Bowsher prior Enhanced Unsupervised PET image Denoising

Zhongxue Wu, Jiankai Wu, Jianan Cui *, Yuanjing Feng, and Zan Chen *

the College of Information Engineering, Zhejiang University of Technology, Hangzhou, 310023, China
{jianancui, zanchen2}@zjut.edu.cn

Abstract. Positron Emission Tomography (PET) is an advanced nuclear medicine imaging technique widely used in the diagnosis and treatment of oncology and neurological diseases. However, PET images suffer from high noise levels due to statistical fluctuations and physical degradation factors during image acquisition. Recently, deep learning-based denoising methods have shown great performance for PET image quality enhancement. Most of these methods attempt to incorporate high-quality anatomical image (such as CT or MR), as network input to provide prior information into the PET denoising process. However, directly using CT or MR image as network input has limited effectiveness and lacks interpretability due to the significant differences between two modalities. Exploring how to make better use of anatomical prior remains a valuable research direction. In this study, we proposed an unsupervised PET image denoising framework that leverages the Bowsher prior to achieving cross-modality fusion and anatomical information extraction. Specifically, we compute the Bowsher prior using the denoised result from the Conditional Deep Image Prior (CDIP) method and the corresponding MR image. The Bowsher prior and MR image are concatenated along the channel dimension and then fed into a designed Spatial Attention Network (SA-Net) to enhance PET image quality. Experiments on both simulation and clinical datasets demonstrated that the proposed framework can effectively utilize Bowsher prior to generating high-quality PET image.

Keywords: PET image denoising · Anatomical prior · Unsupervised deep learning.

1 Introduction

Positron Emission Tomography (PET) is an effective nuclear medicine imaging technique which can quantify physiological metabolic processes by tracking the distribution of radioactive tracers in the body. It has wide applications in oncology [8], cardiology [20], and neurology [21]. However, due to the stochastic

* Corresponding authors
All authors contributed equally to this work. Readers may contact any of the authors listed above.

decay process of radiotracers and limitations of imaging equipment, PET image is inherently affected by statistical fluctuations and physical degradation factors [1]. As a result, PET image suffer from the low signal-to-noise ratio (SNR) which affects its detection and quantification accuracy, especially for small structures. Furthermore, to reduce patient radiation exposure during medical procedures, researchers are actively developing low-dose and short-duration PET imaging protocols [26], these approaches further amplify noise levels, making it increasingly challenging to obtain high-quality PET image [19].

In recent years, numerous PET enhancement algorithms have been developed to improve PET image quality. Early approaches, such as Gaussian filtering [12], adaptive diffusion filtering [22], bilateral filtering [10], guided filtering [25], non-local means (NLM) filtering [7], and block-matching 4D filtering (BM4D) [15], are robust but often suppress high-frequency details, resulting in over-smoothed images. With the rapid advancement of deep learning, end-to-end PET enhancement networks have emerged, achieving significant performance improvements and showing great potential [11,14,4]. Additionally, many methods leverage anatomical images from other modalities (e.g., CT and MR) as prior information to further enhance the denoising performance [16,17,3].

Conditional Deep Image Prior (CDIP) [5] is a representative method that utilizes patient's high-quality anatomical image (CT or MR) as input to the deep learning network and noisy PET image as training label. By leveraging the network's architecture and prior anatomical information, the model learns intrinsic structural features from prior image and generates enhanced PET image. This approach eliminates the need for paired data, which is typically required by deep learning-based denoising methods. However, since the training label is noisy image, the selection of network training stop condition is essential to prevent overfitting or underfitting, both of which will degrade output image quality.

