European Journal of Nuclear Medicine and Molecular Imaging
https://doi.org/10.1007/s00259-021-05197-3

ORIGINAL ARTICLE

Artificial intelligence enables whole-body positron emission tomography scans with minimal radiation exposure

Yan-Ran (Joyce) Wang[1] · Lucia Baratto[1] · K. Elizabeth Hawk[1] · Ashok J. Theruvath[1] · Allison Pribnow[2] · Avnesh S. Thakor[1] · Sergios Gatidis[3] · Rong Lu[4] · Santosh E. Gummidipundi[4] · Jordi Garcia-Diaz[1] · Daniel Rubin[1,2] · Heike E. Daldrup-Link[1,2] ·

Received: 26 October 2020 / Accepted: 10 January 2021

Abstract

Purpose To generate diagnostic ^{18}F-FDG PET images of pediatric cancer patients from ultra-low-dose ^{18}F-FDG PET input images, using a novel artificial intelligence (AI) algorithm.

Methods We used whole-body ^{18}F-FDG-PET/MRI scans of 33 children and young adults with lymphoma (3–30 years) to develop a convolutional neural network (CNN), which combines inputs from simulated 6.25% ultra-low-dose ^{18}F-FDG PET scans and simultaneously acquired MRI scans to produce a standard-dose ^{18}F-FDG PET scan. The image quality of ultra-low-dose PET scans, AI-augmented PET scans, and clinical standard PET scans was evaluated by traditional metrics in computer vision and by expert radiologists and nuclear medicine physicians, using Wilcoxon signed-rank tests and weighted kappa statistics.

Results The peak signal-to-noise ratio and structural similarity index were significantly higher, and the normalized root-mean-square error was significantly lower on the AI-reconstructed PET images compared to simulated 6.25% dose images ($p < 0.001$). Compared to the ground-truth standard-dose PET, SUV_{max} values of tumors and reference tissues were significantly higher on the simulated 6.25% ultra-low-dose PET scans as a result of image noise. After the CNN augmentation, the SUV_{max} values were recovered to values similar to the standard-dose PET. Quantitative measures of the readers' diagnostic confidence demonstrated significantly higher agreement between standard clinical scans and AI-reconstructed PET scans (kappa = 0.942) than 6.25% dose scans (kappa = 0.650).

Conclusions Our CNN model could generate simulated clinical standard ^{18}F-FDG PET images from ultra-low-dose inputs, while maintaining clinically relevant information in terms of diagnostic accuracy and quantitative SUV measurements.

Keywords Pediatric cancer imaging · PET/MRI · Whole-body PET reconstruction · PET denoising · Deep learning

Yan-Ran (Joyce) Wang and Lucia Baratto are co-first authors

This article is part of the Topical Collection on Advanced Image Analyses (Radiomics and Artificial Intelligence)

Key points QUESTIONCan artificial intelligence augment whole-body PET scans with minimal radiation exposure to the quality of standard-dose PET?

✉ Daniel Rubin
drubin@stanford.edu

✉ Heike E. Daldrup-Link
heiked@stanford.edu

[1] Department of Radiology, Molecular Imaging Program at Stanford, Stanford University, 725 Welch Road, CA 94304 Stanford, USA

[2] Department of Pediatrics, Pediatric Oncology, Lucile Packard Children's Hospital, Stanford University, Stanford, CA 94304, USA

[3] Department of Diagnostic and Interventional Radiology, University Hospital Tuebingen, Tuebingen, Germany

[4] Quantitative Sciences Unit, School of Medicine, Stanford University, Stanford, CA 94304, USA

Published online: 01 February 2021

Introduction

In many patients with cancer, the metabolic information from ^{18}F-FDG PET/CT scans is required to provide accurate tumor diagnoses and to monitor response to treatment [1–3]. However, diagnostic ^{18}F-FDG PET/CT scans involve considerable radiation exposure [4, 5]. Several groups independently reported that the radiation exposure from diagnostic CT scans is associated with an increased risk of developing secondary cancers later in life [6–8]. This is particularly concerning for children, as they are more sensitive to radiation effects than adults [9]. For example, a patient with lymphoma who undergoes five PET/CT scans (including an 18F-FDG PET and diagnostic CT scan as often required for pediatric treatment protocols) will be exposed to 10–15 mSv of ionizing radiation per scan, which is 50–75 mSv in total [10]. Direct evidence from human population studies showed that doses of 50–100 mSv (protracted exposure) or 10–50 mSv (acute exposure) increase the risk of developing secondary cancers later in life [9]. While advances in cancer therapy have increased the number of pediatric cancer survivors, these patients now live long enough to encounter secondary cancers [5, 9, 11, 12]. Therefore, the *Image Gently* campaign advocates for practitioners to provide the least possible radiation exposure when examining pediatric patients [13]. Integrated ^{18}F-FDG PET/MRI saves radiation by replacing CT with radiation-free MRI scans [14, 15]. This addresses the radiation exposure from CT scans. A clinical standard ^{18}F-FDG PET/CT scan is associated with about 6–7 mSv of radiation exposure for a clinical CT scan and 6–7 mSv for the ^{18}F-FDG PET [12, 14]. While many studies have focused on replacing CT with MRI for anatomical co-registration of ^{18}F-FDG data, the reduction of the injected radiotracer dose has received less attention thus far. A major bottleneck to reducing radiotracer doses for ^{18}F-FDG PET scans is increased image noise [16] and resultant decreased diagnostic accuracy of ultra-low-dose scans [17]. We hypothesized that this problem could be solved by training a deep convolutional neural network (CNN) to integrate information from ultra-low-dose PET images with anatomical information from simultaneously acquired MRI images to generate simulated standard-dose PET images.

The main innovation of our work is fourfold: (1) We explored ultra-low-dose ^{18}F-FDG PET imaging using CNN-based image data augmentation. While previous studies investigated CNN for augmentation of ^{18}F-FDG PET scans for adult patients [18–21], we focused on pediatric patients and young adults, for whom radiation safety is particularly important. (2) We utilized information from simultaneously acquired MRI scans to reconstruct ultra-low-dose whole-body ^{18}F-FDG PET scans. This approach has been previously applied for image reconstruction of brain images [19, 20, 22], but not whole-body images. The inclusion of MRI data into the CNN improves the depiction of anatomical detail that could be missed if only the low-dose ^{18}F-FDG PET was used as input [22]. (3) We incorporated an attention-weighted loss function to enhance sensitivity of our model to reconstruct the significant regions where lesions occur. Pioneering work in this area has been done by Ouyang et al. [23]. The approach contributes more to the loss function during training in regions where lesions occur, such as perivascular areas where lymph nodes are common. Making the model pay more attention to the anatomical regions with high frequency of pathology could protect the loss and computed gradients from overwhelming by relatively irrelevant pixels in whole-body scans. In this manner, less training data are required, which is critical in the domain of pediatric cancer imaging where imaging data are relatively sparse. (4) We conducted a task-specific region-based clinical evaluation. The reconstructed PET images were not only evaluated by traditional metrics in computer vision but also assessed by expert radiologists and nuclear medicine physicians in terms of the overall image quality, diagnostic accuracy, and diagnostic confidence. To date, no comprehensive region-based clinical evaluation was conducted in such whole-body PET image enhancement studies. Thus, the purpose of our study was to generate diagnostic ^{18}F-FDG PET images of pediatric cancer patients from ultra-low-dose ^{18}F-FDG PET input images, using a novel CNN algorithm.

