



King's Research Portal

DOI:
[10.1109/TRPMS.2020.3004408](https://doi.org/10.1109/TRPMS.2020.3004408)

Document Version
Peer reviewed version

[Link to publication record in King's Research Portal](#)

Citation for published version (APA):
Mehranian, A., & Reader, A. J. (2020). Model-Based Deep Learning PET Image Reconstruction Using Forward-Backward Splitting Expectation Maximisation. *Transactions on Radiation and Plasma Medical Sciences*.
<https://doi.org/10.1109/TRPMS.2020.3004408>

Citing this paper

Please note that where the full-text provided on King's Research Portal is the Author Accepted Manuscript or Post-Print version this may differ from the final Published version. If citing, it is advised that you check and use the publisher's definitive version for pagination, volume/issue, and date of publication details. And where the final published version is provided on the Research Portal, if citing you are again advised to check the publisher's website for any subsequent corrections.

General rights

Copyright and moral rights for the publications made accessible in the Research Portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognize and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the Research Portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the Research Portal

Take down policy

If you believe that this document breaches copyright please contact librarypure@kcl.ac.uk providing details, and we will remove access to the work immediately and investigate your claim.

Model-Based Deep Learning PET Image Reconstruction Using Forward-Backward Splitting Expectation Maximisation

Abolfazl Mehranian[†] and Andrew J. Reader

[†]**Abstract**– We propose a forward backward splitting algorithm to integrate deep learning into maximum-a-posteriori (MAP) PET image reconstruction. The MAP reconstruction is split into regularisation, expectation-maximisation (EM) and a weighted fusion. For regularisation, use of either a Bowsher prior (using Markov-random fields) or a residual learning unit (using convolutional-neural networks) were considered. For the latter, our proposed forward backward splitting expectation-maximisation (FBSEM), accelerated with ordered-subsets (OS), was unrolled into a recurrent-neural network in which network parameters (including regularisation strength) are shared across all states and learned during PET reconstruction. Our network was trained and evaluated using PET-only (FBSEM-p) and PET-MR (FBSEM-pm) datasets for low-dose simulations and short-duration in-vivo brain imaging. It was compared to OSEM, Bowsher MAPEM and a post-reconstruction U-Net denoising trained on the same PET-only (Unet-p) or PET-MR (Unet-pm) datasets. For simulations, FBSEM-p(m) and Unet-p(m) nets achieved a comparable performance, on average, 14.4% and 13.4% normalised root-mean square error (NRMSE), respectively; and both outperformed OSEM and MAPEM methods (with 20.7% and 17.7% NRMSE respectively). For in-vivo datasets, FBSEM-p(m), Unet-p(m), MAPEM and OSEM methods achieved average root-sum-of-squared errors of 3.9%, 5.7%, 5.9% and 7.8% in different brain regions, respectively. In conclusion, the studied U-Net denoising method achieved a comparable performance to a representative implementation of FBSEM net.

Index Terms– Deep learning, image reconstruction, PET, MRI.

I. INTRODUCTION

Model-based image reconstruction of positron emission tomography (PET) has now almost superseded conventional reconstruction methods by accounting for all statistical and physical processes of data acquisition in the image reconstruction. Founded on a Bayesian framework, these techniques can even model the prior probability distribution of the unknown activity distribution. Different image priors have been proposed in the literature, particularly to suppress noise in the reconstructed images without compromising image quality [1]. Based on Markov random fields, the majority of these priors aim to assign a low probability to images that have large local intensity differences between their voxels based on the hypothesis that those differences are due to noise. The major limitation of these hypothesis-driven priors is that they might not only suppress noise but also legitimate image details and

boundaries, depending on the strength of hyperparameters chosen before reconstruction. Thus, edge-preserving and anatomically informed priors have been used to reduce noise while preserving PET details [2-5]. However, their performance highly depends on their functional form and hyperparameters, which are often hand-engineered and selected before reconstruction.

Machine learning and deep learning (DL) and techniques have recently shown promise in many aspects of PET imaging from photon detection to image reconstruction and quantification [6, 7]. In particular, deep convolutional neural networks (CNNs) have an immense potential to learn most representative image features from a multi-modal training space and hence give rise to data-driven priors which can surpass hypothesis-driven ones.

Recent developments for leveraging supervised deep learning techniques in PET image reconstruction can be categorised into three groups: i) direct mapping of PET sinograms to PET images using end-to-end neural networks [8, 9], ii) image enhancement of PET images in terms of noise [10, 11] or convergence [11, 12] and iii) model-based deep learning reconstruction, which combines DL with conventional model-based reconstruction methods [13]. Direct techniques aim to learn the whole process of image reconstruction including the PET system matrix, using fully connected as well as convolutional layers, resulting in a complex learning task for which a large and diverse training corpus is presumably required. Image enhancement techniques aim to map low-dose or low-resolution or under-converged images to their target full-dose, high-resolution and fully converged images using CNNs. On the other hand, DL reconstruction networks aim to merge the power of model-based Bayesian algorithms with neural networks through unrolling an iterative optimization algorithm, which provides an elegant theoretical foundation for designing robust data correction and image prior models.

Gong *et al* [13] proposed an unrolled network, based on the alternating direction method of multipliers (ADMM) algorithm, which alternates between PET maximum likelihood expectation maximisation (MLEM) reconstruction and supervised learning of a deep image prior using a U-Net model [14]. In [15], the authors further explored unsupervised learning of this network using MR images and noisy PET images as inputs and targets, respectively. To ensure the convergence of the network, the ADMM's penalty parameter was

A. Mehranian and A.J Reader are with the School of Biomedical Engineering and Imaging Sciences, Department of Biomedical Engineering, King's College London, St. Thomas' Hospital, London, UK. [†]Email: Abolfazl.Mehranian@kcl.ac.uk

Copyright (c) 2020 IEEE. Personal use of this material is permitted. However, permission to use this material for any other purposes must be obtained from the IEEE by sending a request to pubs-permissions@ieee.org.

experimentally chosen and U-Net’s parameters were initialised using separately reconstructed PET images. In [16], the authors used a similar deep image prior for unsupervised de-noising of PET images. For optimally chosen number of training iterations (epochs), it was shown this method outperformed a number of known de-nosing methods.

Lim *et al* [17] proposed a deep learning reconstruction network by unrolling a block coordinate descent (BCD) algorithm, which alternates between MLEM PET reconstruction and iteration-dependent denoising modules that are composed of convolution, soft-thresholding (as an activation function) and deconvolution layers. The network’s regularisation hyperparameter was finely tuned to achieve the highest contrast-to-noise ratio.

In this work, we i) propose an optimization algorithm for Bayesian maximum *a posteriori* (MAP) PET image reconstruction, which generalizes De Pierro’s MAPEM algorithm [18] for any differentiable convex prior, ii) unroll the resulting algorithm into a model-based deep reconstruction network in which CNNs are used to learn image features while activity images are reconstructed from emission data, iii) learn any hyperparameter from data, iv) optionally incorporate anatomical side information into PET reconstruction without substantial suppression of PET unique features as seen with conventional MR-guided reconstruction algorithms, and importantly v) investigate whether deep reconstruction methods can outperform DL-based denoising methods and whether redesign of the current reconstruction workflow for PET scanners is justified, which would be required for clinical deployment. In this paper, the proposed deep reconstruction network was evaluated using realistic 3D brain simulations and in-vivo PET-MR scans and was compared with conventional ordered subsets expectation maximisation (OSEM), MR-guided MAPEM and DL-based denoising using PET only and PET-MR input channels.

II. MATERIAL AND METHODS

A. Forward-backward splitting expectation maximisation (FBSEM)

The Bayesian maximum *a posteriori* (MAP) reconstruction of PET emission data is obtained by the following maximisation:

$$\hat{\mathbf{x}} = \operatorname{argmax}_{\mathbf{x}} \{L(\mathbf{y}|\mathbf{x}) - \beta R(\mathbf{x})\}$$

$$L(\mathbf{y}|\mathbf{x}) = \sum_i y_i \log([\mathbf{H}\mathbf{x}]_i + \bar{b}_i) - ([\mathbf{H}\mathbf{x}]_i + \bar{b}_i) \quad (1)$$

where L is the Poisson log-likelihood of measured data, \mathbf{y} , given an activity distribution, \mathbf{x} . \mathbf{H} is a system matrix and $\bar{\mathbf{b}}$ is the expected accidental coincidences. R is a penalty function that imposes prior information about \mathbf{x} , controlled by the regularization parameter β . Eq. (1) can be solved using optimisation transfer techniques as long as a separable, differentiable and convex surrogate can be defined for R . Consequently, a monotonically convergent MAP expectation maximisation (EM) algorithm is obtained [18]. In this work, we use a forward-backward splitting (FBS) algorithm [19] for solving Eq. (1) for any differentiable convex prior. The FBS

algorithm in fact generalises the projected gradient descent (a.k.a. Landweber algorithm) by substituting its projection operator with a proximal mapping operator. As a result, the optimisation is performed in the following steps:

$$\mathbf{x}_{Reg}^{(n)} = \mathbf{x}^{(n-1)} - \gamma\beta\nabla R(\mathbf{x}^{(n-1)}) \quad (2)$$

$$\mathbf{x}^{(n)} = \operatorname{argmax}_{\mathbf{x}} \left\{ L(\mathbf{y}|\mathbf{x}) - \frac{1}{2\gamma} \|\mathbf{x} - \mathbf{x}_{Reg}^{(n)}\|^2 \right\} \quad (3)$$

