



# RED: Residual estimation diffusion for low-dose PET sinogram reconstruction



Xingyu Ai<sup>a,1</sup>, Bin Huang<sup>b,1</sup>, Fang Chen<sup>a</sup>, Liu Shi<sup>a</sup>, Binxuan Li<sup>c</sup>, Shaoyu Wang<sup>a,\*</sup>, Qiegen Liu<sup>a,\*</sup>

<sup>a</sup> School of Information Engineering, Nanchang University, Nanchang 330031, China

<sup>b</sup> School of Mathematics and Computer Sciences, Nanchang University, Nanchang 330031, China

<sup>c</sup> Institute of Artificial Intelligence, Hefei Comprehensive National Science Center, Hefei 230000, China

## ARTICLE INFO

### Keywords:

Low-dose PET  
Sinogram reconstruction  
Diffusion model  
Non-Gaussian noise  
Drift correction

## ABSTRACT

Recent advances in diffusion models have demonstrated exceptional performance in generative tasks across various fields. In positron emission tomography (PET), the reduction in tracer dose leads to information loss in sinograms. Using diffusion models to reconstruct missing information can improve imaging quality. Traditional diffusion models effectively use Gaussian noise for image reconstructions. However, in low-dose PET reconstruction, Gaussian noise can worsen the already sparse data by introducing artifacts and inconsistencies. To address this issue, we propose a diffusion model named residual estimation diffusion (RED). From the perspective of diffusion mechanism, RED uses the residual between sinograms to replace Gaussian noise in diffusion process, respectively sets the low-dose and full-dose sinograms as the starting point and endpoint of reconstruction. This mechanism helps preserve the original information in the low-dose sinogram, thereby enhancing reconstruction reliability. From the perspective of data consistency, RED introduces a drift correction strategy to reduce accumulated prediction errors during the reverse process. Calibrating the intermediate results of reverse iterations helps maintain the data consistency and enhances the stability of reconstruction process. In the experiments, RED achieved the best performance across all metrics. Specifically, the PSNR metric showed improvements of 2.75, 5.45, and 8.08 dB in DRF4, 20, and 100 respectively, compared to traditional methods. The code is available at: <https://github.com/yqx7150/RED>.

## 1. Introduction

Medical imaging techniques are essential to modern medicine in disease diagnosis (Beyer et al., 2000; Gunn et al., 2015; Mahac, 2005). Positron emission tomography (PET) is one of these imaging techniques, which reveals changes in metabolic activity and biochemical processes by detecting tracers that are injected into the body. Tracers such as FDG accumulate in organs and lesions over time. Due to its radioactivity, reducing the injection dose helps mitigate the harm to human body. However, it introduces additional noise and artifacts to the imaging results, which impacts diagnostic accuracy (Cui et al., 2023; Xu et al., 2021). When these tracers are annihilated by electrons, pairs of gamma photons are produced and subsequently detected by PET scanner to generate sinograms.

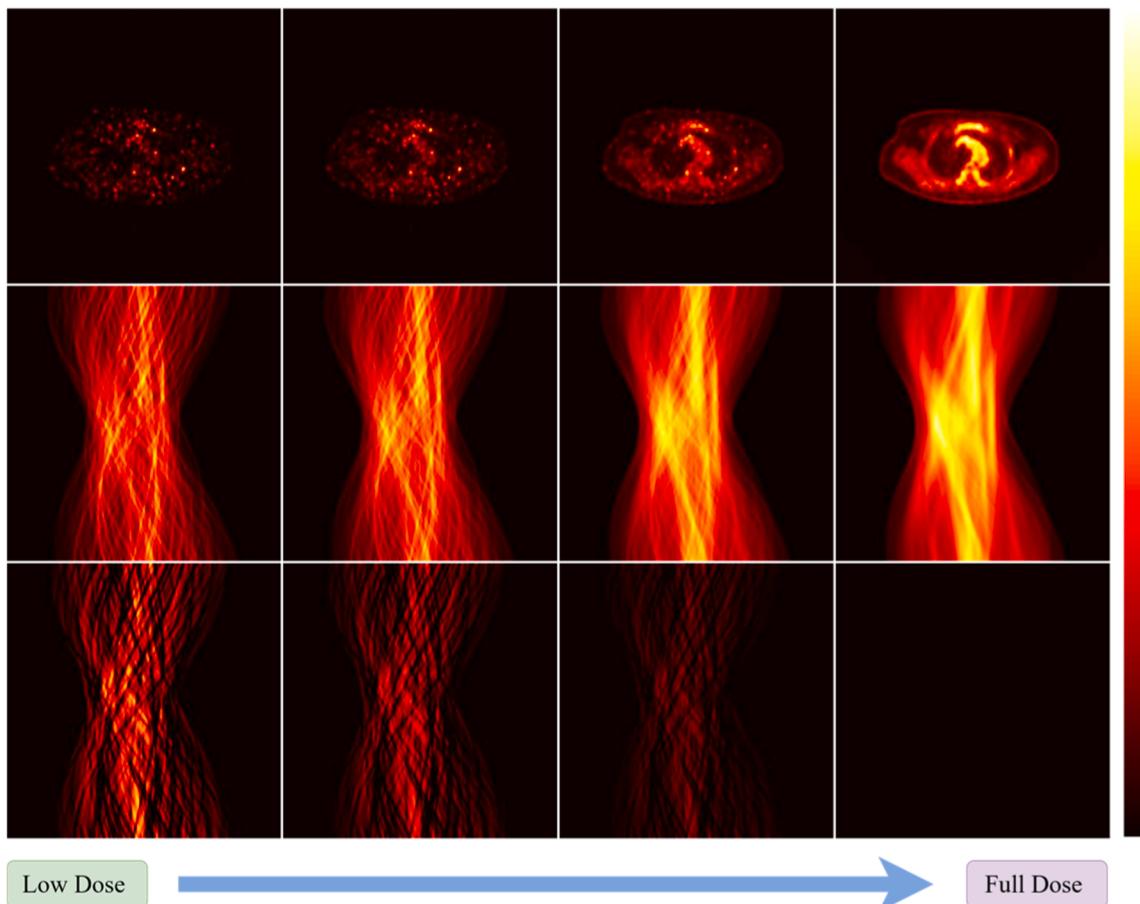
Sinograms can be transferred to image domain via back-projection algorithms. Due to the presence of random noise in raw data, PET

reconstruction is an ill-posed inverse problem. Numerous studies have demonstrated the capability to reconstruct images from sinograms. Filtered back projection (FBP) reconstructs images by back-projecting filtered projection data, achieving rapid image reconstruction. Still, its performance declines in high-noise conditions. In contrast, maximum likelihood expectation maximization (MLEM) optimizes the likelihood of sinograms through iterative reconstruction, but it requires substantial time to converge. Although ordered subset expectation maximization (OSEM) (Hudson and Larkin, 1994) enhances the iteration efficiency, it struggles to compensate for missing information in low-dose situation. Hence, finding effective methods to improve the quality of low-dose sinograms become a crucial area of research.

With the rapid development of machine learning, an increasing number of deep learning-based methods have been applied to PET imaging (Yang et al., 2024; Wang et al., 2024; Gong et al., 2019; Gong et al., 2022). Cui et al. presented a method that integrates sinograms,

\* Corresponding authors.

<sup>1</sup> These authors contributed to this work equally.



**Fig. 1.** The low-dose data (left) has worse quality compared to the full-dose data (right). Graininess and discontinuities increase as the lines in the sinogram become sparser in the projection domain.

images, and frequency data to perform low-dose PET reconstruction (Cui et al., 2024). Hong et al. developed a data-driven sinogram method for single-image super-resolution based on a novel deep residual convolutional neural network, enhancing its effectiveness in PET imaging (Hong et al., 2018). Whiteley and Gregor (2019) presented a sinogram repair network that effectively addresses the missing data caused by block detector failures. Meanwhile, generative models have presented strong capabilities in medical imaging. Generative adversarial networks (GANs) (Goodfellow et al., 2020) involve generating and discriminating networks to enhance generation quality. These networks have seen extensive use in the medical imaging domain (Cui et al., 2024; Hu et al., 2021; Gong et al., 2021; Bera and Biswas, 2024). Nonetheless, GANs face challenges due to their unstable training process, which is prone to training failures (Arjovsky and Bottou, 2017).

Recently, diffusion models have emerged as the leading generative models (Özbey et al., 2023; Li et al., 2024; Luo et al., 2024; Zhang et al., 2024; Dhariwal and Nichol, 2021), which reconstruct images by predicting and removing noise (Yang et al., 2023). Denoising diffusion probabilistic model (DDPM) (Ho et al., 2020) generates images directly from noise. This method iteratively generates high-quality images, rather than producing them in a single prediction. Guan et al. (2024) proposed the generative modeling in sinogram domain (GMSD) approach to perturb sinograms with multi-scale noise and interpolate data. Shen et al. (2024) introduced a novel approach by adding Gaussian noise in the projection domain to simulate the formation of PET noise. Huang et al. (2024) pre-trained an encoding network, then applied a latent diffusion model (Rombach et al., 2022) to enhance the quality of latent embedding. Because DDPM requires a large number of iterative steps, denoising diffusion implicit model (DDIM) (Song et al., 2020)

significantly improves sampling efficiency by generalizing DDPM into a non-Markovian process. These modifications allow the reverse process to achieve similar outcomes with fewer sampling steps and without the need to add noise.

Traditional diffusion models erode input images by adding Gaussian noise through the forward process. Although the noise mechanism helps enhance the diversity of generation results, excessive variability in PET imaging can decrease the fidelity of reconstruction results. Furthermore, adding Gaussian noise to sinograms can produce artifacts after back-projecting to image domain (Oehmigen et al., 2014). Artifacts and noise can lead to misdiagnoses, further reducing the reliability of results. This effect is more pronounced in low-dose sinograms, where the added artificial noise makes already incomplete information even harder to discern.