Materials and methods

Patients and image acquisition

This Health Insurance Portability and Accountability Act–compliant clinical study was approved by our respective institutional review boards and was performed as a secondary analysis of prospectively acquired data. Written informed consent was obtained from all adult patients and all parents of pediatric patients. In addition, children were asked to give their assent. Between July 2015 and June 2019, we enrolled 33 children and young adults (14 female, 9 male) with lymphoma at two centers (University of Tübingen, Germany, and Stanford University, CA, USA) who underwent integrated ^{18}F-FDG PET/MRI scans for tumor staging. Twenty-three patients enrolled at Stanford had a mean age of 17 ± 7 years (range: 6–30 years), and 11 patients enrolled at Tübingen had a mean age of 14 ± 5 years (range: 3–18 years). The patients at Stanford underwent a whole-body integrated ^{18}F-FDG PET/MRI scan on a 3 T Signa PET/MRI scanner (GE Healthcare, Milwaukee, WI, USA) at 1 h after intravenous injection of ^{18}F-FDG at a dose of 3 MBq/kg, using a 32-channel torso phased array coil and an eight-channel, receive-only head coil. PET data were acquired simultaneously with contrast-enhanced T1-weighted gradient echo scans, using a 25-cm transaxial FOV and 3:30-min acquisitions per PET bed. Tübingen patients underwent a whole-body integrated ^{18}F-FDG PET/MRI scan on a 3 T Signa PET/MRI scanner

(Siemens Healthineers, Erlangen, Germany), using a 16-channel torso phased array coil and a 16-channel head coil. PET data were acquired simultaneously with contrast-enhanced T1-weighted gradient echo MRI scans, using a 25-cm transaxial FOV and 4-min acquisitions per PET bed. Radiotracer input data were used to generate 100% dose ^{18}F-FDG PET images. Moreover, 6.25% (0.18 mBq/kg) low-dose ^{18}F-FDG PET images were simulated by unlisting the PET list-mode data and reconstructing them based on the percentage of used counts [24].

CNN architecture

We trained and cross-validated a CNN reconstruction model to augment whole-body ^{18}F-FDG PET/MRI scans of 23 subjects with lymphoma. The inputs for the model are axial simulated ultra-low-dose ^{18}F-FDG PET images and simultaneously acquired axial contrast-enhanced T1-weighted MRI image. The outputs are AI-reconstructed ^{18}F-FDG PET images, which should resemble a standard-dose ^{18}F-FDG PET scan (Fig. 1).

We designed the reconstruction network based on an enhanced deep super-resolution network (EDSR) [25] – the state-of-the-art image reconstruction network. However, our model is significantly different from EDSR and particularly tailored for the whole-body ^{18}F-FDG PET reconstruction. Our network is different in four key ways:

(1) We utilized information from simultaneously acquired MRI scans. We hypothesized that using only a single simulated ultra-low-dose ^{18}F-FDG PET image as input may not provide enough information to reconstruct detailed anatomical structures and may generate hallucination image artifacts. The information from simultaneously acquired MRI scans can be used to provide anatomical information.
(2) We applied middle fusion to integrate MRI and ^{18}F-FDG PET images. Rather than concatenating the MRI and ^{18}F-FDG PET at the input level, we combine them in the feature space. We assume that early fusion might lose information as the characteristics of ^{18}F-FDG PET and MRI modalities are quite different. The benefit of mid-fusion as opposed to early fusion was observed in the initial experiment. The two modalities are integrated after the fourth main residual block. Note that the PET branch is built upon a residual block, while the MRI branch is built upon pure convolutional layers.
(3) A skip connection between the ultra-low-dose ^{18}F-FDG PET input and the final prediction layer is added to alleviate the burden of carrying identity information in the reconstruction network.
(4) The reconstruction network is a slice-wise model, which considers multi-slice inputs. The input consists of 5-slice LAVA MRI and 5-slice ultra-low-dose ^{18}F-FDG PET images, and the output is a synthetic standard-dose ^{18}F-FDG PET slice. Such input scheme provides the network with 2.5D information, reduces image noise, and ensures vertical spatial consistency. The proposed network consists of 44 convolution layers in total.

Preprocessing steps: For each baseline scan, we used ITK-SNAP to obtain the foreground body areas with the MRI image as the reference. The foreground body mask was used to zero out background signals of MRI, standard-dose PET, and simulated ultra-low-dose PET. Then, zero-mean (the mean value of the nonzero region) and unit-variance normalization was applied within the foreground body area before feeding them to the CNN model. The input consists of 5-slice LAVA MRI and 5-slice ultra-low-dose PET image. The size of each slice is 512×512 (fixed).

Model training: The CNN reconstruction model was optimized using Adam with parameters: learning rate, 1e-6; β1, 0.9; β2, 0.999; and batch size, 1. Each trainable node in the CNN was regularized with L1 loss and weight decay of 0.0001. The weighing parameter to balance content loss and regularization loss is 0.001. In the initial experiment, early stopping (when the validation accuracy does not improve, the training will be stopped) was used to set the number of epochs for training. In our experiment, it is ten epochs and we applied it in the leave-one-out cross-validation experiments. The proposed network consists of 44 convolution layers in total. The filter size is 3×3 and the number of filters is 64 for each convolutional layer besides the final layer.

Attention-weighted loss

Weighted loss function was initially proposed to tackle the common issue of imbalanced data in background/foreground classification. By weighing underrepresented categories, a weighted loss function compensates the bias of training loss for the minority categories. In our work, we designed the attention-weighted loss by augmenting the loss function with a weight value corresponding to the significant regions of whole-body scans, specifically visceral organs and lymph node regions that are common areas where tumors occur. While all ^{18}F-FDG PET data are augmented by our algorithm, the CNN pays particular attention to these areas. This encoded prior knowledge in spatial anatomy could enable the network to converge quickly, simplify the training, and improve quality of enhancing image quality of images of cancer lesions. Figure 1b shows the calculation of the attention-weighted loss function in the training phase.

The generation of attention mask

For the training dataset, we used ITK-SNAP (*32*) to obtain the attention masks that highlight the high-clinical-value regions.

