Published in final edited form as:
Med Image Comput Comput Assist Interv. 2022 September ; 13434: 194–203. doi:10.1007/978-3-031-16440-8_19.

Supervised Deep Learning for Head Motion Correction in PET

Tianyi Zeng[1], **Jiazhen Zhang**[1], **Enette Revilla**[5], **Eléonore V. Lieffrig**[1], **Xi Fang**[2], **Yihuan Lu**[6], **John A. Onofrey**[1,3,4]

[1]Department of Radiology & Biomedical Imaging, Yale University, New Haven, CT, USA

[2]Department of Psychiatry, Yale University, New Haven, CT, USA

[3]Department of Urology, Yale University, New Haven, CT, USA

[4]Department of Biomedical Engineering, Yale University, New Haven, CT, USA

[5]University of California, Davis, CA, USA

[6]United Imaging Healthcare, Shanghai, China

Abstract

Head movement is a major limitation in brain positron emission tomography (PET) imaging, which results in image artifacts and quantification errors. Head motion correction plays a critical role in quantitative image analysis and diagnosis of nervous system diseases. However, to date, there is no approach that can track head motion continuously without using an external device. Here, we develop a deep learning-based algorithm to predict rigid motion for brain PET by lever-aging existing dynamic PET scans with gold-standard motion measurements from external Polaris Vicra tracking. We propose a novel Deep Learning for Head Motion Correction (DL-HMC) methodology that consists of three components: (i) PET input data encoder layers; (ii) regression layers to estimate the six rigid motion transformation parameters; and (iii) feature-wise transformation (FWT) layers to condition the network to tracer time-activity. The input of DL-HMC is sampled pairs of one-second 3D cloud representations of the PET data and the output is the prediction of six rigid transformation motion parameters. We trained this network in a supervised manner using the Vicra motion tracking information as gold-standard. We quantitatively evaluate DL-HMC by comparing to gold-standard Vicra measurements and qualitatively evaluate the reconstructed images as well as perform region of interest standard uptake value (SUV) measurements. An algorithm ablation study was performed to determine the contributions of each of our DL-HMC design choices to network performance. Our results demonstrate accurate motion prediction performance for brain PET using a data-driven registration approach without external motion tracking hardware. All code is publicly available on GitHub: https://github.com/OnofreyLab/dl-hmc_miccai2022.

Keywords

deep learning; supervised learning; data-driven motion correction; image registration; brain; PET

tianyi.zeng@yale.edu .

1 Introduction

Positron emission tomography (PET) allows clinicians and researchers to study physiological or pathological processes in humans, and in particular the brain [10,15]. However, patient movement during scanning presents a challenge for accurate PET image reconstruction and subsequent quantitative analysis [9]. Head motion during brain PET scans reduces image resolution (sharpness), lowers concentrations in high-uptake regions, and causes mis-estimation in tracer kinetic modeling. Even small magnitude head motion may have a large impact on brain PET quantification. The long duration of PET studies exacerbates this problem, where involuntary movements of the patient are unavoidable and the average head motion can vary from 7 mm [1] in clinical scans to triple this amount for longer research scans. Therefore, the ability to track and correct head motion is critical in PET studies.

For motion correction, a straightforward approach is physical head restraint [5]. However, it does not provide correction once motion occurs and it reduces the level of patient comfort, especially for long research scans. Post-reconstruction registration, the most commonly used approach, cannot correct for motion within one scan period. An alternative method, multi-acquisition-frame [14] divides scan frames at times of motion but cannot correct for frequent motion due to low count statistics in one frame. Neither of these methods can perform motion correction in real-time. To date, the most accurate approach is event-by-event correction using motion information measured by a hardware-based motion tracking (HMT) system such as Polaris Vicra [8]. However, HMT is not generally accepted in clinical use since it usually requires attaching a tracking device to the patient and additional setup time. Frames with inaccurate motion estimates can be excluded, but this increases image noise by discarding data. For HMT, slippage of the attached markers can happen due to non-rigid fixation that can be affected by hair style. Other systems like markerless motion tracking [11] are still under development and have not been validated for PET use. On the other hand, data-driven motion correction methods that are based on PET raw data do not suffer these problems and have been developed and applied in clinical research [13]. Therefore, it is appealing to develop an approach that can perform accurate and robust head motion tracking and estimation based only on PET raw data in real-time during the scan.

