Time-Aware GAN for Uptake Time Correction and Standard Uptake Value Harmonization in Dynamic PET Imaging

Xueqi Guo[1,2] **Vijay Shah**[1] **David Pigg**[1] **Guenther Platsch**[3] **Xiongchao Chen**[1,2]
Huidong Xie[1,2] **Weijie Gan**[1,4] **Nicha Dvornek**[2] **Chi Liu**[2]
Gerardo Hermosillo Valadez[1] **Lauren Partin**[1] **Bruce Spottiswoode**[1]
[1]Siemens Healthineers, USA [2]Yale University, USA
[3]Siemens Healthineers AG, Germany [4]Washington University at St. Louis, USA
{xueqi.guo,bruce.spottiswoode}@siemens-healthineers.com

Abstract

In dynamic positron emission tomography (PET), both standard uptake value (SUV) and standard tumor-to-blood uptake ratio (SUR) are sensitive to scan time, and inconsistent uptake time might lead to inaccurate metabolism quantification. To overcome the limitations in the current analytical method or deep learning models for uptake time correction, we propose a time-aware generative adversarial network (GAN)-based method to correct SUVs of dynamic frames at a different uptake time to the reference time (60-minute). Specifically, the uptake time of the input frame is encoded and embedded into the bottleneck of the generator through a learnable representation and feature-wise linear modulation, and the temporal 2.5D input along the time dimension provides essential time- and kinetics-related context to the model. On a real-patient dataset, the proposed model demonstrated its ability to predict the dynamic frame at the reference time from a different uptake time with desirable visual performance, high quantitative image similarity measurements, and comparable SUV and SUR distributions, outperforming other analytical and generative baselines. The nuclear medicine expert's review of the readings noted comparable visual and noise patterns, along with identified lesions showing no change in interpretation. The potential to shorten the current clinical workflow by reducing uptake time is suggested.

1 Introduction

Whole-body positron emission tomography (PET) injecting radioactive tracer 2-deoxy-2-[^{18}F]fluoro-D-glucose (FDG) has been widely applied for glycolytic metabolism measurement in clinical oncologic, neurologic, and cardiovascular diagnosis (30). The standard uptake value (SUV) is a quantitative glucose metabolic activity measurement defined as the tracer concentration divided by the injected dose normalized by body weight, typically at approximately 60-minute post tracer injection for an interval of 5-15 min, with the unit of g/mL (27; 28). The standard tumor-to-blood uptake ratio (SUR) is defined as the SUV divided by the mean blood pool uptake as an alternate of the SUV measurement (29).

However, both SUV and SUR are highly time-dependent. The variability of uptake quantification time is a significant problem in clinical oncological quantification since it is difficult to ensure a consistent uptake time in clinical practice. The tumor SUV generally increases over time after tracer injection which might lead to incorrect diagnosis and patient management (10; 28). Current work has investigated various methods of scan time correction for SUVs. A general SUV correction formula

GenAI for Health Workshop @ NeurIPS 2024, Vancouver.

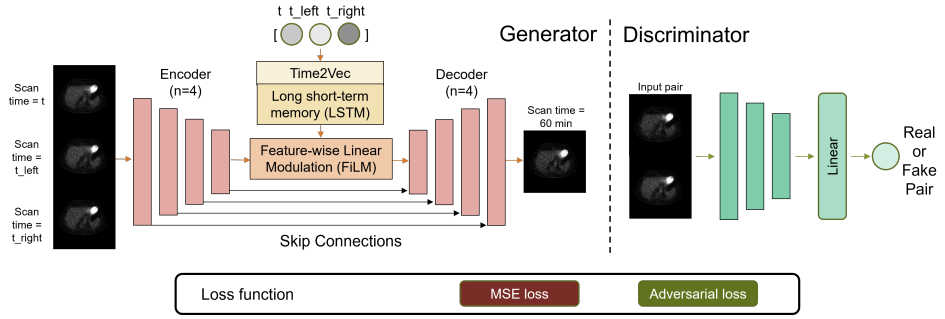

Figure 1: The structure of the proposed time-aware GAN for uptake time correction.

was proposed in (26) but mainly focused on imaging time optimization. Another correction formula was proposed in (1) but was based on purely empirical observation restricted to breast cancer patients only. (28) proposed an uptake time correction method that is validated on whole-body scans, but is based on the irreversible assumption of FDG and requires an arterial input function.

Recently, deep learning has been widely applied in image generation and conversion tasks in PET imaging as a data-driven approach with performance superior to conventional methods. The network structure of a U-Net (3) has been widely applied as a benchmark in generating parametric K_i images from static PET (17), generating virtual high-count PET images (16), and super-resolution brain imaging (20). However, generative adversarial networks (GANs) (6) can provide more realistic results than U-Nets due to the competition between a generator and a discriminator under an adversarial loss in model training. GANs have been successfully implemented for synthesizing PET images from other modalities such as CT (2), MRI (13) and multi-tracer images (31). In dynamic FDG PET, the concurrent state-of-the-art image synthesis method uses single-pairwise conditional GANs to convert all the early dynamic frames to the last frame as a pre-processing step of motion correction (22; 23), but without sufficient guidance on input time, the method requires training multiple models specified for each early time point, which requires impractical time and memory consumption and limits its generalization. As a concurrent method investigating multiple-pair image conversion, a temporally and anatomically informed GAN was proposed for early-to-late frame conversion to improve cardiac PET motion correction (7), outperforming the single-pair benchmark and multiple-pair vanilla GAN model. However, the proposed method of introducing temporal and anatomical information requires additional pre-processing steps specifically designed for cardiac Rubidium-82 imaging, which might not be feasible for other tracers such as FDG. For both concurrent methods, both the input and generated frames were intensity-normalized since the image generation will only provide intermediate results to assist the following motion correction step, but not represent the actual tracer activities. The investigation of predicting the *actual SUV activities* at the reference time using a data-driven generative method is still lacking. Thus, the network for uptake time correction is expected to (a) correct multiple input uptake times through one well-trained model, (b) with input time awareness, and (c) predict frames with intensities representing actual radiotracer uptake activities that can be directly used for SUV evaluation.

