

# Spatiotemporal PET reconstruction using ML-EM with learned diffeomorphic deformation

Ozan Öktem<sup>1</sup>, Camille Pouchol<sup>1</sup>, and Olivier Verdier<sup>1,2</sup>

<sup>1</sup> Department of Mathematics, KTH Royal Institute of Technology, 100 44 Stockholm, Sweden

<sup>2</sup> Department of Computer science, Electrical engineering and Mathematical sciences, Western Norway University of Applied Sciences, Bergen, Norway

**Abstract.** Patient movement in emission tomography deteriorates reconstruction quality because of motion blur. Gating the data improves the situation somewhat: each gate contains a movement phase which is approximately stationary. A standard method is to use only the data from a few gates, with little movement between them. However, the corresponding loss of data entails an increase of noise. Motion correction algorithms have been implemented to take into account all the gated data, but they do not scale well, especially not in 3D. We propose a novel motion correction algorithm which addresses the scalability issue. Our approach is to combine an enhanced ML-EM algorithm with deep learning based movement registration. The training is unsupervised, and with artificial data. We expect this approach to scale very well to higher resolutions and to 3D, as the overall cost of our algorithm is only marginally greater than that of a standard ML-EM algorithm. We show that we can significantly decrease the noise corresponding to a limited number of gates.

**Keywords:** Emission Tomography · Motion Correction · Deep Learning.

## 1 Introduction

Positron emission tomography (PET) is a molecular imaging technology where a radioactive tracer is administered to a patient. The tracer is an x-ray source that emits pairs of photons travelling into opposite directions, and the PET scanner is an arrangement of detectors for detecting such photon pairs (coincidence events). The goal is then to recover the spatial distribution of the tracer (activity map) from these coincidence events.

Acquiring a sufficient amount of coincidence events takes time, typically twenty to forty minutes depending on the detector efficiency and the size of the region being imaged. Organs, such as the heart and lungs, move during the PET data acquisition, so the activity map one seeks to recover in PET imaging is a spatiotemporal quantity. Failure to account for the temporal variability during reconstruction results in a deteriorated PET image.

### 1.1 Survey of existing works

Most approaches that consider motion in PET image reconstruction assume access to gated PET data. Here, PET data is subdivided into subsets where the coincidence data is from the activity map in a specific temporal state. For cardiac and respiratory motion, gated data would correspond to decomposing the entire dataset into parts that represent different breathing and/or cardiac phases. Hence, the activity associated to each gate can be assumed to be stationary, but data in the gates also suffer from a relatively low signal-to-noise ratio since they only contain a small portion of the coincidence events.

A straightforward approach based on gated data is to recover each temporal state of the activity independently of each other (frame-by-frame reconstruction). This does not account for the temporal dynamics of the activity, instead it reduces the spatiotemporal reconstruction problem into a sequence of independent stationary reconstruction problems, which in PET is done by ML-EM [21] (or a variant thereof, like OSEM [14]). Spatiotemporal reconstruction refers to methods that instead take the temporal dynamics into account. Several approaches have been proposed where most rely on estimating a motion model prior to reconstruction, see [8, 10, 18, 19] for survey.

In this paper, the proposed method falls into the family of algorithms that, contrarily to those based on a priori built motion models, jointly estimate the image and motion, directly from the full set of measured data. An objective function is optimised with respect to two arguments: image and motion. Hence, only one image with the full statistic is reconstructed. Given the close relationship between the image reconstruction and motion estimation steps, a simultaneous method of estimating the two is better able to reduce motion blur and compensate for poor signal-to-noise ratios and to improve the accuracy of the estimated motion [11, 12].

In the latter works, one performs a two-step minimisation of a joint energy functional term (which includes both image likelihood and motion-matching terms). The method chosen by Jacobson and Fessler [15, 16], referred to as *joint estimation with deformation modelling*, is based on maximising the likelihood for a parametric Poisson model for gated PET measurements. Motion (from gate to gate) is defined by a set of deformation parameters. A similar motion-aware likelihood function was used by Blume and colleagues [5], although using a distinct optimisation scheme and depicting more convincing results. In this context one may also consult [23], which compares three approaches for joint reconstruction of image and motion.