Furthermore, directly incorporating CT or MR image as network input to provide anatomical prior for PET denoising has limited effectiveness and lacks interpretability due to the inherent differences between two modalities. To address this challenge, we propose the Bowsher prior enhanced unsupervised PET denoising framework. Specifically, we computed the Bowsher prior based on CDIP denoising result and corresponding MR image, and the proposed Spatial Attention Network (SA-Net) leverages the Bowsher prior and MR image as network inputs, with noisy PET image as training label. This approach generates high-quality PET image without clean reference or predefined stop condition. Our work has three main contributions:

- Incorporating the Bowsher prior into PET denoising framework. The Bowsher prior technique aims to promote consistency among anatomically similar PET voxels through MR-derived regularization constraint [24], improving boundary clarity and structural fidelity in PET image while addressing denoising interpretability through multimodal information fusion.
- Proposing an unsupervised PET enhancement network (SA-Net). SA-Net utilizes a specially designed network architecture with reduced parameter

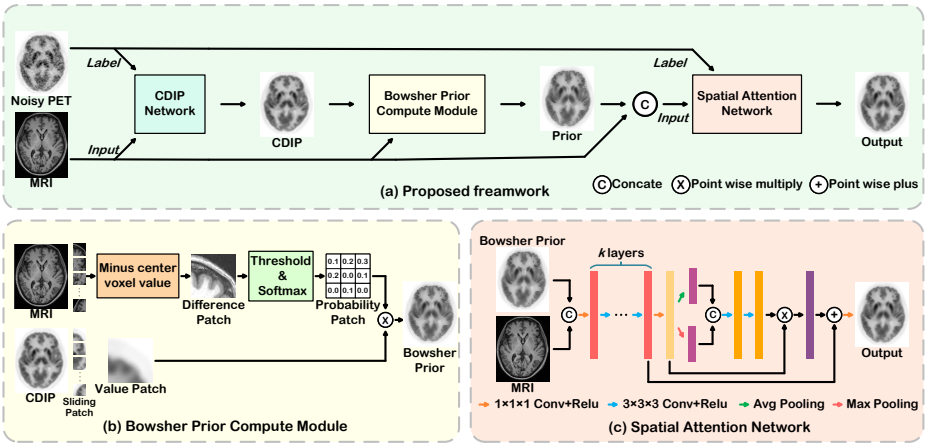

Fig. 1. Overview of the proposed framework. (a) The main structures of the proposed framework and the relationship between them. (b) Calculation of the Bowsher prior. (c) The composition of Spatial Attention Network.

capacity, effectively limiting noise overfitting during training and addressing the stop condition problem inherent in the CDIP method.
– Achieving superior denoising performance on both simulation and clinical datasets, outperforming existing unsupervised denoising approaches in quantitative and qualitative analysis.

2 Methods

The proposed framework is illustrated in Fig. 1(a). CDIP serves as the critical initial denoiser, preventing the propagation of PET noise to the Bowsher prior and SA-Net. CDIP uses the MR image as input to UNet and the noisy PET image as the training label for unsupervised denoising (as we faithfully reproduce the method described in [5], CDIP details are omitted here). The denoised result and corresponding MR image are then used to compute the Bowsher prior image. Bowsher prior bridges the CDIP and SA-Net training processes with patch-based image fusion capability. Subsequently, the Bowsher prior and the MR image are concatenated along the channel dimension and fed as input into SA-Net, with the noisy PET image as the training label. The specially designed SA-Net effectively limits noise overfitting during training and mitigates the stop problem of CDIP. The following sections detail the computation of the Bowsher prior and the structure of SA-Net.

2.1 Bowsher prior Computation

The Bowsher prior [2] incorporates an anatomy-based regularization term into the PET reconstruction process, encouraging adjacent voxels with similar signals in the anatomical image to maintain smoothness in the reconstructed PET

image. In this study, we compute the Bowsher prior using the denoised result obtained from the CDIP method and the corresponding MR image, as shown in Fig. 1(b). To avoid the computationally expensive PET image reconstruction process, we directly compute the Bowsher prior based on 3D image patches from the CDIP result and MR image, which can be expressed as:

$$y_j = \text{Softmax}\left([w_{ji}\,|z_i - z_j|]_{i=1}^{N}\right) \cdot G_j, \tag{1}$$

where z represents the MR image, and z_j corresponds to the j-th voxel of the MR image. G_j denotes a sliding patch of size $3\times l\times l$ extracted from the CDIP image, centered at j-th voxel and containing N voxels. We compute intensity differences between each voxel and the central voxel within the MR image sliding patch. These differences are then thresholded by w_{ji} according to their magnitude and then transformed into a probability patch via the Softmax function. Finally, the probability patch is multiplied with the corresponding CDIP image patch to obtain the value at j-th voxel in the Bowsher prior image.

