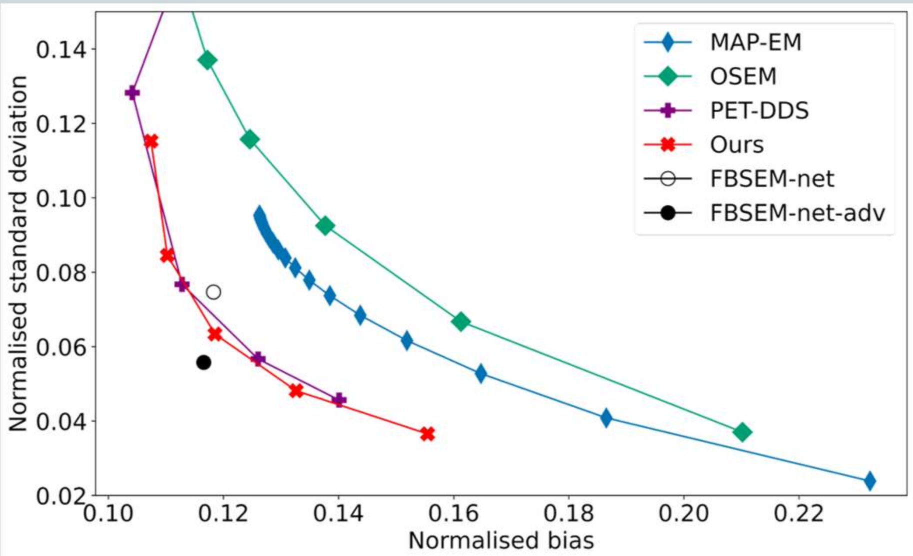
Example reconstructions – 2D realistic simulation data (low count)



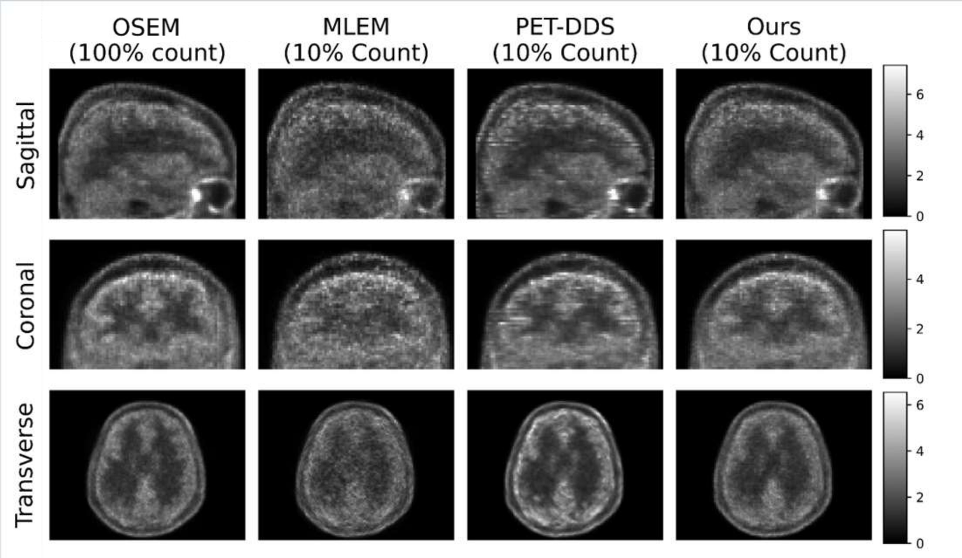
Reconstructions for SGM-based methods are the mean of 5 sample reconstructions.

[6] G. Wang, J. Qi. Penalized likelihood PET image reconstruction using patch-based edge-preserving regularization. IEEE Trans Med Imaging. 2012 Dec;31(12):2194-204.
[7] A. Mehranian, A.J. Reader. Model-Based Deep Learning PET Image Reconstruction Using Forward-Backward Splitting Expectation-Maximization. *IEEE Trans Radiat Plasma Med Sci.* 2020 Jun 23;5(1):54-64.

