An Encoder-Decoder Network For Direct Image Reconstruction On Sinograms of A Long Axial Field of View PET

Ruiyao Ma
Tsinghua University
Jiaxi Hu
University of Bern
Hasan Sari
University of Bern
Song Xu
University of Bern
Clemens Mingels
University of Bern
Marco Viscione
University of Bern
Venkata Sai Sundar Kandarpa
Laboratory of Medical Image Processing (LaTIM), INSERM-UMR 1101
Weibo Li
Helmholtz Zentrum München German Research Center for Environmental Health (GmbH)
Dimitris Visvikis
Laboratory of Medical Image Processing (LaTIM), INSERM-UMR 1101
Rui Qiu (qiurui@tsinghua.edu.cn)
Tsinghua University
Axel Rominger
University of Bern
Junli Li
Tsinghua University
Kuangyu Shi
University of Bern

Research Article

Keywords: Deep Learning, Image Reconstruction, Long Axial Field of View PET

Posted Date: November 29th, 2021

DOI: https://doi.org/10.21203/rs.3.rs-1084555/v1

An Encoder-decoder Network for Direct Image Reconstruction on Sinograms of a Long Axial Field of View PET

（**AI for Direct Image Reconstruction**）

Ruiyao Ma[1, 2, 3], Jiaxi Hu[2], Hasan Sari[2,4], Song Xue[2], Clemens Mingels [2], Marco Viscione[2], Venkata Sai Sundar Kandarpa [5], Weibo Li[3], Dimitris Visvikis [5], Rui Qiu[1*], Axel Rominger[2], Junli Li[1*], Kuangyu Shi[2]

[1] Tsinghua University, Department of Engineering Physics, Key Laboratory of Particle & Radiation Imaging (Tsinghua University), Ministry of Education， Beijing, China

[2] University of Bern, Department of Nuclear Medicine, Bern, Switzerland

[3] Helmholtz Zentrum München German Research Center for Environmental Health (GmbH), Neuherberg, Bavaria, Germany

[4] Advanced Clinical Imaging Technology, Siemens Healthcare AG, Lausanne, Switzerland

[5] Laboratory of Medical Image Processing (LaTIM), INSERM-UMR 1101, Brest, France

Corresponding authors

Prof. Rui Qiu

Building Liuqing Room 404，Tsinghua University, Beijing， P.R China 100084

Fax: 86-010-62797089

Email: qiurui@tsinghua.edu.cn ORCID: 0000-0002-3511-6164

Prof. Junli Li

Building Liuqing Room 406，Tsinghua University, Beijing， P.R China 100084

Fax: 86-010-62797089

Email: lijunli@mail.tsinghua.edu.cn

Abstract

Purpose:

Deep learning is an emerging reconstruction method for positron emission tomography (PET) that can tackle complex PET corrections in an integrated procedure. This study optimized the direct PET reconstruction from sinogram on a long axial field of view (LAFOV) PET.

Methods:

This study developed a new deep learning architecture to reduce the biases during direct reconstruction from sinograms to images. This architecture is based on an encoder-decoder network and perceptual loss is adopted with pre-trained convolutional layers. It is trained and tested on data of 80 patients acquired from recent Siemens Biograph Vision Quadra PET/CT. The patients were randomly split into a training dataset of 60 patients, the validation dataset of 10 patients, and the test data set of 10 patients. The 3D sinograms were converted into 2D sinogram slices and were used as input to the network, and the vendor reconstructed images were considered as ground truths. The proposed method was compared with DeepPET, a benchmark deep learning method for PET reconstruction.

Results:

Compared to the DeepPET, the proposed network significantly reduced the root-mean-squared error (rRMSE) from 0.63 to 0.6 ($p<0.01$), and the structural similarity index (SSIM) and peak signal-to-noise ratio (PSNR) were improved from 0.93 to 0.95($p<0.01$) and from 82.02 to 82.36($p<0.01$), respectively. The reconstruction time was approximately 10s per patient, which was shortened by 36 times compared with the reconstruction using the conventional method. The errors of average standardized uptake values (SUV) for lesions between ground truth and the predicted result was reduced from 33.5% to 18.7% (P=0.03), and the error of max SUV was reduced from 32.7% to 21.8% (P=0.02).

Conclusion:

The results demonstrated the feasibility of using deep learning to reconstruct images with acceptable image quality and short reconstruction time. It showed that the proposed method can improve the quality of deep learning-based reconstructed images without additional CT images for attenuation and scattering corrections. This learning-based approach may provide the potential to accommodate more complex corrections for LAFOV PET.

Key words: Deep Learning, Image Reconstruction, Long Axial Field of View PET

I Introduction

With the development of PET instrumentation, the axial field of view continually increases leading to the new area of long axial field of view (LAFOV) PET or total-body PET. Compared with a current clinically standard of care axial field of view (FOV) PET system (the typical range of 26 cm), the long-axial FOV - PET systems have the larger solid angle coverage and longer axial FOV. Furthermore, a large anatomical region can be covered with one single bed position. As a result, the total acquisition time can be substantially reduced by a large factor due to the increased sensitivity[1-3].

Many image reconstruction algorithms have been previously studied to reconstruct tomographic images using projection data. Conventional methods can solve a mapping function from measurement space to image space based on physical principles. Clinically established methods include analytical and iterative. Analytical methods, such as filtered back projection (FBP)[4], can achieve fast image reconstruction, but image quality is characterized by a high level of noise. The iterative method, for example, the maximum-likelihood expectation-maximization (MLEM)[5], ordered-subset expectation maximization (OSEM)[6] with iteratively back- and forward-projecting, which are clinically accepted standards of care, can obtain images with the lower noise level and good contrast recovery, but the iterative process is time-consuming. In addition, to perform attenuation correction, typically a CT or MRI image is used to estimate attenuation. In recent years, neural networks have been applied to tomographic image reconstruction to achieve higher quality reconstruction results with sparse information and short reconstruction time[7, 8]. The deep learning method has previously been applied. Examples include：(1) noise reduction to enable low-dose PET imaging protocols [9-11]. (2) A neural network was integrated into the iterative process to speed up the convergence speed and improve the reconstruction quality [12, 13]. (3) Train a neural network to directly convert from projection data to image data. For the direct method, an automated method has been developed that utilizes transform by manifold approximation (AUTOMAP) to learn the relationship between sensor domain and image domain[14] While the paper mainly focused on MRI image reconstruction, a PET system application was anticipated in the article [14]. A deep encoder-decoder network named DeepPET was used in direct reconstruction for PET images. PET images and projection data simulated based on XCAT digital phantom were used to train the network [15]. Kandarpa *et al.* proposed a double U-Net to learn the sinogram-to-image transformation, and the deep-learning pipeline consisted of denoising, image reconstruction, and super-resolution segments[16]. William *et al.* proposed a DirectPET network to achieve full-size neural network PET reconstruction from histoimages data, in which the XCT-based attenuation maps were used as additional input for corrections [17]. And the reconstruction from histo-images using a U-net network was also explored by the same team, also CT-based attenuation was required[18].

