

ORIGINAL RESEARCH

Open Access



Deep learning-assisted PET imaging achieves fast scan/low-dose examination

Yan Xing¹, Wenli Qiao¹, Taisong Wang¹, Ying Wang², Chenwei Li², Yang Lv², Chen Xi², Shu Liao³, Zheng Qian² and Jinhua Zhao^{1*} 

*Correspondence:

zhaojinhua1963@126.com

¹ Department of Nuclear Medicine, Shanghai General Hospital, Shanghai Jiaotong University, No. 100 Haining Road, Shanghai 200080, People's Republic of China
Full list of author information is available at the end of the article

Abstract

Purpose: This study aimed to investigate the impact of a deep learning (DL)-based denoising method on the image quality and lesion detectability of ¹⁸F-FDG positron emission tomography (PET) images.

Methods: Fifty-two oncological patients undergoing an ¹⁸F-FDG PET/CT imaging with an acquisition of 180 s per bed position were retrospectively included. The list-mode data were rebinned into four datasets: 100% (reference), 75%, 50%, and 33.3% of the total counts, and then reconstructed by OSEM algorithm and post-processed with the DL and Gaussian filter (GS). The image quality was assessed using a 5-point Likert scale, and FDG-avid lesions were counted to measure lesion detectability. Standardized uptake values (SUVs) in livers and lesions, liver signal-to-noise ratio (SNR) and target-to-background ratio (TBR) values were compared between the methods. Subgroup analyses compared TBRs after categorizing lesions based on parameters like lesion diameter, uptake or patient habitus.

Results: The DL method showed superior performance regarding image noise and inferior performance regarding lesion contrast in the qualitative assessment. More than 96.8% of the lesions were successfully identified in DL images. Excellent agreements on SUV in livers and lesions were found. The DL method significantly improved the liver SNR for count reduction down to 33.3% ($p < 0.001$). Lesion TBR was not significantly different between DL and reference images of the 75% dataset; furthermore, there was no significant difference either for lesions of > 10 mm or lesions in BMIs of > 25 . For the 50% dataset, there was no significant difference between DL and reference images for TBR of lesion with > 15 mm or higher uptake than liver.

Conclusions: The developed DL method improved both liver SNR and lesion TBR indicating better image quality and lesion conspicuousness compared to GS method. Compared with the reference, it showed non-inferior image quality with reduced counts by 25–50% under various conditions.

Keywords: Positron emission tomography and computed tomography (PET/CT), Deep learning, Denoising technique, Image quality

Introduction

Positron emission tomography and computed tomography (PET/CT) is a non-invasive imaging modality widely used in oncology, providing both anatomical and functional information. In oncology, PET/CT is a powerful tool for diagnosis, cancer staging and re-staging, radiation therapy planning, prognosis, and treatment-response monitoring [1–5].

To provide diagnostic PET images with sufficient image quality, a certain activity administered to the patient combined with an adequate acquisition time is recommended in the guideline for oncological ^{18}F -FDG PET imaging [1]. As radiation exposure is always a concern for both patients and operators, especially in radiation-sensitive populations [6]. In addition, the reduction on the administered activity is expected in the clinical management on the patients including those who need multiple PET examinations to monitor the therapy response [7–9]. Over the years, with the advent of advanced PET/CT scanners and image reconstruction algorithms such as time of flight (TOF), activity reduction is possible in pre-clinical studies and clinical practice [10–12]. However, the reduced injected activity/acquisition time always causes increased noise, lower signal-to-noise ratio (SNR), and potentially unnecessary artefacts in PET images.

Many post-processing techniques have been employed to reduce PET image noise, such as Gaussian filtering, Metz filtering [13], wavelet transform [14, 15], non-local mean [16, 17] and BM3D [18]. In general, the noise distribution is modelled in either the transform domain or the spatial domain, and a dedicated filter or a chain of well-designed filters are used to remove the noise. The mathematics behind these conventional methods are explicit, and only a few parameters are involved in the whole process. Thus, these methods are of limited performance in compromising the image noise and lesion detectability. Recently, deep learning methods have achieved great successes in image recognition, segmentation, registration, super-resolution, and image de-noising [19–22]. So far, three commercialized deep learning-based algorithms, designed to reduce the noise in PET images, have been approved by the U.S. Food and Drug Administration (FDA). They are SubtlePET (Subtle Medical), AiCE DLR (Canon Medical Systems Corporation), and HYPER DLR (United Imaging Healthcare).

HYPER DLR has been implemented in more than 10 sites worldwide where various clinical demands emerged according to their specific situations. Undoubtedly, a reduced injected activity is of great importance to reduce the radiation exposure to improve the patient care. On the other hand, due to the limited number of PET/CT scanners per unit population in China and other developing countries, the patient's throughput of the department was large. Thus, a reduced PET acquisition time is expected to reduce the work burden while maintaining an acceptable image quality. In addition, the algorithm has also been used on the mobile PET/CT scanners whose working time for a certain place was critical. Previous studies have been conducted to evaluate the image quality and lesion detectability of commercialized deep learning algorithms with clinical cases [23, 24]. In this study, the nuclear medicine physicians identified all the suspected lesions as in the clinical practice. Furthermore, a detailed analysis on the lesion and patient characteristics, including the patient body mass index (BMI), lesion size and uptake, was performed to comprehensively investigate the performance of the proposed deep learning method. This study aimed to investigate both the image quality and lesion detectability

of HYPER DLR with standard and reduced administered activity in comparison to the Gaussian filter in ordered subset expectation maximization (OSEM) approach. Moreover, several subgroup analyses based on different categories were performed to investigate the performance and limitation of its application in the oncological studies.