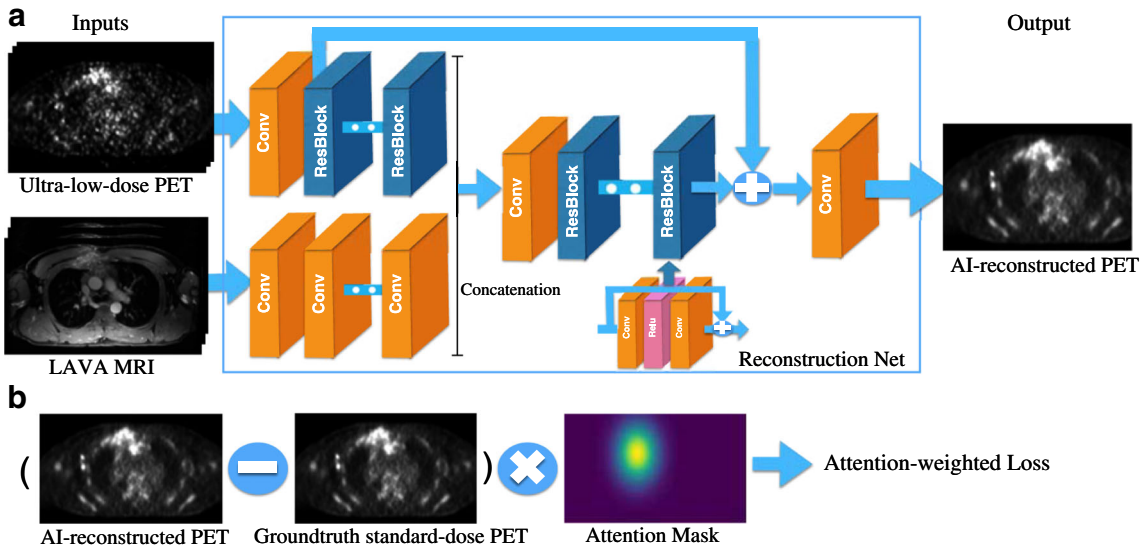

Fig. 1 The pipeline of the CNN reconstruction net. **a** The evaluation framework of the PET reconstruction CNN. It inputs the simulated ultra-low-dose 18F-FDG PET and contrast-enhanced T1-weighted MRI images and outputs the synthetized standard-dose 18F-FDG PET images. The PET and MRI images are integrated at the mid-level after feature extraction; a skip connection is added to the CNN in order to connect the ultra-low-dose 18F-FDG PET images with the final reconstruction. **b** The calculation process of the attention-weighted loss. The attention mask highlights the tumor area

Potential tumor areas along the main vessels, mediastinum, liver, and spleen were segmented as attention areas with the MRI image as the reference. Then, the rigid attention mask was transformed to a soft attention mask via Gaussian distribution. The final attention mask to the loss function is a Gaussian heatmap produced by four variables: the center and standard deviations of the Gaussian distribution of the rough target region mask of the scan, as shown in Supplementary Figure 7.

Training details

We trained our model using Stanford baseline ^{18}F-FDG-PET/MR scans of 23 children and young adults with malignant lymphoma. The large data requirement for training CNN is a limitation for pediatric applications, as there are not many of these studies. Proof-of-concept studies of CNN for pediatric oncology applications are hampered by sparse data [26]. To overcome the challenge, we adopted leave-one-out cross-validation. The dataset was divided into 23-folds. During training, 22 of the folds were used as training set, whereas the remaining one-fold was used for testing. We iterated 23 times to go through all combinations and produce the final AI-reconstructed PET images.

In the initial experiment (18 subjects for training and 5 subjects for validation), early stopping (when the validation accuracy does not improve, the training will be stopped) was used to set the number of epochs for training. In our experiment, it is ten epochs and we applied the same number of epochs (10 epochs) in the leave-one-out cross-validation experiments. For leave-one-out cross-validation, we iterated 23-folds, and for each fold, only one final AI-reconstructed PET image was produced. In this manner, we avoid information leakage during the generation of the final AI-reconstructed PET images.

For the comparison study that should show the superiority of adding the MRI input information and attention-weighted loss, we trained the comparison models on 14 subjects, validated on 3 subjects, and then tested on the remaining 6 subjects (Supplementary Data).

Computational assessment

For the evaluation part, we used three computational metrics in computer vision to evaluate the performance of our networks, including peak signal-to-noise ratio (PSNR), the structural similarity index (SSIM), and the normalized root-mean-square error (NRMSE) (Supplementary Data).

Clinical assessment

To understand the impact of our CNN on tumor detection, three clinical imaging experts (one radiologist, one nuclear medicine physician, and one dual-trained radiologist/nuclear medicine physician) determined the presence or absence of tumor lesions in 20 anatomical regions (Supplementary Table 1) per patient on the ultra-low-dose ^{18}F-FDG PET scan, the AI-augmented ultra-low-dose ^{18}F-FDG PET scan, and the 100% standard-dose ^{18}F-FDG PET scan (in total 20 regions × 23 patients = 460 anatomical regions analyzed by each reviewer). The reviewers were blinded to clinical data and the type of the exam and analyzed the three different scan types in a random order and with an

interval of at least 2 weeks to minimize recall bias. The clinical experts rated the visibility of lesions in these regions according to a Likert scale (1 – tumor definitely not present, 2 – tumor probably not present, 3 – undecided, 4 – tumor probably present, 5 – tumor definitely present). The combination of all clinical imaging tests and biopsies on all available imaging studies was used to generate a standard of reference for these evaluations. The lesion diagnostic metrics on the ultra-low-dose ^{18}F-FDG PET scan, the AI-augmented ultra-low-dose ^{18}F-FDG PET scan, and the 100% standard-dose ^{18}F-FDG PET scan were compared with the standard of reference using *confusionMatrix()* function from R package *caret.*

Measurements of SUV values from tumors in the PET images are important for tumor detection and for quantitative monitoring of tumor therapy response. Reducing the ^{18}F-FDG radiotracer dose can lead to increased image noise and affect these measurements. To evaluate whether flawed SUV measurements could be recovered by our CNN, one nuclear medicine physician measured the SUV_{max} and SUV_{mean} (standardized uptake value) of the lesion with the highest SUV in the 20 refined regions per patient as well as the SUV_{mean}, SUV_{std}, and SUV_{max} of the liver (3-cm ROI) and mediastinal blood pool (2-cm ROI) across the simulated ultra-low-dose ^{18}F-FDG PET scan, the AI-augmented ultra-low-dose ^{18}F-FDG PET scan, and the 100% standard-dose ^{18}F-FDG PET scan. The measurement was obtained using MIM 6.5 (MIM Software, Inc., Cleveland, OH, USA) as SUV_{max} = (tissue tracer activity (mCi/g))/((injected dose (mCi)*patient body weight (kg))). All metrics were compared across the ultra-low-dose ^{18}F-FDG PET scan, the AI-augmented ultra-low-dose ^{18}F-FDG PET scan, and the 100% standard-dose ^{18}F-FDG PET images using a Wilcoxon signed-rank test.

