DeepTOFSino

Sanaat, Amirhossein; Shooli, Hossein; Ferdowsi, Sohrab; Shiri, Isaac; Arabi, Hossein; Zaidi, Habib

Published in:
Neuroimage

DOI:
10.1016/j.neuroimage.2021.118697

Publication date:
2021

NeuroImage 245 (2021) 118697

Contents lists available at ScienceDirect

NeuroImage

journal homepage: www.elsevier.com/locate/neuroimage

DeepTOFSino: A deep learning model for synthesizing full-dose time-of-flight bin sinograms from their corresponding low-dose sinograms

Amirhossein Sanaat[a], Hossein Shooli[b], Sohrab Ferdowsi[c], Isaac Shiri[a], Hossein Arabi[a], Habib Zaidi[a,d,e,f,*]

[a] Division of Nuclear Medicine and Molecular Imaging, Geneva University Hospital, Geneva, Switzerland
[b] Persian Gulf Nuclear Medicine Research Center, Department of Molecular Imaging and Radionuclide Therapy (MIRT), Bushehr Medical University Hospital, Faculty of Medicine, Bushehr University of Medical Sciences, Bushehr, Iran
[c] University of Applied Sciences and Arts of Western, Geneva, Switzerland
[d] Geneva University Neurocenter, University of Geneva, Geneva, Switzerland
[e] Department of Nuclear Medicine and Molecular Imaging, University of Groningen, Groningen, Netherlands
[f] Department of Nuclear Medicine, University of Southern Denmark, Odense, Denmark

ARTICLE INFO

Keywords:
PET/CT
Brain imaging
Low-dose imaging
Deep learning
Time-of-flight

ABSTRACT

Purpose: Reducing the injected activity and/or the scanning time is a desirable goal to minimize radiation exposure and maximize patients' comfort. To achieve this goal, we developed a deep neural network (DNN) model for synthesizing full-dose (FD) time-of-flight (TOF) bin sinograms from their corresponding fast/low-dose (LD) TOF bin sinograms.

Methods: Clinical brain PET/CT raw data of 140 normal and abnormal patients were employed to create LD and FD TOF bin sinograms. The LD TOF sinograms were created through 5% undersampling of FD list-mode PET data. The TOF sinograms were split into seven time bins (0, ±1, ±2, ±3). Residual network (ResNet) algorithms were trained separately to generate FD bins from LD bins. An extra ResNet model was trained to synthesize FD images from LD images to compare the performance of DNN in sinogram space (SS) vs implementation in image space (IS). Comprehensive quantitative and statistical analysis was performed to assess the performance of the proposed model using established quantitative metrics, including the peak signal-to-noise ratio (PSNR), structural similarity index metric (SSIM) region-wise standardized uptake value (SUV) bias and statistical analysis for 83 brain regions.

Results: SSIM and PSNR values of 0.97 ± 0.01, 0.98 ± 0.01 and 33.70 ± 0.32, 39.36 ± 0.21 were obtained for IS and SS, respectively, compared to 0.86 ± 0.02and 31.12 ± 0.22 for reference LD images. The absolute average SUV bias was 0.96 ± 0.95% and 1.40 ± 0.72% for SS and IS implementations, respectively. The joint histogram analysis revealed the lowest mean square error (MSE) and highest correlation (R^2 = 0.99, MSE = 0.019) was achieved by SS compared to IS (R^2 = 0.97, MSE= 0.028). The Bland & Altman analysis showed that the lowest SUV bias (-0.4%) and minimum variance (95% CI: -2.6%, +1.9%) were achieved by SS images. The voxel-wise *t*-test analysis revealed the presence of voxels with statistically significantly lower values in LD, IS, and SS images compared to FD images respectively.

Conclusion: The results demonstrated that images reconstructed from the predicted TOF FD sinograms using the SS approach led to higher image quality and lower bias compared to images predicted from LD images.

1. Introduction

Non-invasive functional imaging using positron emission tomography (PET) is the ultimate technique for the visualization and quantification of events at the cellular and molecular levels. PET is widely used for the assessment of neurodegenerative diseases and other neurological disorders (Zaidi et al., 2010). PET reconstruction algorithms generate a three-dimensional (3D) image representing the radiotracer dis-

PET, Positron Emission Tomography; DNN, Deep Neural Networks; LD, Low-dose; FD, Full-dose; ResNet, Residual network; TOF, Time-of-Flight; SNR, Signal-to-Noise Ratio; OP-OSEM, Poisson ordered subsets-expectation maximization; SSIM, Structural Similarity Index Metric; RMSE, Root Mean Squared Error; SUV, Standardized Uptake Value; STD, standard deviation; IS, Image Space; SS, Sinogram Space.

* Corresponding author at: Geneva University Hospital, Division of Nuclear Medicine and Molecular Imaging, CH-1211 Geneva, Switzerland.
E-mail address: habib.zaidi@hcuge.ch (H. Zaidi).

https://doi.org/10.1016/j.neuroimage.2021.118697.
Received 4 July 2021; Received in revised form 21 September 2021; Accepted 29 October 2021
Available online 4 November 2021.

tribution within the body, providing useful information about biological processes *in vivo*. PET is a relatively noisy imaging modality owing to the low statistics of the acquired events and the Poisson nature of annihilation photons emission and detection processes. The technical aspects, including PET scanner's geometry, photodetection technology, reconstruction methodology, along with physiological considerations, such as patient motion (particularly for patients suffering from dementia that are more susceptible to involuntary motion) influence the quality and quantitative accuracy of PET images. Furthermore, the number of collected true events, which has a direct relationship to the scanning/acquisition time and amount of injected radiotracer, impacts PET image quality. Increasing the scanning time causes additional discomfort to elderly and pediatric patients and decreases the scanner's throughput. Conversely, increasing the amount of injected radiotracer faces concerns with respect to potential radiation hazards, particularly in children and patients undergoing multiple PET examinations at different time intervals for follow-up or monitoring of treatment response (Arabi et al., 2021; Sanaat et al., 2021). Reducing the injected activity or scanning time hampers image quality and decreases the signal-to-noise ratio (SNR), which might jeopardize clinical diagnosis and biased quantification. Hence, techniques enabling to reduce the injected activity or scanning time while maintaining the clinical information of PET images similar to standard dose scans are highly desired.

In recent years, a plethora of deep neural networks (DNN) were proposed to address the issue of noise reduction and improvement of image quality in low-dose (LD) PET imaging while preserving clinical information (Zaidi and El Naqa, 2021). Artificial intelligence algorithms are black box techniques that can learn any type of nonlinear transformation for a large number of tasks, specifically in the image-to-image translation domain. During the last decade, a number of DNN models were developed for image-to-image translation, demonstrating reasonable performance in direct PET image reconstruction mapping the representations from multi-modal images, such as data from sinogram to image domains (Häggström et al., 2019; Sanaat et al., 2020; Sanaat and Zaidi, 2020), cross-modality multisequence MR to PET conversion (Wei et al., 2020), generation of synthetic CT from MR images (Arabi et al., 2019; Han, 2017), internal radiation dosimetry where CT and PET images are fed into the model to predict the dose map (Akhavanallaf et al., 2021), synthesizing PET/CT attenuation corrected from non-attenuated corrected images (Arabi et al., 2020; Yang et al., 2019), and image denoising (Chen et al., 2019; Kang et al., 2015; Kaplan and Zhu, 2019; Ouyang et al., 2019; Schaefferkoetter et al., 2017; Wang et al., 2018; Xiang et al., 2017).