An alternative is to consider motion models derived from deformations modelled by diffeomorphisms, as obtained from example through the LDDMM framework [22]. Here, one can calculate regularising functionals that incorporate such deformations. Finally, [9] provides an overview of variational shape models as applied to the registration and segmentation problems. These could also be coupled with variational regularisation methods for image reconstruction.

The main drawback of all these methods, however, is the relatively high computational costs involved in the joint reconstruction approach.

## 1.2 Proposed method

In this paper, we develop a joint reconstruction method based on the minimisation of a suitable functional. The main novelty of our work is the scalability of the resulting algorithm, as its complexity is of the order of the usual ML-EM algorithm. Images are indeed estimated using a generalised ML-EM algorithm. Motion estimation, with deformations modelled by diffeomorphisms, is based on the unsupervised deep learning framework `voxelmorph` [6]. That is, we make use of a pre-trained neural network which performs direct image registration, i.e., the network finds a diffeomorphism which, given two images, deforms the first one into the second.

Interestingly, one single outer iteration of our algorithm is close to the recently proposed approach [17]. Thus, it generalises the previous work and shows that it can be interpreted in the framework of an optimisation problem.

The results of the proposed method are tested on the Derenzo phantom, and shown to recover a significant part of the information lost when one uses gate-by-gate reconstruction.

## 2 Methods

### 2.1 Mathematical background

**ML-EM algorithm** [21]. Let us consider the statistical model

$$g \sim \text{Poisson}(Af),$$

where  $f$  is the unknown image, and  $g$  is the acquired data—a vector of  $\mathbb{R}^d$ ; this models the physics of *stationary* PET with forward operator  $A$ .

The ML-EM algorithm solves the corresponding *maximum likelihood problem*, which amounts to minimising the divergence  $d_{KL}(g||Af)$ , defined for two non-negative vectors  $u, v$  in  $\mathbb{R}^d$  by

$$d_{KL}(u||v) := \sum_{j=1}^d \left( v_j - u_j - u_j \log \left( \frac{u_j}{v_j} \right) \right).$$

The ML-EM algorithm is an iterative solver with update

$$f^{(n+1)} := \frac{f^{(n)}}{A^T \mathbf{1}} A^T \left( \frac{g}{Af^{(n)}} \right), \quad (1)$$

starting from an initial guess  $f^{(0)}$ , usually  $f^{(0)} = 1$ .

**Diffeomorphisms acting on images.** Viewing images as elements of  $X := L^2(\Omega)$ , i.e., square-integrable functions on a compact  $\Omega \subset \mathbb{R}^p$  with  $p = 2$  or  $p = 3$ , we model motion as an appropriate *group action* of diffeomorphisms onto

images. In this paper, given a diffeomorphism  $\psi: \Omega \rightarrow \Omega$ , we will use the specific definition  $\mathcal{W}_\psi: X \mapsto X$  as the *intensity-preserving* action

$$\mathcal{W}_\psi f(x) := f(\psi^{-1}(x)).$$

Note that our approach is, however, general, and we could have used the *mass-preserving* action instead, namely

$$\widetilde{\mathcal{W}}_\psi f(x) := |D\psi^{-1}(x)|f(\psi^{-1}(x)). \quad (2)$$

We will parameterise diffeomorphisms by exponentials of (stationary) vector fields, i.e.,  $\psi = \exp(v)$ , where the exponential  $\exp(v)$  of a vector field  $v$  is defined as  $\psi(1, \cdot)$ , where  $\psi(t, \cdot)$  solves the differential equation  $\frac{\partial \psi}{\partial t}(t, \cdot) = v(\psi(t, \cdot))$ , with initial condition  $\psi(0, \cdot) = \text{Id}$ .