Statistical analysis

To evaluate the ability of the different scan types to provide clinically relevant information, we compared imaging experts' scan assessments with the ground truth of whether lesions were present in each scan/region or not. The ground truth was defined as the combination of all imaging scans and biopsies obtained to generate a baseline diagnosis. This included biopsy results of the primary tumor, ultrasound, x-rays, CT, MR, ^{18}F-FDG PET, and bone scans. By defining anatomical areas, which are positive (contain tumor) or which are negative (do not contain tumor) through the ground truth, we calculated the positive predictive value (PPV), negative predictive value (NPV), and balanced accuracy (average of sensitivity and specificity). We reported the balanced accuracy because the unbalanced accuracy statistic is very sensitive to the proportion of true positive (contain tumor) and true negative cases (do not contain tumor) in any scan sample with limited sample size. The balanced accuracy estimates are often more stable and more comparable between different studies. All PPV, NPV, and balanced accuracy estimates were calculated using the *confusionMatrix()* function from R package caret. To evaluate the degree of agreement in Likert scale assessments between the 100% standard-dose PET scan and the other 2 scan types, we calculated weighted kappa statistics (using both linear weights and quadratic weights). All kappa estimates were generated using the kappa2() function from R package *irr*. The Wilcoxon signed-rank test was used to compare all image quality metrics and SUV values that are paired between scan types.

Results

CNN reduces image noise of simulated ultra-low-dose PET scans

Figure 2 and Table 1 show the qualitative and quantitative results of the reconstruction model. A major problem of reducing radiotracer doses for ^{18}F-FDG PET scans was an increased image noise [16]. The CNN-augmented ultra-low-dose ^{18}F-FDG PET images demonstrated significantly less noise and better image quality compared to simulated 6.25% dose ^{18}F-FDG PET images, measured by PSNR, SSIM, and NRMSE. PSNR and SSIM were significantly higher, and NRMSE was significantly lower on the CNN-augmented ultra-low-dose ^{18}F-FDG PET as opposed to simulated 6.25% dose ^{18}F-FDG PET images (all pair-wise t-tests $p < 0.001$; Table 1). In addition, the standard deviation of the mean standardized uptake value (SUV) measurements of the liver and mediastinal blood pool was significantly lower on the CNN-augmented ultra-low-dose ^{18}F-FDG PET images compared to simulated 6.25% dose ^{18}F-FDG PET images (Table 3 and Supplementary Figure 4). These qualitative and quantitative results show that the proposed CNN model reduces image noise of the ultra-low-dose ^{18}F-FDG PET scans and reaches an overall image quality on CNN-augmented ^{18}F-FDG PET images, which are similar to 100% dose ^{18}F-FDG PET scans.

Combining MRI and PET improves reconstruction quality

Supplementary Figure 1 shows the qualitative and quantitative performance of our CNN with and without additional MRI inputs. The CNN that predicts the standard-dose ^{18}F-FDG PET with the least error is the CNN trained on both contrast-enhanced MRI and simulated ultra-low-dose ^{18}F-FDG PET images, which demonstrates the benefit of including the MRI modality. The simultaneous ^{18}F-FDG PET and MRI acquisition mode facilitates the integration of input data from the two modalities. The MRI images provided complementary anatomical information for depicting detailed high-resolution

Table 1 Comparison of image quality metrics for simulated 6.25% ultra-low-dose ^{18}F-FDG PET and AI-reconstructed ^{18}F-FDG PET scans

	6.25% ultra-low-dose PET (N =23)	AI-reconstructed PET (N =23)	P value*
PSNR			
Mean (SD)	51.6 (8.50)	55.6 (7.62)	<0.001
Median (Q1, Q3)	53.6 (43.9, 59.3)	58.1 (48.8, 62.3)	
SSIM			
Mean (SD)	0.925 (0.0449)	0.967 (0.0175)	<0.001
Median (Q1, Q3)	0.929 (0.916, 0.957)	0.972 (0.961, 0.978)	
NRMSE			
Mean (SD)	0.257 (0.102)	0.158 (0.0453)	<0.001
Median (Q1, Q3)	0.228 (0.182, 0.305)	0.156 (0.127, 0.191)	

*The AI-reconstructed ^{18}F-FDG PET scan demonstrates improved image quality and significantly less noise for all three metrics compared to the ultra-low dose ^{18}F-FDG PET scan: higher peak signal-to-noise ratio (PSNR), higher structural similarity index (SSIM) and lower normalized root-mean-square error (NRMSE), n = 23 scans per group, Wilcoxon signed-rank tests. P-values were calculated using Wilcoxon signed-rank test

features that could be missed if only the simulated ultra-low-dose ^{18}F-FDG PET images were used as input.

Attention-weighted loss boosts performance

Supplementary Figure 2 shows the performance comparison of two models – one with the attention-weighted loss function and one with a traditional mean-square-error (MSE) loss function. The attention-weighted loss function demonstrated a tendency to boost the performance of the CNN.

AI-reconstructed PET scans enable accurate tumor detection

Figure 3, Supplementary Figure 3, and Table 2 show the qualitative and quantitative results in terms of image diagnostic quality. Since pediatric tumors are highly metabolically active, there were only small differences in the clinical experts' ability to detect tumor lesions on the different imaging scans: The clinical experts detected 225 of 249 tumors on the 100% dose ^{18}F-FDG PET scan, 221 tumors on the simulated ultra-low-dose ^{18}F-FDG PET scan, and 223 tumors on the AI-augmented ultra-low-dose

