Incorporating Dose Level Prior and Texture-Sensitive Contrastive Learning from Frequency Perspective for Multi-dose PET Reconstruction

1st Yuchen Fei
School of Computer Science
Sichuan University
Chengdu, China
13072807126@163.com

2nd Zhenghao Feng
School of Computer Science
Sichuan University
Chengdu, China
fzh_scu@163.com

3rd Jiliu Zhou
School of Computer Science
Sichuan University
Chengdu, China
zhoujiliu@cuit.edu.cn

4th Yan Wang
School of Computer Science
Sichuan University
Chengdu, China
wangyanscu@hotmail.com

***Abstract*—Due to the inherent radiation exposure of positron emission tomography (PET), reconstructing standard-dose PET (SPET) from low-dose PET (LPET) is becoming an attractive alternative to ensure both imaging quality and patient safety. In clinic, the noise levels of LPET images may vary significantly owing to the individual differences among patients, making it challenging to design a universal model for PET reconstruction across multiple low-dose levels. Nevertheless, most existing multi-dose-level PET image reconstruction methods incorporated limited dose-level information and overlooked the potential of frequency domain prior. In this paper, we demonstrate that the frequency domain possesses the capability to differentiate between SPET and LPET. Harnessing this crucial insight, we subtly design a high-frequency component classification network to embed dose prior into the network. Furthermore, considering the critical importance of texture details in clinical applications, we adopt a frequency contrastive learning paradigm that utilizes a hard negative sample strategy to enforce the reconstructed PET (RPET) to preserve more texture details. Meanwhile, we elaborately construct texture-sensitive embeddings that can perceive subtle differences for contrastive learning, enforcing more realistic and robust reconstruction. Extensive experiments conducted on the MICCAI 2022 UDPET dataset and the Phantom Brain Dataset have demonstrated the effectiveness of our proposed method.**

Keywords—Positron Emission Tomography (PET), Frequency Domain, Contrastive learning, Hard Negative Sample, Medical Image Reconstruction

I. Introduction

Positron emission tomography (PET) enables the visualization of physiological and biochemical processes in a non-invasive manner, gaining widespread adoption in clinic [1-2]. However, the radiation hazard associated with the radioactive tracer which is required for PET imaging induces potential health risks to patients. Yet, while reducing the tracer dose will inevitably involve more noise and artifacts, preventing the acquisition of high-quality PET images that meet the clinical requirements [3]. Confronted with this clinical dilemma, a feasible approach is to reconstruct standard-dose PET (SPET) images from low-dose PET (LPET) ones.

Over the past decade, substantial deep learning-based endeavors have been devoted for obtaining diagnostic-quality PET images at low dose [4-15]. For example, Xu *et al.* [6] incorporated multi-slice input strategy to provide additional structural information to PET reconstruction. Spuhler *et al.* [7] combined dilated convolution with multi-scale framework for enlarging the receptive field, thus boosting the quality of reconstructed PET (RPET) image. Wang *et al.* [13] first designed a 3D multi-modality transformer model to leverage the beneficial knowledge from the magnetic resonance imaging (MRI) modality for performance enhancement. Additionally, Cui *et al.* [15] proposed a direct PET reconstruction method to exhaust the complementary information in the spatial, projection, and frequency domains. Despite achieving remarkable results, these methods focus exclusively on processing LPET images at a single dose level. However, different levels of noise and artifacts can be involved by applying different dose reduction factors (DRFs, i.e., the ratio of the original dose to the reduced dose). Meanwhile, the noise levels of PET images are also susceptible due to patient characteristics, scanning time, and so on. Directly applying single-dose-level LPET reconstruction methods to the clinical scenarios with LPET images at various DRFs are prone to generate undesirable results. In light of this, reconstructing SPET from multi-dose-level LPET has garnered increasing attention [16-21]. Along this direction, Li *et al.* [17] constructed an adaptive noise level-aware network, using predicted dose information as prior knowledge to guide multi-dose-level PET reconstruction. Fei *et al.* [19] used a two-stage model to progressively realize realistic reconstruction from LPET at various DRFs. Moreover, Niu et al. [21] incorporate the complementary anatomical constraints derived from CT images for more realistic multi-dose-level PET reconstruction results.

