DOI:
10.1109/TRPMS.2023.3283786

Document Version
Peer reviewed version

Single-Modality Supervised Joint PET-MR Image Reconstruction

Guillaume Corda-D'Incan, Julia A. Schnabel, *Fellow, IEEE*, Alexander Hammers, and Andrew J. Reader

***Abstract*—We present a new approach for deep learned joint PET-MR image reconstruction inspired by conventional synergistic methods using a joint regularizer. The maximum *a posteriori* expectation-maximization algorithm for PET and the Landweber algorithm for MR are unrolled and interconnected through a deep learned joint regularization step. The parameters of the joint U-Net regularizer and the respective regularization strengths are learned and shared across all the iterations. Along with introducing this framework, we propose an investigation of the impact of the loss function selection on network performance. We explored how the network performs when trained with a single or a joint-modality loss. Finally, we explored under which settings a joint reconstruction was beneficial for MR reconstruction by using various undersampling factors. The results obtained on 2D simulated data show that the joint networks outperform conventional synergistic methods and independent deep learned reconstruction methods. For PET, the network trained with only a PET loss achieves a better global reconstruction accuracy than the version trained with a weighted sum of PET and MR loss terms. More importantly, the former further improves the reconstruction of PET-specific features where MR-guided methods show their limit. Therefore, using a single-modality loss to supervise the training while still reconstructing the two modalities in parallel leads to better reconstructions and improved modality-unique lesion recovery in our proposed framework. For MR, while the same effect is observed, joint reconstruction gains only occur in the presence of highly undersampled data. Single-modality loss joint reconstruction results are also demonstrated on 3D clinical PET-MR datasets.**

I. Introduction

POSITRON emission tomography (PET) is a nuclear medicine imaging technique that permits the tracking of *in vivo* biological processes. It is widely used for oncology and neurodegenerative disorder detection and staging. The image reconstruction problem derived from PET measured data is ill-posed. Due to limited detection efficiencies and other physical limitations, sinograms exhibit high noise levels making the inverse problem unstable. The injected radiotracer dose is limited to prevent potentially dangerous side effects on patients and technicians. Furthermore, scanning time is reduced to improve patient comfort. These measures further reduce the number of true annihilation photons collected by the scanner. Iterative methods such as the maximum likelihood expectation-maximization (MLEM) [1] or the ordered subsets EM (OSEM) algorithm [2] are used to reconstruct PET images in routine scans. In the presence of low-count data, these algorithms fit the noise contained in the sinograms and can generate images of poor quality, which could complicate physician diagnoses. Regularization can help to limit the impact of noisy data on the reconstructed images. Early termination of the reconstruction algorithm, post-smoothing of noisy reconstructed images with a Gaussian kernel, or the introduction of prior information into the reconstruction algorithm are clinically used regularization techniques [31]. The latter offers the highest quality for low-count (LC) data reconstruction. The main drawback arising from this strategy is the absence of knowledge of the optimal prior for a given task. Numerous priors have been proposed based on common beliefs and using known functionals with desirable properties [12].

Magnetic resonance imaging (MRI) is a non-invasive imaging technique allowing imaging of tissue morphology with high resolution and various contrasts. It is used to detect a vast range of conditions, from epilepsy to cancer. MRI data acquisition is inherently slow due to anatomical and hardware limitations. Accelerated MRI can be performed by collecting fewer k-space samples, which can introduce aliasing artifacts in the reconstructed MR image. Parallel imaging (PI) permits further reduction of the scanning time where multiple receiver coils collect data on various sub-parts of the object in the scanner. This process allows a reduction in the number of phase-encoding steps performed and, therefore, a reduction in the scanning time. An alternative for accelerating MR data acquisition is compressed sensing (CS). CS permits the collection of fewer measurements than the Nyquist theorem requires by using incoherent undersampling and assuming the signal to be sparse in some domain. Non-linear reconstruction algorithms are needed for CS reconstruction. Similarly to iterative PET reconstruction, regularization methods for PI and CS help achieve higher image quality by removing artifacts and reducing noise but face similar challenges.

Since the introduction of commercial hybrid PET-MR scan-

This work was supported by the EPSRC Centre for Doctoral Training in Smart Medical Imaging [EP/S032789/1] and Siemens Healthineers, Erlangen, Germany. This work was supported by the Wellcome/EPSRC Centre for Medical Engineering [WT 203148/Z/16/Z]. The National Institute for Health Research (NIHR) Biomedical Research Centre at Guy's, St Thomas' NHS Foundation Trust, and King's College London supported this work. The views expressed are those of the authors and not necessarily those of the NHS, the NIHR, or the Department of Health. For the purpose of open access, the author has applied a CC BY public copyright license to any Author Accepted Manuscript version arising from this submission. This work involved human subjects in its research. Approval of all ethical and experimental procedures and protocols was granted by the institution's Research and Development department, the North East – York research ethics committee (reference 15/NE/0203), and the radiation protection committee (ARSAC).

Guillaume Corda-D'Incan and Andrew J. Reader are with the School of Biomedical Engineering and Imaging Sciences, King's College London, UK (e-mail: guillaume.corda@kcl.ac.uk).

Julia A. Schnabel is with the Technical University of Munich, Germany, the Helmholtz Center Munich, Germany, and with the School of Biomedical Engineering and Imaging Sciences, King's College London, UK

Alexander Hammers is with the King's College London & Guy's and St Thomas' PET Centre, and the School of Biomedical Engineering and Imaging Sciences, King's College London, UK.

978-1-6654-8872-3/22/$31.00 ©2023 IEEE

ners a decade ago, the simultaneous acquisition of PET and MR data has mainly been exploited for the generation of attenuation maps [32]. The PET and MR data are independently reconstructed, discarding potentially diagnostically valuable joint and complementary information. The main reason for this omission lies in the complexity of designing joint priors capable of exploiting common information while preserving modality-unique features. Some methods proposed exploiting structural similarity using second-order total generalized variation (TGV) [7] or parallel level sets (PLS) [8]. They have shown better recovery capabilities for modality-specific structures than MR-guided PET reconstruction methods, and have demonstrated that MR benefits from PET guidance under specific settings. However, notable sensitivity to relative signal intensities and contrast of PET-MR images was observed. Mehranian *et al.* [3] proposed a joint prior defined as the sum of weighted multi-modal quadratic priors, providing state-of-the-art results for synergistic PET-MR reconstruction.

Deep learning (DL) recently opened up a new medical image reconstruction era. It allows learning of the form of the priors - or their gradients - embedded in conventional reconstruction algorithms. Whether for PET or MR, deep learned model-based image reconstruction (MBIR) emerged as a promising methodology outperforming conventional methods. The first deep learned joint method, Syn-Net [6], was built based on this idea. Two conventional reconstruction algorithms (one for each modality) were unrolled and interconnected whilst the regularizers and their respective strengths were learned.

We propose a new framework for deep learned joint PET-MR reconstruction through this work. Similarly to Syn-Net, the MAPEM algorithm [26] for PET, and the Landweber algorithm [13] for MR, are unrolled and interconnected through the regularization step. A joint prior is represented as a U-Net [27] whose parameters are shared across all the iterations. The network is afterward trained using a single-modality loss to supervise the joint reconstruction. The limited number of trainable parameters and the simpler architecture offer advantages over previously proposed deep learned joint methods such as Syn-Net.

This article is structured as follows. Section II covers the theory behind PET and MR conventional reconstruction methods as well as joint methods and explores how DL can be used for medical imaging. Section III details the proposed method, the data used, and the various reference methods implemented to compare with our method. Section IV presents the results obtained and the comparison of the various PET and MR reconstruction methods in 2D and 3D. Section V concludes and discusses potential limitations as well as eventual improvements.

II. Background

A. Model-based image reconstruction

Model-based image reconstruction (MBIR) algorithms iteratively reconstruct an unknown image $\mathbf{x} \in \mathbb{R}^N$ from measured data $\mathbf{y} \in \mathbb{R}^M$. The variational problem considered is formulated as follows:

$$\hat{\mathbf{x}} = \underset{\mathbf{x}}{\operatorname{argmin}}\, f(\mathbf{x};\mathbf{y}) + \beta\mathcal{R}(\mathbf{x}) \tag{1}$$

where f is the data fidelity term ensuring consistency between the estimate and the measured data, $\mathcal{R}$ is a penalty term measuring the discrepancy of $\mathbf{x}$ with respect to the set of realistic images based on prior knowledge, and β is the regularization strength.

B. PET image reconstruction

For PET imaging, the data consistency term f used is the negative Poisson log-likelihood defined by:

$$f(\mathbf{x};\mathbf{y}) = -\sum_i y_i \log(\overline{y_i}) + \overline{y_i} + \log(y_i!) \tag{2}$$

with $\overline{\mathbf{y}} = \mathbf{NLPx}+\mathbf{r}+\mathbf{s}$, where $\mathbf{NLP} \in \mathbb{R}^{M\times N}$ is the product of the normalization, attenuation, and x-ray transform matrices, and $\mathbf{r}+\mathbf{s} \in \mathbb{R}^M$ is the expected number of randoms and scatters. Conventionally, regularizers with specific properties, such as convexity, are selected for simplicity and to provide convergence guarantees. When convex and differentiable priors are used, Equation (1) can be solved using the forward-backward splitting (FBS) algorithm [25]:

$$\mathbf{x}^{Reg} = \mathbf{x}_k - \gamma\beta\nabla\mathcal{R}(\mathbf{x}_k) \tag{3a}$$

$$\mathbf{x}_{k+1} = \underset{\mathbf{x}}{\operatorname{argmin}}\, f(\mathbf{x}|\mathbf{y}) - \frac{1}{2\gamma}\left\|\mathbf{x}-\mathbf{x}^{Reg}\right\|^2 \tag{3b}$$

By using a surrogate of the negative log-likelihood [14], the proximal mapping in Equation (3b) can be reformulated, and a closed-form solution is obtained [4]. The maximum *a posteriori* (MAP) estimate is then computed using a three-step algorithm. At each iteration, data consistency (4a), regularization (4b), and fusion steps (4c) are performed. The final forward-backward splitting EM algorithm (FBSEM) reduces to:

$$\mathbf{x}^{Reg} = \mathbf{x}_k - \gamma\beta\nabla\mathcal{R}(\mathbf{x}_k) \tag{4a}$$

$$\mathbf{x}^{EM} = \frac{\mathbf{x}_k}{(\mathbf{NLP})^T\mathbf{1}}(\mathbf{NLP})^T\frac{\mathbf{y}}{\mathbf{NLP}\mathbf{x}_k+\mathbf{r}+\mathbf{s}} \tag{4b}$$

$$\mathbf{x}_{k+1} = \frac{2\mathbf{x}^{EM}}{1-\boldsymbol{\nu}\mathbf{x}^{Reg}+\sqrt{(1-\boldsymbol{\nu}\mathbf{x}^{Reg})^2+4\boldsymbol{\nu}\mathbf{x}^{EM}}} \tag{4c}$$

with $\boldsymbol{\nu} = \dfrac{\mathbf{1}}{\gamma(\mathbf{NLP})^T\mathbf{1}}$

As demonstrated in Bowsher's method [12], anatomical information can be incorporated to improve image quality. Nonetheless, the use of MR images may lead to incorrect PET reconstructions. Despite numerous methods proposed, both traditional and deep learning-based, MR-guided regularization often imposes too much of the anatomical structure. PET-specific lesions may be smoothed out or eliminated, while MR-specific lesions may become visible in the reconstructed PET image, impeding such methods' clinical applicability. Overall, whether anatomical information is used or not, conventional priors are hand-crafted and, therefore, likely not optimal for multiple tasks.

