Dose-aware Diffusion Model for 3D Low-dose PET: Multi-institutional Validation with Reader Study and Real Low-dose Data

Huidong Xie[1,2], Weijie Gan[3], Bo Zhou[1,2], Ming-Kai Chen[1,3], Michal Kulon[4], Annemarie Boustani[4], Benjamin A. Spencer[5], Reimund Bayerlein[5], Xiongchao Chen[1,2], Qiong Liu[1,2], Xueqi Guo[1,2], Menghua Xia[1,4], Yinchi Zhou[1,2], Hui Liu[1,4,14], Liang Guo[1,2], Hongyu An[6], Ulugbek S. Kamilov[3,7], Hanzhong Wang[8], Biao Li[8], Axel Rominger[9], Kuangyu Shi[9], Ge Wang[12], Ramsey D. Badawi[5,13], Chi Liu[1,2,4].

***Abstract*— As PET imaging is accompanied by radiation exposure and potentially increased cancer risk, reducing radiation dose in PET scans without compromising the image quality is an important topic. Deep learning (DL) techniques have been investigated for low-dose PET imaging. However, existing models have often resulted in compromised image quality when achieving low-dose PET and have limited generalizability to different image noise-levels, acquisition protocols, patient populations, and hospitals. Recently, diffusion models have emerged as the new state-of-the-art generative model to generate high-quality samples and have demonstrated strong potential for medical imaging tasks. However, for low-dose PET imaging, existing diffusion models failed to generate consistent 3D reconstructions, unable to generalize across varying noise-levels, often produced visually-appealing but distorted image details, and produced images with biased tracer uptake. Here, we develop DDPET-3D, a dose-aware diffusion model for 3D low-dose PET imaging to address these challenges. Collected from 4 medical centers globally with different scanners and clinical protocols, we extensively evaluated the proposed model using a total of 9,783 ^{18}F-FDG studies (1,596 patients) with low-dose/low-count levels ranging from 1% to 50%. With a cross-center, cross-scanner validation, the proposed DDPET-3D demonstrated its potential to generalize to different low-dose levels, different scanners, and different clinical protocols. As confirmed with reader studies performed by nuclear medicine physicians, the proposed method produced superior denoised results that are comparable to or even better than the 100% full-count images as well as previous DL baselines. The presented results show the potential of achieving low-dose PET while maintaining image quality. Lastly, a group of real low-dose scans was also included for evaluation to demonstrate the clinical potential of DDPET-3D.**

***Index Terms*— Enter about five key words or phrases in alphabetical order, separated by commas.**

Corresponding author: Chi Liu.
Emails: {Huidong.Xie; Chi.Liu}@yale.edu
[1]PET Center, Yale University, USA.
[2]Department of Biomedical Engineering, Yale University, USA.
[3]Department of Computer Science, Washington University in St. Louis, USA.
[4]Department of Radiology and Biomedical Imaging, Yale University, USA.
[5]Department of Radiology, University of California Davis, USA.
[6]Department of Radiology, Washington University in St. Louis, USA.
[7]Department of Electrical & Systems Engineering, Washington University in St. Louis, USA.
[8]Department of Nuclear Medicine, Ruijin Hospital, Shanghai Jiao Tong University School of Medicine, China.
[9]Department of Nuclear Medicine, University of Bern, Switzerland.
[12]Department of Biomedical Engineering, Rensselaer Polytechnic Institute, USA.
[13]Department of Biomedical Engineering, University of California Davis, USA.
[14]Department of Engineering Physics, Tsinghua University, China.

I. Introduction

Positron Emission Tomography (PET) is a functional imaging modality widely used in oncology, cardiology, and neurology studies [1]–[3]. Given the growing concern of radiation exposure and potentially increased cancer risks accompanied with PET scans, reducing the PET injection dose is desirable [4]. However, PET image quality is negatively affected by the reduced injection dose and may affect diagnostic performance such as the identification of low-contrast lesions [5]. Therefore, reconstructing high-quality images from noisy input is an important topic. In fact, making the radiation dose as low as reasonably achievable (the ALARA principle) is a commonly accepted practice in clinical settings.

Iterative methods such as Maximum Likelihood Expectation Maximization (MLEM) [6] and Ordered Subset Expectation Maximization (OSEM) [7] are commonly used for PET reconstructions. However, they are vulnerable to noise in low-dose PET data.

Deep learning (DL) has emerged as a new reconstruction and post-reconstruction processing approach for medical imaging tasks [8], and many different DL methods were proposed for low-dose PET image reconstructions/restorations [9]–[20]. However, existing models often tailored to specific hospital and scanner, patient population, acquisition protocol, and image noise level. They typically have limited generalizability across the diverse PET image acquisition settings in reality. Here, we introduced a diffusion-based model [21], [22] to achieve dose-aware 3D PET image denoising.

Recently, diffusion models have become the new state-of-the-art generative models [23], [24]. They are capable of gen-

erating high-quality samples from Gaussian noise input, and have demonstrated strong potential for low-dose PET imaging. For example, using the DDPM framework [25], Gong *et al.* proposed to perform PET image denoising with MRI as prior information for improved image quality [26]. However, PET-MR systems are not widely available and it would be time-consuming to obtain sequential and paired PET-MR images in reality for network training and image reconstructions. Jiang *et al.* [27] adopted the latent diffusion model [28] for unsupervised PET denoising. Moreover, diffusion models have also been proposed for other imaging modalities, such as CT [29] and MRI [30], [31].

However, these previous diffusion works focus on 2D and do not address the 3D imaging problem. This is particularly important for PET imaging as PET is intrinsically a 3D imaging modality. A 3D diffusion model is desirable, however, due to the hardware memory limit, directly extending the diffusion model to 3D would be difficult. There are a few previous works that aim to address the 3D imaging problems of diffusion models. For example, Chung *et al.* [32] proposed to apply a TV penalty term along the z-axis to remove inconsistencies within each reverse sampling step in the diffusion model. On the other hand, Lee *et al.* [33] utilizes 2 pre-trained perpendicular 2D diffusion models to remove inconsistencies between slices. Nonetheless, presented later in this paper, these previous methods failed to produce satisfactory results for 3D low-count PET(e.g, inaccurate organ boundaries in extreme low-dose settings, inconsistent 3D reconstructions, distorted image features in some cases).

Another challenge with PET image denoising is the high variation of image noise levels among patients. The noise level in PET images can be affected by various factors, including: *(1)* Variations of post tracer injection acquisition start time. *(2)* Occasional tracer injection infiltration in some patients, causing the tracer stuck in the arm, resulting in high image noise in the body. *(3)* The variation of patient's weight and associated weight-based tracer injection protocols. *(4)* Different hospitals using scanners with varying sensitivities have different total scan times for patients. All these factors result in variability of noise levels in the reconstructed images. Because of this, a method to adaptively denoise images with varying noise levels is desirable. However, previously proposed methods mentioned above have limited generalizability to different noise levels [19]. A network trained on one noise level often fails to produce high-quality reconstructions on other noise levels [19]. To address this problem, we previously proposed to combine multiple U-net-based [34] sub-networks with varying denoising power to generate optimal results for any input noise levels [19]. However, training multiple sub-networks requires tedious data pre-processing and long training time. The testing time also linearly increases with the number of sub-networks.

Moreover, different from other imaging modalities, PET was developed as a quantitative tool, being recognized as providing an objective, and more accurate measure for prognosis and response monitoring purposes than visual inspection alone [35]. However, our experimental results showed that, although standard diffusion models (DDPM [25], DDIM [36]) produce visually appealing images, they typically fail to maintain accurate quantification and produced images with biased tracer uptake. Specifically, the tracer distribution in the entire image change after diffusion models. This issue is more severe in whole-body PET scans since the tracer uptake could vary significantly among organs.

In this work, we developed a dose-aware diffusion model for 3D PET Imaging (DDPET-3D) to address these limitations. The main methodological contributions of the proposed DDPET-3D framework are: *(1)* We proposed a dose-embedding strategy that allows noise-aware denoising. DDPET-3D can simultaneously denoise 3D PET images with varying low-dose/count levels. *(2)* We proposed a 2.5D diffusion strategy with multiple fixed noise variables to address the 3D inconsistency issue between slices. DDPET-3D maintains a similar memory burden to 2D diffusion models while achieving high-quality reconstructions. *(3)* To achieve a quantitative-accurate reconstruction, we proposed to use a denoised prior in DDPET-3D. The proposed denoised prior also allows the diffusion model to converge within 25 sampling steps. DDPET-3D can denoise 3D PET images within a reasonable time constraint ($\sim$15 mins on a single GPU). Previous diffusion methods using DDPM sampling would require approximately 6 hrs. Further details on network architecture will be provided later in this paper (see Methods).

The ultimate goal of DL methods for PET image denoising is to achieve low-dose PET without compromising image quality and clinical performance. As outlined in Fig. 1(c), evaluated on real-world, large-scale, cross-center, ultra-low-dose 3D PET data, with confirmation from reader studies with nuclear medicine physicians, DDPET-3D demonstrates superior quantitative and qualitative results that are comparable or even better than the 100% full-count data as well as previous baseline methods across varying low-count-levels. DDPET-3D also showed superior generalizability when directly applied to data acquired at a different hospital without further network fine-tunning. We also evaluated the proposed method on a cohort of human studies with real low-dose injection. Presented results show the potential of DDPET-3D to achieve low-dose PET imaging while maintaining image quality.

II. Results

DDPET-3D takes 3D low-count/low-dose images as input and aims to produce the normal-dose counterparts (Fig. 1(a)). In this paper, with a large-scale, cross-center, cross-scanner dataset, we conducted a comprehensive evaluation of the proposed model to demonstrate the clinical potential (Fig 1(c)). We extensively evaluated the proposed DDPET-3D using a total of 9,783 ^{18}F-FDG studies (1,596 patients) with low-count levels ranging from 1% to 50%. 429 patients were used for network training and validation. The remaining 1,167 patients (5,933 low-count/low-dose image volumes) were used for network testing. All the data were acquired using three different commercial scanners from four different hospitals globally in the USA, China, and Switzerland. In addition to the scanner differences, each hospital implements different scan protocols, and has significantly different patient characteristics. Lastly, in addition to synthetic low-count data rebinned from

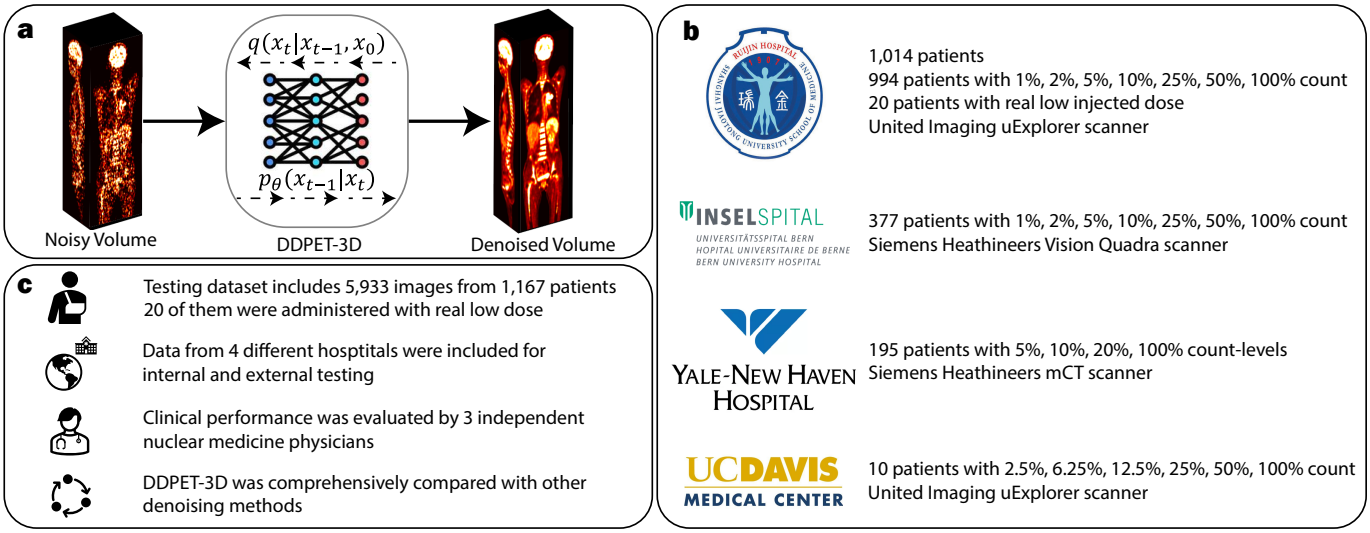

Fig. 1. General overview of the study. a) Model development. The proposed diffusion network, DDPET-3D, was trained in a supervised manner with paired low-count/full-count images. DDPET-3D takes low-count/low-dose 3D volumes as input and outputs the synthesized the corresponding full-count/normal-dose images. Further details on network structure are provided later in this paper (see Methods). b), Dataset overview. 1,596 patient studies obtained at 4 different hospitals with different scanners, clinical protocols, and low-count/low-dose levels were included in this study. c), Model evaluations: 1,167 patients from 4 hospitals were included for internal and external model evaluations. Reader studies were performed to assess the image quality. DDPET-3D was also compared with other denoising methods.

high-count data, where the rates of random events are still the same as those in high-count data, DDPET-3D also presented superior performance on a group of real low-dose PET scans with realistic low random rates. An overview of the dataset is provided in Fig. 1(b).

A. Reader study on three commercial scanners

We performed a reader study to evaluate the network performance. 45 patient studies acquired using three different commercial scanners from three hospitals (15 from each hospital) were randomly selected from the testing dataset and included for the reader study. Three types of commercial scanners are the Siemens mCT scanner, the Siemens Vision Quadra scanner, and the United Imaging uExplorer scanner. Three hospitals are the Yale New Haven Hospital in the US, Shanghai Ruijin Hospital in China, and the University Hospital of Bern in Switzerland.

