



Fast Reconstruction for Deep Learning PET Head Motion Correction

Tianyi Zeng¹, Jiazhen Zhang², Eléonore V. Lieffrig¹, Zhuotong Cai¹,
Fuyao Chen², Chenyu You³, Mika Naganawa¹, Yihuan Lu⁵,
and John A. Onofrey^{1,2,4}✉

¹ Department of Radiology and Biomedical Imaging, Yale University, New Haven, CT, USA

{tianyi.zeng,jiazhen.zhang,eleonore.lieffrig,zhuotong.cai,
mika.naganawa,fuyao.chen,chenyu.you,john.onofrey}@yale.edu

² Department of Biomedical Engineering, Yale University, New Haven, CT, USA

³ Department of Electrical Engineering, Yale University, New Haven, CT, USA

⁴ Department of Urology, Yale University, New Haven, CT, USA

⁵ United Imaging Healthcare, Shanghai, China

yihuan.lu@united-imaging.com

Abstract. Head motion correction is an essential component of brain PET imaging, in which even motion of small magnitude can greatly degrade image quality and introduce artifacts. Building upon previous work, we propose a new head motion correction framework taking fast reconstructions as input. The main characteristics of the proposed method are: (i) the adoption of a high-resolution short-frame fast reconstruction workflow; (ii) the development of a novel encoder for PET data representation extraction; and (iii) the implementation of data augmentation techniques. Ablation studies are conducted to assess the individual contributions of each of these design choices. Furthermore, multi-subject studies are conducted on an ¹⁸F-FPEB dataset, and the method performance is qualitatively and quantitatively evaluated by MOLAR reconstruction study and corresponding brain Region of Interest (ROI) Standard Uptake Values (SUV) evaluation. Additionally, we also compared our method with a conventional intensity-based registration method. Our results demonstrate that the proposed method outperforms other methods on all subjects, and can accurately estimate motion for subjects out of the training set. All code is publicly available on GitHub: https://github.com/OnofreyLab/dl-hmc_fast_recon_miccai2023.

Keywords: Deep Learning · PET fast reconstruction · Data-driven motion correction · Brain PET

1 Introduction

Positron emission tomography (PET) has been widely used in human brain imaging, thanks to the availability of a vast array of specific radiotracers. These

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/978-3-031-43999-5_67.

compounds allow for studying various neurotransmitters and receptor dynamics for different brain targets [11]. Brain PET images are commonly used to diagnose and monitor neurodegenerative diseases, such as Alzheimer’s disease, Parkinson’s disease, epilepsy, and certain types of brain tumors [3]. Head motion in PET imaging reduces brain image resolution, lowers tracer distribution estimation, and introduces attenuation correction (AC) mismatch artifacts [12]. Consequently, the capability to monitor and correct head motion is of utmost importance in brain PET studies.

The first step of PET head motion correction is motion tracking. When head motion information is acquired, either frame-based motion correction or event-by-event (EBE) motion correction methods can be applied in the reconstruction workflow to derive motion-free PET images. EBE motion correction provides better results for real-time motion tracking compared to frame-based methods, as the latter does not allow for correction of motion that occurs within each dynamic frame [1]. Currently, there are two main categories of head motion tracking methods, hardware-based motion tracking (HMT) and data-driven methods. For HMT, head motion is obtained from external devices. Generally, HMT systems offer accurate tracking results with high time resolution. Marker-based HMT such as Polaris Vicra (NDI, Canada) use light-reflecting markers on the patient’s head and track the markers for motion correction [6]. However, Vicra is not routinely used in the clinic, as setup and calibration of the tracking device can be complicated and attaching markers to each patient increases the logistical burden of the scan. In response, some researchers began to use markerless motion tracking systems for brain PET [4, 13]. These methods typically rely on the use of cameras and computer vision algorithms to detect and analyze the movement of a person’s head in real-time, but these methods still require additional hardware setup. In data-driven motion tracking methods, head motion is estimated from PET reconstructions or raw data. With the development of commercial PET systems and technological advancements such as time of flight (TOF), data-driven head PET motion tracking has shown promising results in reducing motion artifacts and improving image quality. For instance, [12] developed a novel data-driven head motion detection method based on the centroid of distribution (COD) of PET 3D point cloud image (PCI). Image registration methods that seek to align two or more images offer a data-driven solution for correcting head motion. Intensity-based registration methods have been used to track head motion using good-quality PET reconstruction frames to achieve stable performance [14]. However, because of the dynamic change in PET images, current registration-based methods need to split the data into several discrete time frames, e.g., 5 min. Therefore, they will introduce a cumulative error when dealing with inter-frame motion. Finally, inspired by the development of deep learning-based registration methods, a deep learning head motion correction (DL-HMC) network using Vicra as ground truth was proposed [15]. This study achieved accurate motion tracking on single subject testing data, but showed less accurate motion predictions for multi-subject motion studies. Meanwhile, the input images were low-resolution PCIs without TOF and had large voxel spacing, which can negatively affect motion tracking accuracy.

