Supplementary

## Appendix 1

### *The neural network architecture and training procedure*

The architecture of the HYPER DLR neural network is derived from the U-Net-based architecture incorporated with residual network (ResNet) blocks and the dense convolutional network (DenseNet) connection techniques (9). *Figure S1* illustrates the network structure. In this network, a long residual path is used to connect the input to the output, which allows for the network to learn the image noise component between the target image and the input image, accelerating the convergence of the deep network. Similar to U-Net, the neural network blocks at higher levels have a larger matrix size and hence a higher spatial resolution. In contrast, the DenseNet connections are included in each level to decrease the information loss. Furthermore, the ResNet blocks are used at the majority of the nodes to avoid vanishing gradient problem and further improving the performance of the deep neural networks.

To train the neural network, 313 positron emission tomography (PET) studies were used, and an additional 80 studies were used for validation. The data were gathered from four sites equipped with PET/computed tomography (PET/CT) scanners manufactured by United Imaging Healthcare. The age of the patients ranged from 18 to 90 years, with a median age of 55 years. The injected dose for the scans varied between 3.1 to 4.3 MBq/kg, and the acquisition time was 90 to 180 seconds per bed for the body torso.

During training, the network used retrospective PET reconstructions with 50% acquisition time as input, resulting in

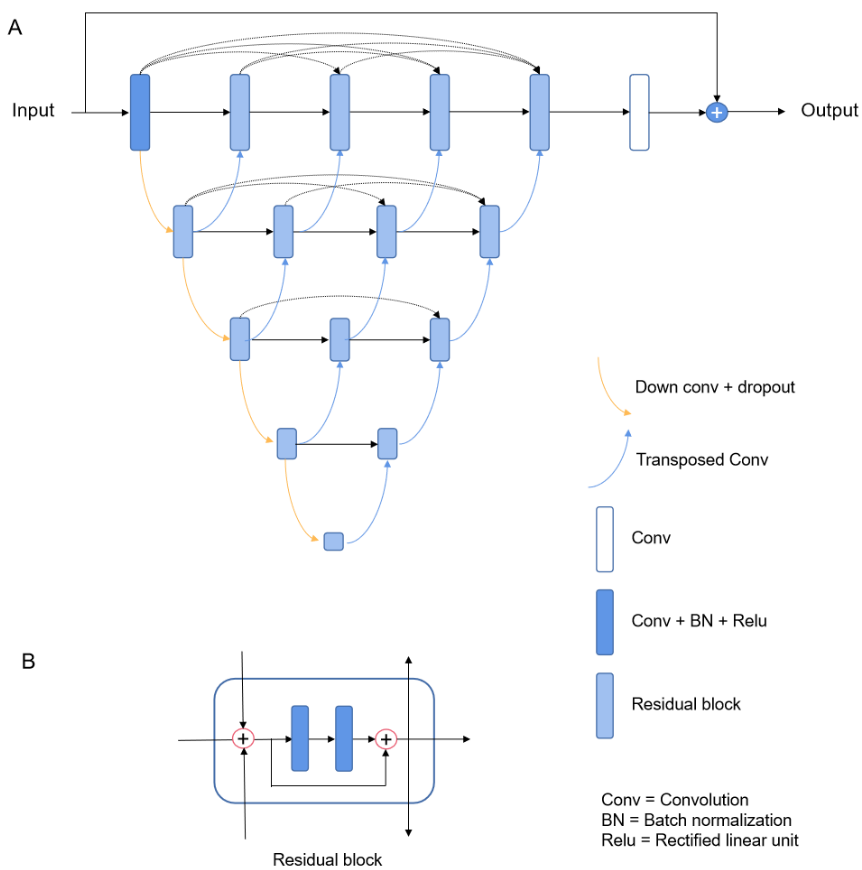



**Figure S1** The diagram of HYPER DLR. (A) The neural network is constructed based on a U-Net architecture with ResNet blocks and DenseNet skip connections. (B) The diagram of the ResNet block used in the neural network. This figure was adapted from Xing *et al.* (9) under the Creative Commons Attribution 4.0 International (CC BY) License. ResNet, residual network; DenseNet, dense convolutional network.

 https://dx.doi.org/10.21037/qims-23-817

high-noise PET images. The training targets were retrospective reconstructions with a full acquisition time, yielding low-noise PET images. All images underwent reconstruction using the ordered subset expectation maximization (OSEM) algorithm with time of flight (TOF) and point spread function (PSF) kernels, which is the standard-of-the-art reconstruction algorithm configuration. The parameters for iteration number and subsets were 2 and 20, respectively. The voxel size of the reconstructed images was $2.34 \times 2.34 \times 2.68\ mm^3$. To enhance the network's robustness and reduce overfitting, data augmentation techniques such as horizontal and vertical flips were applied. All training images were resampled to ensure uniform matrix size and spatial resolution. The image intensities were normalized within the range of [0, 1].