C. MR image reconstruction

For MR reconstruction, a least-squares problem can be formulated, and f is defined as follows:

$$f(\mathbf{u};\mathbf{b}) = \|\mathbf{b} - \mathbf{UFCu}\|_2^2 \tag{5}$$

with $\mathbf{u}$ the MR image, $\mathbf{b}$ the k-space data, $\mathbf{U}$ the undersampling process, $\mathbf{F}$ the discrete Fourier transform and $\mathbf{C}$ the coil sensitivity maps (CSMs). Conventional methods use the Landweber algorithm [8] or iterative SENSE [9] with the conjugate gradient (CG) algorithm (CG-SENSE), amongst others. Similarly to PET reconstruction, the regularized Landweber algorithm can be regarded as three steps:

$$\mathbf{u}^{DC} = (\mathbf{UFC})^*(\mathbf{UFCu}_k - \mathbf{b}) \tag{6a}$$

$$\mathbf{u}^{Reg} = \nabla\mathcal{R}(\mathbf{u}_k) \tag{6b}$$

$$\mathbf{u}_{k+1} = \mathbf{u}_k - \alpha(\mathbf{u}^{Reg} + \lambda\mathbf{u}^{DC}) \tag{6c}$$

where $(\mathbf{UFC})^*$ is the conjugate transpose of the matrix $\mathbf{UFC}$, α and λ are the step size and regularization strength, and DC stands for data consistency. A vast diversity of priors have also been proposed in the literature for MR reconstruction, such as total variation (TV) regularization as well as dictionary learning [17], [18].

D. Conventional joint PET-MR image reconstruction

In joint reconstruction, an optimization problem composed of one data consistency term for each modality and a joint regularization term is considered. The objective function can be expressed as follows:

$$\hat{\mathbf{x}}, \hat{\mathbf{u}} = \underset{\mathbf{x},\mathbf{u}}{\operatorname{argmin}}\, f^{PET}(\mathbf{x};\mathbf{y}) + f^{MR}(\mathbf{u};\mathbf{b}) + \mathcal{R}(\mathbf{x},\mathbf{u};\beta,\lambda) \tag{7}$$

where β and λ are the regularization strengths for PET and MR, respectively. The complexity of joint regularized reconstruction arises from the choice and design of the prior. PET and MR exhibit distinct contrasts and fundamentally different noise distributions. Nonetheless, they are coupled by the underlying anatomy. The joint regularizer has a dual task: exploiting common structures and boundaries to guide one modality with the other while preserving unique features from each modality. Additionally, the prior must not be sensitive to different pixel intensities between PET and MR.

Knoll *et al.* [7] proposed a joint reconstruction framework based on the multi-modal second-order TGV. Since the TV introduces staircase artifacts, the TGV is preferred. Ehrhardt *et al.* [8] proposed a comparison of a joint TV with a Frobenius norm and a joint PLS prior. They demonstrated that the joint PLS prior is superior to the joint TV and that both modalities benefit from joint reconstruction by exploiting common structures. Mehranian *et al.* [3] extended the widely used quadratic prior to synergistic PET and multi-contrast MRI reconstruction. The regularizer is defined as a sum of mutually weighted quadratic priors updated every iteration. The MAPEM algorithm is used for PET, whereas the MR is reconstructed using CG-SENSE. The mutual weights are calculated after every MAPEM and CG-SENSE iteration using the current image estimates. To cope with various resolutions and intensity values, a different set of weights is computed for each modality by first mapping the images used for guidance into the space of the guided image. The weights are obtained by taking the product of Gaussian similarity kernels calculated between voxel j and b in a given neighborhood $\mathcal{N}_j$ for each image. The results reported in [3] set the reference for synergistic PET-MR reconstruction. However, the main drawback of this method is the high number of hyperparameters to fine-tune. The number of sub-iterations for PET and MR (same for T1 and T2), the number of global iterations, the regularization strengths, and the standard deviation of the Gaussian kernels lead to a total of nine hyperparameters. The increase in the number of additional parameters is also observed in independent reconstruction as reported in [22] for CS-MRI. Nonetheless, combined with the long duration of synergistic PET-MR image reconstruction, it becomes extremely time-consuming and complex to optimize this method rigorously. Moreover, based on our implementation, we noticed that a set of hyperparameters leading to the best reconstruction results for one modality does not necessarily give the best results for the other modality. This observation is at the root of the loss function investigation presented in section III.

E. Deep learning in medical image reconstruction

Deep neural networks (NNs), particularly convolutional NNs (CNNs), allow learning an ensemble of kernels and non-linear activation functions to approximate almost any mapping [19]. They show tremendous efficiency in image processing applications. In physics-informed DL, NNs are sparingly used to learn only particular parts of the reconstruction process. Unrolled methods mix conventional model-based Bayesian algorithms with deep neural networks. The gradients of the priors, the priors themselves, or the regularization strength can be learned. Combining an image obtained from a data consistency update and a regularization update at each iteration limits unpredictable NN behaviors compared to direct methods, without fully controlling them, as demonstrated in [23]. Additionally, this methodology achieves faster convergence and better performance with less training data, a crucial asset in medical imaging where datasets are scarce. Various conventional iterative algorithms have been unrolled for PET [4], [16], and MR [5], [10], [15]. For more information, the reader is referred to extensive reviews of DL methods for medical image reconstruction [28], [29]. Only two methods have applied this idea for deep learned joint PET-MR reconstruction to the best of our knowledge. Syn-Net [6], which was later updated in Dense Syn-Net [30], unrolls the FBS and Landweber algorithms as in FBSEM-Net [4] and the variational network [5], respectively. The two regularization steps corresponding to Equations (4a) and (6b) are replaced by two neural networks. PET and MR have separate regularizations, but both are inter-modality guided. In this framework, the learned regularization strengths, as well as the regularizers, are iteration-specific. Dense Syn-Net increases the complexity of the network by using all the previous PET and MR estimates to guide the regularization in a given module. Additionally, it monitors the training of each module by using iteration-dependent

targets. However, these two frameworks have their limitations. First, using iteration-dependent regularizers and regularization strengths widens the gap with theoretical conventional joint reconstruction and notably increases the number of trainable parameters. Second, a different number of iterations is used for PET and MR leading to an asymmetry. Finally, the overall complexity of the architecture complicates the tracking of the origins of the performance increase observed between the two methods. This effect might be attributed to the introduction of more trainable parameters in Dense Syn-Net due to the regularizers' additional input channels.

III. Proposed deep learned joint PET-MR image reconstruction

A. Unrolled network for PET-MR image reconstruction

Based on the observations in sections II.D and II.E, a new framework for deep learned joint reconstruction is proposed. Starting from a similar framework to earlier architectures such as Syn-Net and Dense Syn-Net, the MAPEM and Landweber algorithms are unrolled (Fig. 1). Inspired by the successful approach of conventional joint reconstruction by Mehranian *et al.*, a joint regularizer is used here as opposed to the two aforementioned frameworks. In [3], CG-SENSE is used for MR reconstruction. However, its convergence speed is much higher than the MAPEM algorithm for PET reconstruction. Therefore, the Landweber algorithm is selected due to its comparable convergence speed with MAPEM. To avoid the need to design a joint prior, a NN with a U-Net structure is used instead (Fig. 2). The architecture selected is chosen for its limited number of trainable parameters, 58k in 2D and 160k in 3D, compared to architectures used for direct reconstruction from measurement space. The AUTOMAP [35], and DeepPET [36] frameworks, for instance, use architectures with a number of trainable parameters ranging from 60 to 800 million. Avoiding over-parameterization of the joint regularizer increases the network's generalization ability with datasets of limited size, which is likely to happen for clinical PET-MR datasets. The NN noted $\mathcal{F}$, has three input and output channels, one for PET and two for MR, corresponding to the real and imaginary parts of the image. Its weights are randomly initialized and trained end-to-end, avoiding the need for pre-training. A 60-iteration reconstruction is performed, and the outputs are compared to reference images. The gradients are then backpropagated through the 60 unrolled modules to update the trainable parameters of $\mathcal{F}$ and the regularization strengths. The same network parameters and regularization strengths are shared across all the modules to stay close to the design of conventional joint reconstruction methods and limit the number of trainable parameters. Finally, unlike Syn-Net and Dense Syn-Net, PET and MR are reconstructed with the same number of iterations. The PET reconstruction is accelerated thanks to the use of subsets ($N_{iter} = 10$, $N_{subsets} = 6$), and PSF modeling is included in the system model (FWHM$= 4$ mm). The FWHM was chosen to match the reconstructions of the Siemens e7 tools [33]. Algorithm 1 illustrates the proposed joint unrolled network. The 6 in the index of the MR estimate $\mathbf{u}_{6k+s}$ of Algorithm 1 are due to the use of 6 subsets in the OSEM algorithm. A pass through a single subset for PET reconstruction corresponds to one iteration for MR reconstruction.