To address these aforementioned issues, this study proposes a diffusion mechanism called residual estimation diffusion (RED), which adjusts the diffusion process to be more suitable for low-dose PET sinogram reconstruction. RED constructs the diffusion process by minimizing the residual between low-dose and full-dose sinograms. Instead of adding Gaussian noise to data, RED directly uses low-dose sinogram as the starting point and full-dose sinogram as the endpoint of reverse process. In this case, the low-dose sinogram can be considered as an already eroded version of the full-dose one, and no Gaussian noise needs to be added. Since RED avoids compromising the quality of the input in any form, the reverse process is a deterministic sampling process. Although this process reduces the uncertainty of the generation results, the accumulation of prediction errors can lead to data drift. This phenomenon is caused by inaccurate predictions and is more significant under non-Gaussian noise conditions. Therefore, a cascaded estimation

module (CEM) is designed for the reverse process. CEM first estimates the residuals from current sinogram, and then corrects the accumulated errors of it. When sampling training data, instead of directly using actual residual between sinograms, RED uses the predicted residuals to generate intermediate samples. This strategy simulates the occurrence of drift, improves data consistency during training and reconstruction. The main contributions of this paper can be summarized as follows:

- **Residual Estimation Diffusion Mechanism:** Compared to traditional diffusion models, RED replaces Gaussian noise with the residuals between low-dose and full-dose sinograms and reconstructs sinograms via minimizing the residuals. By setting the low-dose and full-dose sinograms as the starting and ending points of reconstruction, the low-dose sinograms can be viewed as the results of adding residuals to the full-dose sinograms and do not need further degradation. This approach avoids extra Gaussian noise, thereby enhancing the reliability of results.
- **Drift Correction Strategy:** To mitigate the impact of data drift on reconstruction results, this study introduces a drift correction mechanism in the reverse process. During reconstruction, CEM estimates the residual and then calibrates the intermediate sinograms. It is trained in a staged manner. CEM first train a network to predict the accurate residuals. After that, the pre-trained network is used to make a single-step prediction to obtain drifted samples. These samples help reproduce the intermediate results of reconstruction, enhancing data consistency and stability in the reverse process.

The rest of this paper is organized as follows: Section 2 introduces low-dose PET sinograms and the background knowledge of diffusion models. Section 3 presents the internal details of RED. Experimental results are revealed in Section 4, followed by additional discussions in Section 5, with the final conclusion drawn in Section 6.

## 2. Related work

### 2.1. PET sinogram reconstruction

The PET scanner organizes acquired raw data into parallel slices. Each slice includes a list of coincidence events. Each event connects two detectors along a line of response (LOR), indicating the path of positron emissions. The raw data are structured in a sinogram, and the reconstruction process can be framed as an inverse problem:

$$b = Ax + n, \quad (1)$$

where  $x$  is the radioactive activity map,  $b$  is the reconstructed image and  $n$  represents noise.  $A$  is the system matrix that encapsulates tomographic geometry and physical factors of the imaging system.

Various reconstruction algorithms, like MLEM and OSEM, can be applied to process sinograms. In low-dose PET scenarios, a reduced tracer dose leads to fewer photons reaching the detector, compromising imaging accuracy (Karimi et al., 2016). Fig. 1 illustrates the sinograms and images reconstructed by OSEM. As the tracer dose decreases, the residuals between the low-dose and full-dose sinusoidal images become more pronounced, indicating more severe information loss. Traditional reconstruction algorithms struggle to reconstruct low-dose sinograms due to the substantial information loss, while deep learning-based generative models offer new approaches to address these challenges.

### 2.2. Noise removal diffusion mechanism

Diffusion models iteratively reconstruct images degraded by noise. Diffusion mechanism is formed by forward and reverse processes. For the forward process, DDPM operates on the input image  $x_0$  and gradually transforms it into Gaussian noise:  $x_t \sim \mathcal{N}(0, I)$  over  $t$  iterations. Each iteration of this process is described as follows:

$$q(x_t | x_{t-1}) = \mathcal{N}\left(x_t; \sqrt{1 - \beta_t}x_{t-1}, \beta_t I\right), \quad (2)$$

where  $\beta_t$  is a fixed variance schedule. Since adding Gaussian noise step by step is too cumbersome, DDPM simplifies the forward sampling process by utilizing the additivity of Gaussian distributions. Let  $\alpha_t$  denotes  $1 - \beta_t$  and  $\bar{\alpha}_t$  represents  $\prod_{i=1}^t \alpha_i$ ,  $x_t$  can be calculated by:

$$x_t = \sqrt{\bar{\alpha}_t}x_0 + \sqrt{1 - \bar{\alpha}_t}\epsilon, \quad (3)$$

where  $\epsilon$  is a noise sampled from  $\mathcal{N}(0, I)$ . Diffusion model generates  $x_{t-1}$  using current time step data  $x_t$  in the reverse process. A neural network with parameters  $\theta$  is used to approximate this process:

$$p_\theta(x_{t-1} | x_t) = \mathcal{N}\left(x_{t-1}; \mu_\theta(x_t, t), \sum_\theta(x_t, t)\right). \quad (4)$$

Due to the Markov property of the diffusion process, the denoising process also requires generating  $x_{t-1}$  from  $x_t$  at each step. As a result, DDIM proposes a non-Markovian sampling process by redefining the sample generation method. In this context, the reverse process does not need to denoise step by step, and  $x_t$  can be used to compute any  $x_s$  for  $s < t$ . Because of that, by using  $e_\theta(x_t, t)$  to predict  $\epsilon$ ,  $x_s$  can be calculated by:

$$\hat{x}_0 = x_t - \sqrt{1 - \bar{\alpha}_t}\epsilon_\theta(x_t, t) / \sqrt{\bar{\alpha}_t}, \quad (5)$$

$$x_{s-1} = \sqrt{\bar{\alpha}_s}\hat{x}_0 + \sqrt{1 - \bar{\alpha}_s}\epsilon_\theta(x_s, t), \quad (6)$$

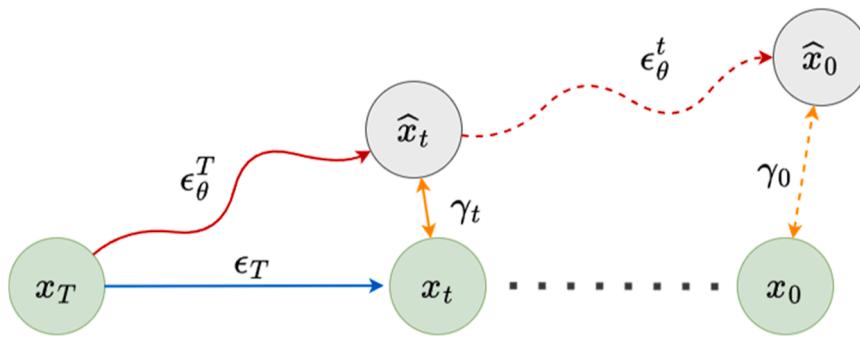
where  $\hat{x}_0$  denotes the current prediction of the original  $x_0$ . Eq. (6) allows DDIM to achieve comparable results with fewer sampling steps, which in turn improves the efficiency of reverse process.

## 3. Proposed RED model

### 3.1. Motivation

The basic idea of RED is to directly utilize low-dose sinograms as the original input to the diffusion model, thereby avoiding compromising the integrity of the original data. Traditional diffusion models transform different data into an approximate prior distribution through Gaussian noise. This mechanism improves the generalization ability, yet adding Gaussian noise to sinograms introduces artifacts into the imaging result. Since low-dose PET sinograms suffer from missing information, extra noise may make existing features harder to distinguish, further complicating the reconstruction. Meanwhile, due to the inherent complexity of PET imaging, the types of noise present in the data may differ from Gaussian noise. Considering these factors, finding a suitable method to model the difference between low-dose and full-dose sinograms can produce a more effective reconstruction process.

Recent studies have validated the feasibility of constructing diffusion models without Gaussian noise. Cold diffusion (Bansal et al., 2023) employs various types of non-Gaussian noise, such as blurring and masking, as degradation operators to process input images. It achieves effective reconstruction across various noise types, yet applying degradation operations will damage the original data. If the residual between two sinograms is used to replace the noise, a given low-dose sinogram can be viewed as the result of a full-dose sinogram with noise already added. Inspired by the iterative noise removal of DDPM, RED replaces the noise in the diffusion process with residuals. In this situation, the low-dose sinogram can be directly used as input without the need of further degradation. In addition, residuals can universally express the differences between sinograms at different doses level without the need to model the specific imaging process. For that reason, RED uses the residual instead of Gaussian noise, and iteratively minimizes the gap between sinograms during the reverse process.



**Fig. 2.** Illustration of data drift in the reverse process. Drift  $\gamma_t$  accumulates due to the imperfect prediction of residual, leading the final reconstruction result to deviate from expectations.

Under optimal conditions, the residuals predicted by RED would precisely match the actual values. However, it is challenging for deep neural networks to achieve perfect predictions in practice (Zhang et al., 2021). Generating content that does not exist in the original data may lead to misdiagnoses, reducing the credibility of predictions in clinical scenarios. Fig. 2 illustrates the data drift in the reverse process. Accumulated errors during reconstruction may cause intermediate results to deviate from training data, which consequently reduces the consistency of data. To mitigate this, RED introduces a drift correction strategy to calibrate the sinogram, thereby reducing sample deviations throughout iterations for a more stable reverse process.