Low-count images were reconstructed by down-sampling the PET listmode events. For the images acquired on the Siemens Vision Quadra and the United Imaging uExplorer scanners, six low-count levels were available, including 1%, 2%, 5%, 10%, 25%, and 50%. For the images acquired using the Siemens mCT scanner, three low-count levels were available, including 5%, 10%, and 20%. Images were reconstructed using vendors' software from United Imaging Healthcare and Siemens Healthineers. Details on image reconstruction is provided later in this paper (see Methods).

Three nuclear medicine physicians from the Yale-New Haven Hospital participated in the reader study independently. Five 3D images from each patient were provided to the readers simultaneously. The readers were asked to rank the presented images based on the overall image quality. To simulate a clinical reality, images were presented to readers in DICOM format and they were able to choose different display views and adjust the display scale if necessary.

We chose images from the higher 2 low-count levels for the reader study since they would be more practical to achieve in clinical settings. For the patient studies acquired on the Siemens Vision Quadra and the United Imaging uExplorer scanners, five images are the 25% and 50% low-count images, the two corresponding denoised images using the proposed DDPET-3D method, as well as the 100% full-count image. For the patient studies acquired on the Siemens mCT scanner, five images are the 10% and 20% low-count images, the two corresponding denoised images using the proposed DDPET-3D method, as well as the 100% full-count image.

Reader study results are summarized in Fig. 2. For data acquired using the Siemens mCT scanner, all three readers preferred the 10% and 20% denoised results generated by the DDPET-3D method over the 100% full-count images with overall superior average rankings.

For the United Imaging uExplorer data, readers #1 and #2 preferred the denoised images over the 100% full-count images. Reader #3 gave similar rankings between 100% full-count images and the 25% denoised images ($p = 0.67$), while rankings of 50% denoised images is still higher than 100% full-count images with statistical significance ($p < 0.05$).

For the Siemens Vision Quadra data, readers #1 and #2 preferred the denoised images over the 100% full-count images. Reader #3 gave overall lower rankings for the 25% denoised images than the 100% full-count images, while at 50% count-level, reader #3 gave DDPET-3D and 100% full-count images similar rankings.

All three readers agreed that the proposed DDPET-3D can generate denoised results that are better or at least comparable to the 100% full-count images in terms of overall image quality across three types of commercial scanners at different

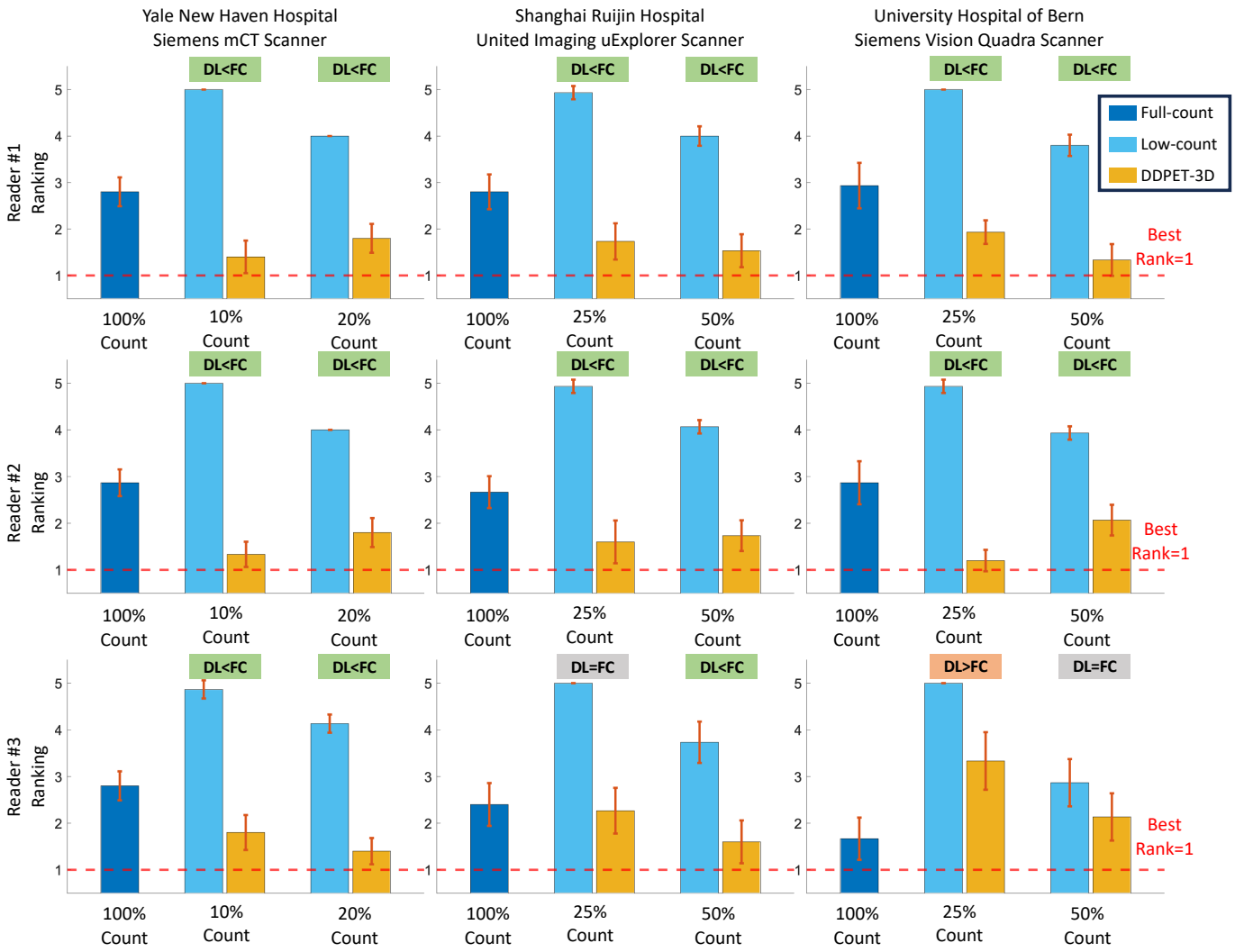

Fig. 2. Comparison of rankings by three readers (#1-3) among images reconstructed with different count-levels and the corresponding denoised images using the proposed DDPET-3D method across three hospitals (Institution #1-3). Error bars indicate the 95% confidence interval. Readers were asked to rank the images based on their overall quality. Red dashed lines indicate the best possible rank (i.e., 1). Compared with 100% full-count images, the text in the box above each plot gives the statistical testing results by a paired t-test at $p < 0.05$. Light red, green, and gray boxes indicate that the DDPET-3D results are worse than, better than, or comparable to the full-count images at 5% significant level, respectively. FC: full-count; DL: the proposed deep learning method (DDPET-3D).

hospitals. Sample denoised images acquired at the Shanghai Ruijin Hospital (United Imaging uExplorer scanner) are presented in Fig. 3. DDPET-3D can generate consistent denoised results across a wide range of low-count levels. DDPET-3D was able to recover fine details in the images (e.g., rib bone in 1% results; myocardium, spine, and femur in 2% results). DDPET-3D also effectively suppressed background noise (e.g., liver in patients shown in Fig. 3), and lesions became clearer in the denoised images (blue arrows in Fig. 3). DDPET-3D effectively removed false-positive hot spots (green arrows in the 25% results) in the images and the true-positive lesions are easier to identified.

B. Reader study on real low-dose scans

While rebinning the PET listmode events to generate low-count data to simulate a low-dose setting is a common practice in the literature [9]–[20], it may affect the random photon rates. To further demonstrate the clinical potential, 20 real low-dose human studies were acquired using a United Imaging uExplorer scanner with an average injected dose of 27.1 ± 5.4MBq (about 10% of full administered dose). Another reader study was performed to assess the image quality. For a more comprehensive comparison, denoised results generated by the Unified Noise-aware Network (UNN) [19] were included in the reader study. UNN was chosen because: (1) similar to DDPET-3D, it also achieves noise-aware denoising; (2) it was among the top 10 winning methods in the 2022 Ultra Low-dose PET Imaging Challenge held at the 2022 IEEE Medical Imaging Conference [1].

Reader study was conducted in the similar way as described previously. Three images were presented to the readers simultaneously and they were asked to rank the images based on their overall image quality. Three images are the real low-

[1]https://ultra-low-dose-pet.grand-challenge.org/leaderboard/

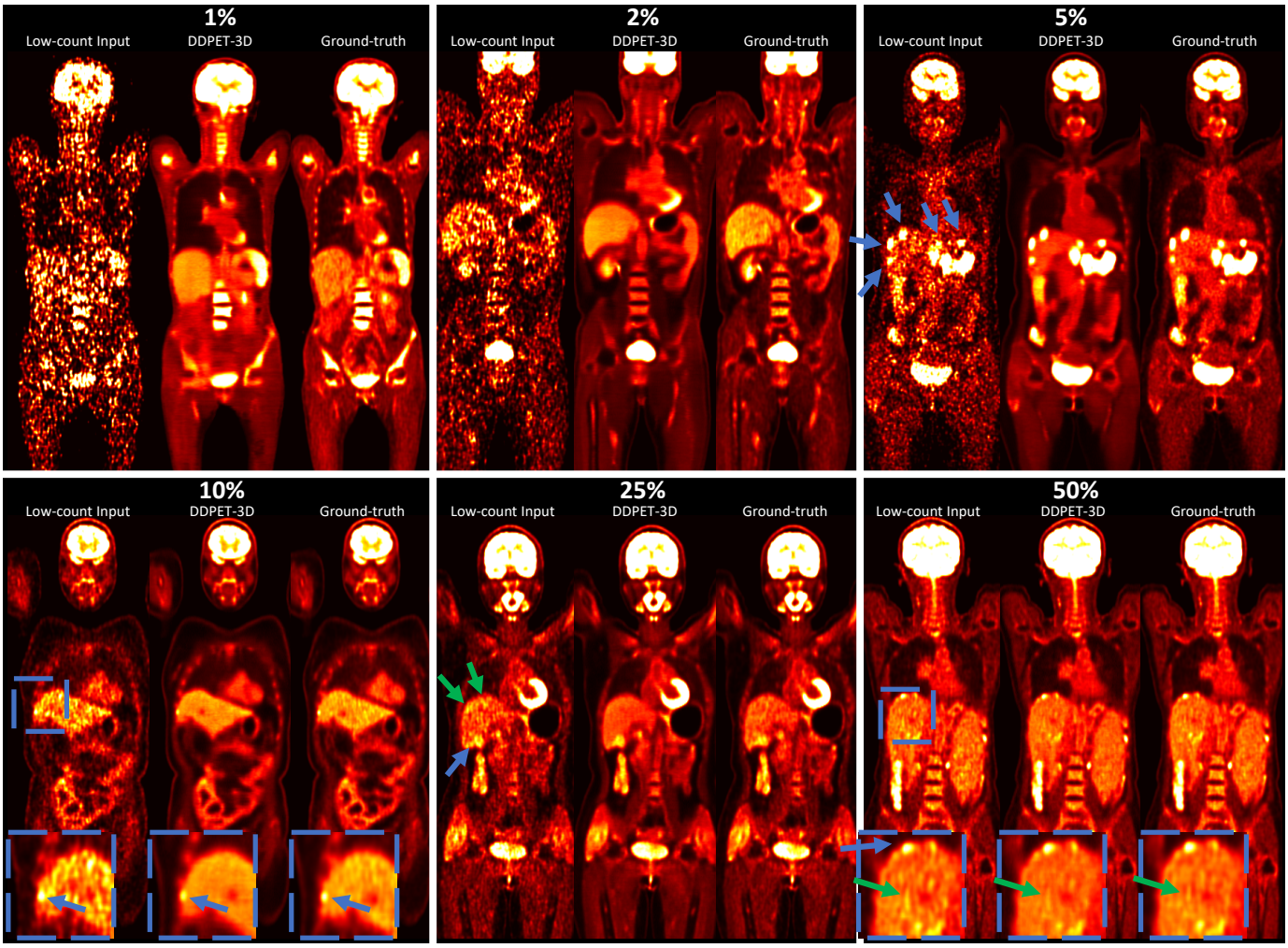

Fig. 3. Sample low-count denoised images reconstructed using the proposed DDPET-3D with different low-count levels. 6 different patients at different low-count levels were randomly chosen from the testing dataset. DDPET-3D can produce consistent denoised results for a wide range of different noise-levels. Ground-truth images were reconstructed using 100% full-count data. Blue arrows point to true-positive lesions or hot spots in different patients. Green arrows point to false-positive hot spots in the images. Blue dashed boxes indicate zoom-in ROI.