Image registration methods [17] that seek to align two or more images offer a data-driven solution for correcting brain motion. While some registration methods use hand-crafted features instead of raw intensities [17], deep learning (DL) techniques, in which neural networks build a hierarchical representation of the data using multiple layers of hidden units [12], allow for the registration methods to learn the features of interest directly from the data. Deep learning methods are of interest because they may be less susceptible to local optima, and they offer highly parallelized implementations conducive to real-time applications.

In this study, we developed a deep learning-based method capable of real-time head motion tracking during brain PET imaging in order to perform rigid motion correction without the aid of external devices. We train our deep learning head motion correction (DL-HMC) network in a supervised manner with one-second 3D point clouds back-projected from clinical patient PET listmode data and use Vicra as gold standard motion estimates. We validate our method in both single subject and multiple subject experiments, and quantitatively compared the synthetic motion information with Vicra gold-standard as well as through qualitative reconstruction image ROI evaluation. We also performed an algorithm ablation study to determine the contributions of each of our strategies.

2 Methods

2.1 Data

We identified 25 ^{18}F-FDG PET scans from a database of brain PET scans acquired on a brain-dedicated Siemens HRRT scanner at the Yale PET Center. The 25 subject group constitutes a diverse patient population that includes 8 cognitive normals, 11 subjects suffering from cocaine dependence, and 6 with cognitive disease. The mean injected activity for these 25 patients is 4.95±0.14 mCi. Eight points forming the vertices of a 10-cm cube centered in the scanner FOV were chosen to describe the motion of the brain, and inter-frame motion was computed by averaging twice the standard deviation of the motion of each of these 8 points [7]. The overall motion of the brain throughout the entire scan (mean±SD) was 12.07±7.12 mm. In addition to the list-mode data, other materials were available such as Polaris Vicra motion tracking information used as motion gold-standard (Sec. 1), T1-weighted MR images and PET-space to MR-space transformation matrices. All PET imaging data is 30 minutes acquired 60 minutes post injection.

2.2 Motion Correction Network Structure

Our DL-HMC network architecture (Fig. 1) consists of the following components: (i) two feature extractor blocks to encode the PET input data, which consists of a reference image I_{ref} and moving image I_{mov}; (ii) a regression block to estimate the six rigid transformation motion parameters; and (iii) feature-wise transformation (FWT) layers [3] to condition the network to tracer time-activity. The proposed network architecture uses DenseNet [6] as a feature extractor with shared weights for two input images and concatenates the output features into a single vector. The encoders effectively reduce the 3D image data volumes down to a vector of size 128. The FWT takes the relative difference in time $\Delta t = t_{\mathrm{mov}} - t_{\mathrm{ref}}$ (in seconds) between the reference and moving images, respectively, as input to a fully connected network and then multiplies the concatenated feature output to this result, which conditions the network to dynamic changes in the PET tracer over time. Finally, the conditioned features are fed into the fully connected regression block to predict the translation and rotation components of the motion $\theta = [t_x, t_y, t_z, r_x, r_y, r_z]$. The network has 11,492,102 trainable parameters.

The input for the network consists of one-second 3D cloud representations of the PET data. The 3D clouds were created by back-projecting the PET listmode data along the line-of-response (LOR) with normalization for scanner sensitivity using MATLAB. We pre-process the 3D cloud data volumes by smoothing and down sampling from 256×256×207 voxels (1.22×1.22×1.23 mm^3 voxel spacing) to 32×32×32 voxels (9.76×9.76×7.96 mm^3 voxel spacing), which reduces the data memory footprint, increases computational efficiency, and removes image noise. Each one-second 3D cloud has corresponding Vicra motion tracking system (rigid transformation matrix) as gold-standard motion information. We implemented the network in Python (version 3.8) using Py-Torch (version 1.7.1) and MONAI (version 0.8). All code is available on GitHub: http://github.com/OnofreyLab/dl-hmc_miccai2022/

