Investigating Intensity Normalisation for PET Reconstruction with Supervised Deep Learning

Imraj Singh, Alexander Denker, Bangti Jin, Kris Thielemans, Simon Arridge

***Abstract*—Deep learning methods have shown great promise in the field of Positron Emission Tomography (PET) reconstruction, but the successful application of these methods depends heavily on the intensity scale of the images. Normalisation is a crucial step that aims to adjust the intensity of network inputs to make them more uniform and comparable across samples, acquisition times, and activity levels. In this work, we study the influence of different linear intensity normalisation approaches. We focus on two popular deep learning based image reconstruction methods: an unrolled algorithm (Learned Primal-Dual) and a post-processing method (OSEMConvNet). Results on the out-of-distribution test dataset demonstrate that the choice of intensity normalisation significantly impacts on generalisability of these methods. Normalisation using the mean of acquisition data or corrected acquisition data led to improved peak-signal-to-noise-ratio (PSNR) and data-consistency (KLDIV). Through evaluation of lesion-specific metrics of contrast recovery coefficients (CRC) and standard deviation (STD) an increase in CRC and STD is observed. These findings highlight the importance of carefully selecting an appropriate normalisation method for supervised deep learning-based PET reconstruction applications.**

***Index Terms*—Image reconstruction, deep learning, positron emission tomography, intensity normalisation**

I. Introduction

POSITRON Emission Tomography (PET) reconstruction is a particularly challenging inverse problem due to the Poisson noise and the complexity of the forward model, typically includes projection, attenuation, resolution, sensitivity, scatter and randoms. Given an image x, the estimate of the mean $\bar{y}$ is given by $\bar{y} = Ax + b$, where the forward model is affine and comprises of a system matrix A and background b. The PET acquisition data y is corrupted by Poisson noise and hence the noise is dependent on intensity-scale and varies with scan times and activity.

In recent years supervised deep learning methods for PET reconstruction have shown promising results [1]. In this work we investigate data-specific intensity normalisation for PET reconstruction with supervised deep learning methods, and show that normalisation can improve generalisability to out-of-distribution samples and noise levels.

This work is supported by the EPSRC-funded UCL Centre for Doctoral Training in Intelligent, Integrated Imaging in Healthcare (i4Health) (EP/S021930/1), and the Deutsche Forschungsgemeinschaft (DFG) within the framework of GRK 2224/1 "π^3: Parameter Identification – Analysis, Algorithms, Applications".

I. Singh (email: imraj.singh.20@ucl.ac.uk) and S. Arridge (email: simon.arridge@cs.ucl.ac.uk) are with the Department of Computer Science, UCL, London, United Kingdom. A. Denker (email: adenker@uni-bremen.de) is with the Center of Industrial Mathematics, University of Bremen. B. Jin (email: b.jin@cuhk.edu.hk) is with the Department of Mathematics, The Chinese University of Hong Kong, Shatin, N.T., Hong Kong. K. Thielemans (email: k.thielemans@ucl.ac.uk) is with Institute of Nuclear Medicine & Centre for Medical Image Computing, UCL, London, United Kingdom.

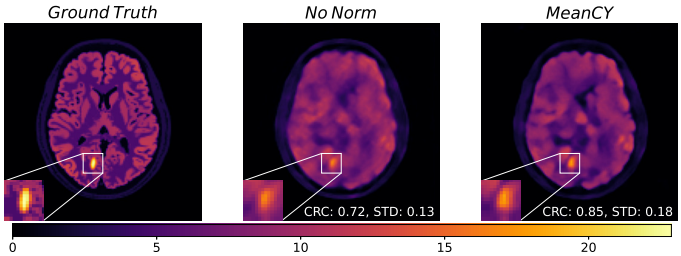

Fig. 1. Reconstruction for LPD with a noise level of 2.5 counts per volume with no normalisation and normalisation by *MeanCY*.