In this study, we proposed a new method to perform deep learning-based brain PET motion prediction across multiple subjects by utilizing high-resolution one-second fast reconstruction images (FRIs) with TOF. A novel encoder and data augmentation strategy was also applied to improve model performance. Ablation studies were conducted to assess the individual contributions of key method components. Multi-subject studies were conducted on a dataset of 20 subject and its results were quantitatively and qualitatively evaluated by MOLAR reconstruction studies and corresponding brain Region of Interest (ROI) Standard Uptake Values (SUV) evaluation.

2 Methods

2.1 Data

We identified 20 ^{18}F -FPEB studies from a database of brain PET scans acquired on a conventional mCT scanner (Siemens, Germany) at the Yale PET center. All subjects are healthy controls and the mean activity injected was 3.90 ± 1.02 mCi. The mean translational and rotational motions across time and scans are 3.75 ± 6.88 mm and $3.30 \pm 8.77^\circ$, respectively. PET list-mode data and Vicra motion tracking information are available for each subject, as well as T1-weighted MR images and MR-space to PET-space transformation matrices. We consider data acquired between 60 and 90 min post injection (30 min total).

2.2 Fast Reconstruction Images

To overcome the challenges when using low-quality, noisy PCI for motion correction, ultra-fast reconstruction techniques [14] that generate one-second dynamic fast reconstruction images (FRIs), can be utilized as input for deep learning motion correction methods. Leveraging the availability of CPU-parallel reconstruction platforms [5], we develop a reconstruction package for one-second FRI. Our proposed method employs a TOF-based projector and utilizes pure maximum-likelihood expectation maximization (MLEM) for reconstruction [10]. Attenuation correction (AC) and scatter correction are turned off to avoid AC mismatch and expensive computation. Normalization correction, random correction and decay correction are applied and the iteration number was set to two. Standard Uptake Value (SUV) calculation was also conducted to normalize the activities between different subjects. The final reconstructed image dimension is $150 \times 150 \times 111$, with voxels spaced at $2.04 \times 2.04 \times 2.00$ mm³. For comparison purposes, we also computed the same resolution PCI by TOF back-projection of the PET list-mode data along the line-of-response (LOR) with normalization for scanner sensitivity. Both FRI and PCI were resized to $96 \times 96 \times 64$, with voxel spacing of $3.18 \times 3.18 \times 3.45$ mm³. As shown in Fig. 1, the resized FRI and PCI from the same time pairs are displayed. Due to the PET corrections, the FRI quality and noise level is superior to the PCI, particularly in areas outside of the head.

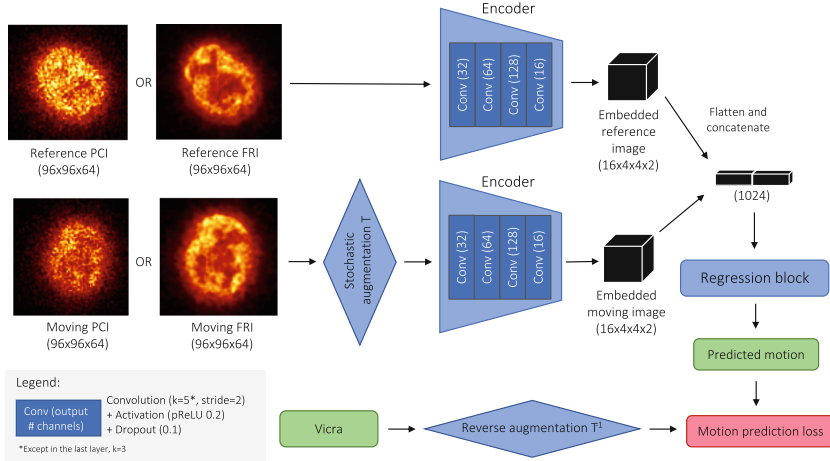


Fig. 1. The proposed network architecture and different network inputs comparison. The proposed network takes two one-second FRIs as input to encoder block with shared weights. Stochastic motion T from the data augmentation block is applied to the moving reconstruction, and the reverse augmentation T^{-1} is applied on corresponding Vicra data to match the augmented FRI pairs. A regression block then estimates the rigid motion transformation parameters. PCI input is shown on the left for visual comparison.