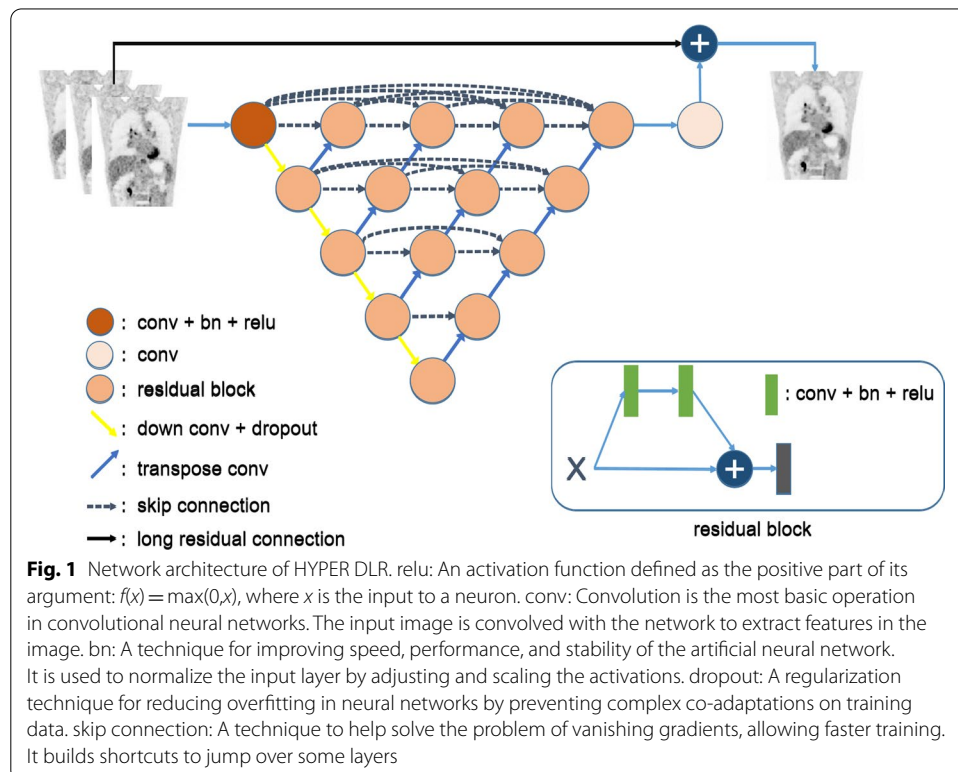
Methods

Network structure

The network used in HYPER DLR is based on the UNet architecture, and adopts the techniques of residual connection and dense connection used in the ResNet [25] and the DenseNet [26]. The structure of the proposed network is shown in Fig. 1. The top layer of the network is a long residual connection from the input to the output, which allows the network to learn the residual, i.e. the noise component, between the target image and the input image and accelerates the convergence of the deep network. From top to bottom, the upper layer has a larger dimension and higher resolution, and a dense connection is used to reduce the loss of the information. By contrast, a sparse connection is adopted in the lower layer. The network uses a large number of residual blocks which can solve the problem of gradient disappearance while improving the performance and ensuring the convergence of the deep network.

Network training and validation

In this network, a total number of 313 studies were used for training and another 80 studies were used for validation. The data were obtained from 4 sites where the PET/CT scanners manufactured by UIH have been installed. The patient age ranged from 18



to 90 years, with a median of 55 years. The injected dose was 3.1–4.3 MBq/kg, while the acquisition time was 90–180 s/bed for the body part. The retrospective PET reconstruction with 50% acquisition time was used as a training input, i.e. high-noise PET images. The retrospective reconstruction with full acquisition time as the training target, i.e. low-noise PET image. Longer acquisition time results in better training target images, and may further improve the performance of the network. All images were reconstructed with OP-OSEM algorithm incorporating TOF and point spread function (PSF). The iteration number and the subsets were fixed at 2 and 20, respectively. The voxel size of the training image was $2.34 \times 2.34 \times 2.68$ mm. In order to further improve the robustness of the network and reduce the effect of overfitting, data augmentation techniques such as horizontal and vertical flips were also used. All training images were resampled such that the images had the same size and resolution. The image intensity was normalized within the range [0, 1]. A 2.5D processing was done for the normalized PET images. That is, five slices of the input images correspond to one slice of the target image. More specifically, the network uses the image patches with a dimension of $64 \times 64 \times 5$ as the training input and the image patches with a dimension of $64 \times 64 \times 1$ as the training target. During the training process, L1 Loss function was adopted. The model was trained with the Adaptive Moment Estimation (ADAM) optimizer. The initial learning rate was set to 10^{-4} and then halved after 20 epochs. The batch size used to train the network was set to 32, and the total number of epochs to train the network was set to 200. Our network was implemented in the PyTorch framework and Python 3.7. The testing was performed on a computer with one Quadra P5000 GPU. The CUDA library was 8.0 and cuDNN version was 7.0.5.

Patients

The study included 52 consecutive patients (female/male: 19/33, age: 24–87 years) with known or suspected malignancies who were referred to the Shanghai General Hospital from 3 to 19 November 2020 for clinical ^{18}F -FDG PET/CT examinations. Their demographic and clinical data are listed in Table 1. All patients had fasted for at least 6 h prior to the PET/CT scans and a blood glucose level of ≤ 10 mmol/mL was confirmed by finger-prick sampling. A weight-based ^{18}F -FDG dose (259 ± 48 MBq, range: 188–407 MBq) was administered as an intravenous bolus to the patient. Patients were instructed to stay in a warm environment and drink 0.5–1.0 L water during the uptake. The uptake time was 66 ± 12 min (range: 45–97 min). This retrospective study was approved by the Institutional Review Board of Shanghai General Hospital, and the need for written informed consent was waived.

PET/CT imaging

Images were acquired on a digital PET/CT scanner (uMI 780, United Imaging Healthcare, China), configured with a 30-cm axial field of view. A CT scan was performed prior to PET imaging for attenuation correction and anatomical localization with a fixed tube voltage of 120 kV and an auto-mAs technique for dose modulation. Subsequently, a whole-body PET scan was performed from the skull to the mid-thigh. Data were acquired in a 3D list-mode with 180-s acquisition for each bed position and auto-adjusted overlap (25–50%).

Table 1 Patient demographic characteristics

Parameter	Value
Age (years)	62.7 ± 14.0 [24, 87] ^a
Weight (kg)	65.6 ± 11.8 [46, 110] ^a
Height (cm)	165.00 ± 8.79 [150, 184] ^a
BMI	23.9 ± 3.3 [16.3, 32.5] ^a
Injected activity (MBq)	259 ± 48 [188, 407] ^a
Uptake time (min)	66 ± 12 [45, 97] ^a
<i>Primary cancer type</i>	
Leukaemia	1 ^b
Nasopharyngeal cancer	2 ^b
Gallbladder cancer	2 ^b
Multiple myeloma	1 ^b
Lung cancer	8 ^b
Liver cancer	1 ^b
Cervical cancer	3 ^b
Laryngeal carcinoma	2 ^b
Colon cancer	4 ^b
Colorectal cancer	1 ^b
Lymphoma	7 ^b
Ovarian cancer	2 ^b
Bladder cancer	2 ^b
Renal cancer	2 ^b
Adrenocortical carcinoma	1 ^b
Oesophageal cancer	3 ^b
Gastric cancer	1 ^b
Pancreatic cancer	4 ^b
Rectal cancer	2 ^b
Endometrial cancer	3 ^b

BMI, body mass index

^a Data are presented as the mean ± standard deviation [minimum, maximum]

^b Number of patients

PET raw data were reconstructed with four different acquisition durations (180, 135, 90, and 60 s) by rebinning the list-mode data to simulate reduced injected activity. Two post-processing methods, Gaussian filter with full width at half maximum of 3 mm and HYPER DLR, were applied to each dataset. For each patient, eight series of images were generated, referred to as GS_180, GS_135, GS_90, GS_60, DLR_180, DLR_135, DLR_90, and DLR_60 groups. For the OSEM algorithm, reconstruction was performed with two iterations, 20 subsets, 192 × 192 matrix, 600 field of view, as well as TOF and PSF. For HYPER DLR, a body model was selected for the whole-body reconstruction instead of the Gaussian filter, while all other reconstruction parameters in the OSEM algorithm remained the same as those in the above-mentioned GS method. All PET reconstructions included standard corrections like decay, scatter, random, dead time, attenuation, and normalization.

Image analysis

Qualitative analysis

The PET image quality was assessed by two nuclear medicine physicians each with more than ten years of experience in interpreting PET/CT images. For each patient, the reading order of the 8 series of PET images was randomized by an independent operator. In addition, the patient's history, as well as the acquisition duration and reconstruction algorithm, were blinded to the reader. A 5-point Likert scale was used for the independent assessment of the image quality in two perspectives: lesion contrast and overall image noise. A score of 1 rated images with excessive noise and unfavourable lesion contrast. Images with sub-optimal noise and lesions with blurring leading to impaired diagnostic confidence were rated with a score of 2. A score of 3 was given if image noise and lesion contrast were substantially equivalent to those in routine oncological ^{18}F -FDG PET/CT images. A score of 4 indicated preferred image noise and good lesion contrast with diagnostically irrelevant blurring, whereas images with optimal noise and clear lesion contrast were rated with a score of 5. Figure 2 shows reference images with different Likert scores. The image quality scores independently assessed by the two raters were documented for the inter-reader agreement test.