II. Methods

Given a supervised dataset $\{y_i, x_{t,i}\}_{i=1}^N$ of N samples, the training loss is given by:

$$\min_\theta \sum_{i=1}^{N} \|M_i^{-1} R_\theta(M_i y_i) - x_{t,i}\|_2^2,$$

where R_θ is the neural network with trainable parameters θ, and M_i the sample-dependent intensity normalisation. Our goal is to normalise the inputs of the supervised deep learning method. To this end, we normalise acquisition data y_i (and/or OSEM reconstructions) to have a similar scale. By normalising y_i, we stabilise the variation of the inputs to the network. For a Gaussian noise model, this corresponds to pre-whitening. In this work we test four normalisation methods and $M_i = I$, i.e. no normalisation (*No Norm*), as a baseline. Here, *MeanY* and *MeanCY* are based on the acquisition data y_i and *MeanOSEM* and *MaxOSEM* are based on an OSEM reconstruction $\hat{x}_i$.

	MeanY	MeanCY	MeanOSEM	MaxOSEM
M_i	$\mathrm{diag}[\frac{1}{\mathrm{Mean}(y_i)}]$	$\mathrm{diag}[\frac{1}{\mathrm{Mean}(y_i - b_i)}]$	$\mathrm{diag}[\frac{1}{\mathrm{Mean}(\hat{x}_i)}]$	$\mathrm{diag}[\frac{1}{\max(\hat{x}_i)}]$

A. Deep Learning Methods

We use two well-established deep learning methods for inverse problems in medical imaging, namely, FBPConvNet [2] and Learned Primal Dual (LPD) [3]. These methods were subsequently adapted for PET reconstruction. The FBP-ConvNet input was changed from a filtered-back-projection to an OSEM reconstruction after 1 epoch with 34 subsets. This post-processing method is referred to as OSEMConvNet. The LPD was adapted by using the OSEM reconstruction as initialisation of primal channels, and by including an affine forward operator with sample-specific multiplicative and additive factors. These sample-specific factors are not included in previous LPD implementations for PET [4]. To reduce computational time a total of 3 unrolled iterations were used. For both methods the network architectures used were changed minimally from DivαL [5]. OSEMConvNet used a UNet architecture with 1,783,249 trainable parameters, and the LPD used a set of convolutional filters at each iteration with 132,300 trainable parameters. Both networks were trained

using a joint dataset corresponding to $5, 10, 50$ true counts per volume. The code for the experiments is publically available[1].

B. BrainWeb Dataset and pyParallelProj

The BrainWeb dataset of 20 anatomical phantoms [6] comprised the test (subject 04) and training sets (other subjects). These high resolution (2 mm^3) phantoms defined ground truth PET and CT volumes. Furthermore, the PET training ground truths were perturbed to give three realisations per-patient [7]. Axial slices of ground-truth data were used to simulate PET acquisition data. The test set included ground-truths with simulated hot elliptical lesions, and associated background Region Of Interest (ROI). True counts were projected using a true forward model defined by pyParallelProj [8]. The true model used a single crystal ring GE Discovery MI acquisition geometry with count-based projector, and resolution, attenuation and sensitivity models were included. The noise level was set by re-scaling the true data to ensure a prescribed true count per emission voxel. For supervised training data the prescribed true counts per emission voxel were $5, 10, 50$, and for test data it was 2.5. After scaling, a constant background based on the mean true counts was added before applying Poisson corruption. The approximate forward model for reconstruction was modelled with low resolution images (8 mm^3) to avoid an inverse crime.

C. Quality metrics

We evaluate the trained models on test data with simulated lesions. In addition to the peak-signal-to-noise value (PSNR), we evaluate two PET specific quality scores computed over a Region of Interest (ROI) that are important for medical use: the contrast recovery coefficient (CRC) and the standard deviation (STD). The CRC measures the ability to reconstruct small differences in tracer concentration between different ROIs. We compute the CRC between lesion L and background B ROIs. Given an ROI Z, we denote the average over the elements of the ROI as $\bar{Z}_r = \frac{1}{N_b}\sum_{k=1}^{N_Z} Z_{r,k}$, and average over realisation is $\bar{Z}_k = \frac{1}{R}\sum_{k=1}^{R} Z_{r,k}$. The CRC is defined:

$$\text{CRC} := \frac{1}{R}\sum_{r=1}^{R}\left(\frac{\bar{L}_r}{\bar{B}_r} - 1\right) / \left(\frac{L_\text{t}}{B_\text{t}} - 1\right), \tag{1}$$

where $\frac{L_\text{t}}{B_\text{t}}$ is the true lesion-to-background contrast. We evaluate the normalised STD on background ROI across realisations, and then averaged over ROI voxels:

$$\text{STD} := \frac{1}{N_B}\sum_{k=1}^{N_B}\sqrt{\frac{1}{R-1}\sum_{r=1}^{R}\frac{(B_{r,k} - \bar{B}_k)^2}{\bar{B}_k}} \tag{2}$$

For our evaluation we used $R = 10$ noise realisations of the acquisition data. In addition, we computed the Kullback-Leibler divergence (KLDIV) between the true measurements y and estimated measurements $\bar{y}$ as a the metric of data consistency, and peak signal-to-noise ratio (PSNR) between ground-truth and reconstructed images.

