Contrastive Diffusion Model with Auxiliary Guidance for Coarse-to-Fine PET Reconstruction

Zeyu Han[1], Yuhan Wang[1], Luping Zhou[2], Peng Wang[1], Binyu Yan[1], Jiliu Zhou[1,3], Yan Wang[1](✉), and Dinggang Shen[4,5](✉)

[1] School of Computer Science, Sichuan University, Chengdu, China
wangyanscu@hotmail.com

[2] School of Electrical and Information Engineering, University of Sydney, Sydney, Australia

[3] School of Computer Science, Chengdu University of Information Technology, Chengdu, China

[4] School of Biomedical Engineering, ShanghaiTech University, Shanghai, China
dinggang.shen@gmail.com

[5] Department of Research and Development, Shanghai United Imaging Intelligence Co., Ltd., Shanghai, China

Abstract. To obtain high-quality positron emission tomography (PET) scans while reducing radiation exposure to the human body, various approaches have been proposed to reconstruct standard-dose PET (SPET) images from low-dose PET (LPET) images. One widely adopted technique is the generative adversarial networks (GANs), yet recently, diffusion probabilistic models (DPMs) have emerged as a compelling alternative due to their improved sample quality and higher log-likelihood scores compared to GANs. Despite this, DPMs suffer from two major drawbacks in real clinical settings, i.e., the computationally expensive sampling process and the insufficient preservation of correspondence between the conditioning LPET image and the reconstructed PET (RPET) image. To address the above limitations, this paper presents a coarse-to-fine PET reconstruction framework that consists of a coarse prediction module (CPM) and an iterative refinement module (IRM). The CPM generates a coarse PET image via a deterministic process, and the IRM samples the residual iteratively. By delegating most of the computational overhead to the CPM, the overall sampling speed of our method can be significantly improved. Furthermore, two additional strategies, i.e., an auxiliary guidance strategy and a contrastive diffusion strategy, are proposed and integrated into the reconstruction process, which can enhance the correspondence between the LPET image and the RPET image, further improving clinical reliability. Extensive experiments on two human brain PET datasets demonstrate that our

Z. Han and Y. Wang—These authors contributed equally to this work.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/978-3-031-43999-5_23.

H. Greenspan et al. (Eds.): MICCAI 2023, LNCS 14229, pp. 239–249, 2023.
https://doi.org/10.1007/978-3-031-43999-5_23

method outperforms the state-of-the-art PET reconstruction methods. The source code is available at https://github.com/Show-han/PET-Reconstruction.

Keywords: Positron emission tomography (PET) · PET reconstruction · Diffusion probabilistic models · Contrastive learning

1 Introduction

Positron emission tomography (PET) is a widely-used molecular imaging technique that can help reveal the metabolic and biochemical functioning of body tissues. According to the dose level of injected radioactive tracer, PET images can be roughly classified as standard-(SPET) and low-dose PET (LPET) images. SPET images offer better image quality and more information in diagnosis compared to LPET images containing more noise and artifacts. However, the higher radiation exposure associated with SPET scanning poses potential health risks to the patient. Consequently, it is crucial to reconstruct SPET images from corresponding LPET images to produce clinically acceptable PET images.

In recent years, deep learning-based PET reconstruction approaches [7,9,13] have shown better performance than traditional methods. Particularly, generative adversarial networks (GANs) [8] have been widely adopted [12,14,15,18,26,27] due to their capability to synthesize PET images with higher fidelity than regression-based models [29,30]. For example, Kand *et al.* [11] applied a Cycle-GAN model to transform amyloid PET images obtained with diverse radiotracers. Fei *et al.* [6] made use of GANs to present a bidirectional contrastive framework for obtaining high-quality SPET images. Despite the promising achievement of GAN, its adversarial training is notoriously unstable [22] and can lead to mode collapse [17], which may result in a low discriminability of the generated samples, reducing their confidence in clinical diagnosis.