In addition, the number of ^{18}F -FDG-avid lesions in each PET series was determined to assess the lesion detectability. The GS_180 group was used as a reference. If the result of the two physicians was not consistent, the result was determined in a consensus meeting held by a third physician experienced in nuclear medicine.

Semiquantitative analysis

Semiquantitative analyses were performed on a dedicated workstation (uWI, United Imaging Healthcare, China). A region of interest (ROI) with a diameter of 30 ± 3 mm was manually drawn at the same position and slice on a homogeneous area of the right liver lobe. The SUVmean, SUVmax, and standard deviation (SD) values of the liver for each series were recorded. The liver SNR, as a measure of image quality, was obtained by dividing the SUVmean by its SD. For identified FDG-avid lesions, SUVmax was recorded by drawing a ROI on the transverse slice with the maximum diameter. Furthermore, the target-to-background ratio (TBR) for each identified lesion, calculated by dividing the lesion SUVmax by the liver SUVmax, was used as a measure of image contrast. The diameters of the lesions were measured on the CT images for further investigation. For any missed lesion that cannot be identified in the image interpretation, a small ROI

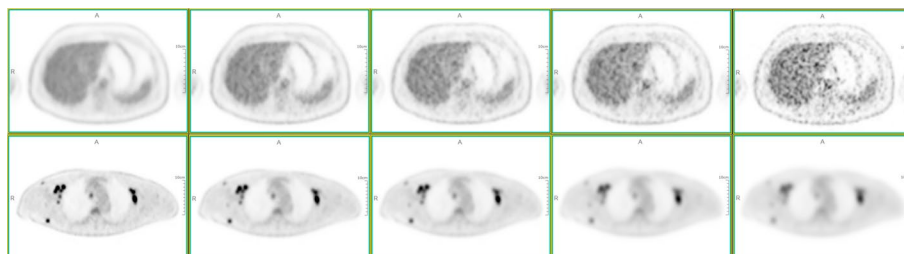


Fig. 2 Illustration of the scores from 5 to 1 (from left to right) in two perspectives. Overall image noise (upper row). Lesion contrast (bottom row)

with a diameter of around 10 mm was carefully drawn adjacent to the lesion to measure the background uptake. The tumour-to-surrounding-background ratio (TsBR) was additionally obtained by dividing the lesion SUVmax by the SUVmax of the surrounding background.

Statistical analysis

Continuous parameters are presented as the mean \pm SD and range. Inter-reader agreement was evaluated using Cohen's kappa test. The subjective scores of the image quality were subsequently compared using the Wilcoxon signed-rank test. All parameters in the semiquantitative analysis were tested for normality using the Kolmogorov–Smirnov test. Bland–Altman plot analyses were performed to assess the agreement of the SUVs between the GS and DLR groups. Concordance correlation coefficient (CCC) and linear regression were used to quantify the agreement on SUVs between groups. Subsequently, the two-tailed paired samples *t*-test was performed to investigate the differences in liver SNR, SD, and lesion TBR between groups. Furthermore, subgroup analyses were performed using the paired samples *t*-test to compare lesion TBRs between groups. Statistical significance was considered for a *p*-value less than 0.05, and all statistical tests were performed using SPSS Statistics, version 25 (IBM, Armonk, NY, USA) and Microsoft Excel.

Results

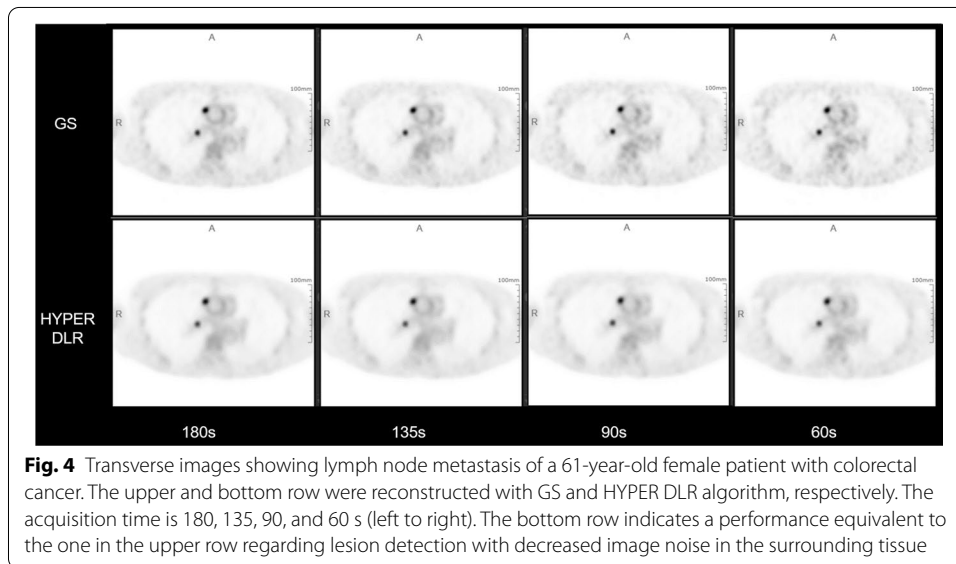
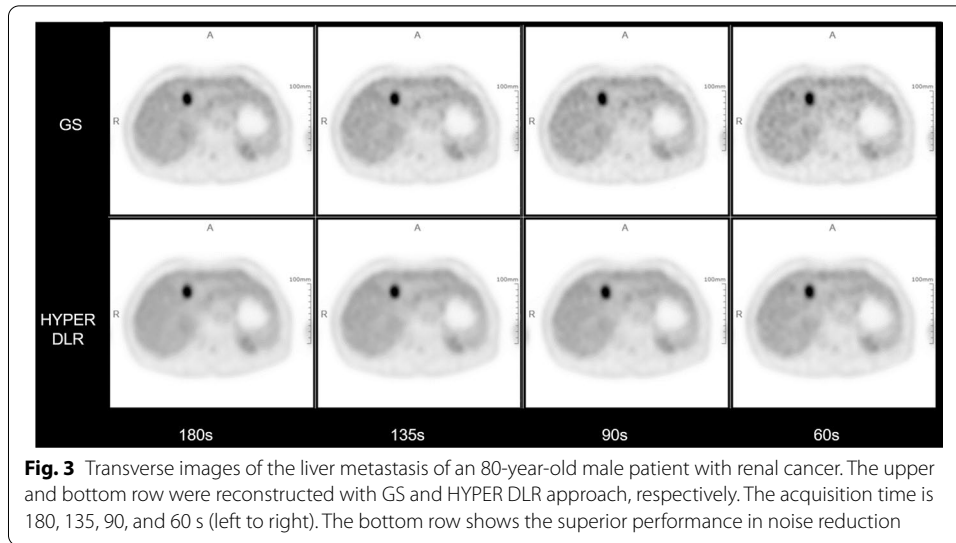
Qualitative image quality

The overall inter-reader agreement regarding lesion contrast and image noise showed kappa values of 0.705 and 0.913, indicating excellent agreement between the readers. The average scores of the two readers for each GS and DLR group are listed in Table 2. Images in all DLR groups showed significantly suppressed image contrast and reduced image noise compared to the reference GS_180 group (all $p < 0.001$). When comparing GS and DLR images with the same acquisition time, DLR method yielded more smoothed images with less noise ($p < 0.05$, Figs. 3 and 4). Moreover, all images of the DLR_180 group were rated with a score of 5 by both physicians, demonstrating the superior performance of HYPER DLR in noise reduction.

