Convolutional micro-networks for MR-guided low dose PET post-processing

Casper O. da Costa-Luis casper.dcl@kcl.ac.uk
Andrew J. Reader andrew.reader@kcl.ac.uk
School of Biomedical Engineering and Imaging Sciences
King's College London, St Thomas' Hospital, London SE1 7EH, United Kingdom

Abstract

We propose a novel micro-network (μ-net) architecture which is a convolutional neural network (CNN) designed to be robust to minimal training data. Applying the μ-net to MR-guided low count PET scans produces impressive qualitative and quantitative cross-validation results despite training on just two noise realisations of the same patient.

1 Introduction

While reducing scan time and radiation dose in positron emission tomography (PET) clearly has safety and cost benefits, it also results in acquired data with fewer overall counts, and thus high levels of noise. Simultaneously acquired magnetic resonance (MR) data can be used in techniques such as non-local means to reduce noise in PET reconstructions [1].

This work proposes an alternative post-processing step informed by deep learning (DL) – specifically, deep convolutional neural networks (CNNs). Uptake of CNNs in medical imaging has been comparatively recent and modest. Some proposals suggest 2-dimensional (2D) patch-based CNNs to reduce noise in low-dose CT (computed tomography) and PET-MR reconstruction [2], [3]. However, such networks typically required large amounts of training data. In contrast, this work focuses on improving image quality through carefully designed intuitive 3D CNNs. Due to the resultant small network size, we propose the term micro-network, or μ-net. These μ-nets have a comparatively small parameter space and thus are robust to extremely small training data sets, in stark contrast to the U-nets found in current literature [4].

2 Methods

2.1 Data

The supervised training process requires known target truths. Rather than scanning an unrealistic phantom (lacking the structure and features of a patient) or relying on a pure simulations (which can also be unrealistic), we use a hybrid approach. Data from two [^{18}F]FDG PET head scans using the Siemens Biograph mMR scanner was reconstructed using maximum likelihood expectation maximisation (MLEM) with the method of sieves [5], accounting for photon attenuation, normalisation, randoms and scatter. Each dataset had $430\,\mathrm{M}$ counts, of which it was estimated $28\,\%$ were scatter and $26\,\%$ were randoms. The simultaneously acquired MR data was reconstructed and downsampled to the same dimensions as the PET images.

Since the target truth is not known, the PET data was sharpened to improve edge contrast (2), and then treated as the target, $\boldsymbol{\tau}$. $\boldsymbol{\tau}$ is subsequently projected to give multiple sinogram measurements $\boldsymbol{m}$ at low ($43\,\mathrm{M}$) counts, which were each reconstructed to produce reference PET images $\boldsymbol{\theta}$.

1st Conference on Medical Imaging with Deep Learning (MIDL 2018), Amsterdam, The Netherlands.

For all the data sets, 300 [100] MLEM iterations were performed with [without] resolution modelling (RM) as per (1). The RM kernel used was a a Gaussian point spread function (PSF) of 4.5mm FWHM (full width at half maximum) in image space. Figure 1 shows the results.

$$\boldsymbol{\theta}^{(k+1)} = \frac{\boldsymbol{\theta}^{(k)}}{\boldsymbol{H}^T\boldsymbol{X}^T\mathbf{1}}\boldsymbol{H}^T\boldsymbol{X}^T\frac{\boldsymbol{m}}{\boldsymbol{X}\boldsymbol{H}\boldsymbol{\theta}^{(k)}}, \tag{1}$$

where $\boldsymbol{\theta}^{(k)}$ is the reconstructed image at the k^{th} iteration,
- $\boldsymbol{H}$ optionally applies an RM kernel,
- $\boldsymbol{X}$ is the rest of the system matrix, and
- $\boldsymbol{m}$ is the sinogram data obtained by projecting $\boldsymbol{\tau}$.

$$\boldsymbol{\tau} = \boldsymbol{\theta}_{real} + a\left(\boldsymbol{\theta}_{real} - G_\sigma(\boldsymbol{\theta}_{real})\right), \tag{2}$$

where $\boldsymbol{\theta}_{real}$ is the real patient reconstruction,
- G_σ applies a gaussian smooth of $\sigma = 0.6$, and
- a is a parameter controlling sharpening amount, chosen to be 0.8.