The weight w_{ji} acts as a thresholding function to exclude voxels that significantly differ from the central voxel. It is determined by the intensity difference between the voxel in the patch and the central voxel, expressed as:

$$w_{ji} = \begin{cases} -1 & \text{If } \frac{|z_i - z_j|}{z_j} < \alpha, \\ -\infty & \text{otherwise} \end{cases} \tag{2}$$

where α is a scaling factor that determines the thresholding mechanism. If the ratio of the difference to the center value is smaller than α (α<1), the corresponding weight will be set to -1, if not the weight will be set to negative infinity. In the Softmax operation, voxels with weights of -1 are assigned probabilities proportional to their intensity differences, with smaller differences resulting in higher probability values. At the same time, voxels with weights of negative infinity are excluded from the Softmax process and assigned a value of 0 in the probability patch. This promotes smoothness among anatomically consistent voxels in PET image.

2.2 Spatial Attention Network

After computing the Bowsher prior image, we concatenate it with the MR image along the channel dimension to create a multi-channel input. This combined input is then fed into our proposed unsupervised PET enhancement network, SA-Net. The noisy PET image act as the training label, and we utilize the Mean Squared Error (MSE) loss for the unsupervised image enhancement process. The architecture of SA-Net is depicted in Fig. 1(c).

SA-Net is initially constructed by stacking k convolutional layers with a kernel size of $3\times3\times3$ resulting in a receptive field of $(2k+1)\times(2k+1)\times(2k+1)$. The features extracted by these layers are then further refined using a spatial attention mechanism, which generates the final denoised output. The formulation of the spatial attention mechanism is as follows:

$$\mathrm{SA}\left(f^k\right) = \mathrm{ReLu}\left(f^k + \mathrm{ReLu}\left(\left[\mathrm{Avg}\left(f^k\right), \mathrm{Max}\left(f^k\right)\right]\right) \odot f^k\right), \tag{3}$$

where f^k represents the features extracted from the input image after k convolutional layers, Avg and Max represent the average and maximum along the channel dimension operation, respectively.

During experiments, we observed that using noisy PET image as training label leads the network to first fit the intensity distribution of the noisy data to rapidly minimize MSE loss, followed by adapting to the noise distribution. Networks with larger parameter capacities show a stronger tendency to overfit noise. To mitigate this, SA-Net utilizes a stacked convolutional layer design to restrict parameter capacity, thereby preventing noise overfitting during training and eliminating the need for predefined stop condition.

2.3 Implementation Details

We validated the framework on simulation and clinical datasets with distinct configurations. The CDIP UNet used a learning rate of 1×10^{-3}, trained for 3,000 epochs (simulation) and 2,000 epochs (clinical), the impact of varying CDIP denoising performance is further analyzed in Section 3.2. For Bowsher prior computation, patch sizes l were 5 ($\alpha = 0.1$) and 11 ($\alpha = 0.2$) for simulation and clinical data respectively. SA-Net employed $k = 2$ (simulation) and $k = 3$ (clinical) convolutional layers, optimized via Adam ($lr = 1 \times 10^{-3}$) with matching epoch numbers.

All implementations used PyTorch on an NVIDIA RTX 4090 GPU. Bowsher prior computation required 5-7 seconds, while SA-Net training completed in 1.5 minutes (simulation) and 10 minutes (clinical).

3 Experiments

3.1 Datasets

In this study, we evaluated the proposed method on both simulation and clinical datasets. For the simulation study, noiseless PET images were generated using 20 normal brain anatomical models from BrainWeb [6]. To simulate lesions, abnormal uptake points were added to the PET images. These images were forward-projected to generate sinogram data, with Poisson noise added at a total count of 1.5×10^7. The simulation PET images with lesions (matrix size, $125 \times 125 \times 105$; voxel spacing, $1 \times 1 \times 1\mathrm{mm}^3$) were reconstructed using the Maximum Likelihood Expectation Maximization (MLEM) algorithm with 40 iterations.