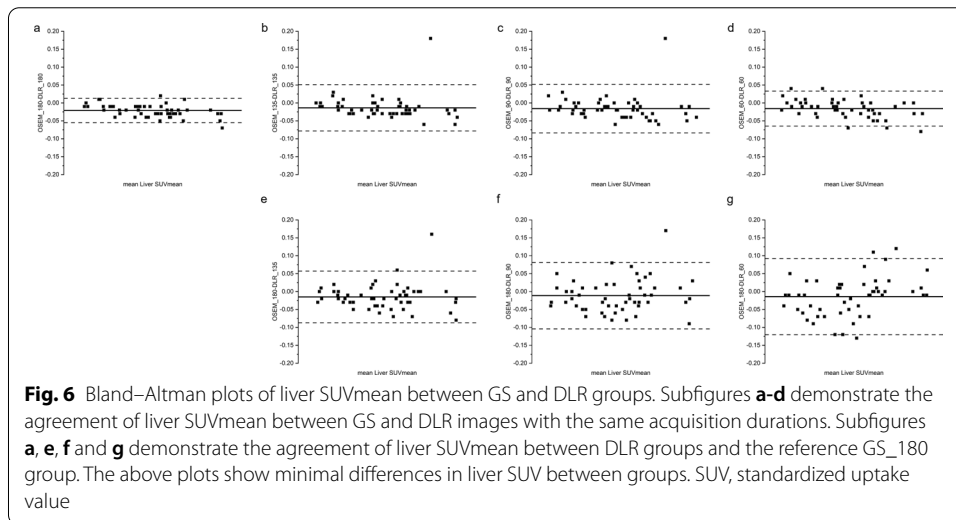
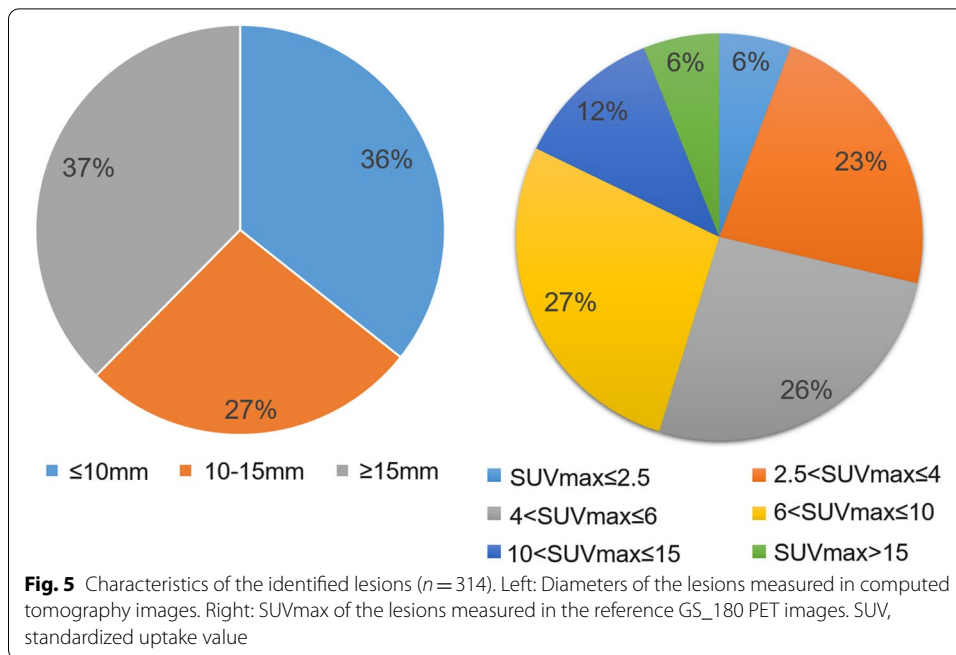
A total number of 314 FDG-avid lesions were identified in the GS_180 group and used as a reference (Fig. 5). For the DLR groups, more than 96.8% of the lesions were

Table 2 Qualitative image quality scores

	Contrast	Noise
GS_180	4.62 \pm 0.53	3.81 \pm 0.44
GS_135	4.64 \pm 0.52	3.70 \pm 0.50
GS_90	4.65 \pm 0.52	3.28 \pm 0.49
GS_60	4.38 \pm 0.83	2.54 \pm 0.54
DLR_180	4.28 \pm 0.65	5.00 \pm 0.00
DLR_135	4.26 \pm 0.70	4.97 \pm 0.17
DLR_90	4.13 \pm 0.66	4.77 \pm 0.45
DLR_60	4.00 \pm 0.19	4.13 \pm 0.41



successfully identified. The DLR_180 and DLR_135 groups had the best performances where only two ^{18}F -FDG-avid lesions were missed. Both lesions were overlooked due to an uptake similar to that of the surrounding tissue, with a low TsBR value ranging from 1.09 to 1.18. In the DLR_90 group, additional two lesions were not detected by the readers. The DLR_60 group showed the worst performance among the 8 groups regarding lesion detection with 10 missed lesions. All missed lesions had a diameter of less than 10 mm. However, the missed lesions had no impact on the cancer staging. In subsequent analyses at the location of the missed lesions, we found that all lesions were metastases, including metastatic lymph nodes and bone and liver metastases. Moreover, all missed lesions had uptake values similar to those of the surrounding tissues. The signals of these metastases were also suppressed although the image noise of the surrounding tissue was reduced.



Semiquantitative image quality

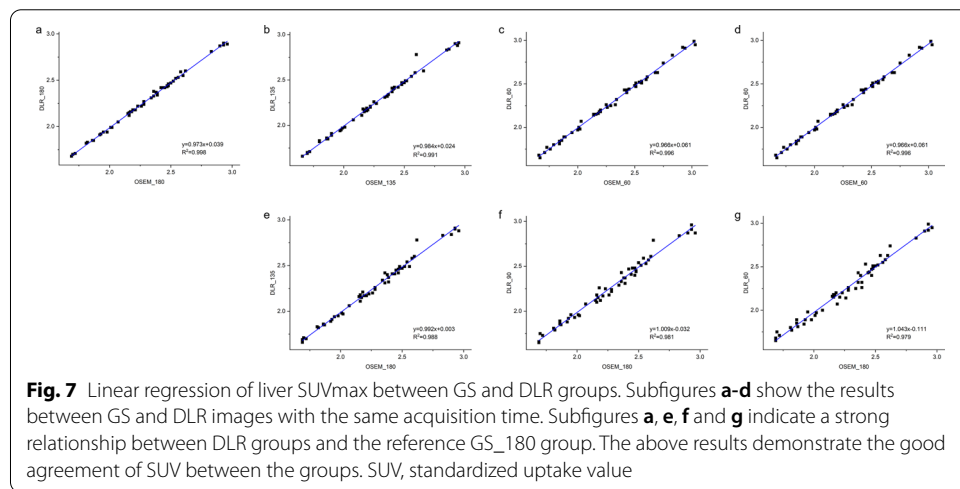
The liver SUVmean between groups agreed well as shown in Bland–Altman plots (Fig. 6). The CCCs of the liver SUVmean were all larger than 98.7% for the DLR groups compared to the GS groups, as shown in Table 3, demonstrating very strong agreement between the DLR and GS groups (Fig. 7). Because the HYPER DLR algorithm was able to suppress image noise, the liver SNR in the DLR groups was significantly higher than that in the reference GS_180 group. In this study, even the DLR_60 group showed a significantly higher SNR value than the reference GS_180 group (Fig. 8).