In this work, we explored the application of the encoder-decoder network to long-axial FOV PET reconstruction using clinical patient data. The work focuses on achieving end-to-end PET reconstruction directly from the detector domain to the images domain; attenuation correction is integrated into the training process.

II Material and Methods

Patients & imaging

Clinical patient list mode data were collected using Biograph Vision Quadra (Siemens Healthineers) at the University of Bern, Switzerland. This system is with a FOV of 106 cm. Preliminary assessments of this scanner's characteristics reveal a sensitivity of 174 cps/kBq and a time of flight (TOF) resolution of 219 ps in ultra-high sensitivity mode[19].

The selected patients were injected with ^{18}F- FDG and performed PET/CT examination, including 80 cases (median age, 63 years; age range, 50-76 years; 44 females; BMI, 73.9±17 kg/m^2) of patients In all cases, subjects fasted for more than 4 h and had a blood glucose of no more than 200 mg/dl, all patients were injected with ^{18}F-FDG with an uptake time of 90 min ± 10%, and the patients without complete PET/CT scan images from above head to below thigh, or poor image quality because of patients' movement were been excluded. The 80 patients were randomly split into a training dataset of 60 patients, the validation dataset of 10 patients, and the test data set of 10 patients (median age, 67 years; age range, 40-81 years; 5 females; BMI, 23.85±2.38 kg/m^2). The study was conducted in accordance with the requirements of the respective local ethics committees in Switzerland (Req-2021-00517).

Data pre-processing

List-mode data obtained from the scanner was reconstructed with industry software (e7-tools, Siemens Healthineers) with CT-based attenuation correction. The used reconstruction method was PSF-TOF with 4 iterations and 5 subsets. 644 slides for each patient were reconstructed and used as ground truth in this work. The image size was 440×440 with voxel dimensions of 1.65×1.65×1.65 mm^3. Furthermore, the list-mode data was transferred and uncompressed into 3-dimensional (3D) sinogram data. The 3D sinogram was converted into 2-dimensional (2D) slices by single-slice rebinning (SSRB)[4]. 644 2D sinogram slices were obtained for each patient, corresponding to 644 reconstructed images.

The input dataset of the network was 2D sinogram slices, and the reconstructed images from e7-

tools were used as training targets. Several images of the starting and ending positions of each patient's data had a low count and were therefore excluded. Each patient retained 599 sets of data (2D sinogram and reconstructed image). 60 patients were designated for training, 10 patients for validation, and 10 for the testing.

Deep Neural Network structure

We developed an encoder-decoder network for direct image reconstruction. It is comprised of two parts, image transform network and perceptual loss[20] network (Fig.1).

The structure of our training network was inspired by DeepPET [15]. The network is consisting of the encoder part, transformation part, and decoder part (Fig.1,Supplemental Fig.1,2)[21]. 31 convolution blocks and one single convolution layer are involved. Each convolution block includes a convolution layer to extract features, a batch normalization (BN) layer to speed up the training and convergence of the network, and a rectified linear unit (ReLU) activation function. In the decoder and transformation part, the convolution filter sizes are selected as 7×7 for the first two blocks，5×5 for followed two blocks, and 3×3 for others. The numbers of extracted features increase from 32 to 1024. The widths and lengths of the feature maps are decreased by the convolution layer with a kernel stride of 2. In the decoder part, the convolution filter sizes are 3×3 and the feature maps are enlarged by upsampling layers. The output layer is a convolution layer with one feature. The 2D sinogram slices are resized as 288×269 as inputs of the network. The outputs of the image transform network are reconstructed images, which are put in the perceptual loss network.

The perceptual loss network adopts the first 3 convolution block of VGG19[22] (Fig.7). The VGG network uses the accumulation of multiple small-scale convolution kernels (3×3) instead of large-scale convolution kernels. This establishment can form multiple non-linear layers to increase the depth of the network and achieve complex feature learning. The convolution blocks in VGG19 are consist of a convolution layer followed by ReLU. The sizes of feature maps are reduced by pooling layers. VGG19 network weights pre-trained on the ImageNet database(image-net.org) are used in this work. The outputs of the first 3 pooling layers are extracted to be used as feature reconstruction loss as expressed below:

$$L_{VGG} = \frac{1}{3}\sum_{i=1}^{3}|VGG(x)_i - VGG(y)_i| \quad (1)$$

where $VGG(x)_i$ and $VGG(y)_i$ represent the output of i-th pooling layer in VGG19 with the input of ground truth and predict image of image transform network.

The other two parts are also involved in the loss function, mean square error (MSE) loss and structural similarity (SSIM) [23]loss. The functions are shown as follows:

$$L_{MSE} = \frac{1}{n}\sum_{i=1}^{n}(x_i - y_i)^2 \quad (2)$$

where *x* is the ground truth, *y* is the predicted image of the image transform network, and n is the total number of image pixels.

$$L_{SSIM} = 1 - \frac{(2u_x u_y + C_1)(2\sigma_{xy} + C_2)}{(u_x^2 + u_y^2 + C_1)(\sigma_x^2 + \sigma_y^2 + C_2)} \quad (3)$$

$$C_1 = (0.01 \cdot max(x))^{\ 2} \quad (4)$$

$$C_2 = (0.03 \cdot max(x))^{\ 2} \quad (5)$$

where u_x and σ_x^2 are mean and variance of ground truth image pixels, u_y and σ_y^2 are mean and variance of ground truth image pixels, and σ_{xy} is the covariance of ground truth and predict image. The $max(x)$ is the maximum of ground truth image value.