Eq. (2) is the gradient descent minimization of R with the step size of γ , whereas Eq. (3) is a proximal mapping [19] associated with the log-likelihood L with $1/\gamma$ as a regularization hyperparameter that controls the data fidelity of \mathbf{x} to \mathbf{y} and its proximity to $\mathbf{x}_{Reg}^{(n)}$. The quadratic prior in Eq. (3) is separable. Following [20], a separable surrogate is then defined for the function L , whereby Eq. (3) can be rewritten as:

$$\mathbf{x}^{(n)} = \operatorname{argmax}_{\mathbf{x}} \sum_j x_{j,EM}^{(n)} \ln(x_j) - x_j - \frac{1}{2\gamma s_j} (x_j - x_{j,Reg}^{(n)})^2 \quad (4)$$

where $\mathbf{x}_{EM}^{(n)}$ is given by the following MLEM update:

$$x_{j,EM}^{(n)} = \frac{x_j^{(n-1)}}{s_j} \sum_i \frac{h_{ij} y_i}{\sum_k h_{ik} x_k^{(n-1)} + \bar{b}_i}, \quad s_j = \sum_i h_{ij} \quad (5)$$

By setting the derivative of the objective function of Eq. (4) to zero, a closed-form solution is obtained as follows [21]:

$$x_j^{(n)} = \frac{2x_{j,EM}^{(n)}}{\left(1 - \delta_j x_{j,Reg}^{(n)}\right) + \sqrt{\left(1 - \delta_j x_{j,Reg}^{(n)}\right)^2 + 4\delta_j x_{j,EM}^{(n)}}} \quad (6)$$

$$\delta_j = \frac{1}{\gamma s_j}$$

Algorithm 1 summarises the resulting FBSEM algorithm, which is accelerated by ordered subsets (OS) method. As a result, the optimisation of Eq. (1) is split into three steps: *regularization* of the previous image estimate, Eq. (7), *EM update* of the previous image estimate, Eq. (8), and *fusion* of the resulting two images, Eq. (9), weighted by γ and the subset-dependent sensitivity image $\mathbf{s}^{(m)}$.

ALGORITHM 1: FBSEM FOR MAP PET IMAGE RECONSTRUCTION

Initialize: $\mathbf{x}^{(0,1)} = \mathbf{1}$, number of iterations (N_{it}) and subsets (N_{sub})

For $n = 1, \dots, N_{it}$

For $m = 1, \dots, N_{sub}$

$$x_{j,Reg}^{(n,m)} = x_j^{(n-1,m)} - \gamma\beta \frac{\partial}{\partial x_j} R(\mathbf{x}^{(n-1,m)}) \quad (7)$$

$$x_{j,EM}^{(n,m)} = \frac{x_j^{(n-1,m)}}{s_j^{(m)}} \sum_{i \in \Omega_m} \frac{h_{ij} y_i}{\sum_k h_{ik} x_k^{(n-1,m)} + \bar{b}_i} \quad (8)$$

$$x_j^{(n,m)} = \frac{2x_{j,EM}^{(n,m)}}{\left(1 - \delta_j x_{j,Reg}^{(n,m)}\right) + \sqrt{\left(1 - \delta_j x_{j,Reg}^{(n,m)}\right)^2 + 4\delta_j x_{j,EM}^{(n,m)}}} \quad (9)$$

$$\delta_j = \frac{1}{\gamma s_j^{(m)}}, \quad s_j^{(m)} = \sum_{i \in \Omega_m} h_{ij}$$

End

End

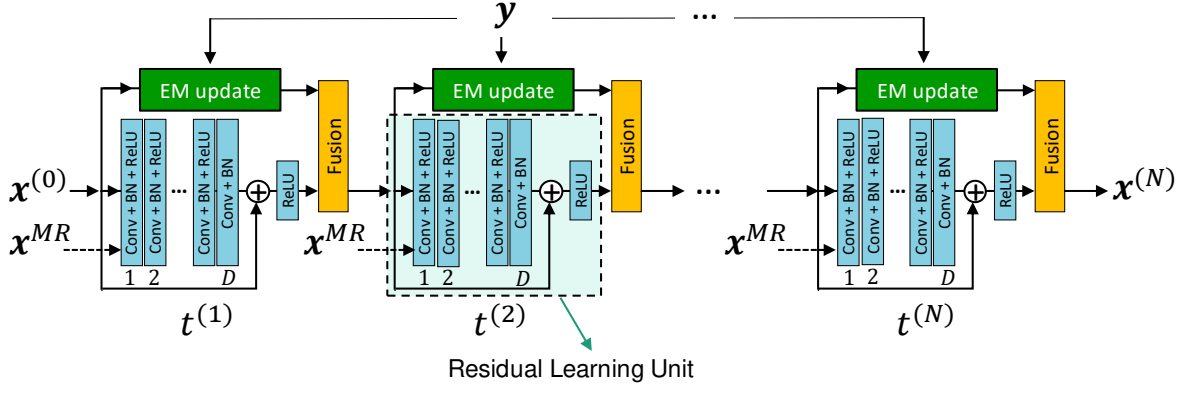


Fig. 1 Architecture of the proposed network using a residual learning unit with D layers of convolution (Conv) filters, batch normalization (BN) and rectified linear unit (ReLU). All trainable parameters including the regularization parameter, which is the only trainable parameter in the fusion block, are shared across all reconstruction states (t). In this network, a co-registered MR image can be optionally used as a second input channel to each state.

Algorithm 1 can be used for the following commonly used quadratic prior, weighted by MR information (w_{jb}):

$$R(x) = \frac{1}{2} \sum_j \sum_{b \in \mathcal{N}_j} w_{jb} (x_j - x_b)^2 \quad (10)$$

where \mathcal{N}_j is a neighbourhood of voxels around j th voxels. For this prior, by setting $\beta = \frac{1}{2}$, $\gamma = \frac{1}{\sum_b w_{jb}}$ in Eq. (7), we obtain:

$$x_{j,Reg}^{(n,m)} = \frac{1}{2 \sum_b w_{jb}} \sum_{b \in \mathcal{N}_j} w_{jb} (x_j^{(n-1,m)} + x_b^{(n-1,m)}) \quad (11)$$

whereby, *Algorithm 1* is reduced to De Pierro's MAPEM algorithm [18]. As $\gamma \rightarrow \infty$, this algorithm reduces to the OSEM algorithm. In this paper, we used a CNN-based model for R and unrolled the FBSEM algorithm into an recurrent neural network (RNN) with $N = N_{it} \times N_{sub}$ reconstruction states, in which model parameters are shared across all states, hence the number of trainable parameters became independent of the number of reconstruction updates [22]. As shown in Fig. 1, a D -layer learning unit with a non-negativity constraint (imposed by a final rectified linear unit (ReLU) layer) was used, whereby Eq. (7) was converted to a residual learning unit [23]. Of course, alternative CNN models such as convolutional encoder-decoders (e.g. U-Net [14]) could also be used.

The proposed network was trained in a supervised manner using a training dataset composed of N_s reference high-definition high-dose (HD) PET images (\mathbf{x}_s^{Ref}), low-definition low-dose (LD) PET sinograms ($\mathbf{y}_s, \bar{\mathbf{b}}_s$) and optionally co-registered MR images (\mathbf{x}_s^{MR}). The training was formulated as the minimization of the mean-squared-error loss function between the network's output ($\mathbf{x}_s^{(N)}$) and the reference image \mathbf{x}_s^{Ref} :

$$\hat{\theta} = \underset{\theta}{\operatorname{argmin}} \frac{1}{N_s} \sum_{s=1}^{N_s} \|\mathbf{x}_s^{(N)} - \mathbf{x}_s^{Ref}\|^2 \quad (12)$$

$$\mathbf{x}_s^{(N)} = \text{FBSEM}_{\theta}(\mathbf{y}_s, \bar{\mathbf{b}}_s, \mathbf{x}^{(0)}, \mathbf{x}_s^{MR}),$$

$$N = N_{it} \times N_{sub}$$

where model parameters, $\theta \in \mathbb{R}^d$, include convolution kernels, biases, batch normalization parameters and γ . $\mathbf{x}^{(0)}$ is an initial

image estimate. Eq. (12) was optimized using the Adam optimiser.

