Annals of Nuclear Medicine

AUTHOR(S)
Jafaritadi, M., Teuho, J., Lehtonen, E., Klén, R., Saraste, A. & Levin, C. S.

TITLE
Deep generative denoising networks enhance quality and accuracy of gated cardiac PET data.

YEAR
2024

DOI
10.1007/s12149-024-01945-1

CITATION
Jafaritadi, M., Teuho, J., Lehtonen, E., Klén, R., Saraste, A., & Levin, C. S. (2024). Deep generative denoising networks enhance quality and accuracy of gated cardiac PET data. *Annals of Nuclear Medicine, 38*(10), 775–788. DOI: https://doi.org/10.1007/s12149-024-01945-1

VERSION
Accepted Manuscript

Deep Generative Denoising Networks Enhance Quality and Accuracy of Gated Cardiac PET Data

Mojtaba Jafaritadi[1], Jarmo Teuho[2,3], Eero Lehtonen[2],
Riku Klen[2,3], Antti Saraste[2,3,4†], Craig S. Levin[1,5,6,7*†]

[1]Department of Radiology, Stanford University, CA, USA.
[2]Turku PET Center, University of Turku, Turku, Finland.
[3]Turku PET Center, Turku University Hospital, Turku, Finland.
[4]Heart Center, Turku University Hospital, Turku, Finland.
[5]Department of Physics, Stanford University, CA, USA.
[6]Department of Electrical Engineering, Stanford University, CA, USA.
[7]Department of Bioengineering, Stanford University, CA, USA.

*Corresponding author(s). E-mail(s): cslevin@stanford.edu;
Contributing authors: mojjaf@stanford.edu; jatateu@utu.fi;
eero.lennart.lehtonen@utu.fi; riku.klen@utu.fi; antti.saraste@utu.fi;
†Original Article

Funding: This research was partially supported by the Stanford Cancer Institute and Finnish Cultural Foundation Postdoctoral Research Fellowship (grant number: 00220062).

Abstract

Background: Cardiac positron emission tomography (PET) can visualize and quantify the molecular and physiological pathways of cardiac function. However, cardiac and respiratory motion can introduce blurring that reduces PET image quality and quantitative accuracy. Dual cardiac- and respiratory-gated PET reconstruction can mitigate motion artifacts but increases noise as only a subset of data is used for each time frame of the cardiac cycle.

Aim: The objective of this study is to create a zero-shot image denoising framework using a conditional generative adversarial networks (cGANs) for improving image quality and quantitative accuracy in non-gated and dual-gated cardiac PET images.

Methods: Our study included retrospective list mode data from 40 patients who underwent an ^{18}F-fluorodeoxyglucose (^{18}F-FDG) cardiac PET study. We initially trained and evaluated a 3D cGAN – known as Pix2Pix – on simulated non-gated low-count PET data paired with corresponding full-count target data, and then deployed the model on an unseen test set acquired on the same PET/CT system including both non-gated and dual-gated PET data.
Results: Quantitative analysis demonstrated that the 3D Pix2Pix network architecture achieved significantly (p-value$<$0.05) enhanced image quality and accuracy in both non-gated and gated cardiac PET images. At 5%, 10%, and 15% preserved count statistics, the model increased peak signal to noise ratio (PSNR) by 33.7%, 21.2%, and 15.5%, structural similarity index (SSIM) by 7.1%, 3.3%, and 2.2%, and reduced mean absolute error (MAE) by 61.4%, 54.3%, and 49.7%, respectively. When tested on dual-gated PET data, the model consistently reduced noise, irrespective of cardiac/respiratory motion phases, while maintaining image resolution and accuracy. Significant improvements were observed across all gates, including a 34.7% increase in PSNR, a 7.8% improvement in SSIM, and a 60.3% reduction in MAE.
Conclusion: The findings of this study indicate that dual-gated cardiac PET images, which often have post-reconstruction artifacts potentially affecting diagnostic performance, can be effectively improved using a generative pretrained denoising network.

Keywords: ^{18}F-FDG PET, Cardiac PET, Cardiac Gating, Conditional GAN, Deep Learning, Dual-Gated PET, Respiratory Gating

1 Introduction

Cardiac positron emission tomography (PET) allows *in vivo* quantification of myocardial blood flow and metabolism [1]. However, the presence of cardiac and respiratory motion can degrade PET image quality by causing blurring, loss of detail, reduced contrast, and inaccurate quantitative measurements [2, 3]. Major sources of motion artifacts in cardiac PET include diaphragmatic displacement, cardiac contraction during elevated heart rates like stress tests, and sudden voluntary patient movement during a scan [3].

Motion compensation techniques like cardiac-gated reconstruction [4] of the diastolic phase or respiratory gating [5, 6] have been applied, but can increase image noise per frame as less data is available per gate [7]. Dual cardiac-respiratory gating [8] can minimize both cardiac and respiratory motion and improve image contrast and resolution synergistically, though at the cost of further amplifying noise due to the smaller subset of data per frame.

Various post-reconstruction image denoising techniques such as Gaussian filtering, non-local mean filtering (NLM) [9, 10], anisotropic diffusion [11], and block-matching 3D (BM3D) [12, 13] have been previously introduced. More recently, deep learning (DL) methods have gained popularity in molecular imaging, showing superior performance in tasks such as PET image reconstruction, segmentation, and denoising [14–18]. Deep learning models, including convolutional neural networks (CNNs) have been applied to denoise non-gated low dose PET images effectively [16]. Unsupervised denoising methods, such as Deep Image Prior, Noise2Noise (N2N), and Noise2Void (N2V), have also been explored for gated PET data [18–24]. However, dual-gated cardiac PET noise reduction using DL is compounded by the lack of high-dose and high-resolution targets for training the supervised models, while non-gated data inherently suffer from motion blurring targets. On the other hand, denoising methods must preserve the motion-frozen qualities of dual-gating images, including reduced blurring, improved resolution, and enhanced structural contrast. This trade-off between motion suppression and low signal-to-noise ratios poses a major challenge for gated cardiac PET imaging [25].

To address these challenges, we propose the use of generative adversarial networks (GAN), specifically conditional GAN (cGAN) models [26], for denoising gated cardiac PET data. We explore the use of cGANs, inspired by the Pix2Pix architecture [26], for the post-reconstruction denoising of non-gated and dual-gated cardiac PET images. Previous work in the realm of cGAN-based cardiac PET/SPECT (single-photon emission computed tomography) denoising generally considered non-gated low-dose to high-dose image correction. For cardiac PET/SPECT specifically, cGAN models have been developed that can effectively reduce noise and improve image quality. Sun *et al* [27] proposed a standard Pix2Pix for myocardial perfusion SPECT denoising with low-dose scans ranging from 5-50% of the full dose scans. In a follow-up study, Sun *et* *al* [28] examined the capability of the cGAN to reduce noise in myocardial perfusion SPECT images, comparing its performance when implemented in both the projection domain before reconstruction and the image domain after reconstruction. Guo *et al* [29] trained a temporally and anatomically aware GAN on cardiac rubidium-82 PET scans to convert early-to-late image frames in dynamic cardiac PET motion correction.

While the majority of previous work on cGAN-based denoising has been in neurology and cancer [30–34], to our knowledge there have not been any studies using a cGAN model to enhance the quality and accuracy of gated cardiac PET data in image or projection domains. Therefore, in this study, we adopted a 3D Pix2Pix cGAN model on non-gated low dose cardiac PET data and subsequently, we applied the pre-trained generative denoising model to dual-gated cardiac PET data to assess its generalizability capabilities on gated reconstructions without requiring high-quality and motion-phase matched targets. To the best of our knowledge, this is the first study on denoising gated PET data using a generative pre-trained model using low-count non-gated data as input. The generated images from this model represent synthetically restored standard dose PET images, and our results demonstrate effective noise reduction while preserving image resolution and anatomical structure. The main new contributions of this work are summarized as follows:

- We formulated a 3D cGAN architecture to process noisy cardiac PET scans to synthesize higher-quality PET images with reduced noise while preserving anatomical details.
- We demonstrate the application of generative pre-trained deep learning models in cardiac PET image restoration specifically for dual-gated PET reconstructions.
- We achieved generalizability of the model using non-gated data as input and tested the model with gated data without needing additional data or longer acquisitions. This enabled high SNR gated cardiac PET imaging data by reducing noise across different motion phases in gated reconstructions.

