The network architecture adopted in this paper is shown in Figure S-1.

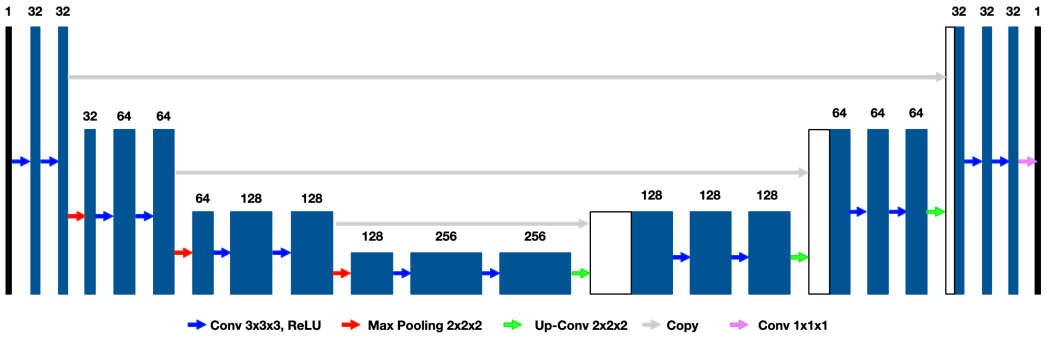

Figure S-1: 3D U-Net architecture. It includes three parts of layers: contracting path, bottleneck and expanding path. In the contracting path, each contracting layer consists of two convolution operations, each of which is a 3×3×3 convolution followed by a rectified linear unit (ReLU) and a 2×2×2 max pooling operation. The bottleneck includes two convolution operations. In the expanding path, each expanding layer includes a 2×2×2 up-convolution operation and two convolution operations. The feature maps extracted from the expanding path were concatenated with copied feature maps from the contracting path. The feature number in the first layer is 32.

Many works have shown that deep learning denising methods outperformed conventional denoising methods. Here we compared with three conventional denoising methods, (1) Gaussian filtering, (2) non-local-means. (3) BM4D. For the Gaussian filter, 1mm, 2mm, 3mm, 4mm and 5mm 3D filter kernels were used for the on the OSEM+5mm Gaussian post-filtering images.

Figure S-2 shows the background noise and lesion SUVmean bias curves of 4 subjects in the dynamic datasets. The SUVmean bias is defined as $\frac{SUV^E_{mean} - SUV^L_{mean}}{SUV^L_{mean}} \times 100\%$, where SUV^E_{mean} is the SUVmean value of the lesion in denoised image, SUV^E_{mean} is the SUVmean value of the lesion in true-high-count image. The background ROI were manually drawn close to the lesions using ITK-SNAP tool.

Figure S-3 and Figure S-4 plot the denoised images using deep learning, Gaussian filtering, NLM, and BM4D methods. It is clearly shown that the deep learning method outperform other conventional methods, with better reduced noise and well-preserved structures. Gaussian filtering, NLM, and BM4D images suffer from details loss when suppressing the noise, as indicated by the white arrows. In addition, quantitative evaluation was performed on the two data in Figure S-3 for different denoised images with the true-high-count image as the reference. The quantitative evaluation results showed that the deep learning method achieves the best performance. We didn't perform quantitative evaluation on the all multi-pass dynamic datasets. The reason is that it is difficult to tune the denoising parameters in Gaussian filtering, NLM, and BM4D methods to tradeoff the denoising performance and detail loss. In addition, it cannot achieve optimal denoising performance when applying the