The clinical dataset was obtained from OpenNeuro [18] including PET/MR data from 30 participants. All participants were scanned with Siemens Biograph mMR 3T PET-MR scanner using ^{18}F-FDG tracer. The PET image were extracted from the last 5-minute frames of a 40-minute scan, with an average intravenous dose of 184 MBq (matrix size, $256 \times 256 \times 256$; voxel sapcing,

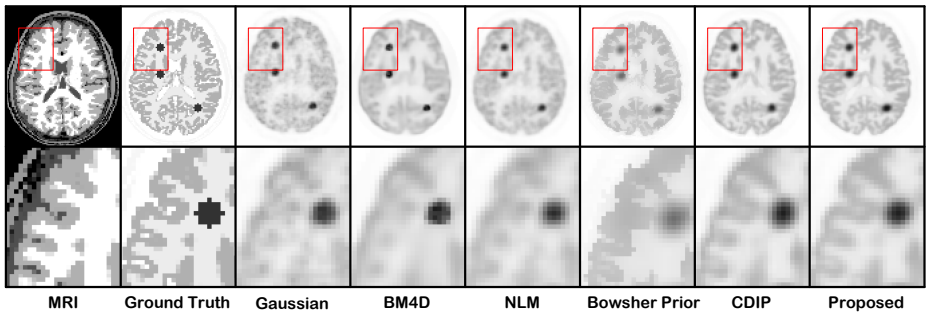

Fig. 2. The denoised simulation images using different methods.

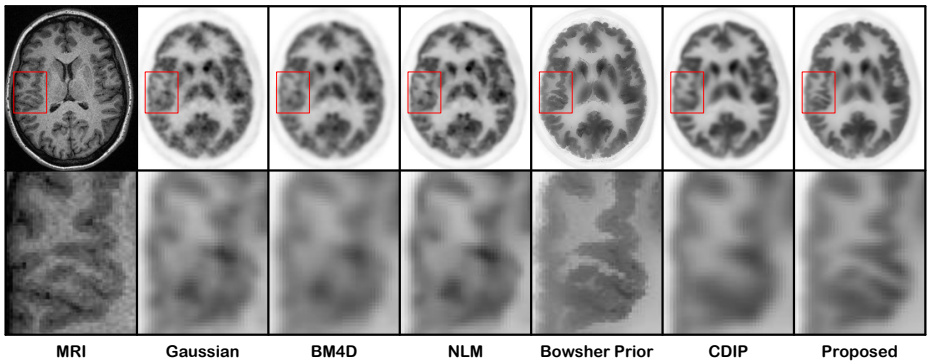

Fig. 3. The denoised clinical images using different methods.

$1.043 \times 1.043 \times 1.015\text{mm}^3$). Reconstruction was performed using the ordered subset expectation maximization (OSEM) algorithm with 14 subsets and 4 iterations. The reconstructed PET images were subsequently aligned with the corresponding T1-weighted MR images using the ANTs toolkit [23].

For the simulation dataset, we took the noiseless PET images as reference images to compute the peak signal-to-noise ratio (PSNR) and structural similarity index measure (SSIM) as quantitative metrics. For the clinical dataset, we calculated the contrast-to-noise ratio (CNR) and CNR improvement ratio based on the noisy PET image for quantitative analysis.

3.2 Comparison experiments

To validate the effectiveness of the proposed framework, we compared its denoising performance with Gaussian filtering[12], BM4D[15], NLM[7], and CDIP[5] techniques. The qualitative and quantitative results on both simulation and clinical datasets are presented in Table 1 and Fig. 2 and 3.

Table 1. Quantitative results of the comparative experiments on the simulation and clinical datasets.

Method	Simulation Dataset		Clinical Dataset
	PSNR [dB] ↑	SSIM ↑	CNR Improvement Ratio ↑
Original	28.316±0.050	0.844±0.002	(Baseline)
Gaussian [12]	29.613±0.032	0.867±0.002	30.02%±15.10%
BM4D [15]	29.506±0.042	0.858±0.001	54.83%±27.75%
NLM [7]	29.287±0.031	0.865±0.001	49.79%±31.19%
CDIP [5]	30.052±0.012	0.898±0.001	106.14%±38.33%
Proposed	**31.147±0.030**	**0.912±0.001**	**147.79%±41.15%**

Quantitative Comparison: As shown in Table 1, the proposed method achieves the highest PSNR and SSIM on the simulation dataset, along with the highest average CNR improvement ratio on the clinical dataset. The Wilcoxon rank sum test results proved that our method was significantly better than Gaussian filtering, NLM, BM4D and CDIP in quantitative metrics ($p<0.001$ for all tests).