- We applied extensive statistical analysis to quantify model performance and generalizability.

2 Materials and Methods

2.1 Retrospective Data

We retrospectively included patients who had cardiac PET and computed tomography (CT) for evaluation of 18F-fluorodeoxyglucose (^{18}F-FDG) metabolism in the myocardium. A total of 40 patients (36 males, 4 females) were included and used during model training and algorithm development (IRB no: ETMK 44/180/2012 at the University of Turku, Finland). Demographic information of the subjects is described in Table 1. Each subject gave informed consent.

2.2 PET/CT System and Data Acquisition

Subjects were scanned with a GE Discovery 690 PET/CT system at Turku PET Centre using a previously described protocol [18] for uptake in the myocardium. The performance evaluation of the scanner with technical details is reported in [35]. The PET system comprises 24 detector rings, each housing a block of detectors with a 9×6 array of 4.2 × 6.3 × 25 mm lutetium-based scintillation (LBS) crystals coupled to a single square position-sensitive photomultiplier tube (PMT) with 4 anodes. The transaxial and axial fields of view (FOV) are 700 mm and 157 mm, respectively. Coincidence timing and energy windows are set at 4.9 ns and 425-650 keV, respectively. The D690 is a fully 3D PET system with a 64-slice Lightspeed CT system [35]. During contrast-enhanced coronary CT angiography (CTA), the patients received 50-100 ml of contrast agent (3.5 ml/s), and 64 parallel slices were acquired simultaneously. In addition to the CT scan, a 3D PET scan of the heart was acquired in list-mode with electrocardiography (ECG) and respiratory gating 24 minutes after the completion of the CTA. CT-based attenuation correction (CTAC) and cinematic CT (CINE-CT) were acquired by using a low-dose CT with a tube voltage of 120 keV.

2.3 Low-dose None-Gated PET Data

We originally acquired all PET data in list-mode to allow flexible data rebinning to simulate scans with varied acquisition times. By extracting only a fraction of counts from the full acquisition, we reconstructed lower-statistic PET images emulating realistic low-dose PET scans with dose reduction factors (DRFs) of 20, 10, and 7. That is, the simulated low-dose PET images contain only 5%, 10%, and 15% of the counts in the original full dose scan. With a DRF of 20, the acquisition time for the non-gated cardiac PET scan was reduced to 1.2 minutes. For a DRF of 10, the acquisition time was 2.4 minutes. And finally, with a DRF of 7, the resulting acquisition time for the non-gated low-dose cardiac PET scan was 3.4 minutes. As the low-dose PET data originate from the same acquisition, they align precisely with the corresponding full-dose images. On average, our training dataset comprised 24 matched low-count and full-dose scan pairs per subject from the list-mode data. 3D PET volumes were reconstructed including non-gated low-count inputs, and full-count targets.

2.4 Dual Gated PET Data

For dual gating, we employed 5 amplitude-based respiratory bins and 5 ECG-gated bins. Respiratory gating was performed using the Real-time Position Monitor (RPM) (Varian Medical Systems, Palo Alto, CA, USA), which is equipped with an infrared camera to track an external marker block positioned on the patient's chest, quantifying the respiratory cycle amplitude. PET and CT respiratory gating equally divided amplitude from end-inspiration to end-expiration into 5 bins. The gating thresholds were set similarly to our previous studies [18, 36]. In short, the respiratory bins were divided equally by amplitude from end-inspiration to end-expiration. The gating thresholds were determined by equidistant sampling. The maximum threshold was defined as the mean plus one standard deviation of the amplitude maxima, whereas the minimum was determined from mean of the amplitude minima of the respiratory cycles. Only the cycles considered as "valid cycles" by the RPM system were used.

Cardiac gating divided the cardiac cycle into 5 non-equal time bins using the R-peak triggers extracted from list-mode data. Specifically, fixed time intervals relative to each R-peak were used to retrospectively assign gates spanning from end-systole to end-diastole phases across ventricular contraction and relaxation. Following the dual gating, we obtained 25 respiratory and cardiac phase-matched PET time-frames – in list mode format – per subject. The dual-gating procedure retained 5-15% of the original acquired counts per cardiac-respiratory phase bin on average. This level of count statistics aligns with the noise levels simulated in our non-gated low-count training data extracted from shorter acquisition times. Notably, while matching realistic noise, the actual gated PET images were not utilized to train the deep learning models. We aimed to test generalizability by only training the network using non-gated data, but testing with both non-gated and gated data.

2.5 Reconstruction of Non-gated PET Data

Static non-gated PET images were reconstructed using up to 24 minutes of acquisition time. PET images were reconstructed with three-dimensional ordered subsets expectation maximization (3D-OSEM) reconstruction, using 2 iterations and 24 subsets. The reconstruction matrix size, voxel size, and transaxial and axial FOV were $256 \times 256 \times 47$, $2.73 \times 2.73 \times 3.27$ mm^3, 350 mm and 157 mm, respectively. To reduce the adverse impact of noise on the PET images, a Gaussian post-filter of 6 mm full width half maximum (FWHM) was applied. All necessary corrections for image quantification such as detector normalization, dead-time, random scatter, and attenuation were applied. All image reconstructions were performed with software provided by GE Healthcare (Research Gating Tool, RGT).

2.6 Reconstruction of Dual-gated PET Data

Dual respiratory (5 bins) and cardiac (5 bins) gated PET images were reconstructed using the 3D-OSEM algorithm with 2 iterations and 24 subsets. In total, 25 sets of gated images each corresponding to a certain cardiac and respiratory phase were reconstructed. Attenuation correction was performed using a respiratory-averaged cinematic CT (CINE CT) over the whole respiratory period [37]. The CINE CT images were compared against all PET images to ensure accurate alignment between the CT and PET scans.

2.7 Data sets

Patient data were split into training (n=32), validation (n=2), and test (n=6) subsets. Paired inputs (noisy) and targets (noise-free) images were prepared in 3D format. Of the 960 simulated (non-gated) low count volumes (24 volumes per subject), 816 volumes were reserved for training (n=768) and validation (n=48). The validation set was used for hyperparameter tuning. For the test set, 144 simulated (non-gated) low-count data, with 24 volumes per test subject, were used to evaluate the 3D cGAN model. Additionally, for each test subject, 25 dual-gated cardiac PET volumes corresponding to gates 1-25 were used to test with the pre-trained model. Analysis of the learning curves during cGAN training showed that the model performance continued to improve without any signs of overfitting. This suggests that enough variability in physiological uptake distributions was present in the training scan data, enabling robust capture of data patterns by the network. Through this ongoing enhancement of image quality, assessment of the cGAN loss metrics indicates the training dataset was sufficiently large and diverse to justify a 3D model.

2.8 Deep Learning Workflow

2.8.1 Conditional GAN: 3D Pix2Pix

We adopted a cGAN architecture called Pix2Pix that learns a mapping from input images to output images, as described in [26]. The Pix2Pix model is trained through a two-player minimax game, where the generator tries to produce samples that are erroneously classified by the discriminator as real, while the discriminator tries to distinguish the real data samples from the generated ones. During training, the generator and the discriminator iteratively improve, allowing the cGAN to learn to generate realistic data samples that satisfy the specified condition.