Various PET image denoising techniques using DNNs in image space were reported. For instance, Chen et al. adopted a residual U-Net architecture for estimating the full-dose (FD) images from an LD with 200th dose reduction (Xu et al., 2017). Other groups explored the possibility of using anatomical information from concurrent imaging modalities, such as CT or MRI in denoising process to enhance the synthesized FD PET image quality. In this regard, Chen et al. used a U-Net model incorporating MR sequences (T1, T2-weighted, and T2-FLAIR) into the learning process to estimate FD ^{18}F-Florbetaben brain PET images from LD images to 1% of the corresponding acquisition time of FD images (Chen et al., 2019).

Most previous studies using deep learning-guided PET image denoising in the image domain. One of the disadvantages of using reconstructed images is that the models are trained for a specific clinical protocol, which limits the applicability of the model to the specific reconstruction protocol adopted (parameters, post-reconstruction filtering, etc.). Hence, switching to a different clinical protocol entails retraining again the DNN model. Conversely, the prediction of FD sinograms in the projection domain makes it possible to apply various image reconstruction protocols. Such an approach was adopted in our recent work where we developed two U-Net models: one conventional to synthesize FD images from LD images and another to synthesize FD non-time-of-flight (TOF) sinograms from LD non-TOF sinograms (Sanaat et al., 2020). The comparison of these two approaches proved that FD images reconstructed from the predicted sinograms effectively reduced the noise level and exhibited superior performance in terms of image quality and quantitative accuracy. Hong et al. (2018) proposed a deep residual sinogram super-resolution network to improve the quality of images produced by a PET scanner equipped with large pixelated crystals. They trained their model using sinograms from a scanner with thin crystals, to generate high-resolution sinograms. Furthermore, more recent work reported on prior knowledge-driven DNN-based approach for sinogram denoising (Lu et al., 2020).

TOF PET imaging provides higher signal-to-noise ratio and overall improved image quality compared to nonTOF PET, particularly for large objects (Surti, 2015). This enables faster/lower dose PET imaging. In this regard, denoising the TOF PET data in the sinogram domain is highly desired to take full advantage of the potential of TOF PET imaging.

One of the limitations of our above referenced previous work is that TOF was not considered, and as such, the technique was not exploited to its full potential (Sanaat et al., 2020). In this work, we set out to develop a model consisting of seven Residual network (ResNet) architectures to predict seven FD sinograms corresponding to various TOF bins (0, ±1, ±2, ±3) from their corresponding LD TOF bin sinograms in an end-to-end fashion. Thereafter, the generated TOF sinograms could be reconstructed using any PET image reconstruction algorithm. We used different ResNet models for training the network in image space. The synthesized images derived using both methodologies (image space and sinogram space) were then compared.

2. Materials and methods

2.1. PET/CT data acquisition

The current study was carried out using a database consisting of 140 ^{18}F-FDG brain PET/CT studies collected between June 2017 and May 2019 at Geneva University Hospital. The dataset covers a wide range of patients with cognitive symptoms of possible neurodegenerative diseases. The patients consisted of 66 males (73 ± 9 yrs) and 74 females (72 ± 11 yrs). The demographic information of patients is summarized in Table 1. The study protocol was accepted by the institution's ethics committee and all patients gave written informed content. List-mode data were acquired on a Biograph mCT scanner (Siemens Healthineers, Erlangen, Germany) about 35 min post-injection. A low-dose CT scan (120 kVp, 20 mAs) was performed for PET attenuation correction. A standard imaging protocol consisting of 20 min scanning time following the injection of 205 ± 10 MBq of ^{18}F-FDG. PET list-mode data were first binned into sinograms and then reconstructed using the e7 tools offline reconstruction toolkit (Siemens Healthcare). An extension of this toolkit, referred to as "decimate.js", enables to produce randomly undersampled TOF sinograms corresponding to a predefined percentage of the total number of collected events, thus allowing to generate low-dose scans from full-dose scans. Each produced TOF sinogram has a matrix size of 400 × 168 × 621 × 13, with the fourth dimension representing the number of time bins for the Biograph mCT having a TOF coincidence time resolution of ~530 ps (0, ±1, ±2, ±3, ±4, ±5, ±6). Since the head is always positioned in the center of the field-of-view and given that the size of the head, it is commonly contained within only 7 bins (0, ±1, ±2, ±3). Hence, we trained 7 models corresponding to these time bins. A random subset containing 5% of the total events was extracted from the list-mode data to produce a LD TOF sinogram. An ordinary Poisson ordered subsets-expectation maximization (OP-OSEM) algorithm considering TOF and point spread function modeling with 5 iterations and 21 subsets was used for reconstruction of both FD and LD PET images with a matrix size of 200 × 200 × 109 and 2.03 × 2.03 × 2.2 mm^3 voxel size. Gaussian post-reconstruction filtering with 2 mm FWHM was applied.

Table 1
Demographic information of patients included in this study.

	Training	Test	Validation
Number	100	30	10
Male/Female	45/55	15/15	4/6
Age (Mean ± SD)	73±8	71±16	69±7.5
Indication/Diagnosis	Cognitive symptoms of possible neurodegenerative an etiology		

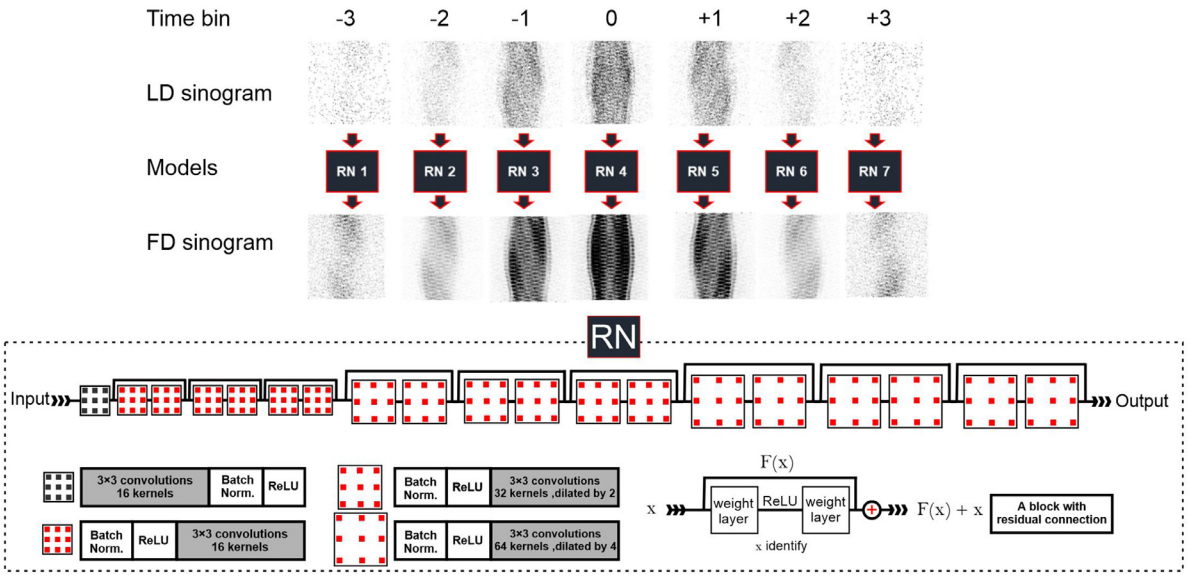

Fig. 1. A schematic of the ResNet network used in this work. In seven separate steps low dose sinograms time bins (−3, −2,−1,0,1,2,3) were fed in 7 separate ResNet models as input and full dose sinograms belong to each time bin predicted. In image space training phase, we fed a ResNet model with TOF LD images to predict TOF FD images.