B. Simulation and in-vivo datasets

T1-weighted MPRAGE MR images of 70 epilepsy and dementia patients, referred for PET-MR brain imaging at PET Centre St. Thomas's Hospital in London, were used for generating realistic brain PET-MR phantoms. The MR image matrix and voxel sizes were $230 \times 230 \times 254$ and $1.04 \times 1.04 \times 1.01 \text{ mm}^3$, respectively. The images were segmented into grey matter (GM), white matter (WM), cerebrospinal fluid (CSF), skull and skin using the SPM12 software¹. For each dataset, an FDG PET phantom was generated as follows. Random uptake values of 96.0 ± 5.0 and 32.0 ± 5.0 (arbitrary units) were respectively assigned to GM and WM regions, leading to uptake ratio of 3:1 between GM and WM. Low values (>16) were assigned to the remaining regions. Four circular lesions with random radii in 2-8 mm (uniformly distributed) and random locations were generated with the hot-to-cold ratio of 50%. The uptake value of 144.0 (1.5 \times of GM) was assigned to hot lesions and 48.0 (0.5 \times of WM) was assigned to cold ones. An attenuation map was also generated by assigning attenuation values of 0.13, 0.0975 and 0 cm^{-1} to skull, tissues and air, respectively. The PET, attenuation map and T1-MR images were resampled into the voxel sizes and field of view of the standard PET images from the Siemens mMR scanner, with matrix and voxel sizes of $344 \times 344 \times 128$ and $2.08 \times 2.08 \times 2.03 \text{ mm}^3$, respectively.

For data augmentation, the resulting images were rotated in the axial direction with 3 random angles within $[0, 15]$ degrees, resulting in 210 datasets. Noisy sinograms were then generated, using image-space point-spread-function (PSF) modelling in the forward model, attenuation, normalisation and Poisson noise. Random and scatter coincidences were not modelled. Each sinogram had a matrix size of 344 (radial bins) \times 252 (azimuthal angles) \times 837 (sinogram plans), as per the standard sinogram format for the mMR scanner.

For each dataset, a high-definition high-dose (HD) sinogram and a low-definition low-dose (LD) sinogram were generated. For HD sinograms, 1 billion counts and a PSF with 2.5 mm full-

¹ <https://www.fil.ion.ucl.ac.uk/spm/>

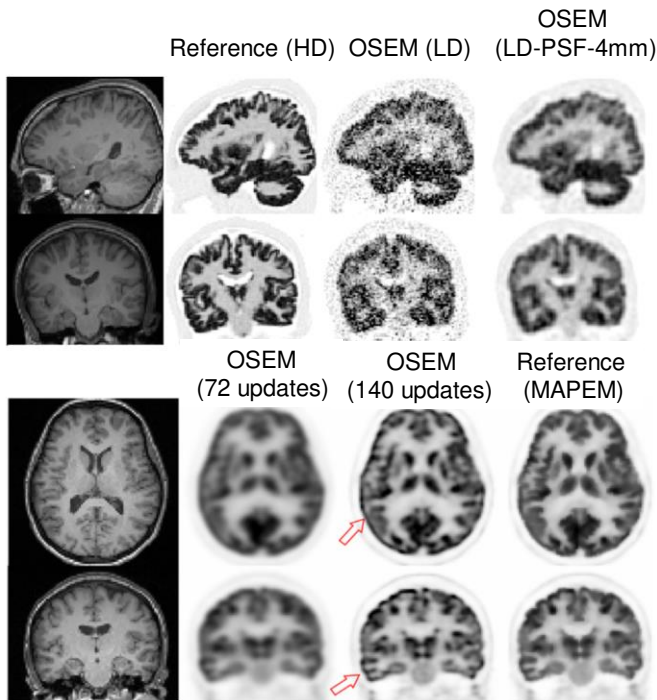


Fig. 2 Top: Reference high-definition high-dose (HD) images and reconstructed low-definition low-dose (LD) images of a sample simulation dataset. Bottom: full-dose (30 min) images of an in-vivo dataset reconstructed using 72 EM updates (vendor’s default) and 140 updates. Increasing the number of updates improved the convergence but led to PSF Gibbs artefacts (see arrows). Thus, MR-guided MAPEM (210 updates) was used as a reference.

width-at-half-maximum (FWHM) Gaussian kernels were considered, while for LD ones randomly chosen count levels in [90-120] million with PSF of 4.5 mm FWHM were considered. The HD data were reconstructed using the OSEM algorithm with $N_{it} = 10$, $N_{sub} = 14$ and $PSF = 2.5\text{ mm}$ to generate a reference image. The LD data were reconstructed using the same number of updates with and without PSF modelling ($PSF = 4\text{ mm}$) and with post-reconstruction Gaussian filtering (4-mm FWHM). Fig. 2, top, shows one simulated example dataset. For in-vivo datasets, 45 PET-MR brain datasets of patients suspected of epilepsy and dementia were retrospectively collected from our PET centre in St. Thomas’s Hospital. Following the injection of $\sim 220\text{ MBq}$ [^{18}F]FDG and uptake time of 60 min, patients underwent a simultaneous T1-MPRAGE MR scan and a 30-min PET scan on a Siemens mMR scanner. MR acquisition and parameters were as follows: repetition time: 1700 ms, echo time: 2.63 ms, inversion time: 900 ms, number of averages: 1, flip angle: 9 degrees and acquisition time of 382 seconds. For PET attenuation correction, a standard Dixon sequence and a UTE sequence was performed. The MR images were rigidly registered to PET images using SPM12 with default co-registration parameters and normalized mutual-information cost function.

The datasets were split into 35 training and 10 test ones. PET list-mode data were histogrammed into full-dose 30-min sinograms and low-dose 2-min sinograms. For one test dataset, the list-mode data were further histogrammed into 1-min and 30-sec sinograms. To obtain a HD reference image, the full-dose sinograms were reconstructed with PSF modelling (using

image-space 4 mm FWHM Gaussian kernels) and a MR-guided MAPEM algorithm with a quadratic prior, Eq. (10), weighted using Bowsher method [24] with \mathcal{N} of $3 \times 3 \times 3$. The FBSEM algorithm was used for optimisation. The regularisation parameter, β , was chosen to be as small as possible while still mitigating PSF Gibbs-like artefacts.

Fig. 2, bottom, shows a full-dose (30 min) brain scan of a subject reconstructed by i) OSEM using 72 updates followed by 5 mm Gaussian smoothing (Siemens e7 tools’ default), ii) OSEM with 140 updates and iii) MR-guided MAPEM with 210 updates. As shown the default reconstruction suffers from over-smoothing and lack of convergence, while increasing the number of updates improved the convergence at the cost of PSF overshoot artefacts (see arrows). The MAPEM image shows less artefacts.

C. Network Training

In this study, the proposed FBSEM was trained and evaluated in two modes: one with PET-only inputs (FBSEM-p) and one with both PET and MR inputs (FBSEM-pm). For comparison purposes, we also included a post-reconstruction U-Net denoising model trained on the same datasets with two modes: using PET only (Unet-p) and then using both PET and MR (Unet-pm) input images. For this purpose, the original U-Net proposed in [14] was extended to 3D with two modifications: inclusion of batch normalisation (BN) before ReLU activation and using trilinear upsampling in the decoder part of the network. The FBSEM and U-Net networks were trained in supervised learning sessions using both simulation and in-vivo training datasets. Each training dataset consists of a low-dose sinogram, attenuation and normalisation correction factors, scatter and random sinograms, reference (HD/full-dose) PET images, low-dose PET images and co-registered MR images. All sinograms were generated using Siemens e7 tools. All reference and low-dose sinograms were corrected for frame length and radionuclide decay before reconstruction, hence the resulting PET images were in counts-per-second units and had a similar dynamic range, which helped accelerate the training of the networks.

Table I summarises the number of training and test datasets used for the training of the networks together with other parameters that were experimentally chosen. In this table, depth refers to the number of layers in FBSEM net (see Fig.1) and number of resolution levels (or scales) in U-Net. Both networks were implemented in PyTorch and trained on a workstation equipped with a Nvidia Quadro k6000 12GB graphic card. Thanks to the parallelism of FBSEM net, the EM-update module was implemented in Python using a GPU-enabled PET projector, while regularization and fusion modules (with trainable parameters) were implemented in PyTorch with GPU acceleration.

The training of unrolled 3D reconstruction networks is extremely time-consuming and memory demanding. To tackle these issues for training of the proposed FBSEM-p(m) nets on both simulations and in-vivo datasets, the following fivefold strategy was used. First, i), the sinograms were radially trimmed by a factor of 3 and accordingly our PET projector was modified, and secondly, ii), data minibatch size was set to 1. Thirdly, iii), the networks were initialised with OSEM PET

TABLE I. THE TRAINING/TEST DATASETS, MODEL ARCHITECTURES AND TRAINING PARAMETERS USED IN THIS STUDY.

EXPERIMENT	MODEL	No. TRAINING DATASETS	No. TEST DATASETS	No. KERNELS IN 1 ST LAYER	KERNEL SIZE	DEPTH *	No. TRAINABLE PARAMETERS	No. BATCHES	LEARNING RATE	No. EPOCHS	OPTIMISER	LOSS FUNCTION
SIMULATION	UNET-P(M)	200	10	32	3×3×3	4	10,043,073 (10,043,937)	1	0.05	100	ADAM	MSE
	FBSEM-P(M)	"	"	16	"	9	49,636 (50068)	"	0.1	100	"	"
IN-VIVO	UNET-P(M)	35	10	70	"	4	48,039,881 (48,041,771)	"	0.005	200	"	"
	FBSEM-P(M)	"	"	37	"	3	76,261 (77,260)	"	0.005	200	"	"

* NUMBER OF DOWN/UP SAMPLING LEVELS FOR THE U-NET, AND NUMBER OF CONVOLUTIONAL LAYERS FOR FBSEM NET.