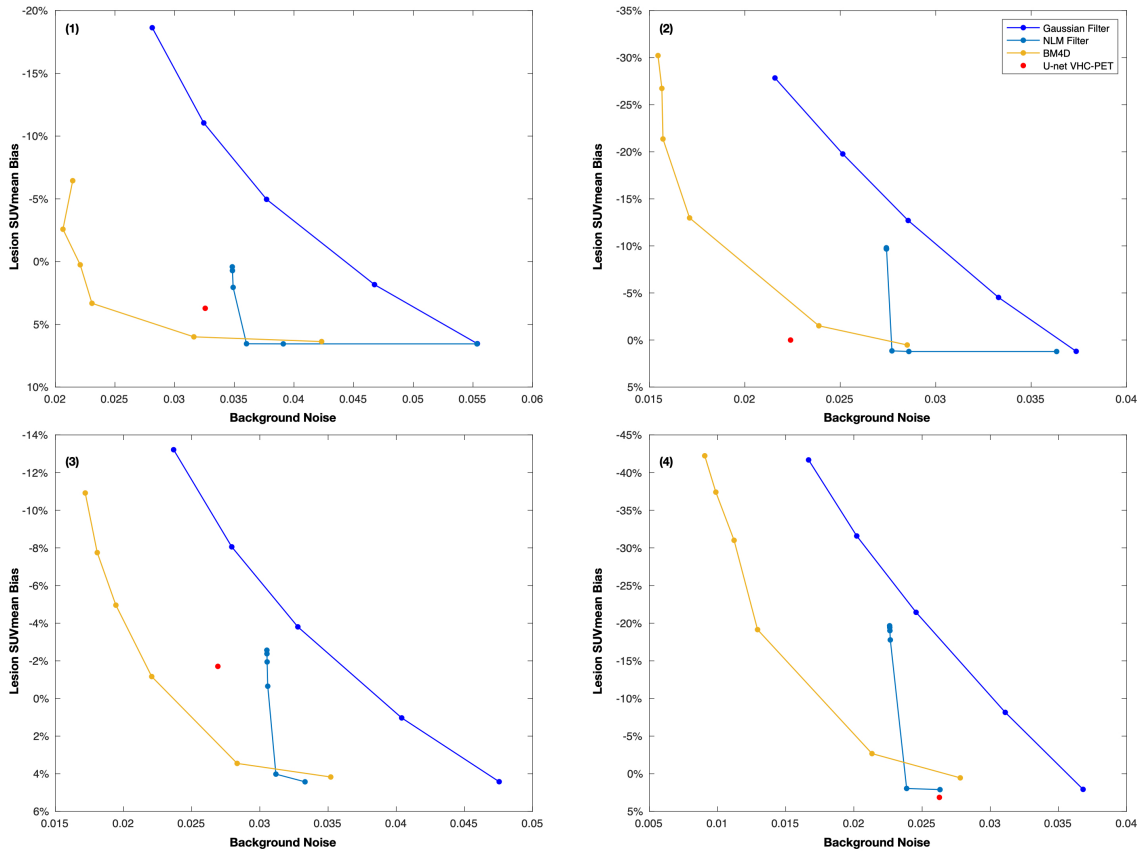

Figure S-2: Noise-bias curve of 4 patients with various noise reduction methods.

same parameters across the whole datasets, which have different noise levels.

In the network training, we used the combination of 20%, 30%, and 40% downsampled data of 27 dynamic data. This can not only augment the training sets but also make the noise distribution of the combination downsampled data better match the noise distribution of 195 clinical data. Here we compared 4 networks, (1) only using 20% downsampled data as training data, (2) only using 30% downsampled data as training data, (3) only using 40% downsampled data as training data, (4) using all 20%, 30%, and 40% downsampled data as training data. The training settings were the same except the training data. After training, the trained models applied on the clinical data to get the virtual high-count images.

Figure S-5, Figure S-6, and Figure S-7 show the different virtual-high-count images from the above four networks. The virtual-high-count images using 20% downsampled data for training showed substantial blurring, which indicates that the unmatched noise could cause over-smoothing when training with higher noise level than the test data with less noisy. Using 30% and 40% downsampled data for training could better preserve the image structures at the cost of slightly remaining noise, which could be observed in Figure S-5 and Figure S-6. The virtual-high-count image using the combined downsampled data have not only reduced