Fortunately, likelihood-based generative models offer a new approach to address the limitations of GANs. These models learn the distribution's probability density function via maximum likelihood and could potentially cover broader data distributions of generated samples while being more stable to train. As an example, Cui *et al.* [3] proposed a model based on Nouveau variational autoencoder for PET image denoising. Among likelihood-based generative models, diffusion probabilistic models (DPMs) [10,23] are noteworthy for their capacity to outperform GANs in various tasks [5], such as medical imaging [24] and text-to-image generation [20]. DPMs consist of two stages: a forward process that gradually corrupts the given data and a reverse process that iteratively samples the original data from the noise. However, sampling from a diffusion model is computationally expensive and time-consuming [25], making it inconvenient for real clinical applications. Besides, existing conditional DPMs learn the input-output correspondence implicitly by adding a prior to the training objective, while this learned correspondence is prone to be lost in the reverse process [33], resulting in the RPET image missing crucial clinical information from the LPET image. Hence, the clinical reliability of the RPET image may be compromised.

Motivated to address the above limitations, in this paper, we propose a coarse-to-fine PET reconstruction framework, including a coarse prediction module (CPM) and an iterative refinement module (IRM). The CPM generates a coarse prediction by invoking a deterministic prediction network only once, while the IRM, which is the reverse process of the DPMs, iteratively samples the residual between this coarse prediction and the corresponding SPET image. By combining the coarse prediction and the predicted residual, we can obtain RPET images much closer to the SPET images. To accelerate the sampling speed of IRM, we manage to delegate most of the computational overhead to the CPM [2,28], hoping to narrow the gap between the coarse prediction and the SPET initially. Additionally, to enhance the correspondence between the LPET image and the generated RPET image, we propose an auxiliary guidance strategy at the input level based on the finding that auxiliary guidance can help to facilitate the reverse process of DPMs, and reinforce the consistency between the LPET image and RPET image by providing more LPET-relevant information to the model. Furthermore, at the output level, we suggest a contrastive diffusion strategy inspired by [33] to explicitly distinguish between positive and negative PET slices. To conclude, the contributions of our method can be described as follows:

- We introduce a novel PET reconstruction framework based on DPMs, which, to the best of our knowledge, is the first work that applies DPMs to PET reconstruction.
- To mitigate the computational overhead of DPMs, we employ a coarse-to-fine design that enhances the suitability of our framework for real-world clinical applications.
- We propose two novel strategies, i.e., an auxiliary guidance strategy and a contrastive diffusion strategy, to improve the correspondence between the LPET and RPET images and ensure that RPET images contain reliable clinical information.

2 Background: Diffusion Probabilistic Models

Diffusion Probabilistic Models (DPMs): DPMs [10,23] define a *forward process*, which corrupts a given image data $x_0 \sim q(x_0)$ step by step via a fixed Markov chain $q(x_t|x_{t-1})$ that gradually adds Gaussian noise to the data:

$$q(x_t|x_{t-1}) = \mathcal{N}(x_t; \sqrt{\alpha_t}x_{t-1}, (1-\alpha_t)I), t = 1, 2, \cdots, T, \tag{1}$$

where $\alpha_{1:T}$ is the constant variance schedule that controls the amount of noise added at each time step, and $q(x_T) \sim \mathcal{N}(x_T; 0, I)$ is the stationary distribution. Owing to the Markov property, a data x_t at an arbitrary time step t can be sampled in closed form:

$$q(x_t|x_0) = \mathcal{N}(x_t; \sqrt{\gamma_t}x_0, (1-\gamma_t)I); x_t = \sqrt{\gamma_t}x_0 + \sqrt{1-\gamma_t}\epsilon, \epsilon \sim \mathcal{N}(0, I), \tag{2}$$

where $\gamma_t = \prod_{i=1}^{t} \alpha_i$. Furthermore, we can derive the posterior distribution of x_{t-1} given (x_0, x_t) as $q(x_{t-1}|x_0, x_t) = \mathcal{N}(x_{t-1}; \hat{\mu}(x_0, x_t), \sigma_t^2 I)$, where $\hat{\mu}(x_0, x_t)$

and σ_t^2 are subject to x_0, x_t and $\alpha_{1:T}$. Based on this, we can leverage the *reverse process* from x_T to x_0 to gradually denoise the latent variables by sampling from the posterior distribution $q(x_{t-1}|x_0, x_t)$. However, since x_0 is unknown during inference, we use a transition distribution $p_\theta(x_{t-1}|x_t) := q(x_{t-1}|\mathcal{H}_\theta(x_t, t), x_t)$ to approximate $q(x_{t-1}|x_0, x_t)$, where $\mathcal{H}_\theta(x_t, t)$ manages to reconstruct x_0 from x_t and t, and it is trained by optimizing a variational lower bound of $log p_\theta(x)$.