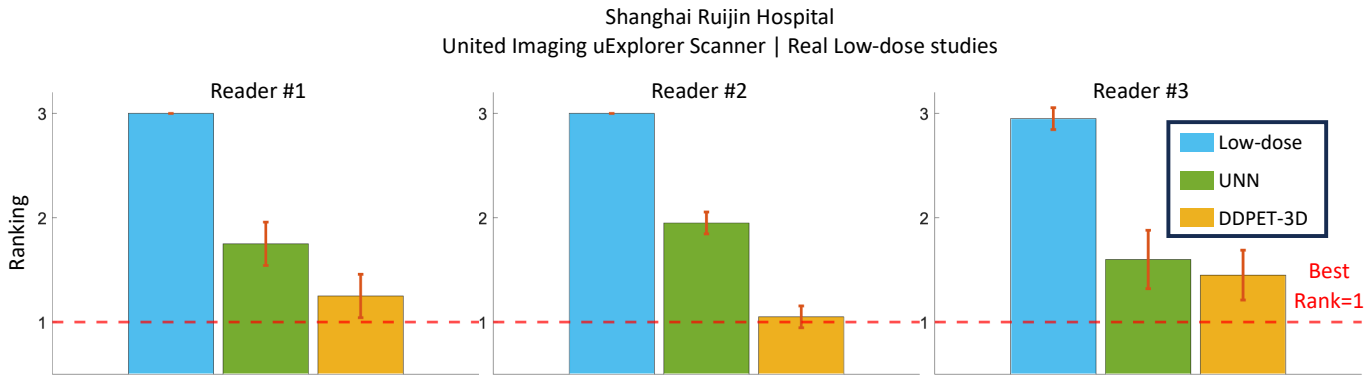

Fig. 4. Comparison of rankings of images with low injected dose as well as the corresponding denoised images using the proposed DDPET-3D and the UNN methods across three readers (#1-3). Error bars indicate the 95% confidence interval. Readers were asked to rank the images based on their overall quality. Red dashed lines indicate the best possible rank (i.e., 1). Compared with low-dose input, both DDPET-3D and UNN produced images with superior average rankings with statistical significance at $\boldsymbol{p < 0.05}$. UNN: Unified Noise-aware Network [19].

dose image, denoised images generated by UNN, and denoised images generated by DDPET-3D. As presented in Fig. 4, all three readers agreed that both DDPET-3D and UNN produced superior denoised results compared with the low-dose images. DDPET-3D produced images with overall better rankings than the UNN with statistical significance ($p < 0.05$) observed in reader #1 and #2. Even though DDPET-3D also had a overall better ranking in reader #3, $p = 0.20$ was observed.

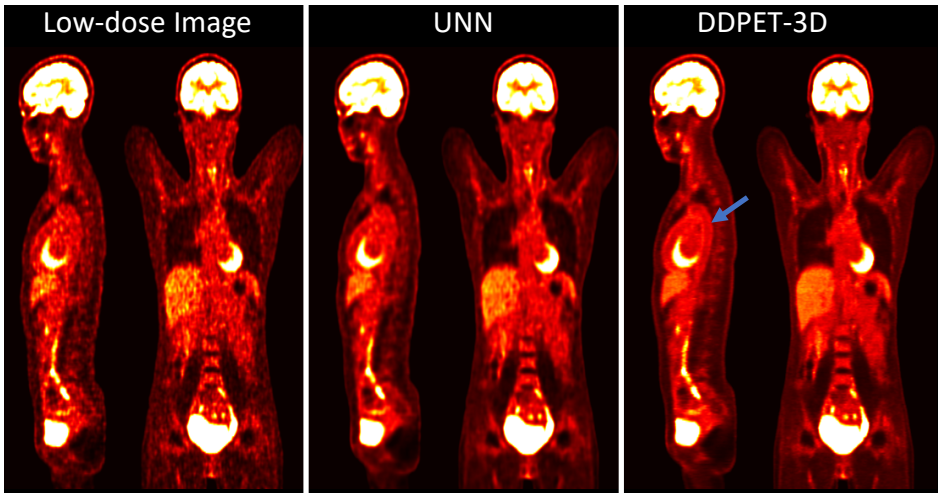

Fig. 5. Sample real low-dose denoised images generated using the proposed DDPET-3D and the UNN [19] methods. Blue arrow points to the aorta wall, which was better reconstructed in DDPET-3D.

Sample real low-dose denoised images are presented in Fig. 5. No obvious lesion was identified in this patient. Compared with other images, DDPET-3D effectively suppressed excessive noise (e.g., in the liver) and produced images with clearer organ boundaries (e.g., the aorta wall pointed by the blue arrow in Fig. 5).

C. Generalizability test in a different hospital

To evaluate the cross-center generalizability of the proposed DDPET-3D, we applied the model trained using the uExplorer data acquired at Shanghai Ruijin Hospital in China to patient studies acquired using the same scanner model at the UC Davis Medical Center in the USA.

In addition to different patient characteristics in different countries, two hospitals also implement different reconstruction parameters, reconstruction voxel sizes, and acquisition protocols (see details in Methods). Images acquired at different hospitals also have varied image contrast due to the differences in scan time and acquisition start time. Shanghai Ruijin Hospital implemented a total of 5 min of scan duration. At UC Davis hospital, they implemented a total of 20 min of scan duration, similar to a typical multi-bed position whole-body scan. Additionally, UC Davis Medical Center also doubled the conventional tracer uptake time from 60 min post-injection to 120 min post-injection for improved tumor to background ratio in oncologic imaging. Shanghai Ruijin Hospital implemented the conventional 60 min tracer post-injection time.

10 patient studies from the UC Davis Medical Center were included in this experiment. Five low-count levels were generated through listmodel rebinning, including 2.5%, 6.25%, 12.5%, 25%, and 50%. Sample denoised results acquired at the UC Davis Medical Center are presented in Fig. 6. The proposed DDPET-3D showed superior generalizability when directly applied to data acquired at a different hospital. Corresponding quantitative measurements are included in Table I.

Another reader study was conducted to evaluate the overall image quality for the 10 UC Davic patients. Similarly, 25% and 50% low-count and denoised images as well as 100% full-count images were included for reader study. As presented in Fig. 7, even though the DDPET-3D denoised images have similar quantitative measurements to low-count images shown in Table I at 25% and 50%, all three readers unanimously gave superior overall rankings to 25% and 50% denoised results than the 100% full-count images. Since the quantitative values shown in Table I were calculated using 100% full-count as reference, they may not correspond to the clinical performance and physicians' preferences. The reader study results presented in Fig. 7 demonstrated the superior generalizability of the proposed DDPET-3D comparing to other neural networks in another hospital with different patient population and clinical protocols. This result suggests that DDPET-3D might be transferable to another medical center without further network fine-tuning, though further investigations are needed.

D. Comparison with other baseline denoising methods

The proposed DDPET-3D was compared with UNN [19], DiffusionMBIR [32], and TPDM [33]. As described previously, UNN was chosen because it was also proposed to achieve noise-aware denoising and was one of the winning method in 2022 Ultra Low-dose PET Imaging Challenge. DiffusionMBIR and TPDM were selected because they were recent works proposed to address the diffusion 3D inconsistencies problem in medical imaging. We also compared with the standard DDIM (Denoising Diffusion Implicit Models) sampling [36]. Detailed implementations of different methods will be provided later in this paper (see Methods).

As presented in Fig. 8, standard DDIM sampling produced images with severe inconsistencies between slices in the sagittal view. Both DiffusionMBIR and TPMD improved the results, but the problem still exist. Proposed DDPET-3D method produced noticeably more consistent reconstructions. Moreover, even though other diffusion models produced visually realistic PET images, they failed to accurately reconstruct the structure of different organs. For example, all other diffusion models produced images with distorted myocardium as shown

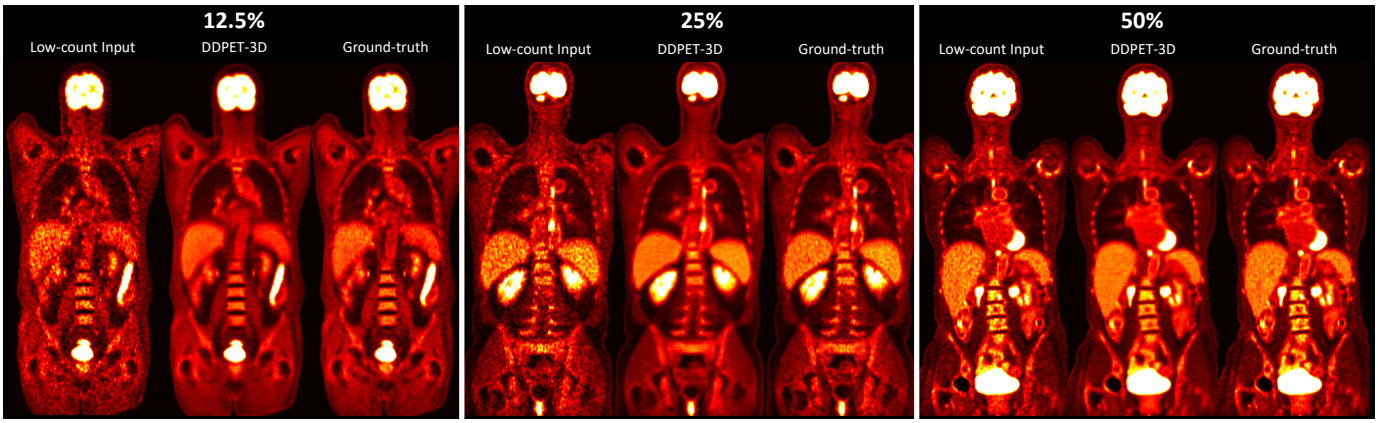

Fig. 6. Low-count denoised images reconstructed using the proposed DDPET-3D with different low-count levels for data acquired at the UC Davis Medical Center. The model trained with Shanghai Ruijin data was directly applied for image reconstructions. DDPET-3D can generalize to different medical centers without further fine-tuning.

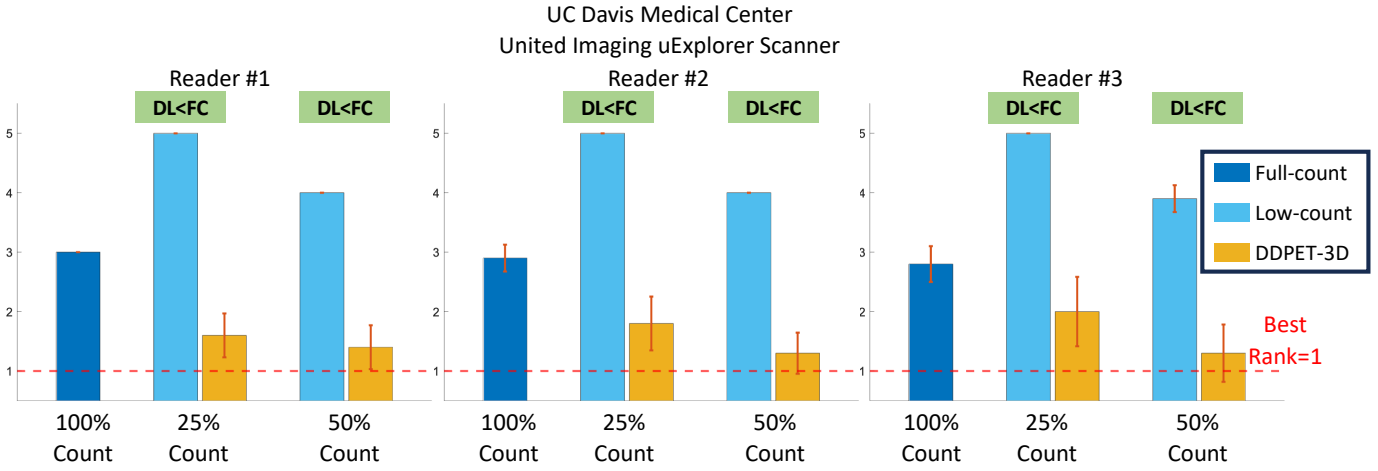

Fig. 7. Low-count denoised images reconstructed using the proposed DDPET-3D with different low-count levels for data acquired using a United Imaging uExplorer scanner at the UC Davis Medical Center. The model trained with another United Imaging uExplorer scanner at Shanghai Ruijin data was directly applied for image reconstructions. DDPET-3D can generalize to different medical centers without further fine-tuning. Error bars indicate the 95% confidence interval. Compared with 100% full-count images, the text in the box above each plot gives the statistical testing results by a paired t-test at $p < 0.05$. Light red, green, and gray boxes indicate that the DDPET-3D results are worse than, better than, or comparable to the full-count images at 5% significant level, respectively. FC: full-count; DL: the proposed deep learning method (DDPET-3D).

in the transverse slice in Fig. 8. The proposed DDPET-3D method also recovered some fine details in the images, such as the hot spot pointed by the blue arrows in Fig. 8, which is either over-smoothed (in UNN) or barely visible (in other diffusion methods).

Another major issue of the diffusion model is inaccurate image quantification. Note that all the presented images were already normalized by the total injected activities. As shown in Fig. 8, the tracer activities in certain organs are completely wrong in other existing diffusion models. For example, the activities in the brain are noticeably lower in DDIM, DiffusionMBIR, and TPDM results; the myocardium is distorted in competing baseline methods. Such differences may affect certain diagnostic tasks such as lesion detection [5]. Even though UNN produced over-smoothed reconstructions, it does not alter the overall tracer activities in different organs. Using UNN output as denoised prior (see details in Methods), the proposed DDPET-3D maintained overall image quantification. Also, using UNN as a denoised prior allows the proposed method to recover some subtle features that are almost invisible in the low-count input (blue arrows in Fig. 8).

Images were quantitatively evaluated using SSIM (Structural Similarity Index), PSNR (Peak signal-to-noise ratio), and RMSE (Root-mean-square error). To facilitate the testing process, we tested all the comparison methods using 20 patients from the entire testing dataset. There were 6 different low-count levels of all the patients, resulting in a total of 120 testing studies. Quantitative results are presented in Table II.