In this cGAN model, we condition on input low dose images and generate corresponding synthetic full dose cardiac PET images. cGANs allow additional input, such as auxiliary data (i.e. high statistics cardiac PET data in this paper) as a label, in contrast to conventional GANs, which only use random noise to create samples. Because of this conditioning, cGANs are especially helpful for tasks where the generated samples must have specified qualities or fall into particular distributions. This helps regulate and guide the formation of specific outputs and mitigates the possibility of mode collapse [26].

The generator in 3D Pix2Pix follows a U-Net architecture with 3D convolutional layers in both the encoder and decoder pathways and a discriminator network represented by a convolutional PatchGAN classifier (see Fig. 1). Each encoder block consists of `3DConv` → `InstanceNorm` → `LeakyReLU` using a filter width of 4 and kernel size of 2^{k+3} where k starts at 3 and increments by 1 in each layer. Every encoder block down-samples the images by a factor of 2.

The decoder blocks use `3DConvTranspose` → `InstanceNorm` → `Dropout` → `ReLU` with kernel size 2^{k+3} where k starts at 6 and decrements by 1. `Dropout` was set with a probability of 0.5. A final `tanh` activation layer provides the output. The networks were trained with random initialization of the weights.

The discriminator also employs 3D convolutional layers with `InstanceNorm` and `LeakyReLU` to classify real versus synthetic volumes in a patch-wise manner.

Compared to 2D operations and batch normalization, this 3D architecture with instance normalization better captures inter-slice PET relationships [38]. The model learns from full volumetric patterns rather than slice-by-slice. This allows the 3D Pix2Pix to effectively process and generate 3D PET data.

2.8.2 Denoising Loss Functions

In the context of image denoising, the generator G aims to translate noisy, low-resolution PET images into realistic, high-quality (high resolution and SNR) reconstructions, while the discriminator D tries to differentiate real versus synthetic images. The training data comprises matched pairs (x_i, y_i), where x_i is a low-dose PET image and y_i is the corresponding high-dose target.

Conditional GAN models the mapping from noisy to clean images via the minimax objective as was proposed in [26]:

$$\min_G \max_D \mathcal{L}_{cGAN}(G, D) \tag{1}$$

where $\mathcal{L}_{cGAN}(G, D)$ is:

$$\mathbb{E}_{(x,y)}[\log D(x, y)] + \mathbb{E}_{(x,z)}[1 - \log D(x, G(x, z))] \tag{2}$$

Here, G aims to minimize the loss while D maximizes it. Alongside deceiving D, G must also generate realistic outputs via an L_1 penalty on deviation from the target y:

$$\mathcal{L}_{L_1}(G) = \mathbb{E}_{(x,y,z)}[||y - G(x, z)||_1] \tag{3}$$

The full objective is:

$$G^* = \arg \min_G \max_D \mathcal{L}_{cGAN}(G, D) + \lambda \mathcal{L}_{L_1}(G) \tag{4}$$

where λ weights the L_1 term. The optimization of the generator and discriminator in a GAN is a min-max framework. The generator tries to "reduce" the discrepancy between the images it creates and the real ground truth images. In addition, the discriminator attempts to “maximize” the rate of correct identification of real examples vs. synthetic ones. This adversarial interplay encourages the generator to produce more photo-realistic outputs as the discriminator struggles to identify synthetic creations. Through the combination of a minimum-loss-based generator and a maximum-loss-based discriminator, the GAN training dynamics are gradually directing the model to create synthetic images that are highly similar to the true images.

2.9 Model training

The cGAN model underwent a training regimen spanning 50 epochs, employing the Adam optimizer [39] with default parameters, specifically $\beta_1 = 0.5$ and $\beta_2 = 0.999$. The learning rate was set at 2e-4, and a batch size of 1 was utilized. The fine-tuning of hyperparameters, including the number of epochs, count of convolutional layers in the encoder-decoder blocks, the dimensions and quantity of convolutional filters, batch size, loss functions, and the selection of optimization algorithms for both the Generator and Discriminator, adhered to the guidelines established in the seminal work by [26].

For regularization purposes, a weight of $\lambda = 5$ was introduced to the generator loss. The parameter λ determines the balance between pixel-level accuracy and photo-realism during generator model updates. A higher λ value emphasizes pixel-level similarity, potentially resulting in overly smoothed or blurred outputs. Conversely, a lower λ value prioritizes sharpness but may sacrifice fine details. The optimal value of λ varies based on factors such as image complexity, noise levels, and dataset size, posing a key challenge in applying cGANs effectively. All Pix2Pix models underwent training using TensorFlow (version 2.6.0) and Python (version 3.9). We utilized a national cluster workstation from CSC Finland, featuring multiple NVIDIA TITAN V100 GPUs with

32GB memory each and 350GB CPU memory. However, we only utilized a single GPU unit and its corresponding CPU memory share for our model training experiments, as the full system capacity was unnecessary for our purposes.

2.10 Evaluation

To objectively assess model performance, we utilized four common image quality metrics: peak signal-to-noise ratio (PSNR) [40] for assessment of image quality, structural similarity index measure (SSIM) [41] for assessment of structural patterns, luminance, and contrast, and mean absolute error (MAE) for assessment of both image quality and bias.

PSNR is the ratio between maximum signal power and distorting noise power. For a reference image f and reconstructed image g of size $M \times N$ pixels,

$$\mathrm{PSNR} = 20 \log_{10} \left(\frac{\mathrm{MAX}_f}{\sqrt{\mathrm{MSE}}} \right), \tag{5}$$

where MAX_f is the maximum intensity and mean square error (MSE) is

$$\mathrm{MSE} = \frac{1}{MN} \sum_{i=1}^{M} \sum_{j=1}^{N} \left(f(i,j) - g(i,j) \right)^2 . \tag{6}$$

Although PSNR is widely used, it has limitations in correlating with perceived visual quality.

In contrast, SSIM considers perceptual factors like luminance, contrast, and structure. For image patches f and g,

$$\mathrm{SSIM}(f,g) = l(f,g) \cdot c(f,g) \cdot s(f,g), \tag{7}$$

where l, c, and s measure luminance, contrast, and structural similarities, respectively. SSIM ranges from 0 to 1, with 1 indicating perfect structural similarity.

2.10.1 Bootstrapping

To characterize the statistical properties of the PET images, we utilized bootstrap resampling to generate multiple simulated datasets from the acquired and predicted patient images [42]. By randomly selecting pixels with replacement, this created `P=100` new bootstrap sample images per patient that capture data variability. Analyzing these bootstrap datasets enabled estimating population statistics like the mean and standard error from our image sample. Bootstrapping thus allowed quantifying uncertainty and distributional properties despite only having a limited sample of patients.

2.10.2 Statistical Analysis

Continuous variables are presented as mean±standard error. The Mann-Whitney test was used to assess the statistical significance of differences between the performance metrics before and after applying the deep learning model. This non-parametric test can handle non-normal data and compare two groups. A significance level of 0.05 was used. All statistical analyses were done using Python SciPy (v1.11.2).

3 Results

3.1 Non-gated PET Denoising

Figure 2 provides a qualitative comparison of restored PET outputs versus the low-dose inputs and high-dose ground truth. The sample axial and coronal slices from patients A and B demonstrate the generative model's ability to learn and generate finer anatomical details and reduce noise compared to the low dose inputs.

Quantitative image quality metrics in Table 2 demonstrate the enhanced performance of the synthetic PET images. Across all non-gated test cases with 5-15% dose reduction, the deep learning models effectively reduced noise, improved structural accuracy, and lowered bias compared to the low-dose inputs. Table 2 details the percentage improvement in PSNR, SSIM, and MAE relative to the low-dose PET for each model and noise level. These quantitative results validate the improvements in image quality and accuracy achieved by the 3D generative model.

Figure 3 shows image quality and bias metrics in three distinct noise levels, each featuring paired low dose and synthetic high-dose data points. The metrics indicate the 3D Pix2Pix model performs well in enhancing image quality from low to higher dose levels. The PSNR, SSIM, and MAE values show statistically significant (P-value<0.001) improvements moving from 5% to 15% dose inputs. Overall, both quantitative metrics and qualitative examples validate the efficacy of the 3D Pix2Pix model for non-gated PET image enhancement.