### 3.2. Residual estimation mechanism

RED utilizes the residual term  $\epsilon$  to represent the difference between low-dose sinogram  $x_L$  and full-dose sinogram  $x_F$  which can be described as:

$$x_L = x_F + \epsilon. \quad (7)$$

As a result,  $x_L$  can be obtained from  $x_F$  by adding the residual  $\epsilon$  through  $T$  iterations. Due to the fact that RED does not rely on Gaussian noise assumption, the forward process can be simplified to adding the residual to the full-dose sinogram. For a given time step  $t$ ,  $x_t$  can be obtained by adding a certain proportion of residual:

$$x_t = x_F + \alpha_t \epsilon, \forall t \in \{1, \dots, T\}, \quad (8)$$

where  $T$  and  $\alpha_1, \dots, \alpha_T \in [0, 1]$  represent the number of diffusion steps and the residual schedule across diffusion steps. In the reverse process,  $x_F$  is

defined as endpoint  $x_0$ , and  $x_L$  is defined as starting point  $x_T$ . When  $t$  reaches the endpoint,  $\alpha_0$  is set to 0, indicating that  $x_L$  has completely transformed into a full-dose sinogram. Based on the forward process, RED correspondingly adjusts the reverse process. Hence sample  $x_{t-1}$  can be obtained by:

$$x_{t-1} = x_0 + \alpha_{t-1} \epsilon. \quad (9)$$

By substituting  $x_0$  using Eq. (8),  $x_{t-1}$  can be computed from  $x_t$  as follows:

$$\begin{aligned} x_{t-1} &= (x_t - \alpha_t \epsilon) + \alpha_{t-1} \epsilon \\ &= x_t - (\alpha_t - \alpha_{t-1}) \epsilon. \end{aligned} \quad (10)$$

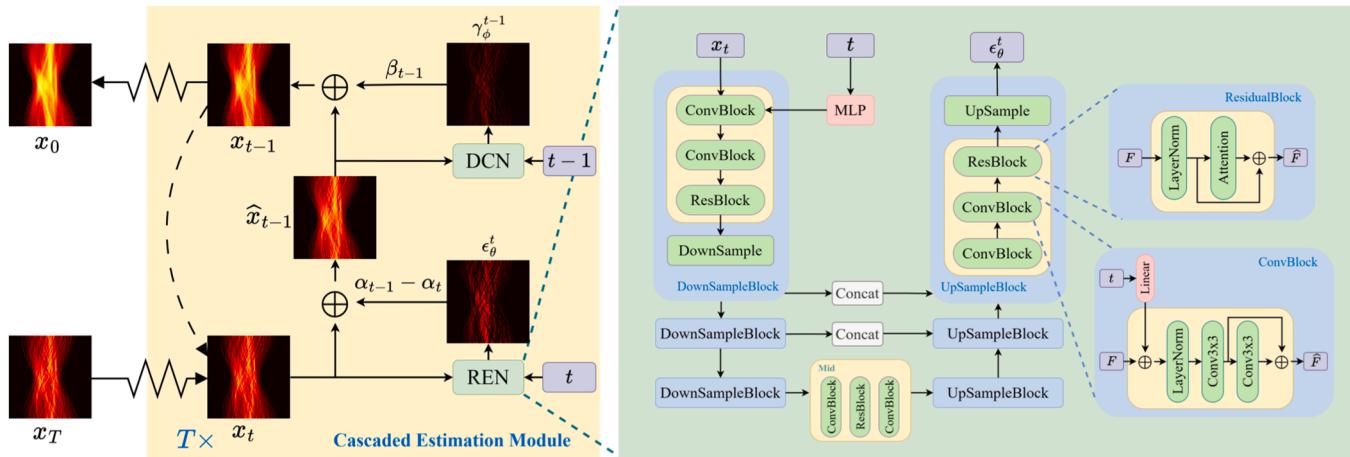
This formula iteratively computes  $x_{t-1}$  from  $x_t$ , forming the reverse process. In this case, the intermediate results can be iteratively computed through the reverse process:

$$x_t = x_T - \sum_{i=T}^t (\alpha_i - \alpha_{i-1}) \epsilon. \quad (11)$$

Meanwhile, the reverse process is deterministic and does not require degradation. Modeling the residual between the data simplifies the diffusion process and establishes a transformation from low-dose to full-dose sinograms.

### 3.3. Drift correction strategy

Due to the value of  $\epsilon$  is unknown in practice, RED employs a pair of cascaded networks to estimate the residual. Fig. 3 illustrates the structure of CEM. It combined by a residual estimation network (REN) with



**Fig. 3.** The overall structure of RED. During the reverse process, REN first estimates the residual  $\epsilon_\theta^t$  to calculate  $\hat{x}_{t-1}$ , where  $\alpha_{t-1} - \alpha_t$  denotes the hyperparameter for adjusting the residual. Then DCN predicts the drift correction term  $\gamma_\phi^{t-1}$  to refine  $\hat{x}_{t-1}$ , adjusted by hyperparameter  $\beta_{t-1}$ . Both REN and DCN use the same network architecture.

parameters  $\theta$  and another drift correction network (DCN) with parameter  $\phi$ .

The prediction of REN can be represented as:

$$\epsilon_\theta^t = REN_\theta(x_t, t). \quad (12)$$

where  $\epsilon_\theta^t$  denotes the predicted residual at time step  $t$ . Thus, the predicted intermediate result  $\hat{x}_t$  can be denote as:

$$\hat{x}_t = x_T - \sum_{i=T}^t (\alpha_i - \alpha_{i-1}) \epsilon_\theta^i. \quad (13)$$

The reverse process of RED divides reconstruction into multiple iterations. As a result, prediction errors accumulate as the number of iterations increases. The actual residual  $\epsilon$  can be represented as:

$$\epsilon = \epsilon_\theta^i - \delta_i, \quad (14)$$

where  $\delta_i$  represents the predicted error at time step  $i$ . Calibrating the errors that arise during the iterative process helps improve the data consistency. Therefore, correction can be conducted by replacing  $\epsilon$  in Eq. (11):

$$\begin{aligned} x_t &= x_T - \sum_{i=T}^t (\alpha_i - \alpha_{i-1})(\epsilon_\theta^i - \delta_i) \\ &= \hat{x}_t + \sum_{i=T}^t (\alpha_i - \alpha_{i-1})\delta_i, \end{aligned} \quad (15)$$

where  $\hat{x}_t$  means the reconstruction result without drift correction. Letting  $\gamma_t = \sum_{i=T}^t (\alpha_i - \alpha_{i-1})\delta_i$ , Eq. (15) can be simplified as:

#### Algorithm 1

Residual Estimation Diffusion (RED).

Training Stage
<b>Require:</b> Schedule $\alpha$ and $\beta$ ; Learning rate $\eta$ , Maximum time step $T$ ; Number of steps $N$ . Parameters $\theta$ and $\phi$ <b>Repeat</b> <b>for</b> $i = 1, 2, \dots, N$ <b>do</b> $x_L \sqcup q(x_L), x_F \sqcup q(x_F), t \sqcup Uniform(0, T)$ Sample $x_t$ according to Eq. (8) $\dot{\alpha}_\theta \leftarrow REN_\theta(x_t, t)$ Take gradient decent step via Eq. (22) $\theta \leftarrow \theta - \eta \nabla_\theta L_{REN}$ <b>end for</b> <b>for</b> $i = 1, 2, \dots, N$ <b>do</b> $x_L \sqcup q(x_L), x_F \sqcup q(x_F), t \sqcup Uniform(0, T)$ Sample $x_t$ according to Eq. (8) $\dot{\alpha}_\theta \leftarrow REN_\theta(x_L, t)$ Sample $\hat{x}_t$ according to Eq. (23) $\gamma'_\phi \leftarrow DCN_\phi(\hat{x}_t, t)$ Take gradient decent step via Eq. (24) $\phi \leftarrow \phi - \eta \nabla_\phi E_{t, (\hat{x}_t, t)} [\ \gamma_t - \gamma'_\phi\ _2^2]$ <b>end for</b> <b>Until converged</b>
Reconstruction Stage
<b>Require:</b> Schedule $\alpha$ and $\beta$ ; Maximum time step $T_{MAX}$ ; Number of iterations $T_s$ ; Trained parameters $\theta^*$ and $\phi^*$ $x_L \sqcup q(x_L)$ $r = T_s / T_{MAX}, t = T_{MAX}, x_t = x_L$ <b>for</b> $s = T_s, \dots, 1, 0$ <b>do</b> $\hat{x}_{t-r} = x_t - (\alpha_t - \alpha_{t-r}) REN_{\theta^*}(x_t, t)$ $x_{t-r} = \hat{x}_{t-r} + \beta_{t-r} DCN_{\phi^*}(\hat{x}_{t-r}, t-r)$ $t = t - r$ <b>end for</b> <b>Return</b> $x_0$

$$\gamma_t = x_t - \hat{x}_t. \quad (16)$$

This formula indicates that the correction of data drift can be obtained by estimating the residual between the current predicted  $\hat{x}_t$  and actual  $x_t$ . RED uses another drift correction network (DCN) with parameter  $\phi$  to predict the drift correction term  $\gamma_\phi^t$ :

$$\gamma_\phi^t = DCN_\phi(\hat{x}_t, t). \quad (17)$$

REN cascades with DCN to form CEM. In the reverse process, CEM first uses REN to calculate  $\epsilon_\theta^t$ , then feeds the predicted  $\hat{x}_{t-1}$  into the DCN to obtain  $\gamma_\phi^{t-1}$ . Hence the reverse process can be represented by the following equations:

$$\hat{x}_{t-1} = x_t - (\alpha_t - \alpha_{t-1}) \epsilon_\theta^t, \quad (18)$$

$$x_{t-1} = \hat{x}_{t-1} + \beta_{t-1} \gamma_\phi^{t-1}, \forall t \in \{1, \dots, T\}, \quad (19)$$

where  $\beta_1, \dots, \beta_T \in [0, 1]$  denotes a schedule to adjust the drift correction. In summary, CEM predicts the overall residual through  $x_t$  and corrects the prediction errors by constraining the drift in  $\hat{x}_{t-1}$ .

#### 3.4. Training and reconstruction

RED performs reconstruction using CEM, which is composed of REN and DCN in cascade. The training of CEM is carried out in stages: first, REN is trained to estimate the residuals, and then the pre-trained REN

generates drifted samples to train DCN.