As presented in Table II, the proposed method outperformed other methods in all the 6 low-count levels. When the input count-level increases, the performance of all the methods gradually improves. Both the proposed method and UNN can achieve noise-aware denoising. However, at higher count-levels (25% and 50%), UNN results were even worse than the input. In contrast, the proposed method produced optimal results at all count-levels. Also, due to inaccurate image quantification, other diffusion models produced images with even worse quantitative results compared to low-count inputs

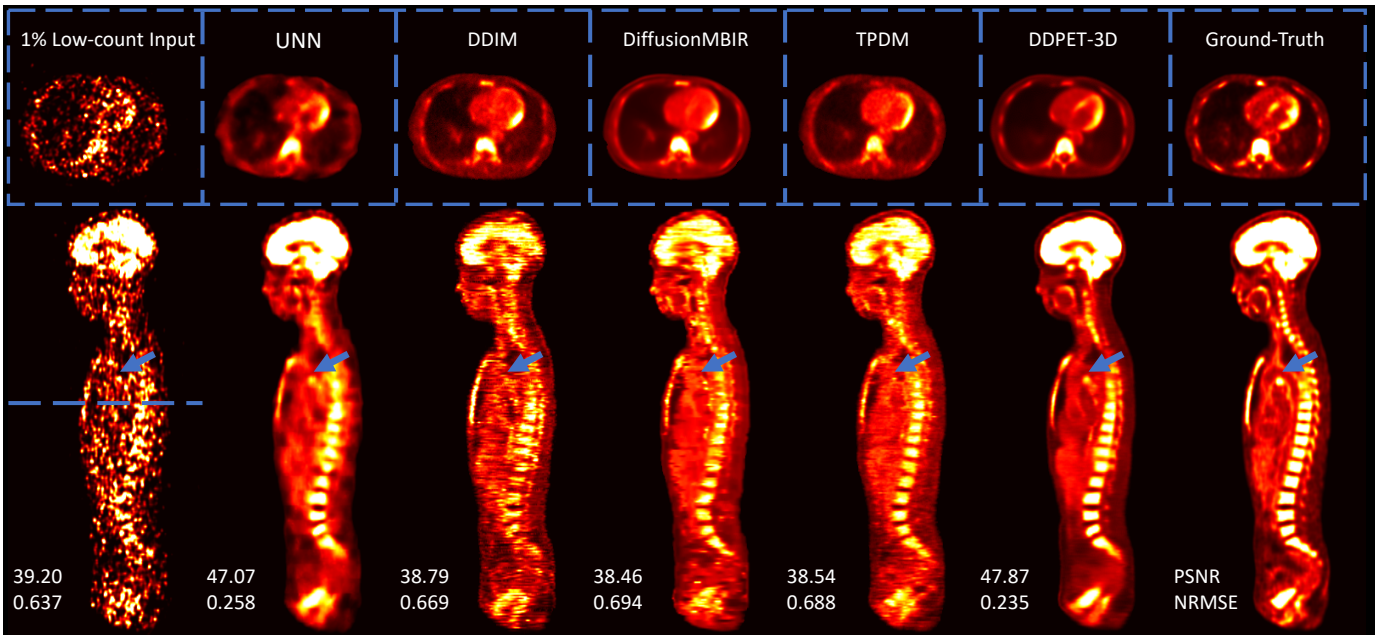

Fig. 8. Comparison between DDPET-3D and other baseline denoising methods. Blue arrows point to a hot spot that was better recovered by the proposed DDPET-3D. Dahsed blue line indicate the location of the transverse slice. All the images were displayed on the same scale and normalized by the total injected activities.

TABLE I

QUANTITATIVE ASSESSMENT FOR DIFFERENT METHODS ON ALL THE TESTING PATIENTS FROM UNITED IMAGING AND SIEMENS SCANNERS. THE MEASUREMENTS WERE OBTAINED BY AVERAGING THE VALUES ON THE TESTING HUMAN STUDIES. THE PROPOSED METHOD CONSISTENTLY PRODUCED PROMISING DENOISED RESULTS REGARDLESS OF INPUT COUNT LEVELS. THE BEST RESULTS AMONG DIFFERENT LOW-COUNT LEVELS ARE MARKED IN RED. NOTE THAT 674 OUT OF 774 TESTING PATIENTS FROM THE SHANGHAI RUIJIN HOSPITAL DO NOT HAVE 50% COUNT RECONSTRUCTIONS.

Shanghai Ruijin Hospital, United Imaging uExplorer Scanner (674 patients × 5 + 100 patients ×6 = 3,970 studies)						
PSNR↑/NRMSE↓/SSIM↑	1% Count Input	2% Count Input	5% Count Input	10% Count Input	25% Count Input	50% Count Input
Input	42.515 / 0.706 / 0.738	47.175 / 0.413 / 0.859	51.734 / 0.245 / 0.939	54.459 / 0.180 / 0.967	57.793 / 0.124 / 0.986	59.805 / 0.103 / 0.995
UNN	50.757 / 0.270 0.946	52.736 / 0.216 / 0.966	54.658 / 0.175 / 0.978	55.763 / 0.155 / 0.983	56.940 / 0.137 / 0.988	57.327 / 0.138 / 0.992
DDPET-3D (proposed)	51.673 / 0.244 / 0.957	53.782 / 0.193 / 0.972	56.035 / 0.151 / 0.981	57.495 / 0.130 / 0.986	59.749 / 0.102 / 0.991	62.516 / 0.080 / 0.997

University Hospital of Bern, Siemens Biograph Vision Quadra Scanner (268 patients × 6 = 1,608 studies)						
PSNR↑/NRMSE↓/SSIM↑	1% Count Input	2% Count Input	5% Count Input	10% Count Input	25% Count Input	50% Count Input
Input	47.376 / 0.523 / 0.856	50.809 / 0.339 / 0.924	54.490 / 0.220 / 0.967	56.769 / 0.174 / 0.982	59.514 / 0.126 / 0.992	61.594 / 0.100 / 0.996
UNN	55.154 / 0.208 / 0.977	56.823 / 0.170 / 0.985	58.470 / 0.141 / 0.989	58.986 / 0.137 / 0.991	60.527 / 0.114 / 0.994	61.877 / 0.098 / 0.995
DDPET-3D (proposed)	55.909 / 0.191 / 0.980	57.630 / 0.154 / 0.987	59.370 / 0.127 / 0.991	60.631 / 0.113 / 0.993	62.229 / 0.093 / 0.995	64.237 / 0.074 / 0.996

Yale New Haven Hospital, Siemens mCT Scanner (95 patients × 3 = 285 studies)			
PSNR↑/NRMSE↓/SSIM↑	5% Count Input	10% Count Input	20% Count Input
Input	44.929 / 0.474 / 0.831	48.040 / 0.331 / 0.900	51.272 / 0.227 / 0.947
UNN	48.975 / 0.291 / 0.933	50.514 / 0.243 / 0.954	51.970 / 0.205 / 0.969
DDPET-3D (proposed)	49.950 / 0.262 / 0.953	51.383 / 0.222 / 0.966	53.056 / 0.184 / 0.976

UC Davis Medical Center, United Imaging uExplorer Scanner (10 patients × 5 = 50 studies)					
PSNR↑/NRMSE↓/SSIM↑	2.5% Count Input	6.25% Count Input	12.5% Count Input	25% Count Input	50% Count Input
Input	45.410 / 0.310 / 0.871	49.331 / 0.197 / 0.939	52.178 / 0.141 / 0.969	54.719 / 0.105 / 0.985	56.860 / 0.082 / 0.994
UNN	49.538 / 0.193 / 0.957	51.222 / 0.157 / 0.971	52.192 / 0.140 / 0.978	52.950 / 0.128 / 0.983	53.370 / 0.122 / 0.986
DDPET-3D (proposed)	50.134 / 0.179 / 0.962	51.964 / 0.145 / 0.973	53.348 / 0.123 / 0.979	54.772 / 0.102 / 0.987	56.607 / 0.085 / 0.991

at higher count-levels.

As presented in Table II, UNN is the second-best method. To comprehensively evaluate the proposed method, We compared the proposed method with UNN for all the 1,137 testing patients across three different scanners. As presented in Table I, DDPET-3D consistently outperformed UNN on different low-count levels.

Several ablated experiments were conducted to demonstrate the effectiveness of different proposed components in DDPET-3D. Results from ablation experiments are provided later in this paper (see Ablation studies).

III. DISCUSSION AND CONCLUSION

We introduced DDPET-3D, a dose-aware diffusion model for 3D low-dose PET imaging. Evaluated on a large-scale, multi-center, and cross-scanner dataset, low-dose/low-count denoised images generated by DDPET-3D performed favourably or comparably relative to the images reconstructed with 100% full-count. Tested on images acquired using three state-of-the-art PET/CT systems from different vendors and

TABLE II

QUANTITATIVE ASSESSMENT FOR DIFFERENT METHODS. THE MEASUREMENTS WERE OBTAINED BY AVERAGING THE VALUES ON THE TESTING HUMAN STUDIES. THE PROPOSED METHOD CONSISTENTLY PRODUCED PROMISING DENOISED RESULTS REGARDLESS OF INPUT COUNT LEVELS. $p < 0.01$ WAS OBSERVED IN ALL GROUPS WHEN COMPARING THE PROPOSED METHOD WITH ALL OTHER METHODS (EXCEPT FOR 50% SSIM VALUES). THE BEST RESULTS AMONG DIFFERENT LOW-COUNT LEVELS ARE MARKED IN RED. BLACK VOXELS WERE REMOVED FOR CALCULATIONS.

Shanghai Ruijin Hospital, United Imaging uExplorer Scanner (20 patients × 6 = 120 studies)						
PSNR↑/NRMSE↓/SSIM↑	1% Count Input	2% Count Input	5% Count Input	10% Count Input	25% Count Input	50% Count Input
Input	44.667 / 0.682 / 0.788	49.260 / 0.404 / 0.895	53.575 / 0.249 / 0.957	56.085 / 0.189 / 0.977	59.346 / 0.131 / 0.991	62.253 / 0.093 / 0.996
UNN	52.034 / 0.293 / 0.953	54.044 / 0.235 0.973	55.838 / 0.194 / 0.982	56.963 / 0.173 / 0.987	58.570 / 0.146 / 0.991	59.885 / 0.126 / 0.994
DDIM	42.550 / 0.852 / 0.904	42.718 / 0.836 / 0.926	42.846 / 0.824 / 0.938	42.942 / 0.815 / 0.944	43.065 / 0.804 / 0.952	43.145 / 0.797 / 0.957
DiffusionMBIR	42.590 / 0.848 / 0.913	42.747 / 0.833 / 0.934	42.868 / 0.822 / 0.944	42.960 / 0.814 / 0.950	43.079 / 0.803 / 0.957	43.156 / 0.796 / 0.961
TPDM	42.691 / 0.839 / 0.906	42.810 / 0.827 / 0.929	42.939 / 0.815 / 0.940	43.051 / 0.805 / 0.946	43.202 / 0.791 / 0.954	43.291 / 0.783 / 0.959
DDPET-3D (proposed)	52.899 / 0.267 / 0.965	54.937 / 0.215 / 0.977	57.119 / 0.171 / 0.985	58.551 / 0.148 / 0.989	60.916 / 0.117 / 0.993	63.804 / 0.088 / 0.996

confirmed with reader studies, DDPET-3D consistently produced superior denoised results. Tested on images acquired at a different hospital, DDPET-3D also demonstrated its potential to be easily transferred to another medical center without further network fine-tuning. In previous literature, about 77% of deep-learning-based healthcare papers are limited to institutional or regionally distributed dataset [37]. Cross-center generalizability would be crucial to implement deep-learning algorithms in clinical practice.

Compared to previous diffusion models which are time-consuming to reconstruct the entire 3D volumes, once the DDPET-3D is fully-trained, the denoising process is efficient and can be finished within 15 min using one NVIDIA A40 GPU. This is sufficiently fast in reality as the clinical PET image reconstruction time is 15-35 min depending on reconstruction parameters [38] on the United Imaging uExplorer scanner.

Compared to most of the previously published PET denoising methods [9]–[18], [20], which learn the mapping from images collected at a specific dose-level to the normal-dose counterparts, the proposed DDPET-3D can be viewed as a major extension that learns to achieve dose-aware denoising by training the network using dataset with a wide range of input count-levels and associating injected dose within the network. Since the noise-level in PET images can be affected by various factors in clinical settings, the ability to generalize to different noise-levels would be beneficial to implement the algorithm for routine clinical use.

Confirmed with reader studies by professional nuclear physicians, the proposed DDPET-3D can provide a similar or better image quality compared with normal-dose images in terms of noise suppression and structural fidelity. DDPET-3D effectively suppressed background noise and false-positive hot spots in certain patients, making true-positive lesion easier to identify.

Because of the above-mentioned benefits, DDPET-3D can be easily integrated with current clinical reconstruction workflow to denoise PET images.

However, this study has several limitations. First, as an overall comparative study, DDPET-3D has not been optimized to a particular body region or patient population. Due to the high variance of tracer uptake in whole-body PET studies, training the network separately for different body regions will help improve the denoising performance. Second, in the cross-center generalizability experiment, DDPET-3D has not been optimized to the specific hospital. In collaboration with a hospital, the proposed DDPET-3D can be further optimized with more in-house patient data. Third, we were not able to obtain a comprehensive clinical conclusion due to the lack of diagnostic comments. In the future, further collaboration with different medical centers would be necessary to evaluate the clinical performance. Lastly, in this study, DDPET-3D was only evaluated on ^{18}F-FDG. Further evaluations on different PET tracers would be helpful to implement the algorithms on various clinical applications. Despite these limitations, our overall results have been encouraging that DDPET-3D is either better or comparable to normal-dose reconstructions.

In conclusion, the proposed DDPET-3D provides better or similar image quality in different low-dose settings compared to images acquired in a normal-dose setting. The experimental results showed that DDPET-3D achieved superior performance compared with previous methods. While this study focuses on PET image denoising, we believe that the proposed method could be easily extended to other 3D reconstruction tasks for different imaging modalities.