Qualitative Comparison: Fig. 2 and Fig. 3 provide visual comparisons for the simulation and clinical datasets. Compared to methods such as Gaussian filtering, BM4D, and NLM, the proposed method generates image with significantly reduced noise and enhanced structural clarity. In contrast to the CDIP method, our approach recovers richer textural details and sharper cortical structures, closely aligning with the anatomical features observed in the MR image. These results demonstrate that our method more effectively leverages anatomical information to enhance PET image denoising performance.

3.3 Ablation experiments

To validate the effectiveness of the proposed framework, we conducted the following ablation experiments: 1) Calculate the Bowsher prior using CDIP results under both overfitting and underfitting conditions, along with the noisy PET images (without CDIP). 2) Use the Bowsher prior image alone (without MR image) as the input to SA-Net. 3) Replace SA-Net with UNet as the enhancement network. 4) Use the Bowsher prior image computed from the noisy PET image to replace the MR image as prior image for the CDIP method.

The quantitative and qualitative results of these ablation studies are presented in Table 2 and Fig. 4. Based on these results, we draw the following observations:

1) The proposed framework is robust to CDIP's denoising performance. SA-Net produces clean PET images of comparable quality regardless of whether CDIP overfits or underfits. However, CDIP remains crucial; its omission introduces residual noise and artifacts in the results. 2) Using only the Bowsher prior image as input degrades the output image quality. This demonstrates that the Bowsher prior alone provides insufficient anatomical guidance. 3) Replacing SA-Net with UNet significantly reduces denoising performance. UNet's extensive

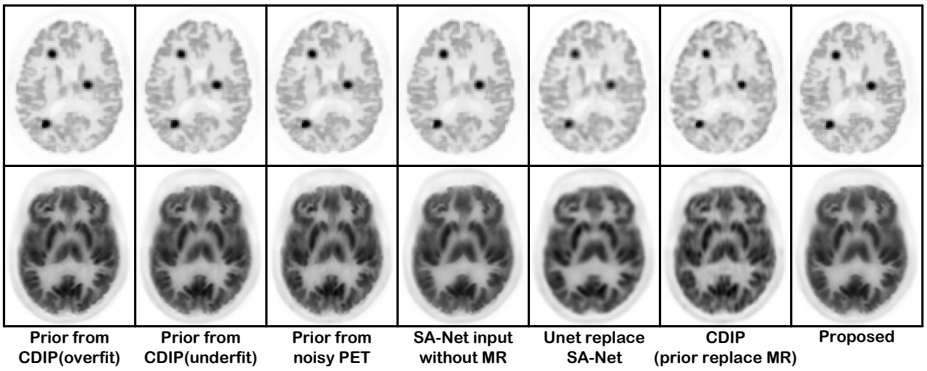

Fig. 4. Simulation (up) and clinical (down) results of ablation experiments.

Table 2. Quantitative results of the ablation experiments on the simulation dataset, in terms of PSNR and SSIM.

Method	PSNR [dB] ↑	SSIM ↑
Prior from CDIP (overfit)	31.070±0.052	0.910±0.001
Prior from CDIP (underfit)	31.139±0.052	0.911±0.001
Prior from noisy PET	31.099±0.047	0.909±0.001
SA-Net input without MR	30.831±0.054	0.901±0.001
UNet replace SA-Net	30.239±0.296	0.894±0.004
CDIP (prior replace MR)	30.273±0.112	0.891±0.002
Proposed	**31.147±0.030**	**0.912±0.001**

parameter capacity, combined with supplementary information from the Bowsher prior, increases its tendency to overfit noise in the noisy PET image. 4) Using the Bowsher prior computed from the noisy PET image (instead of the MR image) to provide anatomical guidance for CDIP increases the network's susceptibility to fitting noise. This indicates that the Bowsher prior cannot provide superior anatomical information compared to MR images.