The total loss function consists of the above three parts.

$$loss = L_{MSE} + L_{SSIM} + L_{VGG}$$

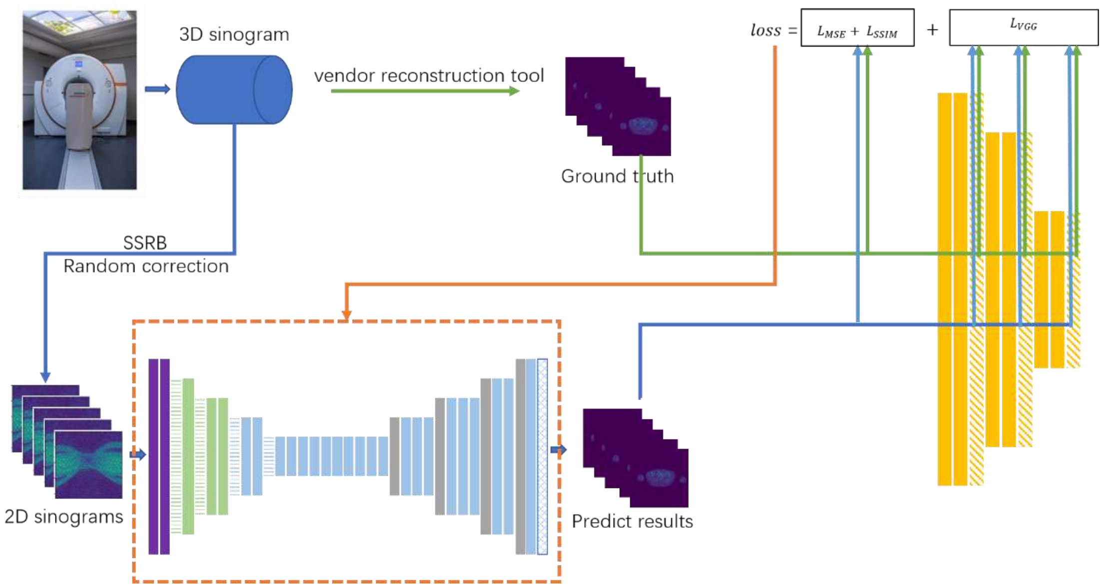

FIGURE.1 The network adopted in this work includes two parts, image transform network, and perceptual loss network.(The detailed structures of image transform network and the perceptual loss network are shown in supplemental FIGURE. 1,2)

Network training and test procedure

The network was implemented in TensorFlow(www.tensorflow.org)[24], the training of the network was performed using a GPU (Tesla V100-PCIe-16GB, NVIDIA) and tested using GeForce RTX 2080 Ti (NVIDIA). Adam optimization method [25] was used as optimized with a learning rate of 10^{-4}. All images in the training dataset (35940 2D sinogram and reconstructed images) were used as input to the network by the batch size of 50 and trained with 300 epochs. The trained network was tested on 10 patients' data (5590 2D sinograms and reconstructed images).

Image quality evaluation

The image quality evaluation was processed with structural similarity index (SSIM), root-mean-squared error (rRMSE), and peak signal-to-noise ratio (PSNR)[23] for the regions of the body. SSIM is an index used to measure the similarity of two images. The mean is used as an estimate of brightness, the standard deviation is used as an estimate of contrast, and covariance is used as a measure of structural similarity. The value range is [0,1], the closer to 1 means the output image is more similar to the target image. The SSIMs were calculated as shown below, and the meaning of the character is consistent with Equation (3-5).

$$SSIM = \frac{(2u_x u_y + C_1)(2\sigma_{xy} + C_2)}{(u_x^2 + u_y^2 + C_1)(\sigma_x^2 + \sigma_y^2 + C_2)} \quad (6)$$

Root-mean-squared error (rRMSE) was calculated based on mean square error (MSE), that is,

$$rRMSE = \frac{\sqrt{MSE}}{\bar{x}} \quad (7)$$

$$MSE = \sum_{i=1}^{n} (x_i - y_i)^2 \quad (8)$$

where the $\bar{x}$ is the average value of all pixels in the ground truth image, x is the ground truth, y is the predicted image of the network.

The SNR was defined as follows:

$$PSNR = 20 \cdot log_{10}\left(\frac{MAX_I}{\sqrt{MSE}}\right) \quad (9)$$

where the MAX_I is the maximum value of the reconstructed images.

Clinical evaluation

The results on the test data set were further evaluated by 2 nuclear medicine physicians. For each patient, a typical lesion was selected and delineated manually. Among the 10 patients, 1 patient was proved to have no lesion. The average standardized uptake values (SUVmean) and max standardized uptake values (SUVmax) were measured of the tracer uptake in a region of interest of selected lesions. The relative errors between ground truth and DeepPET results and this work were calculated and compared.

III Results

The average time cost for the prediction of 1 patient (644 images) is 9.85±0.27s, including 6.41±0.11s for the data preparation process (SSRB) on and 3.44±0.17s for network prediction on NVIDIA GeForce RTX 2080 Ti. The predicted time for a single image was about 5 ms. And the standard industry software (e7-tools, Siemens Healthineers) reconstructed 1 patient's data using about 360s on average. The comparison of the time cost by the two methods is shown in the figure below.

