©RSNA, 2023
10.1148/ryai.220246

# Appendix S1

## Patients and Image Acquisition

Patients in the primary cohort underwent a whole body integrated 18F-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 140 mg/dL. 18F-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 anatomic coregistration. PET data were acquired simultaneously with MRI scans, using a 25 cm transaxial FOV and 3:30 minute acquisitions per PET bed. PET data were 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.

Patients in the external test cohort underwent a whole body integrated 18F-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 140 mg/dL. 18F-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 anatomic coregistration. PET data were acquired simultaneously with MRI scans, using a 25 cm transaxial FOV and 4 minutes acquisitions per PET bed. PET data were 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 18F-FDG PET images. Low dose PET images were simulated by unlisting the PET list-mode data and reconstructing them based on the percentage of used counts. For the primary cohort, the list-mode PET input data collected over a time period of 3:30 and 2 seconds were used to simulate 100% and 1% 18F-FDG dose levels. For data acquired at the external site, PET Acquisition time was 4 minutes per bed position and low-dose PET images were simulated using the same relative dose levels accordingly.

## Image-preprocessing

The preprocessing pipeline aimed to remove the additional burden of the network learning methods to find patterns between scans for final reconstruction. As opposed to traditional single time scan analyses, registration is essential for longitudinal studies to reduce the spatial discrepancies between scans acquired at different times for the same individual. Across all subjects, the follow-up scans were registered to the baseline MRI as the template using affine

transformation. We adopted ANTs, which is considered a state-of-the-art medical image registration toolkit. This ensured all of the scans were registered within each patient's history. In addition, we used ITK-Snap to label the tumor regions in the baseline scan. The top five most prominent tumors (largest lesions) for each individual were delineated with ellipsoid-shaped masks. These tumor masks were used to mask out the tumor area of the baseline PET images to avoid introducing erroneous upstaging signals for the follow-up reconstruction. Top 0.1% of the pixels in PET images were clipped. Of note, the clipping operation is critical in model convergence and stabilizing training as these top pixels possess extreme high values and are outliers of the distribution. Lastly, all scans were normalized between zero and one before feeding them to the deep learning model.

## Proposed False Focal Loss

The loss function is a cornerstone of the neural network model and determines the optimizing direction for model training. We applied the commonly used loss functions for the image restoration task, mean square error (MSE) loss and structural similarity index measure (SSIM) loss, and additionally designed the false focal loss (FF loss) for this study. FF loss is specifically proposed for 1% extremely ultra-low-dose PET reconstruction, as 1% PET images harbor substantial noise, making the model prone to induction of false positive errors. Erroneous upstaging on interim scans would lead to intensified treatment and potential side effects in the absence of viable tumor. FF loss alleviates the impact of these false upstaging focal areas in the output PET images by penalizing incorrect hyperintense pixels during each step of gradient descent in the training process. We formulate the FF loss as below:

$$L_{FF} = E_{(I_{recon}, I_{true})}((I_{recon} - I_{true}) \times 1_{A_\tau}),$$

where $I_{recon}$ and $I_{true}$ refer to the reconstructed output and the ground truth PET, respectively. The indicator function of $A_\tau$, where $A_\tau$ denotes the subset of pixels satisfying $(I_{recon} - I_{true}) > \tau$, is defined as:

$$1_{A_\tau}(x) := \begin{cases} 1, & if\ x \in A_\tau \\ 0, & if\ x \notin A_\tau \end{cases}.$$

Coupled with MSE loss and SSIM loss, the composite loss function (as below) for optimizing *Masked-LMCTrans* encourages the PET reconstruction process to reduce noise, keep textures, and preserve structural details.

$$min L = \lambda_m L_{MSE} + \lambda_s L_{SSIM} + \lambda_f L_{FF}.$$

## Data Augmentation

Model performance improved with increasing training data sample size. We used random rotation, random shifting, and random zoom for data augmentation. During each step of stochastic gradient descent in the training process, we perturbed each training sample (both baseline and follow-up PET/MRI images; four combined inputs) with a random rotation between-20 to 20 degrees and with a random shift between-20%–20% across x and y axis, and with a random zoom between 0.8 to 1.2. Data augmentation resulted in improvement for all

models; around 1% improvement in SSIM metric for Masked-LMCTrans and slight improvement for U-net.

## Training Details

Following ResNet (29) and ViT (30) (vision transformer), we used a learning rate warmup for 5 epochs and then linearly decay the learning rate over the course of training. We trained the models with Adam (31) optimizer, using $\beta_1 = 0.9$ and $\beta_2 = 0.99$. We adopted three-fold cross-validation for the primary cohort. Each fold has 23 paired baseline and follow-up PET/MRI scans (from 14 patients) for training, 8 paired scans for testing, and 3 paired scans for validation (from 7 patients). The training time using four NVIDIA GeForce RTX 3090 GPUs with 24GB VRAM was about 12 hours, and the reference time for each subject was only 10 seconds.

## Lessons from Model Training and Experiments

We examined the difference of using slices from the axial plane or coronal plane and found that axial demonstrates superior performance. More details are provided in fig. S2. For the training scheme, we tried out multitask learning with segmentation added besides the objective reconstruction, but did not notice improvement.

## References


29. He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition 2016; 770–778.

30. Dosovitskiy A, Beyer L, Kolesnikov A, et al. An image is worth 16x16 words: transformers for image recognition at scale. arXiv preprint arXiv:2010.11929. https://arxiv.org/abs/2010.11929. Posted October 22, 2020. Accessed May 2022.

31. Kingma DP, Ba J. Adam: A Method for Stochastic Optimization. arXiv preprint arXiv:1412.6980. https://arxiv.org/abs/1412.6980. Posted December 22, 2014. Accessed May 2022.


### Table S1

### Evaluation Results on 1% Extremely Ultra-low-dose PET Reconstruction-External Test Cohort

| | 1% Ultra-low-dose PET (*n* = 10) | Masked-LMCTrans PET (*n* = 10) | *P* Value |
|---|---|---|---|
| SSIM | | | |
| Mean (SD) | 0.747 (0.045) | 0.899 (0.028) | <0.001 |
| Median (Q1, Q3) | 0.764 (0.699, 0.779) | 0.900 (0.872, 0.920) | |
| PSNR | | | |
| Mean (SD) | 27.4 (0.99) | 34.4 (1.61) | <0.001 |
| Median (Q1, Q3) | 29.2 (28.7, 30.3) | 35.0 (34.0, 35.6) | |
| VIF | | | |
| Mean (SD) | 0.112 (0.010) | 0.254 (0.026) | <0.001 |
| Median (Q1, Q3) | 0.113 (0.106, 0.116) | 0.257 (0.234, 0.270) | |

Note.—The evaluation on the primary Stanford cohort; All comparisons are to the ground truth standard-count PET images; *P* values are calculated using Wilcoxon signed-rank test between the AI-reconstructed PET and the low-count PET. SSIM = structural similarity index; PSNR = signal-to-noise ratio; VIF = visual information fidelity.