2.2. ResNet architecture

The core element to any ResNET is the residual block originally proposed in (He et al., 2016) along with its later variants. The residual connection circumvents the many issues faced while increasing the number of layers of neural networks and efficiently facilitates training very deep networks. This has revolutionized the deep learning practice across many disciplines, including the application targeted in this work, i.e., image-to-image translation.

A general description of a residual connection is depicted in Fig. 1. Let $f_\theta^{[l]}(\cdot)$ the l^{th} module of a DNN, where $l = 1, \cdots, L$ and θ symbolizing the set of all learnable parameters of the neural network, which are updated during training. This usually consists of a couple of convolutional layers with ReLU non-linearities in between. Let $\mathbf{x}^{[l-1]}$ denote the input to this convolutional module, where $\mathbf{x}^{[0]} = \mathbf{x}$ is the input image. The residual block produces the output as follows:

$$\mathbf{x}^{[l]} = \mathbf{x}^{[l-1]} + f_\theta^{[l]}\left(\mathbf{x}^{[l-1]}\right) \tag{1}$$

which is then fed to the next block. This has many benefits, including the fact that the network will have the possibility to circumvent a module if it slows down the optimization, e.g., due to exploding or vanishing gradient issues (Xie et al., 2017). Moreover, from the perspective of convolution, it provides a diverse set of effective receptive fields.

This basic idea is then applied in various ways using modern networks. U-Net proved to be a successful concept in image-to-image translation and is very popular in the medical image processing and analysis domains (Huang et al., 2017). Consider a simplified structure of a U-Net with 3 left modules $f_\downarrow^{[l]}(\cdot)$, as well as 3 right modules $f_\uparrow^{[l]}(\cdot)$. The output of any left module is downsampled and fed to the next left module, in addition to being directly fed into its corresponding right module, which also receives up-sampled signals from the previous right module. Therefore, the final output is formed as:

$$\mathbf{x}^{[L]} = f_\uparrow^{[L]}\left(f_\downarrow^{[1]}\left(\mathbf{x}^{[0]}\right) + \mathbf{x}^{[L-1]}\right) \tag{2}$$

Hence the output image gets directly a signal from the first layer. While this is successful in many tasks (e.g. image segmentation), for tasks where the input-output image pairs are very similar like in auto-encoding, the network uses the easier direct pathway from the input to the output. A trial and error exercise revealed that no matter how many modules exist in the U-Net, in such cases, the signal does not propagate through them and the output relies mostly on the first module.

In our case, where the low-dose and full-dose images are somehow similar, we noticed this phenomenon by not achieving satisfactory results. This led us to chose more sequential ResNet variants. In particular, we adopted the ResNet structure proposed in (Ronneberger et al., 2015) depicted in Fig. 1, also implemented in NiftyNet (Gibson et al., 2018).

This network has a relatively simple residual structure and avoids down-sampling and up-sampling operations, which complicates the implementation process (Li et al., 2017). Instead, it makes extensive use of dilated convolutions that increase effectively the receptive field of the network, so that it learns longer range entities in the image. Another very well-known element in this network is batch-normalization whose combination with residual connections has become a standard and very effective recipe to speed-up the training of deep models.

Each of the first 19 modules of the network exclusively uses convolutional kernels of size $3 \times 3 \times 3$, along with batch-normalization and ReLU. In the first 7 modules, the network uses 16 of these kernels, the following 6 modules use 32 kernels, but with a dilation parameter of 2, and the next 6 modules use 64 kernels with dilation 4. The last layer consists of a convolutional kernel of size $1 \times 1 \times 1$ to adjust the number of output layers to those of the high-dose images.

We used the simple squared l_2 loss function for training, i.e., $l(\mathbf{x}_i, \mathbf{x}_o) = ||\mathbf{x}_i - \mathbf{x}_o)||_2^2$ for both of the pixel and projection domains. If you use l_2 , you should deactivate softmax. The network was trained and tested using NVIDIA 2080Ti GPU with 11 GB random access memory running under windows 10. The training was performed for 300 epochs.The training and validation of the network was performed on 110 patients, while a separate unseen dataset of 30 patients served as the test dataset.

2.3. Evaluation strategy

2.3.1. Quantitative analysis

For estimation of the accuracy of our DNN models, three well-established quantitative metrics, including peak signal-to-noise ratio (PSNR), structural similarity index metric (SSIM) and the root mean squared error (RMSE) were used to compare IS and SS implementations with FD. Moreover, to have an insight about the level of SNR in LD images and figure out how much ResNet can increase SNR, these metrics were also calculated for the LD images.

The standardized uptake values (SUVs) and their standard deviations (STDs) were estimated for 83 brain regions to evaluate the quantitative accuracy between reference FD, LD, and synthesized images (IS and SS) images. The region-wise analysis was performed using PMOD medical image analysis software (PMOD Technologies LLC, Switzerland) by considering a brain atlas consisting of 15 males and 15 females with an average age of 31 years (Hammersmith atlas *n30r83*).

A joint histogram analysis was carried out to assess voxel-wise the correlation between the activity concentration in ground truth FD and LD, and IS and SS images. In addition, Bland & Altman analysis and folded empirical cumulative distribution plot or Mountain plot were drawn to compare the distribution of regions' SUVs for different regions between the reference and synthesized images. A mountain plot was generated by calculation of a percentile for LD, IS, SS images and reference FD images. In this plot, the variation between the two images is depicted on the length of tails. To evaluate the RMSE, SSIM, and PSNR metrics, statistical analysis using pairwise *t*-test was performed between IS vs SS, IS vs LD, and SS images. For all comparisons, the threshold of statistical significance was set at 5%.

2.3.2. Statistical analysis

Original FD, LD, IS and SS images were pre-processed using FSL (FMRIB Software Library v6.0.1, Analysis Group, FMRIB, Oxford, UK). Firstly, the Hammersmith atlas was registered to an ^{18}F-FDG brain PET template (in Montreal Neurological Institute standard space) using 12-affine transformation registration using FLIRT (FMRIB's Linear Image Registration Tool). Subsequently, non-brain regions of the atlas (cerebellum and brain stem) were excluded and binarized to create a Hammersmith atlas-based brain mask. Then, all FD images were registered to the ^{18}F-FDG PET template with 12-affine transformation registration using FLIRT. Afterwards, LD, IS, and SS images of each subject were registered to the template via FLIRT using the same transformation matrix used in FD image registration step for that subject. Finally, all FD, LD, IS, and SS images were masked using the brain mask to exclude non-brain regions in all images. We applied a linear image registration method that do not change voxel's values without smoothing to minimize effect of pre-processing on the results.