**Image registration.** The (direct) image registration problem consists in deforming a *template*  $f_1$  into a *target*  $f_2$ , i.e., finding a diffeomorphism  $\psi$  such that  $\mathcal{W}_\psi f_1 \approx f_2$ . This is usually done by minimising a functional of the form

$$\arg \min_{\psi} d_2(f_2, \mathcal{W}_\psi f_1) + \lambda \mathcal{R}(\psi), \quad (3)$$

where  $d_2$  is the  $L^2$ -distance on  $X$ ,  $\mathcal{R}$  is a regularisation term on diffeomorphisms that is discussed in [subsection 2.3](#), and  $\lambda$  is a regularisation parameter.

## 2.2 General approach

**Modelling.** We are given *gated* data in  $N + 1$  different gates, corrupted by Poisson noise. For  $g_i$  denoting the data,  $f_i$  the images in each gate and  $A$  the forward operator, we thus assume

$$g_i \sim \text{Poisson}(Af_i), \quad i = 0, \dots, N.$$

We also assume that for  $i = 1, \dots, N$ , two consecutive images  $f_{i-1}$  and  $f_i$  are related by the statistical model

$$f_i = \mathcal{W}_{\psi_i} f_{i-1} + e_i,$$

where  $\psi_i: \Omega \rightarrow \Omega$  is the exponential of a vector field following a given probability law (see [\(8\)](#)) and  $e_i$  is a  $X$ -valued random variable.

**Variational problem.** We now define the variational problem associated to the inverse problem of finding both the images  $f_i$  and diffeomorphisms  $\psi_i$  from the data  $g_i$ . It reads

$$\arg \min_{(f_i), (\psi_i)} J(f_0, \dots, f_N, \psi_1, \dots, \psi_N), \quad (4)$$

where

$$J(f_i, \psi_i) := \sum_{i=0}^N d_{KL}(g_i || Af_i) + \sum_{i=1}^N \left( d_2(f_i, \mathcal{W}_{\psi_i} f_{i-1}) + \lambda \mathcal{R}(\psi_i) \right).$$

**General algorithm.** We solve the variational problem (4) by an *intertwined* method, which consists in alternating between estimating the diffeomorphisms (the *motion estimation step*), and the images  $f_i$  (the *reconstruction step*).

The images are first initialised by solving the maximum likelihood problem  $\arg \min_{f_i} (d_{KL}(g_i || Af_i))$ , associated to  $g_i = \text{Poisson}(Af_i)$  in each gate. This is done by the algorithm ML-EM (1), yielding estimates  $f_i^0$ ,  $i = 0, \dots, N$ .

For a given estimate of images  $f_i^k$ , the motion estimation part consists in solving

$$\arg \min_{(\psi_i)} \sum_{i=1}^N (d_2(f_i^k, \mathcal{W}_{\psi_i} f_{i-1}^k) + \lambda \mathcal{R}(\psi_i)),$$

which in turn can be decomposed into  $N$  problems of the form

$$\arg \min_{\psi_i} d_2(f_i^k, \mathcal{W}_{\psi_i} f_{i-1}^k) + \lambda \mathcal{R}(\psi_i), \quad i = 1, \dots, N. \quad (5)$$

Note that each of these becomes an *image registration problem*, as we are looking for a diffeomorphism  $\psi_i^{k+1}$  matching the template  $f_{i-1}^k$  against the target  $f_i^k$ .

For the reconstruction part, we assume  $f_i^k \approx \mathcal{W}_{\psi_i^{k+1}} f_{i-1}^k$  for  $i = 1, \dots, N$  and neglect the  $N$  corresponding  $d_2$  terms. The minimisation problem thus becomes

$$\arg \min_{(f_i)} \sum_{i=0}^N d_{KL}(g_i || Af_i).$$

We then focus on a particular gate, say the zero'th gate, and use  $f_i^k \approx \mathcal{W}_{\psi_i^{k+1}} f_i^k$  to obtain the optimisation problem with  $f_0$  as the only variable:

$$\arg \min_{f_0} \sum_{i=0}^N d_{KL}(g_i || A_i f_0). \quad (6)$$

where

$$A_i := A \mathcal{W}_{\phi_i} \quad (7)$$

and we have used the notation  $\phi_i := \psi_i \circ \dots \circ \psi_1$  for  $i = 1, \dots, N$ . Solving the above yields a next estimate  $f_0^{k+1}$  for  $f_0$ . All the images  $f_i^{k+1}$  are then obtained by  $f_i^{k+1} = \mathcal{W}_{\psi_i^{k+1}} f_{i-1}^{k+1}$ ,  $i = 1, \dots, N$ .