In this work, we propose to use a time-aware generative adversarial network for uptake time correction and SUV harmonization in dynamic PET. The generator is modified to be time-aware through uptake time embedding and temporal 2.5D encoding. To the best of our knowledge, this is the first work using a generative method with time-awareness that predicts actual SUV activities and harmonizes radiotracer uptake time.

2 Methods

2.1 Proposed network

The structure of the proposed time-aware network for uptake time correction is shown in Figure 1. The generator is developed based on a 2-D U-Net structure (21), consisting of four encoding and decoding levels, to predict the frame with the reference uptake time based on the input frame with a different scan time. The generator network has been modified to incorporate temporal encoding to be

time-aware. Next, the real and generated late frames are paired with the corresponding input frame respectively and sent into the discriminator to be distinguished as either real or synthetic pairs. The discriminator utilizes PatchGAN (14) with three encoding levels and one linear output layer.

2.1.1 Time-aware embedding and feature modulation

To account for the considerable tracer distribution variability across different uptake times, we incorporate time-aware embedding related to temporal information and the tracer dynamics into the generator. Specifically, this is achieved by combining the uptake time of the current input frame and that of the two adjacent dynamic frames, i.e., t, t_{left}, and t_{right} respectively. This time vector is embedded by a Time2Vec layer (15) to a learnable feature representation of input time. Subsequently, given the recent achievements in the application of Long Short-Term Memory (LSTM) networks (12) to analyze 1-D medical data sequences (5; 8), an LSTM layer is utilized to further encode this vector of time representation and tasked with mapping the embeddings to the channel-specific parameters denoted as Γ and B. These parameters are subsequently utilized in the Feature-wise Linear Modulation (FiLM) layer (19) inserted at the bottleneck of the generator, as described in equation (1), to introduce time awareness,

$$FiLM(X_j) = \Gamma_j \cdot X_j + B_j, \tag{1}$$

where for the the j^{th} channel of the feature map at the bottleneck X_j, Γ_j and B_j are respectively the two modulation parameters.

2.1.2 Temporal 2.5D encoding

The input of the generator is a dynamic frame acquired at a different uptake time concatenated with the two adjacent frames as a temporal 2.5D encoding on the channel dimension to provide the generator with a richer and more informative input for uptake time awareness. By including the two adjacent frames in addition to the input dynamic frame, the generator has the ability to analyze the temporal context and fuse the enriched temporal information of the dynamic sequence. The tracer uptake distribution change underlying this time interval is also encoded by the generator. This design will also align the network's temporal receptive field with the 15-minute acquisition time of a static SUV frame in clinical practice. This improves the robustness and performance of the network by calibrating uptake increases or decreases and preventing false estimations, giving the prediction results closer to the standard uptake time.

2.1.3 The loss function

The loss function takes account of both the adversarial loss from the classification loss of the discriminator and the similarity loss from the mean squared error (MSE) loss of the generator computed voxel-by-voxel, expressed in equations (2) to (4) as follows:

$$L_{cls} = -log(D(F_R)) - log(1 - D(G(F_j))), \tag{2}$$

$$L_{sim} = \frac{1}{N}\sum_{n=1}^{N}(G(F_j)_n - (F_R)_n)^2, \tag{3}$$

$$L_{total} = L_{cls} + L_{sim}, \tag{4}$$

where L_{cls} represents the adversarial classification loss, L_{sim} is the MSE similarity loss, D denotes the discriminator, G stands for the generator, F_R is the real reference frame, $G(F_j)$ represents the generator-mapped reference frame from the j^{th} input frame F_j, and N represents the number of voxels in each frame. $\hat{G}$ and $\hat{D}$ are obtained by maximizing L_{cls} while simultaneously minimizing $L_{cls} + L_{mse}$ respectively during the adversarial training process.

2.2 Dataset

22 subjects were included in this study. Each subject underwent a dynamic whole-body FDG PET scan on a Biograph Vision (Siemens Healthineers) PET/CT scanner. Following an initial single-bed thoracic scan for the first 6 minutes post-injection, a total of 16 whole-body continuous bed motion (CBM) passes (frames) were reconstructed over the period extending to 70 minutes post-injection.

The dynamic frames are reconstructed under TrueX+TOF, 6 iterations, 5 subsets, no filtering, and relative scatter correction (4), consisting of seven 2-min passes followed by nine 5-min passes. The reconstructed CBM passes have a voxel size of 1.65 $\times$ 1.65 $\times$ 3 mm^3 and resolution of 440 $\times$ 440 on the transverse plane and the height-dependent number of slices in the inferior to superior direction. The SUV standardization is calculated as in (5)-(6),

$$c_{inj} = \frac{ID}{BW}, \tag{5}$$

$$SUV = \frac{c_{img}}{c_{inj}}, \tag{6}$$

where c_{inj} is the whole-body injected radioactivity concentration, ID is the injected dose in the unit of Bq, BW is the body weight in the unit of g, and c_{img} is the calibrated image intensity as the tracer concentration in the unit of Bq/mL. All the PET dynamic frames were well-registered using an in-house non-rigid registration package to prevent inter-frame spatial misalignments. A low-dose CT scan was first performed prior to the PET acquisition and then reconstructed and well-registered to the PET frames for further evaluation.