After pre-processing steps, a mass univariate methodology of Statistical Parametric Mapping (SPM12; Welcome Centre for Human Neuroimaging, UCL, UK) was used to conduct a voxel-by-voxel two-sample *t*-test comparing voxelwise FD images with the corresponding LD, IS, and SS images (FD vs LD, FD vs IS, and FD vs SS) (Friston et al., 1994). This analysis identifies voxels with significant difference with respect to FD images. Statistical significance was defined at a voxel-wise threshold (family-wise error corrected $p<0.05$) and no extent threshold of contiguous voxels was defined. In addition, the number, percentage, and level of significance (T-value) of voxels with statistical significance were calculated in 83 brain regions based on Hammersmith atlas.

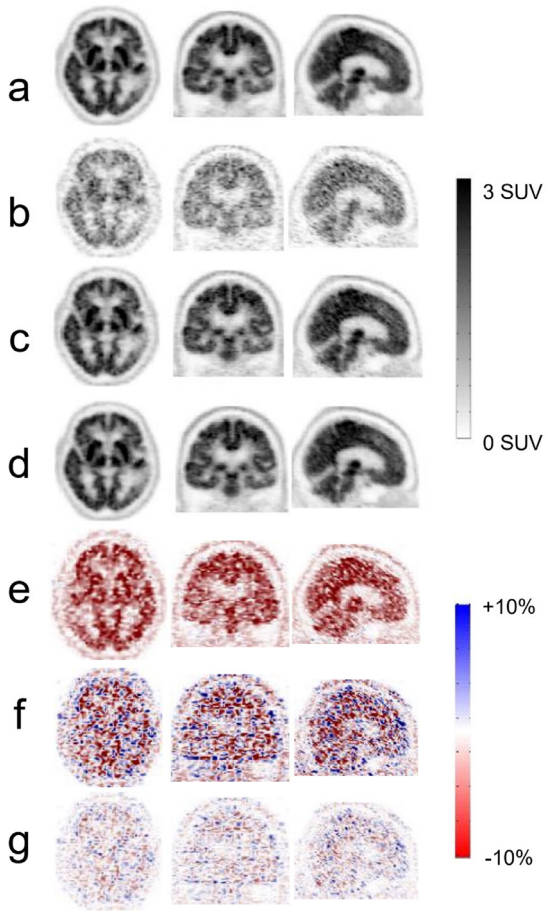

Fig. 2. ^{18}F-FDG brain PET images of a 71-year-old female patient with cognitive symptoms of possible neurodegenerative etiology showing (a) the TOF full-dose image used as reference and (b) low-dose TOF PET image. (c) The predicted standard dose PET images in image space show significantly reduced noise compared with low-dose PET images, while the images generated from sinogram space (d) were superior in reflecting the anatomical features. The bias map for low dose and predicted images in image space and sinogram space with respect to FD images are illustrated in (e), (f), and (g), respectively.

To check sample homogeneity in each dataset, we used "*check sample homogeneity*" from the CAT12 toolbox (Departments of Psychiatry and Neurology, Jena University Hospital, Germany) embedded within SPM12. This tool uses a mean correlation of each image as a measure of homogeneity to identify outlier images in a sample. Accordingly, the correlation is computed between all images across the sample and the mean for every image is calculated. Afterwards, a violin plot was used to visualize the frequency distribution of the mean correlation in each imaging group separately.

3. Results

Overall, the achieved image quality using deep learning approaches both in image space and projection space was almost similar to ground truth FD images, definitely outperforming LD images. Transverse, coronal, and sagittal views of brain images and their corresponding bias maps for FD, LD, IS, and SS of a patient with a large bias between LD and FD were depicted in Fig. 2. The inspection of intensity profiles revealed that the brain anatomy, especially gyrus structures and radioactive uptake patterns, are sharper and are more observable and distinguishable on images reconstructed from predicted TOF sinograms (SS)

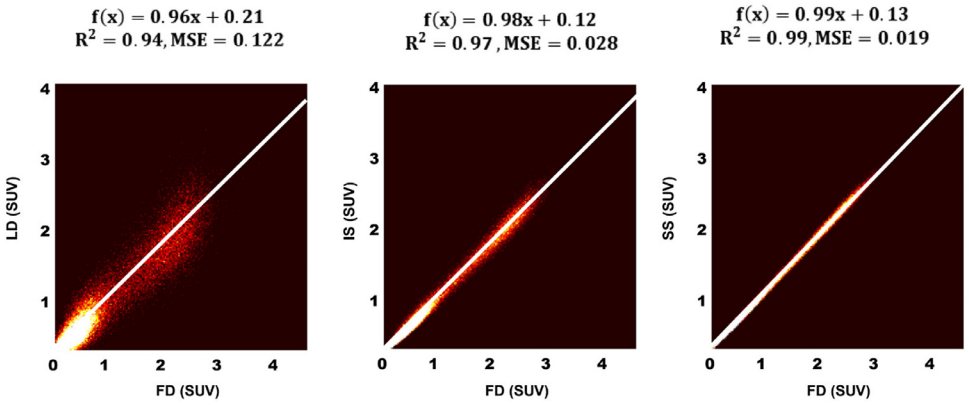

Fig. 3. Joint histograms analysis of low-dose image (left) and predicted images in image space (middle) and predicted full-dose in sinogram space (right) versus standard dose image serving as reference.

Table 2
Statistical analysis of image quality metrics for low dose images, test predicted images in sinogram and image space. SSIM structural similarity index metrics, PSNR, peak signal to noise ratio, RMSE, root mean squared error.

Test dataset	SSIM	PSNR	RMSE
Low-dose (LD)	0.86±0.02	31.12±0.22	0.35±0.06
Image Space (IS)	0.97±0.01	33.70±0.32	0.17±0.03
Sinogram Space (SS)	0.98±0.01	39.36±0.21	0.14±0.09
P-value (IS vs. SS)	**0.041**	**0.046**	**0.038**
P-value (IS vs. LD)	**0.023**	**0.032**	**0.022**
P-value (SS vs. LD)	**0.016**	**0.028**	**0.019**

Table 3
The percentage of average and absolute average of SUV bias and standard deviation for LD, IS, SS for all regions.

	LD	IS	SS
Average SUV bias	0.17±1.96	0.98±1.34	0.22±1.77
Absolute average for SUV bias	1.59±1.03	1.40±0.72	0.96±0.95

relative to those predicted in image space (IS). This vision inspection was supported by quality metrics include PSNR, SSIM, and RMSE were calculated between FD and LD, IS, and SS in Table 2. The noise degradation and quality enhancement of the model were trained with sinogram is significantly higher than the model was trained in images space.

In terms of quantitative accuracy, the average and absolute average SUV bias in all regions for LD, IS and SS are shown in Table 3. Our results show that the lowest absolute average SUV bias (0.96 ± 0.95%) was achieved for SS across all the 83 regions while IS and LD showed an absolute average SUV bias of 1.40 ± 0.72% and 1.59 ± 1.03%, respectively.

The pixel-wise linear regression analysis of ^{18}F-FDG tracer uptake for LD, IS and SS with respect to FD images was illustrated in Fig. 3. The scatter of the data points decreases from LD to IS and then SS and the correlation increased for IS (R^2 = 0.97, MSE = 0.028) compared to LD (R^2 = 0.94, MSE = 0.122). Yet, the highest correlation and lowest scatter were achieved by SS (R^2 = 0.99, MSE = 0.019).