Despite the notable success of current multi-dose-level PET reconstruction methods, there are some limitations worth further investigation. First, the frequency prior of LPET images at various DRFs has been omitted. By visualizing the spectrums in Fig. 1(a), it can be seen that the disparities among PET images at DRFs (e.g., 20, 50, 100) are mainly manifested in high-frequency area, and exhibit similar low-frequency information

Yan Wang is the corresponding author. This work is supported by National Natural Science Foundation of China (NSFC 62371325), Sichuan Science and Technology Program 2025NSFJQ0050, Sichuan Science and Technology Program 2024ZDZX0018, and Key Lab of Internet Natural Language Processing of Sichuan Provincial Education Department (No.INLP202402).

since they correspond to the same anatomical structure. More clearly, by swapping the high-frequency components of SPET and LPET at three DRFs, we find that the swapped SPET is close to LPET, while the swapped LPET exhibit desired SPET-like details. Based on the above observations, we argue that leveraging the frequency prior can facilitate the exploration of the discriminative dose-level related features that are crucial for multi-dose-level PET reconstruction [17]. Second, reconstructing high-quality PET images with rich texture details (e.g., boundaries) constitutes a clinical imperative, as the precise boundary delineation between the normal tissue and pathological lesions reflect the accurate quantification of metabolic heterogeneity which is critical in clinical diagnosis [22]. However, most existing methods based on conventional up-sampling operations may involve inherent noise, thus leading to high-frequency detail distortions.

In this paper, to take full advantage of the frequency-domain prior as well as enhance the reconstruction texture details, we design a novel multi-dose-level PET reconstruction framework from a frequency perspective. Specifically, we verify that the frequency domain possesses superior capability in identifying the degradation level of LPET by visualizing the spectrums. Based on this and the insight that the reconstruction performance can be boosted by incorporating the dose level prior [17], we embed a high-frequency components classifier to promote the model to extract the dose-level-related discriminative features. Moreover, the contrastive learning strategy is used in frequency domain for enhancing the high-frequency details. Concretely, we adopt the hard negative sample strategy to enforce the RPET images far from smooth textures, thus boosting its clinical applicability. The high-frequency components classifier is reused to attain task-friendly embeddings that could perceive subtle differences in high-frequency texture details, thus promoting the generation of more realistic reconstruction results.

The contributions of this paper can be summarized as follows: (1) In contrast to previous spatial domain methods, we investigate the multi-dose-level PET reconstruction task from a frequency perspective. (2) The dose-level prior knowledge embraced in the frequency domain is incorporated to guide more precise reconstruction. (3) We develop a frequency contrastive learning strategy leveraging hard negative samples and task-friendly embedding network to effectively enhance texture details in reconstruction. (4) Experimental results verify the superiority and generalization of our model.

II. Methodology

The architecture of our proposed method is illustrated in Fig. 1(b). Given the inherent modality consistency between LPET and SPET images, we follow the philosophy of multi-task learning and design a dual-branch decoding architecture that respectively realizes LPET self-reconstruction and SPET reconstruction, thus enabling the capture of robust PET image representations. Acknowledging that the dose-level prior embraced in the frequency domain facilitates learning the complex mapping relationship between multi-dose-level LPET and SPET [17], we design a high-frequency components classifier C_h cascaded after the LPET self-reconstruction branch to implicitly promote the encoder to extract dose-level-related discriminative features. In addition, to faithfully preserve the texture information, we adopt the contrastive learning strategy from a frequency perspective to boost RPET image quality and design a hard negative sample strategy by adding blur to the SPET as negative samples, thus enforcing the RPET far from ambiguity. Moreover, the high-frequency components classifier C_h is reused as the embedding network E to construct task-friendly positive/negative/anchor embeddings. More details will be displayed in the following sections.