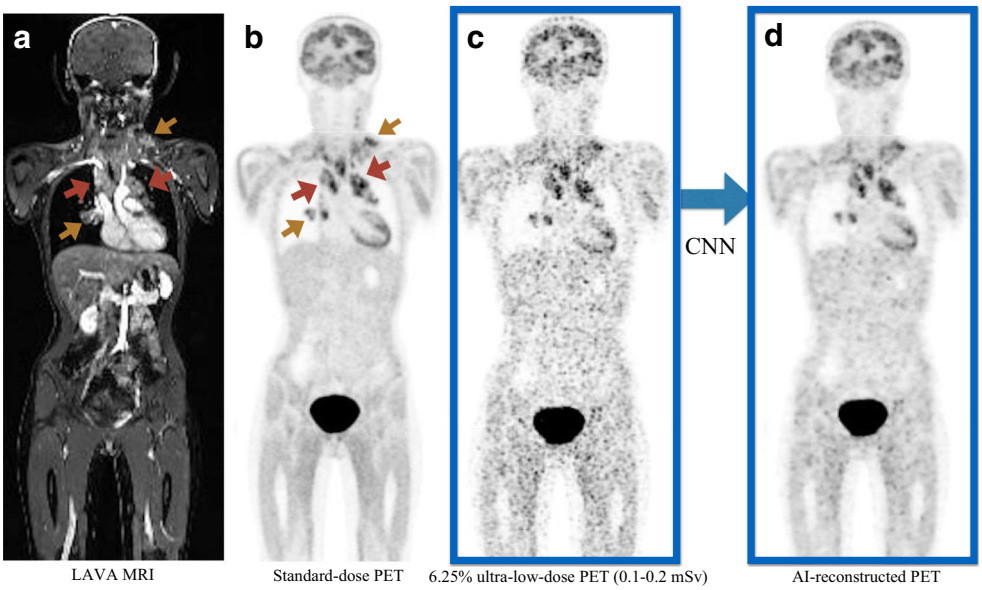

Fig. 2 Representative 18F-FDG PET/MRI scan of a 16-year-old female patient with Hodgkin lymphoma (HL). **a** Coronal contrast-enhanced T1-weighted LAVA (liver acquisition and volume acquisition) MRI, **b** coronal view of a standard-dose 18F-FDG dose PET scan (3 mBq/kg), **c** simulated ultra-low-dose PET scan at 6.25% 18F-FDG dose, and **d** the AI-reconstructed ultra-low-dose 18F-FDG PET image, reconstructed based on the 6.25% ultra-low dose PET and MRI scans as combined inputs. The red arrows point to the hypermetabolic tumors in the mediastinum. Additional hypermetabolic tumors are noted at the right hilum and left lower neck (yellow arrows). All lesions can be detected on all scans, but tumor-to-background contrast and confidence for lesion detection is improved on the AI-reconstructed 18F-FDG PET

^{18}F-FDG PET scans. Most lesions were noted on all scans. Sensitivities, specificities, and diagnostic accuracies were not significantly different for the three imaging modalities (Supplementary Table 2).

However, the confidence of the readers regarding the presence or absence of tumor lesions in specific anatomical regions, as measured by a Likert scale, was significantly different between the different scans. The readers demonstrated a significantly lower confidence in the detection of lesions on 6.25% dose scans as opposed to on the AI-augmented ultra-low-dose ^{18}F-FDG PET scans and the 100% dose ^{18}F-FDG PET scans. The 100% dose ^{18}F-FDG PET scans demonstrated significantly higher agreement with the AI-augmented ultra-low-dose ^{18}F-FDG PET scans (kappa = 0.942) than the 6.25% dose ^{18}F-FDG PET scans (kappa = 0.650). This also applied when considering tumors in different anatomical regions: The 100% dose ^{18}F-FDG PET scans demonstrated significantly higher agreement with the AI-augmented ultra-low-dose ^{18}F-FDG PET scans for tumor detection in the lymph nodes (kappa = 0.955), visceral organs (kappa = 0.910), and bone marrow (kappa = 0.828) compared to corresponding results of the simulated ultra-low-dose ^{18}F-FDG PET scan for tumor detection in the lymph nodes (kappa = 0.702), visceral organs (kappa = 0.573), and bone marrow (kappa = 0.278).

AI-reconstructed PET provides accurate quantitative tumor SUV measurements

Compared to the 100% dose ^{18}F-FDG PET scans, SUV_{max} values of tumors, liver, and mediastinal blood pool were significantly higher on the simulated 6.25% ultra-low-dose PET scans as a result of added image noise. On AI-augmented ultra-low-dose ^{18}F-FDG PET scans, the SUV_{max} values were recovered to values that were similar to the standard-dose PET (Table 3 and Supplementary Figure 4).

Tumor SUV values are often compared to SUV_{mean} values of liver and mediastinal blood pool as an internal standard of reference. Compared to 100% dose ^{18}F-FDG PET scans, liver SUV_{mean} and mediastinal blood pool SUV_{mean} values were higher on 6.25% dose ^{18}F-FDG PET scans (p=0.028 and 0.036, respectively), but were not significantly different on the AI-augmented ultra-low-dose ^{18}F-FDG PET scans (p=0.523 and 0.316, respectively).

The CNN model generalizes in independent data

Next, we evaluated the model's generalization by examining how the model performs when it augments reduced-dose ^{18}F-FDG PET images from PET studies at another institution (Tübingen). We applied our CNN to PET scans of 11 additional subjects. The qualitative and quantitative reconstruction results are shown in Supplementary Tables 3–5 and Supplementary Figure 5. The image quality was significantly improved on AI-augmented ultra-low-dose ^{18}F-FDG PET scans as opposed to the original simulated reduced-dose PET images by 3.7 dB in PSNR, 2.8% in SSIM, and 12.6% in NRMSE, which demonstrates good model generalization across data from different institutions. This is particularly noteworthy as the two institutions used different scanners from different vendors.

When comparing quantitative SUV measurements, we found significantly higher tumor SUVmax in the reduced-dose PET group, compared to 100% dose group (p = 0.008). The CNN effectively corrected this discrepancy: Tumor SUVmax measures on AI-augmented ultra-low-dose ^{18}F-FDG PET scans were not significantly different compared to the 100% dose scans (p = 0.518). The liver SUVmean was not significantly different between reduced-dose and 100% dose PET scans (p = 0.412) or AI-augmented and 100% dose PET scans (p = 0.270). However, the standard deviation of the liver SUV (SUVstd), a quantitative measure of image noise, was significantly higher on reduced-dose PET scans compared to 100% dose PET scans(p < 0.001), while there was no significant difference of liver SUVstd between AI-augmented and 100% dose PET scans (p = 0.818).

Similar to results of the first cohort, the readers showed significantly lower confidence in the detection of lesions on 6.25% dose scans as opposed to AI-augmented ultra-low-dose ^{18}F-FDG PET scans and the 100% dose ^{18}F-FDG PET scans. The 100% dose ^{18}F-FDG PET scans demonstrated significantly higher agreement with the AI-augmented ultra-low-

Table 2 Agreement between expert reviewer tumor diagnoses on 100% standard-dose ^{18}F-FDG PET and simulated 6.25% ultra-low-dose ^{18}F-FDG PET and AI-reconstructed ^{18}F-FDG PET scan

Procedure	Weighted kappa (Linear)	Weighted kappa (Quadratic)
Lymph nodes		
6.25% ultra-low-dose	0.702	0.859
AI-reconstructed PET	**0.955**	0.984
Extralymphatic		
6.25% ultra-low-dose	0.573	0.765
AI-reconstructed PET	**0.910**	0.965
Bone marrow		
6.25% ultra-low-dose	0.278	0.444
AI-reconstructed PET	**0.828**	0.916
Whole body		
6.25% ultra-low-dose	0.650	0.820
AI-reconstructed PET	**0.942**	0.977

Three expert reviewers determined the presence or absence of tumor lesions in 20 anatomical regions per patient according to a Likert scale (1 – tumor definitely not present, 2 – tumor probably not present, 3 – undecided, 4 – tumor probably present, 5 – tumor definitely present). The agreement between 100% standard-dose PET images and 6.25% ultra-low-dose ^{18}F-FDG PET and AI-reconstructed ^{18}F-FDG PET scans was calculated with weighted kappa estimates