2.3 Network Training Strategy

We perform supervised learning of the DL-HMC network using two input 3D point clouds from two different time points t_{ref} and t_{mov} to predict the relative rigid motion transformation with respect to the Vicra gold-standard. Due to the large number of one-second data input pairs available, we developed an efficient data sub-sampling strategy for model training. From each subject in our training set, we select n_{ref} reference time points from a uniform random distribution over the PET scan duration. For each reference time, we then sample n_{mov} different moving image times using a *uniform sampling strategy*, where moving image times are randomly sampled from a uniform distribution such that $t_{\text{mov}} > t_{\text{ref}}$. For n_{sub} training subjects, this process generates $n_{\text{sub}} \times n_{\text{ref}} \times n_{\text{mov}}$ unique training pairs. In both sampling procedures, we calculate the relative motion transformation matrix from the Vicra data and we record the time difference Δt between the reference and moving times. We train the network by minimizing the network's mean square error (MSE) between the predicted motion estimate $\hat{\theta}$ and Vicra θ using Adam optimization with initial learning rate 5e-4, γ=0.98, and exponential decay with step size 200. Because of GPU and CPU memory constraints, a smart caching dataset was used to replace 25% of the data (1,024 samples) for each epoch with new samples.

2.4 Motion Correction Inference

Using the trained DL-HMC model, we perform motion correction by using a single, fixed reference image rather than computing the relative motion between all two consecutive time points. This approach to using a single reference image avoids accumulation of any errors in the motion correction prediction. In this case, we select the first image time point t_{ref}=3,600 (60 minutes post injection) as the reference image and predict the motion from this reference to all subsequent one second image frames in the next 30 minutes (1,800 one-second time points).

3 Results

Due to limitations caused by large amounts of data, hardware memory constraints, and long training times, we performed initial model development using data from a single subject (Sec. 3.1). The results of these pilot experiments informed our model design decisions (Sec. 2.2) as well as training strategy (Sec. 2.3), which we then applied to our multi-subject DL-HMC model (Sec. 3.2).

For all experiments, we quantitatively and qualitatively evaluate motion correction performance. For quantitative assessment, we calculate the MSE between the Vicra gold-standard θ and DL-HMC prediction $\hat{\theta}$ (these are unitless error measurements because they combine translation (mm) and rotation components (degrees)). For qualitative assessment, we reconstruct the PET image using the predicted motion correction information for the whole sequence using Motion-compensation OSEM List-mode Algorithm for Resolution-

Recovery Reconstruction (MOLAR) [2] and compare to the Vicra reconstructed image. The same reconstruction parameters were applied using the different DL-HMC and Vicra motion estimates. For a more comprehensive quantitative analysis of the reconstructed PET images, each subject's co-registered MR image was segmented into 109 regions using FreeSurfer [4], which were then merged and into twelve gray matter brain regions of interest (ROIs) and analysed by calculating standard uptake value (SUV). All computations were performed on a server with Intel Xeon Gold 5218 processors, 256 GB RAM, and an NVIDIA Quadro RTX 8000 GPU (48 GB RAM).

3.1 Single Subject Pilot Experiments

To evaluate the feasibility of the proposed method to accurately predict head motion throughout the PET scan duration, we first applied the DL-HMC in single subject (n_{sub}=1) experiments and perform a rigorous ablation study to determine the contributions of each of our DL-HMC design choices to network performance. For pilot experiments, we split the subject's entire PET scan time course (1,800 one-second images) into three subsets corresponding to 80/10/10% of the data for training/validation/testing. While using data in this manner from the same subject introduces bias, we utilize a single reference point (t_{ref} = 3, 600) and register all t_{mov} in the testing set to this image to calculate experimental error independent of the training data, which provides an informative measure of algorithm performance. From the training set, we randomly selected n_{ref}=1,440 reference time points. DL-HMC ablation experiments (Table 1) consisted of the following design choices: (i) using a different number moving image time points, n_{mov}=1 (1,440 total samples) or n_{mov}=10 (14,400 total samples) (see Sec. 2.3); (ii) the inclusion of time information Δt using the FWT layers; (iii) increased depth of the image encoder by changing the DenseNet growth rate from 4 to 32; and (iv) different data sampling strategies, where we compare our *uniform sampling* strategy to a *normal sampling* with moving image times randomly sampled with respect to a normal distribution (right tail) with FWHM equal to 60 seconds. Our DL-HMC results demonstrated that using a large training dataset (n_{mov}=10), FWTs, deep image encoders, and uniform random data sampling provided accurate motion prediction performance with MSE (mean±SD) 0.035±0.073 in the test set. Typical training for one subject with 14,400 training data samples required 10,000 epochs for convergence. Fig. 2a shows DL-HMC motion correction predictions nearly identical to Vicra. Reconstructed PET images also appear qualitatively similar between DL-HMC and Vicra (Fig. S1) and both demonstrate similar ROI SUV differences compared to reconstruction with no motion correction (NMC).