It now only remains to explain how the optimisation problems (5) and (6) are solved, which is the topic of the next subsections.

### 2.3 Motion estimation

The motion estimation problem (5), can be rewritten for two generic images  $f_1$  and  $f_2$  as

$$\arg \min_v d_2(f_2, \mathcal{W}_{\exp(v)} f_1) + \lambda \mathcal{R}(v), \quad (8)$$

where we parameterise the diffeomorphisms by exponentials of stationary vector fields  $v$ .

To solve this direct image registration problem, we use the `voxelmorph` unsupervised deep learning approach [6], where a neural network parameterises a function  $(f_1, f_2) \mapsto v$ . That neural network is itself based on the network architecture `Unet` [20]. We keep the architecture of `voxelmorph`, with the same hyperparameters and specific regularisation functional  $\mathcal{R}$  given in [6]. Once trained, the network produces a mapping matching any two images  $f_1, f_2$ , which we denote

$$\gamma(f_1, f_2) := \exp(v(f_1, f_2)). \quad (9)$$

**Training.** In [6], the network is trained on tuples of images  $(f_1, f_2)$  coming from brain MRI scans. We use instead *synthetic* data: tuples of images  $(f_1, f_2)$  generated on the fly.

We generate training images as follows. A random image  $f_1$  consists of a Poisson random number of ellipsoids [3, 4]. The centre of each ellipsoid has uniform distribution inside the central part of the domain  $\Omega$ , the principal axes have exponential distribution, and the orientation follows a uniform distribution. We apply a mask vanishing at the boundary to avoid boundary effects. when diffeomorphisms are applied. We generate random vector fields  $v$  using a Gaussian random field with radial basis function kernel, with appropriate scale and typical size. The training image  $f_2$  is then  $f_2 = \mathcal{W}_{\exp(v)} f_1$ . We show in Figure 1 a sample of images  $f_1, f_2$  and vector field  $v$  generated as above.

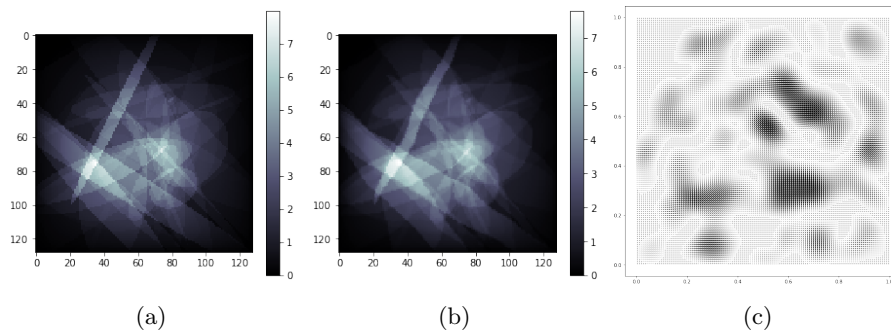


Fig. 1: Example of a 2D synthetic tuple of images  $f_1$  (a) and  $f_2$  (b), related by  $f_2 = \mathcal{W}_{\exp(v)} f_1$  for the intensity-preserving action, with  $v$  plotted in (c).

## 2.4 Reconstruction

We now focus on the reconstruction problem (6) which we solve using a reformulation of ML-EM. Given operators  $A_i$ , we can simply write ML-EM for the compound operator  $A = (A_0, \dots, A_N)$  which yields

$$f_0^{(n+1)} = \frac{f_0^{(n)}}{\sum_{i=0}^N A_i^T 1} \sum_{i=0}^N A_i^T \left( \frac{g_i}{A_i f_0^{(n)}} \right), \quad (10)$$

for an initial guess  $f_0^{(0)}$ . We call this algorithm ‘‘M-ML-EM’’ to avoid the confusion with the vanilla ML-EM algorithm (1). We use this algorithm with  $A_i$  defined in (7). Note that this algorithm has been used in [13] for the particular case of the intensity-preserving action. The computation of  $A_i^T$  requires the computation of  $W_{\phi_i}^T$ . We achieve this by using the identity  $W_{\phi}^T = \widetilde{W}_{\phi^{-1}}$  valid for any diffeomorphism  $\phi$ , where  $\widetilde{W}$  denotes the mass-preserving action (2).