2.3 Motion Correction Framework

Network Architecture. We propose a modified version of the DL-HMC framework to learn rigid head motion in a supervised manner [15] (Fig. 1). Our proposed method uses two FRIs I_{ref} and I_{mov} from two different time points t_{ref} and t_{mov} to predict the relative rigid motion transformation between the reference and moving time points with respect to the Vicra gold-standard. Our encoder consists of 3 convolution layers with kernel size of 5^3 and an intermediate convolution layer with convolution size of 3^3 . PReLU activation layers follow each convolution. In addition, we add dropout in the regression layers with rate 0.3. In this new architecture, the embedding space consists of feature maps of size $16 \times 4 \times 4 \times 2$, which preserves the spatial information in the FRI. No padding is applied to the images or feature maps. The extracted features are fed into the fully connected regression block to predict the six translation and rotation components of the relative rigid motion.

Data Augmentation. To improve the performance and generalizability of our network, we use a task-specific data augmentation strategy to expose it to more varied and diverse training data. As a rule of thumb, translations of 2–5 mm and rotations of 2° – 3° are common and larger magnitudes are expected without restraint or if non-customized supports are used [9]. Due to our sampling strategy during model training, statistically, most of the sampled pairs will have small

relative motion. However, during the inference, the relative motion between the moving frames at late time points and the reference frame at the beginning will be large due to the accumulation of motion. Therefore, the model may not be able to make accurate predictions when facing large relative motions. To take this problem into account, we perform data augmentation by simulating an additional relative motion that can be concatenated with the true relative motion. To be specific, the synthetic translation and rotation are uniformly sampled in the range of $[-10, 10]$ mm and $[-5, 5]^\circ$, respectively. The randomly simulated motion T will be applied to the moving frame to generate a synthetic moving frame $T \circ I_{\text{ref}}$ and be concatenated with the real relative motion to acquire the synthetic relative motion between the reference and the synthetic moving frame. The synthetic moving frame and the synthetic relative motion will be used for training to increase the data variability.

Network Training and Inference. To train the network, we randomly sampled image pairs $(t_{\text{ref}}, t_{\text{mov}})$ under the condition $(t_{\text{ref}} < t_{\text{mov}})$. The network was optimized by minimizing the mean square error (MSE) between the predicted motion estimate and Vicra parameters. More specifically, the prediction error for a given pair of reference and moving clouds is defined as $\mathcal{L}(\hat{\theta}, \theta) = \|\hat{\theta} - \theta\|^2$ with $\theta = [t_x, t_y, t_z, r_x, r_y, r_z]$ the Vicra information for the three translational and three rotational parameters (t_x, t_y, t_z) and (r_x, r_y, r_z) , respectively, and $\hat{\theta}$ the network prediction. After training the model, we perform motion tracking inference by setting the image from first time point $t_{\text{ref}} = 3,600$ (60 min post-injection) as the reference image and predict the motion from this reference image to all subsequent one-second image frames in the next 30 min (1,800 one-second time points).

3 Results

We performed quantitative and qualitative experiments to validate our approach. We evaluated motion correction performance by comparing our proposed method to DL-HMC [15], intensity-based registration using the BioImage Suite (BIS) software package [7] using the one-second FRIs, and ablation studies to demonstrate the effectiveness of our design choices. We qualitatively assessed motion correction performance by reconstructing the PET images with motion tracking result and comparing to DL-HMC and the Vicra gold-standard reconstruction results. We split our dataset of 20 subjects into distinct subsets for training and testing with 14 and 6 subjects, respectively. From the training cohort, we randomly sampled 10% of the time frames to be used as a validation set. Training the network required 6,000 epochs for convergence using a mini-batch size of 64. Adam optimization was used with initial learning rate set to $5e-4$, γ set to 0.98, and exponential decay with a step size of 150. 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). The network was implemented in Python (v3.9) using PyTorch (v1.13.1) and MONAI (v1.0.1).