3.2 Multi-subject Experiments

Based on the results of the pilot experiments (Sec 3.1), we verified the generalizability of the proposed network by training on multiple subjects and avoid bias by evaluating on subjects not included in the training data. Here, we split the patient cohort into training and testing subsets of 20 and 5 subjects, respectively. Training employed the sampling strategy from Sec. 2.3 using n_{sub}=20, n_{ref}=1,800 and n_{mov}=6 (216,000 unique data samples). Using an epoch size 4,096, model training required 20,000 epochs for convergence on the validation set. Fig. 2b shows motion prediction results for three example subjects: one subject included in the training set (to test if the model experiences any loss due to increasing the training population) and two subjects from the testing dataset (a good example and the worst performing example). Table S1 shows quantitative results for all test subjects. The results for Subject 2 (mean MSE 0.02) included in the training set show that the network is capable of accurately predicting motion from training subjects even though the motion relative to the reference frame was never used for training. As expected, performance degrades when evaluated on data from the testing set. DL-HMC tracks the Vicra motion estimates for Subject 3 well except for rotation about y (mean MSE is 0.74), but fails to track motion well in Subject 4 (mean MSE is 6.33). Fig. 3 shows PET reconstruction results using MOLAR comparing DL-HMC motion results to Vicra and to NMC as well as quantitative ROI SUV evaluation. Subject 3 exhibits DL-HMC reconstruction performance similar to Vicra and shows similar ROI SUV values, although caudate and pallidum ROIs exhibit differences in mean SUV. As for the failure case, the DL-HMC reconstruction result is similar with NMC.

4 Discussion and Conclusion

In this work, we proposed a novel structure of deep neural network to perform head motion estimation using supervised learning. Instead of using one-second reconstructed PET images, which is time-consuming (~2 min), we use one-second 3D cloud images that take less than ten seconds to generate. DL-HMC is able to extract the motion information from these 3D cloud data with high noise levels (Fig. 1), which otherwise is extremely challenging for both standard, intensity-based registration methods and unsupervised deep learning approached. In contrast to other supervised registration learning approaches that utilize random (synthetic) transformations of the image data to learn the transformation parameters [16], we use an external device, Vicra, to provide supervision. The motion correction task is further complicated by the dynamic nature of PET imaging caused by tracer kinetics. With the help of FWTs, DL-HMC can capture tracer kinetic changes. The success of the uniform sampling strategy (Table 1) indicates that diversity in the training motion (reference and moving time frames are more likely to have larger motion between them with increasing Δt) is import for creating a robust model. Quantitative and qualitative evaluations of pilot and multi-subject experiments demonstrate that the proposed DL-HMC has potential for accurate PET rigid head motion correction.

Limitations of the work include training with a small number of subjects, which prevents the model from generalizing across diverse subject anatomies and motions. Our model may also benefit from a different data representation. While our initial model results indicate capabilities of predicting motion of magnitude ~1mm, our current pre-processing reduces input image resolution to ~10mm^3, which may limit the model's ability to detect motion with smaller magnitudes. Finally, we train and test DL-HMC using a single tracer (FDG), although this is by far the most commonly used. In the future, we will include more training data and test on a larger subject cohort, experiment with alternative data representations, and evaluate DL-HMC on other tracers.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

Research reported in this publication was supported by the National Institute Of Biomedical Imaging And Bioengineering (NIBIB) of the National Institutes of Health (NIH) under Award Number R21 EB028954. The content is solely the responsibility of the authors and does not necessarily represent the official views of the NIH.