Algorithm 1: Deep learned joint PET-MR image reconstruction

Initialize: $\mathbf{x_0} = \mathbf{1}, \mathbf{u_0} =$ zero filled image, $\gamma \in [0, 1]$, $N_{iters} = 10$, $N_{subsets} = 6$ **for** *k = 0 ...* $N_{iters} - 1$ **do**
  for *s = 0 ...* $N_{subsets} - 1$ **do**

$$\mathbf{x}^{EM} = \frac{\mathbf{x}_{k,s}}{(\mathbf{NLP})^T\mathbf{1}}(\mathbf{NLP})^T\frac{\mathbf{y}}{\mathbf{NLP}\mathbf{x}_{k,s} + \mathbf{r} + \mathbf{s}}$$

$$\mathbf{u}^{DC} = (\mathbf{UFC})^*(\mathbf{UFC}\mathbf{u}_{6k+s} - \mathbf{b})$$

$$\mathbf{x}^{Reg}, \mathbf{u}^{Reg} = \mathcal{F}(\mathbf{x}_{k,s}, \mathbf{u}_{6k+s})$$

$$\mathbf{x}_{k+1,s} = \frac{2\mathbf{x}^{EM}}{\mathbf{1} - \boldsymbol{\nu}\mathbf{x}^{Reg} + \sqrt{(\mathbf{1} - \boldsymbol{\nu}\mathbf{x}^{Reg})^2 + 4\boldsymbol{\nu}\mathbf{x}^{EM}}}$$

with $\boldsymbol{\nu} = \frac{\mathbf{1}}{\gamma(\mathbf{NLP})^T\mathbf{1}}$

$$\mathbf{u}_{6k+s+1} = \mathbf{u}_{6k+s} - \alpha(\mathbf{u}^{Reg} + \lambda\mathbf{u}^{DC})$$

  end
end

B. Single or multi-modality loss for supervised training

Typically, multi-output networks are monitored by multiple loss functions, as shown in Figure 1. Nonetheless, the problem raised at the end of section II.D led us to perform a systematic evaluation of the joint network performance with respect to various losses. The performance when trained with a weighted sum of PET and MR mean square error (MSE) terms (Joint_PM) is compared with the same network trained with only one MSE term for PET (Joint_P) or only one MSE term for MR (Joint_M). When a two-term loss is used, the PET term is rescaled for each mini-batch at epoch ep by a factor defined by $\frac{\text{MSE}^{MR}_{ep-1}}{\text{MSE}^{PET}_{ep-1}}$, such that the relative contributions of the PET and MR losses to the overall loss are similar, and both parts of the network are fairly trained. Based on our experiment, the performance is negatively impacted when the two terms are not rescaled. In addition to different losses, various inputs are used to train the network to explore how PET helps guide the MR regularization under specific settings. Low-count (LC, 500k) and high-count (HC, 100M) data are used for the PET input, whereas for MR, various acceleration factors (R) are used from fully-sampled ($R = 1$ and noise-free noted HQ M) to extremely undersampled ($R = 8$ and noisy noted XUS M) k-space data. The PET and the MR reconstruction parts of the joint network are also independently trained to demonstrate the benefits of joint reconstruction. In both cases, the independent reconstruction performance with and without the guidance of a fully converged image from the other modality is explored. The images used for guidance are HC OSEM images for PET-guided MR reconstruction and

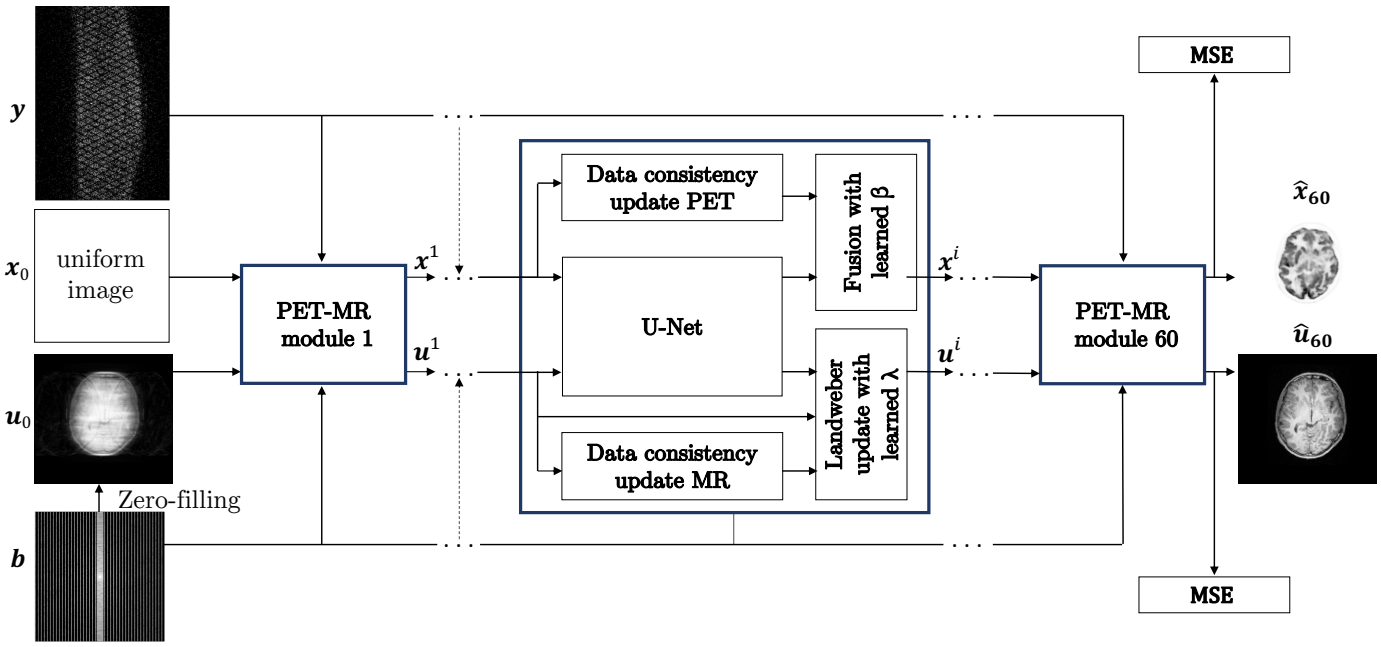

Fig. 1: Joint PET-MR image reconstruction framework proposed with 60 unrolled iterations. The two parts of the network are connected through the deep learned joint regularization represented by a deep neural network with a U-Net structure. DC=Data consistency.

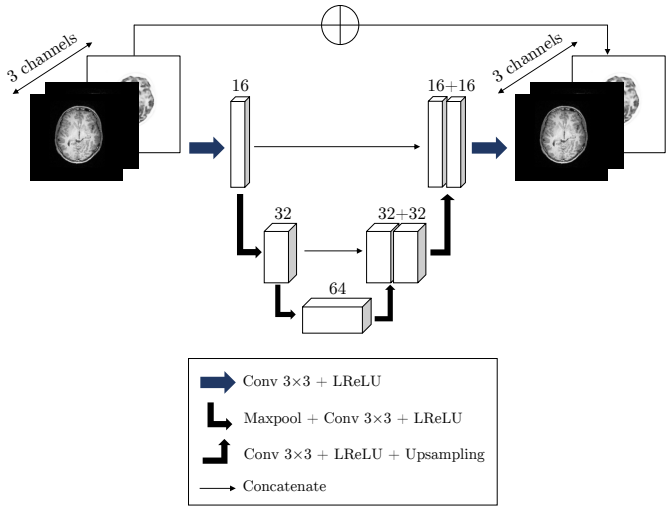

Fig. 2: U-Net architecture used as joint prior. MR images have two channels for real and imaginary parts. The number of parameters is 58k in 2D and 160k in 3D.

fully-sampled noise-free Landweber reconstructions for MR-guided PET reconstruction. In the case of pure independent reconstructions, uniform input channels with a value of 1 are used to ensure that the number of trainable parameters remains the same as for the joint reconstruction. Therefore, all the networks trained have 58k parameters in 2D. The naming of each network is derived as follows: [Recon type]_[Loss]_[PET input]_[MR input]. The recon_type can be joint or independent (noted indep), and the loss can be PET only (noted P), MR only (noted M), or both (noted PM). The PET input is LC or HC data, while the MR input is characterized by the acceleration factor R. Fully sampled data is denoted 1, mildly undersampled data is denoted 4, and highly undersampled is denoted 8. Thus, Joint_P_LC_4 corresponds to a joint reconstruction supervised by only a PET loss term using LC data for the PET input and undersampled k-space data with an acceleration factor of 4 for the MR input data. Indep_M_8_guidedHC corresponds to an independent MR reconstruction using an acceleration factor of 8 and using HC PET reconstructions for guidance. More details on the various models implemented can be found in Tables I & II.

C. 2D hybrid dataset

The 2D datasets were composed of T1-weighted MPRAGE MR images of patients suspected of epilepsy or dementia, previously collected at St Thomas' PET Center in London. To simulate PET data, MR images were first segmented into grey matter (GM), white matter (WM), cerebrospinal fluid (CSF), skull, and skin using the SPM12 software. Random uptake values of 96.0 ± 5.0 and 32.0 ± 5.0 (arbitrary units) were assigned to GM and WM (3:1 ratio). Spherical lesions were inserted with random radii (2-8 mm) and random locations. The uptake value selected was 144.0 (1.5× of GM), and the maximum number of lesions per image was limited to five. Attenuation maps were generated by assigning attenuation values of 0.13, 0.0975, and 0 cm^{-1} to the skull, tissues, and air. Finally, the voxel sizes and shape of the MR images were resampled from 230×230 and 1.04×1.04 mm^2 to 172×172 and 2.08×2.08 mm^2 to match those of the PET. Data augmentation was performed by rotating PET-MR image pairs by five random

TABLE I
SUMMARY OF THE NETWORKS TRAINED FOR JOINT PET-MR RECONSTRUCTION WITH THEIR RESPECTIVE LOSS FUNCTIONS, INPUTS, AND TARGETS

No.	Model name	Loss function	PET input (counts)	PET target** (counts)	MR input (acceleration factor R)	MR target** (R)
1	Joint_PM_LC_4	P+M	500k	100M	4	1*
2	Joint_PM_LC_8	P+M	500k	100M	8	1*
3	Joint_P_LC_4	P	500k	100M	4	-
4	Joint_P_LC_1	P	500k	100M	1*	-
5	Joint_M_LC_4	M	500k	-	4	1*
6	Joint_M_HC_4	M	100M	-	4	1*
7	Joint_M_LC_8	M	500k	-	8	1*
8	Joint_M_HC_8	M	100M	-	8	1*

* If $R = 1$, no noise is introduced in the k-space data.
** PET targets are OSEM reconstructions ($N_{iter} = 10$, $N_{subsets} = 6$, PSF FWHM= 2.5 mm), MR targets are Landweber reconstructions ($N_{iter} = 60$).