IV. METHODS

A. Diffusion models

The general idea of diffusion models is to learn the target data distribution $q(\boldsymbol{x}_0)$ (*i.e.,* full-dose PET images in our case) using neural networks. Once the distribution is learned, we can synthesize a new sample from it. Diffusion models consist of two Markov chains: the forward diffusion process and the learned reverse diffusion process. The forward diffusion process q gradually adds small amount of Gaussian noise to $\boldsymbol{x}_0 \sim q(\boldsymbol{x}_0)$ in each step, until the original image signal is completely destroyed. As defined in [25]

$$\begin{aligned} q(\boldsymbol{x}_{1:T}|\boldsymbol{x}_0) &:= \prod_{t=1}^{T} q(\boldsymbol{x}_t|\boldsymbol{x}_{t-1}) \ , \\ \text{where } q(\boldsymbol{x}_t|\boldsymbol{x}_{t-1}) &:= \mathcal{N}(\boldsymbol{x}_t; \sqrt{1-\beta_t}\boldsymbol{x}_{t-1}, \beta_t\mathbf{I}) \ . \end{aligned} \tag{1}$$

One property of the diffusion process is that, one can sample $\boldsymbol{x}_t$ for any arbitrary time-step t without gradually adding noise to $\boldsymbol{x}_0$. By denoting $\alpha_t := 1-\beta_t$ and $\bar{\alpha}_t := \prod_{s=1}^{t} \alpha_s$, we have

$$\begin{aligned} q(\boldsymbol{x}_t|\boldsymbol{x}_0) &= \mathcal{N}(\boldsymbol{x}_t; \sqrt{(\bar{\alpha}_t)}\boldsymbol{x}_0, (1-\bar{\alpha}_t)\mathbf{I}) \ , \\ \text{and } \boldsymbol{x}_t &= \sqrt{\bar{\alpha}_t}\boldsymbol{x}_0 + \sqrt{1-\bar{\alpha}_t}\boldsymbol{\epsilon} \ , \end{aligned} \tag{2}$$

where $\boldsymbol{\epsilon} \sim \mathcal{N}(0, \mathbf{I})$. The latent $\boldsymbol{x}_T$ is nearly an isotropic Gaussian distribution for a properly designed β_t schedule.

Therefore, one can easily generate a new $\boldsymbol{x}_T$ and then synthesize a $\boldsymbol{x}_0$ by progressively sampling from the reverse posterior $q(\boldsymbol{x}_{t-1}|\boldsymbol{x}_t)$. However, this reverse posterior is tractable only if $\boldsymbol{x}_0$ is known

$$q(\boldsymbol{x}_{t-1}|\boldsymbol{x}_t, \boldsymbol{x}_0) = \mathcal{N}\Big(\boldsymbol{x}_{t-1}; \mu_q(\boldsymbol{x}_t, \boldsymbol{x}_0), \frac{\beta_t(1-\bar{\alpha}_{t-1})}{1-\bar{\alpha}_t}\mathbf{I}\Big) \ , \tag{3}$$

where

$$\mu_q(\boldsymbol{x}_t, \boldsymbol{x}_0) = \frac{\sqrt{\alpha_t}(1-\bar{\alpha}_{t-1})\boldsymbol{x}_t + \sqrt{\bar{\alpha}_{t-1}}(1-\alpha_t)\boldsymbol{x}_0}{1-\bar{\alpha}_t} \ . \tag{4}$$

Note that $q(\boldsymbol{x}_{t-1}|\boldsymbol{x}_t) := q(\boldsymbol{x}_{t-1}|\boldsymbol{x}_t, \boldsymbol{x}_0)$, where the extra conditioning term $\boldsymbol{x}_0$ is superfluous due to the Markov property. DDPM thus proposes to learn a parameterized Gaussian transitions $p_{\boldsymbol{\theta}}(\boldsymbol{x}_{t-1}|\boldsymbol{x}_t)$ to approximate the reverse diffusion posterior (3)

$$p_{\boldsymbol{\theta}}(\boldsymbol{x}_{t-1}|\boldsymbol{x}_t) = \mathcal{N}\Big(\boldsymbol{x}_{t-1}; \mu_{\boldsymbol{\theta}}(\boldsymbol{x}_t, t), \sigma_t^2\mathbf{I}\Big) \ , \tag{5}$$

where

$$\mu_{\boldsymbol{\theta}}(\boldsymbol{x}_t, t) = \frac{1}{\sqrt{\alpha_t}}\Big(\boldsymbol{x}_t - \frac{1-\alpha_t}{\sqrt{1-\bar{\alpha}_t}}\epsilon_{\boldsymbol{\theta}}(\boldsymbol{x}_t, t)\Big) \ . \tag{6}$$

Here, $\epsilon_{\boldsymbol{\theta}}$ denotes a neural network. Through some derivations detailed in [25], the training objective of $\epsilon_{\boldsymbol{\theta}}(\boldsymbol{x}_t, t)$ can be formulated as

$$\mathbb{E}_{\boldsymbol{x},\boldsymbol{\epsilon},t\sim[1,T]}\big[\ \|\boldsymbol{\epsilon} - \epsilon_{\boldsymbol{\theta}}(\boldsymbol{x}_t, t)\|^2\ \big] \ . \tag{7}$$

It is worth noting that the original DDPM [25] set σ_t to a fixed constant value based on the β_t schedule. Recent studies [39], [40] have shown the improved performance by using the learned variance $\sigma_t^2 := \sigma_{\boldsymbol{\theta}}^2(\boldsymbol{x}_t, t)$. We also adopted this approach. To be specific, we have $\sigma_{\boldsymbol{\theta}}(\boldsymbol{x}_t, t) := \exp(v\log\beta_t + (1-v)\log\tilde{\beta}_t)$, where $\tilde{\beta}_t$ refers to the lower bound for the reverse diffusion posterior variances [25], and v denotes the network output. We used a single neural network with two separate output channels to estimate the mean and the variance of (6) jointly. Based on the learned reverse posterior $p_{\boldsymbol{\theta}}(\boldsymbol{x}_{t-1}|\boldsymbol{x}_t)$, the iteration of obtaining a $\boldsymbol{x}_0$ from a $\boldsymbol{x}_T$ can be formulated as follow

$$\boldsymbol{x}_{t-1} = \mu_{\boldsymbol{\theta}}(\boldsymbol{x}_t, t) + \sigma_t\boldsymbol{z}, \ \text{where } \boldsymbol{z}\sim\mathcal{N}(0,\mathbf{I}) \ . \tag{8}$$

B. Conditional PET image denoising

The framework described above only allows unconditional sampling. For the purpose of low-dose PET denoising, instead of generating new samples, the network needs to denoise the images based on input noisy counterparts. This can be achieved by adding additional condition to the neural network. $\epsilon_{\boldsymbol{\theta}}(\boldsymbol{x}_t, t)$ becomes $\epsilon_{\boldsymbol{\theta}}(\boldsymbol{x}_t, t, \boldsymbol{x}_{\text{noisy}})$, where $\boldsymbol{x}_{\text{noisy}}$ denotes the noisy input PET images. Specifically, the input to the neural network becomes a 2-channel input, one is $\boldsymbol{x}_t$, the other one is $\boldsymbol{x}_{\text{noisy}}$. However, we noticed that, this technique results in severe inconsistencies in the reconstructed 3D image volumes. One example is presented in Fig. 8 (DDIM results).

C. Proposed DDPET-3D

The proposed framework is depicted in Fig. 9. The proposed network can observe 3D information from adjacent slices during the training process. This is achieved by using the n neighboring slices as input to predict the central slice. Ablation studies showed that the performance converges at $n = 31$, which was used in this work. One may use 3D convolutional layers to replace the 2D convolutional layers in the diffusion model for an enlarged receptive field and to allow the network to observe 3D structural information. However, we found that it significantly increases memory burden and makes the network difficult to optimize. To alleviate the memory burden and allow faster convergence, we embed neighboring slices in the channel dimension. Specifically, when using only one 2D slice as conditional information ($n = 1$), the input dimension is $N_b \times W \times W \times 2$, where the last dimension is the channel dimension, N_b is input batch size, and W is the width of the images. One channel is conditional 2D slice, and the other one is $\boldsymbol{x}_{t-1}$. When $n = 31$, the input dimension becomes $N_b \times W \times W \times 32$.

Such technique allows the network to observe neighboring slices for 3D image reconstruction with only incremental increase in memory burden. Also, ablation studies showed that such technique also noticeably improves the inconsistency problem for 3D imaging. This is consistent with our expectation. With $n = 1$, when the network tries to predict the next slice, we observe inconsistent reconstructions because both the conditional information and the starting Gaussian noise $p(\boldsymbol{x}_T)$ change. When $n = 31$, the starting Gaussian noise $p(\boldsymbol{x}_T)$ still changes, but all the 31-channel conditional information only shifts a little. The network should produce more consistent output with only subtle changes in the input.

Another reason causing the inconsistency issue is because the starting Gaussian noise of the reverse process is different for different slices. The diffusion sampling strategy starts from a random location in the high-dimensional space (random Gaussian noise), and approximates the data distribution $q(\boldsymbol{x}_0)$ based on the trained neural network, or the score function as described in [41], through many iterations. The denoising problem is ill-posed, and we could generate an infinite number of different denoised images given the same low-count input with different starting Gaussian noise. This is beneficial for generative models to produce a wide variety of different images. However, the stochastic nature of the diffusion model would be problematic for medical image reconstruction problems because we expect neighboring slices to be consistent with each other in the 3D image volume. To address this problem, we proposed to fix the starting Gaussian noise when reconstructing all the slices in the entire 3D volume. Specifically, the starting Gaussian noise at the last time step T, $\boldsymbol{x}_T$ is fixed for all the slices during the sampling process, only the conditional low-count PET images change for different slices.

This approach produced more consistent reconstructions along the z-axis. However, we noticed that by fixing $p(\boldsymbol{x}_T)$, noise-dependent artifacts would propagate to all the slices along the z-axis. One example is presented in the first sub-

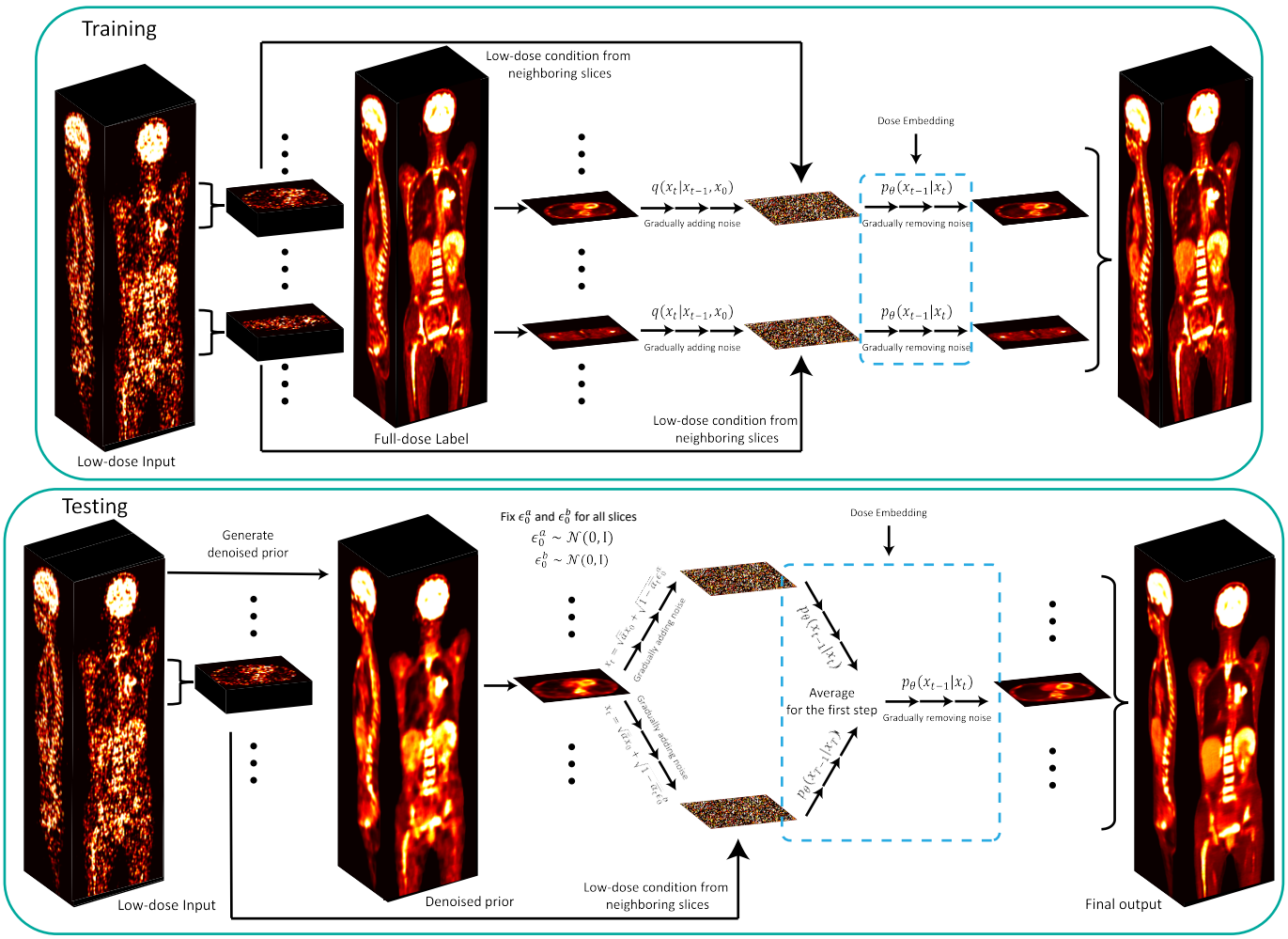

Fig. 9. The training (*top*) and the sampling (*bottom*) pipeline of the proposed DDPET-3D method. The DDPET-3D uses multiple neighboring slices as additional inputs to predict the the central slice, allowing network to observe 3D information during training and testing. The DDPET-3D is also conditioned on the injected dose in order to accommodate inputs with varying noise levels. In sampling, we propose to fix the Gaussian latent for each slice in the 3D volume to address the inconsistency along z-axis. DDPET-3D also use a pre-trained denoiser prior to ensure accurate quantification in sampling.

figure in Fig. 15 (single ϵ image). To address this issue, as presented in Fig. 9, we proposed to initialize 2 different noise variables $\boldsymbol{\epsilon}_0^a$ and $\boldsymbol{\epsilon}_0^b$, and we have 2 different $\boldsymbol{x}_{t-1}$ in the reverse process. Fortunately, having 2 different noise variables does not significantly increase sampling time since we only need to average them at the first reverse step. This technique effectively addresses this issue.