The Mountain and Bland & Altman plots were shown in Fig. 4, reflecting the bias and variance of LD and synthesized images in the 83 brain regions. Each orange circle represents a brain region. The Mountain plots indicate that the tail length decreases significantly in FD – SS plot compared with FD – LD plot. The Bland & Altman plots revealed that the lowest SUV bias (−0.4%) and minimum variance (95% CI: −2.6%, +1.9%) was achieved by predicted SS images. Although LD images show a lower SUV bias (0.1%), the variance is high (95% CI: −3.9, +4.1), reflecting poor image quality. The Bland & Altman and Mountain plots are in good agreement with the joint histogram analysis results.

Table 4
The number of voxels with significant difference in LD, IS and SS compared with FD. The corresponding volume and mean *t*-test value are shown.

	Number of voxels	Volume (mm^3)	Mean t-value
FD - LD	240,845	1,926,760	10.470031
FD - IS	184,448	1,475,584	7.590589
FD - SS	17,322	138,576	5.888064

Fig. 5 shows the SUV bias and STD of ^{18}F-FDG uptake (The STD inside each region) for 83 brain regions extracted from the Hammersmith brain atlas for LD, IS, and SS. To reduce the number of brain regions, we merged the symmetrical left/right sides and reported the results for 44 instead of 83 regions. The graph showed that the magnitude of SUV bias is mostly below 3% for LD, IS, and SS, while LD showed a high STD in each region compared to IS and SS, reflecting a higher noise level in LD images. Moreover, IS presented with a higher variance compared to SS. Although the low-dose images bear overall low SUV bias (owing to zero-mean Poisson noise), the high noise level and local noise-induced bias may impact clinical assessment.

Fig. 6 and Table 4 depict voxel-wise *t*-test analysis for LD, IS and SS. The results showed that there are voxels with significantly lower values in LD, IS, and SS images compared to FD images. However, there were no voxels with significantly higher values in LD, IS, and SS images. In this work, the mean *t*-test was computed over all voxels in the image. The region-wise *t*-test analysis is shown in Supplemental Table 1. Fig. 7 shows a Violin plot depicting the frequency distribution of mean correlation between FD, LD, IS and SS images, separately.

4. Discussion

We evaluated the performance of a deep learning model for synthesizing diagnostic quality FD images from undersampled LD images corresponding to 5% of the injected radiotracer. We trained seven deep learning models separately to generate FD TOF sinogram bins from their corresponding LD TOF sinogram bins and compared the results with a model trained to synthesize FD images from LD images. A well-optimized 2D ResNet DNN was used for image-to-image translation and for each TOF bin separately. We stopped the DNN model at the lowest training loss. Our hypothesis was that if we split a reconstructed TOF image into seven TOF bin sinograms and train seven independent DNN models for each, the final model would be more accurate and robust than a simple model operating in image space. Implementation in projection space involved the use of extended/detailed data (400 × 168 × 621 × 7 = 292′118′400) compared to image space (101 × 101 × 71=724′271).The model trained in projection space (SS) generated images with higher image quality and lower absolute tracer uptake bias in brain regions compared with images synthesized in image

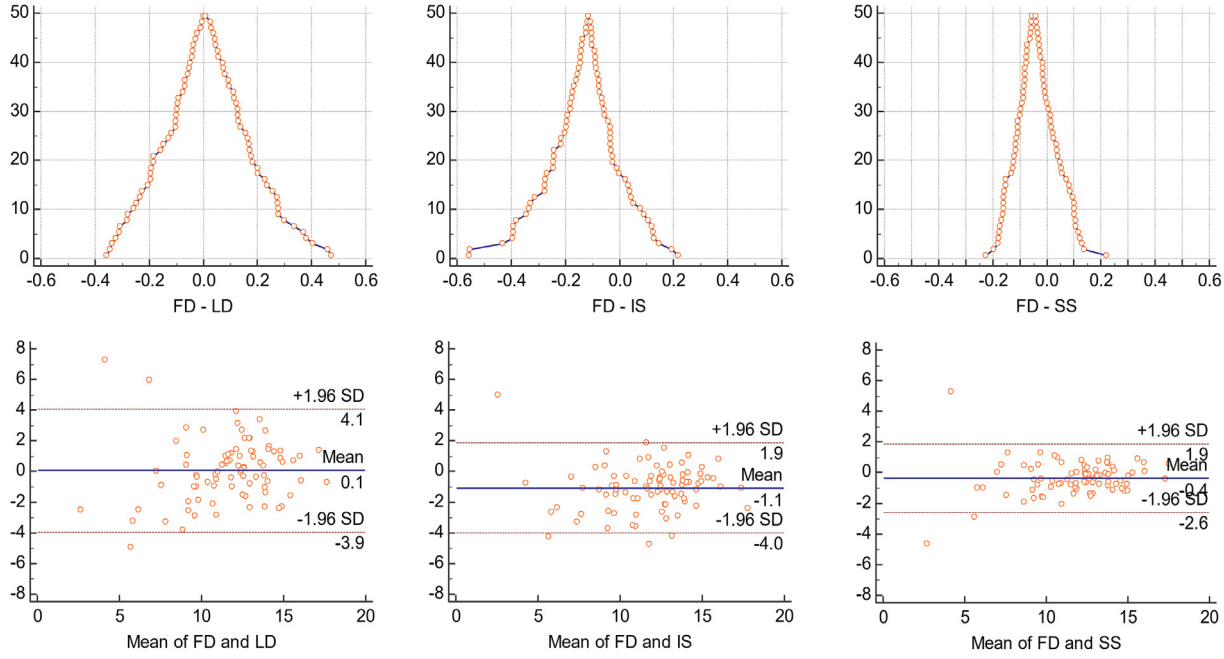

Fig. 4. Mountain and Bland Altman plot of region-wise standardized uptake value comparing FD to LD, IS and SS images across all data sets for the 83 regions. The solid blue and dashed lines denote the mean and 95% confidence interval (CI) of the SUV differences, respectively. In the Mountain plot, the long tails reflect large differences between the methods. For models with lower bias, the mountain will be centered over zero.

space (IS). This confirmed our hypothesis that training seven models for each TOF sinogram bin separately leads to a better performance compared to a single model implementation in image space. We reduced the complexity of the problem by splitting the TOF sinograms into seven separate TOF bins followed by training a model for each bin. Moreover, as different TOF bins contain different signal and noise distributions, independent implementations of denoising networks would lead to more accurate modeling of signal and noise in each TOF bin. In brain PET imaging, since the head is commonly positioned in the center of the FOV, the central TOF bins contain stronger signals and are consequently characterized by higher SNR. Conversely, off-center TOF bins would contain less counts/signal and consequently have lower SNR. In this light, implementation of independent denoising networks for each TOF bin would more accurately model the noise distribution associated with each TOF bin. A model combining all TOF bins would lead to overestimation or underestimation of noise in certain TOF bins.

Overall, by training several models for the different TOF bins, the effect of motion can be reduced, particularly in whole-body PET imaging. Training separate models for each time bin limits the effect of motion to only one time bin and prevents spreading it to all TOF bins (e.g. motion of part of patient's body, such as one arm). This strategy helps to decrease motion blur compared to training a single model for all bins or even training of non-TOF sinograms.