Conditional DPMs: Given an image x_0 with its corresponding condition c, conditional DPMs try to estimate $p(x_0|c)$. To achieve that, condition c is concatenated with x_t [21] as the input of $\mathcal{H}_\theta$, denoted as $\mathcal{H}_\theta(c, x_t, t)$.

Simplified Training Objective: Instead of training $\mathcal{H}_\theta$ to reconstruct the x_0 directly, we use an alternative parametrization $\mathcal{D}_\theta$ named *denoising network* [10] trying to predict the noise vector $\epsilon \sim \mathcal{N}(0, I)$ added to x_0 in Eq. 2, and derive the following training objective:

$$\mathcal{L}_{DPM} = \mathbb{E}_{(c,x_0)\sim p_{train}} \mathbb{E}_{\epsilon\sim\mathcal{N}(0,I)} \mathbb{E}_{\gamma\sim p_\gamma} \|\mathcal{D}_\theta(c, \sqrt{\gamma}x_0 + \sqrt{1-\gamma}\epsilon, \gamma) - \epsilon\|_1, \quad (3)$$

where the distribution p_γ is the one used in WaveGrad [1]. Note that we also leverage techniques from WaveGrad to let the denoising network $\mathcal{D}_\theta$ conditioned directly on the noise schedule γ rather than time step t, and this gives us more flexibility to control the inference steps.

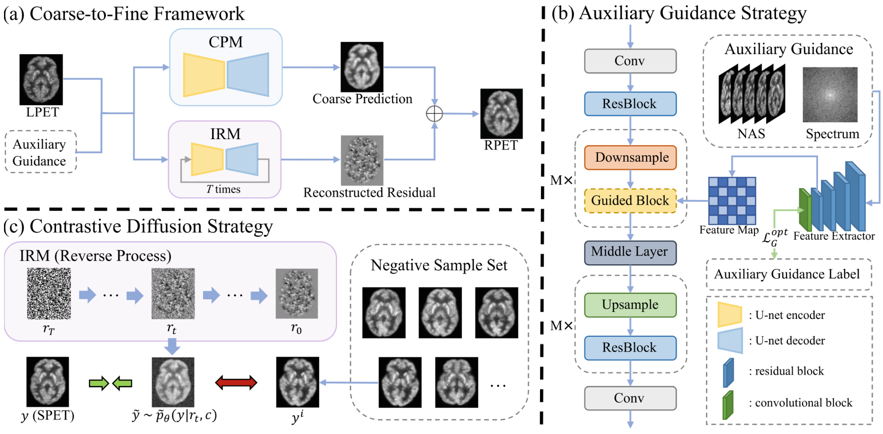

Fig. 1. Overall architecture of our proposed framework.

3 Methodology

Our proposed framework (Fig. 1(a)) has two modules, i.e., a coarse prediction module (CPM) and an iterative refinement module (IRM). The CPM predicts a coarse-denoised PET image from the LPET image, while the IRM models the residual between the coarse prediction and the SPET image iteratively. By combining the coarse prediction and residual, our framework can effectively generate high-quality RPET images. To improve the correspondence between the LPET image and the RPET image, we adopt an auxiliary guidance strategy (Fig. 1(b)) at the input level and a contrastive diffusion strategy (Fig. 1(c)) at the output level. The details of our method are described in the following subsections.

3.1 Coarse-to-Fine Framework

To simplify notation, we use a single conditioning variable c to represent the input required by both CPM and IRM, which includes the LPET image x_{lpet} and the auxiliary guidance x_{aux}. During inference, CPM first generates a coarse prediction $x_{cp} = \mathcal{P}_\theta(c)$, where $\mathcal{P}_\theta$ is the deterministic prediction network in CPM. The IRM, which is the reverse process of DPM, then tries to sample the residual r_0 (i.e., x_0 in Sect. 2) between the coarse prediction x_{cp} and the SPET image y via the following iterative process:

$$r'_{t-1} \sim p_\theta(r_{t-1}|r_t, c), t = T, T-1, \cdots, 1. \tag{4}$$

Herein, the prime symbol above the variable indicates that it is sampled from the reverse process instead of the forward process. When $t = 1$, we can obtain the final sampled residual r_0, and the RPET image y' can be derived by $r'_0 + x_{cp}$.