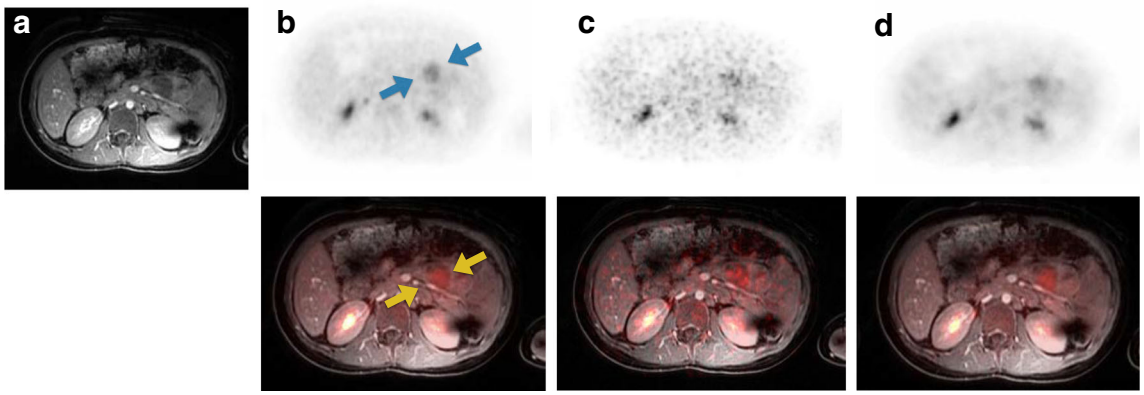

Fig. 3 Illustration of image quality improvement in terms of tumor delineation on the AI-reconstructed 18F-FDG PET scan. Representative 18F-FDG PET/MRI scan of a 10-year-old male patient with diffuse large B-cell lymphoma (DLBCL). **a** Axial T1-weighted contrast-enhanced LAVA MRI; **b** axial standard-dose 18F-FDG PET scan (upper panel), fused with T1-weighted MRI (lower panel); **c** axial simulated 6.25% ultra-low-dose 18F-FDG PET scan, fused with T1-weighted MRI; and **d** axial AI-reconstructed 18F-FDG PET image, fused with T1-weighted MRI. The arrows point to a tumor in the pancreas. The tumor can be well depicted in the original 100% dose 18F-FDG PET scan and the AI-reconstructed 18F-FDG PET, but is nearly invisible on the 6.25% dose 18F-FDG PET scan

Table 3 Standardized uptake values (SUV), as measured on 100% standard-dose ^{18}F-FDG PET, simulated 6.25% ultra-low-dose ^{18}F-FDG PET, and AI-reconstructed ^{18}F-FDG PET scans.

	100% standard-dose PET	6.25% ultra-low-dose PET	AI-reconstructed PET
Liver			
SUV_max			
Mean (SD)	2.65 (1.12)	5.87 (2.44)	2.92 (1.06)
Median (Q1, Q3)	2.47 (2.19, 2.69)	5.49 (4.79, 6.23)	2.71 (2.46, 2.97)
SUV_mean			
Mean (SD)	1.89 (0.817)	1.98 (0.866)	1.90 (0.852)
Median [Q1, Q3]	1.80 (1.54, 1.94)	1.78 (1.56, 1.96)	1.73 (1.48, 1.91)
SUV_std			
Mean (SD)	0.216 (0.0826)	0.774 (0.277)	0.270 (0.0894)
Median [Q1, Q3]	0.200 (0.170, 0.240)	0.710 (0.670, 0.845)	0.250 (0.225, 0.300)
Mediastinal Blood Pool			
SUV_max			
Mean (SD)	2.44 (1.13)	5.11 (2.55)	2.60(1.21)
Median [Q1, Q3]	2.22 (1.68, 2.68)	4.22 (3.29, 6.40)	2.44 (1.77, 2.97)
SUV_mean			
Mean (SD)	1.53 (0.863)	1.61 (0.809)	1.59 (0.794)
Median [Q1, Q3]	1.25 (1.02, 1.67)	1.38 (1.12, 1.82)	1.47 (1.08, 1.70)
SUV_std			
Mean (SD)	0.237 (0.0951)	0.690 (0.316)	0.262 (0.119)
Median [Q1, Q3]	0.220 (0.175, 0.275)	0.580 (0.445, 0.905)	0.240 (0.180, 0.330)
Tumor			
SUV_max			
Mean (SD)	11.9 (6.44)	15.0 (7.19)	10.6 (5.95)
Median [Q1, Q3]	11.6 (7.14, 15.4)	14.2 (9.92, 18.8)	9.77 (6.29, 13.8)
SUV_mean			
Mean (SD)	3.87 (2.33)	3.79 (2.23)	3.63 (2.24)
Median [Q1, Q3]	3.17 (2.32, 5.39)	3.24 (2.37, 4.94)	2.90 (2.23, 4.96)

Data represent mean and median SUV_{max}, SUV_{mean}, *and* SUV_{std} values of representative tumors, liver, and mediastinal blood pool

dose ^{18}F-FDG PET scans (kappa = 0.912) than the 6.25% dose ^{18}F-FDG PET scans (kappa = 0.834).

Discussion

Our data show that our CNN model could generate ^{18}F-FDG PET images of the whole body from ultra-low-dose ^{18}F-FDG PET inputs while maintaining clinically relevant information in terms of diagnostic accuracy and quantitative SUV measurements. Reducing the exposure to ionizing radiation from medical imaging procedures is important to minimize a potential risk of secondary cancer development later in life [4, 5, 12]. Our CNN concept takes advantage of simultaneous ^{18}F-FDG PET and MRI data acquisitions and could substantially advance the development of safer imaging tests for pediatric patients. We found that the inclusion of MRI data in addition to ^{18}F-FDG PET images in the CNN model improves the depiction of anatomical detail in PET reconstruction. The rationale for this is that the MRI data provides detailed anatomic information that helps the CNN improve spatial detail in the reconstructed images. In addition, incorporation of attention-weighted loss into our model emphasizes the high-diagnostic-value regions on medical images, which further boosts the model performance. While several other investigators utilized MRI data to augment low-dose ^{18}F-FDG PET images of the brain [19, 20, 22], we applied this concept to whole-body ^{18}F-FDG PET images of children and young adults.

Most prior PET low-dose reconstruction work focused on the brain. Previous studies [22, 27] showed that a low-dose PET scan of the brain can be obtained by combining 75% dose accelerated PET scans with T1-weighted MRI images as inputs to a CNN model to predict standard-dose brain PET images. Besides CNN, other frequently used neural network models in brain PET reconstruction are generative adversarial networks (GANs) [28, 29], where the generator part of a GAN learns to create synthetic images with the goal of fooling the discriminator, which is designed particularly to distinguish between real and synthetic images. Furthermore, several other attempts have focused on dose reduction below 10% [23, 30, 31]. Chen et al. used ultra-low-dose brain PET images and MRI sequences as inputs to create AI-augmented PET scans of the brain [32]. However, brain image reconstruction is fundamentally different from whole-body image reconstruction, which is a much more challenging task due to its much more variable anatomical detail. Two recent works focused on whole-body PET image reconstruction [33, 34]. A residual CNN was proposed to reconstruct full-dose PET images from 10% low-dose counterparts [33]. However, the research was conducted on only two whole-body scans, which largely limits the generalizability of the developed model. In addition, the model was not tested with regard to its ability to render clinical diagnoses. The second proposed scheme was built upon 50% low-dose images, which save substantially less ionizing radiation than provided by our CNN model [34].