Comparison of the results presented in this work to previous studies should take into consideration some important factors, such as scanning time, injected dose, the time between injection and scanning, scanner sensitivity and count-rate performance and data preparation. Ouyang et al. (2019) proposed a DNN model for synthesizing FD images from LD images with 1% of the full-dose along with three MRI sequences. They reported a SSIM of 0.98 (compared with 0.97 and 0.98 for IS and SS, respectively) and RMSE of 0.14 (compared to 0.17 and 0.14 for IS and SS, respectively). It is worth emphasizing that although they used LD images with five times lower dose percentage relative to our study, the sensitivity of their PET scanner (GE SIGNA PET/MRI) was substantially higher (more than 2 times higher sensitivity than our Biograph mCT). In addition, the amount of radiotracer injected to their patients was substantially higher (330 ± 30 MBq vs 205 ± 10 MBq) and they also incorporated co-registered MR images as support. Compared to our previous work where we focused on non-TOF (Sanaat et al., 2020), this work demonstrated that adding TOF information improves the quality of LD and FD images, which can lead to faster convergence of models trained either in the image or sinogram domain. Conversely, training several DNN models for each time bin can add extra error in each predicted bin that might cause error accumulation in the final TOF sinogram. Hence, the process of optimization and training of the networks is critical and complex.

Conventional PET image denoising techniques, such as non-local mean and bilateral filtering, would inevitably cause signal loss and/or uptake bias within the noise suppression process (Arabi and Zaidi, 2018), particularly in the presence of the high noise levels (Arabi and Zaidi, 2020). However, deep learning-based solutions enable reliable noise suppression without introducing noticeable signal loss and/or uptake bias, which renders them suitable for clinical adoption (Sanaat et al., 2020, 2021). Comparison of various deep learning architectures or algorithms is commended to achieve a reliable model enabling effective noise suppression with minimal signal loss or image artifacts. This effort warrants a thorough investigation and falls beyond the scope of this study. Nevertheless, implementation of deep learning approaches in the sinogram domain should be seriously considered since they enable data—driven selection of the most suitable reconstruction algorithm and/or settings (Arabi and Zaidi, 2021).

The generalizability and robustness of deep learning models is an important factor that determines to what degree the model's output are trustable and robust when testing with new normal/abnormal unseen datasets. We used a heterogeneous dataset consisting of both healthy and abnormal patients with various types of neurodegenerative disorders. Voxel-wise analysis evaluated each voxel in all images (LD, IS, and SS) compared to FD images to determine whether there is a statistically significant difference at the voxel level. This approach offers an overall accuracy measure of the model for accurate prediction independent of subjects' health condition. Therefore, models with less significantly different voxels would be more similar to FD images and

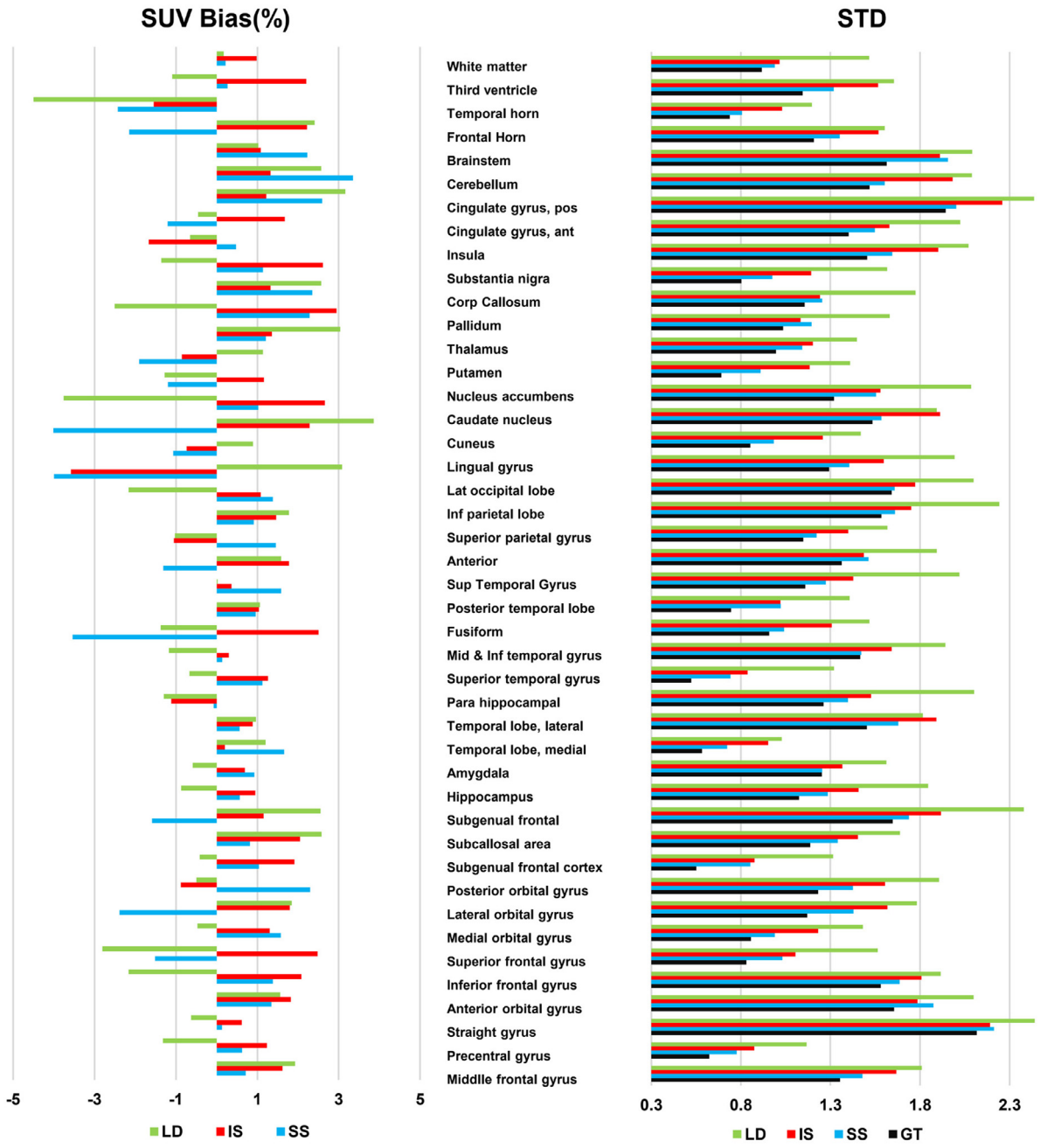

Fig. 5. The SUV bias relative to FD images and standard deviation for LD, IS, SS and FD. To make the plot more readable, we report the average of left and right regions (44 regions).

vice versa. This analysis revealed that the SS model is more accurate than the IS model. However, the number of voxels with significant difference were lower in both IS and SS models than that in LD images. Hence, there is more similarity between the predicted models (IS and SS) and FD group than between LD group and FD group. Moreover, voxel-wise statistical analysis enabled to quantify the number of voxels with significant differences in various regions of the brain in the predicted images (IS and SS) compared to FD images (supplementary Table 1).

