# UDPET: Ultra-low Dose PET Imaging Challenge Dataset

Song Xue[1], Hanzhong Wang[2,3], Yizhou Chen[1], Fanxuan Liu[1], Hong Zhu[4], Marco Viscione[1], Rui Guo[2,3], Axel Rominger[1], Biao Li[2,3], and Kuangyu Shi[1](✉)

[1] Department of Nuclear Medicine, Inselspital, Bern University Hospital, University of Bern
[2] Department of Nuclear Medicine, Ruijin Hospital, School of Medicine, Shanghai Jiao Tong University, Shanghai, China
[3] College of Health Science and Technology, Shanghai Jiao Tong University School of Medicine, Shanghai, China
[4] Department of Informatics, Technical University of Munich, Munich, Germany
**kuangyu@unibe.ch**

**Abstract.** Positron emission tomography (PET) is widely recognized as the most sensitive molecular imaging modality, enabling the in vivo visualization of molecular pathways. Despite its exceptional utility, concerns about ionizing radiation exposure have limited its broader application. A recent breakthrough in total-body PET imaging addresses this limitation by significantly increasing geometric coverage and sensitivity. This innovation reduces radiation exposure to levels comparable to the dose received during a transatlantic flight, achieved through advanced computational techniques. To accelerate progress in this field, we have curated a benchmark dataset specifically designed for developing ultra-low dose PET imaging methodologies. This dataset was pivotal in the Ultra-Low Dose PET Imaging Challenge held in 2022, 2023, and 2024. The challenge aimed to foster innovative computational algorithms capable of recovering high-quality imaging from low-dose scans acquired on total-body PET systems. The dataset includes both standard-dose and simulated low-dose total-body PET images from 1,447 patients. These were acquired using Siemens Biograph Vision Quadra PET/CT and United Imaging uExplorer PET/CT scanners. In addition, we have developed a customized evaluation system to assess the performance of algorithms in recovering image quality from low-dose scans. This paper provides a comprehensive description of the benchmark dataset and evaluation framework, aimed at driving future advancements in ultra-low dose PET imaging. The dataset is available at **https://udpet-challenge.github.io**, subject to the completion of a signed Data Transfer Agreement.

**Keywords:** Positron Emission Tomography (PET), Total-body PET, Image Quality Recovery, Dose Reduction.

## 1 Introduction

Positron emission tomography (PET) is a critical imaging modality in clinical routine procedures of oncology [1], neurology [2], and cardiology [3]. However, the use of PET is limited by the potential harm of ionizing radiation dose [4]. Despite following the

ALARA concept (as low as reasonably achievable) [5], which aims to minimize exposure, patients still receive a considerable amount of equivalent dose, typically exceeding 4 mSv [6]. PET imaging quality is highly influenced by the activity of the injected tracer and the resulting radiation dose, such that a reduction in dose leads to a degradation of imaging quality. This issue presents a challenge in striking a balance between maintaining the imaging quality and minimizing patient exposure.

PET scanners have undergone significant technical advancements in recent years, which has reduced radiation exposure without compromising image quality [7]. These advancements have been achieved through improvements in signal measurement and imaging generation, including advancements in scintillator crystals, photodetectors, acquisition electronics, and image reconstruction techniques [8]. Total-body PET systems for human use have been developed and initial clinical studies have been conducted [9]. These systems have significantly increased the geometric coverage to include the entire body, resulting in a sensitivity boost of about 40 times for total-body imaging or 4-5 times for imaging individual organs such as the brain or heart [7]. This increased sensitivity allows for a reduction of radiation exposure by approximately 10 times [10].

In contrast to technical advancements of hardware, computational methods offer a cost-effective alternative solution to improve image quality for PET imaging. Various denoising methods, such as nonlocal means [11] or multi-scale curvelet and wavelet analysis [12], have been developed to reduce the noise in low-dose PET images. Machine learning-based data-driven approaches have also been employed to recover image quality from low-dose PET imaging. These include random-forest-based regression [13], mapping-based sparse representation [14], semi-supervised tripled dictionary learning [15], and multilevel canonical correlation analysis framework [16]. Recently, deep learning techniques have been developed to better predict textural information in molecular imaging, using network architectures such as fully connected convolutional neural network (CNN) [17], encoder-decoder CNN (U-Net) [18], and generative adversarial network (GAN) [19].

