Supplementary Materials

Table of Contents

1. Patient Cohort and Image Acquisition Description

In this restrospective study, Health Insurance Portability and Accountability (HIPAA)-compliant clinical trial, two participating centers (University of Tübingen, Germany and Stanford University, CA, USA) obtained approval from their institutional review board (IRB). In addition, Stanford University obtained IRB approval to collect de-identified imaging studies in a centralized image registry, along with relevant clinical information (patient age, sex, tumor type). Written informed consent was obtained from all adult patients and parents of pediatric patients. In addition, children were asked to give their assent. Inclusion criteria were comprised of the following: (1) age < 30 years, (2) histologically proven lymphoma and (3) PET/MRI scan at baseline before chemotherapy. Exclusion criteria were (1) MR-incompatible metal implants, (2) claustrophobia, and (3) pregnancy. Between July 2015 and June 2019, we enrolled 22 children and young adults (13 female, 9 male) with lymphoma and a mean age (standard deviation; range) of 17 years (7; range: 6-30 years). Tumor histology consisted of 14 patients with Hodgkin lymphoma, 6 with non-Hodgkin lymphoma and 2 patients with posttransplant lymphoproliferative disorder (PTLD). For Tübingen, 10 patients were enrolled (5 female, 5 male) with a mean age (standard deviation; range) of 14 years (5; range: 3-18 years) and the following distribution of tumor histology: 8 with Hodgkin lymphoma, 2 with non-Hodgkin lymphoma.

Stanford patients underwent a whole body integrated ^{18}F-FDG PET/MRI scan at baseline on a 3T Signa PET/MRI scanner (GE Healthcare, Milwaukee, WI, USA), using a 32-channel torso phased array coil and an eight-channel, receive-only head coil. Before the scan, patients had to fast for at least 4 hours and blood glucose levels had to be below 140mg/dl. ^{18}F-FDG was administered intravenously 60 minutes before the scan at a dose of 3 megabecquerel per kg body weight. The imaging protocol consisted of an axial T1-weighted two-point Dixon Liver Acquisition with Volume Acquisition (LAVA) sequence (repetition time (TR) 4.2 ms, echo time (TE) 1.1, 2.3 ms, flip angle (FA) 5, slice thickness (SL) 5.2 mm) for attenuation correction and a higher-resolution LAVA sequence (TR 4.2 ms, TE 1.7, 3.4 ms, FA 15, SL 3,4 mm) for anatomical co-registration. PET data were acquired simultaneously with MRI scans, using a 25 cm transaxial FOV and 3:30 minute acquisitions per PET bed. PET data was reconstructed using scanner-specific algorithms, (3D OSEM: 28 subsets, 2 iterations, with time of flight and point spread function information), accounting for attenuation from coils and patient cradle.

Tübingen patients underwent a whole body integrated ^{18}F-FDG PET/MRI scan at baseline on a 3T Signa PET/MRI scanner (Siemens Healthineers, Erlangen, Germany), using a 16-channel torso phased array coil and a 16-channel head coil. Before the scan, patients had to fast for at least 4 hours and blood glucose levels had to be below 140mg/dl. ^{18}F-FDG was administered intravenously 60 minutes before the scan at a dose of 3 megabecquerel per kg body weight. The imaging protocol consisted of an axial T1-weighted two-point Dixon Volume Interpolated Breathhold Acquisition (VIBE) sequence (TR 3.95 ms, TEs 1.23, 2.46 ms, FA 10°, SL 3 mm) for attenuation correction and anatomical co-registration. PET data were acquired simultaneously with

MRI scans, using a 25 cm transaxial FOV and 4 minutes acquisitions per PET bed. PET data was reconstructed using scanner-specific algorithms, (3D OSEM: 21 subsets, 2 iterations), accounting for attenuation from coils and patient cradle.