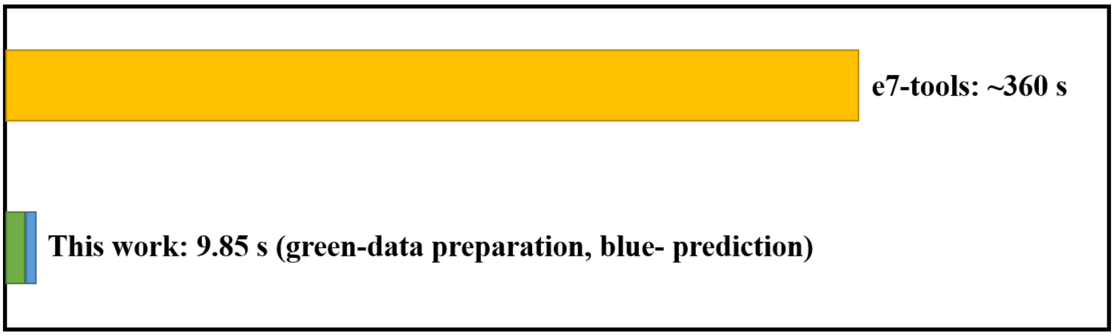

FIGURE.2 Time cost comparison. The data preparation process of this work for 1 patient costs about 6.41 s and the prediction costs about 3.44 s. The e7-tools use about 360 s to reconstruct 1 patient's data.

The predicted images of the network in this work and DeepPET are shown in Fig.3, along with the ground truth and input sinogram. Fig.4 shows the image quality evaluation results of this work for ten test patients, including MSE, rRMSE, PSNR, and SSIM. The same patients' data were also tested by DeepPET. The average results of ten patients are shown in Table 1.

In Fig. 3 we show the input images and reconstruction results from different areas of the body. It can be seen from the images that there is a strong similarity between the results of this work and the ground truth. Especially in regions where tracer uptake is high, such as the head, chest, and heart, the results of this work are in good agreement with the reference values. In the pelvic cavity, legs, and other areas, there are point-like high uptake positions, and the network of this work can also restore them well. At the same time, we have also observed that there are some slight structural differences at the edge regions of some structures, such as the details of the brain, the edge of the heart, which are manifested in the blur of the edges. This is mainly because the true value image is directly reconstructed from the 3D sinogram, and the input used for prediction in this work is 2D sinograms. When the 3D sinogram is converted to a 2D sinogram, a certain amount of information is lost and errors are introduced. Compared with the test results of the original DeepPET structure network, the image structure restoration and detail restoration have been improved, indicating the effectiveness of the introduction of the perceptual loss network.

From the quantitative results statistics in Fig. 4 and Table 1, the SSIM of the original DeepPET structure prediction result compared with the true value is 0.95 ± 0.02, and the network has a 2% improvement in SSIM after the perceptual loss structure is added in this work, which is close to 1. At the same time, from the statistical results of PSNR, the work improves the signal-to-noise ratio from 82.02 ± 0.90 to 82.36 ± 0.87, which is a slight improvement. In addition, MSE decreased from 161.67 ± 84.83 to 132.59 ± 75.01, and rRMSE decreased from 0.63 ± 0.06 to 0.60 ± 0.06, indicating that the reconstructed picture is closer to the true value. The quantitative results of the two networks were statically analyzed with paired *t*-test， all showed significant improvement.

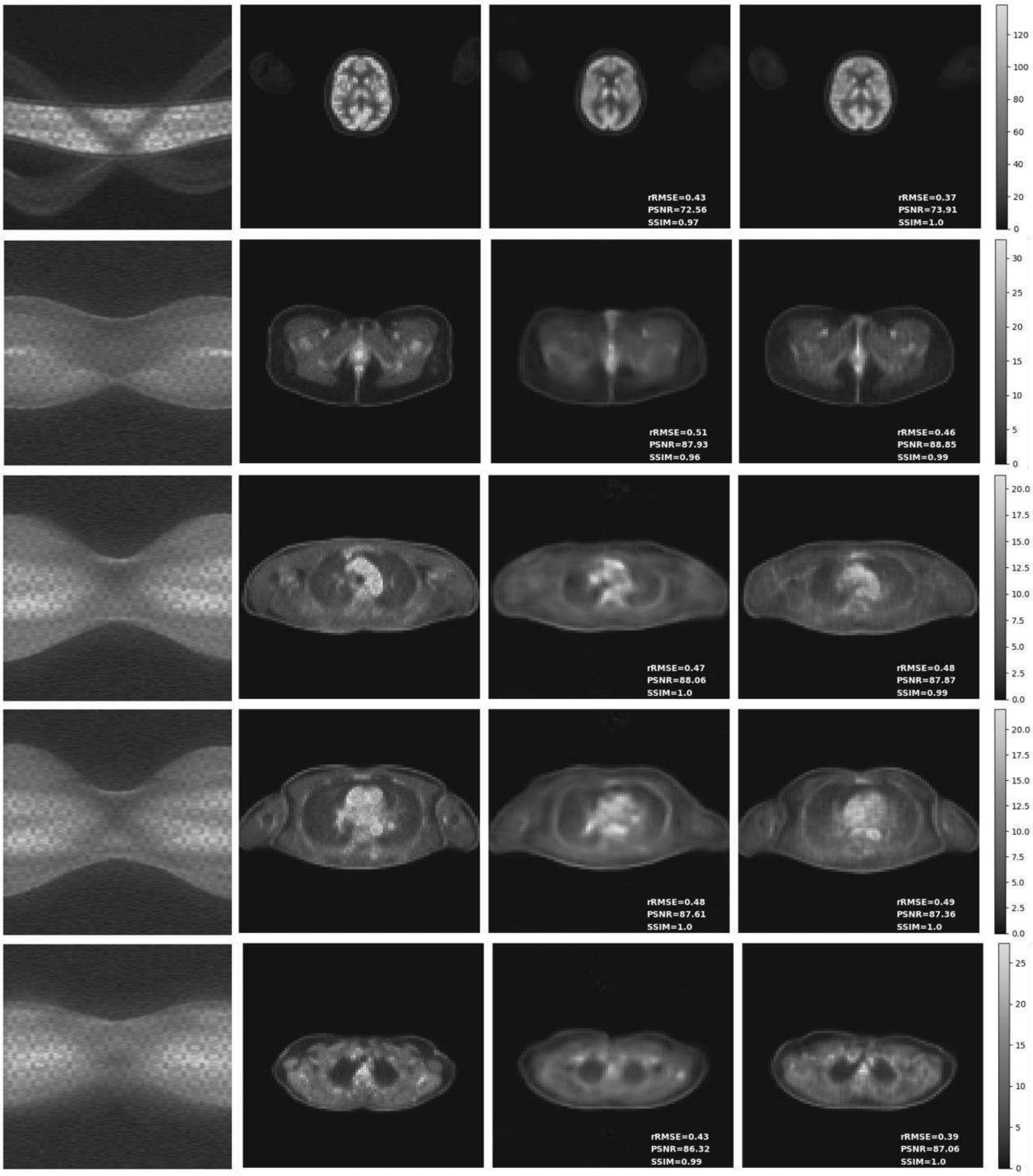

FIGURE.3 test set reconstructions using both DeepPET, as well as the predicted results of this work. Left to right: PET sinogram, ground truth, DeepPET, and results of this work. Images are labeled with SSIM, rRMSE, and PSNR relative ground truth.