References

1. Beyer T, Tellmann L, Nickel I, Pietrzyk U: On the use of positioning aids to reduce misregistration in the head and neck in whole-body pet/ct studies. Journal of Nuclear Medicine 46(4), 596–602 (2005) [PubMed: 15809481]
2. Carson RE, Barker WC, Liow JS, Johnson CA: Design of a motion-compensation osem list-mode algorithm for resolution-recovery reconstruction for the hrrt. In: 2003 IEEE Nuclear Science Symposium. Conference Record (IEEE Cat. No. 03CH37515). vol. 5, pp. 3281–3285. IEEE (2003)
3. Dumoulin V, Perez E, Schucher N, Strub F, Vries H.d., Courville A, Bengio Y: Feature-wise transformations. Distill 3(7), e11 (2018)
4. Fischl B, Van Der Kouwe A, Destrieux C, Halgren E, Ségonne F, Salat DH, Busa E, Seidman LJ, Goldstein J, Kennedy D, et al. : Automatically parcellating the human cerebral cortex. Cerebral cortex 14(1), 11–22 (2004) [PubMed: 14654453]
5. Green MV, Seidel J, Stein SD, Tedder TE, Kempner KM, Kertzman C, Zeffiro TA: Head movement in normal subjects during simulated pet brain imaging with and without head restraint. Journal of Nuclear Medicine 35(9), 1538–1546 (1994) [PubMed: 8071706]
6. Huang G, Liu Z, Van Der Maaten L, Weinberger KQ: Densely connected convolutional networks. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 4700–4708 (2017)
7. Jin X, Mulnix T, Gallezot JD, Carson RE: Evaluation of motion correction methods in human brain PET imaging-a simulation study based on human motion data. Medical Physics 40(10), 102503 (Sep 2013) [PubMed: 24089924]
8. Jin X, Mulnix T, Sandiego CM, Carson RE: Evaluation of frame-based and event-by-event motion-correction methods for awake monkey brain pet imaging. Journal of Nuclear Medicine 55(2), 287–293 (2014) [PubMed: 24434295]
9. Keller SH, Sibomana M, Olesen OV, Svarer C, Holm S, Andersen FL, Højgaard L: Methods for motion correction evaluation using 18f-fdg human brain scans on a high-resolution pet scanner. Journal of Nuclear Medicine 53(3), 495–504 (2012) [PubMed: 22331217]
10. Kuang Z, Wang X, Ren N, Wu S, Gao J, Zeng T, Gao D, Zhang C, Sang Z, Hu Z, et al. : Design and performance of siat apet: a uniform high-resolution small animal pet scanner using dual-ended readout detectors. Physics in Medicine & Biology 65(23), 235013 (2020) [PubMed: 32992302]
11. Kyme AZ, Se S, Meikle SR, Fulton RR: Markerless motion estimation for motion-compensated clinical brain imaging. Physics in Medicine & Biology 63(10), 105018 (2018) [PubMed: 29637899]
12. LeCun Y, Bengio Y, Hinton G: Deep learning. nature 521(7553), 436–444 (2015) [PubMed: 26017442]
13. Lu Y, Gallezot JD, Naganawa M, Ren S, Fontaine K, Wu J, Onofrey JA, Toyonaga T, Boutagy N, Mulnix T, et al. : Data-driven voluntary body motion detection and non-rigid event-by-event correction for static and dynamic pet. Physics in Medicine & Biology 64(6), 065002 (2019) [PubMed: 30695768]
14. Lu Y, Naganawa M, Toyonaga T, Gallezot JD, Fontaine K, Ren S, Revilla EM, Mulnix T, Carson RE: Data-driven motion detection and event-by-event correction for brain pet: Comparison with vicra. Journal of Nuclear Medicine 61(9), 1397–1403 (2020) [PubMed: 32005770]
15. Rodriguez-Vieitez E, Saint-Aubert L, Carter SF, Almkvist O, Farid K, Schöll M, Chiotis K, Thordardottir S, Graff C, Wall A, et al. : Diverging longitudinal changes in astrocytosis and amyloid pet in autosomal dominant alzheimer's disease. Brain 139(3), 922–936 (2016) [PubMed: 26813969]

16. Sloan JM, Goatman KA, Siebert JP: Learning Rigid Image Registration - Utilizing Convolutional Neural Networks for Medical Image Registration. In: Proceedings of the 11th International Joint Conference on Biomedical Engineering Systems and Technologies. pp. 89–99. SCITEPRESS - Science and Technology Publications (2018)

17. Sotiras A, Davatzikos C, Paragios N: Deformable medical image registration: A survey. IEEE transactions on medical imaging 32(7), 1153–1190 (2013) [PubMed: 23739795]

