

# The study of non-complete-ring positron emission tomography (PET) detection method

Yeqi Fang<sup>1,\*</sup> and Rong Zhou<sup>1,†</sup>

<sup>1</sup>College of Physics, Sichuan University, Chengdu, 610065, China

Positron Emission Tomography (PET) is a vital molecular imaging tool widely used in medical diagnosis and treatment evaluation. Traditional PET systems typically rely on complete detector rings to achieve full angular coverage for uniform and statistically robust sampling of coincidence events. However, incomplete-ring PET scanners have emerged in various scenarios due to hardware failures, cost constraints, or specific clinical needs. In such cases, conventional reconstruction algorithms often suffer from performance degradation due to reduced data completeness and geometric inconsistencies. This thesis proposes a coarse-to-fine reconstruction framework for incomplete-ring PET scanners. The framework first employs an Attention U-Net model to recover complete sinograms from incomplete ones, then uses the OSEM algorithm for preliminary reconstruction, and finally applies a two-stage architecture comprising a Coarse Prediction Module (CPM) and an Iterative Refinement Module (IRM) for fine reconstruction. Our approach utilizes neighboring axial slices and spectral transform features as auxiliary guidance at the input level to ensure spatial and frequency domain consistency, and integrates a contrastive diffusion strategy at the output level to improve correspondence between low-quality PET inputs and refined PET outputs. Experimental results on public and in-house brain PET datasets demonstrate that the proposed method significantly outperforms existing approaches in metrics such as PSNR (35.6421 dB) and SSIM (0.9588), successfully preserving key anatomical structures and tracer distribution features, thus providing an effective solution for incomplete-ring PET imaging.

## I. INTRODUCTION

### A. Background and Motivation

Positron Emission Tomography (PET) is one of the most valuable molecular imaging modalities in modern clinical practice. By injecting a radioactive tracer (typically bound to a biologically active molecule) into the patient, PET can non-invasively produce three-dimensional images reflecting physiological, metabolic, and functional activities. In PET imaging, when the injected radioactive tracer decays, it emits a positron that annihilates with an electron, producing a pair of gamma photons traveling in nearly opposite directions (separated by 180°). These annihilation photons are detected by a series of detectors surrounding the patient. As shown in Figure 1, a tracer labeled with a positron-emitting radionuclide (using 18F-labeled amino acid O-([18F]fluoroethyl)-tyrosine as an example, where fluorine-18, carbon, nitrogen, and oxygen atoms are shown as orange, gray, blue, or red spheres, with hydrogen atoms omitted for clarity) is injected into the subject. The subject is then placed inside a PET scanner, which consists of a ring of opposing detectors (represented by blue trapezoids). The biological distribution of the tracer in the body is then tracked by detecting the anti-parallel gamma photons[1]

Traditional PET scanners typically form a complete detector ring, allowing for complete angular sampling of coincidence events. This ensures that each line of re-

sponse (LOR) can be measured with consistent geometric coverage. However, the complete ring structure can be expensive and requires a large number of detectors. In certain clinical or research scenarios, incomplete rings are used due to cost constraints, hardware failures, or specialized system designs. While incomplete ring PET scanners may be less expensive or more flexible for specific applications, they introduce challenges in image reconstruction due to missing data at specific angles or axial segments. Furthermore, reduced detector coverage

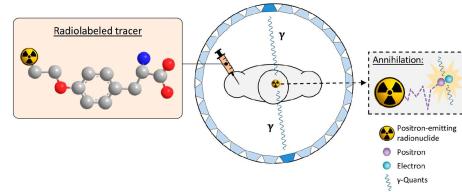


FIG. 1. Principles of Positron Emission Tomography (PET) imaging.

typically translates to lower signal-to-noise ratio (SNR) and more pronounced reconstruction artifacts such as streaks, blurring, or attenuation distortions. Traditional analytic or iterative methods (e.g., filtered backprojection, maximum likelihood expectation maximization) can handle missing data only to a limited extent; the geometry of incomplete rings leads to undersampling, thereby degrading the quality of the final image. Therefore, to make incomplete ring PET a viable and accurate imaging modality, advanced image reconstruction strategies must be developed.