TABLE II
SUMMARY OF THE NETWORKS TRAINED FOR INDEPENDENT PET-MR RECONSTRUCTION WITH THEIR RESPECTIVE LOSS FUNCTIONS, INPUTS, TARGETS, AND IMAGES USED FOR GUIDANCE

No.	Model name	Loss function	PET input (counts)	MR input (acceleration factor R)	PET image for guidance**	MR image for guidance (R)**
1	Indep_P_LC	P	500k	-	-	-
2	Indep_P_LC_guided	P	500k	-	-	1*
3	Indep_M_4	M	-	4	-	-
4	Indep_M_4_guidedHC	M	-	4	100M	-
5	Indep_M_8	M	-	8	-	-
6	Indep_M_8_guidedHC	M	-	8	100M	-
7	Indep_M_8_guidedGT	M	-	8	$\infty^{\dagger}$	-

* If $R = 1$, no noise is introduced in the k-space data.
** PET images used as targets or for guidance are OSEM reconstructions ($N_{iter} = 10$, $N_{subsets} = 6$, PSF FWHM= 2.5 mm). MR images used as targets or for guidance are Landweber reconstructions ($N_{iter} = 60$).
† An infinite number of counts corresponds to the ground truth phantom image.

angles between 0 and 10°. Once the process was completed, LC sinograms were simulated using a PSF of FWHM= 4 mm, attenuation, normalization, and Poisson noise. Randoms and scatters coincidences were not modeled in the 2D simulations for simplicity. Each sinogram was a vector size of 172×252. The targets were obtained by reconstruction of HC data using the OSEM algorithm with $N_{iter} = 10$, $N_{subsets} = 6$, and PSF modeling with FWHM= 2.5 mm. For MR data, the initial MR images were forward-modeled to get the k-space data which was subsequently undersampled. High levels of Gaussian noise were then introduced to demonstrate the potential of the proposed methods. The data obtained were multiplied by eight simulated 2D CSMs to simulate a PI acquisition. The final dataset was composed of 300 training pairs. The test and validation sets comprised 47 and 50 samples, leaving 203 samples for training. Each sample was composed of the sinogram and k-space data, the attenuation and normalization factors, and the PET and MR targets.

D. 3D PET-MR clinical dataset

The clinical data for this study came from a dataset composed of twenty-two patients suspected of epilepsy or dementia who underwent a scan at St Thomas' Hospital's PET center in London with a Siemens Biograph mMR. For PET, an injection of ∼220 MBq [^{18}F]FDG was performed with an uptake time of ∼60 min. A standard Dixon and UTE sequences were performed to generate the PET attenuation map. Simultaneously to the PET data acquisition, a T1-MPRAGE scan was performed with the following acquisition settings: repetition time: 1700ms, echo time: 2.63 ms, inversion time: 900 ms, number of averages: 1, flip angle: 9°, and acquisition time of 382 s. The original fully-sampled k-space data were not available. Therefore, the same process as for the 2D simulated data was used to generate k-space data. In the absence of original MR data, the CSMs had to be simulated, inducing two approximations. First, to generate 3D CSMs, the same CSMs from section III.D were concatenated, leading to a cylindrical geometry instead of a spherical one. Second, CSMs should be patient-dependent, although the same was used for all the patients in our datasets. MR images were subsequently rigidly registered to PET images with the SPM12 software using a normalized mutual-information cost function and default co-registration parameters. For simplicity, the voxel sizes of the MR images were mapped to the PET image voxel sizes. The

PET targets were obtained by reconstructing a full 15 min scan using the OSEM algorithm (10 iterations, 6 subsets, PSF FWHM= 2.5 mm). Only the first two minutes of the scan were used as input to simulate LC data. This choice induces that the LC data have a different distribution than the HC data. The later uptake, as well as higher counts, are inferred.

E. Network training

The networks were trained to map sinograms to OSEM reconstructions of HC data and k-space data to fully sampled T1-weighted MR images. The networks were implemented using Python. PyTorch was used to implement the joint regularizer and the fusion steps, whereas APIRL (GPU-assisted forward and back projectors) was used for the PET data consistency update [24]. The use of different libraries than PyTorch for data consistency prevented the gradients from flowing through these modules. However, this does not notably impact the final performance of the network as discussed in Section V [37]. An 8GB RAM GPU was used to train the networks for 2D data, while a 24GB RAM GPU was used to train the networks for the 3D clinical PET-MR datasets. The training was supervised by either a single MSE loss or by two MSE loss functions. The maximum number of epochs used was 200 for 2D and 3D data. The Adam algorithm was used to optimize the parameters of the network with a learning rate of 5×10^{-4} for 2D and 1×10^{-4} for 3D data. The training was stopped when the validation loss monotonously increased or stagnated for 20 consecutive epochs. The network parameters providing the lowest validation loss for PET and/or MR were selected to evaluate the different methods. For the training of 3D reconstructions, the networks with the maximum unrolled modules fitting in memory were trained end-to-end. This corresponded to 18 unrolled modules (3 iterations, 6 subsets).

F. Evaluation methods

The different versions of the joint and independent networks were evaluated against conventional reconstruction methods over 100 independent noise realizations. For PET reconstruction, the OSEM algorithm (10 iterations, 6 subsets), the MAPEM algorithm with a quadratic prior (Q-MAPEM), and with a Bowsher prior (B-MAPEM) were used. The MAPEM reconstructions were run for 20 iterations and 6 subsets to ensure the reconstruction's convergence. B-MAPEM was set to use $k = 4$ nearest neighbors. All the PET reconstruction methods used PSF modeling in the forward model with FWHM$= 4$ mm. For MR reconstruction, our method was compared to SENSE and SENSE with a quadratic prior (Q-SENSE). We also compared our deep learned joint reconstructions with the synergistic method using mutually weighted quadratic priors from [3] (wQ-Syn). To evaluate the different methods, their performance was assessed using the normalized root mean square error (NRMSE $= \frac{\text{RMSE}}{\overline{y}}$, with $\overline{y}$ the mean of the target dataset). The hyperparameters of the non-deep learned methods were optimized with a grid search based on the MSE criterion. The performance of the various methods was also evaluated using the bias and standard deviation defined as follows: RMSE $= \sqrt{\text{bias}^2 + \text{SD}^2}$ with:

$$\text{bias} = \sqrt{\frac{\sum\limits_{j\in\Omega}(\bar{x_j} - x_j^{Ref})^2}{\sum\limits_{j\in\Omega}(x_j^{Ref})^2}} \tag{8}$$

and:

$$\text{SD} = \sqrt{\frac{1}{S}\frac{\sum\limits_{s=1}^{S}\sum\limits_{j\in\Omega}(\bar{x_j} - x_j^{(s)})^2}{\sum\limits_{j\in\Omega}(x_j^{Ref})^2}} \tag{9}$$

$\bar{x_j}$ being the mean reconstructed value for voxel *j*, obtained by averaging $S = 100$ independent noisy data realizations from the same object, x^{Ref} being the ground truth image, different to the target obtained by reconstruction of high-count data used to train the networks, and Ω representing the set of image voxels. Unrolled networks with 6, 30, and 60 modules were trained in this study.

IV. Results

A. 2D hybrid data

In order to assess the deep learned methods' robustness to various trainable parameter initializations, all the networks were trained three times. Network names will be used in this section to report the results as indicated in Tables I & II, where full details on the training of each network are given. All the networks were trained from scratch in this study. However, the performance of the networks pre-trained with a double-modality loss (Joint_PM) and later fine-tuned with a single-modality loss to obtain Joint_P and Joint_M was evaluated. The same final performance was achieved, whether the single-modality trained networks were pre-trained or not. Based on this observation, it can be hypothesized that the trainable parameter optimization process consistently reaches close local minima of the highly non-convex loss surface.

1) PET reconstruction:

The results demonstrate the superiority of single-modality supervised joint reconstruction for PET with the proposed framework (Figures 3 & 4). Using identical input data, Joint_P_LC_4 performs better than Joint_PM_LC_4 both qualitatively (Figures 3 & 4) and quantitatively (Figures 5 & 6) with better PET-unique feature reconstruction. When high-quality data are used for the MR input (R=1), the benefits of joint reconstruction compared to MR-guided reconstruction are also evident. Joint_P_LC_1 achieves the lowest NRMSE of all methods over entire images. However, these results could lead to incorrect diagnoses due to wrong lesion quantification. When using fully sampled k-space data for joint networks or high-quality MR images for independent reconstruction, risks of misguidance are increased, leading to a potential loss of PET-unique features. The network overuses high-quality MR information. Nonetheless, the joint reconstruction allows better recovery of PET-unique lesions than the MR-guided method

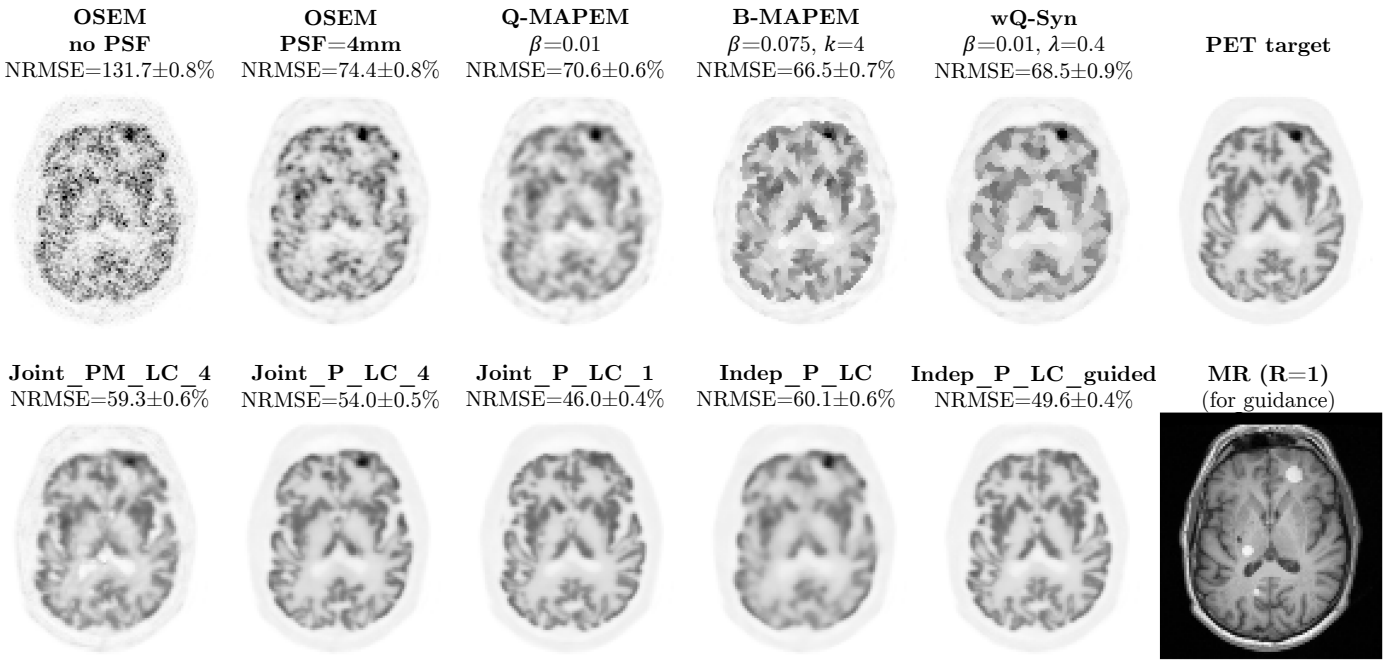

Fig. 3: PET reconstructions of test data, the best mean and standard deviation of the NRMSEs calculated over the entire images from three training runs are reported. Unless stated otherwise, all the methods used a PSF of 4mm in the system model. B-MAPEM used k=4 nearest neighbors. The regularization strength β is reported when relevant. The joint methods on the bottom row unrolled 60 modules.