In practice, both CPM and IRM use the same network architecture shown in Fig. 1(c). CPM generates the coarse prediction x_{cp} by using $\mathcal{P}_\theta$ only once, but the denoising network $\mathcal{D}_\theta$ in IRM will be invoked multiple times during inference. Therefore, it is rational to delegate more computation overhead to $\mathcal{P}_\theta$ to obtain better initial results while keeping $\mathcal{D}_\theta$ small, since the reduction in computation cost in $\mathcal{D}_\theta$ will be accumulated by multiple times. To this end, we set the channel number in $\mathcal{P}_\theta$ much larger than that in the denoising network $\mathcal{D}_\theta$. This leads to a larger network size for $\mathcal{P}_\theta$ compared to $\mathcal{D}_\theta$.

3.2 Auxiliary Guidance Strategy

In this section, we will describe our auxiliary guidance strategy in depth which is proposed to enhance the reconstruction process at the input level by incorporating two auxiliary guidance, i.e., neighboring axial slices (NAS) and the spectrum. Our findings indicate that incorporating NAS provides insight into the spatial relationship between the current slice and its adjacent slices, while incorporating the spectrum imposes consistency in the frequency domain.

To effectively incorporate these two auxiliary guidances, as illustrated in Fig. 1(c), we replace the ResBlock in the encoder with a Guided ResBlock as done in [19]. During inference, the auxiliary guidance x_{aux} is first downsampled by a factor of 2^k as x^k_{aux}, where $k = 1, \cdots, M$, and M is the number of downsampling operations in the U-net encoder. Then x^k_{aux} is fed into a feature extractor $\mathcal{F}_\theta$ to generate its corresponding feature map $f^k_{aux} = \mathcal{F}_\theta(x^k_{aux})$, which is next injected into the Guided ResBlock matching its resolution through 1×1 convolution.

To empower the feature extractor to contain information of its high-quality counterpart y_{aux}, we constrain it with $\mathcal{L}_1$ loss through a convolution layer $\mathcal{C}_\theta(\cdot)$:

$$\mathcal{L}_G^{opt} = \sum_{k=1}^{M} \|\mathcal{C}_\theta(\mathcal{F}_\theta(x_{aux}^k)) - y_{aux}^k\|_1, \tag{5}$$

where *opt* ∈{NAS, spectrum} denotes the kind of auxiliary guidance.

3.3 Contrastive Diffusion Strategy

In addition to the auxiliary guidance at the input level, we also develop a contrastive diffusion strategy at the output level to amplify the correspondence between the condition LPET image and the corresponding RPET image. In detail, we introduce a set of negative samples $Neg = \{y^1, y^2, ..., y^N\}$, which consists of N SPET slices, each from a randomly selected subject that is not in the current batch for training. Then, for the noisy latent residual r_t at time step t, we obtain its corresponding intermediate RPET $\widetilde{y}$, and draw it close to the corresponding SPET y while pushing it far from the negative sample $y^i \in Neg$. Before this, we need to estimate the intermediate residual corresponding to r_t firstly, denoted as $\widetilde{r_0}$. According to Sect. 2, the denoising network $\mathcal{D}_\theta$ manages to predict the Gaussian noise added to r_0, enabling us to calculate $\widetilde{r_0}$ directly from r_t:

$$\widetilde{r_0} = \frac{r_t - (\sqrt{1-\gamma_t})\mathcal{D}_\theta(c, r_t, \gamma_t)}{\sqrt{\gamma_t}}. \tag{6}$$

Then $\widetilde{r_0}$ is added to the coarse prediction x_{cp} to obtain the intermediate RPET $\widetilde{y} = x_{cp} + \widetilde{r_0}$. Note that $\widetilde{y}$ is a one-step estimated result rather than the final RPET y'. Herein, we define a generator $\widetilde{p}_\theta(y|r_t, c)$ to represent the above process. Subsequently, the contrastive learning loss $\mathcal{L}_{CL}$ is formulated as:

$$\mathcal{L}_{CL} = \mathbb{E}_{q(y)}[-log\widetilde{p}_\theta(y|r_t, c)] - \sum_{y^i \in Neg} \mathbb{E}_{q(y^i)}[-log\widetilde{p}_\theta(y^i|r_t, c)]. \tag{7}$$