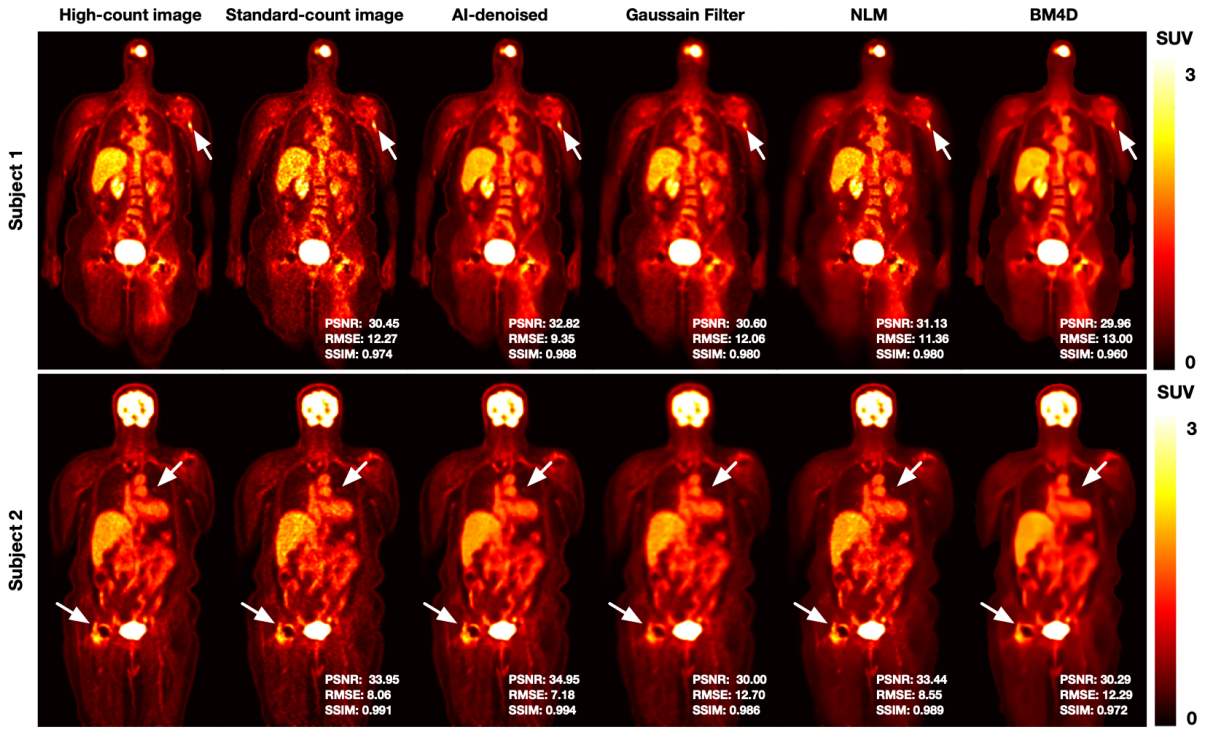

Figure S-3: Visual comparison of different denoising methods in two dynamic data. The white arrows indicates that the AI-denoised images could substantially suppress the noise and preserve important structures, while the Gaussian filtering, NLM, and BM4D reduce the noise but at the cost of blurring. The quantitative metrics also indicates that the AI-denoised images have the best quantitative metric scores.

noise but also well-preserved details.