In recent years, deep learning-based reconstruction methods have received widespread attention for their

\* fangyeqi@stu.scu.edu.cn

† zhourong@scu.edu.cn

ability to learn robust priors directly from data. Convolutional Neural Networks (CNNs), Generative Adversarial Networks (GANs), and more recently, Diffusion Probabilistic Models (DPMs) have achieved state-of-the-art results in many image-to-image translation tasks. Nevertheless, applying these methods to incomplete ring PET remains challenging due to the uniqueness of data acquisition and the complexity of system geometry.

This thesis explores and extends a novel reconstruction paradigm based on *diffusion probabilistic models* with a *coarse-to-fine* approach for PET reconstruction. Our method primarily contains two core objectives: on one hand, it is dedicated to improving the fidelity of reconstruction under incomplete data, accurately recovering high-quality PET images even when the scanner has large angular segment losses or defects, thus addressing the fundamental limitations of incomplete ring data; on the other hand, it focuses on reducing computational overhead, targeting the problem that traditional DPMs, while capable of producing state-of-the-art results, have large inference computational requirements due to their iterative reverse process. We overcome this burden through a customized coarse-to-fine design where a deterministic coarse prediction is refined by a relatively small diffusion model that only needs to reconstruct the residual signal.

This thesis makes several innovative contributions in the field of incomplete ring PET imaging, including proposing a coarse-to-fine reconstruction framework consisting of a Coarse Prediction Module (CPM) and an Iterative Refinement Module (IRM), which effectively reduces the burden on the diffusion-based IRM and shortens inference time by providing a deterministic initial reconstruction through the CPM; introducing an auxiliary guidance strategy that fuses adjacent axial slices and spectral features, providing spatial and frequency domain constraints to enhance image fidelity; innovatively integrating contrastive learning objectives into the diffusion sampling steps to enhance similarity between correct LPET-RPET pairs; proposing a simulation pipeline for generating complete ring and incomplete ring data from 3D brain volumes, and demonstrating robust performance under severely incomplete geometries; and finally, through extensive experiments on public 3D brain PET datasets and in-house datasets, demonstrating the method's superiority over existing methods in terms of reconstruction accuracy, robustness, and clinical reliability.

## B. Thesis Organization

The remainder of this thesis is organized as follows:

**Section II** (*Background on PET Imaging and Incomplete Ring Geometry*) provides an overview of PET physics, principles of coincidence detection, and the main challenges introduced by incomplete rings. We also summarize classical and modern reconstruction methods.

**Section III** (*Fundamentals of Diffusion Probabilistic*

*Models*) details the background of Denoising Diffusion Probabilistic Models (DDPM) and how they apply to various image reconstruction tasks, establishing the theoretical foundation for our proposed method. *Proposed Coarse-to-Fine Reconstruction Framework*) describes the detailed workflow of our method, including complete and incomplete ring geometric modeling, creation of list-mode and sinogram data, the coarse-to-fine design, auxiliary guidance modules, contrastive diffusion learning objectives, and implementation details.

**Section IV** (*Experiments and Results*) presents our extensive experimental setups, hyperparameter choices, evaluation metrics, ablation studies, and comparisons with state-of-the-art methods. We also present results of cross-dataset evaluation on in-house datasets.

**Section V** (*Conclusion and Future Work*) summarizes our findings, discusses limitations, and suggests directions for future research, including potential improvements for real-world incomplete ring PET scanners.

**Appendices** provide additional experiments, extended qualitative results, and tables of hyperparameters used in the thesis.

## II. BACKGROUND ON PET IMAGING AND INCOMPLETE RING GEOMETRY

This section provides a brief but comprehensive overview of the fundamental aspects of PET imaging, focusing on the detection process, data acquisition methods, and conventional reconstruction algorithms. We then discuss incomplete ring PET scanners in detail, highlighting the geometry of missing detectors, their impact on data quality, and current solutions and unresolved research challenges.