The network used a 2.5D processing approach, in which five slices of input images corresponded to one slice of the target image. Specifically, image patches with a dimension of 64×64×5 were used as the training input, while patches with a dimension of 64×64×1 served as the training target.

During the training process, the L1 loss function was adopted, and the model was trained using the adaptive moment estimation (ADAM) optimizer. The initial learning rate was set to $1 \times 10^{-4}$ and was halved after 20 epochs. The network was trained with a batch size of 32 for over a total of 200 epochs.

The implementation of the network was carried out using the PyTorch framework and Python 3.7. For testing, a computer with a single Quadra P5000 GPU was used with CUDA library version 8.0 and cuDNN version 7.0.5.

***The percent limits of agreement of the quantitative measurement for the lesions (Figure S2)***

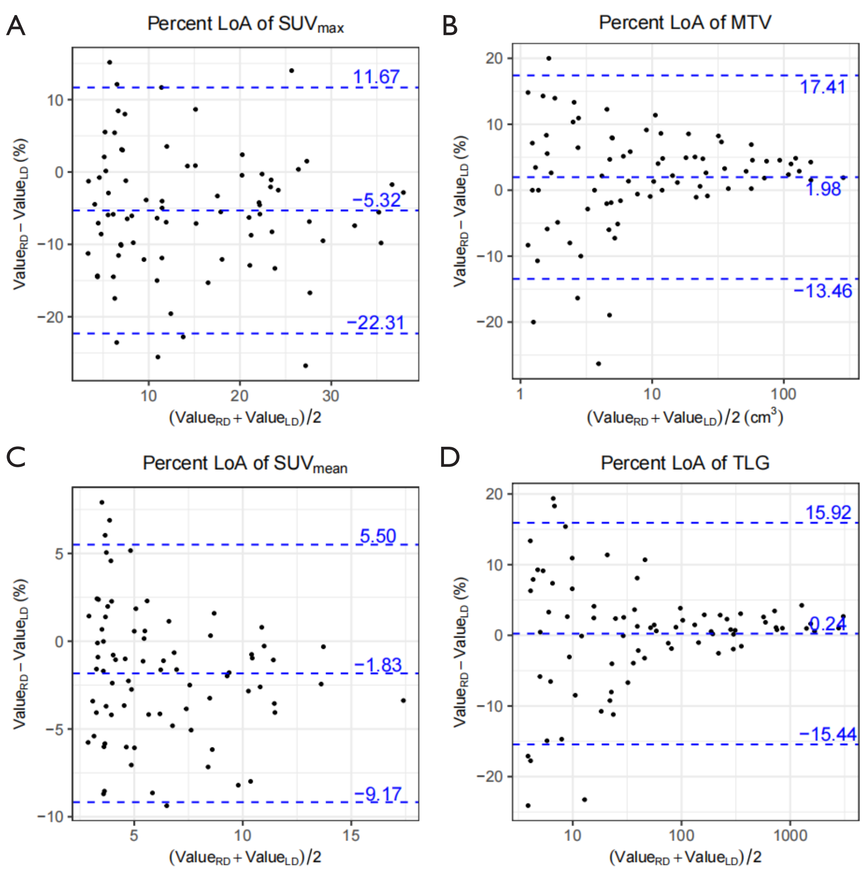
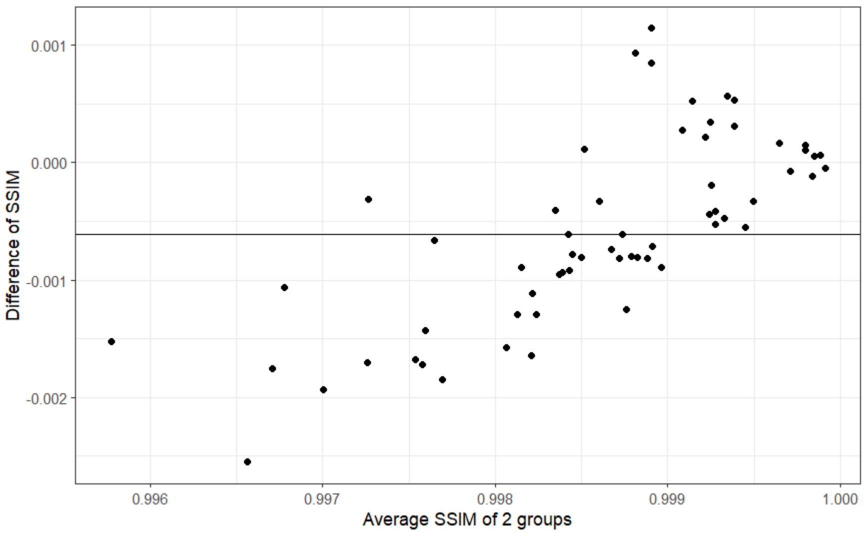