2.3 Model training and evaluation

As a comparison with the single-pair PET generation state-of-the-art method (22), we trained one baseline GAN model using a 2-D U-Net generator converting frames acquired at 40 minutes to the reference 60-minute frame. Training single-pair conversion models for all frame combinations wasn't deemed practical so only a single conversion pair was considered. As a comparison with other popular architectures with multiple-pair conversions, we included a 2-D U-Net, a baseline GAN, and a denoising diffusion probabilistic model (DDPM, (11)) as the benchmark comparisons, following the same pre-processing steps. To comprehensively assess the introduced temporal information and time-aware techniques, we included two variants of GAN with either time-aware embedding only or with the temporal 2.5D encoding only for evaluation. All the 5-minute frames were included for model training to minimize the effect of frame duration-related noise level mismatch in uptake time correction. Prior to model input, all frame intensities were converted to SUV activities. Due to the limitation of GPU memory, the training and evaluation were based on 2D slices, and the 3D volumes were then constructed by concatenating all the 2D slices. A random 80/20 split was applied subject-wise for training and evaluation, resulting in 5,213 input pairs in training and 1,349 pairs in evaluation. A central cropping of voxel size [320,320] is implemented to exclude background, and data augmentation with random translation in the range of [-5,5] voxels, and random rotation in the range of [-45°,45°] was applied during training. We developed all the models using PyTorch employing the Adam optimizer with learning rates G=2 $\times$ 10^{-4} and D=1 $\times$ 10^{-4}.

For qualitative evaluation, we visualized the generated and the real 60-minute frames for comparison. The absolute error maps and joint histograms were also visualized. For quantitative analysis, we calculated the Mean Squared Error (MSE), Normalized Mutual Information (NMI), Peak Signal-to-Noise Ratio (PSNR), and Structural Similarity Index (SSIM) between the generated and actual reference frames. The paired two-tailed t-test was conducted for the significance test. We also plotted the distributions of SUVs and SURs along with the comparison with the analytical correction method proposed by (29) in whole-body and in regions of interest (ROIs). The ROI segmentations of organs were automatically extracted using the automated learning and parsing of human anatomy (ALPHA) package developed by Siemens Healthineers (24) from the CT scans that are well-aligned with the PET frames. The ROI segmentations of organs were manually extracted from a nuclear medicine expert for further analysis. The nuclear medicine expert also carefully read the original and generated dynamic frames and reported reading impressions.

3 Results

3.1 Visualizations of generative time correction

The SUV frames at the input time, reference time, and corrected time are visualized in Figure 2. Compared to the 40-minute input frame, the 60-minute reference frame has lower tracer concentration in the liver and clearer boundary in the heart, with a higher hotspot contrast ratio to the background. The method of Sundar et al. (22) produced a result closer to the input frame instead of the reference

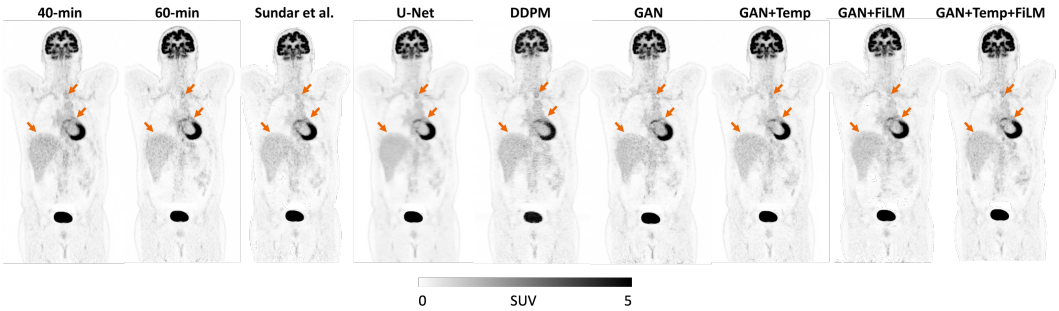

Figure 2: A visualization comparison of the SUV frames before and after uptake time correction.

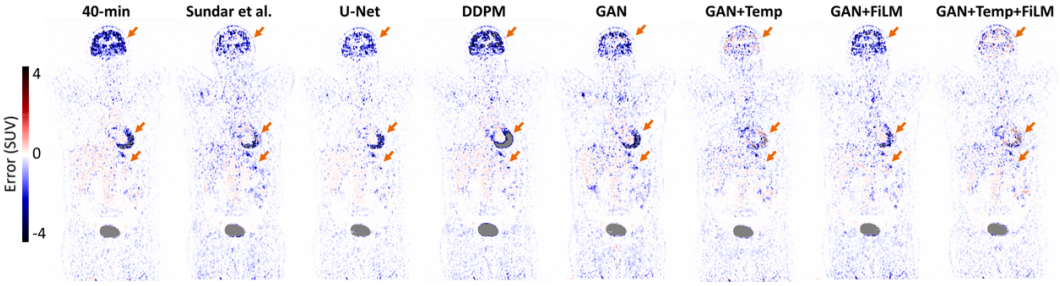

Figure 3: Difference maps of SUVs before and after uptake time correction.