## 2.5 Full algorithm

We summarise the algorithm with all the necessary details in [Algorithm 1](#).

---

### Algorithm 1 Full Algorithm

---

Choose the outer number of iterates  $n_{\text{outer}}$ , the inner number of iterates  $n_{\text{inner}}$  for M-ML-EM, and  $n_{\text{init}}$ , the number of iterates for vanilla ML-EM.

```

for  $i \leftarrow 0, \dots, N$  do
     $f_i \leftarrow \text{ML-EM}(A, g_i, n_{\text{init}})$  ▷ Iterates of (1)
end for
for  $k \leftarrow 1, \dots, n_{\text{outer}}$  do
    for  $i \leftarrow 1, \dots, N$  do
         $\psi_i \leftarrow \gamma(f_{i-1}, f_i)$  ▷ Network registration (9)
    end for
     $W_0 \leftarrow \text{Id}$ 
    for  $i \leftarrow 1, \dots, N$  do
         $W_i \leftarrow W_{\psi_i} W_{i-1}$ 
         $A_i \leftarrow A W_i$ 
    end for
     $A_0 \leftarrow A$ 
     $f_0 = \text{M-ML-EM}(\{A_j\}_{j=0, \dots, N}, \{g_j\}_{j=0, \dots, N}, n_{\text{inner}})$  ▷ Iterates of (10)
    for  $i \leftarrow 1, \dots, N$  do
         $f_i \leftarrow W_i f_0$ 
    end for
end for

```

The outcome is  $f_0$ .

---

## 2.6 Complexity

Evaluating vector fields with the network is negligible when compared to ML-EM or M-ML-EM iterations. Each iteration is itself controlled by the time  $t$  required to compute an expression of the form  $A^T(\frac{g}{Af})$ . Since M-ML-EM sums these quantities  $N$  times, an iteration of it is of the order of  $N \times t$ . Note that evaluating the denominator in (10) (which involves sums of  $A^T 1$ ) does not take more time than evaluating the denominator in ML-EM since  $A^T 1$  can be computed off-line.

### 3 Results

#### 3.1 Derenzo phantom

We present experiments with the Derenzo phantom, with image size  $192 \times 192$ . Although this phantom is made of ellipses, we stress that they are very different from the data used to train the network, compare [Figure 1](#) and [Figure 2a](#).

This phantom is then deformed successively with the intensity preserving action by exponentials of vector fields, where each vector field is drawn from the same distribution used to train the network. For the experiments, we use  $N = 3$ , which amounts to four gates, and we want to recover the image in the initial gate. The resulting four phantoms are presented in [Figure 2](#).

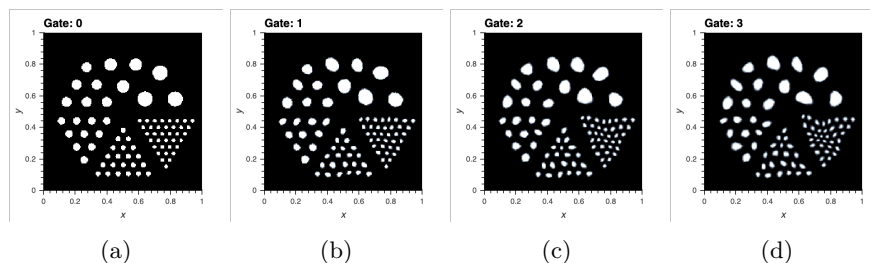


Fig. 2: Derenzo phantom in four different gates.

The forward operator  $A$  is a 2D PET operator with 108 angles (views) and 250 tangential positions. The noisy data is  $\text{Poisson}(A(tf))$  for each image  $f$ , where  $t$  is the acquisition time and thus controls the noise level.