4 Conclusion and Limitations

In this study, we proposed an unsupervised PET image denoising framework, demonstrating its superior performance on both simulation and clinical PET/MR datasets. The framework innovatively incorporates the Bowsher prior into the PET denoising process, leveraging its excellent image fusion capability to integrate the denoised results from the CDIP method with MR images, which effectively utilizes the anatomical structural information from MR images to enhance PET denoising. Furthermore, the proposed SA-Net significantly improves the quality of the Bowsher prior image and maintains robustness to the CDIP denoising results, thereby addressing CDIP's inherent dependence on training-stopping criteria. Comparative experiments (against Gaussian filtering, NLM,

BM4D, and CDIP) and ablation studies confirm that the proposed framework effectively integrates anatomical information to enhance PET image denoising performance.

While Deep Image Prior (DIP)-based denoising methods demonstrate effectiveness in preserving subtle pathological features during denoising [9,13], anatomical image-guided PET denoising approaches remain susceptible to potential misregistration. Simulation experiments confirm that the proposed method can successfully reconstruct simulated lesions in PET image despite the absence of MR visible pathology (Fig. 2). This capability suggests inherent robustness of the proposed framework against to moderate misregistration. Future work will rigorously evaluate the method's robustness to misregistration and explore its applicability to multi-modal datasets, such as PET/CT.

Acknowledgments. This work was sponsored in part by the National Natural Science Foundation of China (Grant No.U23A20334, U22A2040, 62101488), the Zhejiang Provincial Special Support Program for High-Level Talents (No.2021R52004), and the "Leading Goose" R & D Program of Zhe jiang Province (Grant No.2024C03093).

Disclosure of Interests. The authors have no competing interests to declare that are relevant to the content of this article.

References

1. Bousse, A., Kandarpa, V.S.S., Shi, K., Gong, K., Lee, J.S., Liu, C., Visvikis, D.: A review on low-dose emission tomography post-reconstruction denoising with neural network approaches. IEEE Transactions on Radiation and Plasma Medical Sciences, **8**(4), 333–347 (2024)
2. Bowsher, J.E., Yuan, H., Hedlund, L.W., Turkington, T.G., Akabani, G., Badea, A., Kurylo, W.C., Wheeler, C.T., Cofer, G.P., Dewhirst, M.W., et al.: Utilizing mri information to estimate f18-fdg distributions in rat flank tumors. In: IEEE Symposium Conference Record Nuclear Science 2004. vol. 4, pp. 2488–2492. IEEE (2004)
3. Chen, Z., Peng, C., Guo, W., Xie, L., Wang, S., Zhuge, Q., Wen, C., Feng, Y.: Uncertainty-guided transformer for brain tumor segmentation. Medical & Biological Engineering & Computing, **61**(12), 3289–3301 (2023)
4. Chen, Z., Xie, L., Chen, Y., Zeng, Q., ZhuGe, Q., Shen, J., Wen, C., Feng, Y.: Generative adversarial network based cerebrovascular segmentation for time-of-flight magnetic resonance angiography image. NeuroComputing, **488**, 657–668 (2022)
5. Cui, J., Gong, K., Guo, N., Wu, C., Meng, X., Kim, K., Zheng, K., Wu, Z., Fu, L., Xu, B., et al.: Pet image denoising using unsupervised deep learning. European Journal of Nuclear Medicine and Molecular Imaging, **46**, 2780–2789 (2019)
6. Database, B.S.B.: Brainweb: Simulated brain database. https://brainweb.bic.mni.mcgill.ca/anatomic_normal_20.html (2025), accessed: 2025-02-21
7. Dutta, J., Leahy, R.M., Li, Q.: Non-local means denoising of dynamic pet images. PloS One, **8**(12), e81390 (2013)
8. Fletcher, J.W., Djulbegovic, B., Soares, H.P., Siegel, B.A., Lowe, V.J., Lyman, G.H., Coleman, R.E., Wahl, R., Paschold, J.C., Avril, N., et al.: Recommendations on the use of 18f-fdg pet in oncology. Journal of Nuclear Medicine, **49**(3), 480–508 (2008)