The equivalent tests comparing lesion SUVmax between GS and DLR groups are presented in Table 3. The CCCs of the lesion SUVmax demonstrated a very strong

Table 3 Differences in SUV of liver and lesion tissues

Parameter	Comparison group A–B	Pearson coefficient R^2	CCC
Liver SUVmean	GS_180-DLR_180	0.998	0.997
	GS_180-DLR_135	0.991	0.993
	GS_180-DLR_90	0.990	0.990
	GS_180-DLR_60	0.996	0.987
	GS_135-DLR_135	0.988	0.995
	GS_90-DLR_90	0.981	0.994
	GS_60-DLR_60	0.979	0.997
Lesion SUVmax	GS_180-DLR_180	0.994	0.991
	GS_180-DLR_135	0.991	0.989
	GS_180-DLR_90	0.984	0.985
	GS_180-DLR_60	0.975	0.977
	GS_135-DLR_135	0.993	0.989
	GS_90-DLR_90	0.990	0.985
	GS_60-DLR_60	0.986	0.978

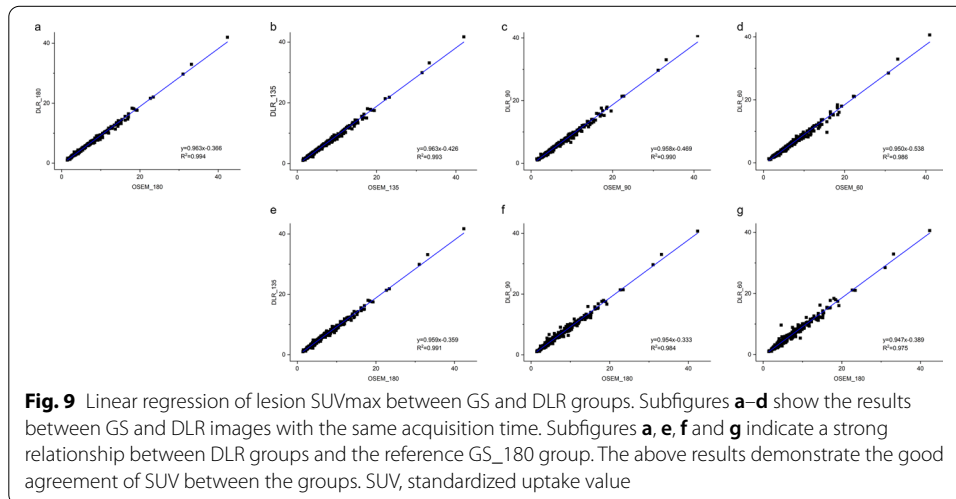
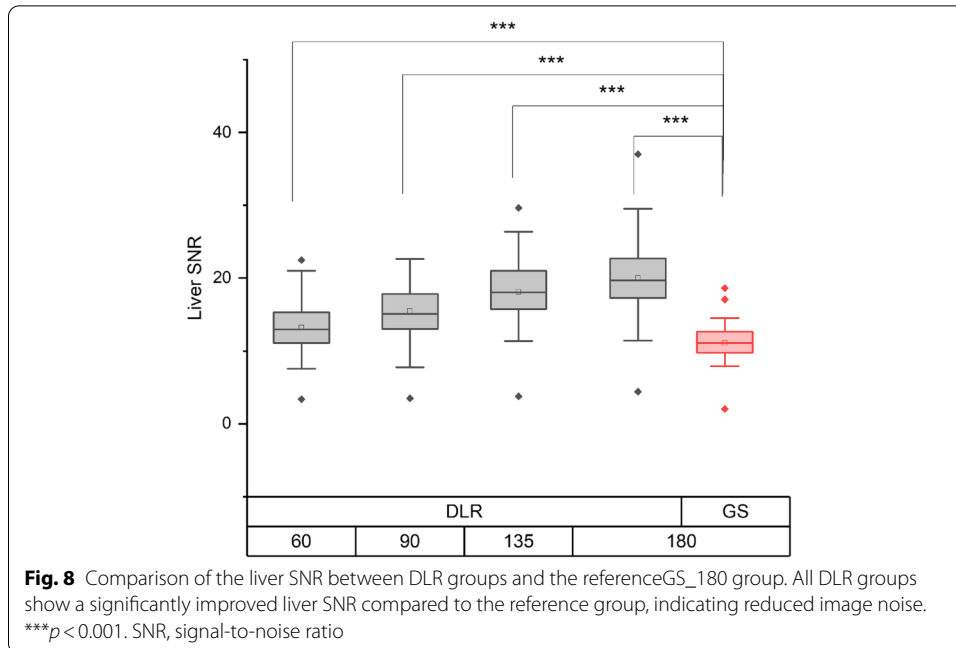
CCC, concordance correlation coefficient; SUV, standardized uptake value



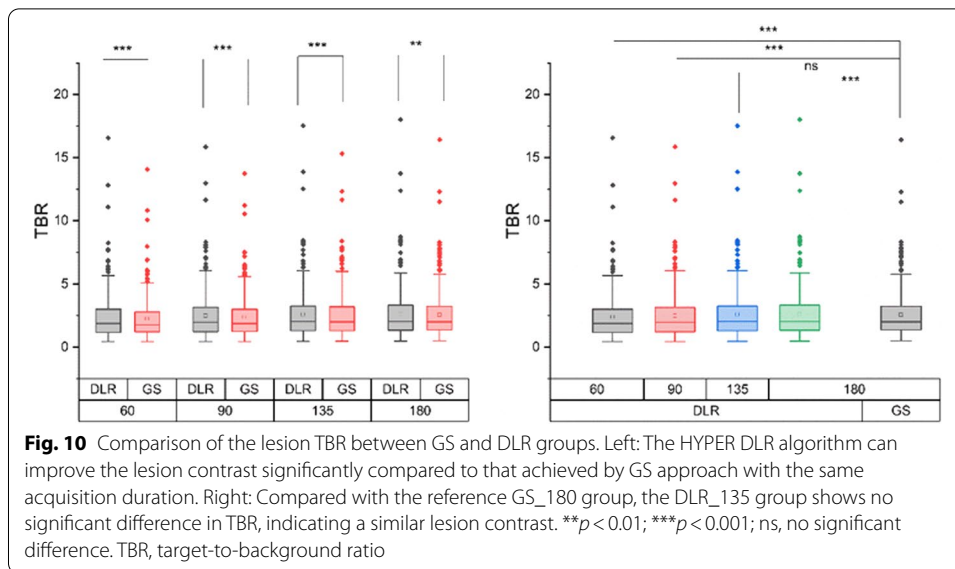
agreement between the GS and DLR groups with all values higher than 97.7% (Table 3 and Fig. 9). The DLR groups also showed a better performance in the TBR compared to the corresponding GS group with the same acquisition time (Fig. 10). There was no significant difference between the TBR values of the DLR_135 group and the reference GS_180 group ($p=0.713$), demonstrating its improvement on lesion contrast with lower counts. Both DLR_90 and DLR_60 groups showed a significantly lower TBR value than the reference (Fig. 10).