**1) REN Training:** REN uses a modified U-Net architecture to estimate residuals based on the input sinogram and time step. To construct training samples, a low-dose sinogram  $x_L$  and its corresponding full-dose sinogram  $x_F$  are selected from the dataset. A time step  $t$  is then sampled from a uniform distribution between 0 and the maximum time step  $T$ . Next, REN utilizes Eq. (8) to calculate  $x_t$  as the training sample. Time step  $t$  and  $x_t$  are fed into REN to generate predicted residuals. During training, predicted  $\epsilon_\theta^t$  is compared to the actual  $\epsilon$ , and the network is optimized by minimizing the mean squared error (MSE) loss:

$$\mathcal{L}_{MSE} = \mathbb{E}_{t,(x_L,x_F)} \left[ \| \epsilon - \epsilon_\theta^t \|_2^2 \right]. \quad (20)$$

At the same time, the training introduces another loss function based on structural similarity (SSIM) to evaluate the restoration of the estimated residuals:

$$\mathcal{L}_{SSIM}(x, y) = 1 - \frac{(2\mu_x\mu_y + c_1)(2\sigma_{xy} + c_2)}{(\mu_x^2 + \mu_y^2 + c_1)(\sigma_x^2 + \sigma_y^2 + c_2)}, \quad (21)$$

where  $x$  and  $y$  denote the two input images,  $c_1$  and  $c_2$  are two coefficients.  $\mu_x$  and  $\mu_y$  measure the brightness of input image.  $\sigma_x$  and  $\sigma_y$  are the variances related to contrast.  $\sigma_{xy}$  is the covariance between images  $x$  and  $y$  used to measure the structure similarity. Thus, by minimizing these two loss functions, the total loss and objective function can be written in the following form:

$$\mathcal{L}_{REN} = \mathcal{L}_{MSE} + \mathcal{L}_{SSIM}(\hat{x}_0, x_0); \theta^* = \operatorname{argmin}_{\theta} \mathcal{L}_{REN}, \quad (22)$$

where  $\hat{x}_0$  is the reconstructed sinogram. By minimizing the error of prediction, the similarity between low-dose and full-dose sinograms can be enhanced.

**2) DCN Training:** RED utilizes DCN for data drift correction, and DCN uses a pre-trained REN to generate drifted training samples. For a random time step  $t$ , samples can be generated through reverse process. However, performing a complete reverse process would significantly increase the computational complexity. To improve training efficiency, DCN uses REN for a one-step prediction and uses the predicted residual to generate drifted training sample  $\hat{x}_t$ . REN first estimates the residual of  $x_T$ . The predicted  $\epsilon_\theta^T$  is then interpolated with  $\epsilon$  using a random coefficient  $\lambda \sim Uniform(0, 1)$ , and the result is substituted into Eq. (8) to generate drifted training samples:

$$\hat{x}_t = x_F + (1 - \alpha_t)(\lambda\epsilon_\theta^T + (1 - \lambda)\epsilon). \quad (23)$$

Since  $x_L$  suffers more information loss compared to the intermediate results, it can produce more pronounced drifted samples for training. When the predictions of REN become more accurate, the drift will correspondingly decrease. The deviation between the generated sample and the actual value can be calculated by Eq. (16), and the optimization objective is to minimize MSE between the predicted and the actual drift:

$$\phi^* = \operatorname{argmin}_{\phi} \mathbb{E}_{t,(\hat{x}_t,x_t)} \left[ \| \gamma_t - \gamma_\phi^t \|_2^2 \right]. \quad (24)$$

**3) RED Reconstruction:** Since the forward process of RED is not a Markov chain, accelerated sampling methods can be employed in the reverse process to speed up sampling. For a given number of

iterations  $T_s$ , the skipped sampling steps  $r$  can be obtained by calculating the ratio of  $T_s$  to the original maximum time step  $T_{MAX}$ .

$$r = T_s / T_{MAX}. \quad (25)$$

By modifying Eq. (18), for a given sample  $x_t$ , the previous sample  $\hat{x}_{t-r}$  can be calculated by:

$$\hat{x}_{t-r} = x_t - (\alpha_t - \alpha_{t-r})\epsilon_\theta^t. \quad (26)$$

In this case,  $t - r$  represents the next sampling time step after skipping a certain ratio of reverse iterations. Algorithm 1 presents the specific process of training and reconstruction.

This training process can be iterated multiple times to enhance the generalization ability and improves the final reconstruction performance.

#### 4. Experiments

To validate the performance of RED, this section compares it with OSEM (Hudson and Larkin, 1994), U-Net (Ronneberger et al., 2015), DDIM (Song et al., 2020), cold diffusion (CD) (Bansal et al., 2023) and denoising diffusion bridge model (DDBM) (Zhou et al., 2023) in reconstruction tasks. The raw data used in the dataset was obtained through OSEM, while other methods first take low-dose sinograms as input and then utilize the FBP algorithm to reconstruct image results (Häggström et al., 2019). The experiment introduces DDIM and CD as different types of diffusion models to compare their effectiveness with RED. Additionally, U-Net and DDBM were included in the comparison to assess the performance of RED with supervised methods to provide more comprehensive comparisons. All models were trained using the same device and dataset to ensure consistency and fairness.

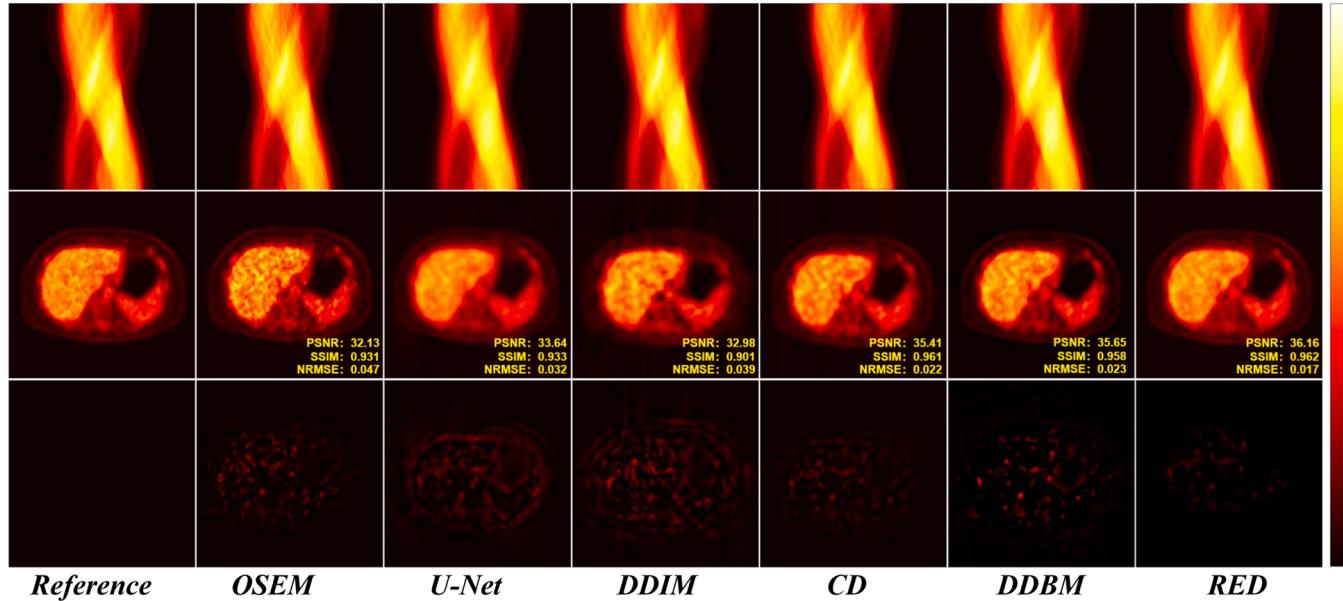
##### 4.1. Data and experimental setting

**1) Dataset:** Two different datasets were used in the experiments (UDPET and Patient Data). **UDPET** dataset is a public dataset from MICCAI 2022 Ultralow-Dose PET (UDPET) Imaging Challenge. Each PET image corresponds to a low-dose PET with a specific dose reduction factor (DRF) and was reconstructed from counts of a time window resampled at the midpoint of acquisition. Each patient includes imaging results under different doses and each dose comprising 673 slices. This experiment utilized  $^{18}\text{F}$ -FDG PET imaging subjects, acquired using uEXPLORER under DRFs of 4, 20, and 100, with each case comprising 500 patients, resulting in a total of 336,500 2D slices for training purposes. Additionally, 1000 extra slices were selected separately to validate the effectiveness at different dose levels. Due to the significant differences in values between original images, the images were first normalized based on the injection dose and weight of patient to calculate the standardized uptake value (SUV) of radioactive activity. The images in the dataset were reconstructed using OSEM, and the corresponding sinograms were obtained by back-projecting the images (Xue et al., 2022). To better accommodate deeper network structures, the sinograms were slightly resized from  $360 \times 360$  to  $352 \times 352$ . All these training data were normalized before being input into the network. During training and prediction, the maximum value of the low-dose sinogram is used for normalization. These parameters are retained after prediction to restore the output to original data range. **Patient Data** selected from 10 patients were used to further assess the generalization capability of these models. The data was sampled at 25% dose from DigitMI 930 PET/CT scanner for validation. This scanner was developed by RAYSOLUTION Healthcare Co., Ltd. and features All-Digital PET detectors with an axial field-of-view (AFOV) of 30.6 cm. Each scan covered 4 to 8 beds, with scan times ranging from 45

**Table 1**

Average Quantitative Reconstruction Quality of Different Methods at DRF 4, 20, 100.