As mentioned in the introduction section, despite the visually appealing results, the diffusion model usually produced images with inaccurate quantification. We noticed that, even though U-net-based methods failed to produce satisfactory results, especially for extremely low-dose settings, they typically maintained overall image quantification much better than diffusion models. To take advantage of that, instead of starting from random Gaussian noise, we first generate a denoised prior using a pre-trained U-net-based network, and then add Gaussian noise to it based on Equation 2. This added-noise denoised prior were used as the starting point of the reverse process. We adapted the Unified Noise-aware Network (UNN) proposed by Xie *et al.* [19] to generate the denoised priors.

For injected dose embedding, the values of the administered dose were first converted to Becquerel (Bq). Both $\sin$ and $\cos$ functions were then used to encode the injected dose and add with the diffusion time-steps (i.e., $t + \sin(\text{dose}) + \cos(\text{dose})$, where t is the diffusion time-step). Time-step t was also encoded using $\sin$ and $\cos$ functions. The encoded values were then fed into 2 linear layers to generate the embedding with Sigmoid linear unit (SiLU) in between.

For diffusion sampling, DDIM sampling enables faster convergence. But we noticed that using DDIM alone would tend to produce over-smoothed images. To alleviate this issue, adapted from the DDIM paper [36], we proposed to have an interpolated sampling between DDIM and DDPM. We inserted DDPM samplings every 5 steps in between DDIM samplings to prevent the images from becoming over-smoothed. Algorithm 1 displays the complete training procedure of the proposed method. Algorithm 2 displays the complete sampling procedure of the proposed method.

In summary, compared to previous methods, DDPET-3D has the following contributions: *(1)* 3D reconstructions with

Algorithm 1 Training

Repeat
$\boldsymbol{x}_0 \sim q(\boldsymbol{x}_0)$ ▷ Sample single-slice full-dose data
$\boldsymbol{x}_{\text{noisy}} \sim q(\boldsymbol{x}_{\text{noisy}})$ ▷ Sample multi-slice low-dose data
$\text{dose}_{\boldsymbol{x}_{\text{noisy}}}$ ▷ Injected dose of $\boldsymbol{x}_{\text{noisy}}$
$t \sim \text{Uniform}(1, ..., \text{T})$ ▷ Sample diffusion time-step
$\boldsymbol{\epsilon} \sim \mathcal{N}(0, \mathbf{I})$
$\nabla_\theta ||\boldsymbol{\epsilon} - \epsilon_{\boldsymbol{\theta}}(\sqrt{\bar{\alpha}_t}\boldsymbol{x}_0 + \sqrt{1-\bar{\alpha}_t}\boldsymbol{\epsilon}, t, \boldsymbol{x}_{\text{noisy}}, \text{dose}_{\boldsymbol{x}_{\text{noise}}})$
▷ Take gradient descent step
Until convergence

Algorithm 2 Testing

$\boldsymbol{x}_{\text{noisy}} \sim q(\boldsymbol{x}_{\text{noisy}})$ ▷ Get entire low-dose image volume
$\text{dose}_{\boldsymbol{x}_{\text{noisy}}}$ ▷ Injected dose of $\boldsymbol{x}_{\text{noisy}}$
$\boldsymbol{x}_{\text{prior}} = \text{UNN}(\boldsymbol{x}_{\text{noisy}})$ ▷ Generate denoised prior using a pre-trained network (UNN) [19]
$\boldsymbol{x}_{\text{recon}} \leftarrow \text{zeros}(\boldsymbol{x}_{\text{noisy}})$ ▷ Initialize $\boldsymbol{x}_{\text{recon}}$
$\boldsymbol{\epsilon}_0^a \sim \mathcal{N}(0, \mathbf{I})$, $\boldsymbol{\epsilon}_0^b \sim \mathcal{N}(0, \mathbf{I})$ ▷ Obtain 2 noise variables

while $s = 1, ..., S$ **do** ▷ Total number of slices
$\boldsymbol{x}_{\text{noisy}}[s]$ ▷ Conditioned multi-slice low-dose input
$\boldsymbol{x}_{\text{prior}}[s]$ ▷ Conditioned single-slice denoised prior
$\boldsymbol{x}_{T'}^a = \sqrt{\bar{\alpha}_{T'}}\boldsymbol{x}_{\text{prior}}[s] + \sqrt{1-\bar{\alpha}_{T'}}\boldsymbol{\epsilon}_0^a$
$\boldsymbol{x}_{T'}^b = \sqrt{\bar{\alpha}_{T'}}\boldsymbol{x}_{\text{prior}}[s] + \sqrt{1-\bar{\alpha}_{T'}}\boldsymbol{\epsilon}_0^b$

while $t = T', ..., 1$ **do**
$z \sim \mathcal{N}(0, \mathbf{I})$
$\boldsymbol{x}_{t-1}^a = \text{Sampler}(\boldsymbol{x}_t^a, \boldsymbol{x}_{\text{noisy}}[s], \text{dose}_{\boldsymbol{x}_{\text{noisy}}})$
if $t == T'$ **then**
$\boldsymbol{x}_{t-1}^b = \text{Sampler}(\boldsymbol{x}_t^b, \boldsymbol{x}_{\text{noisy}}[s], \text{dose}_{\boldsymbol{x}_{\text{noisy}}})$
$\boldsymbol{x}_{t-1}^a = (\boldsymbol{x}_{t-1}^a + \boldsymbol{x}_{t-1}^b)/2$
end if
end while
$\boldsymbol{x}_{\text{recon}}[s] = \boldsymbol{x}_1^a$

end while
Return $\boldsymbol{x}_{\text{recon}}$

fine details and addresses the 3D inconsistency issue in previous diffusion models. This was achieved by using multiple neighboring slices as conditional information and enforcing the same Gaussian latent for all slices in sampling. *(2)* DDPET-3D maintains accurate image quantification by using a pre-trained denoised prior in sampling. *(3)* DDPET-3D achieves dose-aware denoising. It can be generalized to different low-count levels. *(4)* With all the proposed strategies, DDPET-3D converges within 25 diffusion steps, allowing fast reconstructions. It takes roughly 15 minutes to reconstruct the entire 3D volume, while DDPM sampling takes about 6 hours.

D. PET data acquisitions and reconstructions

The collection of the patient datasets in each medical center was approved by the institutional review board (IRB) at each institution. All data in this study were de-identified prior to model training, testing and reader studies.

Shanghai Ruijin Hospital Dataset: 994 subjects with ^{18}F-FDG tracer were included in this dataset. All data were acquired using a United Imaging Healthcare uExplorer total-body PET/CT system. Images were reconstructed using the OSEM algorithm with 4 iterations and 20 subsets. A 5mm FWHM Gaussian filter was applied after reconstructions. The reconstruction matrix size was $673 \times 360 \times 360$ with a $2.89 \times 1.67 \times 1.67\text{mm}^3$ voxel size. We randomly selected 210 subjects for training, 10 subjects for validation, and 774 subjects for testing. The average full administered activity (in Megabecquerel) was 217.4 ± 53.3 MBq with scan duration of 5 min and post-injection tracer uptake time of 60 min. 1%, 2%, 5%, 10%, 25%, and 50% low-count levels were available through listmode rebinning. 50% low-count images were available only for 320 subjects. **Real Low-dose Dataset:** 20 real low-dose scans were also acquired at Shanghai Ruijin Hospital with average administered activity of 27.1 ± 5.4 MBq.

University Hospital of Bern Dataset: 377 subjects with the ^{18}F-FDG tracer were included in this dataset. All data were acquired using a Siemens Biograph Vision Quadra total-body PET/CT system. Images were reconstructed using OSEM with 6 iterations and 5 subsets. A 5mm FWHM Gaussian filter was applied after reconstructions. The reconstruction matrix size was $644 \times 440 \times 440$ with a $1.65 \times 1.65 \times 1.65\text{mm}^3$ voxel size. We randomly selected 99 subjects for training, 10 subjects for validation, and 268 subjects for testing. The average full administered activity was 218.7 ± 49.4 MBq with scan duration of 10 min and post-injection tracer uptake time of 30 min. 1%, 2%, 5%, 10%, 25%, and 50% low-count levels were available through listmode rebinning.

Yale-New Haven Hospital Dataset: 195 subjects with the ^{18}F-FDG tracer were included in this dataset. All data were acquired using a Siemens Biograph mCT PET/CT system. The whole-body scan protocol with continuous-bed motion scanning was used. Images were reconstructed using OSEM with 2 iterations and 21 subsets. A 5mm FWHM Gaussian filter was applied after reconstructions. The voxel size of the reconstructed image was $2.04 \times 2.04 \times 2.03\text{mm}^3$. We randomly selected 90 subjects for training, 10 subjects for validation, and 95 subjects for testing. The image size was 400×400 in the transverse plane and varied in the axial direction depending on patient height. The average full administered activity was 256.3 ± 16.2MBq with scan duration of 5 min every bed position and post-injection tracer uptake time of 60 min. 5%, 10%, and 20% low-count levels were available through listmode rebinning.

UC Davis Medical Center Dataset: 10 subjects with the ^{18}F-FDG tracer were included in this dataset. All data were acquired using a United Imaging Healthcare uExplorer total-body PET/CT system. Images were reconstructed using the OSEM algorithm with 4 iterations and 20 subsets. No post-smoothing was applied. The reconstruction matrix size was $828 \times 256 \times 256$ with a $2.34 \times 2.34 \times 2.34\text{mm}^3$ voxel size. The average full administered activity was 300.6 ± 9.0 MBq with scan duration of 20 min and post-injection tracer uptake time of 120 min. 2.5%, 6.25%, 12.5%, 25%, and 50% low-count levels were available through listmode rebinning. Images were interpolated to match the voxel size of Shanghai Ruijin

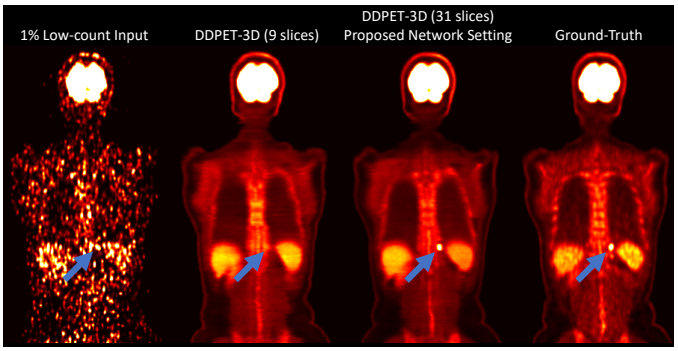

Fig. 10. Images reconstructed using DDPET-3D with different numbers of conditional slices. Note that all the images were normalized by the total injected activities. Blue arrows point to a possible lesion that is better recovered by DDPET-3D with more conditional slices.

Hospital data when applying the trained network.

All the images were reconstructed using vendors' software provided by United Imaging Healthcare and Siemens Healthineers. All the reconstruction parameters are the same as those used in clinical settings in the respective hospitals.

V. Ablation Studies

We performed several ablated experiments to demonstrate the effectiveness of different proposed components in the proposed DDPET-3D.

Impact of Number of Conditional Slices: We evaluated the DDPET-3D with different numbers of neighboring conditional slices. Three variants of DDPET-3D networks were trained using 9, 21, and 41 neighboring slices as conditional information. These networks are denoted as "DDPET-3D (n slice)", where n is the number of conditioned neighboring slices. Results showed that $n = 31$ performed the best in most quantitative metrics. Therefore, $n = 31$ was used in this paper. Quantitative results are presented in Table III. As presented in Fig. 10, using more conditioned slices helped the network to recover some subtle details in the images (blue arrows in Fig. 10).

For comparison, we also extended the comparison diffusion models (DiffusionMBIR, TPDM, and DDIM), and retrained them using neighboring slices as conditional information (denoted as DiffusionMBIR+2.5D, TPDM+2.5D, and DDIM+2.5D). Sample reconstruction results are presented in Fig. 11. Simply adding neighboring slices as conditional information does not necessarily lead to a better performance in other baseline models. Using the same testing patients in Table II, corresponding quantitative evaluations are presented in Table IV.

Impact of Denoised Prior: We tested the DDPET-3D without the denoised prior during the sampling process to demonstrate the improvement in image quantification using the proposed denoised prior. This network is denoted as "DDPET-3D (no prior)" in Table III. As presented in Fig. 12, we can see that diffusion models produced images with inaccurate tracer activities in different organs without the proposed denoised prior (especially in the brain and liver). Note that the images were already normalized by the total injected activities of the entire 3D volume. Also, without the denoised prior, many details in the images were not able to be recovered, as presented in the brain slice in Fig. 12. In addition, we implemented the denoised prior in other competing methods for comparison, so that all the methods have the same starting point in the diffusion process. As shown in Fig. 13, DDPET-3D still outperformed other competing methods, demonstrating the effectiveness of other proposed techniques in DDPET-3D. With the denoised prior, DiffusionMBIR and TPDM produced images with better PSNR and RMSE measurements than the networks without it, but inconsistencies between slices still exist. Lastly, as shown in Fig. 14, without the denoised prior, DDPET-3D still produced more consistent reconstructions compared with other methods. Note that the denoised prior is one of the key contributions proposed in this work.