2.2 Micro-network

Each layer j of our network encodes its input vector $\boldsymbol{\alpha}^{(j)}$ as shown in (3).

$$\boldsymbol{\alpha}_i^{(j+1)} = A(\boldsymbol{W}^{(j)}\boldsymbol{\alpha}_i^{(j)} + \boldsymbol{b}^{(j)}), \tag{3}$$

where $\boldsymbol{\alpha}_i^{(j)}$ is the i^{th} channel input for layer j, where $\boldsymbol{\alpha}^{(1)}$ represents the input volumes,
- $\boldsymbol{W}^{(j)}$is a block diagonal matrix containing the convolutional kernel weights,
- $\boldsymbol{b}^{(j)}$ are the biases, and
- A is a non-linear element-wise activation function, here chosen to be $A(x) = 1/(1+e^{-x})$.

$\boldsymbol{\alpha}^{(j)}$ represents a multi-channel tensor. In the case of the network's input, $\boldsymbol{\alpha}^{(1)}$, each channel could be a reconstructed modality volume. Here, it includes $\boldsymbol{\theta}^{(T1)}$ as well as low-count PET reconstructions $\boldsymbol{\theta}_{RM}^{(300)}$ and $\boldsymbol{\theta}^{(100)}$.

Element-wise products of $\boldsymbol{\theta}^{(T1)}$ and the PET reconstructions were also provided. This allows for modulation of the PET data by the MR intensities, thereby sharing edge information. Such an operation would normally require a dense (fully connected) layer in the network. However this would unnecessarily greatly increase the number of optimisation parameters, thus increasing computational cost and the likelihood of over-fitting on limited training data sets. Operating directly on full 3D volumes circumvents problems associated with edge effects on small patches or 2D slices; and allows the network to make use of adjacent voxel information, thus resulting in an enhanced ability to compensate for the partial volume effect (PVE) and differentiate between signal and noise.

$\boldsymbol{W}^{(j)}$ corresponds to n_j different multi-channel kernels of size $n_{j-1} \times s_j \times s_j \times s_j$, each operating on the n_{j-1}-channel input $\boldsymbol{\alpha}^{(j)}$ to produce a corresponding single-channel 3D output $\in \boldsymbol{\alpha}^{(j+1)}$. The network used here consists of three layers, with $n_j = \{9, 18, 1\}$, and $s_j = \{5, 3, 1\}$. The rationale is that the first layer performs detection of up to 9 different features, and the second recombines these feature maps in different ways to produce 18 viable clean PET volumes. The final layer performs a weighted average over these volumes.

The sensitivity of each feature-detecting kernel is controlled by the combination of $\boldsymbol{b}^{(j)}$ and A (nonlinear thresholding). A is often chosen to be rectified linear units ($ReLU$) – setting negative values to zero – which performs fast thresholding by simply discarding data. However, in the micro-network proposed here, such discarding is not desirable as it would result in minimal speed improvements at the cost of accuracy. The sigmoidal activation function chosen here ensures that information is retained as it propagates through the network.

The network therefore has comparatively few parameters ($\mathcal{O}(10^4)$). Because the number of parameters is much less than the size of the training data (even if compressed), there is no chance of

over-fitting, since the network is incapable of memorising the training data. This helps ensure that the network only performs feature recognition, as desired.

The network was trained[1] to map two low-count noise realisations of the same patient to the corresponding label $\boldsymbol{\tau}$. Two realisations were used in order to increase robustness to noise.