For the phantoms in [Figure 2](#), we choose  $t = 60$ . This noise level gives rise to typical optimal numbers of iterates for ML-EM which are of the same order of magnitude as the ones in clinic applications. Note that all images are multiplied by the same time factor, which amounts to assuming that acquisition time is roughly the same in each gate.

#### 3.2 Methods without motion correction

We compare our method with two simple reconstruction methods (simple because without motion correction) for images with gated data:

- Either one aggregates the whole data and reconstructs from ML-EM, leading to blurry results because of the movement.
- Or one tries and limit blur by focusing on one gate (say the first) and reconstructing only from that. Since there is less data, the result is noisier.

In order to quantitatively compare these strategies, we use ML-EM for the data obtained from taking gate zero only, aggregating gates zero and one, and

so on up until aggregating all the four gates. Finally, we can also estimate the best reconstruction one could hope for, that is, if there were no movement. This amounts to acquiring the phantom in the 0th gate four times longer.

The results are given in [Figure 3](#), where the PSNR between the estimated image and the real image in gate zero is computed at each iteration.

The results show that aggregating the gates progressively induces a drop in image quality, as measured by the PSNR. Compared to gate zero acquired four times longer, the best possible achievable gain is about 2.2 dB.

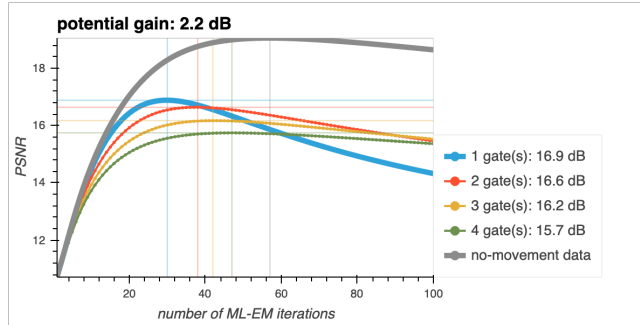


Fig. 3: PSNR for different ML-EM strategies without motion correction, and comparison with "no-movement" data, reconstructing from the initial gate acquired four times longer.

### 3.3 Proposed method

We apply [Algorithm 1](#) to the data above. It turns out that a single outer iteration is responsible for most of the improvement, so we focus on that case for presenting experiments. In other words:

1. we initialise by running some ML-EM iterations in each gate,
2. we then match the resulting images to estimate the diffeomorphisms,
3. we finally run some M-ML-EM iterations.

We plot the PSNR between the image reconstructed (in the initial gate zero) and the real image, for a given number `em_iter` of ML-EM iteration followed by a given number `diff_iter` of M-ML-EM iterations. These results are presented in [Figure 4](#).

We find that the optimal strategy is to iterate only a few times (six iterations in this specific experiment) with ML-EM before estimating the diffeomorphisms through M-ML-EM (42 iterations in this specific experiment). Note that this yields a total of 48 iterations which is higher than the 29 ML-EM iterations which would be optimal for reconstructing from the gate zero.

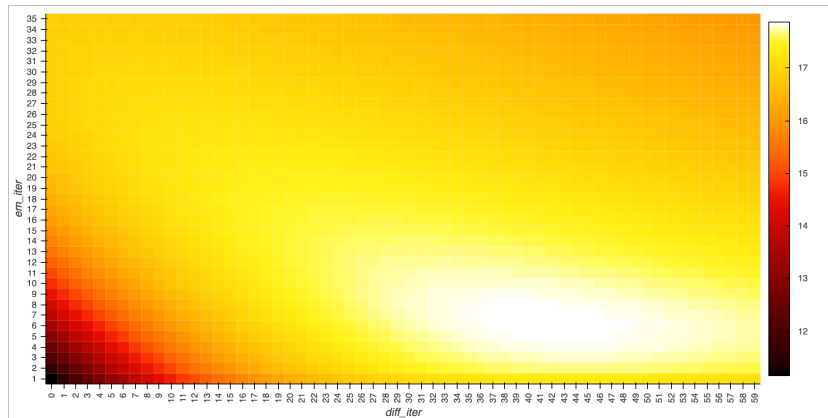


Fig. 4: PSNR for various choices of number `em_iter` of initial ML-EM iterates and number `diff_iter` of M-ML-EM iterates.