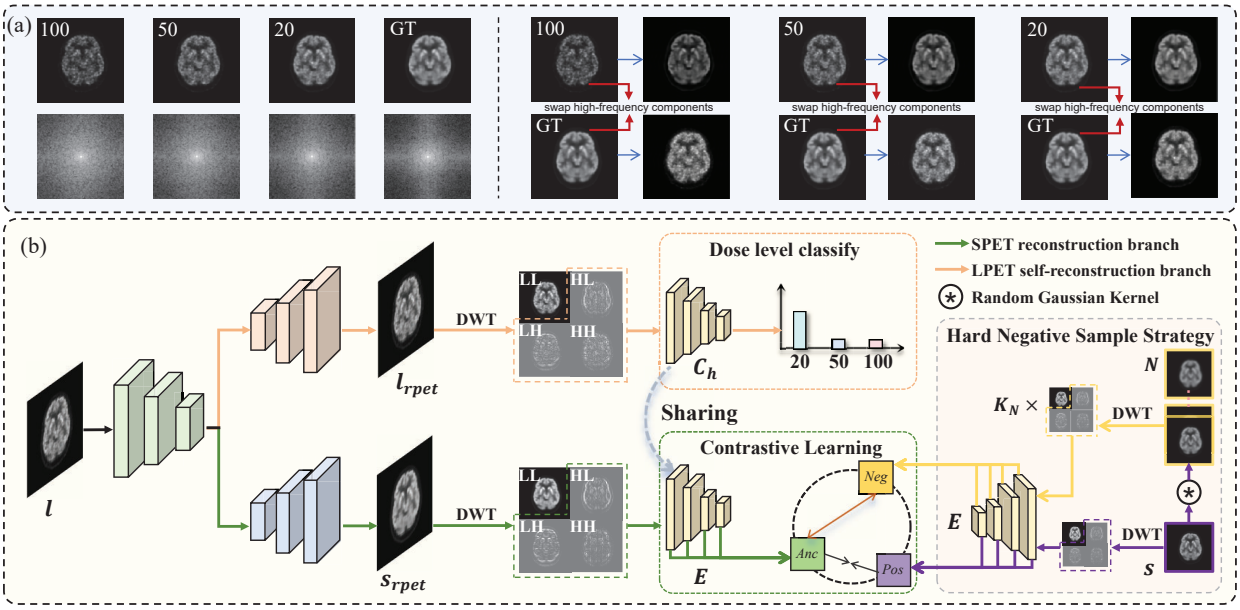

Fig. 1. (a) Motivation: frequency domain possesses the capability to differentiate SPET and LPET at various DRFs. (b) Overall architecture of the proposed framework. Specifically, we design a dual-branch for capturing robust PET image representations and adopt a high-frequency components classifier to embed dose prior into the model. The frequency contrastive learning paradigm utilizing hard negative sample strategy and texture-sensitive embeddings is also devised to enforce the RPET to preserve more texture details.

A. Prior-embedded framework

Considering the shared content and structure information between LPET and SPET images, we design a shared encoder fed with LPET and a dual-branch decoder for RPET reconstruction and auxiliary LPET self-reconstruction, respectively. In this way, the model could faithfully capture the image representations that embed rich content and structure information. Specifically, the encoder employs three down-sampling blocks while the decoder takes symmetrical structure comprising three up-sampling blocks to gradually restore the high-dimensional feature maps to the image size. Aligned with [23], the skip connection is adopted in the RPET reconstruction branch for mitigating spatial information degradation, but is dropped in the self-reconstruction branch to enforce the decoder to better learn the mapping relationship between high-dimensional features and LPET images. To enforce the consistency of the generated image and ground truth while encouraging less blurry, we adopt L1 loss as below:

$$L_{lrec} = \left\| l - l_{rpet} \right\|_1, L_{srec} = \left\| s - s_{rpet} \right\|_1 \quad (1)$$

where l, s are the LPET and SPET, l_{rpet} and s_{rpet} are the self-reconstructed LPET (RLPET) and the RPET, respectively.

High-frequency Components Classifier. Considering that noise level information is conductive to learning the mapping relationship between the input image with various noise levels and the target high-quality image [17], we also strive to explore the dose-level related discriminative features for effective multi-dose-level PET reconstruction. Specifically, as illustrated in Fig. 1(a), LPET images at varying DRFs show consistency in the low-frequency components, but exhibit distinctive disparities in terms of high-frequency components. With this in mind, we construct a classifier from the frequency perspective after the auxiliary LPET self-reconstruction branch, thus promoting the model to explore the discriminative features related to each DRF under the guidance of learned dose prior. Specifically, we employ the Haar discrete wavelet transformation (DWT) [24] to obtain four orthogonal sub-bands: i.e., LL (low-low), LH (low-high), HL (high-low), and HH (high-high). The LL sub-band capturing the overall structure information represents the low-frequency components of the image, the last three detail sub-bands embody high-frequency information and are concatenated as discriminative inputs to the classifier C_h. The classifier is structured like an CNN encoder, employing four down-sampling modules based on strided convolution to gradually extract the features. After that, the features are flattened into a one-dimensional vector which is passed through a fully connected layer to estimate the dose level property of the l_{rpet}. The cross entropy (CE) loss is used to force the model to accurately predict the category of l_{rpet}, which can be expressed as:

$$L_{ce} = -\sum_{i=1}^{j} c_{real}^{i} * log c_{pre}^{i} \quad (2)$$

where c_{real}^{i} and c_{pre}^{i} are the real DRF labels and the probability predicted by the model that l_{rpet} belong to category i, j is the total number of DRF categories.

B. Frequency domain Contrastive Learning

Acknowledging that multi-dose-level LPET images involve various noise levels and exhibit obvious differences from SPET images in high-frequency components, the reconstruction process is expected to recover the lost high-frequency texture details. In light of this, we adopt the contrastive learning strategy in frequency domain for enhancing the high-frequency components learning while mitigating the spectrum distortion.