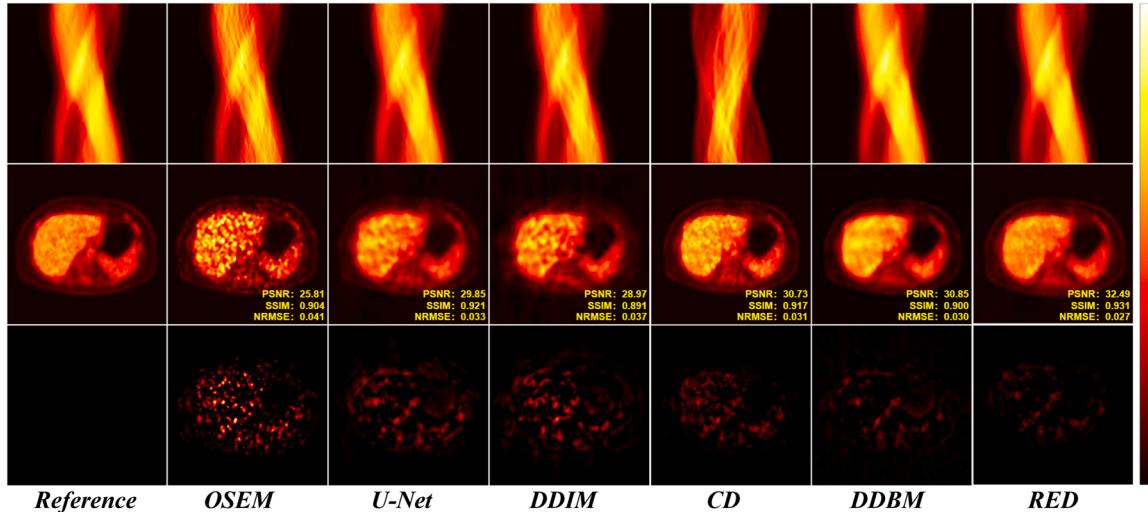
Method	DRF 4 (25%)			DRF 20 (5%)			DRF 100 (1%)		
	PSNR↑	SSIM↑	NRMSE↓	PSNR↑	SSIM↑	NRMSE↓	PSNR↑	SSIM↑	NRMSE↓
OSEM (Hudson and Larkin, 1994)	36.82	0.956	0.037	29.48	0.896	0.063	20.13	0.802	0.109
U-Net (Ronneberger et al., 2015)	37.58	0.967	0.015	32.93	0.929	0.023	24.30	0.844	0.042
DDIM (Song et al., 2020)	36.95	0.937	0.024	31.19	0.906	0.022	23.89	0.839	0.046
CD (Bansal et al., 2023)	38.21	0.972	0.022	33.21	0.937	0.020	25.98	0.851	0.031
DDBM (Zhou et al., 2023)	38.96	0.975	0.018	33.68	0.929	0.021	24.83	0.846	0.044
RED	<b>39.57</b>	<b>0.977</b>	<b>0.012</b>	<b>34.93</b>	<b>0.952</b>	<b>0.018</b>	<b>28.21</b>	<b>0.877</b>	<b>0.028</b>

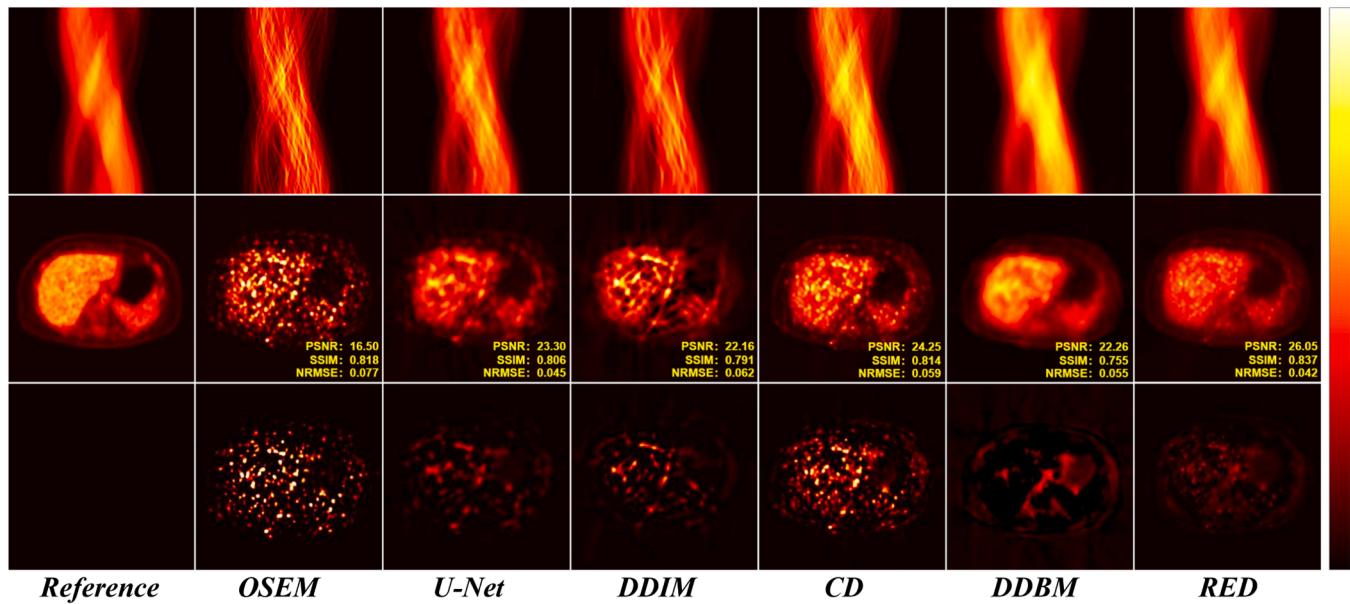
**Fig. 4.** Comparison of reconstruction under DRF 4 using different methods. The first row illustrates the sinograms, the second row presents the reconstruction results, and the last row displays the residuals.

seconds to 3 minutes per bed. Low-dose PET data was obtained by resampling the listmode data into 2 ms intervals, retaining 1 ms of data per cycle while discarding the rest.

2) **Experimental Setup:** RED was trained on a GPU (NVIDIA GeForce RTX 4090 24GB) using the PyTorch framework. The AdamW optimizer (Loshchilov and Hutter, 2017) was used, with an initial

learning rate set to  $1 \times 10^{-4}$ , which was gradually decayed by half every  $1 \times 10^5$  iterations. The maximum sampling step  $T_{MAX}$  was set to 500 and  $T_s$  was set to 20. The parameters  $\alpha$  and  $\beta$  represent the mixing ratio at each timestep and are used to control the reverse process.  $\alpha$  was calculated by subtracting the squared value of the normalized timestep from 1, and  $\beta$  was set to 0.05. Various

**Fig. 5.** Comparison of reconstruction under DRF 20 using different methods. The first row illustrates the sinograms, the second row presents the reconstruction results, and the last row displays the residuals.



**Fig. 6.** Comparison of reconstruction under DRF 100 using different methods. The first row illustrates the sinograms, the second row presents the reconstruction results, and the last row displays the residuals.

**Table 2**  
Average Quantitative Quality of Sinogram at DRF 20.

Method	DRF 20 (5%)		
	PSNR↑	SSIM↑	NRMSE↓
OSEM (Hudson and Larkin, 1994)	30.67	0.824	0.072
U-Net (Ronneberger et al., 2015)	32.54	0.837	0.035
DDIM (Song et al., 2020)	30.51	0.863	0.039
CD (Bansal et al., 2023)	31.42	0.916	0.029
DDBM (Zhou et al., 2023)	33.70	0.930	0.032
<b>RED</b>	<b>35.71</b>	<b>0.947</b>	<b>0.027</b>

evaluation metrics were employed in the experiment: peak signal-to-noise ratio (PSNR), structural similarity index (SSIM), and normalized root mean squared error (NRMSE). These evaluation metrics in the experiment section are all represented by the average metrics calculated on the test set.

#### 4.2. Reconstruction experiments

**1) Imaging Result Comparisons:** The evaluation results are presented in [Table 1](#), with RED achieving the best performance across all evaluation metrics.

Due to the high quality of the initial data, all models perform relatively well under DRF 4, with RED achieving the highest PSNR value of 39.57 dB. Furthermore, RED consistently delivered the best results under DRF 20. This trend became even more pronounced at DRF 100, where RED outperforms OSEM by 8.08 dB and surpasses the second-ranked CD by 2.23 dB in terms of PSNR. [Figs. 4–6](#) display the sinograms and reconstruction results at different dose levels. Overall, the results of RED are smoother while retaining more details. At DRF 4, images of RED exhibit fewer noise and artifacts, whereas other methods signify more graininess within the organs. Under DRF 20, the data quality degraded, and the unsupervised diffusion methods perform poorly. Although U-Net manages to reconstruct the lost information, it produces blurry images. In contrast, results of RED are more similar to the full-dose images, achieving smooth interiors while preserving sharp edges.