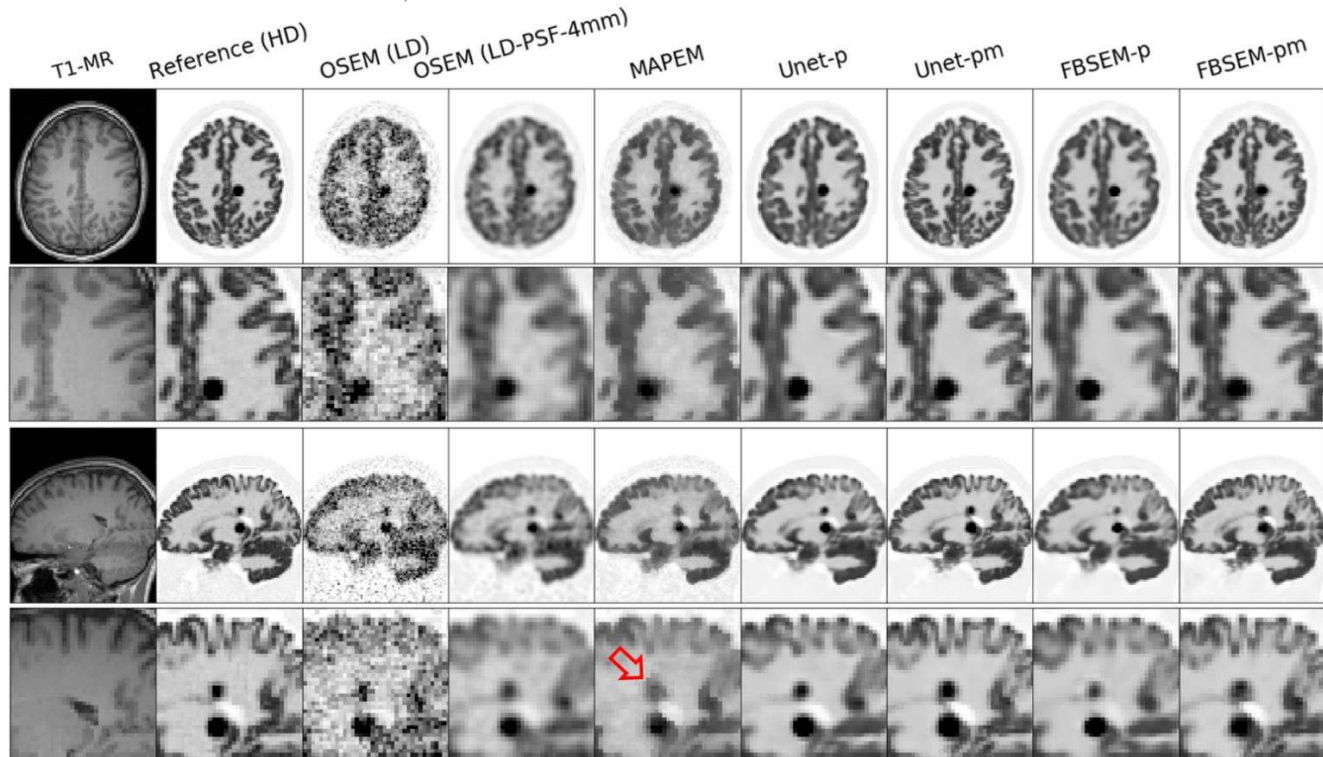


Fig. 3 Reconstruction results of a test simulation dataset with two adjacent hot lesions for different reconstruction methods. The arrow shows MAPEM notably suppresses the lesion.

images (10 iterations and 4 subsets), and iv) the networks were unrolled for 12 reconstruction states (3 iterations, 4 subsets). Nonetheless, it's important to note that our initial 2D simulations (not shown in this paper, see [25]) demonstrated that fully-unrolled FBSEM nets initialised by uniform images perform well irrespective of the initial estimate. Finally, the fifth acceleration strategy to reduce training time: v) validation datasets, that are often used to choose an optimal epoch at which the model has the minimum generalisation error, were not used in this study.

D. Evaluation

For each test dataset, six different methods were evaluated including conventional OSEM, MR-guided MAPEM, Unet-p, Unet-pm, FBSEM-p and FBSEM-pm. For simulations, the performance of these methods was evaluated based on i) contrast-to-noise ratio (CNR) between GM and WM tissues,

that is, the mean activity in GM minus mean activity in WM, divided by variance of activity in WM and ii) quantification errors of hot lesions and normalised root mean square error (NRMSE) across whole brain. For in-vivo datasets, the high-resolution MR images were parcellated into different cortical and subcortical regions using FreeSurfer software suite¹. The reconstructed low-dose PET images were mapped into the MR space for region-wise quantifications with respect to reference image in terms of mean (μ), standard deviation (SD) and their root-sum-square, $RSS = \sqrt{\mu^2 + SD^2}$.

III. RESULTS

Fig. 3 shows the reconstruction results of a test simulation dataset for different methods considered in this study, from standard OSEM and conventional MAPEM to new deep learning image denoising and reconstruction methods. Among the methods that only reply on PET data, the results show that

¹ <https://surfer.nmr.mgh.harvard.edu/>

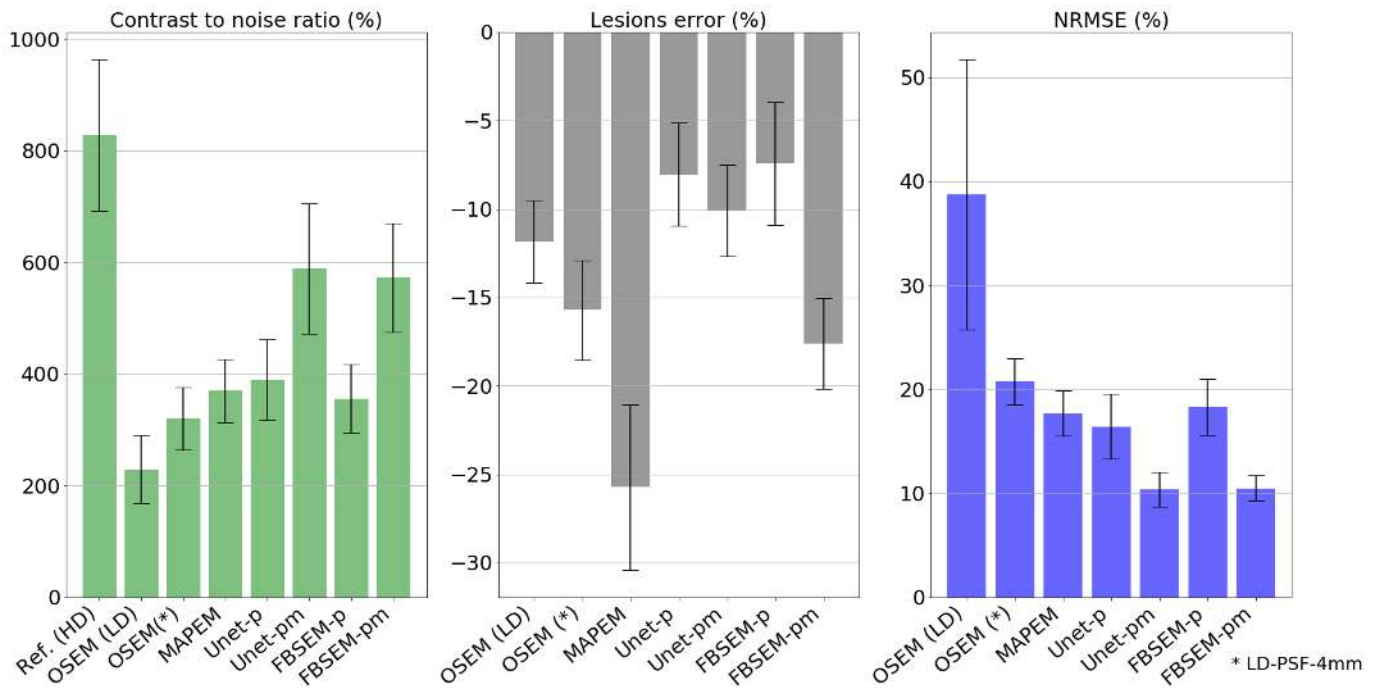


Fig. 4 Quantitative performance of different reconstruction methods averaged on test simulation datasets.