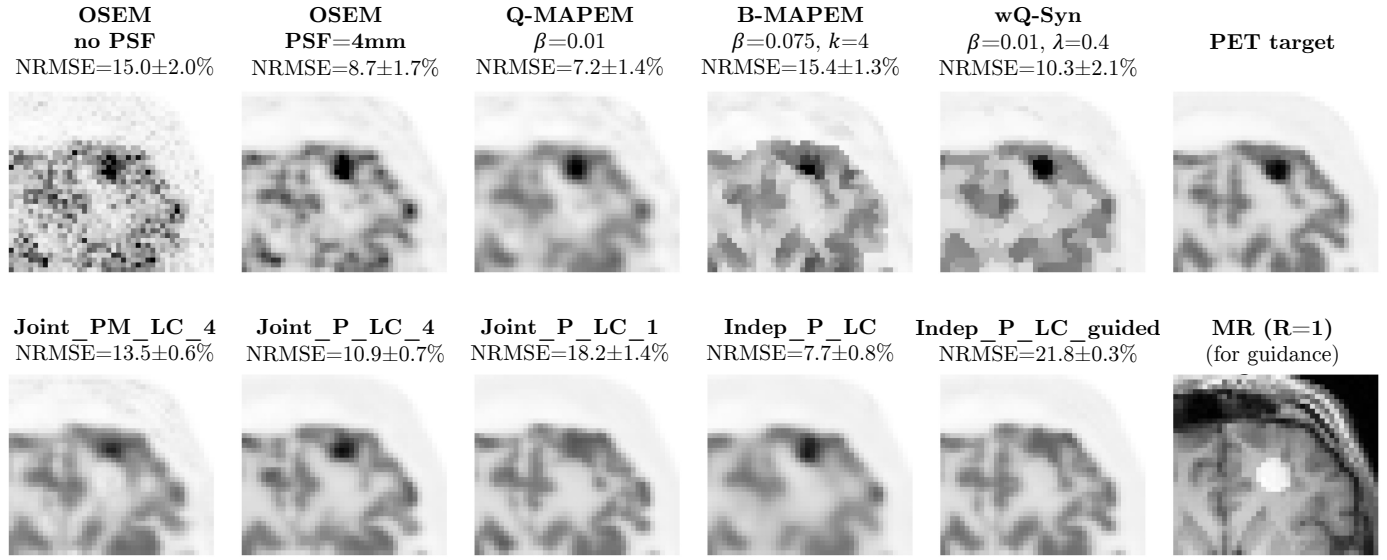

Fig. 4: PET reconstructions of test data, the best mean and standard deviation of the NRMSEs calculated over the PET lesion from three training runs are reported. The PET and MR lesion only overlap over a few pixels.

Indep_P_LC_guided. Having access to all the intermediate updates of the MR ensures better guidance compared to independent MR-guided methods. On the other hand, Joint_P_LC_4, with lower-quality MR input data, achieves accurate PET lesion recovery (Figures 4 & 6). The MR data simultaneously reconstructed is not of high quality in that case. Thus, the joint network learns to use it sparingly for guidance. The cause of this behavior lies in the way the network is trained. Its main goal is to reduce the PET MSE, which is a global loss, as much as possible. The network will tend to overuse MR information, as this process allows for the fastest and most efficient training loss decrease.

Figure 7 shows the trade-off for the PET reconstruction methods between global and local accuracy. Five regions of interest (ROIs) of match and five ROIs of mismatches of size 5×5 pixels, caused by modality-unique lesions, were selected in the test images. The NRMSE was computed for each method and averaged over the five ROIs in the two cases. For the deep learned methods, the best value from the three training runs was retained. The average NRMSE for the regions of match was then plotted against those of the regions of mismatch. Three distinct clusters can be distinguished from this graph.

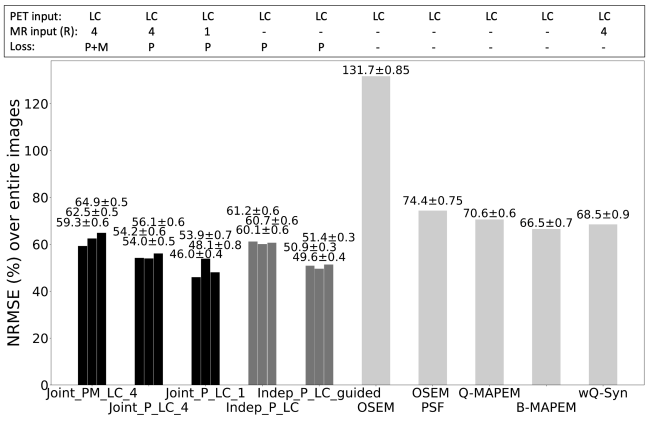

Fig. 5: NRMSE for PET test data over entire images. The 3 bars for the deep learned methods represent three different training runs.

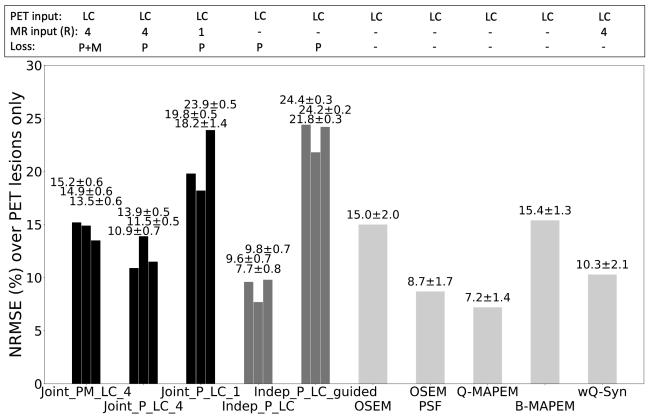

Fig. 6: NRMSE for test data over PET lesion only. The 3 bars for the deep learned methods represent three different training runs.

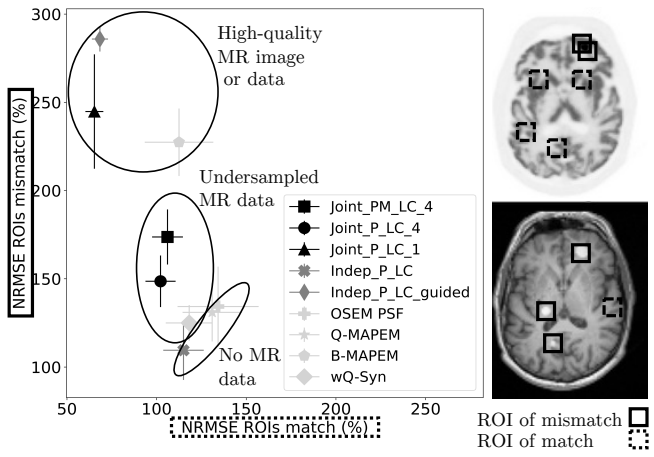

Fig. 7: NRMSEs averaged over 5 regions of interest of match and mismatch for PET test data. Three distinct clusters are visible: methods using high-quality k-space or MR images, methods using undersampled MR data, and methods with no MR guidance.

First, the methods that benefit from high-quality MR images or fully sampled data. Composed of conventional and deep learned methods, whether independent or joint, all the methods in that group perform strongly for match areas but fail to reconstruct regions of mismatch. A second group on the bottom

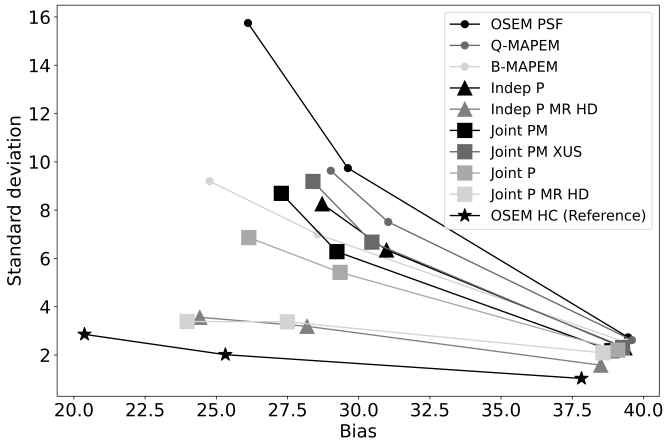

Fig. 8: Bias-standard deviation plot for PET reconstruction. Three markers for each curve correspond to the results for 6, 30, and 60 iterations (from right to left). The uncertainties were too small to be reported on this graph.

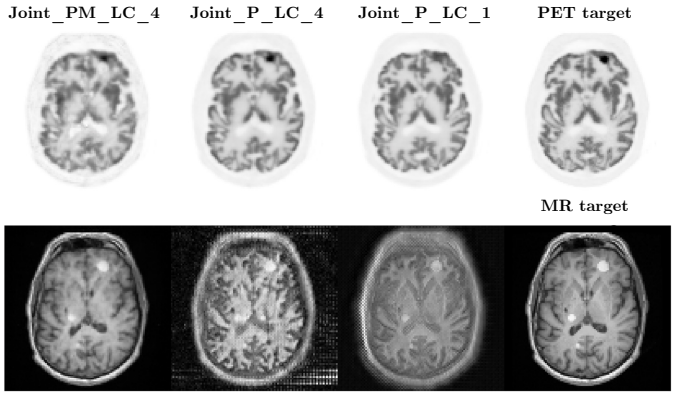

Fig. 9: PET-MR outputs for joint reconstruction. Joint_P_LC_4 and Joint_P_LC_1 learn how to exploit MR information to guide the PET without supervision. Joint_P_LC_4 partly removes aliasing artifacts.

right corner contains methods with no MR guidance, hence the high performance for ROIs of mismatch. However, with no anatomical guidance, the NRMSE over ROIs of match is poor. Joint methods (deep learned and conventional) using mildly undersampled data can be found in between. With Joint_P_LC_4 performing the best, these methods find the best balance between local and global accuracy.