Intuitively, as illustrated in Fig. 1(b), the $\mathcal{L}_{CL}$ aims to minimize the discrepancy between the training label y and the intermediate RPET $\widetilde{y}$ at each time step (first term), while simultaneously ensuring that $\widetilde{y}$ is distinguishable from the negative samples, i.e., the SPET images of other subjects (second term). The contrastive diffusion strategy extends contrastive learning to each time step, which allows LPET images to establish better associations with their corresponding RPET images at different denoising stages, thereby enhancing the mutual information between the LPET and RPET images as done in [33].

3.4 Training Loss

Following [28], we modify the objective $\mathcal{L}_{DPM}$ in Eq. 3, and train CPM and IRM jointly by minimizing the following loss function:

$$\mathcal{L}_{main} = \mathbb{E}_{(c,y)\sim p_{train}} \mathbb{E}_{\epsilon\sim\mathcal{N}(0,I)} \mathbb{E}_{\gamma\sim p_\gamma} \|\mathcal{D}_\theta(c, \sqrt{\gamma}(y - \mathcal{P}_\theta(c)) + \sqrt{1-\gamma}\epsilon, \gamma) - \epsilon\|_1. \tag{8}$$

In summary, the final loss function is:

$$\mathcal{L}_{total} = \mathcal{L}_{main} + m\mathcal{L}_{G}^{NAS} + n\mathcal{L}_{G}^{spectrum} + k\mathcal{L}_{CL}, \tag{9}$$

where m, n and k are the hyper-parameters controlling the weights of each loss.

3.5 Implementation Details

The proposed method is implemented by the Pytorch framework using an NVIDIA GeForce RTX 3090 GPU with 24GB memory. The IRM in our framework is built upon the architecture of SR3 [21], a standard conditional DPM. The number of downsampling operations M is 3, and the negative sample set number N is 10. 4 neighboring slices are used as the NAS guidance and the spectrums are obtained through discrete Fourier transform. As for the weights of each loss, we set $m = n = 1$, and $k = 5e{-}5$ following [33]. We train our model for 500,000 iterations with a batch size of 4, using an Adam optimizer with a learning rate of $1e{-}4$. The total diffusion steps T are 2,000 during training and 10 during inference.

4 Experiments and Results

Datasets and Evaluation: We conducted most of our low-dose brain PET image reconstruction experiments on a public brain dataset, which is obtained from the Ultra-low Dose PET Imaging Challenge 2022 [4]. Out of the 206 18F-FDG brain PET subjects acquired using a Siemens Biograph Vision Quadra, 170 were utilized for training and 36 for evaluation. Each subject has a resolution of $128 \times 128 \times 128$, and 2D slices along the z-coordinate were used for training and evaluation. To simulate LPET images, we applied a dose reduction factor of 100 to each SPET image. To quantify the effectiveness of our method, we utilized three common evaluation metrics: the peak signal-to-noise (PSNR), structural similarity index (SSIM), and normalized mean squared error (NMSE). Additionally, we also used an in-house dataset, which was acquired on a Siemens Biograph mMR PET-MR system. This dataset contains PET brain images collected from 16 subjects, where 8 subjects are normal control (NC) and 8 subjects are mild cognitive impairment (MCI). To evaluate the generalizability of our method, all the experiments on this in-house dataset are conducted in a cross-dataset manner, i.e., training exclusively on the public dataset and inferring on the in-house dataset. Furthermore, we perform NC/MCI classification on this dataset as the clinical diagnosis experiment. ***Please refer to the supplementary materials for the experimental results on the in-house dataset***.

Comparison with SOTA Methods: We compare the performance of our method with 6 SOTA methods, including DeepPET [9] (regression-based method), Stack-GAN [27], Ea-GAN [31], AR-GAN [16], 3D CVT-GAN [32] (GAN-based method) and NVAE [3] (likelihood-based method) on the public

Table 1. Quantitative comparison results on the public dataset. *: We implemented this method ourselves as no official implementation was provided.