frame, likely due to the limited sample size in the training set under the most specific single-pair mapping. For all the multiple-pair baselines, the result of U-Net has an artifact of over-smoothness, and DDPM introduced artifacts likely related to the Gaussian noise. Since the forward and inverse processes of DDPM require Gaussian noises N(0,1), the Diffusion Models might perform better in natural images with a fixed or standardized intensity range [-1,1] but might not gain comparable performance when predicting the actual SUV intensities, where the intensity range varies from 0 to $>$100. Besides, due to the iterations of the inverse process including typically 1000 steps for one 2D slice, the inference time of DDPM for constructing a 3D frame ($\sim$8 hours) is substantially longer than other CNN-based methods ($\sim$1 minute), making the application of DDPM less practical. Our evaluation of DDPM revealed limitations in performance and applicability to PET imaging tasks due to the inherent characteristics of PET data and the prohibitive computational demands of these models. Despite these challenges, we remain open to exploring the potential of Diffusion Models, specifically addressing their limitations in future research. The GAN-produced frame has the most realistic visual characteristics, but without sufficient temporal guidance, the boundaries of the hotspot and the upper liver are still unclear. After introducing either temporal 2.5D encoding or time-point embedding, the generation result was further improved. The proposed time-aware model with both techniques gave the result closest to the 60-minute reference frame. Due to memory limitations, all the current methods are under 2D training using axial slices, resulting in potential slice inconsistency and stitching artifacts in the visualization of coronal views. Under the same setting of training, the proposed method showed its improvement in reducing this slice-inconsistency-related artifact.

The SUV difference maps before and after each uptake time correction method are shown in Figure 3. The frame at 40-minute generally has a lower uptake than the 60-minute reference frame. After combining the temporal 2.5D encoding and the time awareness, the proposed model generally reduced the high SUV differences, especially in the brain, heart, and gastrointestinal tract as highlighted by the arrows.

The joint histograms between the reference frame and the frame before and after each uptake time correction method are shown in Figure 4. Note that the joint histogram of DDPM showed distorted distributions, potentially indicating the lack of ability to recover SUV intensities in a wide range. After uptake time correction using the proposed model, the joint histogram distributions are closest to the identity line.

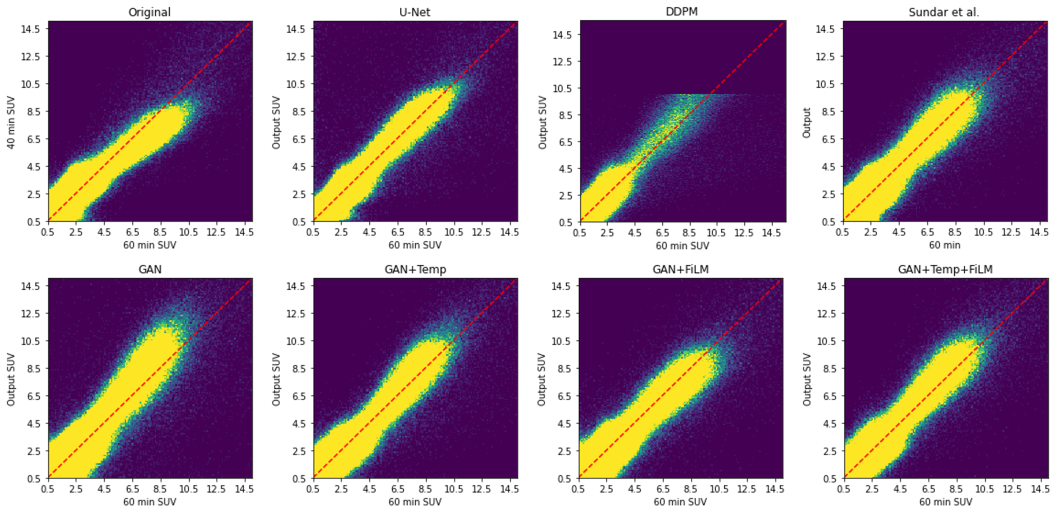

Figure 4: The joint histograms between the reference frame and the frame before and after each uptake time correction method. The distributions of the joint histograms after uptake time correction by the proposed model are closest to the identity line.

3.2 Quantitative metrics of image similarity

Table 1: Quantitative assessment of image similarity measurements for single-pair uptake time correction from 40-minute input (mean ± standard deviation) with the best results listed **in bold**.

Methods	MSE	SSIM	PSNR	NMI
Original (40-minute)	0.474±0.415*	0.872±0.025	42.72±3.57*	0.616±0.038*
Sundar et al.	0.492±0.344*	0.850±0.025*	42.06±3.12*	**0.958±0.004**
U-Net	0.401±0.259*	**0.879±0.023**	42.83±3.00*	0.955±0.0042*
DDPM	1.984±1.774*	0.851±0.027*	37.75±0.97*	0.956±0.0043*
GAN	0.566±0.404*	0.850±0.027*	41.49±3.50*	**0.958±0.004**
GAN+Temp	0.382±0.312*	0.854±0.027*	43.45±3.07*	**0.958±0.004**
GAN+FiLM	0.413±0.169*	0.850±0.026*	42.32±3.31*	**0.958±0.004**
GAN+Temp+FiLM (Proposed)	**0.345±0.223**	0.861±0.027	**43.50±3.34**	**0.958±0.004**

*$P < 0.05$ between the current method and the proposed method (paired two-tailed t-test).

Table 2: Quantitative assessment of image similarity measurements for multiple-pair uptake time correction from 25 min to 65 min input (mean ± standard deviation) with the best results listed **in bold**.

Methods	MSE	SSIM	PSNR	NMI
Original (25-65 min)	0.498±0.525*	0.871±0.026*	42.99±4.61*	0.613±0.037*
U-Net	0.431±0.376*	0.879±0.024*	43.14±4.49*	0.956±0.0042*
DDPM	2.021±1.782*	0.850±0.028*	37.67±1.34*	0.956±0.0042*
GAN	0.519±0.413*	0.846±0.027*	41.99±4.18*	**0.958±0.004**
GAN+Temp	0.317±0.341*	0.883±0.059*	46.48±7.68*	**0.958±0.004**
GAN+FiLM	0.447±0.328*	0.847±0.028*	42.57±4.45*	0.957±0.005*
GAN+Temp+FiLM (Proposed)	**0.288±0.232**	**0.894±0.064**	**46.72±8.16**	**0.958±0.004**

*$P < 0.05$ between the current method and the proposed method (paired two-tailed t-test).