Figure 8 depicts the bias-standard deviation plot. It illustrates how incorporating MR data for guidance enhances the performance of the different techniques by reducing excessive increases in standard deviation. The deep learned networks were trained for 6, 30, and 60 iterations (markers from right to left) to match the corresponding iteration of HC PET data reconstruction. The bias is reduced when the number of iterations increases. The standard deviation increase with the number of iterations is limited compared to other methods for Indep_P_LC_guided and Joint_P_LC_1.

Figure 9 shows the PET and MR outputs obtained in joint reconstruction for Joint_PM_LC_4 and when the joint reconstruction is only supervised by a single-modality loss for PET. Without any supervision of the MR reconstruction, joint

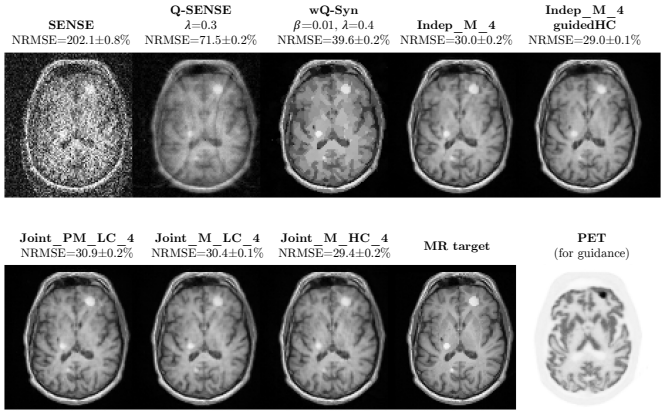

Fig. 10: MR reconstructions of test data for noisy mildly undersampled data (R=4). λ is the regularization strength.

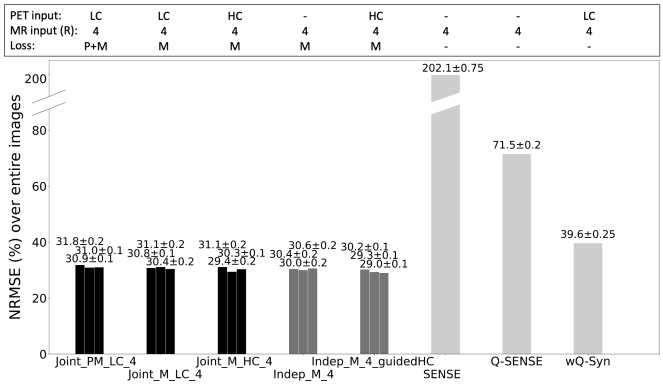

Fig. 11: NRMSE of MR test data over entire images for noisy mildly undersampled data (R=4). The multiple bars for the deep learned methods represent three different training runs.

networks learn how to exploit MR data to optimally guide the PET reconstruction. In the case of Joint_P_LC_4, the network partially removes aliasing artifacts caused by undersampled data (R=4). Additionally, the analysis of these MR outputs indicates that the current way to perform MR-guided PET reconstruction using high-quality MR images might not be optimal. The results suggest that a better alternative would be to perform two pre-processing steps on the high-quality MR image: 1) blurring and 2) increasing contrast between GM and WM.

2) MR reconstruction:

The benefits of PET guidance for R=4 are null. Figures 10 and 11 show that joint and independent methods achieve similar performance. The PET information is not exploited by the networks. The best method for low or moderate acceleration factors is independent deep learned MR reconstruction.

The case of highly undersampled data (R=8) demonstrates the benefits of inserting PET information into the reconstruction process. Visually, the structure becomes much clearer. Figures 12 & 13 show that the best methods evaluated over

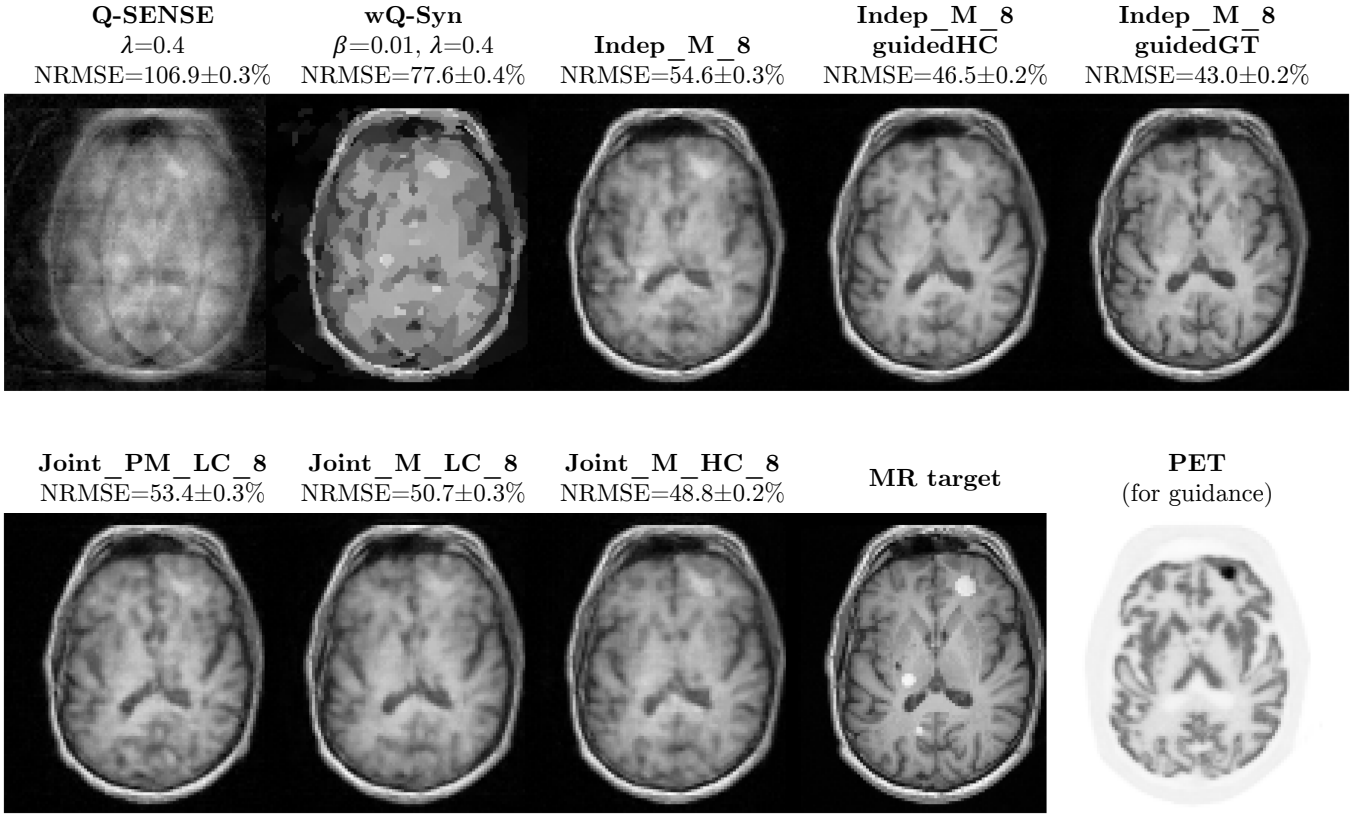

Fig. 12: MR reconstructions of test data with R=8. The best mean, and standard deviation of the NRMSE calculated over the entire images from three training runs are reported. The regularization strength λ is reported when relevant. The joint methods on the bottom row used 60 modules (i.e. 60 iterations).

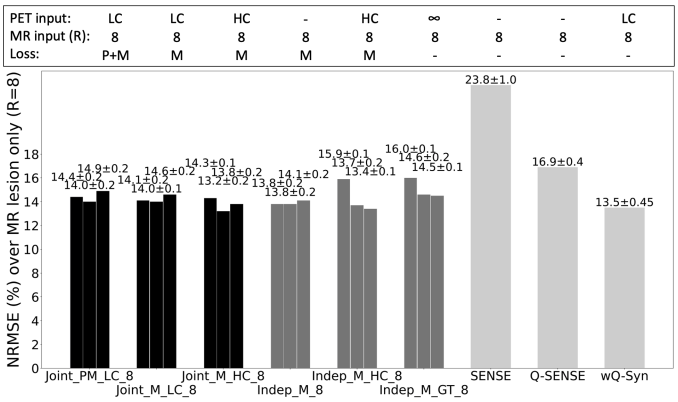

Fig. 13: NRMSE of MR test data over MR-unique lesions for noisy, highly undersampled data (R=8). The multiple bars for the deep learned methods represent three different training runs.

entire images are the ones using PET guidance. A ground truth PET phantom was used to assess the impact of the guidance of (unrealistically) high PET quality. Figure 12 shows how such PET data could potentially impact the quality of the reconstructed MR images. When the NRMSE is evaluated over entire images, Indep_M_8_guidedGT is the best method, as expected. Nonetheless, this method shows its limitations when ROIs of MR-unique lesions are considered (Figure 13). Similarly to MR-guided PET reconstruction, the guidance becomes too strong, and MR-specific lesions are at risk of being removed. The guidance of PET images does not induce the same issues for joint methods demonstrating their superiority for modality-specific features reconstruction again. PET image quality being inherently low, the networks learn not to overuse it for guidance. Therefore, no pre-processing for PET images is required.

B. 3D clinical data

Unrolled network training is computationally and memory-demanding. As a consequence, a joint network composed of 60 modules did not fit in the memory of a 24GB GPU. In order to perform end-to-end training similar to 2D data, the maximum number of modules fitting in memory was unrolled. The resulting joint and independent networks consisted of 18 modules, corresponding to 3 iterations and 6 subsets for accelerated PET reconstruction. To speed up the training, the unrolled networks were initialized with OSEM reconstructions of LC data with 7 iterations and 6 subsets (PSF modeling with FWHM$=4$ mm) and with Landweber reconstructions for MR using 42 iterations. The results shown in Figures 14 & 15 are consistent with the results obtained for 2D data.