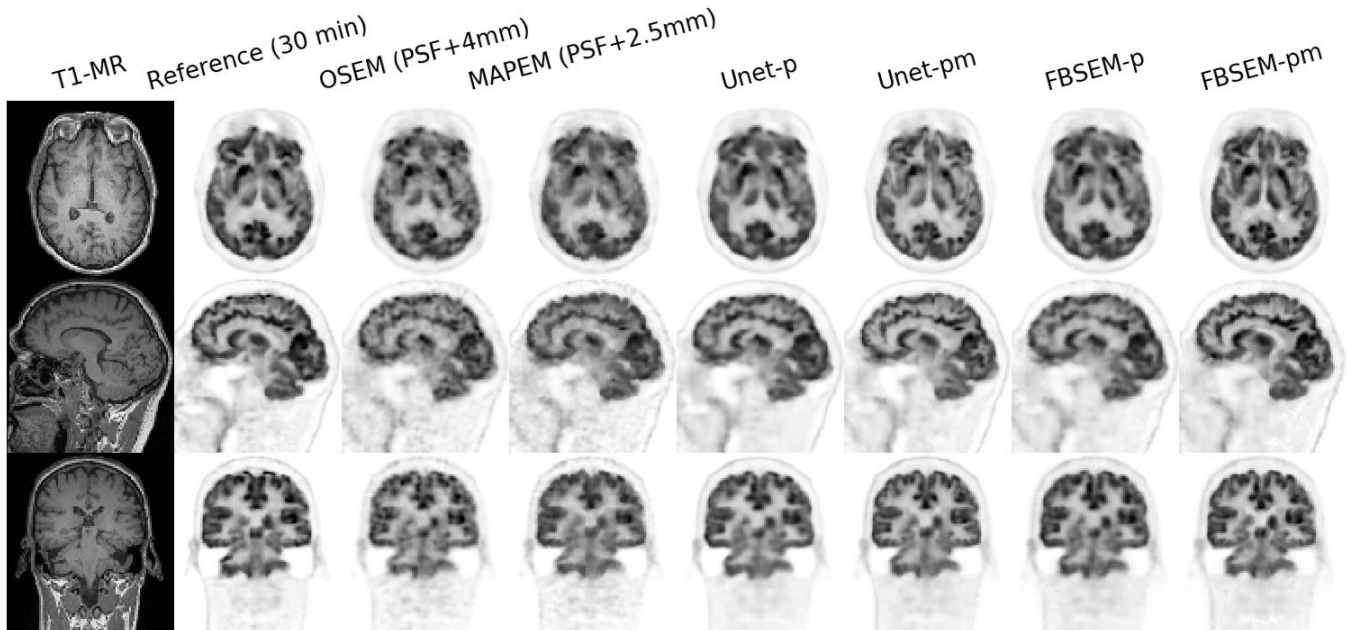


Fig. 5 Reconstruction results of different methods for a 2-min in-vivo dataset in comparison with their reference 30 min scan

both Unet-p and FBSEM-p improve upon the OSEM reconstruction by reducing the noise and contrast to a large extent. For those method that utilise the additional MR information, Unet-pm and FBSEM-pm both outperform the MAPEM algorithm, which suffers from lack of convergence and suppression of the PET lesions (see the arrow in the coronal view). These results show that Unet-p(m) and FBSEM-p(m) perform similarly on most of the anatomical regions except over the small lesions, as shown in this example test dataset. The performance of the reconstruction methods was the objectively evaluated for all simulation test datasets based on CNR between GM and WM, quantification errors in hot lesions and NRMSE

in the whole brain. As shown in Fig. 4, the Unet-pm and FBSEM-pm show the highest CNR as these methods reduced noise and at the same time improved the convergence. Unet-p showed slightly higher CNR than MAPEM and FBSEM-p, while as could be expected OSEM method achieved the lowest CNR. For hot lesions, the MAPEM and FBSEM-p resulted in the highest (-25.7%) and lowest (-7.4%) quantification errors with respect to reference images. The results show that Unet-pm notably outperformed FBSEM-pm net over the lesions by achieving errors of 10.0% versus 17.6%. The NRMSE performance of the methods show that Unet-pm and FBSEM-pm networks result in the lower overall errors.

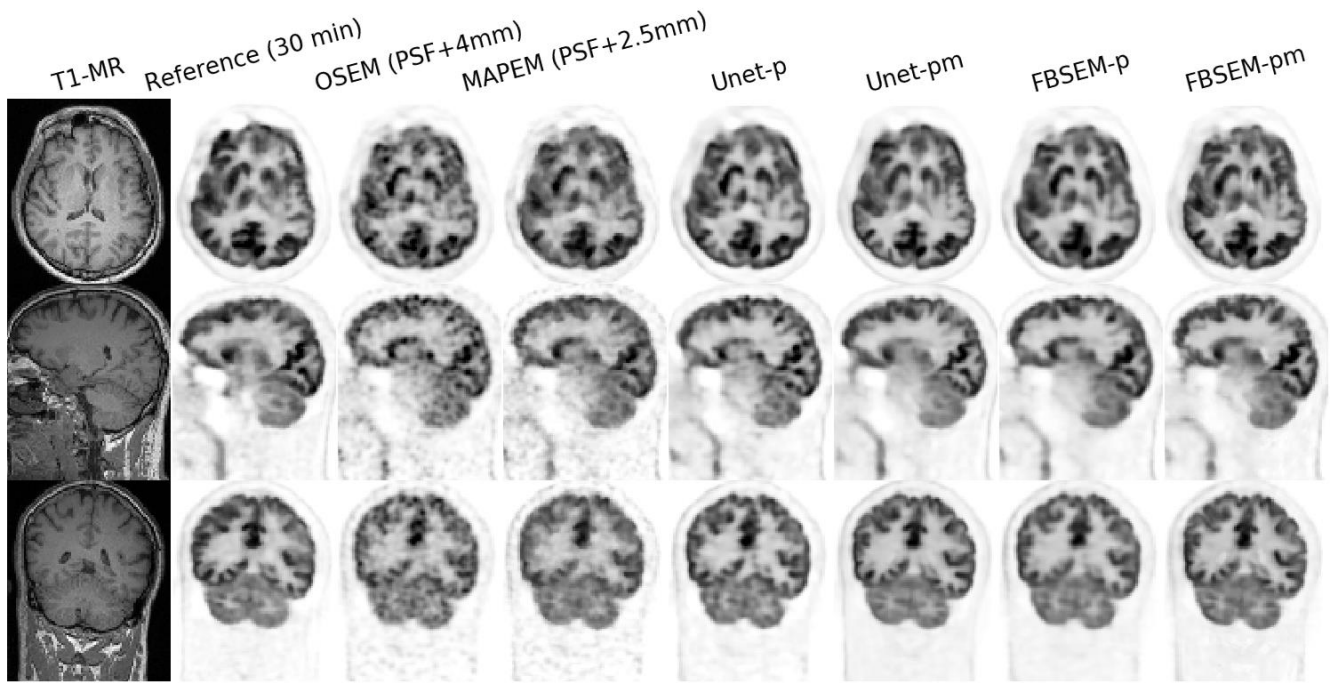


Fig. 6 Similar to Fig. 5, but for PET data of another subject.

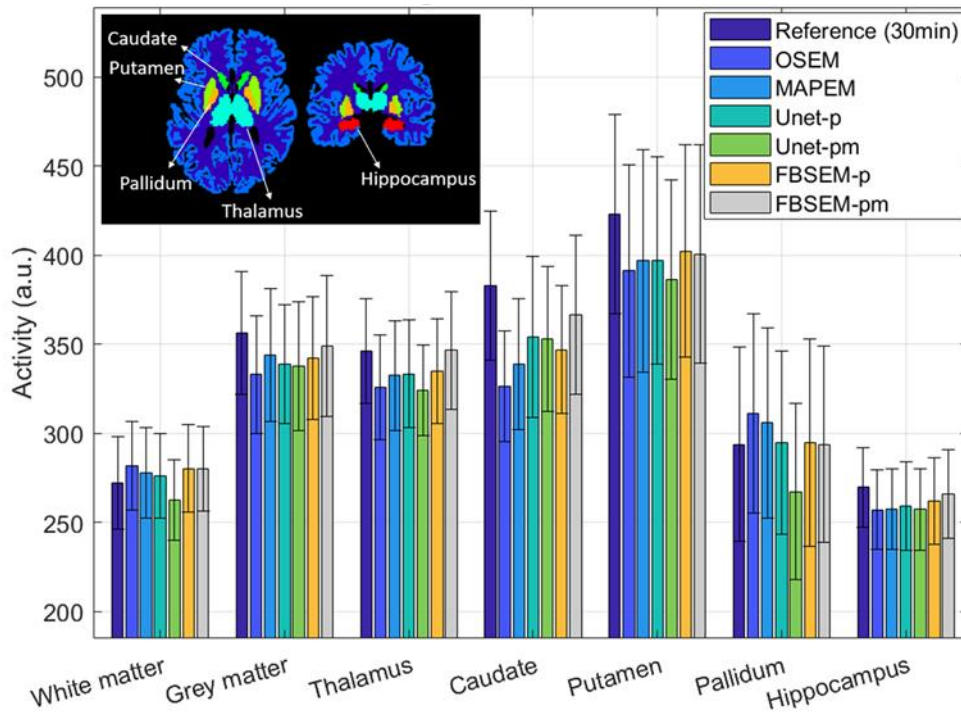


Fig. 7 Quantitative evaluation of different reconstruction methods in terms of mean activity in different regions of the brain averaged across 10 test in-vivo datasets.

For our simulations, training parameters and schedules chosen according to Table I, these results show that Unet-p outperforms FBSEM-p.

Based on these simulation results, in the training of FBSEM net for in-vivo datasets, as shown in Table I, we increased the number of kernels and reduced its depth (due to GPU memory limitations), which resulted in 1.5 times more trainable parameters compared to simulations. At the same time, we

pushed the U-net to its limit of performance, by increasing the number of its kernels (based on the capacity of our GPU memory), resulting in ~5 time more trainable parameters compared to simulations.