3.2 Dual-gated PET Denoising

After training the model on non-gated data, we evaluated the model performance on the more challenging task of inferring with the trained 3D Pix2Pix. For each dual-gated cardiac PET study, we reconstructed image data for 25 gates, with each gate containing 1% to 15% of the total counts. These low-count gated images were input to the pre-trained model. For this experiment, we considered the model which was trained on the non-gated data with as low as 5% count rates. We did not use any non-gated low-dose (5% to 15%) instances of the subjects reserved for the testing phase. This ensured that our model remained entirely blind when engaged in generating synthetic high-dose (denoised) dual-gated PET images, as visually illustrated in Figure 4.

Figure 4 shows the denoising performance of the 3D Pix2Pix in axial view for six different gates, ranging from bins 1 to 25, for patients A and B. As it shows, in all these gates, regardless of the state of the motion, the 3D Pix2Pix model remarkably improved the signal-to-noise ratio and anatomical characteristics of the heart. Furthermore, the 3D Pix2Pix denoising model maintained its effectiveness even when the gate being considered had a significantly lower data-count fraction than that considered in training. In gates 1 through 10, for example, images are reconstructed with significant low count statistics ($< 5\%$), and the model was still able to restore synthetic high-dose images with high resolution and quality.

Figure 5 shows image quality metrics, PSNR, SSIM, and MAE measurements, for the input (noisy dual-gated), and restored data (synthetic standard dose PET images generated by 3D Pix2Pix model) as compared to the standard dose non-gated PET data. In order to objectively assess the performance of the denoising model, we used the matching non-gated high-dose image as the target to calculate PSNR, SSIM, and MAE. Through gates number 1 to 25, the 3D Pix2Pix model significantly improved signal-to-noise ratio and perceptual quality. Notably, the gates with the lowest count-data fraction improved the most.

Table 3 summarizes the model performance on the dual-gated test cases across all 25 gates. Testing on dual-gated data is more difficult due to the reduced counts per gate image compared to a full-dose non-gated image. Nonetheless, the pre-trained model demonstrated robust image enhancement even when applied to new dual-gated image data not seen during training. This highlights the versatility of the deep learning approach to generalize across varying input types. For a comprehensive evaluation of the performance of our trained generative model, encompassing all 25 sets of cardiac and respiratory gated images, we have presented a detailed summary in Figure 5. The PSNR, SSIM, and MAE values show statistically significant (P-value$<$0.001) improvements moving from Gate 1 to Gate 25.

3.3 Line Profile Analysis

Figure 6 and 7 show uptake profiles through the myocardium wall. Compared to noisy dual-gated PET data, the 3D Pix2Pix method achieves higher contrast recovery and maintains the improved spatial resolution and reduced motion blur of the dual-gated images.

4 Discussion

We presented a generative denoising scheme using a 3D cGAN and transferring learned generative parameters from non-gated low-dose data to the dual-gated reconstructions of the unseen test cases. Low-dose PET scans included ultra count statistics ranging from 5-15% of the total full count scans. As demonstrated in Figs. 2 and 3, trained models on non-gated data effectively generated high-quality images by improving signal to noise ratio without causing any noticeable side effects or artifacts. Figure 4 shows cardiac PET images across different gates, comparing original noisy gated PET images to those denoised by cGAN. The dual-gated PET images, which suffered from an extensive amount of post-reconstruction noise due to the low statistics, were effectively denoised while retaining the sharpness and distinct anatomy of the myocardial regions. For comparison purposes, we also added non-gated full-dose PET images to show the differences between gated and non-gated PET. Non-gated images provide a single static visualization of tracer uptake over the entire scan duration. In contrast, gated images are temporally aligned with the cardiac and respiratory cycles, capturing uptake dynamically across multiple gate frames. Intensity line profiles (Fig. 6 and Fig. 7) also demonstrate a noticeable increase in signal intensity for the cGAN-based denoised images across different gates compared to the noisy non-gated and dual-gated images.

The proposed generative denoising model enables a more comprehensive analysis of dual-gated cardiac PET images across different cardiac and respiratory phases (see Fig. 5). By generating high-quality dual-gated volumes across all the gates, clinicians can obtain a precise and accurate multi-phase cardiac PET assessment with full motion characterization. As seen in Fig. 4, the 3D Pix2Pix model's robustness across physiological motion states enables the efficient generation of high-quality gated PET images. This highlights cGAN capabilities beyond just denoising that offer additional benefits for cardiac PET analysis. For instance, the model could denoise dynamic PET imaging over the cardiac cycle, as well as studies using ultra-short half-life radiotracers like Oxygen-15 or Rubidium-82 that produce very low-count PET data. By effectively handling varying noise levels and motion, the generative approach expands the utility of low-dose and gated cardiac PET acquisitions.

Generative deep learning models can potentially improve the diagnostic accuracy of dual-gated cardiac PET imaging. With these models, high-quality and accurate images can be generated from noisy gated data, enabling uncovering of subtle changes in the heart and blood flow that aren't apparent in the non-gated images or in the conventional single- or dual-gated dynamic studies. This can lead to a more accurate diagnosis, improved patient outcomes, and reduced costs associated with repeat imaging procedures. Unlike the other studies in the realm of applying deep learning to enhance myocardial PET images that were described in Section 1 [27–29, 34], we trained a 3D conditional GAN (cGAN) model on non-gated cardiac PET images comprising varying count statistics to simulate varying noise conditions. We then tested the model's generalizability by applying it to synthesize noise-reduced versions of independent dual-gated cardiac PET datasets suffering from low count statistics and motion-related artifacts.

Additionally, as seen in Fig. 2, generative models can help improve the SNR and accuracy of ultra-low dose PET imaging. Automating the image generation workflow can reduce the amount of time and manual effort required to acquire PET scans, allowing medical professionals to focus on interpreting the results and making faster decisions based on the data. This can involve generating PET images with a significantly less amount of radiotracer dose injection and improving the resulting image by learning parameters from existing images to better understand the dynamics of the heart and blood flow. It is particularly useful in dual-gated PET, in which cardiac- and respiratory-gated images can be combined to create a more accurate and detailed image of the heart.

This paper focused on denoising low-dose cardiac PET data – simulated by subsampling high dose scans using cGANs. While our method showed promising results, in practice, it would be challenging to confirm different noise levels with the same radioactivity distribution using patient data alone. However, by performing numerical simulations, such as Monte Carlo simulations using digital phantoms, it becomes possible to validate our findings under different noise levels while maintaining the same underlying radioactivity distribution. Monte Carlo simulations, for example, using toolkits like GATE [43] or SimSET [44], offer a controlled and reproducible environment, allowing for systematic evaluation and fine-tuning of the proposed methods. Therefore, numerical simulations will be an interesting direction for future research, enabling us to comprehensively assess the robustness and performance of our cGAN-based denoising approach across various noise scenarios while ensuring a consistent ground truth radioactivity distribution.