Radiotracer input data were used to generate images. Full-dose (3 MBq/kg) PET data were acquired in list mode, which helps detect coincidence events across the entire duration of the PET bed time (3 minutes 30 seconds). Low- dose PET images were retrospectively simulated by unlisting the PET list-mode data and reconstructing them based on the percentage of coincidence events (22). The list-mode PET input data collected over a time period of the first block of 3 minutes 30 seconds, 2 minutes 38 seconds, 1 minute 45 seconds, 53 seconds, 26 seconds, 13 seconds, and 2 seconds, were used to simulate 100%, 75%, 50%, 25%, 12.5%, 6.25%, and 1% ^{18}F-FDG dose levels. For data acquired at the Tübingen site, PET Acquisition time was four minutes per bed position and low-dose PET images were simulated using the same relative dose levels accordingly.

2. Image Preprocessing

DICOM to NifTI conversion was performed with "dcm2niix" command. "dcm2niix" (http://manpages.ubuntu.com/manpages/bionic/man1/dcm2niix.1.html) is designed to convert neuroimaging data from the DICOM format to the NIfTI format and can be performed using a simple command-line interface from Ubuntu system. The pre-processing pipeline aimed to remove the additional burden of the network learning methods to find patterns between images for final reconstruction. We adopted the Convert3D (http://www.itksnap.org/c3d/)command-line tool for PET and MRI re-slicing. This ensures all of the scans are in the same image dimension. In addition, we used ITK-Snap[1] to label the foreground body area in the scan. These body masks are used to mask out the background area of the PET and MRI images to avoid introducing unnecessary noise for reconstruction. Then, top 0.1% of the pixels in PET images are clipped. Note that the clipping operation helps model convergence and stabilize training as these top pixels possess extreme high value and are outliers of the distribution. The majority of the top 0.1% pixels are located within the bladder and the brain regions. Thus, the SUV values of the significant regions, including liver and lesions, are not much affected after the preprocessing. Finally, all scans are normalized between zero and one before feeding them to the neural network model and denormalized after the neural network reconstruction.

3. Deep Learning Models and Training Details

For the model design in low-count PET reconstruction, a skip connection between the low-dose ^{18}F-FDG PET input and the final prediction layer is added to alleviate the burden of carrying identity information in the reconstruction network. We adopted four-fold cross-validation for the Stanford cohort. Each fold has 36 PET/MRI scans (from 17 patients; except 18 patients in fold #4) for training, and 8 scans for testing as well as 4 scans for validation (from 6 patients; except 5 patients in fold #4). In terms of the training strategy for the five AI algorithms, we experimented the optimal configuration for each of the algorithms. For U-net, we trained the model with AdamW[2], using $\beta_1 = 0.9$ and $\beta_2 = 0.99$, with a linearly decay learning rate schedule (initialized as 1e-3, decay step-size = 8 epochs, decay gamma= 0.8). For EDSR, the setting is the same as U-net, besides learning rate initialized as 1e-4. SwinIR shares the same configure as EDSR. For EDSR-ViT, we initialized the EDSR encoder part with the EDSR model trained on the Stanford PET/MRI cohort, and initialized the ViT part with the ViT pretrain on ImageNet. Then following[3], we used a warm-up learning rate (5 epochs) and then linearly decay the learning rate over the course of EDSR-ViT training. For GAN, the linear decay learning rates were initialized as 5e-4 and 1e-3 for generator and discriminator respectively (decay step-size = 8 epochs, decay gamma= 0.8). AdamW optimizer was adopted with $\beta_1 = 0.9$ and $\beta_2 = 0.99$. While training the generator of GAN, the weights of the discriminator network are kept constant, and vice versa. The parameter updating of generator and discriminator are processed alternately every other iteration. The training was performed on four NVIDIA GeForce RTX 3090 GPUs with 24GB VRAM.