**Figure S2** The percent LoA for the lesion $SUV_{max}$, $SUV_{mean}$, MTV, and TLG. The dotted lines are the mean of the difference and its 95% CI. The values for the mean and 95% CI are noted next the lines. The x-axis of c and d is plotted in the logarithmic scale for better data visualization. LoA, limits of agreement; $SUV_{max}$, maximum standardized uptake value; RD, routine-dose; LD, low-dose; $SUV_{mean}$, mean standardized uptake value; MTV, metabolic tumor volume; TLG, tumor lesion glycolysis; CI, confidence interval.

 https://dx.doi.org/10.21037/qims-23-817

*Structural similarity index measure (SSIM) and peak signal to noise ratio (PSNR) (Figure S3)*



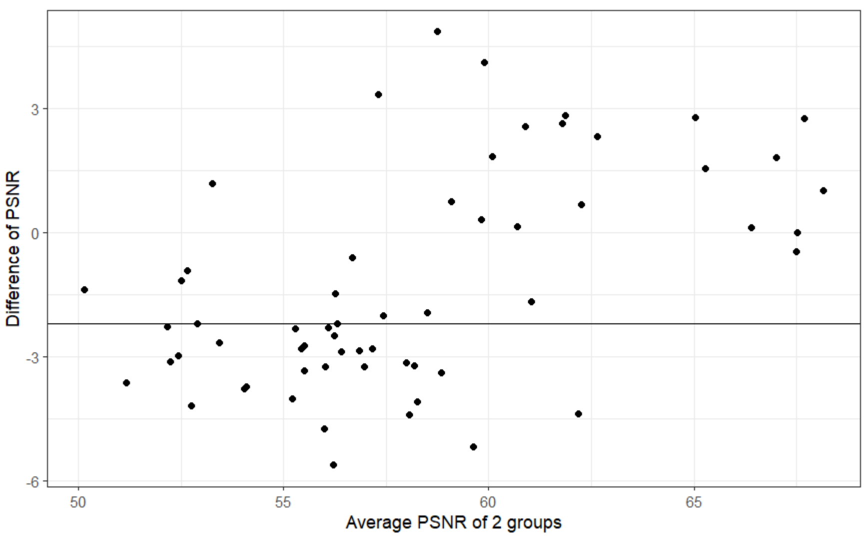




**Figure S3** The comparison of SSIM and PSNR between LD images derived from the DL denoising technique and RD images without DL (standard reconstruction). The y-axis and x-axis are the difference and mean between LD image group with DL and the RD image group without DL, respectively. The horizontal solid lines represent the median of the difference. Both lines are below zero, which indicate that the value of the SSIM and PSNR was slightly lower in the LD images with DL compared with the RD images without DL when RD images with DL were used as the reference. SSIM, structural similarity index measure; PSNR, peak signal to noise ratio; LD, low-dose; DL, deep learning; RD, routine-dose.

 https://dx.doi.org/10.21037/qims-23-817

*The results of the preliminary study*

We conducted a preliminary study before starting the final study presented in the paper. The aims of the preliminary study were to (I) to help inform the standards for image quality scores and (II) to determine a duration for the reconstructions; that is, the level of the dose reduction. In the preliminary study, five patients with lymphoma undergoing fluorine-18 fluorodeoxyglucose ([$^{18}$F]FDG) PET/CT were enrolled. The data of these patients were not used in the final study. To reach agreement on the standard of image quality score, three nuclear radiologists reviewed the PET images and scored the image quality using a 5-point scale in a joint session. The radiologists were the same readers who scored the images in the final study.