Sample Construction. The core idea of contrastive learning is to minimize the distance between the positive sample pair while maximizing the distance between negative sample pair [25-26]. Therefore, it is of vital importance to construct effective positive and negative pairs. Tailored to the PET reconstruction task, we establish the RPET as anchor sample and designate the target SPET (GT) as positive sample. Nevertheless, directly utilizing LPET images with significant image quality degradation as negative samples may leads the model to easily distinguish the anchor sample from the negative sample, providing very limited knowledge for contrastive learning. To tackle this issue, we inject the spirit of hard negative sample strategy to the conventional contrastive learning paradigm. To faithfully preserve the texture details in reconstruction, it is practical to treat some blurred SPET images as negative samples and strengthen the sharpness in prediction by enlarging the distances between the prediction and blurred SPET images. Specifically, we apply Gaussian blur to the target SPET image, thus obtaining the slightly blurry PET (BPET) images as hard negative samples. For the k-th image s_k, the negative sample set N_k can be formulated as:

$$N_k = \{n_t | n_t = Blur(s_k)\}_{t=1}^{K_N}, \quad (3)$$

where K_N is the number of negative samples, $Blur(\cdot)$ denotes the blurring function.

Embedding Network. Currently, most methods employ pretrained vision encoder network predominantly encoding high-level semantic features and generate embeddings which may be not sufficient for low-level tasks. As a result, the generated embeddings fail to capture the subtle differences of PET images with varying levels of degradation, thus limiting the performance of contrastive learning. In light of this, we reuse the high-frequency components classifier which is texture-sensitive to generate the PET embeddings. Specifically, since the contrastive learning strategy strives to boost reconstruction performance by explicitly modeling inter-sample similarities/dissimilarities and thus capturing subtle noise-level variations, the discriminative representations derived from the high-frequency components classifier inherently encode the high-frequency texture features. Thereby, the task-friendly and computationally-efficient embedding network E could provide critical prior knowledge to the contrastive learning, ensuring the reconstruction of RPET images with preserved high-frequency fidelity. Specifically, embedding network E compresses the prediction s_{rpet}, positive GT sample s, and negative blurred samples n_t in N_k into compact prediction representations p, GT representations g, and blurred representations b_t. To make full advantage of the samples, the multi-scale hierarchical features

extracted from the down-sampling pathway are integrated for constructing the contrastive loss, which is illustrated as below:

$$L_{InfoNCE} = -\sum_{v=1}^{V} log \frac{\exp\left(\frac{sim(p^v, g^v)}{\tau}\right)}{\exp\left(\frac{sim(p^v, g^v)}{\tau}\right)+\sum_{t=1}^{K_N}\exp\left(\frac{sim(p^v, b_t^v)}{\tau}\right)}, \quad (4)$$

where τ is the temperature parameter, v is the number of down-sampling layers, p^v, g^v, b_t^v are the v-th layer features extracted from the high-frequency components of anchor sample RPET s_{rpet}, positive sample SPET s, negative sample n_t, respectively. $sim(\cdot)$ is the pixel-wise cosine similarity function. In this way, the texture details are enforced to be far away from smooth and become closer to the SPET images through contrastive optimization. Finally, the overall loss function L_{total} can be derived as:

$$L_{total} = L_{srec} + \lambda_1 L_{lrec} + \lambda_2 L_{InfoNCE} + \lambda_3 L_{ce}, \quad (5)$$

where λ_1, λ_2 and λ_3 are hyper-parameters to balance the loss terms.

C. Implementation Details

Experiments are implemented utilizing the Pytorch framework with the NVIDIA GeForce GTX 2070 with 8GB memory. Specifically, we employ the Adam optimizer to train the model for 100 epochs, with learning rate of 2×10^{-4} and batch size of 8 to enable an efficient convergence. Empirically, λ_1 and λ_2 are separately set to 1 and 0 in the first 20 epochs to ensure the quality of RLPET images and enable the high-frequency components classifier to capture the dose level prior. Then, we respectively decreased λ_1 to 0.1 and increased λ_2 to 0.001 according to the convergence state, which enforces the model to utilize the learned discriminative features to effectively generate high-quality RPET images. λ_3 in equation (5) is set to 1, and the number of negative sample set K_N is set as 4 [32]. In this study, we incorporate three categories of LPET images for the experiments, i.e., the number of the DRF category i is 3.

III. Experiments and Results

We validate our method on the MICCAI 2022l UDPET Dataset [27] and the Phantom Brain Dataset, which contains PET data from 206 subjects utilizing the Biograph Vision Quadra System and 20 subjects which are simulated from the BrainWeb database [28-29], respectively. For the MICCAI 2022 UDPET Dataset, we randomly selected 170 subjects for training and utilized the remaining 36 subjects for testing. All data are acquired in list mode, allowing for the rebinding of data to simulate different acquisition times and obtain equivalent low-dose subjects. In our work, LPET image data at DRFs of 20, 50, and 100 are incorporated for the experiments. Each 3D brain image of size 128×128×128 was sliced to 128 2D slices with size 128 ×128 along the axial direction to enrich the training samples. For the Phantom Brain Dataset, the simulated LPET is equivalent to a quarter of the standard count level. Considering the relatively small dose difference between LPET and SPET, the LPET image is adopted as the hard negative sample in contrastive learning. We also adopt image slicing to increase the samples and the five-fold cross-validation strategy to ensure the generalization of training. We adopt the peak signal-to-noise ratio (PSNR), structural similarity index (SSIM), normalized root-mean-square error (NRMSE) as the evaluation metric to quantitatively measure the performance.