Impact of Fixing Noise Variables: We analyzed the DDPET-3D without and with fixing the 2 noise variables to demonstrate its effectiveness in producing consistent 3D reconstructions. Specifically, ϵ_0^a and ϵ_0^b are sampled from $\mathcal{N}(0, \mathbf{I})$ for every slice in the 3D image volume, instead of fixing them for all slices. This network is denoted as "DDPET-3D (no fix ϵ)" in Table III. Although the differences in quantitative evaluations without and with fixing noise variables are small, and not fixing noise even led to better quantitative measurements in certain cases, we noticed significant improvements in visual quality, as presented in Fig. 15. Not fixing noise variables produced images with inconsistent slices and unclear organ boundaries, which are unfavorable in clinical settings.

Impact of Dose Embedding: We tested the DDPET-3D without dose embedding, which is denoted as "DDPET-3D (no dose)" in Table III. Experimental results showed that DDPET-3D with dose embedding produced images with better quantitative results.

Acknowledgment

This work was supported by the National Institute of Health (NIH) under grant numbers R01EB025468 and R01CA224140. Parts of the data used in the preparation of this article were obtained from the University of Bern, Department of Nuclear Medicine and School of Medicine, Ruijin Hospital. As such, the investigators contributed to the design and implementation of DATA and/or provided data but did not participate in the analysis or writing of this report. A complete listing of investigators can be found at: https://ultra-low-dose-pet.grand-challenge.org/Description/

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.

References

[1] E. M. Rohren, T. G. Turkington, and R. E. Coleman, "Clinical applications of pet in oncology," *Radiology*, vol. 231, no. 2, pp. 305–332, 2004.

[2] M. Schwaiger, S. Ziegler, and S. G. Nekolla, "Pet/ct: challenge for nuclear cardiology," *Journal of Nuclear Medicine*, vol. 46, no. 10, pp. 1664–1678, 2005.

TABLE III

QUANTITATIVE ASSESSMENT FOR DIFFERENT ABLATION STUDIES. THE BEST RESULTS AMONG DIFFERENT LOW-COUNT LEVELS ARE MARKED IN RED. THE SECOND BEST RESULTS ARE MARKED IN BLUE. BLACK VOXELS WERE REMOVED FOR CALCULATIONS.

United Imaging uExplorer Scanner (20 patients × 6 = 120 studies)						
PSNR↑/NRMSE↓/SSIM↑	1% Count Input	2% Count Input	5% Count Input	10% Count Input	25% Count Input	50% Count Input
DDPET-3D (9 slices)	52.563 / 0.277 / 0.964	54.666 / 0.215 / 0.977	56.194 / 0.162 / 0.984	57.688 / 0.137 / 0.988	60.011 / 0.106 / 0.993	62.267 / 0.082 / 0.995
DDPET-3D (21 slices)	52.502 / 0.278 / 0.960	54.509 / 0.224 / 0.974	56.574 / 0.181 / 0.981	57.949 / 0.157 / 0.985	60.137 / 0.125 / 0.990	62.338 / 0.100 / 0.994
DDPET-3D (41 slices)	51.866 / 0.262 / 0.964	53.926 / 0.208 / 0.977	56.147 / 0.162 / 0.985	57.618 / 0.138 / 0.989	59.948 / 0.107 / 0.993	62.438 / 0.081 / 0.995
DDPET-3D (no prior)	41.497 / 0.961 / 0.868	41.638 / 0.945 / 0.898	41.704 / 0.938 / 0.913	41.715 / 0.937 / 0.920	41.747 / 0.934 / 0.927	41.767 / 0.932 / 0.932
DDPET-3D (no fix ϵ)	52.991 / 0.264 / 0.961	55.049 / 0.212 / 0.973	57.225 / 0.170 / 0.980	58.695 / 0.146 / 0.983	61.026 / 0.116 / 0.988	63.916 / 0.087 / 0.991
DDPET-3D (no dose)	52.817 / 0.269 / 0.960	54.822 / 0.217 / 0.972	56.903 / 0.175 / 0.979	58.259 / 0.152 / 0.983	60.360 / 0.123 / 0.987	62.871 / 0.096 / 0.990
DDPET-3D (single ϵ)	52.289 / 0.285 / 0.959	54.261 / 0.231 / 0.974	56.324 / 0.185 / 0.983	57.701 / 0.161 / 0.987	59.917 / 0.128 / 0.992	62.548 / 0.098 / 0.995
DDPET-3D (proposed)	52.899 / 0.267 / 0.965	54.937 / 0.215 / 0.977	57.119 / 0.171 / 0.985	58.551 / 0.148 / 0.989	60.916 / 0.117 / 0.993	63.804 / 0.088 / 0.996

TABLE IV

QUANTITATIVE ASSESSMENT FOR DIFFERENT METHODS. THE MEASUREMENTS WERE OBTAINED BY AVERAGING THE VALUES ON THE TESTING HUMAN STUDIES. THE PROPOSED METHOD CONSISTENTLY PRODUCED PROMISING DENOISED RESULTS REGARDLESS OF INPUT COUNT LEVELS. $p < 0.01$ WAS OBSERVED IN ALL GROUPS WHEN COMPARING THE PROPOSED METHOD WITH ALL OTHER METHODS (EXCEPT FOR 50% SSIM VALUES). THE BEST RESULTS AMONG DIFFERENT LOW-COUNT LEVELS ARE MARKED IN RED. BLACK VOXELS WERE REMOVED FOR CALCULATIONS.

United Imaging uExplorer Scanner (20 patients × 6 = 120 studies)						
PSNR↑/NRMSE↓/SSIM↑	1% Count Input	2% Count Input	5% Count Input	10% Count Input	25% Count Input	50% Count Input
Input	44.667 / 0.682 / 0.788	49.260 / 0.404 / 0.895	53.575 / 0.249 / 0.957	56.085 / 0.189 / 0.977	59.346 / 0.131 / 0.991	62.253 / 0.093 / 0.996
UNN	52.034 / 0.293 / 0.953	54.044 / 0.235 / 0.973	55.838 / 0.194 / 0.982	56.963 / 0.173 / 0.987	58.570 / 0.146 / 0.991	59.885 / 0.126 / 0.994
DDIM50	42.550 / 0.852 / 0.904	42.718 / 0.836 / 0.926	42.846 / 0.824 / 0.938	42.942 / 0.815 / 0.944	43.065 / 0.804 / 0.952	43.145 / 0.797 / 0.957
DDIM50+2.5D	43.636 / 0.754 / 0.910	43.851 / 0.736 / 0.924	43.856 / 0.735 / 0.930	43.829 / 0.738 / 0.932	43.805 / 0.740 / 0.936	43.820 / 0.738 / 0.938
DiffusionMBIR	42.590 / 0.848 / 0.913	42.747 / 0.833 / 0.934	42.868 / 0.822 / 0.944	42.960 / 0.814 / 0.950	43.079 / 0.803 / 0.957	43.156 / 0.796 / 0.961
DiffusionMBIR+2.5D	42.615 / 0.846 / 0.913	42.834 / 0.825 / 0.935	42.892 / 0.819 / 0.945	42.911 / 0.818 / 0.948	42.931 / 0.816 / 0.952	42.950 / 0.814 / 0.955
TPDM	42.691 / 0.839 / 0.906	42.810 / 0.827 / 0.929	42.939 / 0.815 / 0.940	43.051 / 0.805 / 0.946	43.202 / 0.791 / 0.954	43.291 / 0.783 / 0.959
TPDM+2.5D	42.334 / 0.873 / 0.887	42.458 / 0.861 / 0.912	42.483 / 0.859 / 0.923	42.499 / 0.857 / 0.927	42.528 / 0.854 / 0.932	42.536 / 0.853 / 0.935
DDPET-3D (proposed)	52.899 / 0.267 / 0.965	54.937 / 0.215 / 0.977	57.119 / 0.171 / 0.985	58.551 / 0.148 / 0.989	60.916 / 0.117 / 0.993	63.804 / 0.088 / 0.996

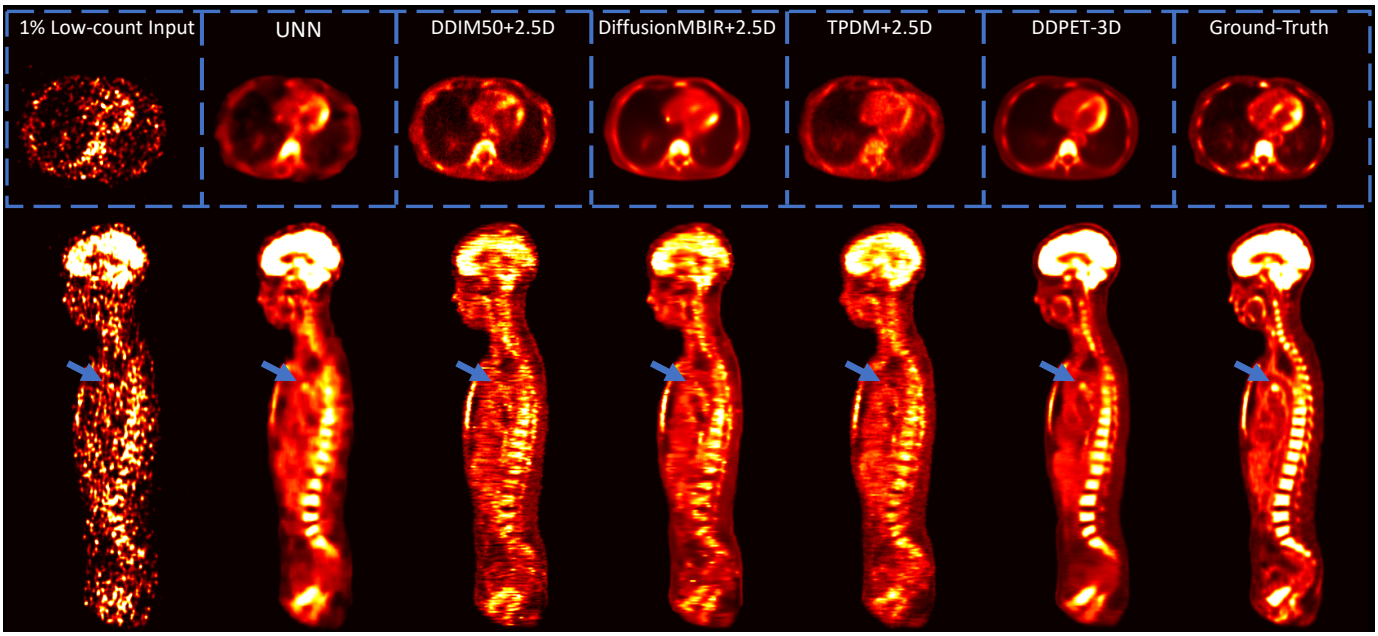

Fig. 11. Images reconstructed using DDPET-3D with different numbers of conditional slices. Note that all the images were normalized by the total injected activities. Blue arrows point to a possible lesion that is better recovered by DDPET-3D with more conditional slices.

[3] C. M. Clark, M. J. Pontecorvo, T. G. Beach, B. J. Bedell, R. E. Coleman, P. M. Doraiswamy, A. S. Fleisher, E. M. Reiman, M. N. Sabbagh, C. H. Sadowsky, *et al.*, "Cerebral pet with florbetapir compared with neuropathology at autopsy for detection of neuritic amyloid-β plaques: a prospective cohort study," *The Lancet Neurology*, vol. 11, no. 8, pp. 669–678, 2012.

[4] E. Robbins, "Radiation risks from imaging studies in children with cancer," *Pediatric blood & cancer*, vol. 51, no. 4, pp. 453–457, 2008.

[5] J. D. Schaefferkoetter, J. Yan, D. W. Townsend, and M. Conti, "Initial assessment of image quality for low-dose pet: evaluation of lesion detectability," *Physics in Medicine & Biology*, vol. 60, no. 14, p. 5543, 2015.

[6] L. Shepp and Y. Vardi, "Maximum Likelihood Reconstruction for Emission Tomography," *IEEE Transactions on Medical Imaging*, vol. 1, pp. 113–122, Oct. 1982.

[7] H. Hudson and R. Larkin, "Accelerated image reconstruction using ordered subsets of projection data," *IEEE Transactions on Medical Imaging*, vol. 13, pp. 601–609, Dec. 1994.

[8] G. Wang, "A Perspective on Deep Imaging," *IEEE Access*, vol. 4, pp. 8914–8924, 2016.

[9] J. Xu, E. Gong, J. Pauly, and G. Zaharchuk, "200x Low-dose PET Reconstruction using Deep Learning," 2017.

[10] L. Zhou, J. D. Schaefferkoetter, I. W. Tham, G. Huang, and J. Yan, "Supervised learning with cyclegan for low-dose fdg pet image denoising,"

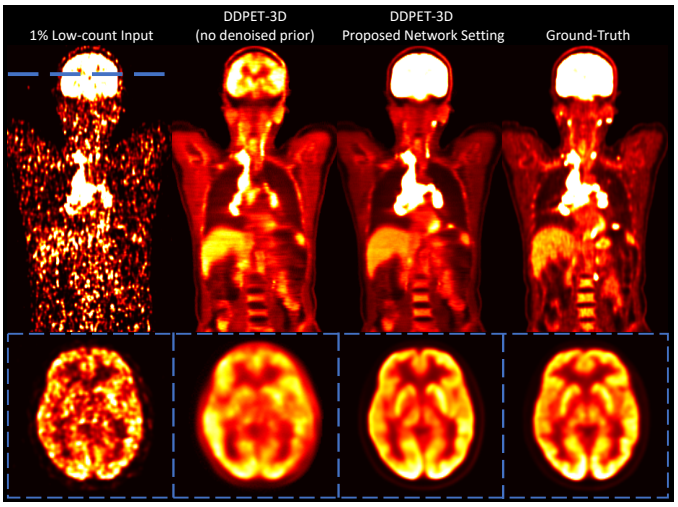

Fig. 12. Visual comparison of images reconstructed using the proposed DDPET-3D methods with and without denoised priors. Note that all the images were normalized by the total injected activities. Note the tracer activity differences in different organs. For the brain transverse slice, due to quantification bias, image intensity in the "DDPET-3D (no prior)" method was manually tuned for visual comparison. Many image details were lost with no prior. Dashed blue line indicates the location of the transverse slice.