The Mountain and Bland-Altman analysis revealed that the model trained in projection space led to lower bias and variance relative to image space. In the Mountain plot, the median of the differences is close to zero for LD and then shift to left for IS and SS, reflecting the overestimation of tracer uptake of both models. The shorter tails were observed for SS, IS, and LD images, respectively, demonstrating the level of difference with reference FD images.

One of the drawbacks of the present study was that the LD images were produced by random undersampling the list-mode data instead of re-injecting the patient with 5% of the amount of radiotracer and rescanning. Another limitation of or work is the limited size of GPU's RAM which prevented training the model using the whole TOF sinogram simultaneously. In addition, patient motion spatially for pediatric and elderly patients who suffer from dementia may lead to a mismatch between CT and PET images and cause the image quality degradation for both LD and FD PET images. Training the model in 2D is another limitation of this work. Extension of the same approach for whole body imaging would be time-consuming and if separate models are trained for each sinogram bin.

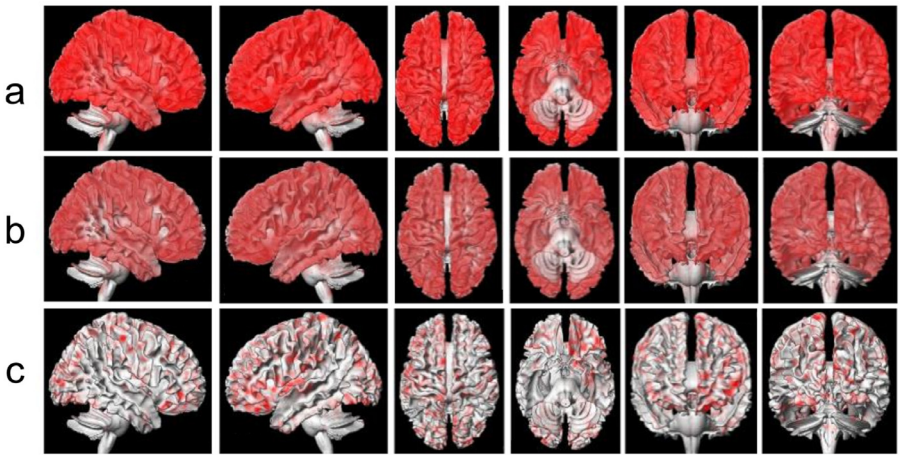

Fig. 6. 3D-rendered views of voxel-wise analysis of LD (a), IS (b), and SS (c) images in comparison with original FD images. The images depict regional differences in three image groups (LD, ISS, and SS) compared with FD images. The red regions represent voxels with significant difference, while white regions indicate voxels without significant difference.

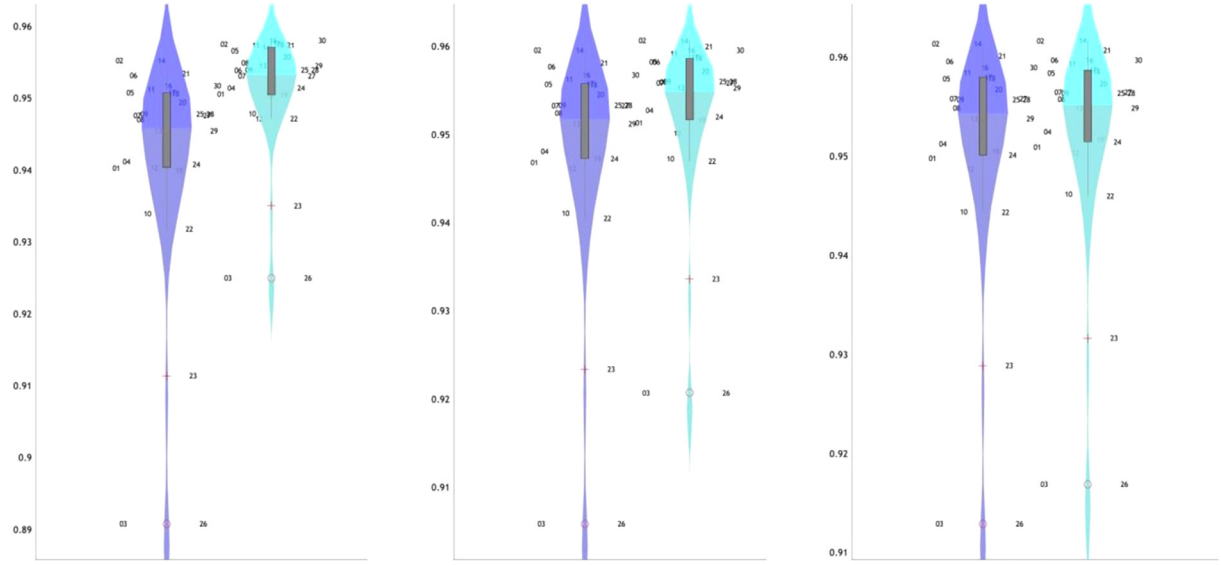

Fig. 7. Violin plot of the mean correlation FD, LD (left panel), IS (middel panel), and SS (right panel). The violin plot depicts the frequency distribution of mean correlation in different images. The distribution pattern of the data is considered as the similarity measure in order to assess the level of difference between groups. In addition, it can be regarded as a measure to evaluate the accuracy of models' prediction. According to this plot, the SS model offers a higher accuracy compared to IS model for predicting ground truth FD images correctly.

5. Conclusion

We demonstrated that training separate ResNet algorithms for mapping the LD TOF sinogram bins to their corresponding FD TOF sinogram bins enables to generate high-quality ^{18}F-FDG brain PET images. The results revealed the superior performance of the TOF model implemented in projection space compared to the implementation in image space. The predicted images in projection space led to better noise characteristics and overall lower absolute tracer uptake bias.

Data and code availability statement

The MatLab/Python code is available from the authors upon request.

Credit authorship contribution statement

Amirhossein Sanaat: Conceptualization, Methodology, Software, Writing – review & editing, Writing – original draft. **Hossein Shooli:** Data curation, Writing – review & editing, Writing – original draft, Methodology. **Sohrab Ferdowsi:** Conceptualization, Writing – review & editing, Writing – original draft. **Isaac Shiri:** Conceptualization, Writing – review & editing, Writing – original draft. **Hossein Arabi:** Conceptualization, Writing – review & editing, Writing – original draft. **Habib Zaidi:** Writing – review & editing, Writing – original draft, Methodology, Conceptualization.

Acknowledgments

This work was supported by the Swiss National Science Foundation under grant SNRF 320030_176052, the Eurostars program of the European commission under grant E! 114021 ProVision and the Private Foundation of Geneva University Hospitals under Grant RC-06–01.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.neuroimage.2021.118697.

References

Akhavanallaf, A., Shiri, I., Arabi, H., Zaidi, H., 2021. Whole-body voxel-based internal dosimetry using deep learning. Eur. J. Nucl. Med. Mol. Imaging 48, 670–682.