		PSNR↑	SSIM↑	NMSE↓	MParam.
regression-based method	DeepPET [9]	23.078	0.937	0.087	**11.03**
GAN-based method	Stack-GAN [27]	23.856	0.959	0.071	83.65
	Ea-GAN [31]	24.096	0.962	0.064	41.83
	AR-GAN [16]	24.313	0.961	0.055	43.27
	3D CVT-GAN [32]	25.080	0.971	0.039	28.72
likelihood-based method	*NVAE [3]	23.629	0.956	0.064	58.24
	Ours	25.638	0.974	0.033	34.10
	Ours-AMS	**25.876**	**0.975**	**0.032**	34.10

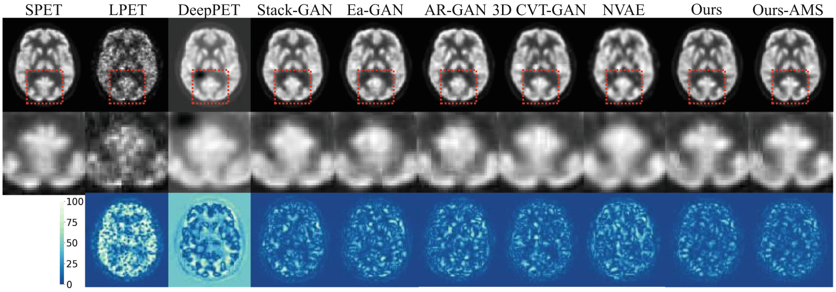

Fig. 2. Visual comparison with SOTA methods.

dataset. Since the IRM contains a stochastic process, we can also average multiple sampled (AMS) results to obtain a more stable reconstruction, which is denoted as Ours-AMS. Results are provided in Table 1. As can be seen, our method significantly outperforms all other methods in terms of PSNR, SSIM, and NMSE, and the performance can be further amplified by averaging multiple samples. Specifically, compared with the current SOTA method 3D CVT-GAN, our method (or ours-AMS) significantly boosts the performance by 0.558 dB (or 0.796 dB) in terms PSNR, 0.003 (or 0.004) in terms of SSIM, and 0.006 (or 0.007) in terms of NMSE. Moreover, 3D CVT-GAN uses 3D PET images as input. Since 3D PET images contain much more information than 2D PET images, our method has greater potential for improvement when using 3D PET images as input. Visualization results are illustrated in Fig. 2. Columns from left to right show the SPET, LPET, and RPET results output by different methods. Rows from top to bottom display the reconstructed results, zoom-in details, and error maps. As can be seen, our method generates the lowest error map while the details are well-preserved, consistent with the quantitative results.

Ablation Study: To thoroughly evaluate the impact of each component in our method, we perform an ablation study on the public dataset by breaking down

Table 2. Quantitative results of the ablation study on the public dataset.

	Single Sampling					Averaged Multiple Sampling			
	PSNR↑	SSIM↑	NMSE↓	MParam.	BFLOPs	PSNR↑	SSIM↑	NMSE↓	SD
(a) baseline	23.302	0.962	0.058	128.740	5973	23.850	0.968	0.052	6.16e−3
(b) CPM	24.354	0.963	0.049	**24.740**	**38**	–	–	–	–
(c) +IRM	24.015	0.966	0.044	31.020	132	24.339	0.967	0.041	3.78e−3
(d) +NAS	24.668	0.969	0.046	33.040	140	24.752	0.970	0.044	3.41e−3
(e) +spec	25.208	0.972	0.044	34.100	145	25.376	0.973	0.043	3.30e−3
(f)+$\mathcal{L}_{CL}$	**25.638**	**0.974**	**0.033**	34.100	145	**25.876**	**0.975**	**0.032**	**2.49e−3**

our model into several submodels. We begin by training the SR3 model as our baseline (a). Then, we train a single CPM with an L2 loss (b), followed by the incorporation of the IRM to calculate the residual (c), and the addition of the auxiliary NAS guidance (d), the spectrum guidance (e), and the $\mathcal{L}_{CL}$ loss term (f). Quantitative results are presented in Table 2. By comparing the results of (a) and (c), we observe that our coarse-to-fine design can significantly reduce the computational overhead of DPMs by decreasing MParam from 128.740 to 31.020 and BFLOPs from 5973 to 132, while achieving better results. The residual generated in (c) also helps to improve the result of the CPM in (b), leading to more accurate PET images. Moreover, our proposed auxiliary guidance strategy and contrastive learning strategy further improve the reconstruction quality, as seen by the increase in PSNR, SSIM, and NMSE scores from (d) to (f). Additionally, we calculate the standard deviation (SD) of the averaged multiple sampling results to measure the input-output correspondence. The standard deviation (SD) of (c) (6.16e−03) is smaller compared to (a) (3.78e−03). This is because a coarse RPET has been generated by the deterministic process. As such, the stochastic process IRM only needs to generate the residual, resulting in less output variability. Then, the SD continues to decrease (3.78e−03 to 2.49e−03) as we incorporate more components into the model, demonstrating the improved input-output correspondence.