Bias-variance assessment



Fully 3D reconstruction (real [^{18}F]-DPA714 data)

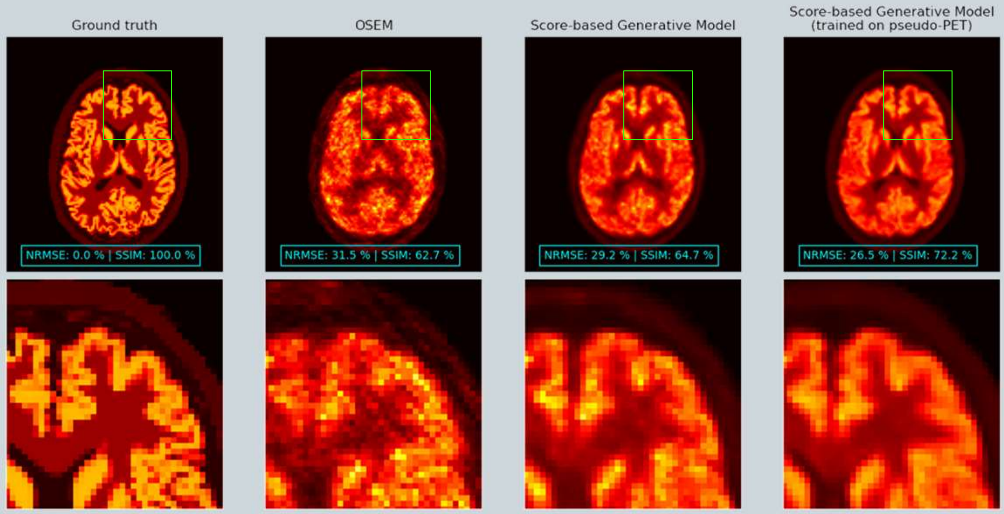


29

Alternative training data

POSTER
M-03-154
Webber *et al.*

- The quality an SGM-based method is limited by the choice of training data
- If we can provide better training data, we can obtain better results!



30

Self-Supervised Deep-Learned Fully 3D Filtered Backprojection for Image Reconstruction Objectives with a Poisson Likelihood