Despite advancements in computational methods for image quality recovery, the progress in this field remains less than ideal, primarily due to the dearth of valuable datasets. The main bottlenecks of publicly available datasets encompass:

- Limited total quantity available.
- Lack of diversity in the datasets.
- Absence of datasets for cutting-edge total body PET.
- Absence of standard evaluation system.

To advance computational methods for recovering high-quality images from low-dose scans on total-body PET scanners, we carefully curated a benchmark dataset. This dataset served as the foundation for the Ultra-Low Dose PET Imaging Challenge (UDPET), held in 2022, 2023, and 2024. The challenge aimed to inspire the development of cutting-edge techniques for enhancing image quality in low-dose total-body PET scans. In this paper, we provide a detailed overview of the benchmark dataset and introduce a customized scoring system designed to evaluate the performance of these computational methodologies.

# 2 Materials and Methods

## 2.1 Dataset Description

Our benchmark dataset includes 1447 subjects referred to $^{18}$F-Fluorodeoxyglucose ($^{18}$F-FDG) total-body PET for various neurological or oncological diseases, acquired from Siemens Biograph Vision Quadra (n=433) and United Imaging uExplorer (n=1060) PET/CT scanners. Details of patient demographics is shown in Table 1.

**Table 1.** Patient demographics from different PET/CT scanners.

| PET/CT Scanner | Siemens Biograph Vision Quadra | United Imaging uExplorer |
|---|---|---|
| Tracer | $^{18}$F-FDG | $^{18}$F-FDG |
| Number of patients | 387 | 1060 |
| Total dose (MBq) | 219.2 ± 50.9 | 185.0 ± 77.3 |
| Weight (kg) | 73.0 ± 16.4 | 62.9 ± 11.1 |

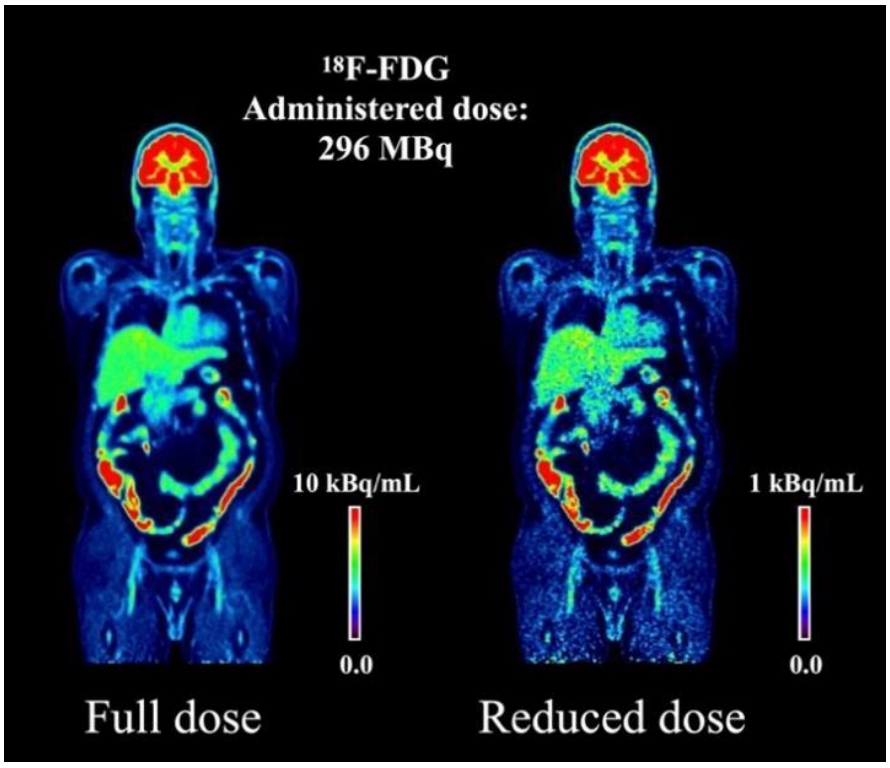


**Fig. 1.** Example of a coronal view of $^{18}$F-FDG PET imaging acquired from Quadra (Biograph Vision Quadra) PET/CT scanner.