A. Comparison with Existing State-of-the-art Methods

To demonstrate the advancement of our method, seven SOTA methods are implemented for comparison, including Auto-Context [5], StackGAN [11], Ea-GAN [22], AR-GAN [10], M-cGAN [27], CGSRGAN [18], RefineNet [19] and AdaIR [30]. Note that, the U-Net [31] models were trained individually on LPET images at each DRF to verify the generalization of our model. The quantitative results are shown in Table II. As observed, our model significantly outperforms the first four classical medical image reconstruction methods, suggesting our capability to account for the disparities among input multi-dose-level LPET images. On the contrary, our method acquires the optimal performance with the PSNR metric achieving 30.544dB, 32.270dB and 34.652dB, for DRF = 100, 50 and 20, respectively. Besides, the following general reconstruction methods and the separate trained U-Net attain superior results compared to classical medical image reconstruction methods, but our method still ranks first among all general approaches. In addition, the t-test results show that the p-values in most cases are smaller than 0.05, indicating that the improvements gained by our model are statistically significant. We also provide the visual comparison results in Fig. 3. It can be seen that the classical methods obtain relatively poor reconstruction results with blurred structure, while the general reconstruction methods effectively enhance the RPET image quality at all three DRFs. Among all multi-dose-level

TABLE I
QUANTITATIVE COMPARISON OF THE PROPOSED METHOD WITH STATE-OF-THE-ART APPROACHES IN TERMS OF PSNR ↑, SSIM ↑ AND NRMSE ↓, RESPECTIVELY. (*: T-TEST IS CONDUCTED ON PROPOSED METHOD AND THE COMPARISON METHODS.)

Model	MICCAI 2022 UDPET Dataset									Phantom Brain Dataset		
	DRF=100			DRF=50			DRF=20			\		
	PSNR	SSIM	NRMSE	PSNR	SSIM	NRMSE	PSNR	SSIM	NRMSE	PSNR	SSIM	NRMSE
U-Net	29.608	0.954	0.242*	31.589	0.971	0.176*	33.856	0.982	0.130*	\	\	\
Auto-Context	27.354*	0.958	0.216*	28.158*	0.966*	0.197*	28.685*	0.972*	0.185*	29.733*	0.965*	0.112*
StackGAN	28.778*	0.958	0.197*	29.913*	0.968	0.171*	30.801*	0.977*	0.153*	30.823*	0.966*	0.099*
Ea-GAN	29.278*	0.961	0.192*	30.511*	0.969	0.168*	31.596*	0.977*	0.148*	30.991*	0.968*	0.101*
AR-GAN	29.448*	0.959	0.193*	30.770*	0.970	0.162*	32.067*	0.978*	0.139*	31.693	0.969	0.092
M-cGAN	29.747	0.961	0.199*	31.111*	0.972	0.166*	32.841*	0.981	0.134*	31.126*	0.968*	0.099*
CGSRGAN	29.336*	0.850*	0.206*	31.320	0.864*	0.153*	33.867	0.952*	0.111	31.515	0.977	0.093
RefineNet	30.187	0.960	0.173	31.962	0.973	0.144	34.192	0.982	0.108	31.336*	0.977	0.095
AdaIR	30.270	0.959	0.164	31.368	0.971	0.144	32.604	0.983	0.125	**31.812**	0.969	0.092
Ours	**30.544**	**0.962**	**0.165**	**32.270**	**0.974**	**0.135**	**34.652**	**0.984**	**0.104**	31.768	**0.977**	**0.092**

reconstruction methods, the RPET images reconstructed by our method are most analogous to the ground truth, presenting sharper textures and richer details. Both the quantitative and visual results demonstrate the effectiveness of our method.

Moreover, to further verify the generalizability of our method, we also conduct the comparison experiments on the Phantom Brain Dataset. The quantitative and visual results are respectively displayed in Table II, and Fig. 4. Considering that the Phantom Brain Dataset includes the LPET images at only a single DRF, the dose classification task embedded in the multi-dose-level PET reconstruction comparation methods are removed in this experimental setting. At the same time, since our model reuses the dose classifier to generate contrast learning embeddings, we construct classification tasks for distinguishing the self-reconstruction LPET and SPET to enable the dose classifier to learn the discriminative representations. In addition, as mentioned above, the LPET image whose DRF is equivalent to 4 in Phantom Brain Dataset is adopted as the hard negative sample. As shown in Table II, our model obtains the comprehensive optimal reconstruction performance. From the visual results, we can also find that the RPET images reconstructed by our model are most analogous to the ground truth. These comparison results demonstrate not only the superiority of our method, but also the capability of our method to generalize to the single-dose-level reconstruction task.