There are several limitations of our study. Some small lymph nodes can be less well delineated on the AI-reconstructed PET compared to the original standard-dose PET. In Supplementary Figure 6, we can see that sub-centimeter hypermetabolic lymph nodes are better delineated on the AI-augmented scan than on the 6% dose scan. The AI-augmented scan does not discriminate each individual lesion as well as the original 100% standard-dose scan. This technical limitation will be addressed with further improvements of our algorithm. The limitation of FDG-PET for the detection of sub-centimeter lesions is a well-described problem not only for our studies but for the field in general. Another limitation is the need for simultaneously acquired MRI. There might be situations where only PET images are present or if the PET and MRI data are acquired separately on different scanners. It is worth noting that the proposed model could potentially be applied in a PET/CT scenario where the CT provides the anatomical information for PET reconstruction. While this approach would not save irradiation, it could be used to save time by acquiring ultra-fast PET scans and augmenting them with CT data. Related problems due to sequential rather than simultaneous PET and CT data acquisition would have to be investigated. Meanwhile, hallucination signals could be introduced during reconstruction due to the lack of performance guarantee in deep learning models.

Conclusion

We have demonstrated that high-quality ^{18}F-FDG PET images can be reconstructed from ultra-low-dose inputs using a new CNN that includes the simultaneously acquired MRI data in addition to PET data. The AI-augmented ultra-low-dose ^{18}F-FDG PET images maintain clinically relevant information in terms of diagnostic accuracy, diagnostic confidence, and quantitative SUV measurements.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s00259-021-05197-3.

Code availability Code details will be made available.

Author contributions Guarantors of integrity of entire study: Yan-Ran (Joyce) Wang, Daniel Rubin, Heike E. Daldrup-Link

Study concepts/study design: Yan-Ran (Joyce) Wang, Rong Lu, Daniel Rubin, Heike E. Daldrup-Link

Clinical studies: Lucia Baratto, K Elizabeth Hawk, Allison Pribnow, Avnesh S Thakor, Ashok J Theruvath, Sergios Gatidis, Jordi Garcia-Diaz, Heike E. Daldrup-Link

Data acquisition: Yan-Ran (Joyce) Wang, Lucia Baratto, K Elizabeth Hawk, Ashok J Theruvath, Sergios Gatidis, Jordi Garcia-Diaz, Heike E. Daldrup-Link

Data analysis/interpretation: all authors

Literature research: Yan-Ran (Joyce) Wang, Lucia Baratto, Ashok J Theruvath, Jordi Garcia-Diaz

Statistical analysis: Rong Lu, Santosh E Gummidipundi

Manuscript drafting or manuscript revision for important intellectual content: all authors

Approval of final version of submitted manuscript: all authors

Agrees to ensure any questions related to the work are appropriately resolved: Daniel Rubin, Heike E. Daldrup-Link

Funding This study was supported by a grant from the Eunice Kennedy Shriver National Institute of Child Health and Human Development (NICHD, R01 HD081123A) and the Andrew McDonough B+ Foundation. Statistical analysis for this work was also partially supported by the Biostatistics Shared Resource, which is funded by the Cancer Center Support Grant, P30CA124435.

Data availability Data of this project will be made available.

Declarations

Conflict of interest The authors declare no conflict of interest.

Ethics approval and consent to participate Research PET/MR imaging studies have been approved by the Institutional Review Board at Stanford University and the University of Tübingen. All patients provided written informed consent to participate in a research PET/MR study and the results of this research will be published.

Pertinent findings Using a cohort study of 23 clinical whole-body ^{18}F-FDG PET/MRI subjects, we demonstrated that the AI-reconstructed ultra-low-dose ^{18}F-FDG PET images resemble high similarity with standard-dose ^{18}F-FDG PET images, based on both quantitative and qualitative clinical evaluations. The proposed PET reconstruction model also generalizes in an independent cohort study of 11 clinical whole-body ^{18}F-FDG PET/MRI subjects.

Implications for patient care We anticipate that our proposed model will enable a new generation of imaging exams for children that can be widely applied to interrogate health and disease without the risk of secondary cancer development later in life.