In Table 1, quantitative evaluations of single-pair uptake time correction from the 40-minute input to the 60-minute reference frame are presented. The method of Sundar et al. (22) is the most specific uptake time correction pair (40-minute to 60-minute) but didn't achieve desirable quantitative results, possibly related to the insufficient sample size and lack of temporal information in dynamic PET. Although the baseline U-Net model achieved lower MSE and higher SSIM and PSNR than

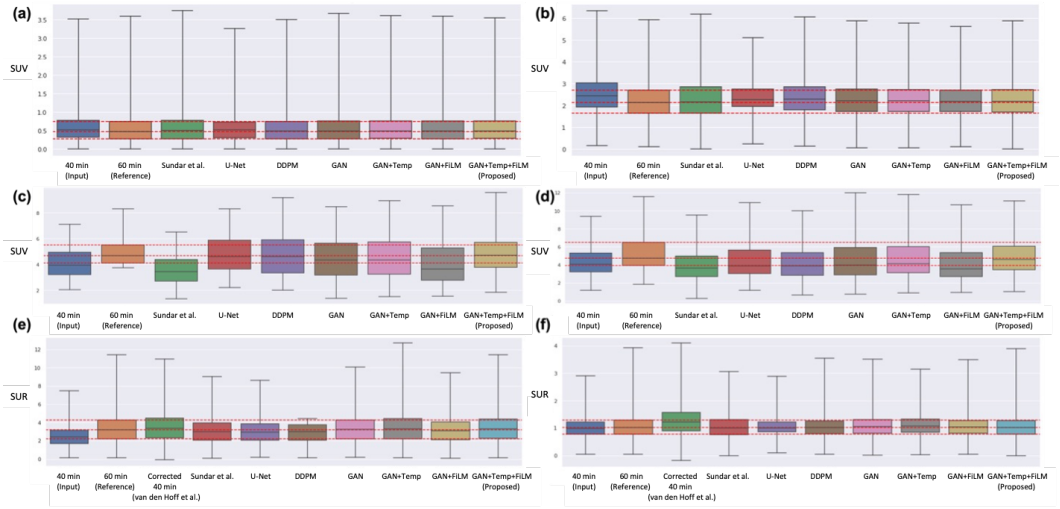

Figure 5: The SUV distributions of (a) whole body, (b) liver, (c) a sample lesion, and (d) all the lesions before and after each uptake time correction method, and the SUR distributions of (e) brain and (f) liver after each analytical and generative uptake time correction method.

other baseline generative models DDPM and GAN, considering the over-smoothness in the U-Net predictions, the visual quality is not satisfactory compared to GAN. Even though this subset of data contains a specific time difference in mapping, adding either temporal 2.5D encoding or time-aware embedding is able to enhance the ability of the network to estimate the reference point SUVs closer to the real 60-minute data. Table 2 provides a comprehensive quantitative assessment of image similarity in uptake time correction from 25-minute input to 65-minute input. Similarly, after adding either temporal 2.5D encoding or time-aware embedding, the quantitative measurements are improved. Notably, the proposed approach with both temporal 2.5D encoding and time-aware embedding consistently demonstrates superior performance compared to other methods and achieved statistically significant differences in every metric compared to the state-of-the-art single-pair method and in MSE and PSNR compared to all the other multiple-pair baselines, potentially suggesting the efficacy and robustness of the proposed method.

3.3 Distributions of SUV and SUR values

Figure 5(a)-(d) shows the SUV distributions in the whole body, liver, a sample lesion, and all the lesions before and after uptake time correction. The SUV distributions are slightly different between the 40-minute and 60-minute frames for the whole body. Both the single-pair baseline (22) and the multiple-pair U-Net baseline showed a discrepancy in the outputs for both the interquartile range and the minimum and maximum values. Both the DDPM and GAN baselines corrected the interquartile range but there are still errors for the minimum and maximum values. Although GAN+Temp and GAN+FiLM showed comparable SUV distributions with the proposed GAN+Temp+FiLM, for the organ of interest (e.g., the liver), these two variants showed residual SUV correction errors. The correction bias is the most sensitive in the lesions since it was originally segmented on the reference dynamic frame and it's also the most sensitive to minor mislocation or shape difference errors in uptake time correction. The proposed method achieved the interquartile range closest to the reference time both for the sample lesion and across all the lesions.

Figure 5(e)-(f) presents the SUR distributions in the brain and liver before and after each analytical and generative uptake time correction. For the brain following the irreversible assumption of FDG, the analytical method of SUR correction proposed by (28) was able to achieve comparable results to the 60-minute reference, but in the liver where the irreversible assumption isn't applicable, this method shows its limitation. Similarly, for both organs regardless of this non-irreversible assumption, the proposed model was able to correct the SUR values to the reference time for both interquartile range and minimum and maximum values, outperforming other baselines and two variants.

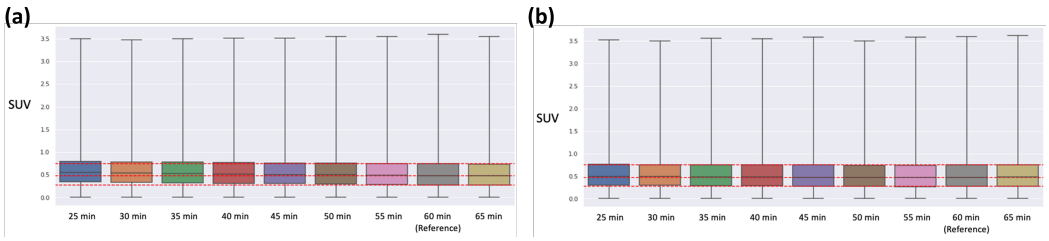

Figure 6: The whole-body SUV distributions over time (a) without uptake time correction and (b) after uptake time correction using the proposed method.