Subgroup analyses

Subsequently, a subgroup analysis was performed with different parameters. To investigate the performance of HYPER DLR regarding image contrast, a subgroup analysis was performed by dividing the lesions into two groups based on the lesion diameter measured on diagnostic CT images: a small lesion group with diameters of less than



15 mm and a large lesion group with diameters of at least 15 mm [27]. In the small lesion group ($n = 196$), the TBR was not significantly different between the DLR_180 group and the reference GS_180 group ($p = 0.051$) whereas the TBR values in other DLR groups were significantly lower than that in the reference GS_180 group (all $p < 0.001$). In the large lesion group ($n = 118$), the TBR values in the DLR_180 and DLR_135 groups were significantly improved compared to that in the reference GS_180 group ($p < 0.001$ and $p = 0.028$, respectively), and there was no significant difference between the DLR_90 group and the reference group ($p = 0.19$). This indicates that for these relatively larger lesions, the HYPER DLR algorithm has a better contrast enhancement.



In addition, we further divided the lesions using another threshold since all missed lesions had a diameter of less than 10 mm: a small lesion group with a diameter below 10 mm and a large lesion group with a diameter of 10 mm or above. In the small lesion group ($n = 112$), the TBR values of all DLR groups were significantly lower than that of the reference GS_180 group (all $p < 0.001$). However, in the large lesion group ($n = 202$), there was no significant difference in TBR between the DLR_135 group and the reference GS_180 group.

In a subsequent subgroup analysis, the lesions were divided based on whether their mean uptake value was lower than the liver uptake in the reference group. Group G1 comprised lesions whose uptake values were lower than the liver uptake values whereas all other lesions were classified as Group G2. Group G1 ($n = 37$) revealed a limitation of the HYPER DLR algorithm for lesions with lower uptake because the TBR values in all DLR groups were significantly lower than that in the reference GS_180 group (all $p < 0.001$). In Group G2 ($n = 277$), there was no significant difference between the DLR_135 group and the reference GS_180 group ($p = 0.208$).

Finally, a subgroup analysis was performed based on patient habitus. The image quality of overweight or obese patients is decreased according to a previous study [28]. In our subgroup analysis, the enrolled patients were according to WHO criteria divided into a Group B1 with a body mass index (BMI) of less than 25 kg/m² and a Group B2 with a BMI of at least 25 kg/m². In Group B1 with a total number of 220 lesions, the DLR groups showed a significant improvement in TBR compared to the corresponding GS groups with the same acquisition time. However, there was no significant difference between the DLR_135 group and the reference GS_180 group ($p = 0.501$). In Group B2 with a total number of 94 lesions, there was no significant difference between the reference GS_180 group and the DLR_180, DLR_135, and DLR_90 groups ($p = 0.094, 0.685, \text{ and } 0.098$, respectively). These findings suggest that the HYPER DLR algorithm has the capability of improving the lesion contrast in images of inferior quality.

Discussion

Deep learning-based denoising algorithms provide an effective solution for maintaining image quality with shortened acquisition time or reduced injected activity. Studies have been performed to evaluate the performance of deep learning-based methods either approved by authority organization that can be implemented in clinical use [23, 24] or still within the research field at different noise levels [29–34]. The performance of the algorithms depends on the network structure, the training strategy and the training dataset. In a recent real-world low-dose PET study on a deep learning-based denoising technique, the patients were injected with a reduction of ^{18}F -FDG by 33%, and it demonstrated the non-inferiority of AI processed image quality and lesion detectability compared with the standard method [23]. Moreover, this study included an assessment of a business case. In another study with a deep learning enhancement (DLE) model, images reconstructed with block sequential regularised expectation maximisation (BSREM) was used as the target images during the training. This study compared the results of DLE methods for different acquisition time with the full-duration BSREM method and found the models may allow a reduction in acquisition time/injected activity by 50% [34]. Our study lacked such a business case assessment, and used the standard OSEM images during the network training. However, subgroup analyses on the lesion detectability and the characteristics of the enrolled patients, lesions (size and uptake) were performed to comprehensively evaluate the performance and limitation of the HYPER DLR method. In our study, the enrolled patients were from an independent site and presented various conditions regarding age, weight, height, and cancer type to provide a heterogeneous sample reflecting common clinical practice. Short-duration datasets were reconstructed with the OSEM and the proposed HYPER DLR method, of which the image quality and lesion detectability were compared. In our clinical evaluation, the HYPER DLR method had a superior capability for noise reduction compared to the GS method from both subjective and semiquantitative perspectives. Moreover, it can improve the image quality with a shorter acquisition time, indicating its potential in studies with lower administered activity. According to our study results, the administered activity can be reduced by 25% without compromising quantitative SUV and TBR values compared to the standard dose.

In addition, the HYPER DLR algorithm can provide better performance in patients with higher BMI. It is well-known that the PET image quality deteriorates with increasing patient weight for linear weight-based ^{18}F -FDG dose regimens [26]. However, our study demonstrates that the administered activity in overweight and obese patients can even be reduced by 50% without compromising image quality as compared to the standard FDG administration. This will be helpful in reducing the total activity administered to overweight and obese patients and will reduce radiation exposure-induced risks for patients and operators [35, 36].

In oncology, lesion detectability is important in the early diagnosis and staging of patients. Early detection of lesions can lead to a positive patient prognosis, so the ability to detect small, low-intensity/uptake lesions is of high importance and was evaluated in the study. HYPER DLR was able to maintain a lesion detectability over 98.6% even if the acquisition time was reduced by 66.7% of that during standard acquisition. If the acquisition time was only reduced by 25%, the lesion detectability can be up to 99.4%, with only

two lesions missed in our study. These two lesions were adjacent to the aorta and the scalp, respectively, where the physiological uptake of the surrounding tissues was relatively high. After checking the uptake of the missed lesions, the maximum SUV_{max} was up to 3.64, measured in a liver metastasis of a 74-year-old female with gallbladder carcinoma (Additional file 1: Fig. S1). In both DLR₉₀ and DLR₆₀ images, this metastasis was overlooked due to high image noise and respiratory motion artefacts. Moreover, the liver was found to have a relatively high uptake which may also have impacted the identification of this metastasis. Similarly, all missed lesions had uptake values approximately comparable with those of the surrounding tissues, with TsBR values from 1.004 to 1.595. We further investigated the lesion with the maximum TsBR value and found it to be a pelvic metastatic lymph node adjacent to the intestinal tract (Additional file 2: Fig. S2). It was challenging to distinguish this small lesion with a diameter of 6.1 mm from the physiological uptake observed in the adjacent intestinal tract. The missed lesions had an average diameter of 8.6 ± 1.3 mm (range: 6.1–10.0 mm). According to our subgroup analysis, TBR values, as a measure of contrast, of small lesions with a diameter of less than 10 mm were suboptimal compared to those in GS groups. This may impede the lesion identification in HYPER DLR PET images.

The subjective assessment of the image quality performed by nuclear medicine physicians demonstrated the superior noise reduction performance of the HYPER DLR algorithm. Even if the acquisition time was reduced by 66.7% in the HYPER DLR group, the performance of this group regarding image noise was superior to that of the reference GS₁₈₀ group. The noise reduction in PET images can help physicians better interpret PET/CT images where false-positive uptakes were reduced. However, regarding lesion contrast, the subjective assessment showed a better performance of the GS method compared to the HYPER DLR algorithm which was not consistent with the later semi-quantitative analysis in this study. This bias may mainly be due to the long-term experience in using the GS method. The subjective assessment might improve when physicians get used to the assessment of HYPER DLR images.