B. Ablation Studies

To verify whether each component can yield performance improvement, we design the ablation groups as: (A) the U-Net baseline, (B) the dual-branch decoding model, (C) incorporating the high-frequency component classifier, and our proposed method (Ours). Quantitative results are presented in Table I. By comparing (A) and (B), we observe that variant (B) adopting dual-branch decoding path achieves higher PSNR and lower NRMSE across all DRFs. As for the decreased SSIM metric, this may be explained as that the LPET images at ultra-low dose levels exhibit substantial noise and the shared encoder may misidentify high-intensity noise as fine structural information and thus be interfered with, which is consistent with the minimum metric degradation observed at DRF = 20. In the third row, we can observe that the variant (C) incorporating the dose classifier further promotes the exploration of the dose-level-related discriminative features and boosts all metrics to varying degrees. By applying frequency contrastive learning, our method rises PSNR remarkably by 0.427dB, 0.576dB and 0.461dB for DRF =100, 50, 20, respectively. In addition, we further construct variant models by (I) leveraging hard negative samples in image domain, and (II) adopting LPET images as negative samples. Compared with (I), we can clearly see that pulling RPET and SPET closer from a frequency perspective contribute to boosting RPET image quality, proving the effectiveness of our design. Moreover, it can be observed that variant (II) provides limited knowledge for contrastive learning, while our model achieves better reconstruction performance by constructing Gaussian-blurred SPET as hard negatives.

IV. Conclusion

In this work, we verify that frequency domain enables degradation-level characterization of LPET and design the multi-dose-level PET reconstruction network from a frequency

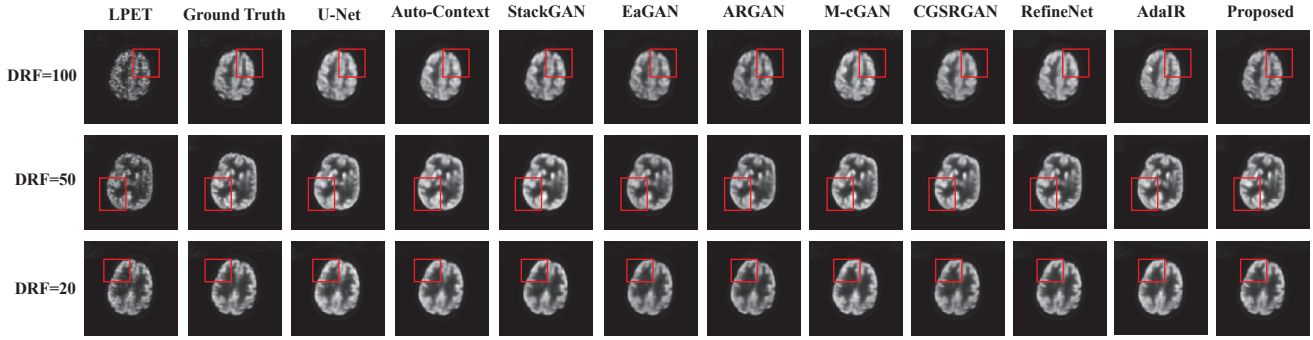

Fig. 3. Visual comparison of the proposed method with state-of-the-art approaches on MICCAI 2022 UDPET Dataset.

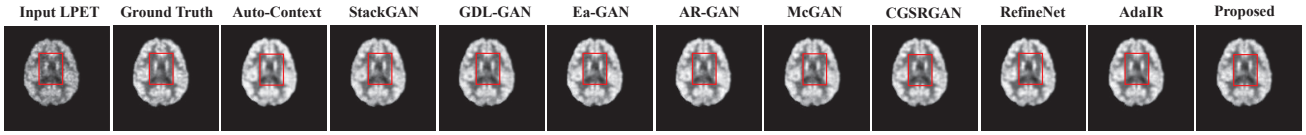

Fig. 4. Visual comparison of the proposed method with state-of-the-art approaches on Phantom Brain Dataset.

TABLE II
QUANTITATIVE COMPARISON OF THE PROPOSED METHOD WITH THREE VARIANT MODELS IN TERMS OF PSNR ↑, SSIM ↑ AND NRMSE ↓, RESPECTIVELY.