9. Gong, K., Catana, C., Qi, J., Li, Q.: Pet image reconstruction using deep image prior. IEEE transactions on medical imaging **38**(7), 1655–1665 (2018)
10. Hofheinz, F., Langner, J., Beuthien-Baumann, B., Oehme, L., Steinbach, J., Kotzerke, J., van den Hoff, J.: Suitability of bilateral filtering for edge-preserving noise reduction in pet. EJNMMI Research, **1**, 1–9 (2011)
11. Jiang, C., Pan, Y., Liu, M., Ma, L., Zhang, X., Liu, J., Xiong, X., Shen, D.: Pet-diffusion: Unsupervised pet enhancement based on the latent diffusion model. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 3–12. Springer (2023)
12. Kitamura, K., Iida, H., Shidahara, M., Miura, S., Kanno, I.: Noise reduction in pet attenuation correction using non-linear gaussian filters. IEEE Transactions on Nuclear Science, **47**(3), 994–999 (2000)
13. Liu, Q., Tsai, Y.J., Gallezot, J.D., Guo, X., Chen, M.K., Pucar, D., Young, C., Panin, V., Casey, M., Miao, T., et al.: Population-based deep image prior for dynamic pet denoising: A data-driven approach to improve parametric quantification. Medical Image Analysis **95**, 103180 (2024)
14. Luo, Y., Zhou, L., Zhan, B., Fei, Y., Zhou, J., Wang, Y., Shen, D.: Adaptive rectification based adversarial network with spectrum constraint for high-quality pet image synthesis. Medical Image Analysis, **77**, 102335 (2022)
15. Maggioni, M., Katkovnik, V., Egiazarian, K., Foi, A.: Nonlocal transform-domain filter for volumetric data denoising and reconstruction. IEEE Transactions on Image Processing, **22**(1), 119–133 (2012)
16. Mehranian, A., Belzunce, M.A., Prieto, C., Hammers, A., Reader, A.J.: Synergistic pet and sense mr image reconstruction using joint sparsity regularization. IEEE Transactions on Medical Imaging, **37**(1), 20–34 (2017)
17. Onishi, Y., Hashimoto, F., Ote, K., Ohba, H., Ota, R., Yoshikawa, E., Ouchi, Y.: Anatomical-guided attention enhances unsupervised pet image denoising performance. Medical Image Analysis, **74**, 102226 (2021)
18. OpenNeuro: Dataset ds004513, version 1.0.3. https://openneuro.org/datasets/ds004513/versions/1.0.3 (2025), accessed: 2025-02-21
19. Pain, C.D., Egan, G.F., Chen, Z.: Deep learning-based image reconstruction and post-processing methods in positron emission tomography for low-dose imaging and resolution enhancement. European Journal of Nuclear Medicine and Molecular Imaging, **49**(9), 3098–3118 (2022)
20. Schwaiger, M., Ziegler, S., Nekolla, S.G.: Pet/ct: challenge for nuclear cardiology. Journal of Nuclear Medicine, **46**(10), 1664–1678 (2005)
21. Tai, Y.F., Piccini, P.: Applications of positron emission tomography (pet) in neurology. Journal of Neurology, Neurosurgery & Psychiatry, **75**(5), 669–676 (2004)
22. Tauber, C., Stute, S., Chau, M., Spiteri, P., Chalon, S., Guilloteau, D., Buvat, I.: Spatio-temporal diffusion of dynamic pet images. Physics in Medicine & Biology, **56**(20), 6583 (2011)
23. Tustison, N.J., Yassa, M.A., Rizvi, B., Cook, P.A., Holbrook, A.J., Sathishkumar, M.T., Tustison, M.G., Gee, J.C., Stone, J.R., Avants, B.B.: Antsx neuroimaging-derived structural phenotypes of uk biobank. Scientific Reports, **14**(1), 8848 (2024)
24. Vunckx, K., Nuyts, J.: Heuristic modification of an anatomical markov prior improves its performance. In: IEEE Nuclear Science Symposuim & Medical Imaging Conference. pp. 3262–3266. IEEE (2010)
25. Yan, J., Lim, J.C.S., Townsend, D.W.: Mri-guided brain pet image filtering and partial volume correction. Physics in Medicine & Biology, **60**(3), 961 (2015)

26. Zhao, F., Li, D., Luo, R., Liu, M., Jiang, X., Hu, J.: Self-supervised deep learning for joint 3d low-dose pet/ct image denoising. Computers in Biology and Medicine, **165**, 107391 (2023)