Quadra data was acquired from the Department of Nuclear medicine, University of Bern, Switzerland, and uExplorer data was acquired from the Department of Nuclear Medicine, Ruijin Hospital, Shanghai Jiao Tong University School of Medicine, Shanghai, China. All PET data was acquired in list mode where the scanner recorded every detection event individually, including the precise time and position of each detected gamma photon pair, which enabled the simulation of different acquisition time windows by rebinding the data. Quadra PET data was reconstructed using OSEM (Ordered Subsets-Expectation Maximization) with PSF+TOF, 4 iterations and 5 subsets (4i5s), a 2 mm Gaussian filter, and a voxel size of 1.650 × 1.650 × 5.000 mm. In contrast, uExplorer PET data was reconstructed using OSEM with 3D iterative TOF PSF, no Gaussian filter, and a voxel size of 1.667 × 1.667 × 2.886 mm. Normalization, attenuation, scatter and random correction were performed after the reconstruction, but the corresponding CT scans were not provided. Each simulated low-dose PET image, with a specific dose reduction factor (DRF), was reconstructed from counts in a time window sampled at the midpoint of the acquisition, with the corresponding reduced acquisition time, as shown in Fig. 1. Low-dose images were generated for DRFs of 4, 10, 20, 50, and 100, along with full-dose images. Since these low-dose PET images are produced by subsampling a portion of the full scan, they are perfectly aligned with the full-dose PET images.

### 2.2 Data Management

All procedures involving human participants in this study were conducted in compliance with the ethical standards of the institutional and/or national research committee, as well as the 1964 Helsinki Declaration and its subsequent amendments or equivalent ethical guidelines. Informed consent was obtained from all participants included in the study. Access to the dataset was provided upon receipt of the duly signed Data Transfer Agreement (DTA) by the organizers. The dataset is securely stored using cloud storage services: the dataset is hosted on Microsoft OneDrive. This data acquisition process adheres rigorously to the ethical guidelines and regulations stipulated by the pertinent local ethics committees in Switzerland and China.

## 3 Evaluation Metrics

Evaluation of the generated PET image is complex because clinicians focus more on the regions with high uptake values, especially tumors. Here, we proposed a customized comprehensive scoring system with both global and local metrics to evaluate computational methods. The idea behind this system is to inspire the development of new evaluation metrics that consider the unique clinical requirements of PET imaging.

The global metrics include Normalized Root Mean Squared Error (NRMSE), Peak Signal-to-Noise Ratio (PSNR), and Structural Similarity Index Measurement (SSIM), for overall image quality and fidelity assessment. The NRMSE is computed as

$$\text{NRMSE} = \frac{\sqrt{\sum_{i=1}^{n} \frac{\left(y_{true} - y_{pred}\right)^2}{n}}}{max(y_{true}) - min(y_{true})} \quad (1)$$

where $y_{true}$ is the full dose PET and $y_{pred}$ is the predicted PET image, and it measures the overall pixel-wise intensity deviation between these two PET images. The PSNR is defined as

$$\text{PSNR} = 10 \log_{10} \frac{VR^2}{\left\| y_{true} - y_{pred} \right\|_2^2} \quad (2)$$

where $V$ is the total amounts of voxels and $R$ represents the range of the intensity of the full dose PET image, and $\left\| y_{true} - y_{pred} \right\|_2^2$computes the mean squared error between it and the predicted PET image. The SSIM is defined as

$$\text{SSIM}(x, y) = \frac{(2\mu_x\mu_y + c_1)(2\sigma_{xy} + c_2)}{(\mu_x^2 + \mu_y^2 + c_1)(\sigma_x^2 + \sigma_y^2 + c_2)} \quad (3)$$

where$\mu_x$, $\mu_y$are the averages of images full dose PET and predicted PET, $\mu_x^2$ and $\mu_y^2$ are the variances of the two images, $\sigma_{xy}$ is the covariance between the two images, and $c_1$ and $c_2$ are two positive constants to avoid a null denominator. Theoretically, image with lower NRMSE, higher PSNR, and SSIM closer to 1 represent higher image synthesis quality.

For the local metrics, spherical volumes of interest (VOIs) can be delineated within lesions and target organs, such as livers, kidney, and heart. The mean and maximum of standardized uptake value (SUV) of the VOI are extracted and PSNR is computed locally. Some selected radiomics features are then calculated for that region. Absolute percentage error is calculated for these local metrics except PSNR between the generated PET image and its ground truth.