References

1. Baum SH, Fruhwald M, Rahbar K, Wessling J, Schober O, Weckesser M. Contribution of PET/CT to prediction of outcome in children and young adults with rhabdomyosarcoma. J Nucl Med. 2011;52:1535–40. https://doi.org/10.2967/jnumed.110.082511.
2. Cheng G, Chen W, Chamroonrat W, Torigian DA, Zhuang H, Alavi A. Biopsy versus FDG PET/CT in the initial evaluation of bone marrow involvement in pediatric lymphoma patients. Eur J Nucl Med Mol Imaging. 2011;38:1469–76. https://doi.org/10.1007/s00259-011-1815-z.
3. Kleis M, Daldrup-Link H, Matthay K, Goldsby R, Lu Y, Schuster T, et al. Diagnostic value of PET/CT for the staging and restaging of pediatric tumors. Eur J Nucl Med Mol Imaging. 2009;36:23–36. https://doi.org/10.1007/s00259-008-0911-1.
4. Huang B, Law MW, Khong PL. Whole-body PET/CT scanning: estimation of radiation dose and cancer risk. Radiology. 2009;251:166–74. https://doi.org/10.1148/radiol.2511081300.
5. Brenner DJ, Hall EJ. Computed tomography–an increasing source of radiation exposure. N Engl J Med. 2007;357:2277–84. https://doi.org/10.1056/NEJMra072149.
6. Meulepas JM, Ronckers CM, Smets A, Nievelstein RAJ, Gradowska P, Lee C, et al. Radiation exposure from pediatric CT scans and subsequent cancer risk in the Netherlands. J Natl Cancer Inst. 2018. https://doi.org/10.1093/jnci/djy104.
7. Pearce MS, Salotti JA, Little MP, McHugh K, Lee C, Kim KP, et al. Radiation exposure from CT scans in childhood and subsequent risk of leukaemia and brain tumours: a retrospective cohort study. Lancet. 2012;380:499–505. https://doi.org/10.1016/S0140-6736(12)60815-0.
8. Mathews JD, Forsythe AV, Brady Z, Butler MW, Goergen SK, Byrnes GB, et al. Cancer risk in 680,000 people exposed to computed tomography scans in childhood or adolescence: data linkage study of 11 million Australians. Bmj. 2013;346:f2360. https://doi.org/10.1136/bmj.f2360.
9. Brenner DJ, Doll R, Goodhead DT, Hall EJ, Land CE, Little JB, et al. Cancer risks attributable to low doses of ionizing radiation: assessing what we really know. Proc Natl Acad Sci U S A. 2003;100:13761–6. https://doi.org/10.1073/pnas.2235592100.
10. Klenk C, Gawande R, Uslu L, Khurana A, Qiu D, Quon A, et al. Ionising radiation-free whole-body MRI versus (18)F-fluorodeoxyglucose PET/CT scans for children and young adults with cancer: a prospective, non-randomised, single-centre study. Lancet Oncol. 2014;15:275–85. https://doi.org/10.1016/S1470-2045(14)70021-X.
11. Robbins E. Radiation risks from imaging studies in children with cancer. Pediatr Blood Cancer. 2008;51:453–7. https://doi.org/10.1002/pbc.21599.
12. Chawla SC, Federman N, Zhang D, Nagata K, Nuthakki S, McNitt-Gray M, et al. Estimated cumulative radiation dose from PET/CT in children with malignancies: a 5-year retrospective review. Pediatr Radiol. 2010;40:681–6. https://doi.org/10.1007/s00247-009-1434-z.
13. Applegate KE, Frush DP. Image gently: a decade of international collaborations to promote appropriate imaging for children. J Am Coll Radiol. 2017;14:956–7. https://doi.org/10.1016/j.jacr.2017.04.039.
14. Muehe AM, Theruvath AJ, Lai L, Aghighi M, Quon A, Holdsworth SJ, et al. How to provide gadolinium-free PET/MR cancer staging of children and young adults in less than 1 h: the Stanford approach. Mol Imaging Biol. 2018;20:324–35. https://doi.org/10.1007/s11307-017-1105-7.
15. Theruvath AJ, Siedek F, Muehe AM, Garcia-Diaz J, Kirchner J, Martin O, et al. Therapy response assessment of pediatric tumors with whole-body diffusion-weighted MRI and FDG PET/MRI. Radiology. 2020;192508. https://doi.org/10.1148/radiol.2020192508.
16. Karakatsanis NA, Fokou E, Tsoumpas C. Dosage optimization in positron emission tomography: state-of-the-art methods and future prospects. Am J Nucl Med Mol Imaging. 2015;5:527–47.
17. Kim SM, Alessio AM, De Man B, Asma E, Kinahan PE. Direct reconstruction of CT-based attenuation correction images for PET with cluster-based penalties. IEEE Nucl Sci Symp Conf Rec (1997). 2013;2013. https://doi.org/10.1109/NSSMIC.2013.6829245.
18. Dong X, Wang T, Lei Y, Higgins K, Liu T, Curran WJ, et al. Synthetic CT generation from non-attenuation corrected PET images for whole-body PET imaging. Phys Med Biol. 2019;64:215016. https://doi.org/10.1088/1361-6560/ab4eb7.
19. Kang J, Gao Y, Shi F, Lalush DS, Lin W, Shen D. Prediction of standard-dose brain PET image by using MRI and low-dose brain [18F]FDG PET images. Med Phys. 2015;42:5301–9. https://doi.org/10.1118/1.4928400.
20. Wang Y, Zhang P, An L, Ma G, Kang J, Shi F, et al. Predicting standard-dose PET image from low-dose PET and multimodal MR images using mapping-based sparse representation. Phys Med Biol. 2016;61:791–812. https://doi.org/10.1088/0031-9155/61/2/791.

21. Cui J, Gong K, Guo N, Wu C, Meng X, Kim K, et al. PET image denoising using unsupervised deep learning. Eur J Nucl Med Mol Imaging. 2019;46:2780–9. https://doi.org/10.1007/s00259-019-04468-4.
22. Xiang L, Qiao Y, Nie D, An L, Wang Q, Shen D. Deep auto-context convolutional neural networks for standard-dose PET image estimation from low-dose PET/MRI. Neurocomputing. 2017;267:406–16. https://doi.org/10.1016/j.neucom.2017.06.048.
23. Ouyang J, Chen KT, Gong E, Pauly J, Zaharchuk G. Ultra-low-dose PET reconstruction using generative adversarial network with feature matching and task-specific perceptual loss. Med Phys. 2019;46:3555–64. https://doi.org/10.1002/mp.13626.
24. Sekine T, Delso G, Zeimpekis KG, de Galiza Barbosa F, Ter Voert E, Huellner M, et al. Reduction of (18)F-FDG dose in clinical PET/MR imaging by using silicon photomultiplier detectors. Radiology. 2018;286:249–59. https://doi.org/10.1148/radiol.2017162305.
25. Lim B, Son S, Kim H, Nah S, Mu Lee K. Enhanced deep residual networks for single image super-resolution. Proc IEEE Conf Comput Vis Pattern Recognit. 2017;136–44.
26. Banerjee I, Crawley A, Bhethanabotla M, Daldrup-Link HE, Rubin DL. Transfer learning on fused multiparametric MR images for classifying histopathological subtypes of rhabdomyosarcoma. Comput Med Imaging Graph. 2018;65:167–75. https://doi.org/10.1016/j.compmedimag.2017.05.002.
27. Zaharchuk G. Next generation research applications for hybrid PET/MR and PET/CT imaging using deep learning. Eur J Nucl Med Mol Imaging. 2019;46:2700–7. https://doi.org/10.1007/s00259-019-04374-9.
28. Wang Y, Yu B, Wang L, Zu C, Lalush DS, Lin W, et al. 3D conditional generative adversarial networks for high-quality PET image estimation at low dose. Neuroimage. 2018;174:550–62. https://doi.org/10.1016/j.neuroimage.2018.03.045.
29. Wang Y, Zhou L, Yu B, Wang L, Zu C, Lalush DS, et al. 3D auto-context-based locality adaptive multi-modality GANs for PET synthesis. IEEE Trans Med Imaging. 2019;38:1328–39. https://doi.org/10.1109/TMI.2018.2884053.
30. Sanaat A, Arabi H, Mainta I, Garibotto V, Zaidi H. Projection space implementation of deep learning-guided low-dose brain PET imaging improves performance over implementation in image space. J Nucl Med. 2020;61:1388–96. https://doi.org/10.2967/jnumed.119.239327.
31. Xu J, Gong E, Pauly JM, Zaharchuk G. 200x low-dose PET reconstruction using deep learning. Computer Vision and Pattern Recognition. Cornell University Library. 2018;arXiv:1712.04119.
32. Chen KT, Gong E, de Carvalho Macruz FB, Xu J, Boumis A, Khalighi M, et al. Ultra-low-dose (18)F-florbetaben amyloid PET imaging using deep learning with multi-contrast MRI inputs. Radiology. 2019;290:649–56. https://doi.org/10.1148/radiol.2018180940.
33. Kaplan S, Zhu YM. Full-dose PET image estimation from low-dose PET image using deep learning: a pilot study. J Digit Imaging. 2019;32:773–8. https://doi.org/10.1007/s10278-018-0150-3.
34. Whiteley W, Luk WK, Gregor J. DirectPET: full-size neural network PET reconstruction from sinogram data. J Med Imaging (Bellingham). 2020;7:032503. https://doi.org/10.1117/1.JMI.7.3.032503.