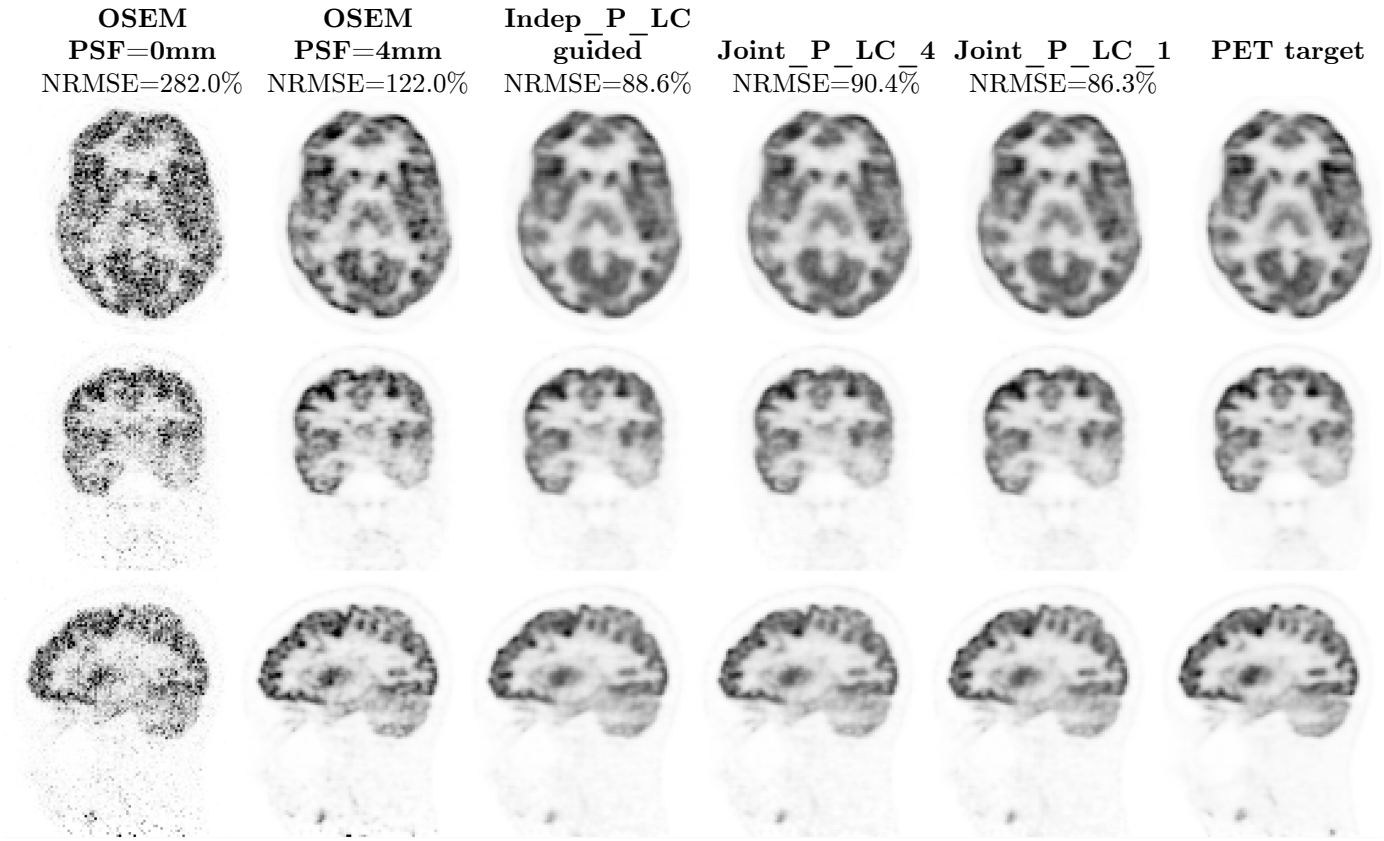

Fig. 14: PET real data reconstructions. The target was built by reconstruction of the full scan (15 minutes). The input of the various algorithms was the first two minutes of the scan to simulate LC data. Only 18 modules were unrolled for real data. The reconstructions were initialized with OSEM reconstructions (7 iterations, 6 subsets). The performance of the various methods is consistent with the results obtained for 2D simulated data.

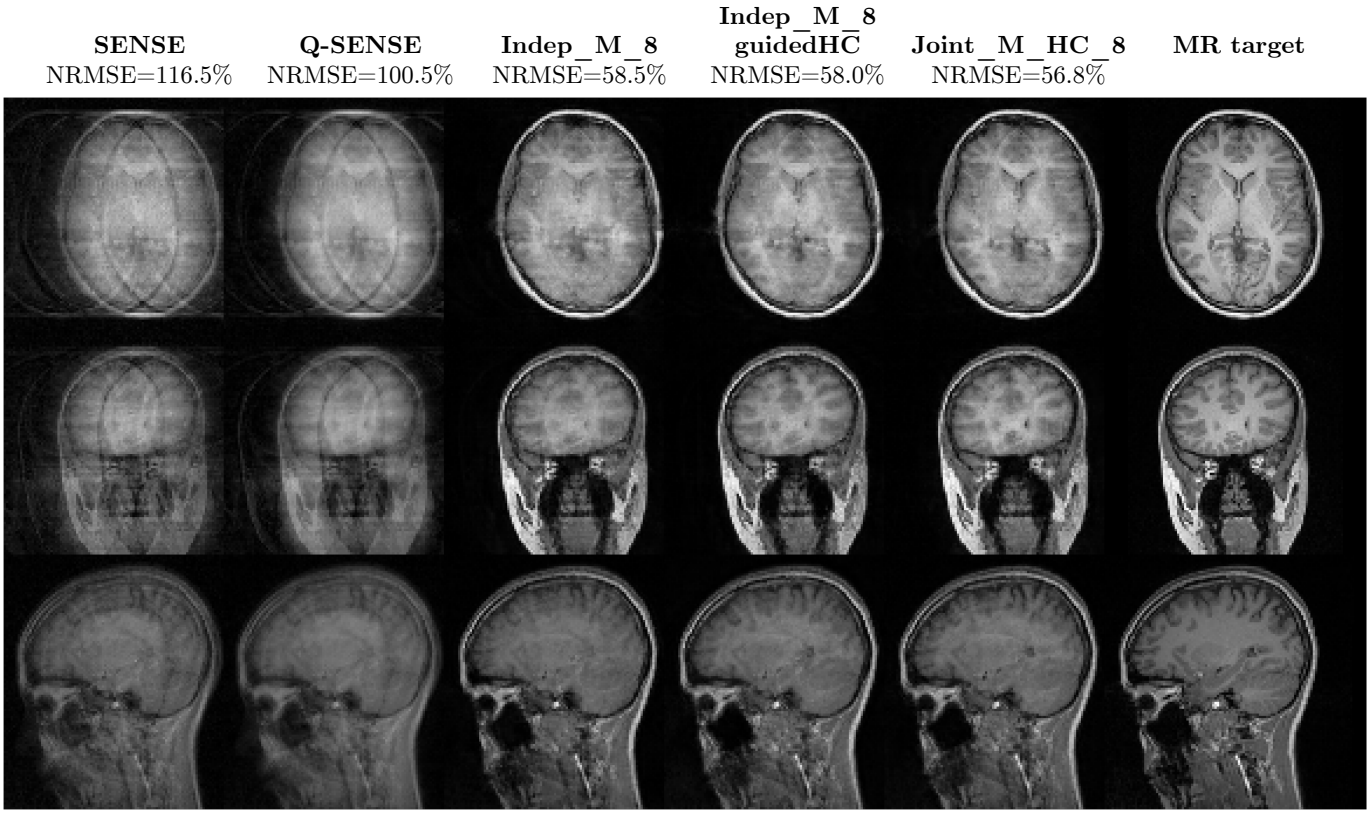

Fig. 15: MR real data reconstructions. The target was built by reconstruction of the noise-free fully-sampled k-space. The input of the various methods corresponded to noisy, undersampled data with an acceleration factor of 8 using parallel imaging (8 coils). Only 18 modules were unrolled for real data. The reconstructions were initialized with Landweber reconstructions (42 iterations). The performance of the various methods is consistent with the results obtained for 2D simulated data.

The method with the lowest reconstruction error over entire images is Joint_P_LC_1. Since the complete data for the scan was not accessible, the HC targets of the clinical datasets were reconstructed using only 15 minutes out of the 60-minute scan. Consequently, high-quality features were lacking in the targets, and so utilization of MR information by the networks for reconstructing the LC data was not as strong as for the simulated data case, where the targets were of much higher quality. Higher-quality targets can allow a network to exploit high-quality MR data to improve the reconstruction of LC data. As no known areas of mismatch were present in the training and test data, an analysis similar to the one performed on 2D data is needed to assess the accuracy of the reconstruction on ROIs of mismatch.

V. Discussion & Conclusion

A new method for joint deep learned PET-MR reconstruction is presented in this work. It is demonstrated that better results are obtained for the proposed joint reconstruction framework when the network is trained with a single-modality loss. Joint networks trained for PET reconstruction only achieve better accuracy than MR-guided methods globally. It is shown that using mildly undersampled data instead of fully sampled k-space for the joint networks allows for better PET-unique feature recovery. The results from joint reconstruction also indicate that the current methods for MR-guided PET reconstruction use suboptimal MR images for this task. The results also show that joint methods become beneficial for MR reconstruction only for highly undersampled data. Deep learned independent methods remain the best for mildly undersampled MR data. In future work, a small component performing a mapping between PET and MR space could be added in order to reconstruct MR images with their native high resolution. Finally, the undersampling pattern could also be learned to obtain an optimal acquisition trajectory followed by an optimized reconstruction [21], [34].

In this work, the various components of each module were not all implemented in PyTorch. Thus, the backpropagation of the gradients did not pass through the data consistency updates of PET and MR. In order to assess the impact of truncated backpropagation, forward and back projectors were implemented in PyTorch. These were used to train the unrolled PET reconstruction Indep_P_LC, and the results were compared with the same network trained without backpropagation through the EM update. Three observations were made from this investigation. The first was that the two networks achieved very similar performance. Less than 3% of difference in NRMSE was noted in favor of the network trained with the correct backpropagation over three training runs. The second observation was made on the loss function. As stated before, the training and validation loss curves reached a very close minimum, but the network with correct backpropagation

reached that minimum a few epochs earlier. Finally, the difference in training time between the two implementations justified the choice not to backpropagate through the EM module. The naive implementations of the forward and back projectors in PyTorch used were not optimized for fast execution. Developing efficient projectors was out of the scope of this article. Therefore, the forward and back projection operations were notably longer in PyTorch than in Python. Training joint networks for PET-MR reconstruction took approximately 15 hours, with 60 modules unrolled for 2D data (203 samples in the training set) and no backpropagation through the EM module. The training time was multiplied by five when the PyTorch projectors were used. The current study necessitated the training of more than 70 unrolled networks, it would have become highly time-consuming if the PyTorch projectors were used. More results on the impact of truncated backpropagation versus accurate backpropagation can be found in [37] for SPECT reconstruction.

It would be envisioned that with a much larger training set in tandem with a more highly parameterized joint network, better results would be achieved. However, medical data is scarce, and this work emphasized the generalization of the proposed method to unseen data. Under this particular regime, only shallow networks with a limited number of trainable parameters can perform well and reliably. Using deep networks with fewer parameters offers significant advantages, such as greater speed, but also limits the expressiveness of the network.