Table 1. Quantitative motion correction results. We compared our proposed approach using fast reconstruction image (FRI) as input with DL-HMC and with using point cloud image (PCI) as input. An ablation study quantifies the effect of stochastic data augmentation (DA) and to standard intensity-based registration (BIS). Reported values are MSE (mean \pm SD) comparing motion estimates to Vicra gold-standard.

Method	Val. loss	Test Set		
		Total loss	Translation (mm)	Rotation ($^{\circ}$)
DL-HMC PCI	0.60 ± 1.18	0.72 ± 0.18	1.05 ± 0.75	0.38 ± 0.43
DL-HMC FRI	0.33 ± 0.58	0.87 ± 0.42	1.14 ± 1.03	0.59 ± 1.03
Proposed PCI	0.35 ± 0.28	0.32 ± 0.22	0.28 ± 0.26	0.35 ± 0.31
Proposed FRI	0.18 ± 0.12	0.30 ± 0.19	0.35 ± 0.31	0.25 ± 0.19
Proposed w/o DA	0.20 ± 0.24	0.41 ± 0.25	0.61 ± 0.49	0.21 ± 0.20
BIS	N/A	8.40 ± 12.34	2.48 ± 2.35	1.15 ± 1.05

Quantitative Evaluation. For quantitative comparisons of motion tracking, we compare MSE loss in the validation set and test set. We calculate MSE for the 6 parameter rigid motion as well as the translation and rotational components separately. To verify feasibility of traditional intensity-based registration method on FRIs, we use BIS with a multi-resolution hierarchical representation (3 levels) and minimize the sum of squared differences (SSD) similarity metric (Fig. S1 shows a BIS result on a example testing subject). Compared to Vicra gold-standard, BIS fails to predict the motion. We evaluate the following motion prediction methods (Table 1): (i) DL-HMC with PCI as input (DL-HMC PCI); (ii) DL-HMC with FRI as input (DL-HMC FRI); (iii) proposed network with PCI as input (Proposed PCI); (iv) proposed network with FRI as input (Proposed FRI); and (v) proposed method with FRI but without the data augmentation module (Proposed w/o DA); Results demonstrate that the proposed network with FRI input provides the best motion tracking performance in both validation and testing data. We also observe that using FRI yields a lower loss for the proposed network, indicating that high image quality enhanced the motion correction performance. For testing translation results, Proposed PCI outperforms Proposed FRI and has similar total motion loss, which indicates that the proposed network can still estimate motion on testing subjects even with noisy input. Figure 2 shows motion prediction results for different variations of the proposed method in a single test subject. These results show that the proposed FRI method is more similar to Vicra than the other methods, especially for translation in the x and z directions. However, the Proposed FRI method exhibits higher variance than other methods, which may be a result of the data augmentation distribution. Overall, our experiments demonstrate that the strategies in proposed FRI method enhance the motion tracking performance of the network.

Qualitative Reconstruction Evaluation. After inference, the 6 rigid degrees of freedom transformation estimated from the network were used to reconstruct

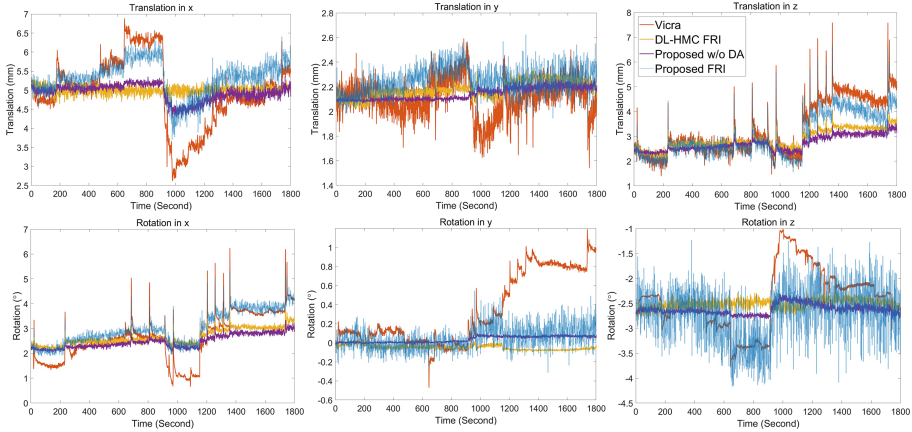


Fig. 2. Motion information comparison for testing subject inference. Columns show rigid transformation parameters from Vicra (Orange), DL-HMC FRI (yellow), proposed network without data augmentation (purple), and proposed FRI method (blue). (Color figure online)