At DRF 100, the result of OSEM becomes barely recognizable, and U-

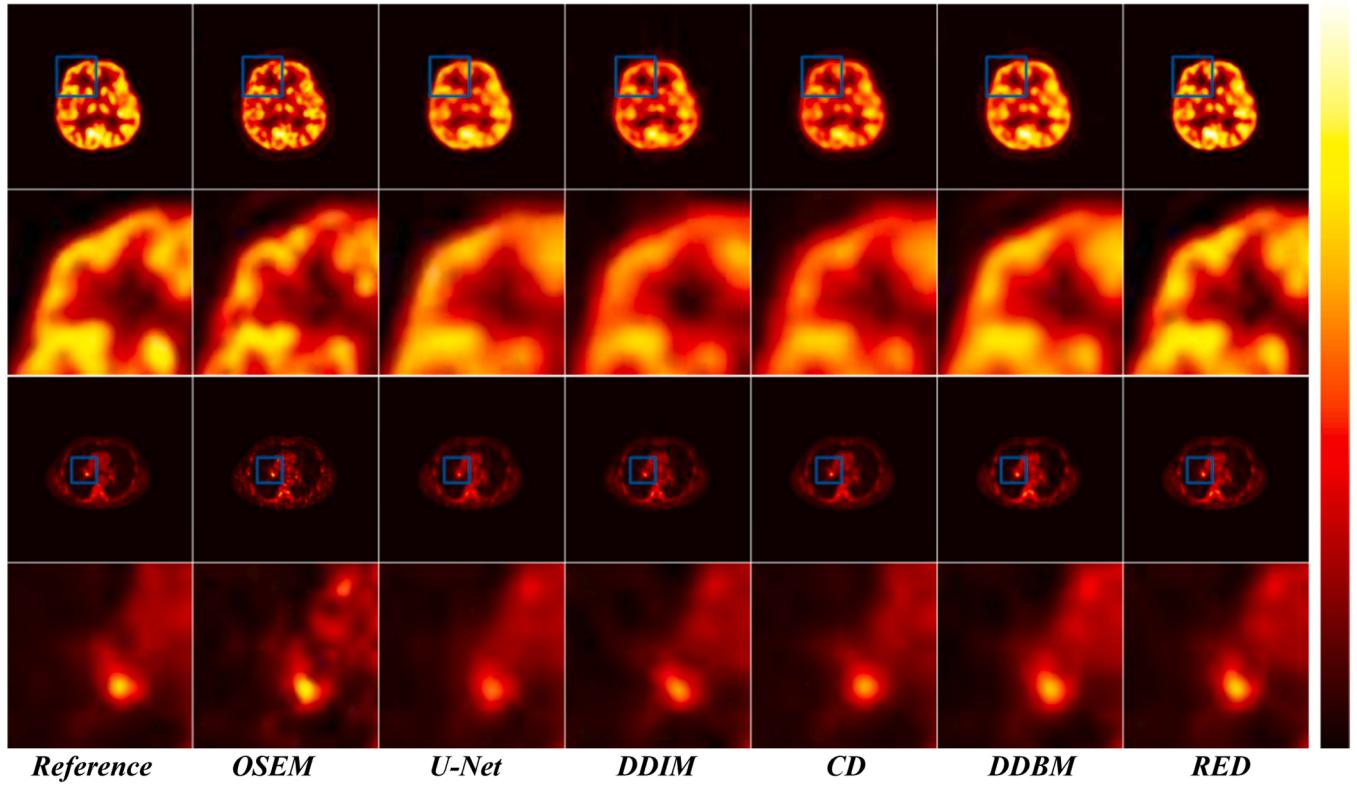
**Table 3**  
Average Quantitative Quality at Critical Area.

Method	DRF 20 (5%)		
	PSNR↑	SSIM↑	NRMSE↓
OSEM (Hudson and Larkin, 1994)	28.97	0.883	0.034
U-Net (Ronneberger et al., 2015)	31.18	0.904	0.029
DDIM (Song et al., 2020)	30.22	0.891	0.032
CD (Bansal et al., 2023)	31.42	0.916	0.029
DDBM (Zhou et al., 2023)	32.17	0.923	0.028
<b>RED</b>	<b>32.78</b>	<b>0.957</b>	<b>0.026</b>

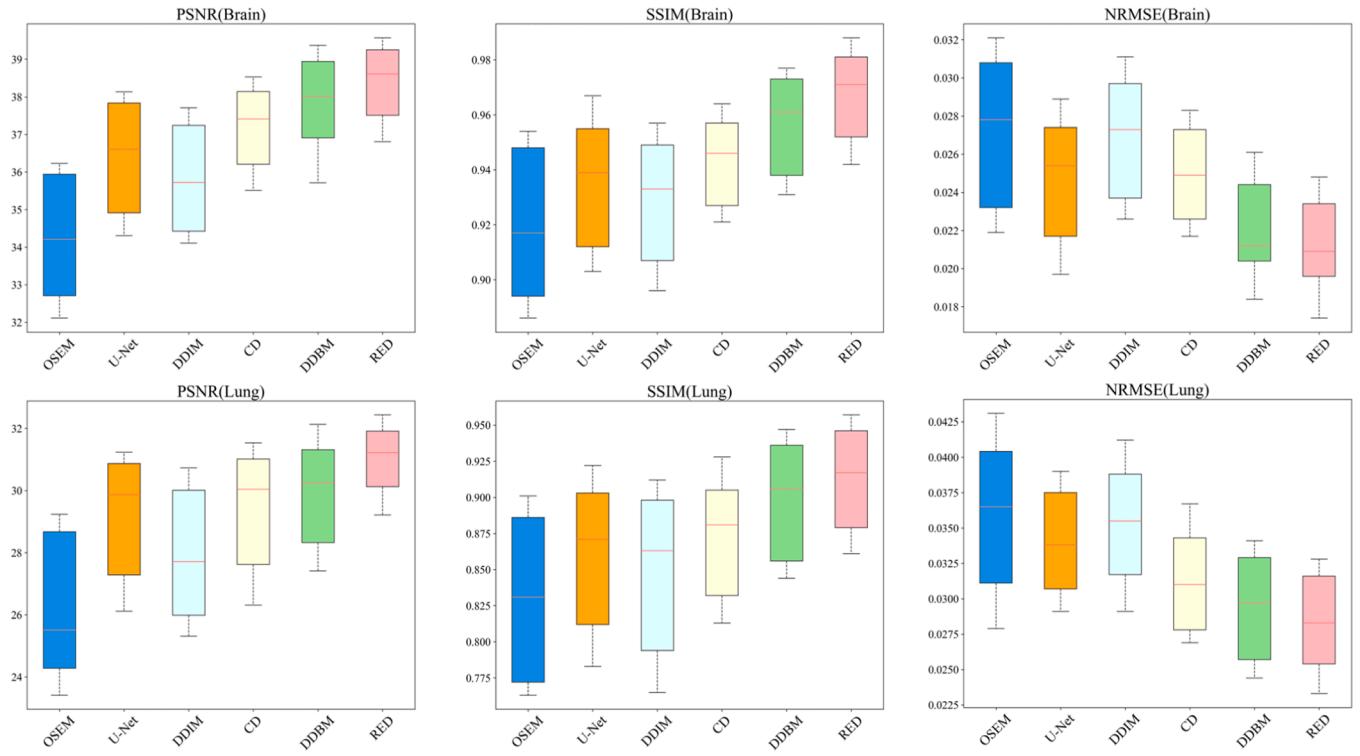
Net introduces considerable blurring. Both DDIM and CD recover few details. The performance of DDBM degrades noticeably in this scenario, with the reconstructed results showing over-smoothing, along with an increased presence of artifacts and distortions. Despite the extremely poor conditions, RED still manages to restore part of the information to maintain structural integrity. The comparison of reconstructed sinograms from different methods provides more intuitive evaluation. [Table 2](#) directly presents the evaluation metrics in the projection domain under DRF 20. RED achieved the best results across all evaluation metrics and demonstrated excellent capability with a PSNR improvement of 5.04 dB compared to OSEM. While other methods indicated lower similarity, which may cause artifacts and noise after back-projection.

**2) Lesion Comparisons:** PET imaging is widely used for disease diagnosis in the brain ([Farmanfarma et al., 2004](#)) and lung ([Coleman, 1999](#)). For this experiment, slices 80 to 450 from each patient were selected as the critical area for comparison. This range includes brain and torso, excluding the leg areas and blank scanning slices.

The evaluation of critical area under DRF 20 is presented in [Table 3](#). RED achieved the best performance and demonstrated its effectiveness of imaging critical areas. Furthermore, [Fig. 7](#) provides a more intuitive comparison of the imaging results for the brain and lung regions. Under this circumstance, RED not only preserved image details but also avoided excessive smoothing. Additionally, [Fig. 8](#) presents the box plots of evaluation metrics for brain and lung region of patient. The PSNR box plot of RED shows a smaller interquartile range for both the brain and



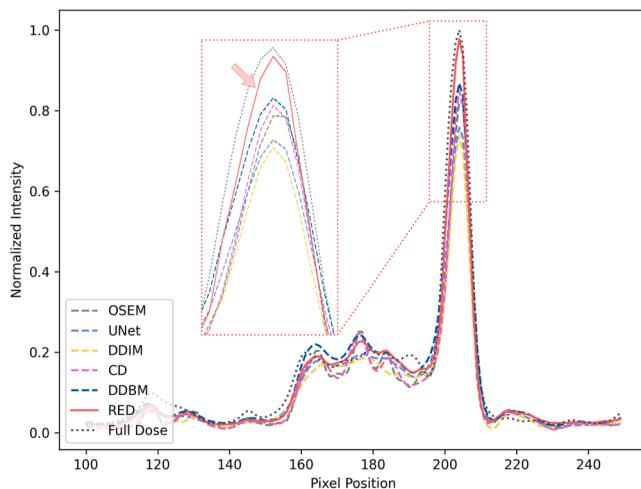
**Fig. 7.** Illustration of the brain and lung regions under DRF 20. RED preserves more details while also producing smoother results.



**Fig. 8.** The box plots for the brain and lungs under DRF20, with the maximum and minimum values marked by dots.

lungs, indicating more stable predictions. In contrast, traditional algorithms show a wider range, with greater fluctuations in the evaluation metrics. Similar conclusions can be drawn from the SSIM and RMSE box plots.

The profile of the lesion area is expressed in Fig. 9. The imaging results of RED are closest to the true values. This experiment proves the reliability of RED, whereas other methods exhibited more noticeable discrepancies.



**Fig. 9.** The profile lines of each model after normalization. The result shows that the recovery values of RED are closest to the real data.

**Table 4**  
Impact of Different Components at DRF 20.

Metrics	(w/o) DC+ SL	(w/o) DC	(w/o) SL	RED
PSNR↑	32.12	33.72	33.91	<b>34.93</b>
SSIM↑	0.914	0.947	0.933	<b>0.952</b>
NRMSE↓	0.035	0.023	0.021	<b>0.018</b>

#### 4.3. Ablation study

RED employs DCN to correct the iterative process during reconstruction and introduces an SSIM-based loss function when training the REN. To evaluate the contribution of these strategies, Table 4 presents the changes in evaluation metrics under DRF 20 after removing the drift correction (DC) and the SSIM-based loss (SL). The results signify that removing either of these measures led to a decline in performance, whereas the RED model, which simultaneously utilized both methods, achieved the best test results, demonstrating the effectiveness of both data correction and the reconstruction loss function.