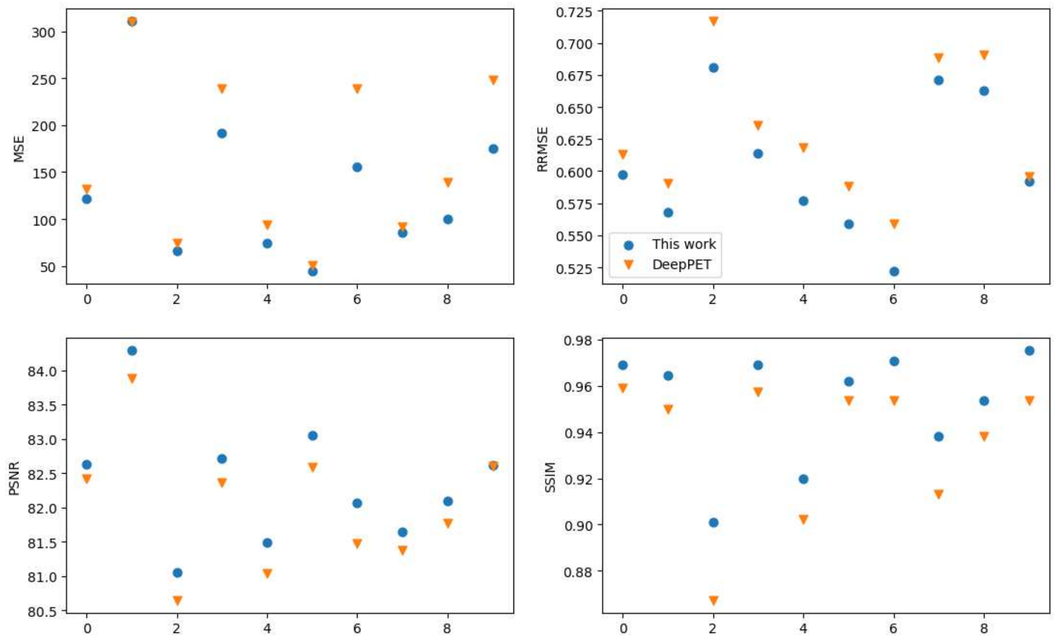

FIGURE.4 The image quality evaluation results of this work for ten test patients: including mean square error (MSE), root-mean-squared error (rRMSE), peak signal-to-noise ratio (PSNR), and structural similarity index (SSIM).

The average standardized uptake values (SUV) and max standardized uptake values (SUV) were measured of the tracer uptake in a region of interest of lesions(Fig. 5) for test sets (median age, 67 years; age range, 40-81 years; 5 females; BMI, 23.85±2.38 kg/m2). The relative errors between ground truth and DeepPET results and this work were calculated and shown in Table 2. From the comparison for smaller lesions, for example, lesions 1 and 6, the reconstruction results of this work are closer to the ground truth value. For larger lesions, such as lesions 3 and 7, the recovery of this work and DeepPET on SUV (max) is slightly worse, but comparing the shape and contour of the lesion in the figure, the results of this work are more similar to the ground truth. For some cases, the two reconstruction results are not enough to the details, and lesion 8 is not significantly separated. In addition, a good sign is this work shows superior performance in anatomical structure with non-intensive uptaken, e.g. in the same layer of lesion 7, this work shows better in displaying the nonuptaken area in the liver, which the DeepPET is not clearly shown. In general, compared with the DeepPET, the SUV(average) and SUV(max) of lesions of this work are closer to the ground truths. This indicates that the prediction results of this work can provide a better clinical reference at the level of the lesion. The currently trained network and DeepPET both have the possible degradation of the small lesions, for example, the lesion 2.