While cGAN models have achieved success in applications like image translation, limitations remain. Such generative denoising models can produce highly realistic synthetic high-dose images through adversarial training. However, they generally struggle to accurately encode and decode spatial and texture patterns without incorporating anatomically matched prior information, such as CT data or segmentation maps. The lack of this additional guidance makes it challenging for cGANs to faithfully reproduce intricate anatomical structures and spatial relationships present in the original data [45]. In addition, convolutional-based cGAN-generated images may lack fine details or introduce artifacts due to limited diversity in training data, instability from mode collapse [46], underfitting on small localized features [47], and sensitivity to hyperparameters that require careful tuning [48]. By incorporating a label for training, cGANs are designed to mitigate mode collapse, but there is a potential trade-off as they may encounter challenges in capturing the entire spectrum of variations linked to diverse conditions [49]. Ongoing efforts are improving models through advanced architectures such as denoising diffusion probabilistic models [50] and vision transformers [51]. Diffusion models present a promising future direction for advancing the image denoising capabilities demonstrated in this work. As generative models leveraging stochastic processes, diffusion frameworks inherently handle complex spatially correlated noise patterns within target data distributions. Recent advances in denoising diffusion model architectures achieving state-of-the-art results across perceptual image synthesis and reconstruction tasks [52, 53] motivate their exploration for PET image enhancement. Adapting current techniques including score-based diffusion models [54] may offer performance improvements while generalizing robustly across imaging domains. With further research into tailoring initialization, score modeling, and sampling procedures, diffusion-based approaches could soon complement or build upon the capabilities of adversarial methods highlighted in this paper for reliable PET data denoising of low-count scans in both non-gated and motion-affected reconstructions.

One limitation of this study is that it relied on a relatively small dataset of FDG PET scans for training the generative model. In the future, we plan to explore the use of different radiotracers, particularly those with significantly shorter half-lives, and extend the size of training data. Another limitation of this study is the absence of prior information integration, such as CT data, during the model training process. Moving forward, we will explore the incorporation of early fusion techniques to include priors in the network.

PET image correction via generative modeling remains an interesting field warranting further research. New directions include exploring vision transformers to achieve fine-grained synthetic standard dose PET data, incorporating anatomical priors, e.g. 4D-CT or CINE-CT, to preserve the structure, fusing gating data to provide additional signal, optimizing execution time for real-time use, and investigating pre-trained models for gate-to-gate translation. As generative techniques mature, they will likely play an increasing role in PET imaging and other medical applications.

5 Conclusion

We designed and evaluated a 3D Pix2Pix architecture for zero-shot denoising of dual-gated and non-gated cardiac PET images using real-world clinical data. Quantitative and qualitative results demonstrated effective noise suppression by the generative denoising model while maintaining perceptual quality, sharpness, resolution, and anatomical fidelity in PET images. Further statistical evaluations on both gated and non-gated PET data validated the denoising capabilities. Our findings indicate that deep generative denoising model can substantially enhance cardiac PET imaging by reducing noise and artifacts while maintaining high resolution and contrast recovery. This may improve diagnostic accuracy, especially for dual gated studies where normally conventionally shorter gate intervals better minimize motion at the cost of increased noise. As generative models continue to advance, ultra-low dose PET acquisitions could also be effectively denoised to boost image quality and diagnostic potential while lowering radiation exposure.

Acknowledgments

The authors would like to express their gratitude for the assistance and computing resources provided by the CSC-Puhti supercomputer, a non-profit state company operated by the Finnish government and higher education institutions in Finland. Dr. Jafaritadi is grateful for support from Finland State Research Fund (grant number: 11125), Ulla Tuominen Foundation, the Stanford Cancer Institute, and Finnish Cultural Foundation Postdoctoral Research Fellowship (grant number: 00220062). Dr. Antti Saraste discloses grants from the Academy of Finland and Finnish Foundation for Cardiovascular Research; and fees for lectures or consultancy from Abbott, Astra Zeneca, Bayer, Novartis and Pfizer.

References

[1] Schindler TH, Schelbert HR, Quercioli A, Dilsizian V. Cardiac PET imaging for the detection and monitoring of coronary artery disease and microvascular health. JACC: Cardiovascular Imaging. 2010;3(6):623-40.

[2] Loghin C, Sdringola S, Gould KL. Common artifacts in PET myocardial perfusion images due to attenuation–emission misregistration: clinical significance, causes, and solutions. Journal of Nuclear Medicine. 2004;45(6):1029-39.

[3] Burrell S, MacDonald A. Artifacts and pitfalls in myocardial perfusion imaging. Journal of nuclear medicine technology. 2006;34(4):193-211.

[4] Nehmeh SA, Erdi YE, Rosenzweig KE, Schoder H, Larson SM, Squire OD, et al. Reduction of respiratory motion artifacts in PET imaging of lung cancer by respiratory correlated dynamic PET: methodology and comparison with respiratory gated PET. Journal of Nuclear Medicine. 2003;44(10):1644-8.

[5] Livieratos L, Rajappan K, Stegger L, Schafers K, Bailey DL, Camici PG. Respiratory gating of cardiac PET data in list-mode acquisition. European Journal of Nuclear Medicine and Molecular Imaging. 2006;33:584-8.

[6] Büther F, Dawood M, Stegger L, Wübbeling F, Schäfers M, Schober O, et al. List mode–driven cardiac and respiratory gating in pet. Journal of Nuclear Medicine. 2009;50(5):674-81.

[7] Lassen ML, Kwiecinski J, Slomka PJ. Gating approaches in cardiac PET imaging. PET clinics. 2019;14(2):271-9.

[8] Kokki T, Sipilä HT, Teräs M, Noponen T, Durand-Schaefer N, Klén R, et al. Dual gated PET/CT imaging of small targets of the heart: Method description and

testing with a dynamic heart phantom. Journal of nuclear cardiology. 2010;17:71-84.

[9] Chan C, Fulton R, Feng DD, Meikle S. Median non-local means filtering for low SNR image denoising: Application to PET with anatomical knowledge. In: IEEE Nuclear Science Symposuim & Medical Imaging Conference. IEEE; 2010. p. 3613-8.

[10] Joshi N, Jain S, Agarwal A. An improved approach for denoising MRI using non local means filter. In: 2016 2nd International Conference on Next Generation Computing Technologies (NGCT). IEEE; 2016. p. 650-3.

[11] Wang Y, Guo J, Chen W, Zhang W. Image denoising using modified Perona–Malik model based on directional Laplacian. Signal Processing. 2013;93(9):2548-58.

[12] Danielyan A, Katkovnik V, Egiazarian K. BM3D frames and variational image deblurring. IEEE Transactions on image processing. 2011;21(4):1715-28.

[13] Dabov K, Foi A, Katkovnik V, Egiazarian K. Image restoration by sparse 3D transform-domain collaborative filtering. In: Image Processing: Algorithms and Systems VI. vol. 6812. International Society for Optics and Photonics; 2008. p. 681207.

[14] Xiang L, Qiao Y, Nie D, An L, Lin W, Wang Q, et al. Deep auto-context convolutional neural networks for standard-dose PET image estimation from low-dose PET/MRI. Neurocomputing. 2017;267:406-16.

[15] Xu J, Gong E, Pauly J, Zaharchuk G. 200x low-dose PET reconstruction using deep learning. arXiv preprint arXiv:171204119. 2017.

[16] Gong K, Guan J, Liu CC, Qi J. PET image denoising using a deep neural network through fine tuning. IEEE Transactions on Radiation and Plasma Medical Sciences.

2018;3(2):153-61.

[17] Ote K, Hashimoto F, Onishi Y, Isobe T, Ouchi Y. List-mode PET image reconstruction using deep image prior. IEEE Transactions on Medical Imaging. 2023.

[18] Gambin JR, Tadi MJ, Teuho J, Klén R, Knuuti J, Koskinen J, et al. Learning to Denoise Gated Cardiac PET Images Using Convolutional Neural Networks. IEEE Access. 2021;9:145886-99.

[19] Ulyanov D, Vedaldi A, Lempitsky V. Deep image prior. International Journal of Computer Vision. 2020;128(7).

[20] Gong K, Catana C, Qi J, Li Q. PET image reconstruction using deep image prior. IEEE transactions on medical imaging. 2018;38(7):1655-65.

[21] Zhang K, Li Y, Zuo W, Zhang L, Van Gool L, Timofte R. Plug-and-Play Image Restoration with Deep Denoiser Prior. arXiv preprint. 2020.

[22] Lehtinen J, Munkberg J, Hasselgren J, Laine S, Karras T, Aittala M, et al. Noise2noise: Learning image restoration without clean data. arXiv preprint arXiv:180304189. 2018.