The PET images were reconstructed using six protocols. The standard-of-care protocol in our department was the OSEM reconstruction with 2 minutes (min) list-mode data (OSEM 2 min). Furthermore, the deep learning (DL) denoising neural network was applied to five protocols with 0.5-, 0.75-, 1.0-, 1.25-, and 1.5-min duration data (DL 0.5 min, DL 0.75 min, DL 1 min, DL 1.25 min, and DL 1.5 min, respectively).

The visual image quality scores are shown in the *Figure S4*, which shows that the median of the visual image quality score for D-processed images decreased when the duration decreased from 1 to 0.5 min, while it became flat at durations of 1, 1.25, and 1.5 min. We concluded that a duration of 1 min with DL reconstructions can obtain superior image quality compared with OSEM reconstructions. Although the median score of 0.75 min was on par with OSEM 2 min, DL 1 min was selected due to preferred image quality (*Figure S5*) and substantially reduced duration.

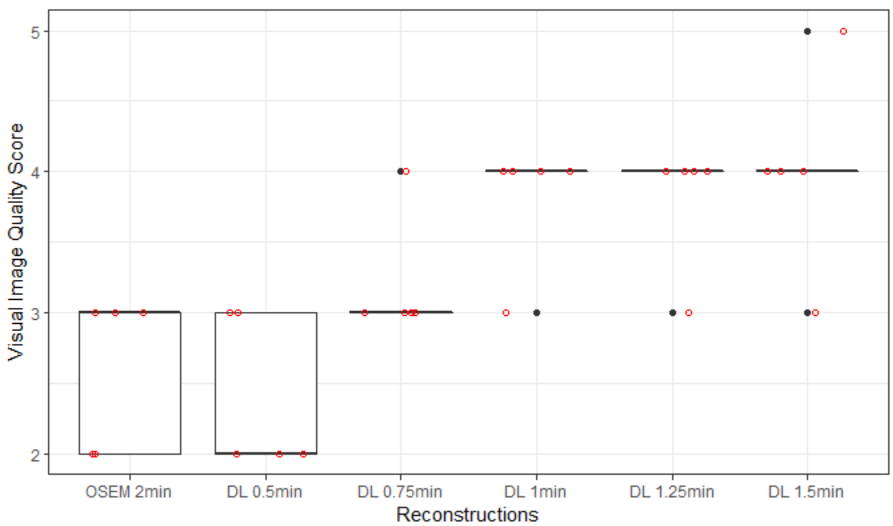



**Figure S4** The box plot of visual image quality scores in the preliminary study. The red circles are the image quality scores. The dots are the outliers of the box plot.

 https://dx.doi.org/10.21037/qims-23-817

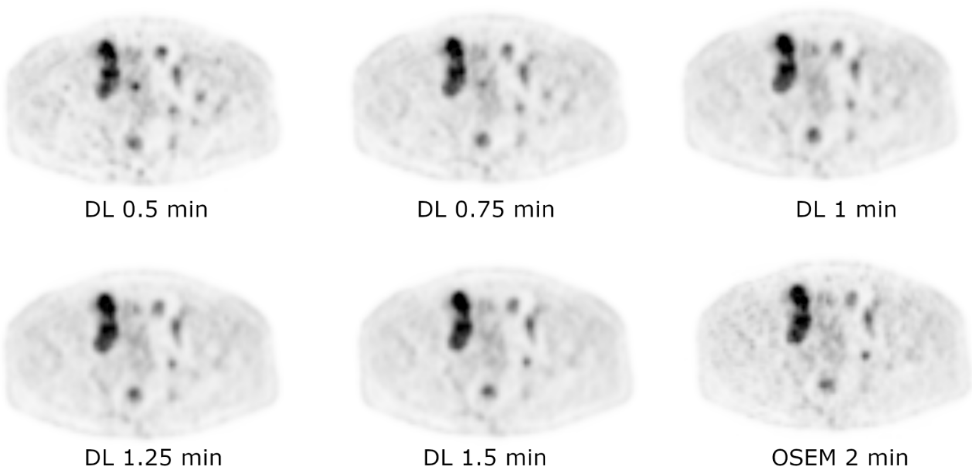



**Figure S5** Illustration of the DL denoising technique and the reconstruction duration. This preliminary result indicated that a duration of 1 min with the deep learning denoising technique is appropriate to acquire the correct diagnosis and avoid false-positive findings (arrowheads). DL, deep learning; OSEM, ordered subset expectation maximization.

 https://dx.doi.org/10.21037/qims-23-817