the PET images using Motion-compensation OSEM List-mode Algorithm for Resolution-Recovery Reconstruction (MOLAR) algorithm [5]. We applied PET head motion correction using the Proposed FRI model and compared with Vicra and no motion correction (NMC) reconstruction results. Figure 3 shows the reconstruction results for the same testing subject in quantitative evaluation. Based on the tracer distribution of ^{18}F -PEB, we selected some frames from reconstructed images to illustrate the proposed FRI motion correction performance. In general, the Proposed FRI results in qualitatively enhanced anatomical interpretation of PET images. In Fig. 3, NMC reconstruction has motion blurring on margins of the brain, while Proposed FRI reconstruction shows well-defined gyrus and sulcus comparable to Vicra reconstruction. ^{18}F -FPEB tracer is a metabotropic glutamate 5 receptor antagonist with moderate to high accumulation in multiple brain regions such as insula/caudate nucleus, thalamus and temporal lobe. Thus, we compared visualization of these regions among NMC, Proposed FRI and Vicra reconstructions (Fig. 3). Specifically, our Proposed FRI method yields clear delineation of the insula/caudate nucleus, thalamus, and temporal lobe nearly indistinguishable from Vicra reconstructed images.

In addition, brain region of interest (ROI) analyses were also performed for quantitative use. Each subject's MR image was segmented into 74 regions using FreeSurfer software [2]. These regions were then merged into twelve large grey matter (GM) ROIs. For all testing subjects, the SUV difference of 12 ROIs from DL-HMC, Proposed PCI, and Proposed FRI reconstruction were calculated for comparison with Vicra reference (Fig. 3 right). The proposed FRI method yields the lowest absolute difference (0.8%) from Vicra images in SUV, while the absolute SUV difference for DL-HMC FRI is 1.5% and for NMC is 1.9%.

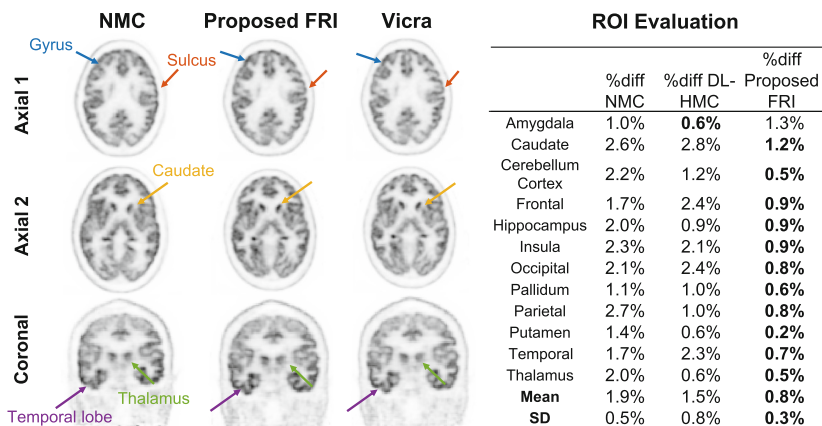


Fig. 3. PET image reconstruction and ROI evaluation results. MOLAR reconstructed images using Vicra gold-standard motion correction, the proposed FRI motion correction, and no motion correction (NMC). The table on the right shows quantitative SUV difference values with respect to Vicra.

Specifically, results of proposed FRI method are closest to Vicra results in regions such as thalamus, temporal lobe, insula (absolute activity difference are 0.5%, 0.7%, 0.9%, respectively), which are the target areas of ^{18}F -FPEB tracer with highest empirical tracer accumulation. The reconstruction and ROI evaluation results indicate that the proposed FRI method holds potential to improve clinical applicability through amelioration of PET motion correction accuracy.