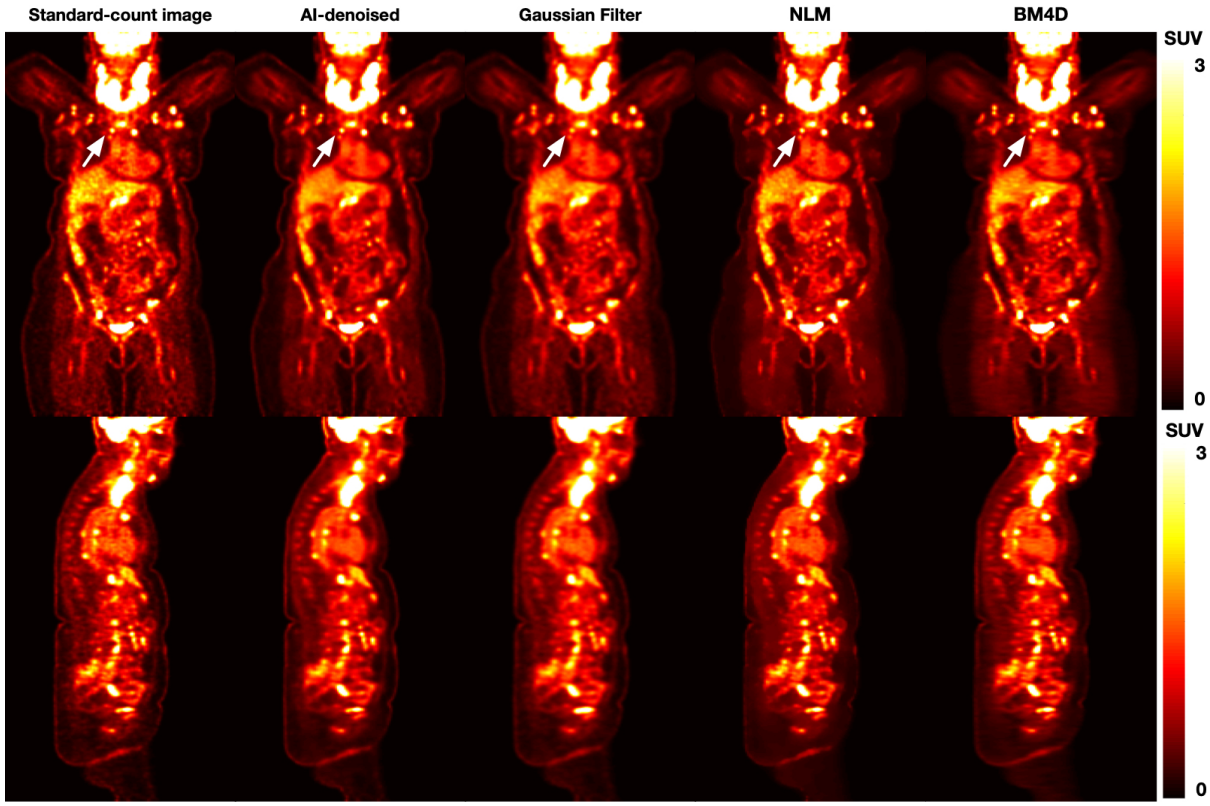

Figure S-4: Visual comparison of different denoising methods in one clinical data. The AI-denoised images could substantially suppress the noise and preserve important structures (white arrows). The Gaussian filtering, NLM and BM4D reduce the noise but at the cost of blurring.