Arabi, H., AkhavanAllaf, A., Sanaat, A., Shiri, I., Zaidi, H., 2021. The promise of artificial intelligence and deep learning in PET and SPECT imaging. Phys. Med. 83, 122–137.
Arabi, H., Bortolin, K., Ginovart, N., Garibotto, V., Zaidi, H., 2020. Deep learning-guided joint attenuation and scatter correction in multitracer neuroimaging studies. Hum. Brain Mapp. 41, 3667–3679.
Arabi, H., Zaidi, H., 2018. Improvement of image quality in PET using post-reconstruction hybrid spatial-frequency domain filtering. Phys. Med. Biol. 63, 215010.
Arabi, H., Zaidi, H., 2020. Spatially guided nonlocal mean approach for denoising of PET images. Med. Phys. 47, 1656–1669.
Arabi, H., Zaidi, H., 2021. Non-local mean denoising using multiple PET reconstructions. Ann. Nucl. Med. 35, 176–186.
Arabi, H., Zeng, G., Zheng, G., Zaidi, H., 2019. Novel adversarial semantic structure deep learning for MRI-guided attenuation correction in brain PET/MRI. Eur. J. Nucl. Med. Mol. Imaging 46, 2746–2759.
Chen, K.T., Gong, E., de Carvalho Macruz, F.B., Xu, J., Boumis, A., Khalighi, M., Poston, K.L., Sha, S.J., Greicius, M.D., Mormino, E., Pauly, J.M., Srinivas, S., Zaharchuk, G., 2019. Ultra-low-dose (18)F-Florbetaben amyloid PET imaging using deep learning with multi-contrast MRI inputs. Radiology 290, 649–656.
Friston, K.J., Holmes, A.P., Worsley, K.J., Poline, J.P., Frith, C.D., Frackowiak, R.S., 1994. Statistical parametric maps in functional imaging: a general linear approach. Hum. Brain Mapp. 2, 189–210.
Gibson, E., Li, W., Sudre, C., Fidon, L., Shakir, D.I., Wang, G., Eaton-Rosen, Z., Gray, R., Doel, T., Hu, Y., 2018. NiftyNet: a deep-learning platform for medical imaging. Comput Meth Prog Biomed 158, 113–122.
Häggström, I., Schmidtlein, C.R., Campanella, G., Fuchs, T.J., 2019. DeepPET: a deep encoder–decoder network for directly solving the PET image reconstruction inverse problem. Med. Image Anal. 54, 253–262.
Han, X., 2017. MR-based synthetic CT generation using a deep convolutional neural network method. Med. Phys. 44, 1408–1419.
He, K., Zhang, X., Ren, S., Sun, J., 2016. Deep residual learning for image recognition. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 770–778.
Hong, X., Zan, Y., Weng, F., Tao, W., Peng, Q., Huang, Q., 2018. Enhancing the image quality via transferred deep residual learning of coarse PET sinograms. IEEE Trans. Med. Imaging 37, 2322–2332.
Huang, G., Liu, Z., Van Der Maaten, L., Weinberger, K.Q., 2017. Densely connected convolutional networks. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 4700–4708.
Kang, J., Gao, Y., Shi, F., Lalush, D.S., Lin, W., Shen, D., 2015. Prediction of standard-dose brain PET image by using MRI and low-dose brain [18F]FDG PET images. Med. Phys. 42, 5301–5309.
Kaplan, S., Zhu, Y.M., 2019. Full-dose PET image estimation from low-dose PET image using deep learning: a pilot study. J. Digit. Imaging 32, 773–778.
Li, W., Wang, G., Fidon, L., Ourselin, S., Cardoso, M.J., Vercauteren, T., 2017. On the compactness, efficiency, and representation of 3D convolutional networks: brain parcellation as a pretext task. In: International Conference on Information Processing in Medical Imaging. Springer, pp. 348–360.
Lu, S., Tan, J., Gao, Y., Shi, Y., Liang, Z., 2020. Prior knowledge driven machine learning approach for PET sinogram data denoising. Medical Imaging 2020: physics of Medical Imaging. Int. Soc. Opt. Photonics, 113124A.
Ouyang, J., Chen, K.T., Gong, E., Pauly, J., Zaharchuk, G., 2019. Ultra-low-dose PET reconstruction using generative adversarial network with feature matching and task-specific perceptual loss. Med. Phys. 46, 3555–3564.
Ronneberger, O., Fischer, P., Brox, T., 2015. U-net: convolutional networks for biomedical image segmentation. In: International Conference on Medical image computing and computer-assisted intervention. Springer, pp. 234–241.
Sanaat, A., Arabi, H., Mainta, I., Garibotto, V., Zaidi, H., 2020. Projection-space implementation of deep learning-guided low-dose brain PET imaging improves performance over implementation in image-space. J. Nucl. Med. 61, 1388–1396.
Sanaat, A., Mirsadeghi, E., Razeghi, B., Ginovart, N., Zaidi, H., 2021. Fast dynamic brain PET imaging using stochastic variational prediction for recurrent frame generation. Med. Phys. 48 (9), 5059–5071. doi:10.1002/mp.15063.
Sanaat, A., Shiri, I., Arabi, H., Mainta, I., Nkoulou, R., Zaidi, H., 2021. Deep learning-assisted ultra-fast/low-dose whole-body PET/CT imaging. Eur. J. Nucl. Med. Mol. Imaging 48, 2405–2415.
Sanaat, A., Zaidi, H., 2020. Depth of interaction estimation in a preclinical PET scanner equipped with monolithic crystals coupled to SiPMs using a deep neural network. Appl. Sci. 10, 4753.
Schaefferkoetter, J.D., Yan, J., Soderlund, T.A., Townsend, D.W., Conti, M., Tam, J.K., Soo, R., Tham, I., 2017. Quantitative accuracy and lesion detectability of low-dose FDG-PET for lung cancer screening. J. Nucl. Med. 58, 399–405.
Surti, S., 2015. Update on time-of-flight PET imaging. J. Nucl. Med. 56, 98–105.
Wang, Y., Yu, B., Wang, L., Zu, C., Lalush, D.S., Lin, W., Wu, X., Zhou, J., Shen, D., Zhou, L., 2018. 3D conditional generative adversarial networks for high-quality PET image estimation at low dose. Neuroimage 174, 550–562.
Wei, W., Poirion, E., Bodini, B., Tonietto, M., Durrleman, S., Colliot, O., Stankoff, B., Ayache, N., 2020. Predicting PET-derived myelin content from multisequence MRI for individual longitudinal analysis in multiple sclerosis. Neuroimage 223, 117308.
Xiang, L., Qiao, Y., Nie, D., An, L., Wang, Q., Shen, D., 2017. Deep auto-context convolutional neural networks for standard-dose PET image estimation from low-dose PET/MRI. Neurocomputing 267, 406–416.
Xie, S., Girshick, R., Dollár, P., Tu, Z., He, K., 2017. Aggregated residual transformations for deep neural networks. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 1492–1500.
Xu, J., Gong, E., Pauly, J., Zaharchuk, G., 2017. 200x low-dose PET reconstruction using deep learning. arXiv preprint arXiv:1712.04119.
Yang, J., Park, D., Gullberg, G.T., Seo, Y., 2019. Joint correction of attenuation and scatter in image space using deep convolutional neural networks for dedicated brain (18)F-FDG PET. Phys. Med. Biol. 64, 075019.
Zaidi, H., El Naqa, I., 2021. Quantitative molecular Positron Emission Tomography imaging using advanced deep learning techniques. Annu. Rev. Biomed. Eng. 23, 249–276.
Zaidi, H., Montandon, M.-L., Assal, F., 2010. Structure-function based quantitative brain image analysis. PET Clin. 5, 155–168.