Deep learning methodologies, although achieved superior performance against conventional methods, still have room to improve. Firstly, the network architecture should be re-designed. For example, using a fully 3D architecture instead of 2.5D or incorporating the attention mechanism into the network. Secondly, a large and high quality training dataset is helpful. To train a stable and reliable network, the training dataset is expected to cover a large population with a variety of age, sex, weight, races, types of diseases, and covers typical clinical scenarios such as acquisition time and injected activity. Moreover, the learning target is expected to have the lowest image noise and the highest image contrast. To achieve this goal, the acquisition time should be long enough to collect adequate counts. It is very hard for the conventional PET/CT scanners because of their limited sensitivity. Benefits from its ultra-high sensitivity, total-body or long axial field-of-view (LAFOV) PET/CT scanners such as uEXPLORER [37], PennPET [38] and Biograph Vision Quadra [39] have demonstrated extraordinary image quality than conventional PET/CT scanners. So if the training data are from the total-body or LAFOV PET/CT scanners, the network performance may be greatly improved. Thirdly, the denoising step could be integrated into the iterative reconstruction algorithm, not just work after the image reconstruction. Contrast loss is inevitable when post-processing

the high noise images. This is because the small lesions are almost overwhelmed by the background noise. The signal features extracted by the network are too weak to be differed from the noise. To tackle this problem, combination of iterative reconstruction algorithm and neural networks has been actively studied [39–42]. This may inspire the development of deep learning methodologies with both reduced image noise and improved image contrast.

This study has several limitations. It was a single-centre study with limited retrospective data from one PET/CT system. Further studies on a larger number of patients in a clinical setting using different PET/CT systems may be required. The cohort of overweight and obese patients enrolled in this study was small, which may introduce bias into quantitative analyses for this group of our study population. Detectability of small lesions is affected by many factors, e.g. lesion size, uptake, shape, location, respiratory motion, and, most importantly, the acquisition time. The analysis of statistical fluctuations of reduced acquisition time in small lesion measurements is necessary, but it is beyond the scope of this study. This study used shorted acquisition time to simulate the scenario of reduced activity, and future study can be performed with a real-world reduced injected activity to validate our findings. Up to now, HYPER DLR method has only been trained with 18F-FDG PET images, and thus can only be applied to 18F-FDG PET applications. However, PET imaging with other tracers, such as prostate specific membrane antigen (PSMA) and Fibroblast activation protein inhibitor (FAPI), is becoming more widely used. The HYPER DLR can be applied in other tracers if the network is trained in the future.

Conclusion

In PET/CT oncological images used in this study, the developed deep learning algorithm HYPER DLR showed improvements in image quality compared to that post-processed by Gaussian filter. HYPER DLR reconstruction yielded improved results in liver SNR even if the acquisition duration was shortened to 33.3% of that in the reference group, indicating the superior noise reduction of this approach. In addition, HYPER DLR can generate images with uncompromised lesion contrast with an acquisition duration shortened by 25–50%. Thus, it offers potential applications in PET imaging with reductions in required FDG activity or acquisition time.

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s40658-022-00431-9>.

Additional file 1. Fig. S1: Liver metastasis of a 74-year-old female with gallbladder carcinoma. The lesion (red arrow) was identified in the reference OSEM_180 images with a measured SUVmax of 3.64 (left) but missed in the DLR_90 and DLR_60 images due to the high image noise and respiratory motion artefacts. SUV, standardized uptake value.

Additional file 2. Fig. S2: The lesion with the maximum TsBR value among all missed lesions was found to be a pelvic metastatic lymph node of a 66-year-old male with bladder cancer. It was challenging to distinguish the lesion with a diameter of 6.1 mm from the physiological uptake observed in the intestinal tract (line intersection, right). TsBR, tumour-to-surrounding-background ratio.

Authors' contributions

Study conception and design: JZ and CL. Material preparation and data collection: YL, CX and SL. Data analysis: YX, WQ, TW and ZQ. Drafting of the manuscript: YX and YW. All authors commented on previous versions of the manuscript and read and approved the final manuscript.

Funding

This study is supported by Shanghai Science and Technology Commission (17411953200) and Shanghai Municipal Health Commission (201840076).

Availability of data and materials

Data are available on request to the corresponding author.

Code availability

Not applicable.

Declarations**Ethics approval and consent to participate**

All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki Declaration and its later amendments or comparable ethical standards. The need for written informed consent was waived by the Institutional Review Board of Shanghai General Hospital.

Consent for publication

Patients signed informed consent regarding publishing their data and photographs for scientific use.

Competing interests

Ying Wang, Chenwei Li, Yang Lv, Chen Xi and Zheng Qian are all full-time employees of United Imaging Healthcare, China. Shu Liao is a full-time employee of Shanghai United Imaging Intelligence, China. All other authors have no conflicts of interest to report.

Author details

¹Department of Nuclear Medicine, Shanghai General Hospital, Shanghai Jiaotong University, No. 100 Haining Road, Shanghai 200080, People's Republic of China. ²United Imaging Healthcare, Shanghai, People's Republic of China. ³Shanghai United Imaging Intelligence Co. Ltd, Shanghai, People's Republic of China.

Received: 15 June 2021 Accepted: 12 January 2022

Published online: 04 February 2022

References

1. Boellaard R, Delgado-Bolton R, Oyen WJ, et al. FDG PET/CT: EANM procedure guidelines for tumour imaging: version 2.0. *Eur J Nucl Med Mol Imaging*. 2015;42:328–354.
2. Czernin J, Allen-Auerbach M, Nathanson D, Herrmann K. PET/CT in Oncology: Current Status and Perspectives. *Curr Radiol Rep*. 2013;1:177–90.
3. Fletcher JW, Djulbegovic B, Soares HP, et al. Recommendations on the use of 18F-FDG PET in oncology. *J Nucl Med*. 2008;49:480–508.
4. Weber WA. Use of PET for monitoring cancer therapy and for predicting outcome. *J Nucl Med*. 2005;46:983–95.
5. Cuaron J, Dunphy M, Rimmer A. Role of FDG-PET scans in staging, response assessment, and follow-up care for non-small cell lung cancer. *Front Oncol*. 2013;2:208.
6. Pearce MS, Salotti JA, Little MP, et al. Radiation exposure from CT scans in childhood and subsequent risk of leukaemia and brain tumours: a retrospective cohort study. *Lancet*. 2012;380(9840):499–505.
7. Barrington SF, Mikhaeel NG, Kostakoglu L, et al. Role of imaging in the staging and response assessment of lymphoma: consensus of the International Conference on Malignant Lymphomas Imaging Working Group [published correction appears in *J Clin Oncol*. 2016 Jul 20;34(21):2562]. *J Clin Oncol*. 2014;32(27):3048–3058.
8. Iovoli AJ, Farrugia MK, Ma SJ, Chan JM, Markiewicz MR, McSpadden R, Wooten KE, Gupta V, Kuriakose MA, Hicks WL Jr, Singh AK. Role of Repeat PET/CT Imaging in Head and Neck Cancer Following Initial Incomplete PET/CT Response to Chemoradiation. *Cancers*. 2021;13(6):1461.
9. Castello A, Rossi S, Lopci E. 18F-FDG PET/CT in Restaging and Evaluation of Response to Therapy in Lung Cancer: State of the Art. *Curr Radiopharm*. 2020;13(3):228–37. <https://doi.org/10.2174/1874471013666191230144821>.
10. Akamatsu G, Ishikawa K, Mitsumoto K, et al. Improvement in PET/CT image quality with a combination of point-spread function and time-of-flight in relation to reconstruction parameters. *J Nucl Med*. 2012;53:1716–22.
11. Buzhan P, Dolgoshein B, Filatov LA, et al. Silicon photomultiplier and its possible applications. *Nuclear Instruments & Methods in Physics Research Section A-accelerators Spectrometers Detectors and Associated Equipment*. 2003;504:48–52.
12. Conti M. State of the art and challenges of time-of-flight PET. *Phys Med*. 2009;25:1–11.
13. Jacobson M, Levkovitz R, Ben-Tal A, et al. Enhanced 3D PET OSEM reconstruction using inter-update Metz filtering. *Phys Med Biol*. 2000;45:2417–39.
14. Bousson N, Cheze Le Rest C, Hatt M, Visvikis D. Incorporation of wavelet-based denoising in iterative deconvolution for partial volume correction in whole-body PET imaging. *Eur J Nucl Med Mol Imaging*. 2009;36:1064–75.
15. Bal A, Banerjee M, Sharma P, Maitra M. An efficient wavelet and curvelet-based PET image denoising technique. *Med Biol Eng Comput*. 2019;57(12):2567–98.
16. Dutta J, Leahy RM, Li Q. Non-local means denoising of dynamic PET images. Muñoz-Barrutia A, editor. *PLoS One* 2013;8:e81390.
17. Arabi H, Zaidi H. Non-local mean denoising using multiple PET reconstructions. *Ann Nucl Med*. 2021;35:176–86.