Movindu Dassanayake[1], Julia A. Schnabel and Andrew J. Reader

School of Biomedical Engineering and Imaging Sciences, King's College London, UK

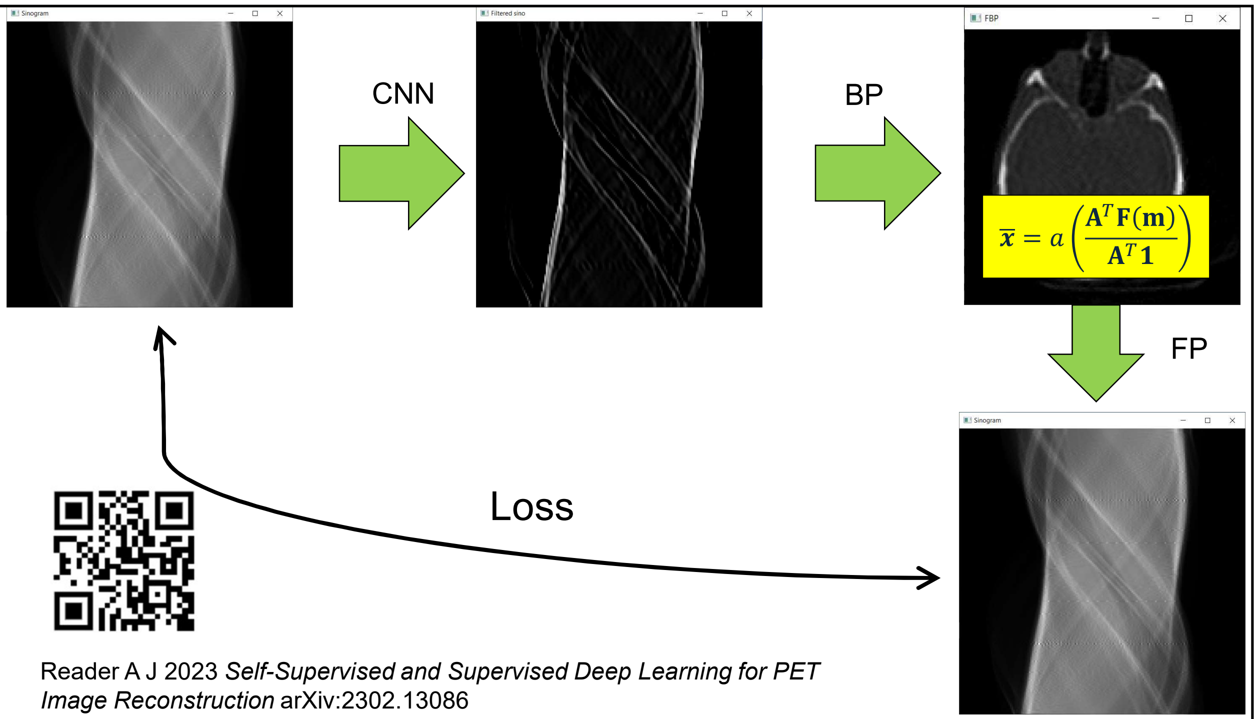
[1]movindu.dassanayake@kcl.ac.uk

KING'S College LONDON

Poster **M-03-175**

SIEMENS Healthineers · EPSRC Centre for Doctoral Training Smart Medical Imaging · UKRI Engineering and Physical Sciences Research Council