5 Conclusion

In this paper, we propose a DPM-based PET reconstruction framework to reconstruct high-quality SPET images from LPET images. The coarse-to-fine design of our framework can significantly reduce the computational overhead of DPMs while achieving improved reconstruction results. Additionally, two strategies, i.e., the auxiliary guidance strategy and the contrastive diffusion strategy, are proposed to enhance the correspondence between the input and output, further improving clinical reliability. Extensive experiments on both public and private datasets demonstrate the effectiveness of our method.

Acknowledgement. This work is supported by the National Natural Science Foundation of China (NSFC 62371325, 62071314), Sichuan Science and Technology Program 2023YFG0263, 2023YFG0025, 2023NSFSC0497.

References

1. Chen, N., Zhang, Y., Zen, H., Weiss, R.J., Norouzi, M., Chan, W.: WaveGrad: estimating gradients for waveform generation. arXiv preprint arXiv:2009.00713 (2020)
2. Chung, H., Sim, B., Ye, J.C.: Come-closer-diffuse-faster: accelerating conditional diffusion models for inverse problems through stochastic contraction. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 12413–12422 (2022)
3. Cui, J., et al.: Pet denoising and uncertainty estimation based on NVAE model using quantile regression loss. In: Wang, L., Dou, Q., Fletcher, P.T., Speidel, S., Li, S. (eds.) MICCAI 2022, Part IV. LNCS, vol. 13434, pp. 173–183. Springer, Cham (2022). https://doi.org/10.1007/978-3-031-16440-8_17
4. MICCAI challenges: Ultra-low dose pet imaging challenge 2022 (2022). https://doi.org/10.5281/zenodo.6361846
5. Dhariwal, P., Nichol, A.: Diffusion models beat GANs on image synthesis. Adv. Neural. Inf. Process. Syst. **34**, 8780–8794 (2021)
6. Fei, Y., et al.: Classification-aided high-quality pet image synthesis via bidirectional contrastive GAN with shared information maximization. In: Wang, L., Dou, Q., Fletcher, P.T., Speidel, S., Li, S. (eds.) MICCAI 2022, Part VI. LNCS, vol. 13436, pp. 527–537. Springer, Cham (2022). https://doi.org/10.1007/978-3-031-16446-0_50
7. Gong, K., Guan, J., Liu, C.C., Qi, J.: Pet image denoising using a deep neural network through fine tuning. IEEE Trans. Radiat. Plasma Med. Sci. **3**(2), 153–161 (2018)
8. Goodfellow, I., et al.: Generative adversarial networks. Commun. ACM **63**(11), 139–144 (2020)
9. Häggström, I., Schmidtlein, C.R., Campanella, G., Fuchs, T.J.: DeepPET: a deep encoder-decoder network for directly solving the pet image reconstruction inverse problem. Med. Image Anal. **54**, 253–262 (2019)
10. Ho, J., Jain, A., Abbeel, P.: Denoising diffusion probabilistic models. Adv. Neural. Inf. Process. Syst. **33**, 6840–6851 (2020)
11. Kang, S.K., Choi, H., Lee, J.S., Initiative, A.D.N., et al.: Translating amyloid pet of different radiotracers by a deep generative model for interchangeability. Neuroimage **232**, 117890 (2021)
12. Kaplan, S., Zhu, Y.M.: Full-dose pet image estimation from low-dose pet image using deep learning: a pilot study. J. Digit. Imaging **32**(5), 773–778 (2019)
13. Kim, K., et al.: Penalized pet reconstruction using deep learning prior and local linear fitting. IEEE Trans. Med. Imaging **37**(6), 1478–1487 (2018)
14. Lei, Y., et al.: Whole-body pet estimation from low count statistics using cycle-consistent generative adversarial networks. Phys. Med. Biol. **64**(21), 215017 (2019)
15. Luo, Y., et al.: 3D transformer-GAN for high-quality PET reconstruction. In: de Bruijne, M., et al. (eds.) MICCAI 2021, Part VI. LNCS, vol. 12906, pp. 276–285. Springer, Cham (2021). https://doi.org/10.1007/978-3-030-87231-1_27
16. Luo, Y., et al.: Adaptive rectification based adversarial network with spectrum constraint for high-quality pet image synthesis. Med. Image Anal. **77**, 102335 (2022)