Figs. 5 and 6 compare the reconstruction results of the studied methods for two 2-min in-vivo scans in comparison with their reference 30-min scans (i.e. 15 times longer scan). As shown, the OSEM reconstruction notably suffer from noise,

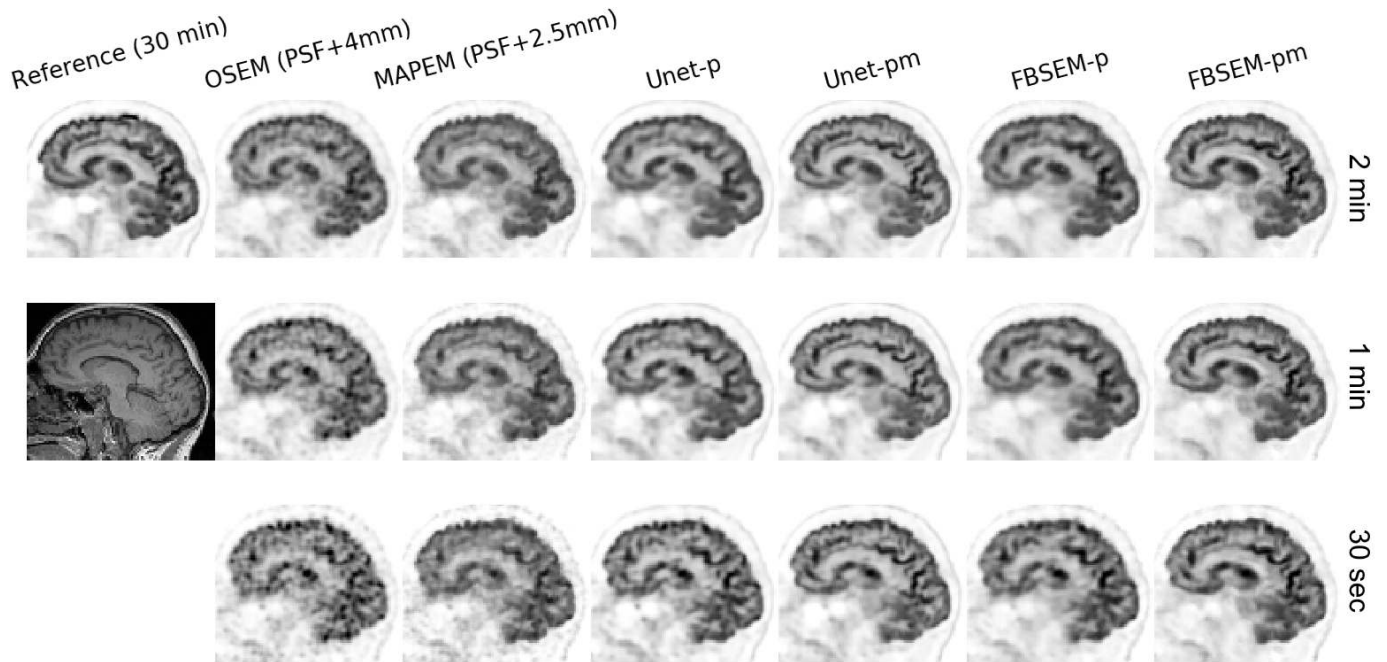


Fig. 8 Real-data performance of the studied methods for reconstruction of reduced scan times (2 min, 1 min and 30 sec.) of a subject with respect to their reference 30 min scan. Note the Unet-p(m) and FBSEM-p(m) networks are trained only with scan datasets of 2 min duration.

TABLE II. ERROR PERCENTAGE OF MEAN ACTIVITY IN DIFFERENT REGIONS OF IN-VIVO DATASETS TOGETHER WITH THE MEAN, SD AND RSS OF ALL REGIONAL ERRORS

	OSEM	MAPEM	UNET-P	UNET-PM	FBSEM-P	FBSEM-PM
WHITE MATTER	3.6	2.2	1.6	-3.4	3.1	2.9
GREY MATTER	-6.6	-3.5	-5.0	-5.3	-4.0	-2.1
THALAMUS	-5.9	-4.0	-3.7	-6.5	-3.3	0.0
CAUDATE	-14.8	-11.6	-7.6	-7.8	-9.4	-4.4
PUTAMEN	-7.5	-6.2	-6.2	-8.7	-4.9	-5.3
PALLIDUM	5.9	4.1	0.4	-9.1	0.3	0.0
HIPPOCAMPUS	-4.7	-4.6	-4.0	-4.6	-2.9	-1.4
MEAN	-4.3	-3.4	-3.5	-6.5	-3.0	-1.5
SD	6.5	4.8	3.1	2.0	3.6	2.6
RSS	7.8	5.9	4.7	6.8	4.7	3.0

while MAPEM shows lack of convergence, despite its regularisation parameter was chosen fairly low; even after 2.5 mm Gaussian filtering the images show some background noise. The results show that Unet-p and FBSEM-p networks achieve fairly comparable performance. Likewise, Unet-pm and FBSEM-pm networks performed similarly and produced images that are visually close to their reference images.

Fig. 7 compares the performance of these methods based on mean FDG uptake in WM, cortical GM and subcortical GM regions averaged over all in-vivo test datasets. Table II summarises the error percentage of mean activity in each anatomical region averaged over the test datasets, reporting the mean, SD and RSS of mean and SD for all regions. As seen in Fig. 7, all methods underestimated the mean activity in cortical and subcortical GM regions, except for the pallidum. For these datasets, FBSEM-p(m) nets achieve the closest mean activity to reference scans in most of GM regions. The results in Table II shows that Unet-p and FBSEM-p both achieve RSS errors of 4.7%, with slight difference in mean and SD errors, and

outperform the OSEM method. Among the methods using MR side information, FBSEM-pm shows the lowest RSS (3.0%) compared to Unet-pm (6.8%) and MAPEM (5.9%).

In this work, the DL methods were trained for mapping 2-min data to their reference 30-min data. In order to evaluate their generalisation and performance for shorter scan durations, we applied them to an in-vivo dataset with scan durations of 2 min, 1 min and 30 sec (i.e. 15 \times , 30 \times and 60 \times shorter than their reference scan, respectively). Fig. 8 compares the results for all methods. Note the regularisation parameters of the FBSEM-p(m) nets were not modified despite they have been trained for 2-min datasets. Likewise, the regularisation parameters of MAPEM for 1-min and 30-sec datasets was set to the one chosen for 2-min dataset of this subject. As seen, with shortening scan duration, noise notably dominates OSEM and MAPEM reconstructions. The Unet-p and FBSEM-p both show similar qualitative performance for 2- and 1-min datasets, however for 30-sec one, FBSEM-p tends to show less residual noise. For this subject, both Unet-pm and FBSEM-pm nets

demonstrate a consistent performance across all three scan durations, which shows their ability to generalise to datasets that they have never been trained for.

IV. DISCUSSION

In this study, we applied a proximal splitting technique for MAPEM reconstruction. For a specific regularisation parameter (β) and step size (γ), the resulting optimisation algorithm reduces to De Pierro’s MAPEM algorithm which is known to be monotonically convergent. However similar to Green’s one-step-late algorithm [26], for an arbitrarily large β this algorithm may not converge to a global maximum. A possible solution could be imposing a non-negativity constrain on Eq. (7). In fact, as shown in Fig.1, the residual learning unit used in our FBSEM net applies a ReLU activation function to the sum of the input image and the output of the CNN layers in order to explicitly ensure the non-negativity of the output. Moreover, our proposed FBSEM algorithm makes use of ordered subsets for acceleration which is known to cycle over a number of image estimates, especially for unbalanced subsets. A possible solution is to upgrade the OSEM update in Eq. (8) by a row-action maximum likelihood algorithm (RAMLA) [27], which is a convergent OS algorithm. Moreover, since the FBSEM algorithm is based on an optimisation transfer approach, convergence can be slow. In this study, we used a CNN-based prior for regularisation in the FBSEM net. Depending on whether the learned prior is convex or not, the trained FBSEM net can be convergent (if $N_{sub} = 1$) or non-convergent. The convexity of the learned priors which do not have an explicit functional form can potentially be tested using non-parametric techniques [28], which is behind the scope of this paper.

Our proposed reconstruction network has a number of advantages. Compared to the recently proposed BCD net [17] or EM net [29], which alternate between a MLEM reconstruction and a CNN-based image denoising module, model parameters in FBSEM net are shared across all reconstruction states (similar to RNNs), while BCD and EM nets employ separate networks for each reconstruction states. Sharing model parameters not only notably reduces the number of trainable parameters but also allows the trained network to be used with different number of iterations during inference [22]. Unlike BCD net and Gong *et al* [13], the regularization (penalty) parameter is learned from the data and the network can be initialised with a uniform image estimate. Unlike EM net and similar to BCD net, the data-fidelity based EM update and the CNN-based regularisation operations are performed in parallel in our network, and during training the backpropagation was set to pass only through the regularisation and fusion steps eliminating the need for computationally intensive differentiation of PET forward- and back-projections. In addition, our proposed network operates in PET-only and PET-MR modes.