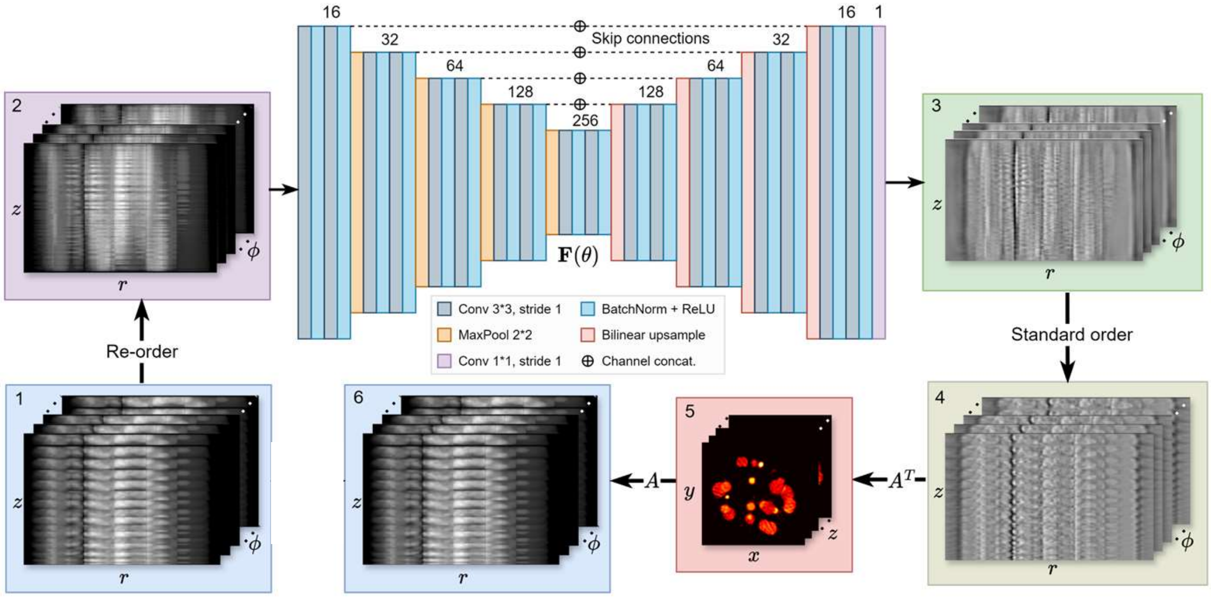
31



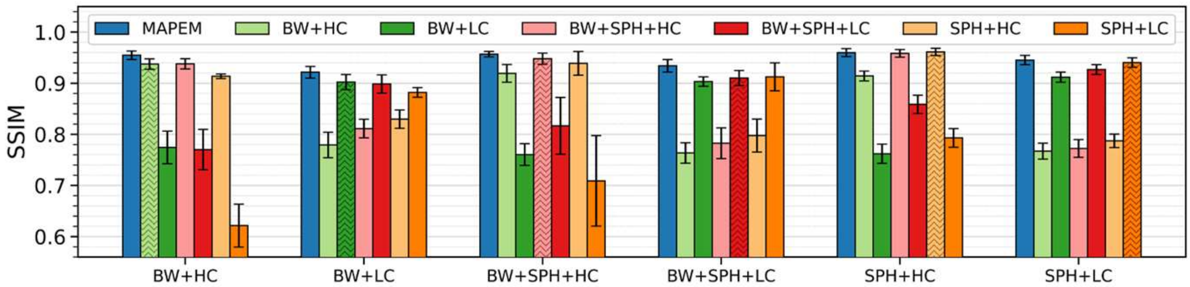
Reader A J 2023 *Self-Supervised and Supervised Deep Learning for PET Image Reconstruction* arXiv:2302.13086

32

Network

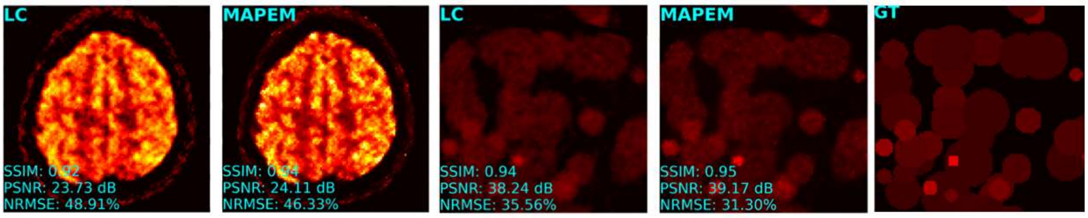


Results: MAPEM & DL-FBP-PLL (PLL+RDP)



DL-FBP-RDP generalizability between domains (at LC)

Similar performance to MAPEM is obtained with **high generalizability** between vastly different domains



Results: MAPEM & DL-FBP-PLL (PLL+RDP)

- Gaussian post-smoothing applied with increasing sigma: 0 – 5 x 5 x 5 mm³
- DL-FBP approaches a smaller bias at clinical post-smoothing (4 x 4 x 4 mm³) when compared to MLEM and MAPEM

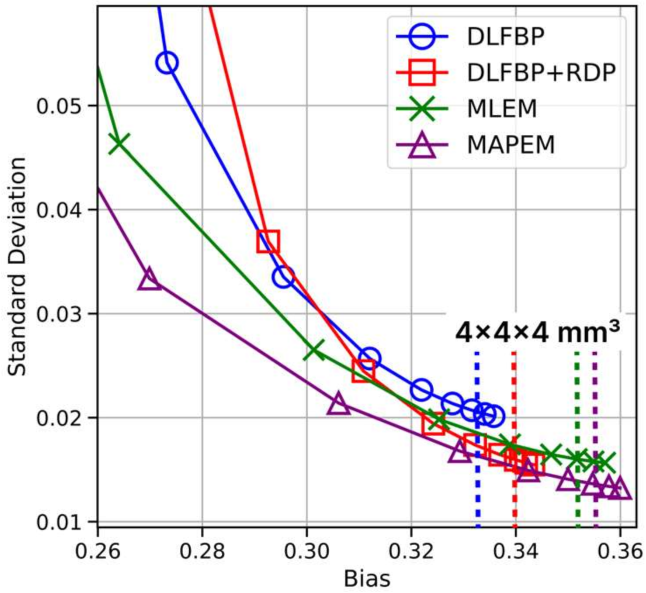
- DL-FBP-RDP can reconstruct

~**50x faster** than MAPEM (100 iterations)

~5x faster than MLEM (30 iterations)

- Reconstruction involves just one CNN,

one backprojection



KING'S College LONDON

Engineering and Physical Sciences Research Council

Thank you