3 Results

For cross-validation, a single low-count dataset from the second patient (not used during the training stage) was given to the network to make a prediction. Figure 1 shows results for a central slice. The μ-net reduces the normalised root mean square error of $28\,\%$ (b) and $24\,\%$ (c) to $16\,\%$ (d).

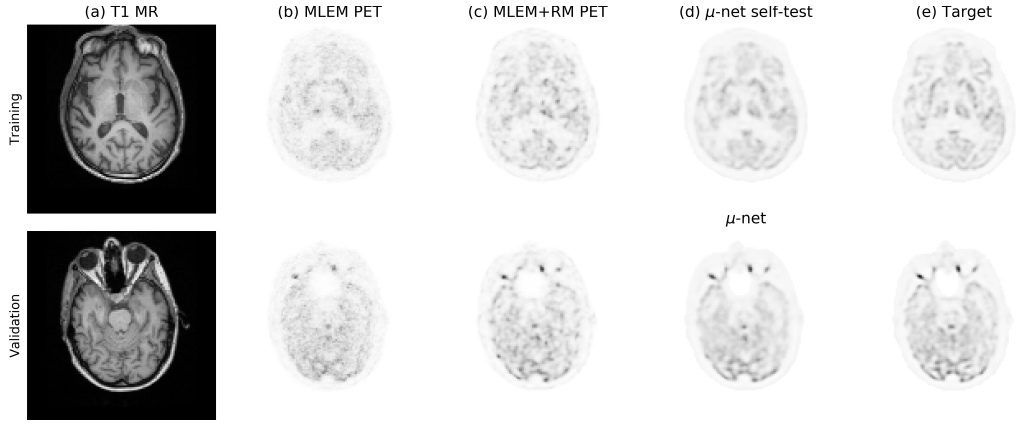

Figure 1: Central slices from training (top row) and validation (bottom row) volumes. From left to right: a) T1-weighted MR, $\boldsymbol{\theta}^{(T1)}$, b) 100 MLEM iterations $\boldsymbol{\theta}^{(100)}$, c) 300 iterations with RM $\boldsymbol{\theta}^{(300)}_{RM}$, d) CNN prediction $\boldsymbol{\alpha}^{(4)}$ (based on inputs (a)-(c)), and e) Target $\boldsymbol{\tau}$.

4 Conclusion and future work

The proposed μ-net clearly can simultaneously suppresses noise, partial volume, and ringing effects, and improve resolution with minimal training data. Future work will compare performance with alternative state-of-the-art networks and guided reconstruction methods.

References

[1] M. S. Tahaei et al., "MR-guided PET image denoising," in *2016 IEEE medical imaging conference (MIC)*, 2016, pp. 1–3 [Online]. Available: http://ieeexplore.ieee.org/document/8069564/

[2] H. Chen et al., "Low-dose CT via convolutional neural network," *Biomed. Opt. Express*, vol. 8, no. 2, pp. 679–694, Feb. 2017.

[3] L. Xiang et al., "Deep auto-context convolutional neural networks for standard-dose PET image estimation from low-dose PET/MRI," *Neurocomputing*, vol. 267, pp. 406–416, Dec. 2017.

[4] J. Xu et al., "200x Low-dose PET Reconstruction using Deep Learning," Dec. 2017 [Online]. Available: http://arxiv.org/abs/1712.04119

[5] D. L. Snyder and M. I. Miller, "The use of sieves to stabilize images produced with the em algorithm for emission tomography," *IEEE Trans Nuc. Sci.*, vol. 32, no. 5, pp. 3864–3872, Oct. 1985.

[6] D. P. Kingma and J. Ba, "Adam: A Method for Stochastic Optimization," Dec. 2014 [Online]. Available: http://arxiv.org/abs/1412.6980

[1] Trained using `Theano v0.8.2` (http://www.deeplearning.net/software/theano/) on an NVidia GeForce GTX 1070 Max-Q, using the adaptive moment estimation (Adam) optimiser [6] and an ℓ^2-norm objective.