Model	DRF=100			DRF=50			DRF=20		
	PSNR	SSIM	NRMSE	PSNR	SSIM	NRMSE	PSNR	SSIM	NRMSE
(A)	29.232	0.953	0.191	30.258	0.963	0.163	30.314	0.968	0.158
(B)	29.852	0.864	0.178	31.465	0.881	0.145	33.829	0.962	0.111
(C)	30.117	0.943	0.173	31.694	0.960	0.143	34.191	0.981	0.107
(I)	30.413	0.957	0.175	32.184	0.968	0.140	34.122	0.981	0.108
(II)	30.157	0.962	0.170	32.018	0.974	0.140	34.241	0.984	0.109
Ours	**30.544**	**0.962**	**0.165**	**32.270**	**0.974**	**0.135**	**34.652**	**0.984**	**0.104**

perspective. Considering the inherent modality consistency between LPET and SPET images, we design a dual-branch decoding architecture to learn more robust features. Meanwhile, in view of the fact that the disparities among PET images at different DRFs are mainly manifested in high-frequency components, we explore the inherent dose prior thorough a high-frequency component classification task to guide more precise reconstruction. Moreover, to alleviate the high-frequency detail-distorted reconstruction, we apply the frequency contrastive learning, utilizing task-friendly embeddings to reduce subtle high-frequency texture discrepancies between RPET and SPET images. In contrast to the common paradigms that directly apply LPET images as negative samples, we incorporate blur to the SPET images drawing inspiration from the hard negative sample strategy. The experimental results validate the superiority and practicality of our model. Limitations include ignoring some 3D partial information when using 2D slices and the inclusion of limited degrees of degradation; future work will leverage 3D spatial information and extend the model to diverse degradation scenarios and multiple scanners.

REFERENCES

[1] K. A. Johnson et al., "Tau positron emission tomographic imaging in aging and early Alzheimer disease," Annals of Neurology, vol. 79, no. 1, pp. 110–119, Jan. 2016.

[2] S. Daerr et al., "Evaluation of early-phase [18F]-florbetaben PET acquisition in clinical routine cases," NeuroImage: Clinical, vol. 14, pp. 77–86, Jan. 2017.

[3] B. Huang, M. W.-M. Law, and P.-L. Khong, "Whole-Body PET/CT Scanning: Estimation of Radiation Dose and Cancer Risk," Radiology, vol. 251, no. 1, pp. 166–174, Apr. 2009.

[4] K. Gong, J. Guan, C. Liu, J. Qi, "PET image denoising using a deep neural net-work through fine tuning," IEEE Transactions on Radiation and Plasma Medical Sciences, vol. 3, no. 2, pp. 153-161, Mar. 2019.

[5] L. Xiang et al., "Deep auto-context convolutional neural networks for standard-dose PET image estimation from low-dose PET/MRI," Neurocomputing, pp. 406–416, Dec. 2017.

[6] J. Xu, E. Gong, J. Pauly, G.W. Zaharchuk, "200x low-dose PET reconstruction using deep learning," 2017, arXiv:1312.6114. [Online]. Available: https://arxiv.org/abs/1712.04119.

[7] K. Spuhler, M. Serrano-Sosa, R. Cattell, C. DeLorenzo, C. Huang, "Full-count PET recovery from low-count image using a dilated convolutional neural network," Medical physics, vol. 47, no. 10, pp. 4928–4938, Oct. 2020.

[8] Y. Wang et al., "3D Auto-Context-Based Locality Adaptive Multi-Modality GANs for PET Synthesis," IEEE Transactions on Medical Imaging, pp. 1328–1339, Jun. 2019.

[9] Y. Fei et al., "Classification-Aided High-Quality PET Image Synthesis via Bidirectional Contrastive GAN with Shared Information Maximization," in Medical Image Computing and Computer Assisted Intervention – MICCAI 2022, 2022, pp. 527–537.

[10] Y. Luo et al., "Adaptive rectification based adversarial network with spectrum constraint for high-quality PET image synthesis," Medical Image Analysis, p. 102335, Apr. 2022.

[11] Y. Wang et al., "3D conditional generative adversarial networks for high-quality PET image estimation at low dose," NeuroImage, pp. 550–562, Jul. 2018.

[12] A. Sano, N. Nishio, T. Masuda, K. Karasawa, "Denoising PET images for proton therapy using a residual U-net," Biomedical Physics & Engineering Express, vol. 7, no. 2, pp. 025014, Feb. 2021.

[13] Y. Wang et al., "3D multi-modality Transformer-GAN for high-quality PET reconstruction," Medical Image Analysis, vol. 91, pp. 102983, Jan. 2024.