4 Discussion and Conclusion

In this work, we propose a new head motion correction approach using fast reconstructions as input. The proposed method outperforms other methods in a multi-subject cohort, and ablation studies demonstrate the effectiveness of our strategies. We apply our proposed FRI motion correction to get motion-free reconstruction using MOLAR. The proposed FRI method achieves good image quality and similar ROI evaluation results compared to Vicra gold-standard HMT. In this study, we showed that conventional intensity-based registration fails at performing motion tracking on FRI data. This is likely due to the PET dynamic changes and non-optimal registration parameters. Compared with previous deep learning motion correction [15], the training speed and GPU memory usage of the proposed method are much better thanks to the proposed shallower encoder architecture and our efficient training and testing strategies. Though HMT method such as Vicra achieves good accuracy and time resolution for PET head motion tracking, two common types of Vicra failure may occur: slipping and wobbling. Our method would be robust enough to compensate for the

Vicra failure. Because of the limited Vicra data, in the future, we will develop semi-supervised deep learning methods for PET head motion correction. Our study used TOF PET data because it can yield high signal to noise ratio (SNR) for both FRI and PCI due to the better location identification of photons, thus the one-second FRI still retains some essential brain structures. Limitations of this work include partial limited tracking time and low time resolution compared to HMT methods mentioned in Sect. 1. In the future, with the development of PET techniques such as depth-of-interaction [8], higher resolution and sensitivity PET will be available. Such PET will give data-driven PET motion correction a revolutionary opportunity to have more accurate tracking and higher time resolution. We plan to apply the proposed method to other datasets, developing a generalized model for multi-tracer and multi-scanner PET data.

Acknowledgments. 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. Carson, R.E., Barker, W.C., Liow, J.S., Johnson, C.A.: 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)
2. Fischl, B.: Freesurfer. *Neuroimage* **62**(2), 774–781 (2012)
3. Hoffman, J.M., et al.: FDG PET imaging in patients with pathologically verified dementia. *J. Nucl. Med.* **41**(11), 1920–1928 (2000)
4. Iwao, Y., Akamatsu, G., Tashima, H., Takahashi, M., Yamaya, T.: Marker-less and calibration-less motion correction method for brain pet. *Radiol. Phys. Technol.* **15**(2), 125–134 (2022)
5. Jin, X., et al.: List-mode reconstruction for the biograph MCT with physics modeling and event-by-event motion correction. *Phys. Med. Biol.* **58**(16), 5567 (2013)
6. Jin, X., Mulnix, T., Sandiego, C.M., Carson, R.E.: Evaluation of frame-based and event-by-event motion-correction methods for awake monkey brain pet imaging. *J. Nucl. Med.* **55**(2), 287–293 (2014)
7. Joshi, A., et al.: Unified framework for development, deployment and robust testing of neuroimaging algorithms. *Neuroinformatics* **9**(1), 69–84 (2011)
8. Kuang, Z., et al.: Design and performance of SIAT APET: a uniform high-resolution small animal pet scanner using dual-ended readout detectors. *Phys. Med. Biol.* **65**(23), 235013 (2020)
9. Kyme, A.Z., Fulton, R.R.: Motion estimation and correction in SPECT, PET and CT. *Phys. Med. Biol.* **66**(18), 18TR02 (2021)
10. Lange, K., Carson, R., et al.: EM reconstruction algorithms for emission and transmission tomography. *J. Comput. Assist. Tomogr.* **8**(2), 306–16 (1984)
11. Phelps, M.E., Mazziotta, J.C.: Positron emission tomography: human brain function and biochemistry. *Science* **228**(4701), 799–809 (1985)

12. Revilla, E.M., et al.: Adaptive data-driven motion detection and optimized correction for brain pet. *Neuroimage* **252**, 119031 (2022)
13. Slipsager, J.M., et al.: Markerless motion tracking and correction for PET, MRI, and simultaneous PET/MRI. *PLoS ONE* **14**(4), e0215524 (2019)
14. Spangler-Bickell, M.G., Deller, T.W., Bettinardi, V., Jansen, F.: Ultra-fast list-mode reconstruction of short pet frames and example applications. *J. Nucl. Med.* **62**(2), 287–292 (2021)
15. Zeng, T., et al.: Supervised deep learning for head motion correction in pet. In: Wang, L., Dou, Q., Fletcher, P.T., Speidel, S., Li, S. (eds.) *MICCAI 2022. LNCS*, vol. 13434, pp. 194–203. Springer, Cham (2022). https://doi.org/10.1007/978-3-031-16440-8_19