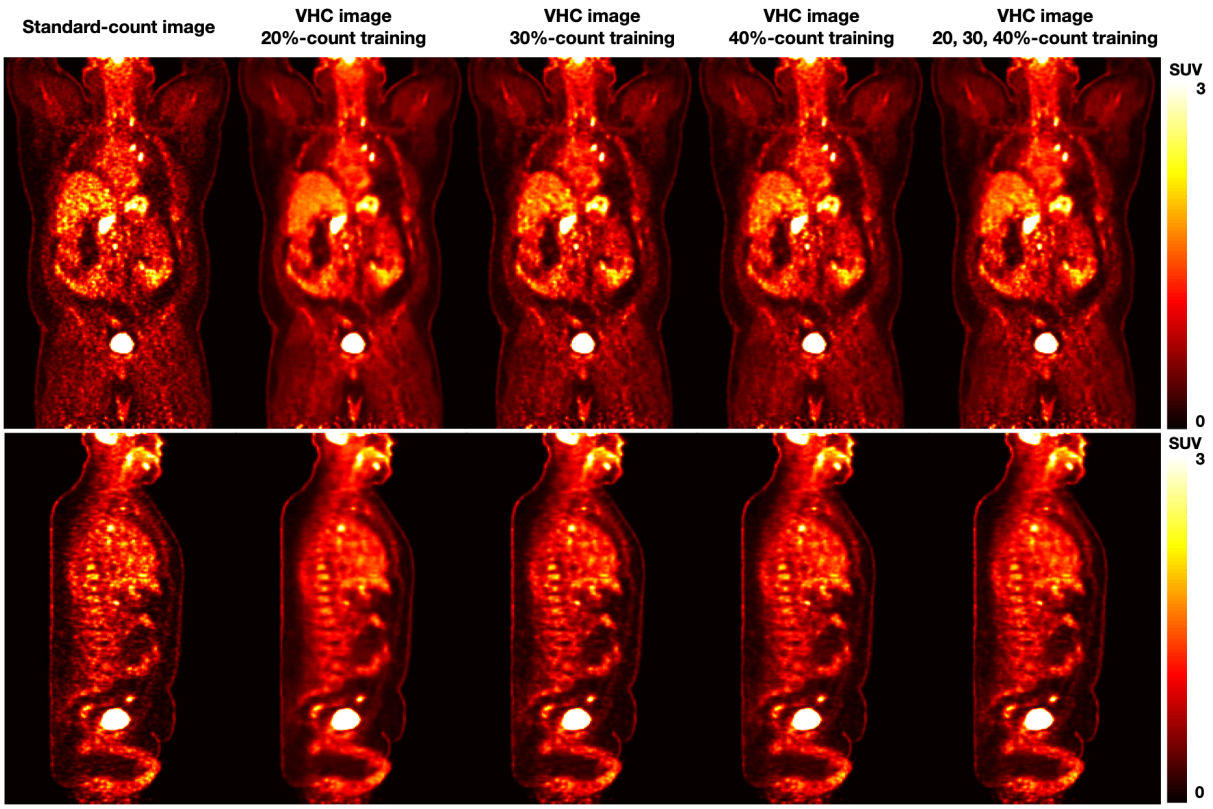

Figure S-5: Visual comparison of the standard-count and virtual-high-count using different training data. All virtual-high-count images suppressed the noise and showed less noisy than the standard-count images. The virtual-high-count images using 20% downsampled data for training showed substantial blurring, while using 30% and 40% downsampled data for training could better preserve the image structures at the cost of remaining noise. The virtual-high-count image using the combined downsampled data showed reduced noise and well-preserved details. (Patient: male, BMI 48.1 $kg \cdot m^2$).

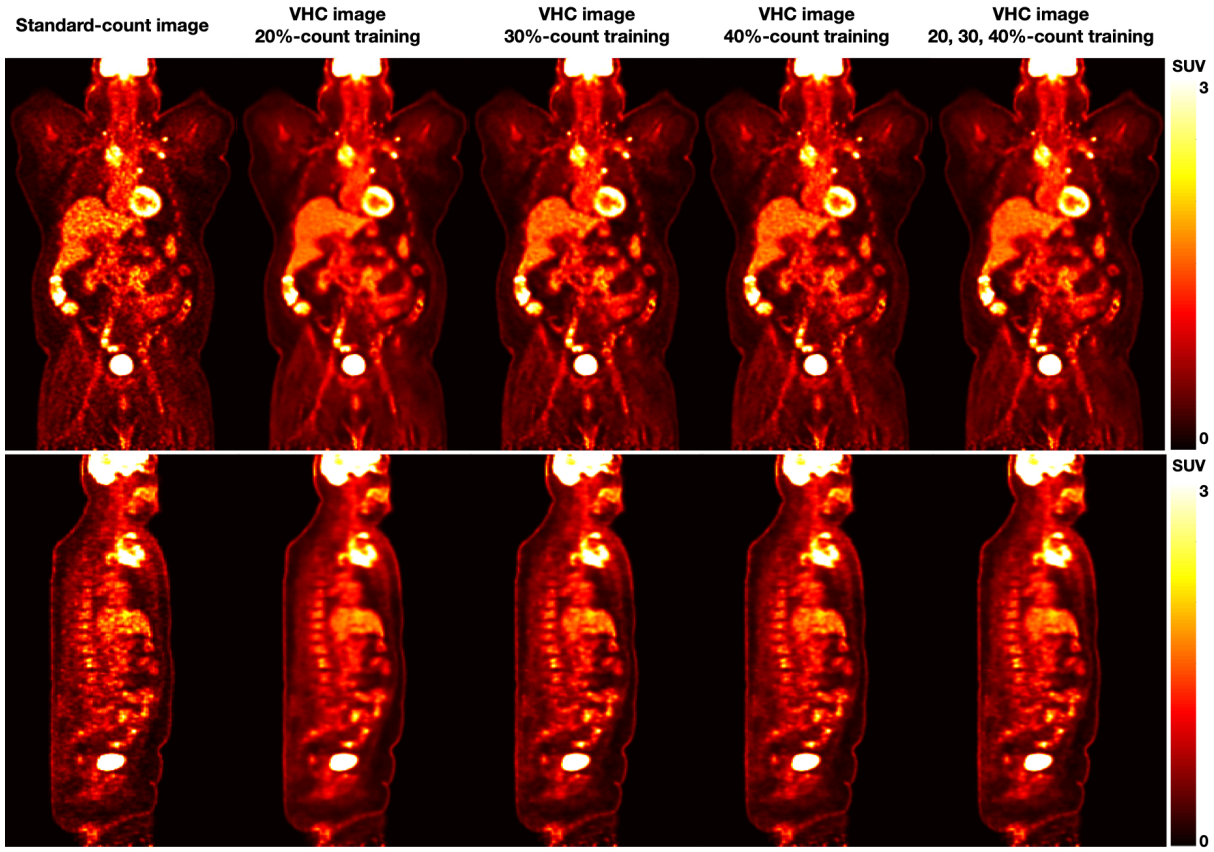

Figure S-6: Visual comparison of the standard-count and virtual-high-count using different training data. All virtual-high-count images suppressed the noise and showed less noisy than the standard-count images. The virtual-high-count images using 20% downsampled data for training showed substantial blurring, while using 30% and 40% downsampled data for training could better preserve the image structures at the cost of remaining noise. The virtual-high-count image using the combined downsampled data showed reduced noise and well-preserved details. (Patient : male, BMI 37.9 $kg \cdot m^2$).