[23] Liu J, Yang Y, Wernick MN, Pretorius PH, King MA. Deep learning with noise-to-noise training for denoising in SPECT myocardial perfusion imaging. Medical physics. 2021;48(1):156-68.

[24] Krull A, Buchholz TO, Jug F. Noise2void-learning denoising from single noisy images. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition; 2019. p. 2129-37.

[25] Vaquero JJ, Kinahan P. Positron emission tomography: current challenges and opportunities for technological advances in clinical and preclinical imaging systems. Annual review of biomedical engineering. 2015;17:385-414.

[26] Isola P, Zhu JY, Zhou T, Efros AA. Image-to-image translation with conditional adversarial networks. In: Proceedings of the IEEE conference on computer vision and pattern recognition; 2017. p. 1125-34.

[27] Sun J, Du Y, Li C, Wu TH, Yang B, Mok GS. Pix2Pix generative adversarial network for low dose myocardial perfusion SPECT denoising. Quantitative Imaging in Medicine and Surgery. 2022;12(7):3539.

[28] Sun J, Jiang H, Du Y, Li CY, Wu TH, Liu YH, et al. Deep learning-based denoising in projection-domain and reconstruction-domain for low-dose myocardial perfusion SPECT. Journal of Nuclear Cardiology. 2023;30(3):970-85.

[29] Guo X, Shi L, Chen X, Zhou B, Liu Q, Xie H, et al. TAI-GAN: Temporally and Anatomically Informed GAN for early-to-late frame conversion in dynamic cardiac PET motion correction. In: International Workshop on Simulation and Synthesis in Medical Imaging. Springer; 2023. p. 64-74.

[30] Wang Y, Yu B, Wang L, Zu C, Lalush DS, Lin W, et al. 3D conditional generative adversarial networks for high-quality PET image estimation at low dose. Neuroimage. 2018;174:550-62.

[31] Lu W, Onofrey JA, Lu Y, Shi L, Ma T, Liu Y, et al. An investigation of quantitative accuracy for deep learning based denoising in oncological PET. Physics in Medicine & Biology. 2019;64(16):165019.

[32] Zhou L, Schaefferkoetter JD, Tham IW, Huang G, Yan J. Supervised learning with cyclegan for low-dose FDG PET image denoising. Medical image analysis.

2020;65:101770.

[33] Xue H, Teng Y, Tie C, Wan Q, Wu J, Li M, et al. A 3D attention residual encoder–decoder least-square GAN for low-count PET denoising. Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment. 2020;983:164638.

[34] Zhou B, Tsai YJ, Chen X, Duncan JS, Liu C. MDPET: a unified motion correction and denoising adversarial network for low-dose gated PET. IEEE transactions on medical imaging. 2021;40(11):3154-64.

[35] Bettinardi V, Presotto L, Rapisarda E, Picchio M, Gianolli L, Gilardi M. Physical Performance of the new hybrid PET/CT Discovery-690. Medical physics. 2011;38(10):5394-411.

[36] Schultz J, Siekkinen R, Tadi MJ, Teräs M, Klén R, Lehtonen E, et al. Effect of respiratory motion correction and CT-based attenuation correction on dual-gated cardiac PET image quality and quantification. Journal of Nuclear Cardiology. 2022;29(5):2423-33.

[37] Alessio AM, Kohlmyer S, Branch K, Chen G, Caldwell J, Kinahan P. Cine CT for attenuation correction in cardiac PET/CT. Journal of nuclear medicine. 2007;48(5):794-801.

[38] Cirillo MD, Abramian D, Eklund A. Vox2Vox: 3D-GAN for brain tumour segmentation. In: Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries: 6th International Workshop, BrainLes 2020, Held in Conjunction with MICCAI 2020, Lima, Peru, October 4, 2020, Revised Selected Papers, Part I 6. Springer; 2021. p. 274-84.

[39] Kingma DP, Ba J. Adam: A method for stochastic optimization. arXiv preprint arXiv:14126980. 2014.

[40] An L, Zhang P, Adeli E, Wang Y, Ma G, Shi F, et al. Multi-level canonical correlation analysis for standard-dose PET image estimation. IEEE Transactions on Image Processing. 2016;25(7):3303-15.

[41] Wang Z, Bovik AC, Sheikh HR, Simoncelli EP. Image quality assessment: from error visibility to structural similarity. IEEE transactions on image processing. 2004;13(4):600-12.

[42] Johnson RW. An introduction to the bootstrap. Teaching statistics. 2001;23(2):49-54.

[43] Jan S, Santin G, Strul D, Staelens S, Assié K, Autret D, et al. GATE: a simulation toolkit for PET and SPECT. Physics in Medicine & Biology. 2004;49(19):4543.

[44] Lewellen TK, Harrison RL, Vannoy S. Monte carlo calculations in nuclear medicine. Institute of Physics Publishing: Philadelphia. 1998:309.

[45] Wang TC, Liu MY, Zhu JY, Tao A, Kautz J, Catanzaro B. High-resolution image synthesis and semantic manipulation with conditional gans. In: Proceedings of the IEEE conference on computer vision and pattern recognition; 2018. p. 8798-807.

[46] Bau D, Zhu JY, Wulff J, Peebles W, Strobelt H, Zhou B, et al. Seeing what a gan cannot generate. In: Proceedings of the IEEE/CVF International Conference on Computer Vision; 2019. p. 4502-11.

[47] Wang W, Sun Y, Halgamuge S. Improving MMD-GAN training with repulsive loss function. arXiv preprint arXiv:181209916. 2018.

[48] Lucic M, Kurach K, Michalski M, Gelly S, Bousquet O. Are gans created equal? a large-scale study. Advances in neural information processing systems. 2018;31.

[49] Adiga S, Attia MA, Chang WT, Tandon R. On the tradeoff between mode collapse and sample quality in generative adversarial networks. In: 2018 IEEE global conference on signal and information processing (GlobalSIP). IEEE; 2018. p. 1184-8.

[50] Ho J, Jain A, Abbeel P. Denoising diffusion probabilistic models. Advances in Neural Information Processing Systems. 2020;33:6840-51.

[51] Luo Y, Wang Y, Zu C, Zhan B, Wu X, Zhou J, et al. 3D transformer-GAN for high-quality PET reconstruction. In: Medical Image Computing and Computer Assisted Intervention–MICCAI 2021: 24th International Conference, Strasbourg, France, September 27–October 1, 2021, Proceedings, Part VI 24. Springer; 2021. p. 276-85.

[52] Saharia C, Chan W, Chang H, Lee C, Ho J, Salimans T, et al. Palette: Image-to-image diffusion models. In: ACM SIGGRAPH 2022 Conference Proceedings; 2022. p. 1-10.

[53] Rombach R, Blattmann A, Lorenz D, Esser P, Ommer B. High-resolution image synthesis with latent diffusion models. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition; 2022. p. 10684-95.

[54] Song Y, Sohl-Dickstein J, Kingma DP, Kumar A, Ermon S, Poole B. Score-based generative modeling through stochastic differential equations. arXiv preprint arXiv:201113456. 2020.