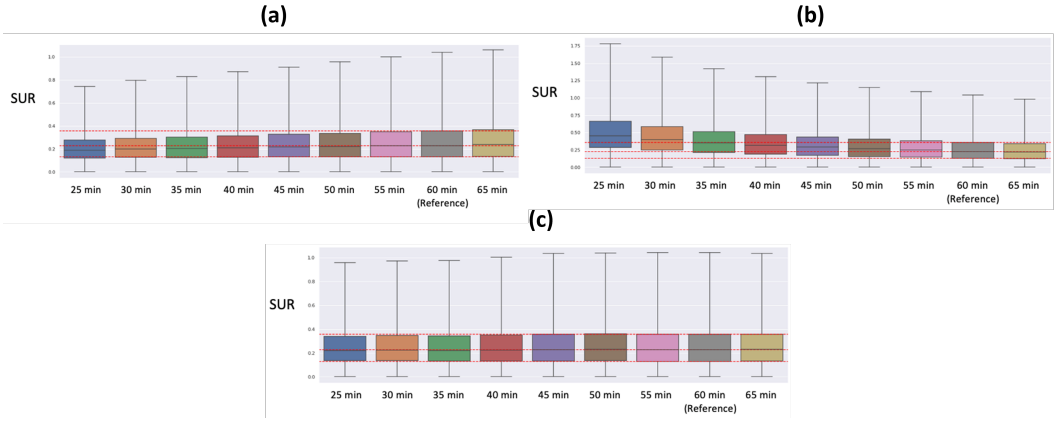

Figure 7: The whole-body SUR distributions over time (a) without uptake time correction, (b) after analytical uptake time correction, and (c) after uptake time correction using the proposed method.

The boxplots of SUV and SUR distribution change across the time before and after uptake time correction by the proposed model are shown in Figure 7 and Figure 6. Before uptake time correction, the SUV distributions show a general decay across the time, and after uptake time correction, the SUV distributions are all consistent with the reference frame at 60 minutes. In contrast, the SUR values generally increase over time. Due to the inconsistency of the irreversible assumption of FDG across different body organs, the analytical uptake time correction method shows unsatisfactory results. The proposed generative method was able to correct SUR values acquired at different uptake time points.

3.4 Reading impressions from the nuclear medicine expert

Figure 8 showed two sample slices comparing uptake time correction results of the proposed method with the real reference frame at 60 minutes. In Figure 8(a), the input frame is at 40 minutes. It's reported that both the 40-minute input frame and the 60-minute reference frame are noisy, but the 40-minute frame exhibits different local details. In the input frame at 40 minutes, though the relevant findings are all available, the SUVs are generally lower, expected due to an earlier scan time, which might introduce quantification error. The lesion in the anterior mediastinum is visible at both time points, and this lesion is well-recovered in the generated frame. In Figure 8(b), the input scan is at 25 min. At this earlier scan time, the tracer uptake and filling patterns show the time dependency, with higher liver uptake and lower tumor and brain uptake. The shape of the lesion in the abdomen is also slightly different in this 25-minute frame. The generated frame reduced the blood pool and liver uptake reasonably. The lesion in the generated frame is recovered with an increase of uptake, though with minor changes of form, there is no change of report or interpretation in the reading impression. An artificial-like hot spot can be found in the generated frame adjacent to the original lesion, possibly to be reported as true findings.

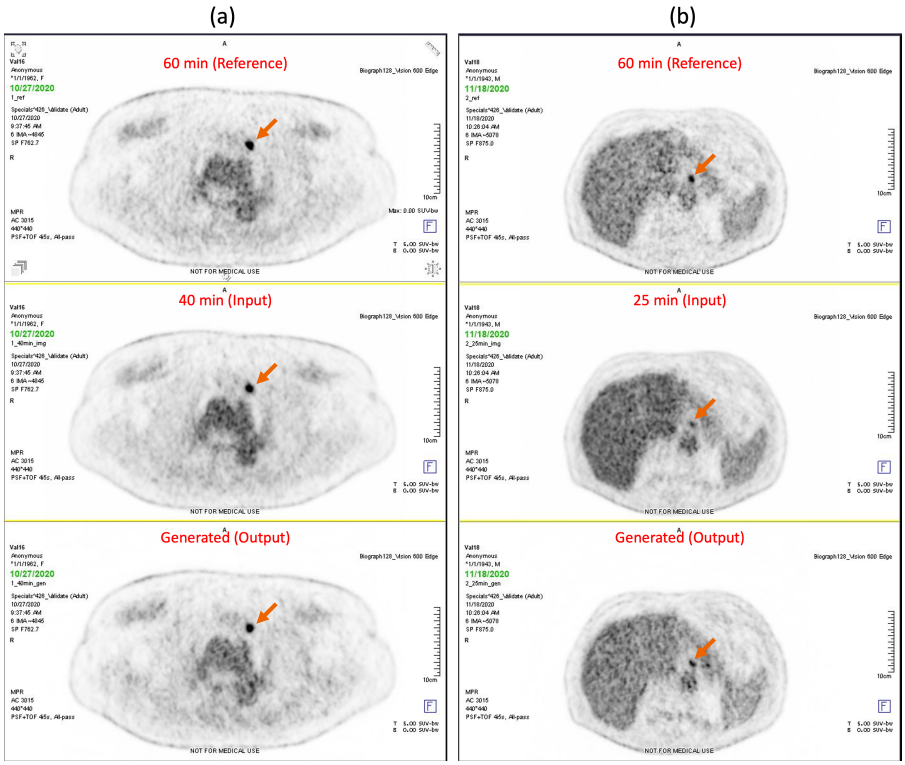

Figure 8: Two sample slices comparing uptake time correction results of the proposed method with input frame and the real reference frame, with pointed lesions from the nuclear medicine expert reading impressions.