Each subject's score was calculated using the sum of weighted scores for each evaluation metric that has been normalized within the range between 0 and 1. For the scoring weight, 50% of the total weight was assigned to the global metrics, with 40% allocated to NRMSE, 40% to PSNR, and 20% to SSIM. The remaining 50% was assigned to the local metrics, with 60% for SUV parameters, 10% for PSNR, and the final 30% evenly distributed across the selected radiomics parameters. A detailed breakdown of the evaluation metrics and their corresponding weights can be found in Table 2.

**Table 2.** Global and local quantitative metrics for evaluation of generated PET images

| | Metric Name | Intra-group Weights | Inter-group Weights |
|---|---|---|---|
| **Global** | NRMSE | 30% | 30% |
| | PSNR | 30% | |
| | SSIM | 40% | |
| **Local** | $SUV_{mean}$ | 35% | 35% (Targeted organs) + 35% (Lesions) |
| | $SUV_{max}$ | 25% | |
| | PSNR | 10% | |
| | First order: Root Mean Squared | 5% | |
| | First order: 90 Percentile | 5% | |
| | First order: Median | 5% | |
| | GLRLM: High Gray Level Run Emphasis | 5% | |
| | GLSZM: Zone Percentage | 5% | |
| | GLCM: Joint Average | 5% | |

## 4 Discussion and Conclusion

One major challenge in developing AI algorithms is the scarcity of high-quality training data, as highlighted in several studies [20]. While various efforts have been directed toward techniques like fine-tuning [21], and data augmentation [22], it is evident that the inclusion of additional data could significantly enhance the accuracy of these models. In this paper, a first-of-its-kind benchmark dataset was introduced to enhance low-dose total-body PET image quality recovery algorithms. It aims at addressing the key bottleneck of limited datasets in the development of low-dose PET image enhancement algorithms.

The limited diversity in training cases can also hinder the generalizability of AI algorithms. PET imaging, in particular, exhibits significant variability across instrumentation and imaging protocols [23]. These variations include differences in detector capabilities, data correction methods, and system calibration. Therefore, some concerns have been raised regarding the reliability and stability of such AI models when applied to external datasets. Our dataset aims to address this issue by incorporating data from

the only two commercially available total-body PET systems: the United Imaging Healthcare uExplorer and the Siemens Biograph Vision Quadra. By using this dataset, developers can better evaluate the generalizability of their algorithms, especially for computational methods targeting PET instrumentation. Notably, algorithms developed with total-body PET data can potentially be adapted for use with older PET systems, paving the way for advancements over the next decade.

Despite its strengths, the dataset has several limitations. First, the dataset includes only PET images with the $^{18}$F-FDG tracer, which may limit its usage for developing enhancement methods across different radiotracers. Additionally, the absence of accompanying CT scans restricts the potential for developing multimodal methods. Moreover, the dataset may be skewed by demographic or clinical biases depending on the population from which the scans were acquired, potentially limiting its applicability to broader or underrepresented patient groups. These factors should be carefully considered when interpreting results or generalizing findings from models trained on this dataset.

Currently, there is no standardized system for evaluating the quality of PET images generated by AI algorithms. With this work, we aim to spark a discussion about developing such an evaluation system. Assessing PET image quality differs significantly from evaluating natural images, as regions with high uptake values often carry critical clinical importance and should receive greater attention. Our intention is to encourage further exploration in this area, fostering comparisons among developers and driving the optimization of AI methods tailored to the unique challenges of PET imaging.

**Acknowledgments.** This work was supported by National Natural Science Foundation of China (82171975) and Natural Science Foundation of Shanghai (23ZR1439600). All authors declare that they have no known conflicts of interest in terms of competing financial interests or personal relationships that could have an influence or are relevant to the work reported in this paper.

**Disclosure of Interests.** The authors have no competing interests.

## References

1. Rowe, S.P., Pomper, M.G.: Molecular imaging in oncology: Current impact and future directions. *CA Cancer J Clin.* **72**(4), 333–352 (2022). https://doi.org/10.3322/caac.21713
2. Scheltens, P., et al.: Alzheimer's disease. *Lancet.* **397**(10284), 1577–1590 (2021). https://doi.org/10.1016/s0140-6736(20)32205-4
3. Kobayashi, R., Chen, X., Werner, R.A., Lapa, C., Javadi, M.S., Higuchi, T.: New horizons in cardiac innervation imaging: Introduction of novel 18F-labeled PET tracers. *Eur J Nucl Med Mol Imaging.* **44**(13), 2302–2309 (2017). https://doi.org/10.1007/s00259-017-3828-8
4. Knuuti, J., et al.: Risks and benefits of cardiac imaging: An analysis of risks related to imaging for coronary artery disease. *Eur Heart J.* **35**(10), 633–638 (2014). https://doi.org/10.1093/eurheartj/eht512