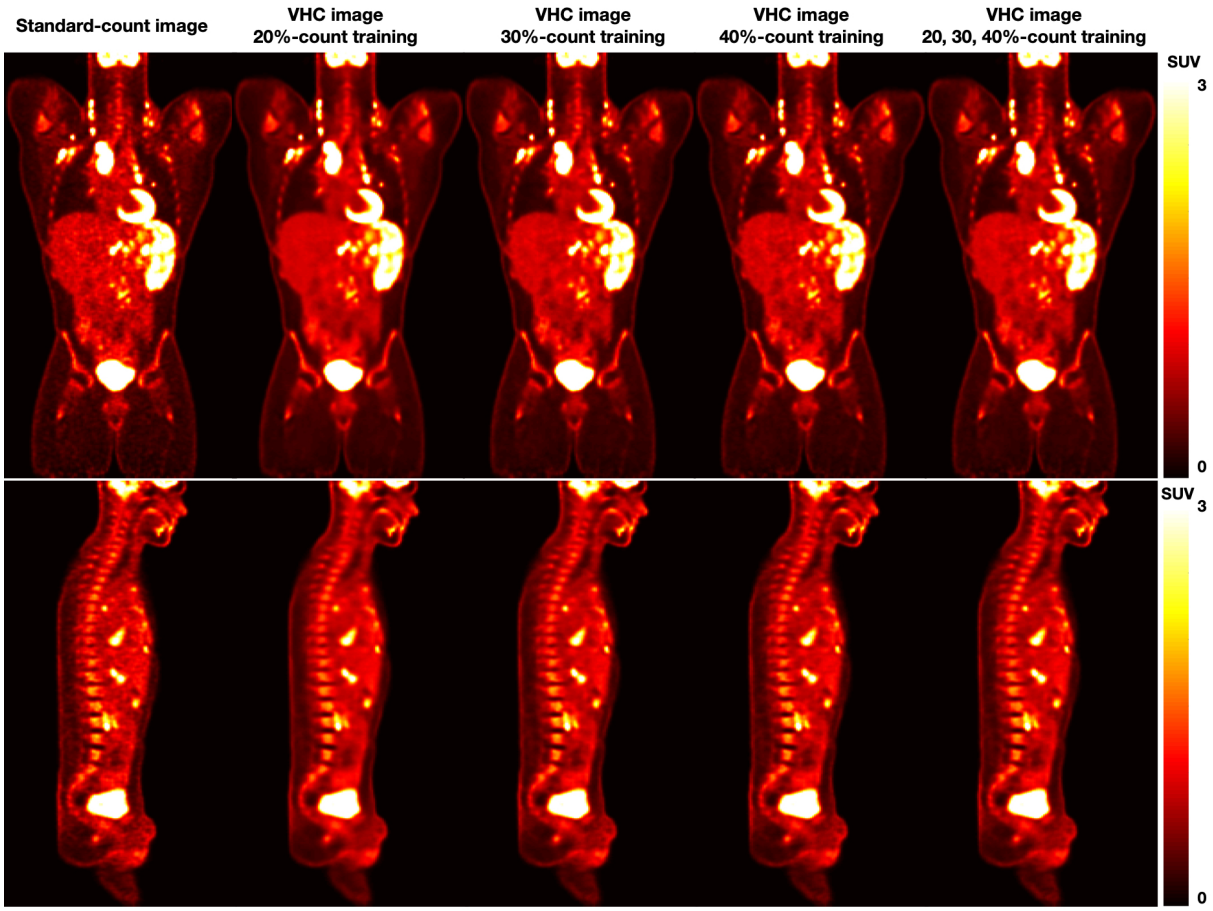

Figure S-7: Visual comparison of the standard-count and virtual-high-count using different training data. All virtual-high-count images suppressed the noise and showed less noisy than the standard-count images. The virtual-high-count images using 20% downsampled data for training showed substantial blurring, while using 30% and 40% downsampled data for training could better preserve the image structures at the cost of remaining noise. The virtual-high-count image using the combined downsampled data showed reduced noise and well-preserved details. (Patient 3: male, BMI 26.7 $kg \cdot m^2$).