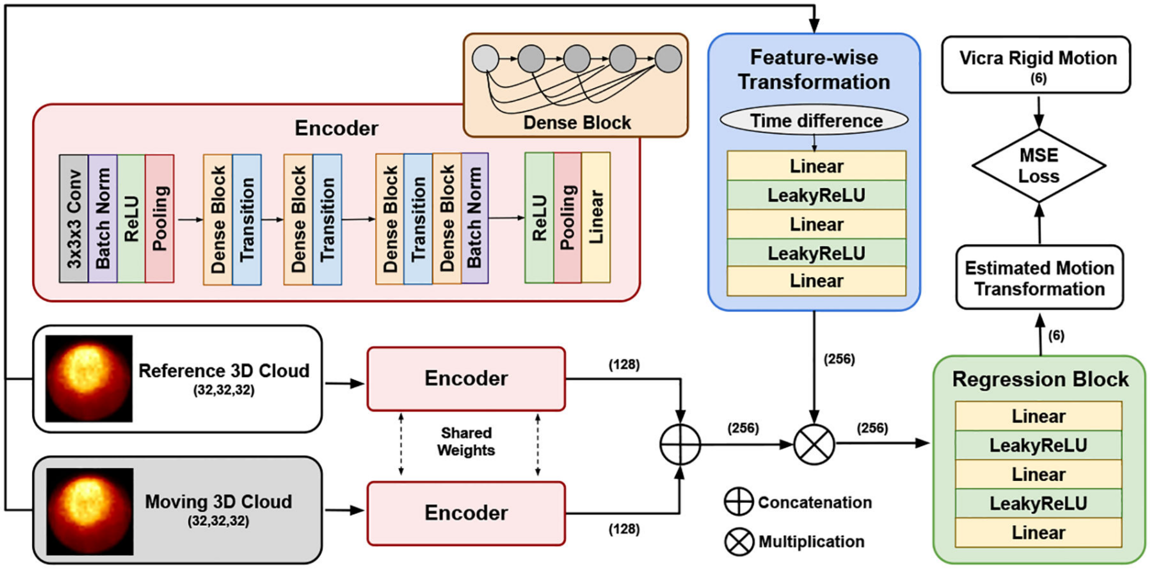


Fig. 1. DL-HMC network architecture.

The network takes two one-second 3D point cloud images as input to a feature extraction block (with a DenseNet structure) with shared weights between the reference and moving images. A regression block then estimates the rigid motion transformation parameters. A feature-wise transformation block conditions the network to relative time difference bewteen the reference and moving images.

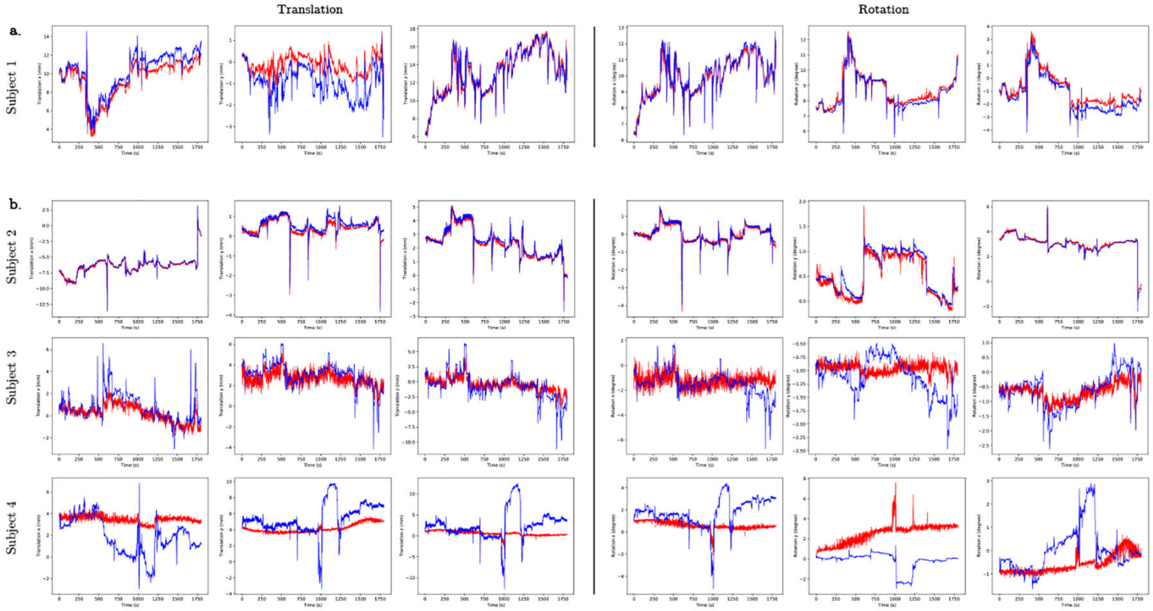
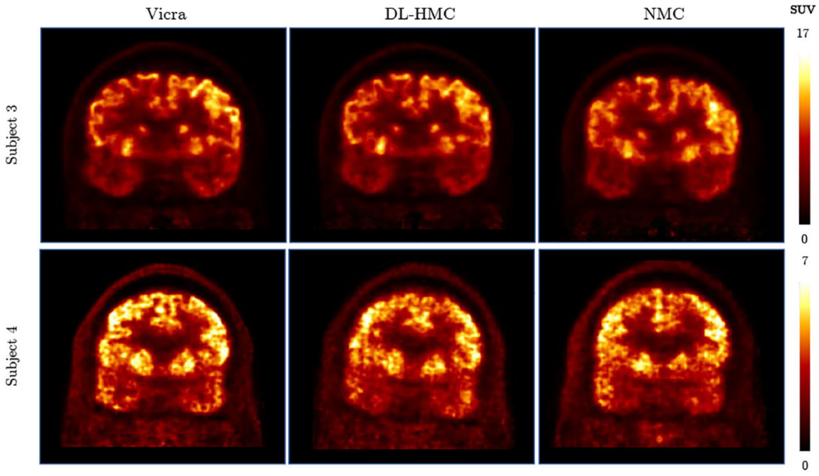