Medical image analysis, vol. 65, p. 101770, 2020.

[11] J. Ouyang, K. T. Chen, E. Gong, J. Pauly, and G. Zaharchuk, "Ultra-low-dose PET reconstruction using generative adversarial network with feature matching and task-specific perceptual loss," *Medical Physics*, vol. 46, no. 8, pp. 3555–3564, 2019.

[12] B. Zhou, Y.-J. Tsai, X. Chen, J. S. Duncan, and C. Liu, "Mdpet: A unified motion correction and denoising adversarial network for low-dose gated pet," *IEEE Transactions on Medical Imaging*, 2021.

[13] B. Zhou, Y.-J. Tsai, J. Zhang, X. Guo, H. Xie, X. Chen, T. Miao, Y. Lu, J. S. Duncan, and C. Liu, "Fast-mc-pet: A novel deep learning-aided motion correction and reconstruction framework for accelerated pet," *arXiv preprint arXiv:2302.07135*, 2023.

[14] B. Zhou, T. Miao, N. Mirian, X. Chen, H. Xie, Z. Feng, X. Guo, X. Li, S. K. Zhou, J. S. Duncan, and C. Liu, "Federated transfer learning for low-dose PET denoising: A pilot study with simulated heterogeneous data," *IEEE Transactions on Radiation and Plasma Medical Sciences*, vol. 7, no. 3, pp. 284–295, 2023.

[15] B. Zhou, H. Xie, Q. Liu, X. Chen, X. Guo, Z. Feng, J. Hou, S. K. Zhou, B. Li, A. Rominger, *et al.*, "Fedftn: Personalized federated learning with deep feature transformation network for multi-institutional low-count pet denoising," *Medical image analysis*, vol. 90, p. 102993, 2023.

[16] Q. Liu, H. Liu, N. Mirian, S. Ren, V. Viswanath, J. Karp, S. Surti, and C. Liu, "A personalized deep learning denoising strategy for low-count pet images," *Physics in Medicine & Biology*, vol. 67, no. 14, p. 145014, 2022.

[17] K. Gong, J. Guan, C.-C. Liu, and J. Qi, "PET Image Denoising Using a Deep Neural Network Through Fine Tuning," *IEEE Transactions on Radiation and Plasma Medical Sciences*, vol. 3, pp. 153–161, Mar. 2019.

[18] Y. Onishi, F. Hashimoto, K. Ote, H. Ohba, R. Ota, E. Yoshikawa, and Y. Ouchi, "Anatomical-guided attention enhances unsupervised PET image denoising performance," *Medical Image Analysis*, vol. 74, p. 102226, Dec. 2021.

[19] H. Xie, Q. Liu, B. Zhou, X. Chen, X. Guo, H. Wang, B. Li, A. Rominger, K. Shi, and C. Liu, "Unified Noise-Aware Network for Low-Count PET Denoising With Varying Count Levels," *IEEE Transactions on Radiation and Plasma Medical Sciences*, 2023.

[20] H. Liu, H. Yousefi, N. Mirian, M. Lin, D. Menard, M. Gregory, M. Aboian, A. Boustani, M.-K. Chen, L. Saperstein, D. Pucar, M. Kulon, and C. Liu, "PET Image Denoising Using a Deep-Learning Method for Extremely Obese Patients," *IEEE Transactions on Radiation and Plasma Medical Sciences*, vol. 6, pp. 766–770, Sept. 2022.

[21] Z. Chang, G. A. Koulieris, and H. P. H. Shum, "On the Design Fundamentals of Diffusion Models: A Survey," Oct. 2023.

[22] H. Cao, C. Tan, Z. Gao, Y. Xu, G. Chen, P.-A. Heng, and S. Z. Li, "A Survey on Generative Diffusion Models," *IEEE Transactions on Knowledge and Data Engineering*, pp. 1–20, 2024.

[23] F.-A. Croitoru, V. Hondru, R. T. Ionescu, and M. Shah, "Diffusion models in vision: A survey," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2023.

[24] A. Kazerouni, E. K. Aghdam, M. Heidari, R. Azad, M. Fayyaz, I. Hacihaliloglu, and D. Merhof, "Diffusion models for medical image analysis: A comprehensive survey," *arXiv preprint arXiv:2211.07804*, 2022.

[25] J. Ho, A. Jain, and P. Abbeel, "Denoising diffusion probabilistic models," *arXiv preprint arxiv:2006.11239*, 2020.

[26] K. Gong, K. Johnson, G. El Fakhri, Q. Li, and T. Pan, "PET image denoising based on denoising diffusion probabilistic model," *European Journal of Nuclear Medicine and Molecular Imaging*, 2023.

[27] C. Jiang, Y. Pan, M. Liu, L. Ma, X. Zhang, J. Liu, X. Xiong, and D. Shen, "PET-diffusion: Unsupervised PET enhancement based on the latent diffusion model," in *Medical Image Computing and Computer Assisted Intervention MICCAI 2023*, Lecture Notes in Computer Science, pp. 3–12, Springer Nature Switzerland, 2023.

[28] R. Rombach, A. Blattmann, D. Lorenz, P. Esser, and B. Ommer, "High-resolution image synthesis with latent diffusion models," *arXiv preprint arXiv:2112.10752*, 2022.

[29] Q. Gao, Z. Li, J. Zhang, Y. Zhang, and H. Shan, "CoreDiff: Contextual error-modulated generalized diffusion model for low-dose CT denoising and generalization," *IEEE Transactions on Medical Imaging*, pp. 1–1, 2023.

[30] A. Gngr, S. U. Dar, . ztrk, Y. Korkmaz, H. A. Bedel, G. Elmas, M. Ozbey, and T. ukur, "Adaptive diffusion priors for accelerated MRI reconstruction," *Medical Image Analysis*, vol. 88, p. 102872, 2023.

[31] H. Chung and J. C. Ye, "Score-based diffusion models for accelerated MRI," *Medical Image Analysis*, vol. 80, p. 102479, 2022.

[32] H. Chung, D. Ryu, M. T. McCann, M. L. Klasky, and J. C. Ye, "Solving 3d inverse problems using pre-trained 2d diffusion models," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 22542–22551, June 2023.

[33] S. Lee, H. Chung, M. Park, J. Park, W.-S. Ryu, and J. C. Ye, "Improving 3d imaging with pre-trained perpendicular 2d diffusion models," 2023.

[34] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional Networks for Biomedical Image Segmentation," in *Medical Image Computing and Computer-Assisted Intervention MICCAI 2015*, pp. 234–241, Springer International Publishing, 2015.

[35] R. Boellaard, "Standards for PET image acquisition and quantitative data analysis," *Journal of Nuclear Medicine*, vol. 50, pp. 11S–20S, 2009.

[36] J. Song, C. Meng, and S. Ermon, "Denoising diffusion implicit models," *arXiv preprint arXiv:2010.02502*, 2022.

[37] J. Yang, A. A. S. Soltan, and D. A. Clifton, "Machine learning generalizability across healthcare settings: insights from multi-site COVID-19 screening," *npj Digital Medicine*, vol. 5, pp. 1–8, June 2022.

[38] B. A. Spencer, K. McBride, H. Hunt, T. Jones, S. R. Cherry, and R. D. Badawi, "Practical Considerations for Total-Body PET Acquisition and Imaging," in *Positron Emission Tomography: Methods and Protocols* (T. H. Witney and A. J. Shuhendler, eds.), pp. 371–389, New York, NY: Springer US, 2024.

[39] P. Dhariwal and A. Nichol, "Diffusion models beat gans on image synthesis," *Advances in neural information processing systems*, vol. 34, pp. 8780–8794, 2021.

[40] A. Q. Nichol and P. Dhariwal, "Improved denoising diffusion probabilistic models," in *International Conference on Machine Learning*, pp. 8162–8171, PMLR, 2021.

[41] Y. Song, J. Sohl-Dickstein, D. P. Kingma, A. Kumar, S. Ermon, and B. Poole, "Score-based generative modeling through stochastic differential equations," *arXiv preprint arXiv:2011.13456*, 2021.

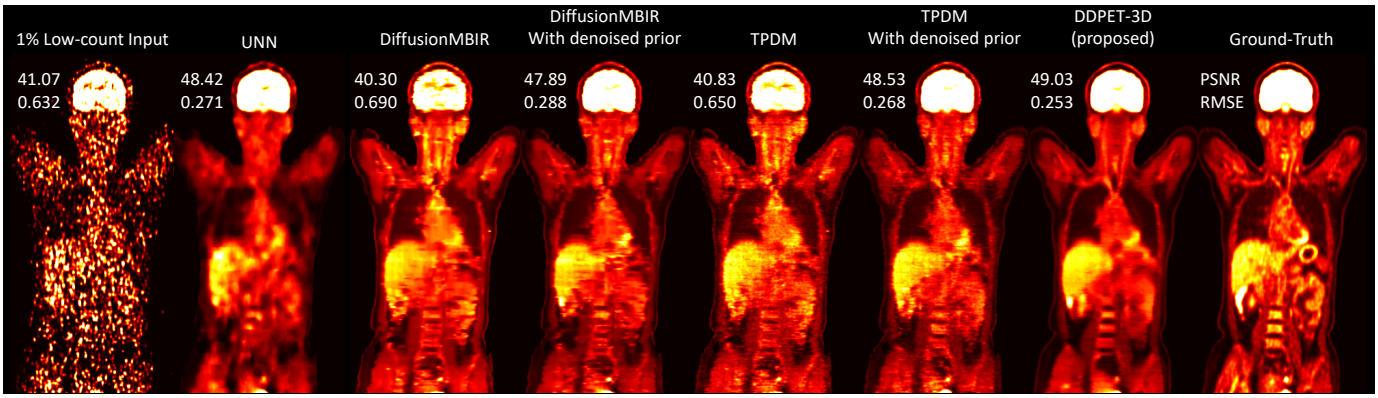

Fig. 13. Results produced by competing methods with denoised prior. The proposed DDPET-3D still outperformed other competing methods. Note that the denoised prior is one of the main contributions proposed in this work.

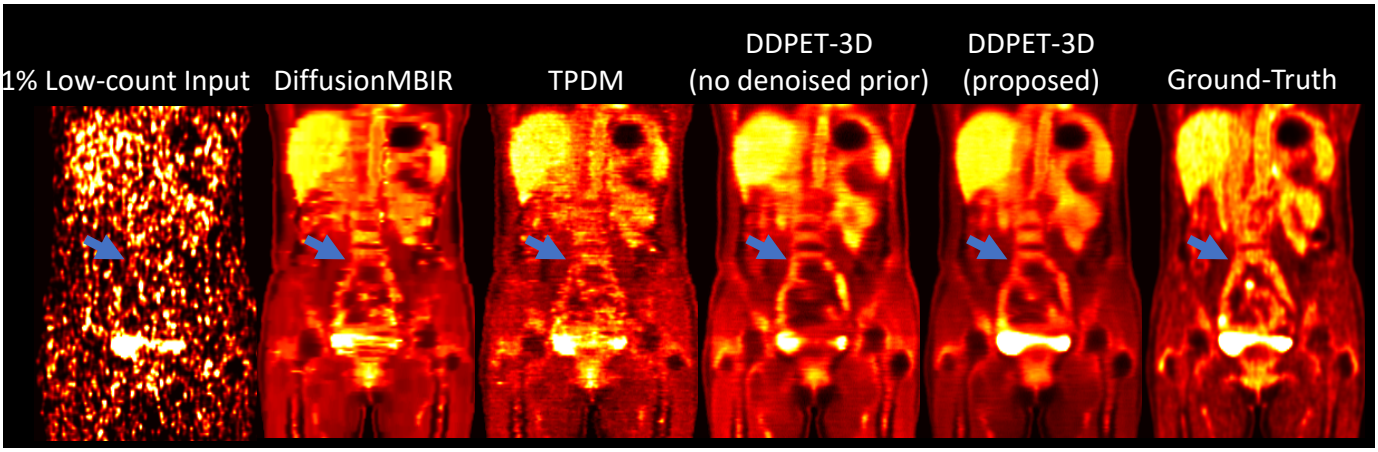

Fig. 14. Without the denoised prior, proposed DDPET-3D still produced more consistent reconstructions compared with other methods. Blue arrows point to regions with noticeable inconsistencies between slices in competing methods.

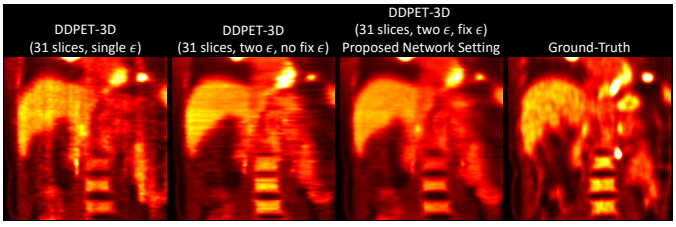

Fig. 15. Images reconstructed with different settings of noise variables. Note that all the images were normalized by the total injected activities. Note the undesired artifacts along the z-direction by using only one noise variable ϵ. Also note the inconsistencies of difference slices without fixing ϵ. Despite similar quantitative values, proposed network settings produced images with more consistent 3D reconstructions and less artifacts.