[1]https://github.com/Imraj-Singh/pet_supervised_normalisation

TABLE I
OSEMCONVNET: MEAN (STANDARD ERROR) OF THE QUALITY METRICS ON 80 SAMPLE TEST SET

	No Norm	MeanY	MeanCY	MeanOSEM	MaxOSEM
CRC (↑)	0.749 (0.014)	**0.805***** (0.015)	0.788*** (0.014)	0.803*** (0.015)	0.758*** (0.017)
STD (↓)	**0.216** (0.005)	0.248 (0.007)	0.244 (0.007)	0.247 (0.008)	0.264 (0.008)
PSNR (↑)	27.55 (0.38)	27.93*** (0.40)	**27.94***** (0.40)	**27.94***** (0.39)	27.08* (0.43)
KLDIV (↓)	68128 (127)	**63068***** (246)	63115*** (245)	63069*** (246)	63667*** (268)

$^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$

TABLE II
LEARNED PRIMAL-DUAL: MEAN (STANDARD ERROR) OF THE QUALITY METRICS ON 80 SAMPLE TEST SET

	No Norm	MeanY	MeanCY	MeanOSEM	MaxOSEM
CRC (↑)	0.749 (0.016)	0.876*** (0.013)	0.862*** (0.013)	**0.893***** (0.011)	0.874*** (0.011)
STD (↓)	**0.163** (0.005)	0.230 (0.008)	0.227 (0.008)	0.278 (0.006)	0.250 (0.008)
PSNR (↑)	27.94 (0.44)	28.38*** (0.44)	**28.50***** (0.44)	27.89* (0.41)	28.00** (0.43)
KLDIV (↓)	63925 (239)	63153*** (225)	63027*** (243)	**62958***** (238)	63089*** (249)

$^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$

III. RESULTS AND DISCUSSION

We show results for out-of-distribution test data (2.5 true counts per volume with lesions) in Tables I and II for OSEMConvNet and LPD respectively. In addition, Figure 1 gives a qualitative comparison of sample with *No Norm* and *MeanCY*. We find that models with normalisation benefit from higher CRCs and lower KLDIV. The STD is higher with normalisation, meaning reconstructions are more sensitive to perturbations in noise, and is a result of being more data-consistent. Normalisation based on OSEM images shows instability with regards to PSNR and can under-perform as compared with *No Norm*. More work is needed to conclusively establish the best normalisation practice. We leave non-linear normalisation techniques, and mixed/weighted loss functions for further work.

REFERENCES

[1] A. J. Reader et al. Deep learning for PET image reconstruction, *IEEE Trans. Radiat. Plasma Med. Sci.*, **5.1** (2020): 1-25.

[2] K. Jin et al. Deep convolutional neural network for inverse problems in imaging, *IEEE Trans. Image Process.* 26.9 (2017): 4509-4522.

[3] J. Adler et al. Learned primal-dual reconstruction, *IEEE Trans. Med. Imaging*, **37.6**, (2018): 1322-1332.

[4] A. Guazzo et al. Learned Primal Dual Reconstruction for PET, *J. Imaging* **7.12** (2021): 248.

[5] J. Leuschner et al. DIVal Library Zenodo, https://doi.org/10.5281/zenodo.3970517 (2020).

[6] B. Aubert-Broche et al. Twenty new digital brain phantoms for creation of validation image data bases, *IEEE Trans. Med. Imaging*, **25.11**, (2006): 1410-1416.

[7] G. Schramm, Simulated brainweb PET/MR data sets for denoising and deblurring Zenodo, https://doi.org/10.5281/zenodo.4897350 (2021).

[8] G. Schramm, PARALLELPROJ – An open-source framework for fast calculation of projections in tomography, arXiv, https://doi.org/10.48550/arXiv.2212.12519 (2022).