18. Peltonen S, Tuna U, Ruotsalainen U, Low count PET sinogram denoising. IEEE Nuclear Science Symposium and Medical Imaging Conference Record (NSS/MIC). Anaheim, CA, USA. 2012;2012:3964–7.
19. Zhou L, Schaefferkoetter JD, Tham IWK, Huang G, Yan J. Supervised learning with cyclegan for low-dose FDG PET image denoising. *Med Image Anal.* 2020;65:101770.
20. Lu W, Onofrey JA, Lu Y, et al. An investigation of quantitative accuracy for deep learning based denoising in oncological PET. *Phys Med Biol.* 2019;64:165019.
21. Cui J, Gong K, Guo N, et al. PET image denoising using unsupervised deep learning. *Eur J Nucl Med Mol Imaging.* 2019;46:2780–9.
22. Ronneberger O, Fischer P, Brox T. U-Net: Convolutional Networks for Biomedical Image Segmentation. In *Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention*, Munich, Germany. 2015;5–9:234–41.
23. Katsari K, Penna D, Arena V, et al. Artificial intelligence for reduced dose 18F-FDG PET examinations: a real-world deployment through a standardized framework and business case assessment. *EJNMMI Phys.* 2021;8(1):25. Published 2021 Mar 9. doi:<https://doi.org/10.1186/s40658-021-00374-7>
24. Tsuchiya J, Yokoyama K, Yamagiwa K, et al. Deep learning-based image quality improvement of 18F-fluorodeoxyglucose positron emission tomography: a retrospective observational study. *EJNMMI Phys.* 2021;8(1):31. Published 2021 Mar 25. doi:<https://doi.org/10.1186/s40658-021-00377-4>
25. He K, Zhang X, Ren S, Sun J. Deep Residual Learning for Image Recognition. *IEEE Conference on Computer Vision and Pattern Recognition (CVPR).* 2016;2016:770–8. <https://doi.org/10.1109/CVPR.2016.90>.
26. Huang G, Liu Z, Van Der Maaten L, Weinberger KQ. Densely Connected Convolutional Networks. *IEEE Conference on Computer Vision and Pattern Recognition (CVPR).* 2017;2017:2261–9. <https://doi.org/10.1109/CVPR.2017.243>.
27. Nguyen NC, Vercher-Conejero JL, Sattar A, et al. Image Quality and Diagnostic Performance of a Digital PET Prototype in Patients with Oncologic Diseases: Initial Experience and Comparison with Analog PET. *J Nucl Med.* 2015;56:1378–85.
28. Sánchez-Jurado R, Devis M, Sanz R, Aguilar JE, del Puig CM, Ferrer-Rebolleda J. Whole-body PET/CT studies with lowered ¹⁸F-FDG doses: the influence of body mass index in dose reduction. *J Nucl Med Technol.* 2014;42:62–7.
29. Sanaat A, Shiri I, Arabi H, Mainta I, Nkoulou R, Zaidi H. Deep learning-assisted ultra-fast/low-dose whole-body PET/CT imaging [published online ahead of print, 2021 Jan 25]. *Eur J Nucl Med Mol Imaging.* 2021;<https://doi.org/10.1007/s00259-020-05167-1>. doi:<https://doi.org/10.1007/s00259-020-05167-1>
30. Schaefferkoetter J, Yan J, Ortega C, et al. Convolutional neural networks for improving image quality with noisy PET data. *EJNMMI Res.* 2020;10(1):105. Published 2020 Sep 21. doi:<https://doi.org/10.1186/s13550-020-00695-1>
31. 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(9):1388–96. <https://doi.org/10.2967/jnumed.119.239327>.
32. Cui J, Gong K, Guo N, et al. PET image denoising using unsupervised deep learning. *Eur J Nucl Med Mol Imaging.* 2019;46(13):2780–9. <https://doi.org/10.1007/s00259-019-04468-4>.
33. Wang YJ, Baratto L, Hawk KE, et al. Artificial intelligence enables whole-body positron emission tomography scans with minimal radiation exposure [published online ahead of print, 2021 Feb 1]. *Eur J Nucl Med Mol Imaging.* 2021;<https://doi.org/10.1007/s00259-021-05197-3>. doi:<https://doi.org/10.1007/s00259-021-05197-3>
34. Mehranian A, Wollenweber SD, Walker MD, et al. Image enhancement of whole-body oncology [18F]-FDG PET scans using deep neural networks to reduce noise [published online ahead of print, 2021 Jul 28]. *Eur J Nucl Med Mol Imaging.* 2021;<https://doi.org/10.1007/s00259-021-05478-x>. doi:<https://doi.org/10.1007/s00259-021-05478-x>
35. Huang B, Law MW, Khong PL. Whole-body PET/CT scanning: estimation of radiation dose and cancer risk. *Radiology.* 2009;251:166–74.
36. Botkin CD, Osman MM. Prevalence, challenges, and solutions for (18)F-FDG PET studies of obese patients: a technologist's perspective. *J Nucl Med Technol.* 2007;35:80–3.
37. Badawi RD, Shi H, Hu P, et al. First Human Imaging Studies with the EXPLORER Total-Body PET Scanner. *J Nucl Med.* 2019;60(3):299–303. <https://doi.org/10.2967/jnumed.119.226498>.
38. Pantel AR, Viswanath V, Daube-Witherspoon ME, et al. PennPET Explorer: Human Imaging on a Whole-Body Imager. *J Nucl Med.* 2020;61(1):144–51. <https://doi.org/10.2967/jnumed.119.231845>.
39. Alberts I, Hünermund JN, Prenosil G, 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(8):2395–404. <https://doi.org/10.1007/s00259-021-05282-7>.
40. Kim K, Wu D, Gong K, et al. Penalized PET Reconstruction Using Deep Learning Prior and Local Linear Fitting. *IEEE Trans Med Imaging.* 2018;37(6):1478–87. <https://doi.org/10.1109/TMI.2018.2832613>.
41. Reader AJ, Corda G, Mehranian A, Costa-Luis CD, Ellis S, Schnabel JA. Deep Learning for PET Image Reconstruction. *IEEE Transactions on Radiation and Plasma Medical Sciences.* 2020;5(1):1–25.
42. Lv Y, Xi C. PET image reconstruction with deep progressive learning. *Phys Med Biol.* 2021;66(10):<https://doi.org/10.1088/1361-6560/abfb17>. Published 2021 May 14. doi:<https://doi.org/10.1088/1361-6560/abfb17>

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.