17. Metz, L., Poole, B., Pfau, D., Sohl-Dickstein, J.: Unrolled generative adversarial networks. arXiv preprint arXiv:1611.02163 (2016)
18. Ouyang, J., Chen, K.T., Gong, E., Pauly, J., Zaharchuk, G.: Ultra-low-dose pet reconstruction using generative adversarial network with feature matching and task-specific perceptual loss. Med. Phys. **46**(8), 3555–3564 (2019)
19. Ren, M., Delbracio, M., Talebi, H., Gerig, G., Milanfar, P.: Image deblurring with domain generalizable diffusion models. arXiv preprint arXiv:2212.01789 (2022)
20. 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, pp. 10684–10695 (2022)
21. Saharia, C., Ho, J., Chan, W., Salimans, T., Fleet, D.J., Norouzi, M.: Image super-resolution via iterative refinement. IEEE Trans. Pattern Anal. Mach. Intell. **45**, 4713–4726 (2022)
22. Salimans, T., Goodfellow, I., Zaremba, W., Cheung, V., Radford, A., Chen, X.: Improved techniques for training GANs. Advances in Neural Inf. Process. Syst. **29** (2016)
23. Sohl-Dickstein, J., Weiss, E., Maheswaranathan, N., Ganguli, S.: Deep unsupervised learning using nonequilibrium thermodynamics. In: International Conference on Machine Learning, pp. 2256–2265. PMLR (2015)
24. Song, Y., Shen, L., Xing, L., Ermon, S.: Solving inverse problems in medical imaging with score-based generative models. arXiv preprint arXiv:2111.08005 (2021)
25. Ulhaq, A., Akhtar, N., Pogrebna, G.: Efficient diffusion models for vision: a survey. arXiv preprint arXiv:2210.09292 (2022)
26. Wang, Y., et al.: 3D conditional generative adversarial networks for high-quality pet image estimation at low dose. Neuroimage **174**, 550–562 (2018)
27. Wang, Y., et al.: 3D auto-context-based locality adaptive multi-modality GANs for pet synthesis. IEEE Trans. Med. Imaging **38**(6), 1328–1339 (2018)
28. Whang, J., Delbracio, M., Talebi, H., Saharia, C., Dimakis, A.G., Milanfar, P.: Deblurring via stochastic refinement. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 16293–16303 (2022)
29. Xiang, L., et al.: Deep auto-context convolutional neural networks for standard-dose pet image estimation from low-dose PET/MRI. Neurocomputing **267**, 406–416 (2017)
30. Xu, J., Gong, E., Pauly, J., Zaharchuk, G.: 200x low-dose pet reconstruction using deep learning. arXiv preprint arXiv:1712.04119 (2017)
31. Yu, B., Zhou, L., Wang, L., Shi, Y., Fripp, J., Bourgeat, P.: EA-GANs: edge-aware generative adversarial networks for cross-modality mr image synthesis. IEEE Trans. Med. Imaging **38**(7), 1750–1762 (2019)
32. Zeng, P., et al.: 3D CVT-GAN: a 3D convolutional vision transformer-GAN for pet reconstruction. In: Wang, L., Dou, Q., Fletcher, P.T., Speidel, S., Li, S. (eds.) MICCAI 2022, Part VI. LNCS, vol. 13436, pp. 516–526. Springer, Cham (2022). https://doi.org/10.1007/978-3-031-16446-0_49
33. Zhu, Y., Wu, Y., Olszewski, K., Ren, J., Tulyakov, S., Yan, Y.: Discrete contrastive diffusion for cross-modal and conditional generation. arXiv preprint arXiv:2206.07771 (2022)