The networks were trained with a global loss but were evaluated in local areas. The idea of introducing a new term in the loss function dealing with local accuracy could be considered. Nevertheless, it is challenging to design a local loss and correctly balance its importance against the global loss. Besides, the MSE loss heavily penalizes outliers. The wrong quantification of a lesion should already be taken care of with this loss alone. It could be of interest to look at regional average errors to give an assessment of quantification within a lesion in future work.

The various unrolled methods presented have not been compared to post-processing networks. Since such results were already proposed in the work of Mehranian and Reader [4] for PET image reconstruction and in the work of Hammernik *et al.* [20] for MR image reconstruction, it was decided to build upon those previous findings rather than revisit the question here.

Finally, the choice of the reconstruction algorithm for both PET and MR is based on two previously published independent reconstruction methods using unrolled networks: FBSEM-Net for PET [4], and the VN for MR [5] unfolding Landweber's algorithm. However, in order to choose the best combination of methods (proximal mapping, gradient descent, etc) for joint PET-MR reconstruction of brain data, a systematic evaluation of various methods such as in [20] should be performed.

ACKNOWLEDGMENT

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. Data supporting this article can only be shared in anonymized form due to the type of participant consent obtained; please contact the corresponding author if required. All authors declare that they have no known conflicts of interest in terms of competing financial interests or personal relationships that could have an influence or are relevant to the work reported in this paper.

REFERENCES

[1] L. A. Shepp and Y. Vardi, "Maximum Likelihood Reconstruction for Emission Tomography," in IEEE Transactions on Medical Imaging, vol. 1, no. 2, pp. 113-122, Oct. 1982, doi: 10.1109/TMI.1982.4307558.

[2] Hudson HM and Larkin RS. Accelerated image reconstruction using ordered subsets of projection data. IEEE Trans Med Imaging. 1994;13(4):601-9. doi: 10.1109/42.363108. PMID: 18218538.

[3] Mehranian, A, Belzunce, MA, McGinnity, CJ, et al. Multi-modal synergistic PET and MR reconstruction using mutually weighted quadratic priors. Magn Reson Med. 2018; 00: 2120– 2134. https://doi.org/10.1002/mrm.27521

[4] A. Mehranian and A. J. Reader, "Model-based deep learning pet image reconstruction using forward-backward splitting expectation– maximization," IEEE Trans. Radiat. Plasma Med. Sci., vol. 5, no. 1, pp. 54–64, Jan. 2021, doi: 10.1109/TRPMS.2020.3004408.

[5] Hammernik, K., Klatzer, T., Kobler, E., Recht, M.P., Sodickson, D.K., Pock, T. and Knoll, F. (2018), Learning a variational network for reconstruction of accelerated MRI data. Magn. Reson. Med., 79: 3055-3071. https://doi.org/10.1002/mrm.26977

[6] Corda-D'Incan G., Schnabel J. A. and Reader A. J. "Syn-Net for Synergistic Deep Learned PET-MR Reconstruction", IEEE 2020 NSS MIC Proceedings.

[7] F. Knoll, M. Holler, T. Koesters, K. Bredies and D. K. Sodickson, "Simultaneous PET-MRI reconstruction with vectorial second order total generalized variation," 2015 IEEE Nuclear Science Symposium and Medical Imaging Conference (NSS/MIC), 2015, pp. 1-4, doi: 10.1109/NSSMIC.2015.7582009.

[8] M. J. Ehrhardt et al. "Joint reconstruction of PET-MRI by exploiting structural similarity." Inverse Problems 31 (2015): 015001.

[9] Pruessmann KP, Weiger M, Scheidegger MB, Boesiger P. "SENSE: Sensitivity encoding for fast MRI", Magn Reson Med 1999.

[10] Chun, Il and Fessler, Jeffrey. (2018). "Deep BCD-Net Using Identical Encoding-Decoding CNN Structures for Iterative Image Recovery", 2018 IEEE 13th Image, Video, and Multidimensional Signal Processing Workshop (IVMSP).

[11] Corda-D'Incan G., Schnabel J. A. and Reader A. J. "Memory-Efficient Training for Fully Unrolled Deep Learned PET Image Reconstruction with Iteration-Dependent Targets", IEEE Transactions on Radiation and Plasma Medical Sciences.

[12] Bowsher J E et al, "Utilizing MRI information to estimate F18-FDG distributions in rat flank tumors", IEEE NSS MIC Proceedings 2004.

[13] Landweber L. "An Iteration Formula for Fredholm Integral Equations of the FirstKind", American Journal of Mathematics 1951.

[14] A. R. De Pierro, "On the relation between the ISRA and the EM algorithm for positron emission tomography," IEEE Trans. Med. Imag., vol. 12, no. 2, pp. 328–333, Jun. 1993, doi: 10.1109/42.232263.

[15] J. Schlemper, J. Caballero, J. V. Hajnal, A. N. Price and D. Rueckert, "A deep cascade of convolutional neural networks for dynamic MR image reconstruction", IEEE Transactions on Medical Imaging, 2017.

[16] K. Gong et al., "MAPEM-Net: An unrolled neural network for Fully 662 3D PET image reconstruction," in Proc. SPIE, 2019, Art. no. 110720O, 663 doi: 10.1117/12.2534904.

[17] Hongyang Lu, Jingbo Wei, Qiegen Liu, Yuhao Wang, Xiaohua Deng, "A Dictionary Learning Method with Total Generalized Variation for MRI Reconstruction", International Journal of Biomedical Imaging, vol. 2016, Article ID 7512471, 13 pages, 2016. https://doi.org/10.1155/2016/7512471

[18] Liu Y., Zhan Z., Cai J., Guo D., Chen, Z. and Qu X. "Projected iterative soft-thresholding algorithm for tight frames in compressed sensing magnetic resonance imaging", IEEE Trans. Med. Imaging 2016, 35, 2130–2140.

[19] Zhou Ding-Xuan, "Universality of deep convolutional neural networks", Applied and Computational Harmonic Analysis. (2019) 48. 10.1016/j.acha.2019.06.004.

[20] Hammernik, K, Schlemper, J, Qin, C, et al. ”Systematic evaluation of iterative deep neural networks for fast parallel MRI reconstruction with sensitivity-weighted coil combination”. Magn Reson Med. 2021; 86: 1859– 1872. https://doi.org/10.1002/mrm.28827

[21] C. D. Bahadir, A. Q. Wang, A. V. Dalca and M. R. Sabuncu, ”Deep-Learning-Based Optimization of the Under-Sampling Pattern in MRI,” in IEEE Transactions on Computational Imaging, vol. 6, pp. 1139-1152, 2020, doi: 10.1109/TCI.2020.3006727.

[22] Gu, Hongyi & Yaman, Burhaneddin & Moeller, Steen & Ellermann, Jutta & Ugurbil, Kamil & Akçakaya, Mehmet. (2022). Revisiting 1-wavelet compressed-sensing MRI in the era of deep learning. Proceedings of the National Academy of Sciences. 119. 10.1073/pnas.2201062119.

[23] Antun, Vegard & Renna, Francesco Poon, Clarice & Adcock, Ben & Hansen, Anders. (2020). On instabilities of deep learning in image reconstruction and the potential costs of AI. Proceedings of the National Academy of Sciences. 117. 201907377. 10.1073/pnas.1907377117.

[24] Belzunce, M.A. and Reader, A.J. (2017), Assessment of the impact of modeling axial compression on PET image reconstruction. Med. Phys., 44: 5172-5186. https://doi.org/10.1002/mp.12454

[25] 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, USA: Springer, 2011, pp. 185–212.

[26] A.R. De Pierro, ”A modified expectation-maximization algorithm for penalized likelihood estimation in emission tomography”, IEEE Trans. Med. Imaging. 1995;14:132–137.

[27] Ronneberger, Olaf & Fischer, Philipp & Brox, Thomas. (2015). U-Net: Convolutional Networks for Biomedical Image Segmentation. LNCS. 9351. 234-241. 10.1007/978-3-319-24574-4_28.

[28] A. J. Reader, G. Corda, A. Mehranian, C. d. Costa-Luis, S. Ellis and J. A. Schnabel, ”Deep Learning for PET Image Reconstruction,” in IEEE Transactions on Radiation and Plasma Medical Sciences, vol. 5, no. 1, pp. 1-25, Jan. 2021, doi: 10.1109/TRPMS.2020.3014786.

[29] Sai Ravishankar, Jong Chul Ye, J A Fessler, ”Image reconstruction: from sparsity to data-adaptive methods and machine learning”, Proc. IEEE, 108(1):86-109, Jan. 2020.

[30] G. Corda-D’Incan, J. A. Schnabel and A. J. Reader, ”Dense Syn-Net: Inter-Modal and Self-Guided Deep Learned PET-MR Reconstruction,” 2021 IEEE Nuclear Science Symposium and Medical Imaging Conference (NSS/MIC), 2021, pp. 1-5, doi: 10.1109/NSS/MIC44867.2021.9875927.

[31] Ross S. Q.Clear (GE Healthcare White Paper), 2015.

[32] Akbarzadeh, A., Ay, M.R., Ahmadian, A., Riahi Alam, N., Zaidi, H., 2013a. MRI-guided attenuation correction in whole-body PET/MR: assessment of the effect of bone attenuation. Ann. Nucl. Med. 27, 152–162.

[33] M. A. Belzunce, A. Mehranian, Z. Chalampalakis, and A. J. Reader, “Evaluation of shift-invariant image-space PSFs for the biograph mMR PET scanner,” in Proc. PSMR Conf., Lisbon, Portugal, May 2017.

[34] Wang, Guanhua & Luo, Tianrui & Nielsen, Jon-Fredrik & Noll, Douglas & Fessler, Jeff. (2022). B-spline Parameterized Joint Optimization of Reconstruction and K-space Trajectories (BJORK) for Accelerated 2D MRI. IEEE Transactions on Medical Imaging. PP. 1-1. 10.1109/TMI.2022.3161875.

[35] Zhu B, Liu JZ, Cauley SF, Rosen BR, Rosen MS. Image reconstruction by domain-transform manifold learning. Nature. 2018 Mar 21;555(7697):487-492. doi: 10.1038/nature25988. PMID: 29565357.

[36] Häggström I, Schmidtlein CR, Campanella G, Fuchs TJ. DeepPET: A deep encoder-decoder network for directly solving the PET image reconstruction inverse problem. Med Image Anal. 2019 May;54:253-262. doi: 10.1016/j.media.2019.03.013. Epub 2019 Mar 30. PMID: 30954852; PMCID: PMC6537887.

[37] Z. Li, Y. K. Dewaraja and J. A. Fessler, ”Training End-to-End Unrolled Iterative Neural Networks for SPECT Image Reconstruction,” in IEEE Transactions on Radiation and Plasma Medical Sciences, doi: 10.1109/TRPMS.2023.3240934.