4 Conclusion and Future Directions

In conclusion, we propose a generative method under adversarial training with auxiliary temporal 2.5D encoding and time-aware embedding techniques to correct the uptake time of dynamic PET frames for SUV and SUR harmonization. The proposed network is able to map the dynamic PET frame acquired at a different time to the 60-minute reference time, with high image similarity both qualitatively and quantitatively. The proposed model demonstrates its superiority in mapping the SUV and SUR distributions to the reference compared to other analytical and generative baselines. Reading impressions from the nuclear medicine expert reported comparative visualization and noise patterns as well as recovered lesions without change of interpretation, but the possibility of introducing a few false positives warrants further improvement of the trustworthiness and robustness of the network. The proposed network can potentially improve current clinical workflows by harmonizing and shortening uptake time.

Several future directions are worth investigating. First, to avoid slice inconsistency, if the memory limitation is overcome, we will include spatial 2.5D encoding by concatenating multiple adjacent slices as the additional information or train a 3D network to ensure the receptive field of the network covers the whole body. Second, the current design of bottleneck FiLM and Time2Vec embedding aligns with the best practice and concurrent research (7; 18; 9; 25), balancing model complexity and performance. Future work includes investigating alternative configurations of FiLM and time-awareness. Also, following the collaboration with the nuclear medicine expert, we will conduct a downstream oncological analysis to evaluate the impact and potential for clinical decision-making. Last, since dynamic PET is still an emerging application, current method is developed and validated on an in-house dataset due to the challenges in acquiring large-scale data. Future work will cover evaluations on a larger number of cases including a wider range of data (20-120 minutes) and variations in scanning protocols.

References

[1] Sylvain Beaulieu, Paul Kinahan, Jeffrey Tseng, Lisa K Dunnwald, Erin K Schubert, Pam Pham, Barbara Lewellen, and David A Mankoff. Suv varies with time after injection in 18f-fdg pet of breast cancer: characterization and method to adjust for time differences. *Journal of Nuclear Medicine*, 44(7):1044–1050, 2003.

[2] Lei Bi, Jinman Kim, Ashnil Kumar, Dagan Feng, and Michael Fulham. Synthesis of positron emission tomography (pet) images via multi-channel generative adversarial networks (gans). In *Molecular Imaging, Reconstruction and Analysis of Moving Body Organs, and Stroke Imaging and Treatment: Fifth International Workshop, CMMI 2017, Second International Workshop, RAMBO 2017, and First International Workshop, SWITCH 2017, Held in Conjunction with MICCAI 2017, Québec City, QC, Canada, September 14, 2017, Proceedings 5*, pages 43–51. Springer, 2017.

[3] Özgün Çiçek, Ahmed Abdulkadir, Soeren S Lienkamp, Thomas Brox, and Olaf Ronneberger. 3d u-net: learning dense volumetric segmentation from sparse annotation. In *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2016: 19th International Conference, Athens, Greece, October 17-21, 2016, Proceedings, Part II 19*, pages 424–432. Springer, 2016.

[4] André H Dias, Anne M Smith, Vijay Shah, David Pigg, Lars C Gormsen, and Ole L Munk. Clinical validation of a population-based input function for 20-min dynamic whole-body 18f-fdg multiparametric pet imaging. *EJNMMI physics*, 9(1):1–18, 2022.

[5] Nicha C Dvornek, Pamela Ventola, Kevin A Pelphrey, and James S Duncan. Identifying autism from resting-state fmri using long short-term memory networks. In *International Workshop on Machine Learning in Medical Imaging*, pages 362–370. Springer, 2017.

[6] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial networks. *Communications of the ACM*, 63(11):139–144, 2020.

[7] Xueqi Guo, Luyao Shi, Xiongchao Chen, Bo Zhou, Qiong Liu, Huidong Xie, Yi-Hwa Liu, Richard Palyo, Edward J. Miller, Albert J. Sinusas, Bruce Spottiswoode, Chi Liu, and Nicha C. Dvornek. 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.

[8] Xueqi Guo, Sule Tinaz, and Nicha C Dvornek. Characterization of early stage parkinson's disease from resting-state fmri data using a long short-term memory network. *Frontiers in Neuroimaging*, 1:952084, 2022.

[9] Xueqi Guo, Bo Zhou, David Pigg, Bruce Spottiswoode, Michael E Casey, Chi Liu, and Nicha C Dvornek. Unsupervised inter-frame motion correction for whole-body dynamic pet using convolutional long short-term memory in a convolutional neural network. *Medical Image Analysis*, 80:102524, 2022.

[10] Leena M Hamberg, George J Hunter, Nathaniel M Alpert, Noah C Choi, John W Babich, and Alan J Fischman. The dose uptake ratio as an index of glucose metabolism: useful parameter or oversimplification? *Journal of Nuclear Medicine*, 35(8):1308–1312, 1994.

[11] Jonathan Ho, Ajay Jain, and Pieter Abbeel. Denoising diffusion probabilistic models. *Advances in neural information processing systems*, 33:6840–6851, 2020.

[12] Sepp Hochreiter and Jürgen Schmidhuber. Long short-term memory. *Neural computation*, 9(8):1735–1780, 1997.