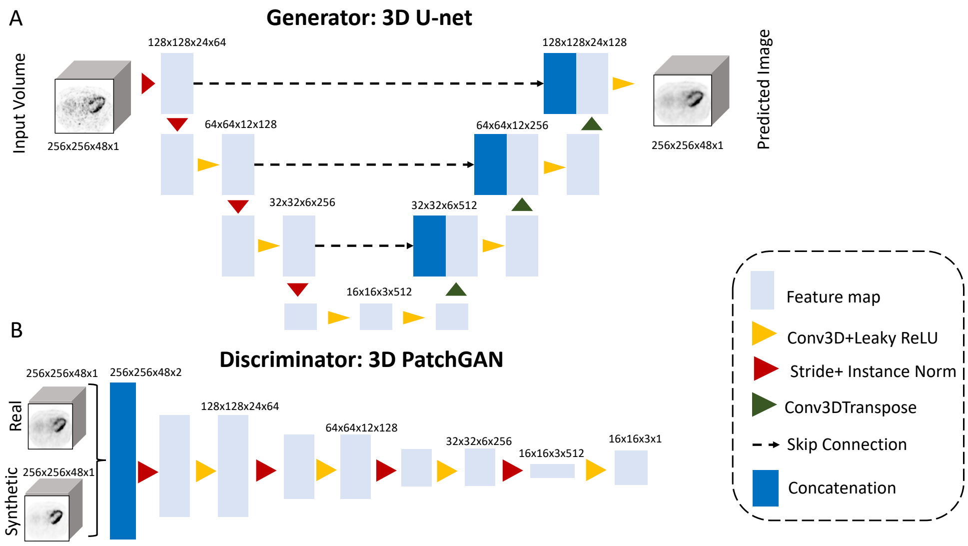

Fig. 1: 3D Pix2Pix model. (A) Generator networks with a U-net architecture and Discriminator (B) with a patchGAN classifier.

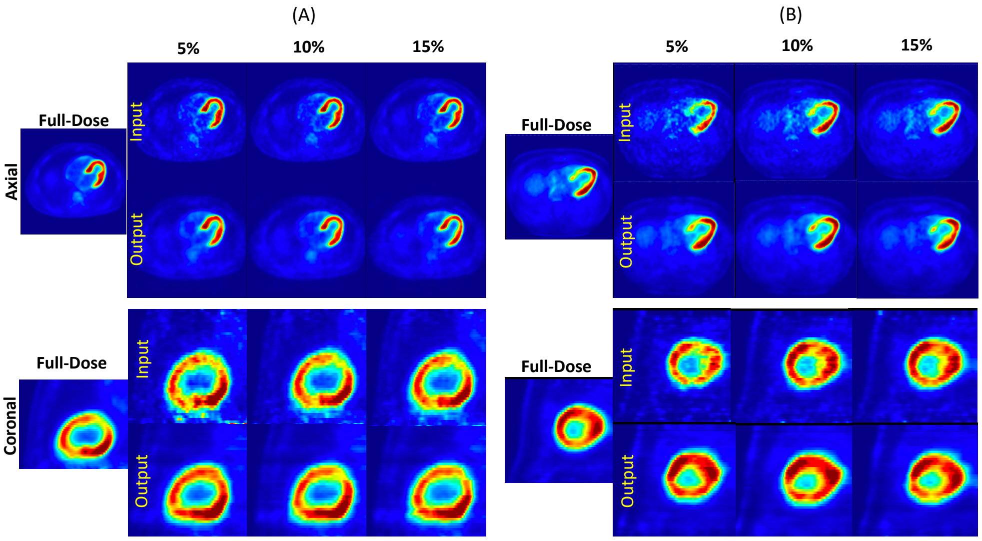

Fig. 2: Reconstructed original high dose target, low dose inputs of 5%,10%,15%, and 3D Pix2Pix generated synthetic high dose for patients (A) and (B) in axial and coronal views. A zoomed-in view of the Coronal view is shown for better visualization.

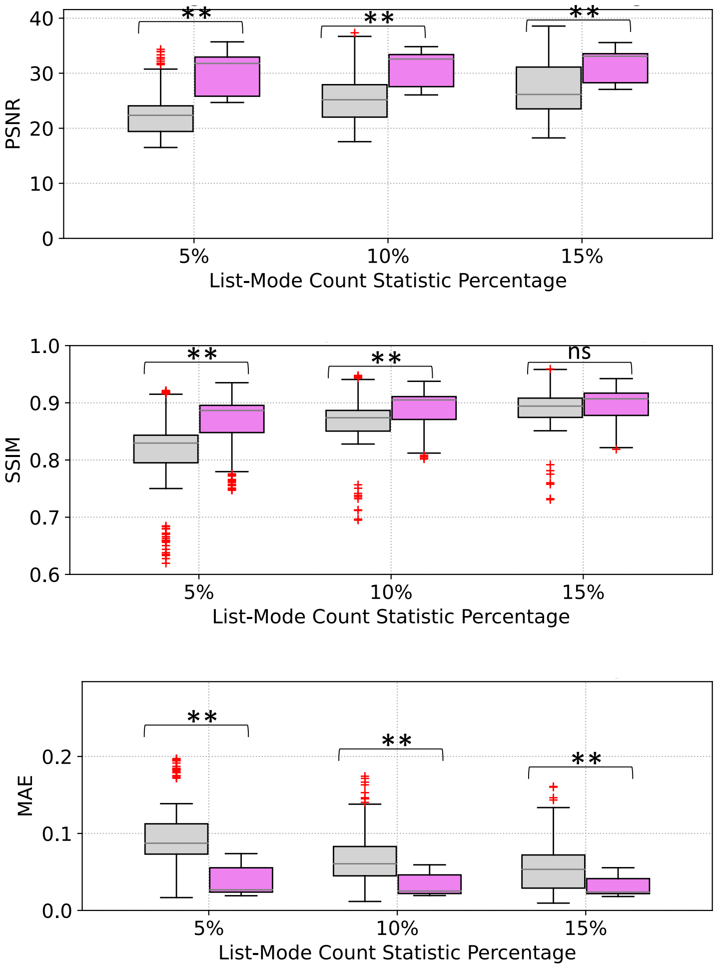

Fig. 3: Image quality assessment including the PSNR, SSIM and MAE measurements when comparing non-gated low-dose (5%,10%, and 15%) inputs (gray) and corresponding 3G Pix2Pix restored (synthetic high-dose) images (pink) against the non-gated high dose target ($* =$ P-value<0.05, $** =$ P-value<0.001, ns = non-significant).

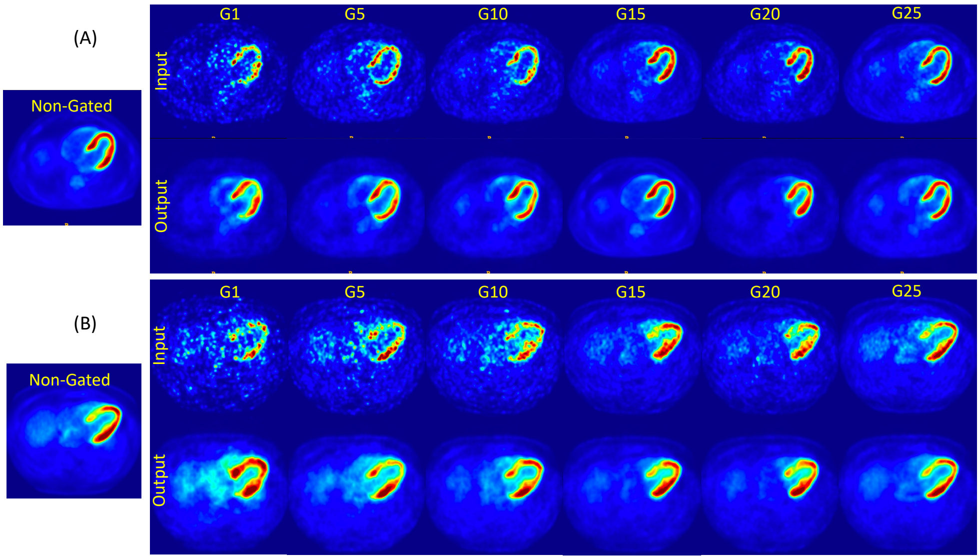

Fig. 4: Dual gated cardiac (input) and generative deep learning enhanced (output) PET images in five different cardiac and respiratory phases for patients (A) and (B) in axial view.