4. Composite Loss Function

The loss function is a cornerstone of neural network models and determines the optimizing direction for model training. We applied the commonly used loss functions for the image restoration task –MSE (mean square error) loss and SSIM (structural similarity index measure) loss.The SSIM loss encourages production of output images that are structurally similar to the target image. Together with MSE loss and SSIM loss, the composite loss function (as below) for optimizing deep learning reconstruction models encourages the PET reconstruction process to reduce noise, keep textures, and preserve structural details. The PIQA (PyTorch Image Quality Assessment, version 1.1.7) implementation of SSIM was used to compute the SSIM loss term.

$$\min \mathcal{L} = \lambda_m \mathcal{L}_{MSE} + \lambda_s \mathcal{L}_{SSIM}$$

In the generative adversarial networks (GAN), the generator and discriminator are trained simultaneously in an adversarial process. Pairs of real PET images and outputs from the generator are fed into the discriminator. The discriminator aims to distinguish real images and predictions from the generator. The loss function of the discriminator is the cross entropy loss of the classification labels. The loss function of the generator consists of three components: 1). the mean square error (MSE) loss of the generated images and the real PET images; 2). the SSIM loss of the generated images and the real PET images; 3). the adversarial loss computed from the output of the discriminator. Together, the composite loss function encourages the generator to produce plausible translations of the source images. The loss function for the generator from GAN is formulated as below.

$$\min \mathcal{L} = \lambda_m \mathcal{L}_{MSE} + \lambda_s \mathcal{L}_{SSIM} + \lambda_d \mathcal{L}_{discrimintor}$$

5. Lessons from Model Training and Experiments

As the radiotracer dose reduces, the AI model gets more sensitive to the hyper-parameters. We found that results can be significantly improved with careful hyperparameter choice – e.g. the initializing learning-rate - for dose below 12.5%. While for the dose above 12.5%, the AI models possess more robustness towards the hyperparameter configurations. For all algorithms, we found that initializing the convolutional layers with the orthogonal initialization[4] enables more efficient convergence as opposed to other initialization methods including Kaiming uniform initialization[5] and Xavier uniform initialization[6].

6. Quantitative Assessment

For evaluation, three quantitative metrics were adopted to measure the quality of the reconstructed PET images, including SSIM (the structural similarity index), PSNR (peak signal-to-noise ratio), and VIF (Visual information fidelity). The higher the SSIM, PSNR and VIF, the better degraded image has been reconstructed to match the original image. The code for calculating the performance was written with Python and using Scikit-image toolkit, as below. The VIF metric refers to the implementation of https://github.com/aizvorski/video-quality/blob/master/vifp.py.

```
psnr = skimage.metrics.peak_signal_noise_ratio(y_true, y_pred)
ssim = skimage.metrics.structural_similarity(y_true, y_pred, multichannel=True, K1=0.0001, K2=0.0003)
vif = VIF(groundtruth, test)
```

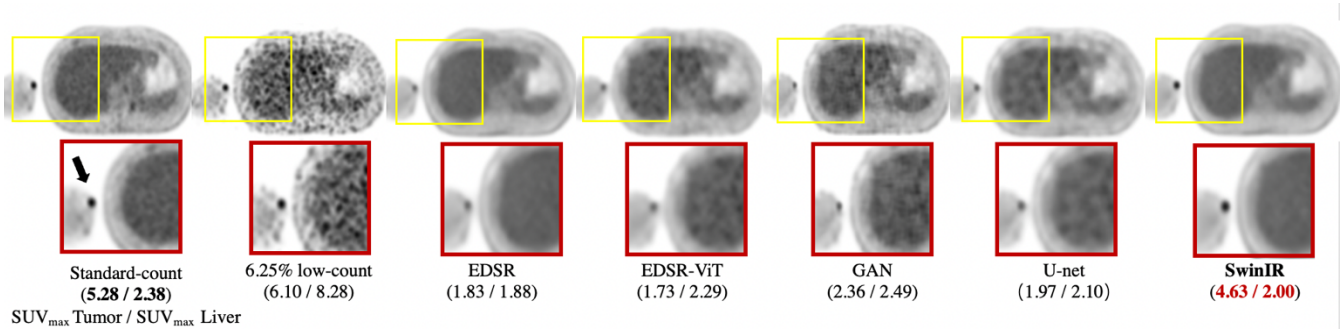
The SUV metric in the study was measured using OsiriX version 12.5.1. (OsiriX software). The SUV_{max} of the target lesions and SUV_{max} of liver were measured by placing a three-dimensional volume of interest over tumor lesions, and liver. SUV values were calculated based on patient body weight by using the following equation: SUV = tissue tracer activity (in millicuries per milliliter)/[injected dose (in millicuries)/patient body weight (in grams)].