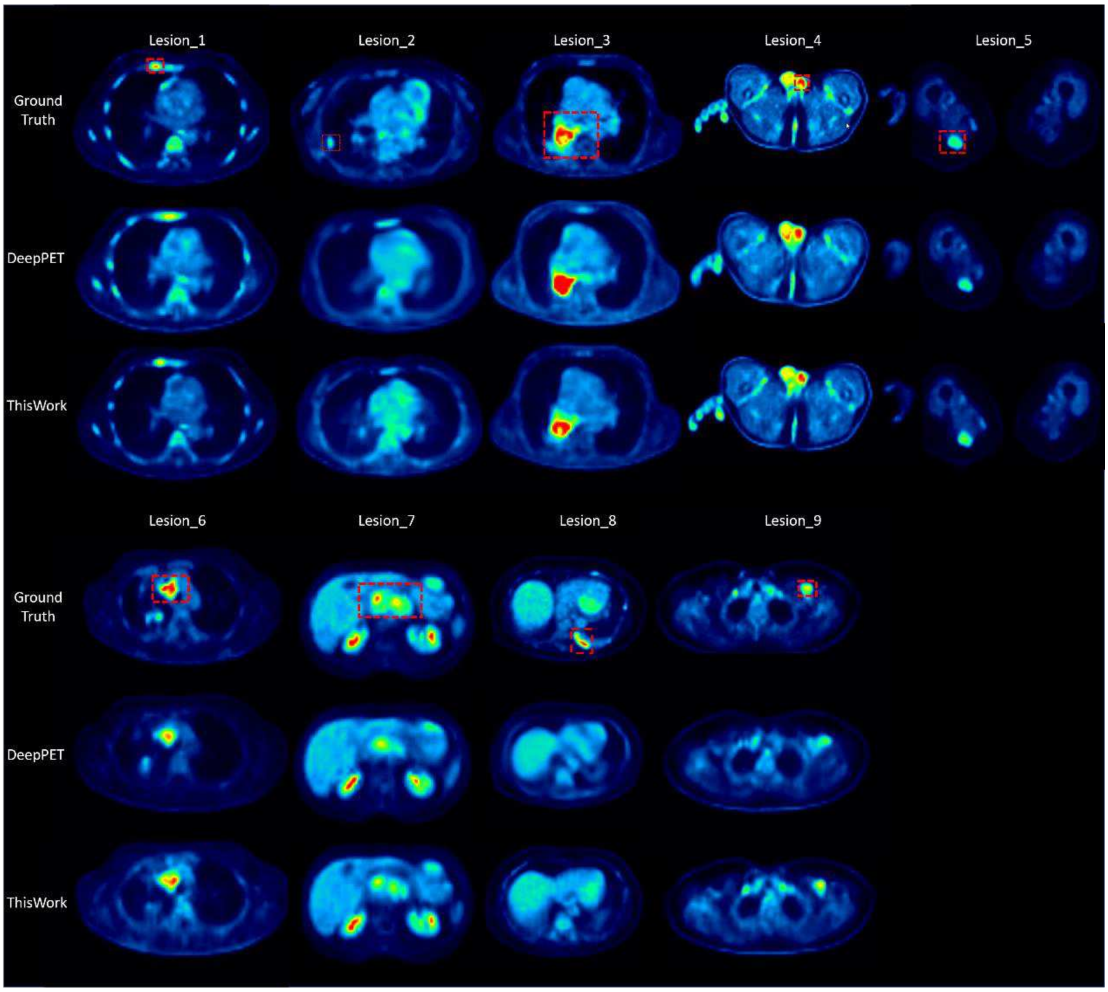

FIGURE.5 The average standardized uptake values (SUV) and max standardized uptake values (SUV) were measured of the tracer uptake in a region of interest of lesions

IV Discussion

This study followed the mainstream of AI development for PET reconstruction and focused on the direct reconstruction from Sinogram. In contrast to most previous research on Sinogram data from phantom-based simulation [15] or anthropomorphic simulation by projecting real patient data [26], the training and test in this study were directly on real PET measurements. A critical concern for AI development is its reproducibility and extensibility to complexity in real applications [27]. Compared to the development on simulated sinogram data, the development on real measurement data in this study can better tackle the challenges of physical and physiological complexity and enhance the translational potential of data-driven methods.

An advanced LAFOV PET scanner was employed in the development and test of the AI-based direct reconstruction from Sinogram data in this study. Although the Sinogram of LAFOV PET did not consider all the rich information in the acquisition of this powerful scanner leading to downgraded measurement data, the preliminary results confirmed that the deep neural network can reconstruct PET images with corrections of attenuation and scattering directly from Sinogram without the input of CT. This potential of AI in complex reconstruction with various corrections may bring benefits for the reconstruction of LAFOV PET considering the increased complexity in the reconstruction of LAFOV PET [28]. The LAFOV of total-body PET increases the probability of the detection of LORs to increase the sensitivity. However, the high obliqueness of the LORs between distant rings suffers from the parallax error [1], and introduced large heterogeneity in the image quality [29, 30]. The increased Compton scattering and the ratio between multiple over single scattered photons is another critical bottleneck for the reconstruction of LAFOV PET [28]. The fraction of multiple scatters changes heterogeneously in LAFOV PET [31]. The fractions of random events are also dependent on the difference of rings in LAFOV PET [30]. The correction of heterogeneity of random and multiple scattered events makes the reconstruction more difficult than in conventional scanners.

Although the current study did not consider all the challenging issues such as larger and heterogenous solid angle in LAFOV PET reconstruction, the advantage of AI methods may dealing with complexity and heterogeneity encourage the development of such technology. At this stage, the AI-based reconstruction may be less advanced and accurate compared with physics-based reconstruction. Further improvement of the input Sinogram and the training data with more accurate corrections may enhance the performance of this data-driven approach in LAFOV PET reconstruction and eventually may reach or even outperform the physics-based reconstruction.

Due to the huge number of LORs received in LAFOV PET, the storage and processing of the acquisition data is daunting [28, 31]. For example, the 2-m LAFOV system has roughly 10 times of data to process compared to a 20-cm PET system. But with more oblique LORs used, there could be a 40-fold increase [31]. Remarkably, the prompts count rate peaked at 10 million events, a few orders of magnitude larger than for a traditional PET scanner [28]. Conventional PET reconstruction algorithms are inefficient in the processing of the vast data of LAFOV PET reconstruction. Although our test was performed on Sinogram data for reconstruction, the results confirmed that deep learning can significantly shorten (up to 36 times) the reconstruction time for whole-body imaging compared with a conventional iterative algorithm. The potential in accelerating the computational speed may bring advantages for the practice of LAFOV PET in the clinical routine.

The preliminary results demonstrated that the improvement of deep learning architecture can improve the performance of AI-based reconstruction. In response to the challenge of real data training, this work employed the perceptual loss network structure to optimize the neural network. The pre-trained VGG network was used to extract the feature map from predicted images and the ground truth. The perceptual loss was added into the loss function calculation, which can improve the training efficiency and effect of the network. The comparison of the prediction results shows that the similarity of the reconstructed image structure and the signal-to-noise ratio is improved. This is because the perceptual loss function can calculate the distance between the predicted image and the target image from the feature level, but not from pixel level, so the structure of the image can be reconstructed better in a larger area[20].