Following the proposal and implementation of the FBSEM net, which can potentially improve upon prior networks owing to the above advantages, our next goal was to compare its performance with the best of post-reconstruction DL-based denoising. In this work, U-Net was chosen as a widely used encoder-decoder CNN. In [30], Lu *et al* recently showed that an

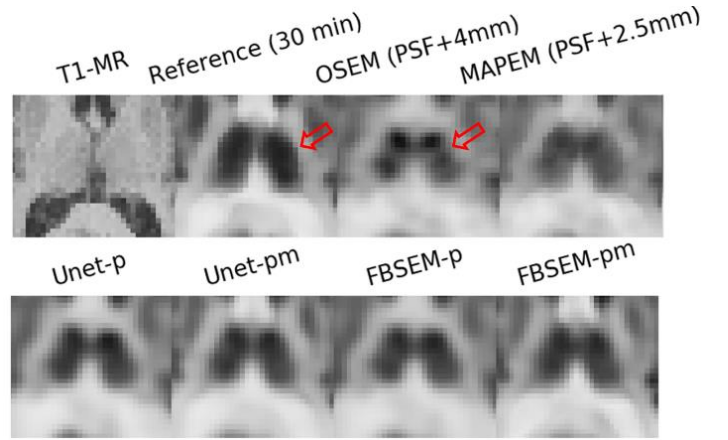


Fig. 9 Physiology mismatches in thalamus between a reference 30-min scan and a 2 min scan (obtained by the replay of 30 min list-mode data). Deep learning methods (second row) have increased the thalamus’ 2-min uptake toward its 30-min reference uptake.

optimised 3D U-Net could outperform a convolutional autoencoder network and a generative adversarial net for lung nodule quantification in reduced dose scans. Our results in Fig. 4 showed that Unet-p(m) net has a relatively comparable performance to FBSEM-p(m) net, even in its PET-MR mode, Unet-pm outperformed FBSEM-pm in preserving PET lesions in our simulations, despite their NRMSE over whole brain was comparable. This can be attributed to the fact that U-Net extracts and captures features at a multi-resolution level, and that the employed Unet-pm in our simulations has ~ 200 times more trainable parameters than the FBSEM-pm net. For in-vivo datasets, we increased the number of convolution kernels and decreased the number of reconstruction states and learning rate of FBSEM-p(m) net. At the same time, we increased the number of kernels for Unet-p(m) nets to potentially improve its performance even further, which resulted in ~ 600 times more trainable parameters compared to FBSEM nets. The in-vivo results showed that FBSEM-p and Unet-p perform comparatively for PET mode, however for PET-MR mode, FBSEM-pm outperforms Unet-pm on average.

The results in Fig. 4 have been averaged across 10 test datasets; since the reference HD images have fairly low noise (see Fig. 2 and 3), the variability of CNR for HD images represents the fact that our phantoms were generated from MR images of patients suspected of epilepsy and dementia, for which there may be cortical atrophy and partial volume effects of differing degrees. The results in Fig.4 show that the networks that used MR images (i.e. FBSEM-pm and Unet-pm) are able to capture that variability to some extent despite their mean CNRs being notably lower.

Similar to EM net, a residual U-Net could be used as the regularisation module in FBSEM net. However since U-Net usually employs a large number of trainable parameters the training of the resulting FBSEM network would be tremendously memory demanding. Note that the inference of FBSEM net or generally any network is notably less memory demanding, as automatic differentiation (autograd) will be inactive and hence tensors’ gradient will not be tracked and stored in memory. Our initial 2D simulation results presented in [25] showed that FBSEM net with the smaller residual learning unit (RLU) architecture achieves a comparable

performance to when a residual U-Net is used inside FBSEM net. Hence, in this study, we opted for the less memory demanding RLU network, on the understanding that the FBSEM net results would be representative also of the case of when a U-Net is used instead of a RLU. Furthermore, our previous 2D results showed that post-reconstruction denoising using a U-Net outperforms a residual learning unit. Therefore, given these initial results, and furthermore also those reported in literature (e.g. [30]), we chose a U-Net to best represent performance of post-reconstruction denoising networks, just as using a RLU in FBSEM net best represents its performance as well.

The number of parameters in our U-Net trained for in-vivo datasets is nearly 48 M parameters, which is in the range used in modern CNNs; from ~40 M in Inception-v4 to ~140 M in VGG net. However, for a fixed amount of training data, the large number of parameters can increase the chance of overfitting and generalisation error. In Fig. 8, we used the models trained on only 2-min data for testing on even shorter scans, to assess performance on a domain different to that of the training data. Given that the Unet-p model achieves a comparable performance to FBSEM-p for a 2-min test dataset and that these models have not been trained for 1-min and 30-sec scans, the slightly poorer performance of the Unet-p for 30-sec scan should not be interpreted as overfitting but better domain adaptation capabilities of FBSEM net. This can be attributed to the fact that noise is iteratively amplified during OSEM reconstruction, whereas FBSEM net can suppress the noise from early stages.

In this work, we considered DL image enhancement and reconstruction of reduced-duration and full-dose scans instead of reduced-dose and full-duration ones. Because we believe the immediate clinical test of DL methods will be for reduced duration studies as they can be done retrospectively and will have less complications for ethics approval compared to reduced dose studies which are prospective and require modification to clinical acquisition protocols. Hence, in this study, the 30-min list-mode FDG data were resampled to emulate 2-min scans and DL networks were trained to reduce noise in the 2-min PET images and improve their image quality towards their reference 30-min images. However, since a 15 \times scan-time reduction was considered in order to make noise vividly dominant, there is a chance for physiological mismatches between the two scans over brain regions with rapid/delayed glucose metabolism in some patients. For these cases, the DL networks will not only reduce noise but also predict what the physiology of those regions could be if the scan time had been prolonged for another 28 min. Fig. 9 illustrates this for the thalamus in a subject. As shown, there is a physiology mismatch between thalamus' uptake after 2- and 30-min scans. MAPEM reconstruction flattens uptake in the thalamus by reducing noise or PSF artefacts, but the DL methods not only flatten the uptake but also increase it towards its reference 30-min uptake. Another physiology mismatch can be seen in Fig. 6, sagittal view, where the tracer's uptake has been washed out in the caudate during the 30-min course of scan compared the 2-min scan. As seen in the transverse view, this patient has hypometabolism in their right hemisphere. It should be noted that for whole-body scans with 3-4 min acquisition per bed, scan-time reductions up to 4 will make

noise vividly dominant, hence the chance of physiology changes will be relatively lower.

In our FBSEM networks, the number of kernels of all D layers is the same. As summarized in Table I, for in-vivo datasets we chose a taller and shallower network (37 kernels and 3 layers) whereas for simulated datasets we chose a shorter and deeper one (16 kernels and 9 layers). The reason is that our simulations in Fig. 4 showed that the FBSEM-p(m) nets are not quantitatively as good as the Unet-p(m) networks. Given that these networks operate on 3D images, our intuition was that the 16 kernels in the first layer of an FBSEM net might not be sufficient to capture 3D edges, hence we opted for a taller network at the compromise of making the network shallower due to GPU memory limitations. Since our simulations were made as realistic as possible, the improved quantitative performance of the FBSEM nets for in-vivo datasets compared to simulations implies that our intuition might be correct. In general, for a specific deep learning task, a network's hyperparameters can be chosen based on the network's performance on validation datasets. As mentioned earlier, we could not afford the computational load of the validation process, nonetheless our results for unseen test datasets were acceptable. Had we included the validation; FBSEM net's performance could have been potentially even better.

The in-vivo brain PET and MR images are often well aligned, nonetheless there is chance of head drift between the different PET and MR acquisition time windows, therefore in this study we assured their alignment using SPM co-registration. In regard to misalignments, our simulation results with lesion mismatches between PET and MR showed that Unet-pm and FBSEM-pm nets outperformed the conventional Bowsher MAPEM algorithm in lesion quantification, which indicates the potential ability of DL methods in dealing with mismatches. This can be investigated in a future work. Future work would also include evaluation of the number of reconstruction states and depth of the FBSEM net on its performance and investigation of memory-efficient reconstruction algorithms and strategies to train a fully unrolled FBSEM net for 3D whole-body PET-MR image reconstruction.

V. CONCLUSION

A model-based deep learning reconstruction network was designed by unrolling an optimisation algorithm that we proposed in this study for MAPEM image reconstruction. The proposed FBSEM net was evaluated in PET-only and PET-MR modes in comparison with the state-of-the-art U-Net denoising and conventional MR-guided MAPEM and standard OSEM methods. Our simulation and in-vivo results showed that both DL based techniques outperform the conventional methods. It was found that for the chosen network parameters and training schedules the Unet-p and FBSEM-p net achieve a fairly comparable performance for both simulation and in-vivo datasets. For simulations, Unet-pm net showed lower quantification errors for PET unique lesions while achieving a similar NRMSE to FBSEM-pm net. Whereas for in-vivo datasets, the FBSEM-pm outperformed Unet-pm and achieved the lowest quantification error amongst all reconstruction methods. It can be concluded that DL-based post-reconstruction denoising methods can potentially perform as good as DL-based reconstruction methods.