7. More Image Comparison of PET Reconstructions

Supplementary Table 1. Peak signal-to-noise ratio (PSNR) and Standardized uptake values (SUV), as measured on 100% standard-dose ^{18}F-FDG PET, original low-count PET scans, and AI-reconstructed PET scans (U-Net and SwinIR).

PET Modality	Liver. PSNR Mean (SD)	Liver SUVmean Mean (SD)	Liver SUVmax Mean (SD)
100% standard-dose PET	N/A	1.42 (0.44)	1.58 (0.47)
75% low-count PET	47.1 (2.36)	1.43 (0.43)	1.61 (0.42)
SwinIR	40.5 (4.29)	1.25 (0.38)	1.43 (0.39)
U-net	38.8 (4.07)	1.22 (0.36)	1.39 (0.37)
50% low-count PET	39.4 (1.81)	1.42 (0.44)	1.80 (0.50)
SwinIR	39.1 (2.55)	1.26 (0.40)	1.52 (0.47)
U-net	36.9 (3.38)	1.22 (0.35)	1.48 (0.43)
25% low-count PET	36.83 (1.77)	1.45 (0.45)	1.85 (0.50)
SwinIR	37.0 (2.82)	1.19 (0.39)	1.50 (0.44)
U-net	35.6 (3.16)	1.17 (0.32)	1.48 (0.33)
12.5% low-count PET	32.8 (1.82)	1.45 (0.45)	2.06 (0.71)
SwinIR	35.9 (2.55)	1.19 (0.35)	1.54 (0.49)
U-net	35.3 (2.95)	1.22 (0.36)	1.65 (0.50)
6.25% low-count PET	29.62 (1.80)	1.50 (0.45)	2.07 (0.62)
SwinIR	36.5 (2.68)	1.29 (0.37)	1.50 (0.49)
U-net	35.6 (2.19)	1.28 (0.35)	1.62 (0.40)
1% low-count PET	21.6 (1.60)	1.92 (0.51)	3.80 (0.89)
SwinIR	29.5 (2.16)	1.62 (0.64)	2.47 (1.19)
U-net	33.24 (2.13)	1.60 (0.38)	2.68 (0.79)

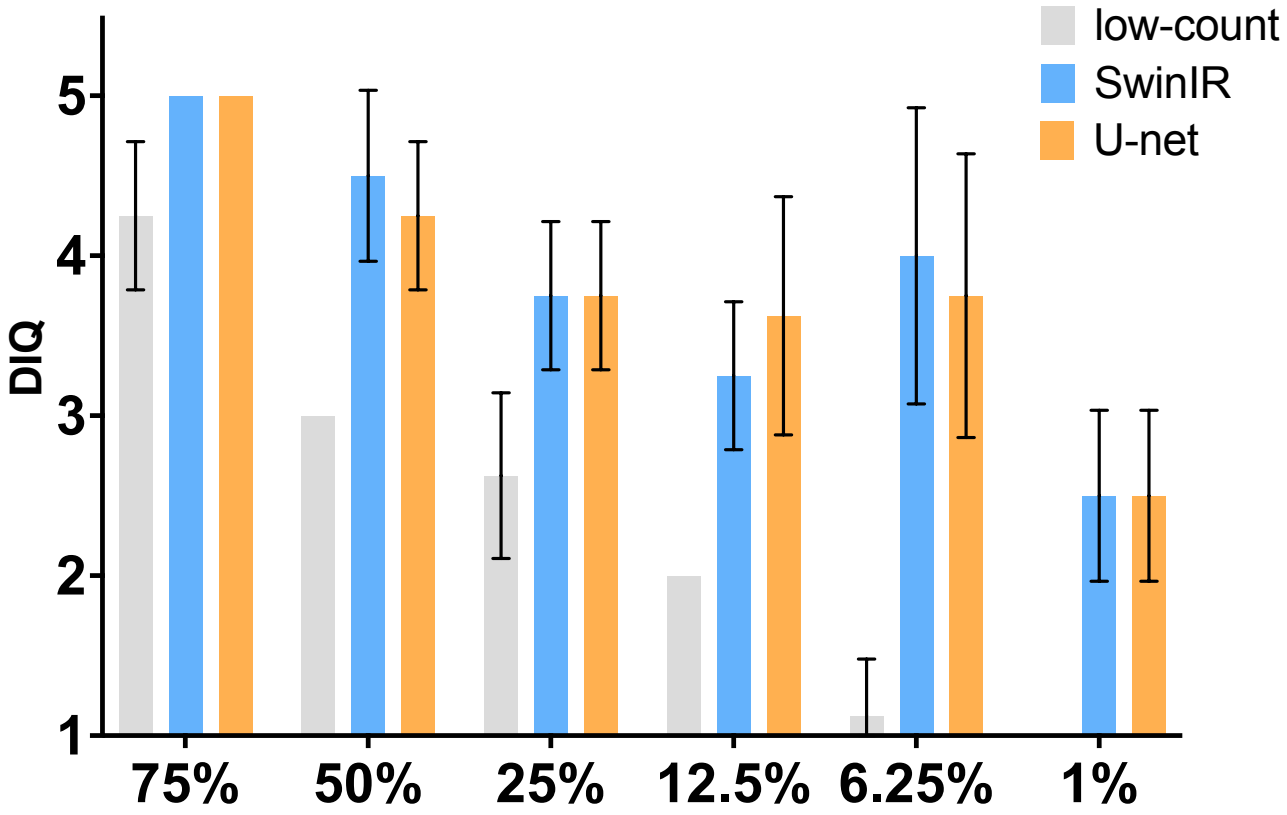
*SUV values were measured on the liver (3-cm ROI) using OsiriX software.