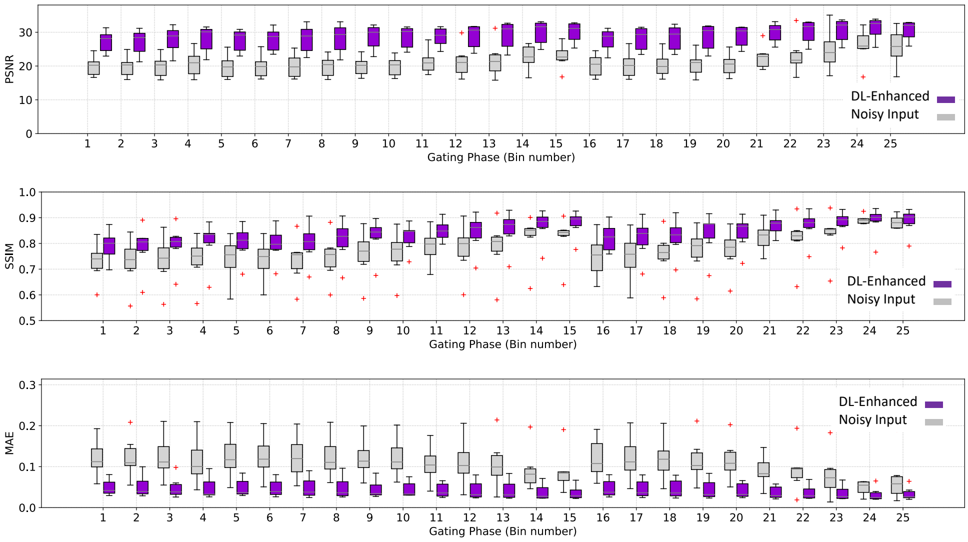

Fig. 5: Image quality assessment including the PSNR, SSIM and MAE measurements obtained from all 25 gates when comparing noisy dual gated (gray) input and 3D Pix2Pix restored (synthetic high-dose) images (purple) against the non-gated high dose. When comparing cGAN-denoised data and noisy dual-gated inputs against original standard high statistics targets, the images from cGAN showed significantly (P-value$<$0.05) higher PSNR and SSIM, along with lower MAE.

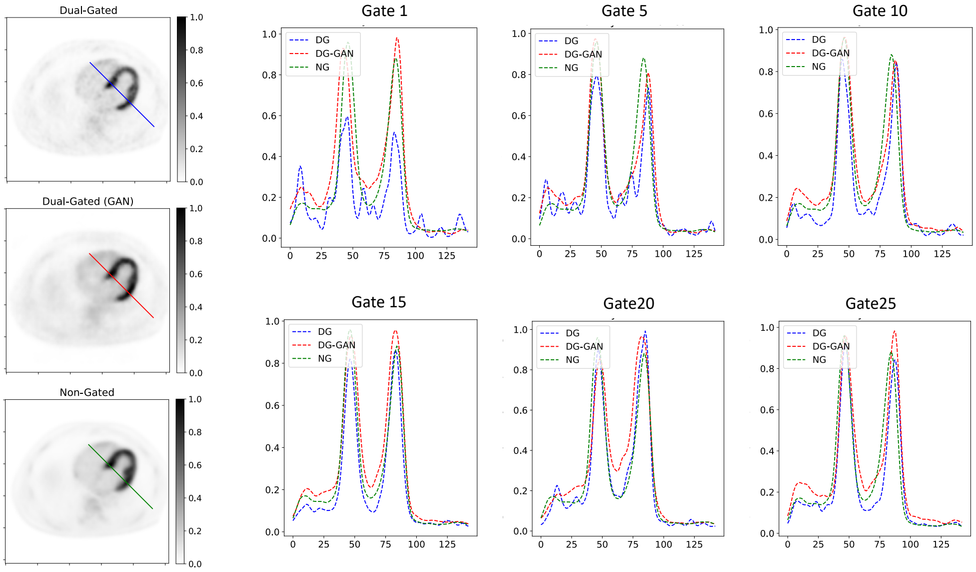

Fig. 6: Normalized uptake profile analysis across the myocardium wall in patient A. The peak activity is higher in GAN-enhanced PET images than the input dual-gated in all motion phases.

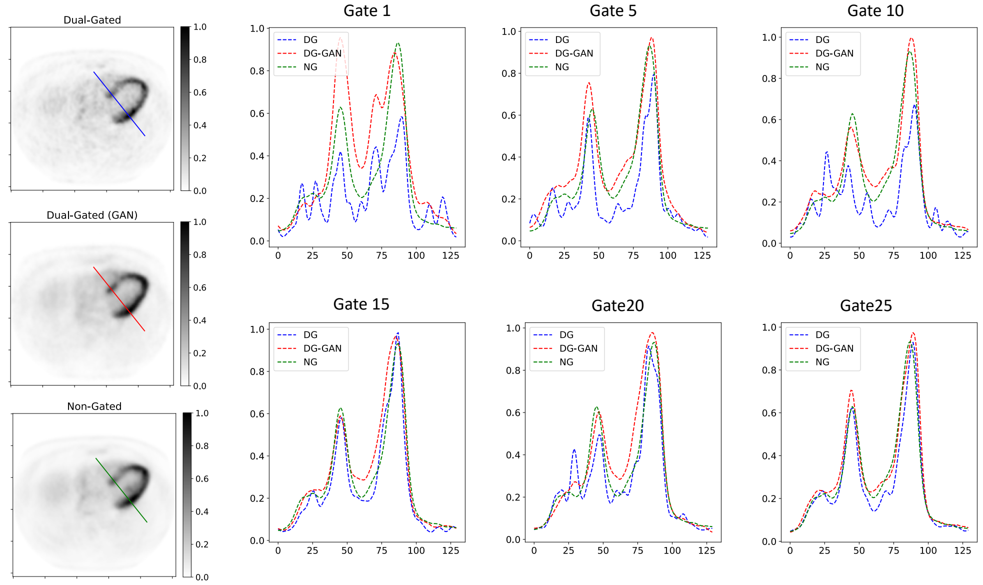

Fig. 7: Normalized uptake profile analysis across the myocardium wall in patient B. The peak activity is higher in GAN-enhanced PET images than the input dual-gated in all motion phases.

Table 1: Subject characteristics with mean, standard deviations, and ranges.

Demographics	**Mean±SD**	**Range**
Age (years)	64 ± 9	$44 - 84$
Weight (kg)	86 ± 15	$47 - 116$
Height (m)	1.75 ± 0.09	$1.53 - 2.00$
Dose (MBq)	309 ± 26	$277 - 400$

MBq: Megabecquerel
SD: standard deviation

Table 2: Summary of denoising performance (bootstrapped) over non-gated PET data. Values are in Mean±Standard Error and % format.

DRF	5%			10%			15%		
	Low-dose input	Synthetic	Enhanced	Low-dose input	Synthetic HD	Enhanced	Low-dose input	Synthetic HD	Enhanced
PSNR	22.6 ± 0.4	30.3± 0.3	33.7%	25.7 ±0.7	31.17 ± 0.38	21.2%	27.4±0.9	31.6 ± 0.41	15.5%
SSIM	0.81 ± 0.007	0.87 ± 0.004	7.1%	0.86 ± 0.008	0.89 ± 0.005	3.3 %	0.88±0.009	0.90±0.005	2.2%
MAE	0.100 ± 0.004	0.037 ± 0.002	61.4%	0.07 ± 0.006	0.032 ± 0.002	54.3 %	0.06 ± 0.007	0.030 ± 0.002	49.7%

Table 3: Summary of denoising performance (bootstrapped) over all 25 gates. Values are in Mean±Standard Error and % format.

Model	3D Pix2Pix		
Metric	Dual Gated Input	Synthetic	Enhanced
PSNR	21.4 ± 1.45	28.8 ± 1.36	34.7%
SSIM	0.77 ± 0.034	0.84 ± 0.025	7.8%
MAE	0.11 ± 0.02	0.04 ± 0.008	60.3%
NRMSE	0.02 ± 0.006	0.004 ± 0.001	82.0%