[13] Jyoti Islam and Yanqing Zhang. Gan-based synthetic brain pet image generation. *Brain informatics*, 7:1–12, 2020.

[14] Phillip Isola, Jun-Yan Zhu, Tinghui Zhou, and Alexei A Efros. Image-to-image translation with conditional adversarial networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1125–1134, 2017.

[15] Seyed Mehran Kazemi, Rishab Goel, Sepehr Eghbali, Janahan Ramanan, Jaspreet Sahota, Sanjay Thakur, Stella Wu, Cathal Smyth, Pascal Poupart, and Marcus Brubaker. Time2vec: Learning a vector representation of time. *arXiv preprint arXiv:1907.05321*, 2019.

[16] Juan Liu, Sijin Ren, Rui Wang, Niloufarsadat Mirian, Yu-Jung Tsai, Michal Kulon, Darko Pucar, Ming-Kai Chen, and Chi Liu. Virtual high-count pet image generation using a deep learning method. *Medical Physics*, 49(9):5830–5840, 2022.

[17] Tianshun Miao, Bo Zhou, Juan Liu, Xueqi Guo, Qiong Liu, Huidong Xie, Xiongchao Chen, Ming-Kai Chen, Jing Wu, Richard E Carson, et al. Generation of whole-body fdg parametric ki images from static pet images using deep learning. *IEEE Transactions on Radiation and Plasma Medical Sciences*, 2023.

[18] Tony CW Mok and Albert CS Chung. Conditional deformable image registration with convolutional neural network. In *Medical Image Computing and Computer Assisted Intervention–MICCAI 2021: 24th International Conference, Strasbourg, France, September 27–October 1, 2021, Proceedings, Part IV 24*, pages 35–45. Springer, 2021.

[19] Ethan Perez, Florian Strub, Harm De Vries, Vincent Dumoulin, and Aaron Courville. Film: Visual reasoning with a general conditioning layer. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 32, 2018.

[20] Sijin Ren, Juan Liu, Huidong Xie, Takuya Toyonaga, Niloufarsadat Mirian, Ming-Kai Chen, Mariam Aboian, Richard Carson, and Chi Liu. Super-resolution pet brain imaging using deep learning. In *2021 IEEE Nuclear Science Symposium and Medical Imaging Conference (NSS/MIC)*, pages 1–6. IEEE, 2021.

[21] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: 18th International Conference, Munich, Germany, October 5-9, 2015, Proceedings, Part III 18*, pages 234–241. Springer, 2015.

[22] Lalith Kumar Shiyam Sundar, David Iommi, Otto Muzik, Zacharias Chalampalakis, Eva-Maria Klebermass, Marius Hienert, Lucas Rischka, Rupert Lanzenberger, Andreas Hahn, Ekaterina Pataraia, et al. Conditional generative adversarial networks aided motion correction of dynamic 18f-fdg pet brain studies. *Journal of Nuclear Medicine*, 62(6):871–879, 2021.

[23] Lalith Shiyam Sundar, David Iommi, Benjamin Spencer, Qian Wang, Simon Cherry, Thomas Beyer, and Ramsey Badawi. Data-driven motion compensation using cgan for total-body [18f] fdg-pet imaging, 2021.

[24] Yimo Tao, Zhigang Peng, Arun Krishnan, and Xiang Sean Zhou. Robust learning-based parsing and annotation of medical radiographs. *IEEE transactions on medical imaging*, 30(2):338–350, 2010.

[25] Adam G Tattersall, Keith A Goatman, Lucy E Kershaw, Scott IK Semple, and Sonia Dahdouh. Improving style transfer in dynamic contrast enhanced mri using a spatio-temporal approach. In *International Workshop on Simulation and Synthesis in Medical Imaging*, pages 97–106. Springer, 2023.

[26] Joseph A Thie, Karl F Hubner, and Gary T Smith. Optimizing imaging time for improved performance in oncology pet studies. *Molecular Imaging & Biology*, 4(3):238–244, 2002.

[27] Giampaolo Tomasi, Federico Turkheimer, and Eric Aboagye. Importance of quantification for the analysis of PET data in oncology: review of current methods and trends for the future. *Molecular Imaging and Biology*, 14:131–146, 2012.

[28] Jörg van den Hoff, Alexandr Lougovski, Georg Schramm, Jens Maus, Liane Oehme, Jan Petr, Bettina Beuthien-Baumann, Jörg Kotzerke, and Frank Hofheinz. Correction of scan time dependence of standard uptake values in oncological pet. *EJNMMI research*, 4(1):1–14, 2014.

[29] Jörg van den Hoff, Liane Oehme, Georg Schramm, Jens Maus, Alexandr Lougovski, Jan Petr, Bettina Beuthien-Baumann, and Frank Hofheinz. The pet-derived tumor-to-blood standard uptake ratio (sur) is superior to tumor suv as a surrogate parameter of the metabolic rate of fdg. *EJNMMI research*, 3(1):1–8, 2013.

[30] RL Wahl and JW Buchanan. Principles and practice of positron emission tomography. *Principles and Practice of Positron Emission Tomography (Lippincott, Williams and Wilkins, Philadelphia, PA, 2002)*, 2002.

[31] Bo Zhou, Rui Wang, Ming-Kai Chen, Adam P Mecca, Ryan S O'Dell, Christopher H Van Dyck, Richard E Carson, James S Duncan, and Chi Liu. Synthesizing multi-tracer pet images for alzheimer's disease patients using a 3d unified anatomy-aware cyclic adversarial network. In *Medical Image Computing and Computer Assisted Intervention–MICCAI 2021: 24th International Conference, Strasbourg, France, September 27–October 1, 2021, Proceedings, Part VI 24*, pages 34–43. Springer, 2021.