Fig. 2. Motion prediction results.

Columns show rigid transformation parameters (from left to right: translation in x, y, z directions and rotation about the x, y, z axes) from DL-HMC (red) and gold-standard Vicra motion tracking (blue). (a) Estimates from single-subject (Subject 1) model experiments and from (b) the multi-subject model in three example test subjects: example case in training set (Subject 2); a good example from the test set (Subject 3, Subject C in Table S1); and a failure case in the test set (Subject 4, Subject D in Table S1).



	Vicra	%diff DL-HMC	%diff NMC
Amygdala	5.50	3%	8%
Caudate	8.11	-4%	-24%
Cerebellum Cortex	8.17	-1%	-4%
Cerebellum WM	6.24	-1%	-4%
Cerebral WM	6.96	0%	4%
Frontal	10.30	-2%	-13%
Hippocampus	5.61	0%	-9%
Insula	7.94	0%	0%
Occipital	9.15	1%	-3%
Pallidum	7.05	4%	-2%
Parietal	9.21	0%	-7%
Putamen	9.83	0%	-11%
Temporal	7.11	-1%	-3%
Thalamus	7.71	0%	-11%
Amygdala	3.27	-6%	-7%
Caudate	3.26	1%	-4%
Cerebellum Cortex	4.37	-6%	-7%
Cerebellum WM	3.44	3%	-6%
Cerebral WM	3.62	0%	-1%
Frontal	5.07	-6%	-12%
Hippocampus	3.27	-3%	-3%
Insula	4.06	-4%	-7%
Occipital	5.45	-7%	-6%
Pallidum	3.42	3%	5%
Parietal	5.46	-7%	-10%
Putamen	4.98	-9%	-12%
Temporal	4.39	-11%	-10%
Thalamus	3.85	-6%	-7%

Fig. 3. PET image reconstruction results.

MOLAR reconstructed images using Vicra gold-standard motion tracking, the proposed DL-HMC predicted motion correction, and no motion correction (NMC). Subject 3 is an example of successful motion correction using DL-HMC, and Subject 4 represents a failure case. The table on the right shows quantitative SUV difference values evaluated of twelve brain ROIs.

Table 1.

DL-HMC ablation study results.

Reported values are mean±SD of corresponding dataset.

More Data	FWT	Deep Encoder	Unif. Sampling	Val. MSE	Test MSE
✓	✓	✓	✓	0.070±0.109	**0.035±0.073**
✓	✓	✓	○	**0.029±0.082**	3.530±4.644
✓	✓	○	✓	0.094±0.244	0.058±0.118
✓	✓	○	○	0.139±0.237	0.112±0.179
✓	○	○	✓	0.170±0.227	0.188±0.177
○	✓	○	✓	0.144±0.186	0.143±0.204
○	○	○	○	0.665±0.970	0.786±0.836