Based on lesion demarcation, overall image quality, and visually assessed signal-to-noise ratio, this work displays better quality than the traditional DeepPET approach. Also, the semiquantitative measurement method was used and the result is shown in table 2. In this study, we estimated a series of lesions located in different organs, including the rib, muscle, mediastinum, retroperitoneal space soft tissue, which is shown in Fig. 5. For illustration, lesion 1 and 2 both presents a lesion located in the rib, the image of this work shows a better-outlined shape than the traditional DeepPET, which would easily misdiagnose to be located in the sternum. Considering the purpose of optimizing reconstruction, the outputs show satisfying performance in presenting the morphological character of the primary lesion with elevator uptake values. The lack of structure details leads to misdiagnosis in location and concealing some small lesions in worse cases. This may be due to the limited training cases. The use of more varied and larger capacity training sets can make contributions to improving the accuracy of predictions. Another limitation of this work was the loss in the conversion of 3D sinograms to 2D sinograms. Due to the limited GPU memory, we transferred the vendor output of the 3D sinograms to the 2D sinograms as the input of this network. In the conversion process, certain noises were introduced, and part of the spatial information in the axial direction was lost. This resulted in the inability to fully recover all the details of the ground truth images. The development on 3D sinogram data with more data and computational resources may improve the performance.

V Conclusion

This study developed a network structure combing encoding-decoding and perceptual loss structure to improve the direct PET image reconstruction from projection data. For the first time, such kind of AI-based reconstruction methods were tested on real clinical data on a LAFOV PET. Despite the limitations of AI-based methods, the end-to-end reconstruction process from the sinogram data demonstrates the potential of deep learning to learn complex reconstruction principles such as projection, normalization, attenuation correction and scattering correction. Further optimization and development of AI-based reconstruction may provide an effective solution for complex PET reconstruction such as LAFOV PET.

TABLE. 1 Quality evaluation results of the test database, MSE, rRMSE, PSNR, and SSIM. The quantitative results of the two networks were statically analyzed with paired t-test

	This work	DeepPET	Paired t-test
MSE	132.6±75.0	161.7±84.8	-3.1 ($p<0.01$)
rRMSE	0.60±0.04	0.63±0.04	-6.9 ($p<0.01$)
PSNR	82.4±0.87	82.1±0.90	7.1 ($p<0.01$)
SSIM	0.95±0.02	0.93±0.03	7.3 ($p<0.01$)

TABLE. 2 The average standardized uptake values (SUV) errors and max standardized uptake values (SUV) errors between ground truth and DeepPET results and this work for lesions from test cases.

Lesions	SUV(Ave) Relative Errors	SUV(Max) Relative Errors

	DeepPET	This Work	DeepPET	This Work
Lesion_1	53.0%	2.9%	41.8%	6.4%
Lesion_2	81.3%	77.0%	83.3%	80.3%
Lesion_3	5.4%	2.1%	8.6%	12.1%
Lesion_4	9.1%	8.7%	7.2%	0.9%
Lesion_5	11.7%	2.6%	11.8%	4.5%
Lesion_6	8.7%	2.0%	14.0%	5.0%
Lesion_7	27.6%	16.8%	35.1%	29.7%
Lesion_8	67.7%	43.3%	72.2%	48.1%
Lesion_9	37.1%	13.0%	20.1%	9.4%
Average	33.5%	18.7%	32.7%	21.8%
Paired t-test	P=0.03		P=0.02	

Acknowledgments

This project received funding from the National Natural Science Foundation of China [Grant No. 11875036] and Swiss National Science Foundation [SNSF Grant No. 188350 & 188914].

REFERENCES

1. Vandenberghe S, Moskal P, Karp JS. State of the art in total body PET. EJNMMI Phys. 2020;7:35. doi:10.1186/s40658-020-00290-2.
2. Cherry SR, Jones T, Karp JS, Qi J, Moses WW, Badawi RD. Total-Body PET: Maximizing Sensitivity to Create New Opportunities for Clinical Research and Patient Care. Journal of nuclear medicine : official publication, Society of Nuclear Medicine. 2018;59:3-12. doi:10.2967/jnumed.116.184028.
3. Cherry SR, Badawi RD, Karp JS, Moses WW, Price P, Jones T. Total-body imaging: Transforming the role of positron emission tomography. Science Translational Medicine. 2017;9:eaaf6169.

doi:10.1126/scitranslmed.aaf6169.
4. Defrise M, Kinahan PE, Michel CJ. Image Reconstruction Algorithms in PET. In: Bailey DL, Townsend DW, Valk PE, Maisey MN, editors. Positron Emission Tomography: Basic Sciences. London: Springer London; 2005. p. 63-91.
5. Shepp LA, Vardi Y. Maximum Likelihood Reconstruction for Emission Tomography. IEEE Transactions on Medical Imaging. 1982;1:113-22. doi:10.1109/TMI.1982.4307558.
6. Hudson HM, Larkin RS. Accelerated image reconstruction using ordered subsets of projection data. IEEE Trans Med Imaging. 1994;13:601-9. doi:10.1109/42.363108.
7. Uribe CF, Mathotaarachchi S, Gaudet V, Smith KC, Rosa-Neto P, Benard F, et al. Machine Learning in Nuclear Medicine: Part 1-Introduction. J Nucl Med. 2019;60:451-8. doi:10.2967/jnumed.118.223495.
8. Reader AJ, Corda G, Mehranian A, da Costa-Luis C, Ellis S, Schnabel JA. Deep Learning for PET Image Reconstruction. IEEE Transactions on Radiation and Plasma Medical Sciences. 2020:1-. doi:10.1109/trpms.2020.3014786.
9. Gong K, Guan J, Liu CC, Qi J. PET Image Denoising Using a Deep Neural Network Through Fine Tuning. IEEE Transactions on Radiation and Plasma Medical Sciences. 2019;3:153-61. doi:10.1109/TRPMS.2018.2877644.
10. Cui J, Gong K, Guo N, Wu C, Meng X, Kim K, et al. PET image denoising using unsupervised deep learning. European journal of nuclear medicine and molecular imaging. 2019;46:2780-9.
11. Katsari K, Penna D, Arena V, Polverari G, Ianniello A, Italiano D, et al. Artificial intelligence for reduced dose 18F-FDG PET examinations: a real-world deployment through a standardized framework and business case assessment. EJNMMI Physics. 2021;8:25. doi:10.1186/s40658-021-00374-7.
12. Gong K, Guan J, Kim K, Zhang X, Fakhri G, Qi J, et al. Iterative PET Image Reconstruction Using Convolutional Neural Network Representation. IEEE Transactions on Medical Imaging. 2017;38. doi:10.1109/TMI.2018.2869871.
13. Gong K, Wu D, Kim K, Yang J, Sun T, El Fakhri G, et al. MAPEM-Net: an unrolled neural network for fully 3D PET image reconstruction. 15th International Meeting on Fully Three-Dimensional Image Reconstruction in Radiology and Nuclear Medicine: International Society for Optics and Photonics; 2019. p. 110720O.
14. Zhu B, Liu JZ, Cauley SF, Rosen BR, Rosen MS. Image reconstruction by domain-transform manifold learning. Nature. 2018;555:487-92. doi:10.1038/nature25988.
15. Haggstrom I, Schmidtlein CR, Campanella G, Fuchs TJ. DeepPET: A deep encoder-decoder network for directly solving the PET image reconstruction inverse problem. Med Image Anal. 2019;54:253-62. doi:10.1016/j.media.2019.03.013.
16. Kandarpa VSS, Bousse A, Benoit D, Visvikis D. DUG-RECON: A Framework for Direct Image Reconstruction Using Convolutional Generative Networks. IEEE Transactions on Radiation and Plasma Medical Sciences. 2021;5:44-53. doi:10.1109/trpms.2020.3033172.
17. Whiteley W, Luk WK, Gregor J. DirectPET: full-size neural network PET reconstruction from sinogram data. J Med Imaging (Bellingham). 2020;7:032503. doi:10.1117/1.JMI.7.3.032503.
18. Whiteley W, Panin V, Zhou C, Cabello J, Bharkhada D, Gregor J. FastPET: Near real-time PET reconstruction from histo-images using a neural network. arXiv preprint arXiv:200204665. 2020.
19. Alberts I, Hünermund JN, Prenosil G, Mingels C, Bohn KP, Viscione M, et al. Clinical performance of long axial field of view PET/CT: a head-to-head intra-individual comparison of the Biograph Vision Quadra with the Biograph Vision PET/CT. Eur J Nucl Med Mol Imaging. 2021;48:2395-404. doi:10.1007/s00259-021-05282-7.