5. Fargette, C., Shulkin, B., Jha, A., Pacak, K., Taïeb, D.: Clinical utility of nuclear imaging in the evaluation of pediatric adrenal neoplasms. *Front Oncol.* **12**, 1081783 (2022). https://doi.org/10.3389/fonc.2022.1081783
6. Martí-Climent, J.M., et al.: Effective dose estimation for oncological and neurological PET/CT procedures. *EJNMMI Res.* **7**(1), 37 (2017). https://doi.org/10.1186/s13550-017-0272-5
7. Cherry, S.R., Jones, T., Karp, J.S., Qi, J., Moses, W.W., Badawi, R.D.: Total-Body PET: Maximizing sensitivity to create new opportunities for clinical research and patient care. *J Nucl Med.* **59**(1), 3–12 (2018). https://doi.org/10.2967/jnumed.116.184028
8. Jones, T., Townsend, D.: History and future technical innovation in positron emission tomography. *J Med Imaging.* **4**(1), 011013 (2017).
9. Vandenberghe, S., Moskal, P., Karp, J.S.: State of the art in total body PET. *EJNMMI Phys.* **7**(1), 35 (2020). https://doi.org/10.1186/s40658-020-00290-2
10. Alberts, I., et al.: Clinical performance of long axial field of view PET/CT: A head-to-head intra-individual comparison of the Biograph Vision Quadra with the Biograph Vision PET/CT. *Eur J Nucl Med Mol Imaging.* **48**, 2395–2404 (2021).
11. Dutta, J., Leahy, R.M., Li, Q.: Non-local means denoising of dynamic PET images. *PLoS One.* **8**(12), e81390 (2013).
12. Le Pogam, A., Hanzouli, H., Hatt, M., Le Rest, C.C., Visvikis, D.: Denoising of PET images by combining wavelets and curvelets for improved preservation of resolution and quantitation. *Med Image Anal.* **17**(8), 877–891 (2013).
13. Kang, J., Gao, Y., Shi, F., Lalush, D.S., Lin, W., Shen, D.: Prediction of standard-dose brain PET image by using MRI and low-dose brain [18F] FDG PET images. *Med Phys.* **42**(9), 5301–5309 (2015).
14. Wang, Y., et al.: Predicting standard-dose PET image from low-dose PET and multimodal MR images using mapping-based sparse representation. *Phys Med Biol.* **61**(2), 791 (2016).
15. Wang, Y., et al.: Semisupervised tripled dictionary learning for standard-dose PET image prediction using low-dose PET and multimodal MRI. *IEEE Trans Biomed Eng.* **64**(3), 569–579 (2016).
16. An, L., et al.: Multi-level canonical correlation analysis for standard-dose PET image estimation. *IEEE Trans Image Process.* **25**(7), 3303–3315 (2016).
17. Xiang, L., et al.: Deep auto-context convolutional neural networks for standard-dose PET image estimation from low-dose PET/MRI. *Neurocomputing.* **267**, 406–416 (2017).
18. Chen, K.T., et al.: Ultra–low-dose 18F-florbetaben amyloid PET imaging using deep learning with multi-contrast MRI inputs. *Radiology.* **290**(3), 649–656 (2019).
19. Xue, S., et al.: A cross-scanner and cross-tracer deep learning method for the recovery of standard-dose imaging quality from low-dose PET. *Eur J Nucl Med Mol Imaging.* **48**, 1–14 (2021).
20. Arabi, H., AkhavanAllaf, A., Sanaat, A., Shiri, I., Zaidi, H.: The promise of artificial intelligence and deep learning in PET and SPECT imaging. *Phys Med.* **83**, 122–137 (2021).
21. Lu, W., et al.: An investigation of quantitative accuracy for deep learning-based denoising in oncological PET. *Phys Med Biol.* **64**(16), 165019 (2019).
22. Apostolopoulos, I.D., et al.: Automatic classification of solitary pulmonary nodules in PET/CT imaging employing transfer learning techniques. *Med Biol Eng Comput.* **59**(6), 1299–1310 (2021).
23. Yan, J., et al.: Impact of image reconstruction settings on texture features in 18F-FDG PET. *J Nucl Med.* **56**(11), 1667–1673 (2015).