The gain in PSNR is 1.0 dB, which makes up for about 46 % of the maximal gain of 2.2 dB. Reconstructions obtained from the optimal uncorrected ( $n = 29$  iterations of ML-EM are used on gate zero) and the proposed method with the optimal number of iterations of ML-EM and M-ML-EM are presented in Figure 5. The proposed method seems to give smoother results. The smaller discs towards the middle of the image are also better seen.

We also emphasise that these results (improvement in PSNR and optimal number of iterations) are extremely robust with respect to the randomness involved in the experiments, namely the vector fields drawn randomly as well as the Poisson noise.

### 3.4 Implementation Details

All computations are run in Python and use Operator Discretization Library (`odl`) for manipulating operators [2], `neuron` for warping utilities [7], which itself uses `tensorflow` [1]. The training was performed with `voxelmorph` [6].

## 4 Perspectives

This paper presents a new method for joint motion estimation and image reconstruction in PET. Its main advantage is its cost, similar to that of the usual ML-EM algorithm, making it scalable to clinical 4D data.

Our framework also allows for further modelling such as attenuation correction. In a future work, we consider testing this method with clinical data. This would require training the network on appropriate datasets. We also plan to generalise the approach to other group actions, such as the mass-preserving one, which is more physically relevant.

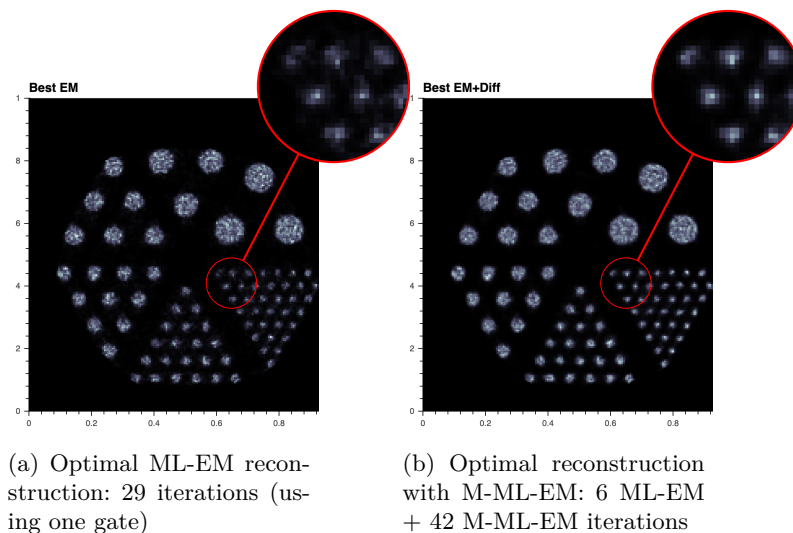


Fig. 5: Optimal reconstructions of the gate zero (measured in PSNR).

## Acknowledgements

We acknowledge support from the Swedish Foundation of Strategic Research grant AM13-004.

## References

1. Abadi, M., et al.: TensorFlow: Large-scale machine learning on heterogeneous systems (2015), <http://tensorflow.org/>, software available from tensorflow.org
2. Adler, J., Kohr, H., Öktem, O.: ODL-a Python framework for rapid prototyping in inverse problems. Royal Institute of Technology (2017)
3. Adler, J., Öktem, O.: Solving ill-posed inverse problems using iterative deep neural networks. *Inverse Problems* **33**(12), 124007 (2017)
4. Adler, J., Öktem, O.: Learned primal-dual reconstruction. *IEEE transactions on medical imaging* **37**(6), 1322–1332 (2018)
5. Blume, M., Martinez-Moller, A., Keil, A., Navab, N., Rafecas, M.: Joint reconstruction of image and motion in gated positron emission tomography. *IEEE Transactions on Medical Imaging* **29**(11), 1892–1906 (2010)
6. Dalca, A.V., Balakrishnan, G., Gutttag, J., Sabuncu, M.R.: Unsupervised learning for fast probabilistic diffeomorphic registration. In: *International Conference on Medical Image Computing and Computer-Assisted Intervention*. pp. 729–738. Springer (2018)
7. Dalca, A.V., Gutttag, J., Sabuncu, M.R.: Anatomical priors in convolutional networks for unsupervised biomedical segmentation. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. pp. 9290–9299 (2018)
8. Dawood, M., Jiang, X., Schäfers, K.P. (eds.): *Correction Techniques in Emission Tomography*. Series in Medical Physics and Biomedical Engineering, CRC Press (2008)