20. Johnson J, Alahi A, Fei-Fei L. Perceptual losses for real-time style transfer and super-resolution. European conference on computer vision: Springer; 2016. p. 694-711.
21. Shan H, Zhang Y, Yang Q, Kruger U, Kalra MK, Sun L, et al. 3-D convolutional encoder-decoder network for low-dose CT via transfer learning from a 2-D trained network. IEEE transactions on medical imaging. 2018;37:1522-34.
22. Simonyan K, Zisserman A. Very Deep Convolutional Networks for Large-Scale Image Recognition. Computer Science. 2014.
23. Wang Z, Bovik AC, Sheikh HR, Simoncelli EP. Image quality assessment: from error visibility to structural similarity. IEEE transactions on image processing. 2004;13:600-12.
24. Abadi M, Agarwal A, Barham P, Brevdo E, Chen Z, Citro C, et al. Tensorflow: Large-scale machine learning on heterogeneous distributed systems. arXiv preprint arXiv:160304467. 2016.
25. Kingma DP, Ba J. Adam: A method for stochastic optimization. arXiv preprint arXiv:14126980. 2014.
26. Hu Z, Xue H, Zhang Q, Gao J, Zhang N, Zou S, et al. DPIR-Net: Direct PET Image Reconstruction Based on the Wasserstein Generative Adversarial Network. IEEE Transactions on Radiation and Plasma Medical Sciences. 2021;5:35-43. doi:10.1109/trpms.2020.2995717.
27. Roberts M, Driggs D, Thorpe M, Gilbey J, Yeung M, Ursprung S, et al. Common pitfalls and recommendations for using machine learning to detect and prognosticate for COVID-19 using chest radiographs and CT scans. Nature Machine Intelligence. 2021;3:199-217.
28. Efthimiou N. New Challenges for PET Image Reconstruction for Total-Body Imaging. PET Clin. 2020;15:453-61. doi:10.1016/j.cpet.2020.06.002.
29. Schmall JP, Karp JS, Werner M, Surti S. Parallax error in long-axial field-of-view PET scanners-a simulation study. Physics in medicine and biology. 2016;61:5443-55. doi:10.1088/0031-9155/61/14/5443.
30. Zhang X, Badawi RD, Cherry SR, Qi J. Theoretical study of the benefit of long axial field-of-view PET on region of interest quantification. Physics in medicine and biology. 2018;63:135010. doi:10.1088/1361-6560/aac815.
31. Zhang X, Zhou J, Cherry SR, Badawi RD, Qi J. Quantitative image reconstruction for total-body PET imaging using the 2-meter long EXPLORER scanner. Physics in medicine and biology. 2017;62:2465-85. doi:10.1088/1361-6560/aa5e46.

Statements and Declarations

Funding This project received funding from the National Natural Science Foundation of China [Grant No. 11875036] and Swiss National Science Foundation [SNSF Grant No. 188350 & 188914].

Competing Interests Hasan Sari is a full-time employee of Siemens Healthcare AG, Switzerland. No other potential conflict of interest relevant to this article was reported.

Author Contributions All authors contributed to the study conception and design. Material preparation, data collection and analysis were performed by Ruiyao Ma, Jiaxi Hu, Hasan Sari, and Song Xue. The first draft of the manuscript was written by Ruiyao Ma, Jiaxi Hu, and Kuangyu Shi, and all authors commented on previous versions of the manuscript. All authors read and approved the final manuscript. Weibo Li, Rui Qiu, Axel Rominger, Junli Li, and Kuangyu Shi supported this work with funding acquisition.

Data Availability The datasets generated during and/or analysed during the current study are available from the corresponding author on reasonable request.

Ethics approval The study was conducted in accordance with the requirements of the respective local ethics committees in Switzerland (Req-2021-00517).

Consent to participate Informed consent was obtained from all individual participants included in the study.

Consent to publish The authors affirm that human research participants provided informed consent for publication of Figure 3,4,5.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- SUPPLEMENT.docx