[14] L. Bi, J. Kim, A. Kumar, D. Feng and M. Fulham, "Synthesis of positron emission tomography (PET) images via multi-channel generative adversarial networks (GANs)," in Lecture Notes in Computer Science, Molecular Imaging, Reconstruction and Analysis of Moving Body Organs, and Stroke Imaging and Treatment, 2017, pp. 43–51.

[15] J. Cui et al., "Prior Knowledge-guided Triple-Domain Transformer-GAN for Direct PET Reconstruction from Low-Count Sinograms," IEEE Transactions on Medical Imaging, vol. 43, no. 12, pp. 4174-4189, Dec. 2024.

[16] L. Zhang et al., "Spatial adaptive and transformer fusion network (STFNet) for low-count pet blind denoising with MRI," Medical Physics, vol. 49, no. 1, pp. 343–356, Jan. 2022.

[17] W. Li et al., "Adaptive 3D noise level-guided restoration network for low-dose positron emission tomography imaging," Interdisciplinary Medicine, vol. 1, no. 3, pp. e20230012, Jun. 2023.

[18] Y. Xue, Y. Peng, L. Bi, D. Feng and J. Kim, "CG-3DSRGAN: A classification guided 3D generative adversarial network for image quality recovery from low-dose PET images," in 2023 45th Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC), 2023, pp. 1-4.

[19] Y. Fei et al., "Two-Phase Multi-Dose-Level Pet Image Reconstruction with Dose Level Awareness," in 2024 IEEE International Symposium on Biomedical Imaging (ISBI), 2024, pp. 1-4.

[20] Z. Tang, C. Jiang, Z. Cui, D. Shen, "Hf-resdiff: High-frequency-guided residual diffusion for multi-dose pet reconstruction," In Medical Image Computing and Computer Assisted Intervention – MICCAI 2024, 2024, pp. 372-381.

[21] X. Niu et al., "MDAA-Diff: CT-Guided Multi-dose Adaptive Attention Diffusion Model for PET Denoising," In Medical Image Computing and Computer Assisted Intervention – MICCAI 2025, 2025, pp. 333–343.

[22] B. Yu, L. Zhou, L. Wang, Y. Shi, J. Fripp, and P. Bourgeat, "Ea-GANs: Edge-Aware Generative Adversarial Networks for Cross-Modality MR Image Synthesis," IEEE Transactions on Medical Imaging, vol. 38, no. 7, pp. 1750-1762, Jul. 2019.

[23] S. Chen, G. Bortsova, A. G. Juárez, G. V. Tulder, M. D. Bruijne, "Multi-task Attention-Based Semi-supervised Learning for Medical Image Segmentation," in Medical Image Computing and Computer Assisted Intervention – MICCAI 2019, 2019, pp. 457-465.

[24] A K. Jain, "Fundamentals of digital image processing," Oct. 1988.

[25] H. Hu, X. Wang, Y. Zhang, Q. Chen, Q. Guan, "A comprehensive survey on contrastive learning," Neurocomputing, vol. 610, pp. 128645, Dec. 2024.

[26] G. Wu, G. Jiang, X. Liu, "A Practical Contrastive Learning Framework for Single-Image Super-Resolution," IEEE Transactions on Neural Networks and Learning Systems, vol. 35, no. 11, pp. 15834-15845, Nov. 2024.

[27] S. Xue et al., "A cross-scanner and cross-tracer deep learning method for the recovery of standard-dose imaging quality from low-dose PET," European Journal of Nuclear Medicine and Molecular Imaging, vol. 49, no. 6, pp. 1843–1856, May 2022.

[28] B. Aubert-Broche, A. C. Evans, and L. Collins, "A new improved version of the realistic digital brain phantom," NeuroImage, vol. 32, no. 1, pp. 138–145, Aug. 2006.

[29] B. Aubert-Broche, M. Griffin, G. B. Pike, A. C. Evans, and D. L. Collins, "Twenty new digital brain phantoms for creation of validation image data bases," IEEE Trans. Med. Imag., vol. 25, no. 11, pp. 1410–1416, Nov. 2006.

[30] Y. Cui, S. Zamir, S. Khan, A. Knoll, M. Shah, and F. Khan, "AdaIR: Adaptive All-in-One Image Restoration via Frequency Mining and Modulation," In International Conference on Learning Representations (ICLR), 2025, pp. 57335-57356, [Online]. Available: https://arxiv.org/abs/2403.14614.

[31] O. Ronneberger, P. Fischer, T. Brox, "U-Net: convolutional networks for biomedical image segmentation," In Medical Image Computing and Computer Assisted Intervention – MICCAI 2015, 2015, pp. 234–241.

[32] G. Wu, J. Jiang, X. Liu, "A Practical Contrastive Learning Framework for Single-Image Super-Resolution," IEEE Transactions on Neural Networks and Learning Systems, vol. 35, no. 11, pp. 15834-15845, Nov. 2024.