ACKNOWLEDGMENT

This work is supported by the Engineering and Physical Sciences Research Council (EPSRC) under grant EP/M020142/1 and The Wellcome EPSRC Centre for Medical Engineering at King's College London (WT 203148/Z/16/Z) as well as the Department of Health via the National Institute for Health Research (NIHR) comprehensive Biomedical Research Centre award to Guy's & St Thomas' NHS Foundation Trust in partnership with King's College London and King's College Hospital NHS Foundation Trust. The authors would like to thank Prof. Alexander Hammers and Dr Colm McGinnity from King's College London & Guy's and St Thomas' PET Centre for providing the clinical brain PET-MR datasets. For reproducible research and compliance with EPSRC's policy framework on research data, PyTorch codes and simulation datasets for training of a 2D FBSEM net will be openly available at <https://github.com/Abolfazl-Mehranian/FBSEM/>

REFERENCES

- [1] J. Qi and R. M. Leahy, "Iterative reconstruction techniques in emission computed tomography," *Phys Med Biol*, vol. 51, no. 15, pp. R541-78, Aug 7 2006, doi: 10.1088/0031-9155/51/15/R01.
- [2] B. Bai, Q. Li, and R. M. Leahy, "Magnetic resonance-guided positron emission tomography image reconstruction," (in eng), *Semin Nucl Med*, vol. 43, no. 1, pp. 30-44, 2013, doi: 10.1053/j.semnuclmed.2012.08.006.
- [3] A. Mehranian *et al.*, "PET image reconstruction using multi-parametric anato-functional priors," *Phys Med Biol*, vol. 62, no. 15, pp. 5975-6007, Jul 6 2017, doi: 10.1088/1361-6560/aa7670.
- [4] J. Bland, M. A. Belzunce, S. Ellis, C. J. McGinnity, A. Hammers, and A. J. Reader, "Spatially-Compact MR-Guided Kernel EM for PET Image Reconstruction," *IEEE Trans Radiat Plasma Med Sci*, vol. 2, no. 5, pp. 470-482, Sep 2018, doi: 10.1109/TRPMS.2018.2844559.
- [5] J. Bland *et al.*, "MR-Guided Kernel EM Reconstruction for Reduced Dose PET Imaging," *IEEE Trans Radiat Plasma Med Sci*, vol. 2, no. 3, pp. 235-243, 2018, doi: 10.1109/TRPMS.2017.2771490.
- [6] G. Wang, J. C. Ye, K. Mueller, and J. A. Fessler, "Image Reconstruction is a New Frontier of Machine Learning," *IEEE Trans Med Imaging*, vol. 37, no. 6, pp. 1289-1296, Jun 2018, doi: 10.1109/TMI.2018.2833635.
- [7] K. Gong, E. Berg, S. R. Cherry, and J. Qi, "Machine Learning in PET: From Photon Detection to Quantitative Image Reconstruction," *Proceedings of the IEEE*, vol. 108, no. 1, pp. 51-68, 2020, doi: 10.1109/JPROC.2019.2936809.
- [8] B. Zhu, J. Z. Liu, S. F. Cauley, B. R. Rosen, and M. S. Rosen, "Image reconstruction by domain-transform manifold learning," *Nature*, vol. 555, no. 7697, pp. 487-492, Mar 21 2018, doi: 10.1038/nature25988.
- [9] I. Haggstrom, C. R. Schmidlein, G. Campanella, and T. J. Fuchs, "DeepPET: A deep encoder-decoder network for directly solving the PET image reconstruction inverse problem," *Med Image Anal*, vol. 54, pp. 253-262, May 2019, doi: 10.1016/j.media.2019.03.013.
- [10] K. Gong, J. H. Guan, C. C. Liu, and J. Y. Qi, "PET Image Denoising Using a Deep Neural Network Through Fine Tuning," (in English), *IEEE Trans Radiat Plasma Med Sci*, vol. 3, no. 2, pp. 153-161, Mar 2019, doi: 10.1109/Trpms.2018.2877644.
- [11] C. C. Liu and J. Qi, "Higher SNR PET image prediction using a deep learning model and MRI image," *Phys Med Biol*, vol. 64, no. 11, p. 115004, May 23 2019. [Online]. Available: <https://www.ncbi.nlm.nih.gov/pubmed/30844784>.
- [12] L. Cheng, S. Ahn, S. Ross, H. Qian, and B. De Man, "Accelerated Iterative Image Reconstruction Using a Deep Learning Based Leapfrogging Strategy," in *Fully 3D*, 2017.
- [13] K. Gong *et al.*, "Iterative PET Image Reconstruction Using Convolutional Neural Network Representation," (in English), *IEEE Trans Med Imaging*, vol. 38, no. 3, pp. 675-685, Mar 2019, doi: 10.1109/Tmi.2018.2869871.
- [14] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional Networks for Biomedical Image Segmentation," in *Medical Image Computing and Computer-Assisted Intervention MICCAI 2015*, Cham, N. Navab, J. Hornegger, W. M. Wells, and A. F. Frangi, Eds., 2015// 2015: Springer International Publishing, pp. 234-241.
- [15] K. Gong, C. Catana, J. Qi, and Q. Li, "PET Image Reconstruction Using Deep Image Prior," *IEEE Trans Med Imaging*, vol. 38, no. 7, pp. 1655-1665, 2019, doi: 10.1109/TMI.2018.2888491.
- [16] J. Cui *et al.*, "PET image denoising using unsupervised deep learning," *European Journal of Nuclear Medicine and Molecular Imaging*, journal article vol. 46, no. 13, pp. 2780-2789, December 01 2019, doi: 10.1007/s00259-019-04468-4.
- [17] H. Lim, Z. Huang, J. A. Fessler, Y. K. Dewaraja, and I. Y. Chun, "Application of trained Deep BCD-Net to iterative low-count PET image reconstruction," in *2018 IEEE Nuclear Science Symposium and Medical Imaging Conference Proceedings (NSSMIC)*, 10-17 Nov. 2018 2018, pp. 1-4, doi: 10.1109/NSSMIC.2018.8824563.
- [18] G. Wang and J. Qi, "Penalized likelihood PET image reconstruction using patch-based edge-preserving regularization," *IEEE Trans Med Imaging*, vol. 31, no. 12, pp. 2194-204, Dec 2012, doi: 10.1109/TMI.2012.2211378.
- [19] P. L. Combettes and J.-C. Pesquet, "Proximal Splitting Methods in Signal Processing," in *Fixed-Point Algorithms for Inverse Problems in Science and Engineering*, H. H. Bauschke, R. S. Burachik, P. L. Combettes, V. Elser, D. R. Luke, and H. Wolkowicz Eds. New York, NY: Springer New York, 2011, pp. 185-212.
- [20] A. R. De Pierro, "On the relation between the ISRA and the EM algorithm for positron emission tomography," *IEEE transactions on medical imaging*, vol. 12, no. 2, pp. 328-333, 1993, doi: 10.1109/42.232263.
- [21] E. Levitan and G. T. Herman, "A Maximum a Posteriori Probability Expectation Maximization Algorithm for Image Reconstruction in Emission Tomography," *Ieee T Med Imaging*, vol. 6, no. 3, pp. 185-192, 1987, doi: 10.1109/TMI.1987.4307826.
- [22] H. K. Aggarwal, M. P. Mani, and M. Jacob, "MoDL: Model-Based Deep Learning Architecture for Inverse Problems," *IEEE Trans Med Imaging*, vol. 38, no. 2, pp. 394-405, 2019, doi: 10.1109/TMI.2018.2865356.
- [23] K. He, X. Zhang, S. Ren, and J. Sun, "Deep Residual Learning for Image Recognition," in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 27-30 June 2016 2016, pp. 770-778, doi: 10.1109/CVPR.2016.90.
- [24] J. E. Bowsher *et al.*, "Utilizing MRI information to estimate F18-FDG distributions in rat flank tumors," in *IEEE Symposium Conference Record Nuclear Science 2004.*, 16-22 Oct. 2004 2004, vol. 4, pp. 2488-2492 Vol. 4, doi: 10.1109/NSSMIC.2004.1462760.
- [25] A. Mehranian and A. J. Reader, "Model-Based Deep Learning PET Image Reconstruction Using Forward-Backward Splitting Expectation Maximisation," presented at the IEEE NSS/MIC, in press, 2020.
- [26] P. J. Green, "On Use of the Em Algorithm for Penalized Likelihood Estimation," *Journal of the Royal Statistical Society: Series B (Methodological)*, vol. 52, no. 3, pp. 443-452, 1990, doi: 10.1111/j.2517-6161.1990.tb01798.x.

- [27] J. Browne and A. B. d. Pierro, "A row-action alternative to the EM algorithm for maximizing likelihood in emission tomography," *IEEE Trans Med Imaging*, vol. 15, no. 5, pp. 687-699, 1996, doi: 10.1109/42.538946.
- [28] A. Juditsky and A. Nemirovski, "On nonparametric tests of positivity/monotonicity/convexity," *Annals of Statistics*, vol. 30, 04/01 2002, doi: 10.1214/aos/1021379863.
- [29] K. Gong *et al.*, *EMnet: an unrolled deep neural network for PET image reconstruction* (SPIE Medical Imaging). SPIE, 2019.
- [30] W. Lu *et al.*, "An investigation of quantitative accuracy for deep learning based denoising in oncological PET," *Phys Med Biol*, vol. 64, no. 16, p. 165019, Aug 21 2019, doi: 10.1088/1361-6560/ab3242.