9. Farag, A.A., Shalaby, A., El Munim, H.A., Farag, A.: Variational shape representation for modeling, elastic registration and segmentation. In: Li, S., Tavares, J.M.T.S. (eds.) *Shape Analysis in Medical Image Analysis, Lecture Notes in Computational Vision and Biomechanics*, vol. 14, chap. 95–122. Springer-Verlag (2014)
10. Gigengack, F., Jiang, X., Dawood, M., Schäfers, K.P.: *Motion Correction in Thoracic Positron Emission Tomography*. Springer-Verlag (2015)
11. Gilland, D.R., Mair, B.A., Bowsher, J.E., Jaszczak, R.J.: Simultaneous reconstruction and motion estimation for gated cardiac ECT. *IEEE Transactions in Nuclear Science* **49**(5), 2344–2349 (2002)
12. Gravier, E., Yang, Y., King, M.A., Jin, M.: Fully 4D motion-compensated reconstruction of cardiac SPECT images. *Physics in Medicine and Biology* **51**(18), 4603–4619. (2006)
13. Hinkle, J., Szegedi, M., Wang, B., Salter, B., Joshi, S.: 4d ct image reconstruction with diffeomorphic motion model. *Medical image analysis* **16**(6), 1307–1316 (2012)
14. Hudson, H.M., Larkin, R.S.: Accelerated image reconstruction using ordered subsets of projection data. *IEEE transactions on medical imaging* **13**(4), 601–609 (1994)
15. Jacobson, M.W., Fessler, J.A.: Joint estimation of image and deformation parameters in motion-corrected PET. In: 2003 IEEE Nuclear Science Symposium Conference Record. pp. 3290–3294 (2003)
16. Jacobson, M.W., Fessler, J.A.: Joint estimation of respiratory motion and activity in 4D PET using CT side information. In: 3rd IEEE International Symposium on Biomedical Imaging: Nano to Macro. Arlington, VA, April 6–9, 2006. pp. 275–278 (2006)
17. Li, T., Zhang, M., Qi, W., Asma, E., Qi, J.: Motion correction of respiratory-gated pet image using deep learning based image registration framework. In: 15th International Meeting on Fully Three-Dimensional Image Reconstruction in Radiology and Nuclear Medicine. vol. 11072, p. 110720Q. International Society for Optics and Photonics (2019)
18. Rahmim, A., Tang, J., Zaidi, H.: Four-dimensional image reconstruction strategies in cardiac-gated and respiratory-gated PET imaging. *PET Clinics* **8**(1), 51–67 (2013)
19. Reader, A.J., Verhaeghe, J.: 4D image reconstruction for emission tomography. *Physics in Medicine and Biology* **59**(22), R371–R418 (2014)
20. Ronneberger, O., Fischer, P., Brox, T.: U-net: Convolutional networks for biomedical image segmentation. In: *International Conference on Medical image computing and computer-assisted intervention*. pp. 234–241. Springer (2015)
21. Shepp, L.A., Vardi, Y.: Maximum likelihood reconstruction for emission tomography. *IEEE transactions on medical imaging* **1**(2), 113–122 (1982)
22. Younes, L.: *Shapes and Diffeomorphisms, Applied Mathematical Sciences*, vol. 171. Springer-Verlag (2010)
23. Zhang, Y., Ghodrati, A., Brooks, D.H.: An analytical comparison of three spatio-temporal regularization methods for dynamic linear inverse problems in a common statistical framework. *Inverse Problems* **21**(1), 357–382 (2005)