#### 4.4. Generalization analysis

To further validate the generalization performance of RED, the experiment employed the Patient Dataset for performance evaluation.

**Fig. 10** manifests a comparison of reconstruction results of different models on this dataset, while Table 5 lists the performance metrics for each model. The experimental results present that RED has strong generalization capability, maintaining good performance even when the dataset is changed.

#### 4.5. Experiment of reverse process

The diffusion models compared in the above experiments involve different diffusion strategies: DDIM reconstructs images from Gaussian noise, and the deblurring CD generates clear images from blurred ones. Both methods erode low-dose images during processing, whereas RED directly reconstructs from the low-dose sinogram. During the diffusion process, DDIM reconstructs sinogram with added noise and spends most of its iterative steps for noise removal. Meanwhile, CD retains initial information, leading to a higher initial performance.

The reverse process for these models is presented in Fig. 11. As the diffusion process continues, CD exhibited excessive deblurring which ultimately decreased the imaging quality. In contrast, since RED retains the original information from the low-dose sinogram, it starts with a high initial similarity and maintains continuous quality improvement throughout the reverse process, achieving the best reconstruction results.

## 5. Discussion

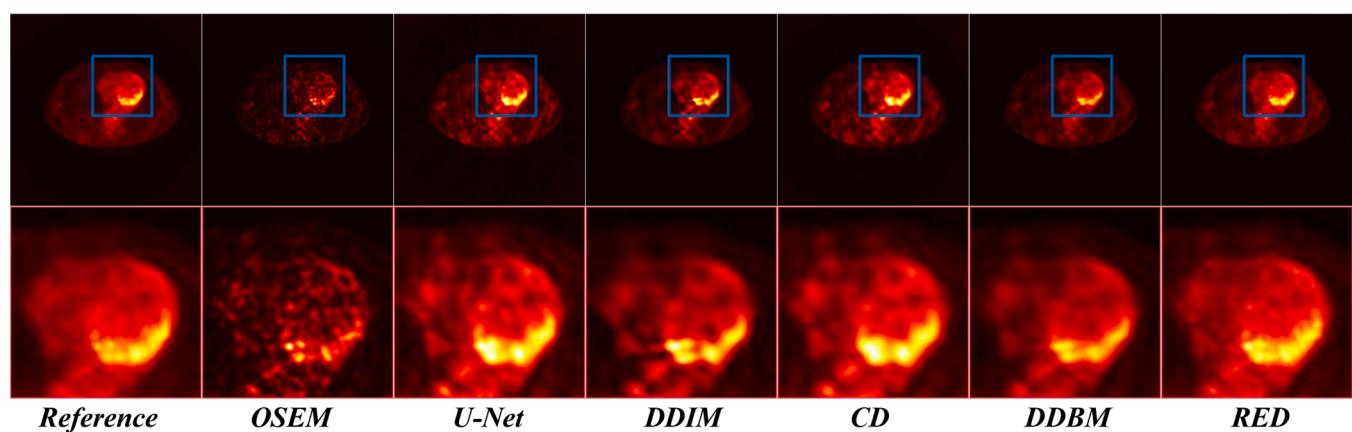
### 5.1. Combine with Gaussian noise

RED demonstrates excellent performance in sinogram reconstruction for low-dose PET. Avoiding the addition of noise and preserving as much of the original information as possible helps alleviate the learning difficulty. Although RED dose not further damaging the original image, experiments have suggested that RED remains effective even when Gaussian noise is added.

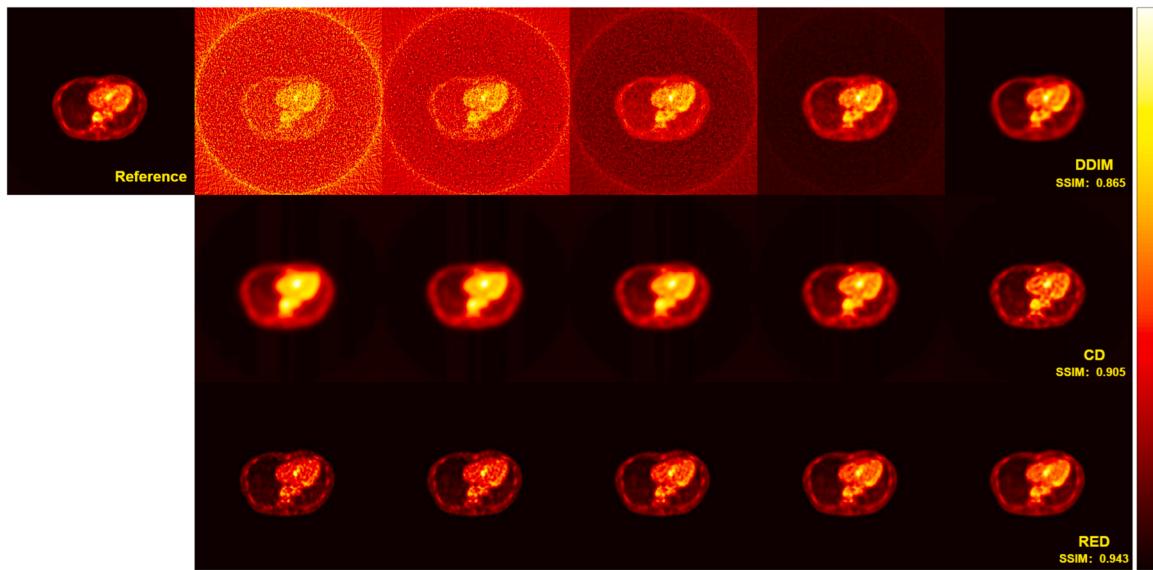
Table 6 demonstrates the evaluation metrics under DRF 20 after

**Table 5**  
Average Quantitative Quality on Patient Dataset.

Method	DRF 20 (5%)		
	PSNR↑	SSIM↑	NRMSE↓
OSEM (Hudson and Larkin, 1994)	27.39	0.875	0.032
U-Net (Ronneberger et al., 2015)	29.08	0.898	0.034
DDIM (Song et al., 2020)	27.57	0.877	0.041
CD (Bansal et al., 2023)	28.02	0.913	0.039
DDBM (Zhou et al., 2023)	29.34	0.920	0.036
<b>RED</b>	<b>31.78</b>	<b>0.931</b>	<b>0.027</b>



**Fig. 10.** Comparison of the reconstruction results using Patient Dataset, where RED outperforms the other comparison schemes in detail addition, showing better continuity.



**Fig. 11.** Comparison of the reverse process under DRF 20. The first column is the full-dose reference image, and the others represent the intermediate results of different diffusion mechanisms.

**Table 6**

Effective Comparisons of Mixed Gaussian Noise at DRF 20.

Metrics	DDIM	Supervised RED	Unsupervised RED	RED
PSNR	31.19	32.79	32.13	<b>34.93</b>
SSIM	0.906	0.914	0.902	<b>0.952</b>
NRMSE	0.022	0.021	0.024	<b>0.018</b>

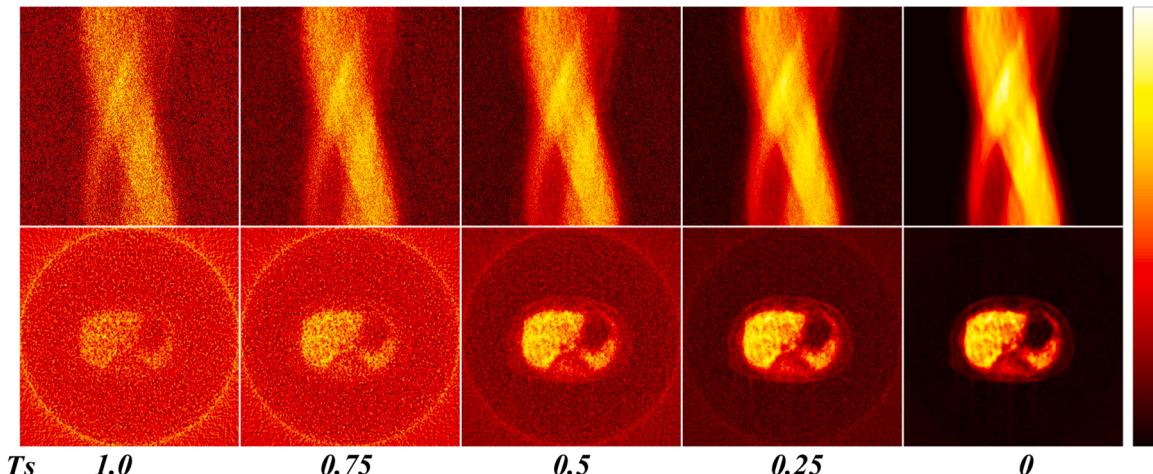
adding Gaussian noise. In this situation, the residual can be defined as the difference between two sinograms with added noise for supervised learning. Alternatively, it can be treated as pure noise, enabling the construction of RED in an unsupervised manner. This characteristic makes the method promising for use in data-limited situations or other specific environments.

Fig. 12 presents the sinogram during the reverse process and the corresponding back-projection results. RED is still able to effectively remove noise even after adding Gaussian noise, and the overall process remains fully functional. This indicates that RED offers a variety of options in selecting degradation operators such as blurring and masking, allowing it to be combined with different types of noise.

### 5.2. Limitations

RED is a diffusion model that operates in 2D projection space, which helps preserve original information of the input sinogram. However, this also results in a higher computational cost, which becomes more significant in 3D PET reconstruction. A feasible approach is to combine the current RED with latent space diffusion, using latent space embeddings to replace the original sinograms. Latent space-based diffusion models have been widely used in the field of natural images. By encoding and compressing sinograms, the computational cost during the diffusion process can be greatly reduced, and it also facilitates integration with other features.

Additionally, RED is designed with separate REN and DCN components to predict the residuals and drift independently, which gives DCN high flexibility in practice. However, since the focus of this work is to validate the effectiveness of the residual diffusion mechanism, other practical approaches for DCN were not extensively explored. For supervised learning, defining an appropriate optimization objective can be challenging. A feasible improvement would be to use GANs to train DCN, such as employing WGAN (Arjovsky et al., 2017). Training DCN with a discriminator network may simplify the design of the loss



**Fig. 12.** The intermediate results of RED after adding Gaussian noise to inputs. RED effectively removes the added noise during the iterative process.

function and enhance its effectiveness.