Supplementary Figure 1: An representative ^{18}F-FDG PET scan of a 14-year-old male patient with Hodgkin lymphoma (HL). The arrow points to the iv line, which is visiable across all reconstructions, but the contrast with the background is only preserved by SwinIR. SwinIR shows superiority in retaining contrast and structural fidelity.

Supplementary Figure 2: This figures shows an representative ^{18}F-FDG PET scan of a 11-year-old male patient with Hodgkin lymphoma (HL). The write arrow (on the MRI scan) points to the lesion of the left neck lymph node. AI-reconstructed PET images (second and third rows) show reduced noise and improved contrast between tumor and liver compared with non-AI-reconstructed original (first row) PET images. SwinIR outperformed U-net on 6.25% PET reconstruction with better image

denoising and more structural details preserved (SwinIR-6.25P and U-net-6.25P). U-net-75p = U-net reconstructed 75% low-count PET; SwinIR-75P = SwinIR reconstructed 75% low-count PET.



Supplementary Figure 3: Diagnostic image quality (DIQ) (mean ± standard deviation) for the original low-count-PET scans, and the AI-restored PET scans (U-net and SwinIR) across the entire dose reduction spectrum. Diagnostic image quality (DIQ) on a 5-point Likert scale (1 = nondiagnostic, 2 = poor, 3 = acceptable, 4 = good, 5 = excellent image quality).

References

1. Yushkevich PA, Gao Y, Gerig G. ITK-SNAP: An interactive tool for semi-automatic segmentation of multi-modality biomedical images. IEEE; 2016:3342-3345.
2. Loshchilov I, Hutter F. Decoupled weight decay regularization. *arXiv preprint arXiv:171105101*. 2017;
3. Dosovitskiy A, Beyer L, Kolesnikov A, et al. An image is worth 16x16 words: Transformers for image recognition at scale. *arXiv preprint arXiv:201011929*. 2020;
4. Hu W, Xiao L, Pennington J. Provable benefit of orthogonal initialization in optimizing deep linear networks. *arXiv preprint arXiv:200105992*. 2020;
5. He K, Zhang X, Ren S, Sun J. Delving deep into rectifiers: Surpassing human-level performance on imagenet classification. *Proceedings of the IEEE international conference on computer vision*. 2015:1026-1034.
6. Glorot X, Bengio Y. Understanding the difficulty of training deep feedforward neural networks. *Proceedings of the thirteenth international conference on artificial intelligence and statistics*. 2010:249-256.