## 6. Conclusion

This study introduced RED as a high-fidelity diffusion mechanism that minimized the residuals between sinograms. Unlike standard diffusion models, RED did not introduce extra noise to low-dose sinograms, thereby preserving the integrity of the original data. To reduce the impact of accumulated prediction error during the iterations, RED introduced a drift correction mechanism and employed estimated samples during training to enhance data consistency. Further exploration of combining RED with Gaussian noise demonstrated great potential for broader applications in other domains. This offered promising directions for future research to expand the application of RED in various medical imaging tasks and beyond.

## CRediT authorship contribution statement

**Xingyu Ai:** Writing – original draft, Visualization, Validation, Software, Methodology, Data curation, Conceptualization. **Bin Huang:** Writing – review & editing, Software, Methodology, Conceptualization. **Fang Chen:** Validation, Investigation, Formal analysis, Data curation. **Liu Shi:** Visualization, Resources, Investigation, Formal analysis. **Bin-xuan Li:** Visualization, Validation, Formal analysis, Data curation. **Shaoyu Wang:** Writing – original draft, Supervision, Methodology, Data curation. **Qiegen Liu:** Supervision, Resources, Investigation, Funding acquisition.

## Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## Acknowledgements

This work was supported in part by the National Key Research and Development Program of China under 2023YFF1204300, 2023YFF1204302, National Natural Science Foundation of China under 621220033 and 62201193.

## Data availability

The authors do not have permission to share data.

## References

- Arjovsky, M., Bottou, L., 2017. Towards principled methods for training generative adversarial networks. *ICLR*.
- Arjovsky, M., Chintala, S., Bottou, L., 2017. Wasserstein Generative Adversarial Networks. *70. PMLR*, pp. 214–223.
- Bansal, A., Borgnia, E., Chu, H., et al., 2023. Cold diffusion: inverting arbitrary image transforms without noise. *NeurIPS* 36, 41259–41282.
- Bera, S., Biswas, P.K., 2024. Axial consistent memory GAN with interslice consistency loss for low dose computed tomography image denoising. *IEEE Trans. Radiat. Plasma Med. Sci.* 8 (2), 173–183.
- Beyer, T., et al., 2000. A combined PET/CT scanner for clinical oncology. *J. Nucl. Med.* 41 (8), 1369–1379.
- Coleman, R.E., 1999. PET in lung cancer. *J. Nucl. Med.* 40 (5), 814–820.
- Cui, J., Wang, Y., Zhou, L., et al., 2024. 3D point-based multi-modal context clusters GAN for low-dose PET image denoising. *IEEE Trans. Circuits Syst. Video Technol.*
- Cui, J., Zeng, P., Zeng, X., et al., 2023. TriDo-former: a triple-domain transformer for direct PET reconstruction from low-dose sinograms. *Med. Image Comput. Comput. Assist. Interv.* 14229, 184–194.
- Cui, J., Zeng, P., Zeng, X., et al., 2024. Prior knowledge-guided triple-domain transformer-GAN for direct PET reconstruction from low-count sinograms. *IEEE Trans. Med. Imaging*.
- Dhariwal, P., Nichol, A., 2021. Diffusion models beat GANs on image synthesis. *NeurIPS* 34, 8780–8794.
- Farmanfarma, K.H.K., Mohammadian, M., Shahabinia, Z., et al., 2004. Brain cancer in the world: an epidemiological review. *World Cancer Res. J.* 49 (11), 2209.
- Gong, K., Catana, C., Qi, J., et al., 2019. PET image reconstruction using deep image prior. *IEEE Trans. Med. Imaging* 38 (7), 1655–1665.
- Gong, K., Catana, C., Qi, J., et al., 2022. Direct reconstruction of linear parametric images from dynamic PET using nonlocal deep image prior. *IEEE Trans. Med. Imaging* 41 (3), 680–689.
- Gong, Y., Shan, H., Teng, Y., et al., 2021. Parameter-transferred wasserstein generative adversarial network (PT-WGAN) for low-dose PET image denoising. *IEEE Trans. Radiat. Plasma Med.* 5 (2), 213–223.
- Goodfellow, I., Abadie, J.P., Mirza, M., 2020. Generative adversarial networks. *ACM* 63 (11), 139–144.
- Guan, B., Yang, C., Zhang, L., et al., 2024. Generative modeling in sinogram domain for sparse-view CT reconstruction. *IEEE Trans. Radiat. Plasma Med. Sci.* 8 (2), 195–207.
- Gunn, R.N., Sifstein, M., Searle, G.E., et al., 2015. Quantitative imaging of protein targets in the human brain with PET. *Phys. Med. Biol.* 60 (22), R363–R411.
- Häggström, I., Schmidlein, C.R., Campanella, G., et al., 2019. DeepPET: A deep encoder-decoder network for directly solving the PET image reconstruction inverse problem. *Med. Image Anal* 54, 253–262.
- Ho, J., Jain, A., Abbeel, P., 2020. Denoising diffusion probabilistic models. *NeurIPS* 33, 6840–6851.
- Hong, X., Zan, Y., Weng, F., et al., 2018. Enhancing the image quality via transferred deep residual learning of coarse PET sinograms. *IEEE Trans. Med. Imaging* 37 (10), 2322–2332.
- Hu, Z., Xue, H., Zhang, Q., et al., 2021. DPIR-Net: direct PET image reconstruction based on the wasserstein generative adversarial network. *IEEE Trans. Radiat. Plasma Med. Sci* 5 (1), 35–43.
- Huang, B., Liu, X., Fang, L., et al., 2024. Diffusion Transformer Model Using Compact Prior For Low-Dose PET Reconstruction *arXiv preprint arXiv:2407.00944*.
- Hudson, H.M., Larkin, R.S., 1994. Accelerated image reconstruction using ordered subsets of projection data. *IEEE Trans. Med. Imaging* 13 (4), 601–609.
- Karimi, D., Deman, P., Ward, R., et al., 2016. A sinogram denoising algorithm for low-dose computed tomography. *BMC Med Imaging* 16 (11), 1–14.
- Li, Y., Shao, H., Liang, X., et al., 2024. Zero-shot medical image translation via frequency-guided diffusion models. *IEEE Trans. Med. Imaging* 43 (3), 980–993.
- Loshchilov, I., Hutter, F., 2017. Fixing Weight Decay Regularization in Adam, 5 *arXiv preprint arXiv:1711.05101*.
- Luo, M., Zhou, N., Wang, T., et al., 2024. Bi-constraints diffusion: A conditional diffusion model with degradation guidance for metal artifact reduction. *IEEE Trans. Med. Imaging*.
- Mahac, J., 2005. Cardiac positron emission tomography imaging. *Semin. Nucl. Med.* 35 (1), 17–36.
- Oehmigen, M., Ziegler, S., Jakoby, B.W., et al., 2014. Radiotracer dose reduction in integrated PET/MR: implications from national electrical manufacturers association phantom studies. *J. Nucl. Med* 55 (8), 1361–1367.
- Özbey, M., Dalmaz, O., Dar, S.U.H., et al., 2023. Unsupervised medical image translation with adversarial diffusion models. *IEEE Trans. Med. Imaging* 42 (12), 3524–3539.
- Rombach, R., Blattmann, A., Lorenz, D., et al., 2022. High-resolution image synthesis with latent diffusion models. In: *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit*, pp. 10684–10695.
- Ronneberger, O., Fischer, P., Brox, T., 2015. U-Net: convolutional networks for biomedical image segmentation. *Med. Image Comput. Comput. Assist. Interv.* 234–241.
- Shen, C., Tie, C., Yang, Z., et al., 2024. Bidirectional condition diffusion probabilistic models for PET image denoising. *IEEE Trans. Radiat. Plasma Med. Sci* 8 (4), 402–415.
- Song, J., Meng, C., Ermon, S., 2020. Denoising diffusion implicit models. *ICLR*.
- Wang, Z., Wu, Y., Xia, Z., et al., 2024. Non-invasive quantification of the brain [18 F] FDG-PET using inferred blood input function learned from total-body data with physical constraint. *IEEE Trans. Med. Imaging* 43 (7), 2563–2573.
- Whiteley, W., Gregor, J., 2019. CNN-based PET sinogram repair to mitigate defective block detectors. *Phys. Med. Biol.* 64 (23), 235017.
- Xu, Y., Li, Z., Zhang, X., et al., 2021. Low-dose PET image denoising based on coupled dictionary learning. *Nucl. Instrum. Methods Phys. Res. Sect. A* 1020, 165908.
- Xue, S., Guo, R., Bohn, K.P., et al., 2022. A cross-scanner and cross-tracer deep learning method for the recovery of standard-dose imaging quality from low-dose PET. *Eur. J. Nucl. Med. Mol. Imaging* 49, 1843–1856.
- Yang, B., Gong, K., Liu, H., et al., 2024. Anatomically guided PET image reconstruction using conditional weakly-supervised multi-task learning integrating self-attention. *IEEE Trans. Med. Imaging* 43 (6), 2098–2112.
- Yang, L., Zhang, Z., Song, Y., 2023. Diffusion models: A comprehensive survey of methods and applications. *ACM Comput. Surv.* 56 (105), 1–39.
- Zhang, C., Bengio, S., Hardt, M., et al., 2021. Understanding deep learning (still) requires rethinking generalization. *Commun. ACM* 64 (3), 107–115.
- Zhang, Y., Li, C., Zhong, L., et al., 2024. DoseDiff: distance-aware diffusion model for dose prediction in radiotherapy. *IEEE Trans. Med. Imaging*.
- Zhou, L., Lou, A., Khanna, S., et al., 2023. Denoising Diffusion Bridge Models *arXiv preprint arXiv: 2309.16948*.