### A. Principles of PET Imaging

In PET imaging, positron-emitting nuclides (typically  $^{18}\text{F}$ ,  $^{15}\text{O}$ ,  $^{11}\text{C}$ , etc.) are usually combined with biologically relevant tracer compounds. After entering the bloodstream, these nuclides decay, with each emitting a positron. When the positron encounters an electron, an annihilation event occurs, producing two gamma photons, each with an energy of approximately 511 keV, traveling in nearly opposite directions (collinearly).

PET scanners are designed to detect coincidences of these 511 keV photon pairs. When two detectors at different positions record gamma photons within a short coincidence time window, it is assumed that these photons came from the same annihilation event. This forms a line of response (LOR) between the two detector elements, as shown in Figure 2. PET data can be acquired in multiple formats: (1) **List-mode data**: Each coincidence event is recorded individually, providing precise detector pair identification and timestamps. (2) **Sinogram data**: By

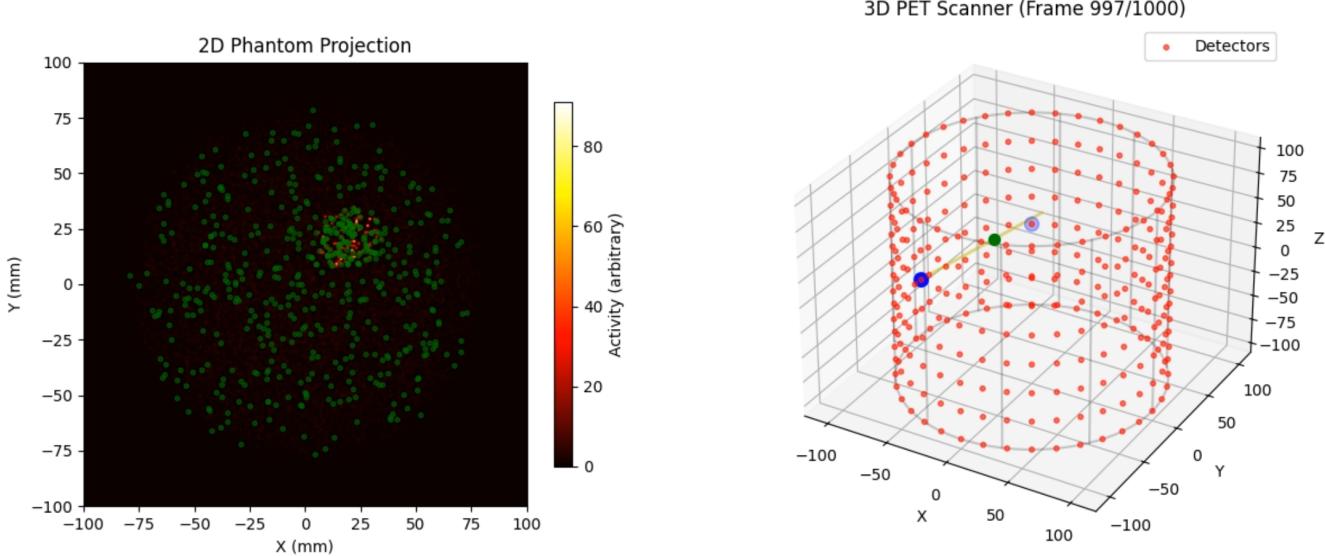


FIG. 2. Schematic diagram of PET detection, where the red dot represents the detector center, the green dot represents the annihilation event location, and the two blue dots represent the centers of the two detectors that detect the gamma rays. This figure is only illustrative; the detector parameters in the figure do not equal the actual simulation parameters.

categorizing coincidence counts into projection bins indexed by radial, angular, and axial coordinates (possibly also by axial ring difference), forming a 2D (or 3D) histogram. In a typical 3D PET system, a scanner with a complete ring covers a  $360^\circ$  view angle around the patient, ensuring uniform sampling of LORs.

### B. Incomplete Ring PET Scanners

In normal PET (Positron Emission Tomography) system design, *incomplete rings* or *partial rings* often arise from various practical considerations and technical constraints. From an economic perspective, reducing the total number of detectors can effectively lower system costs. In certain specialized fields, such as brain imaging or organ-specific examinations, researchers may adopt partially covered designs to accommodate smaller field-of-view requirements. Additionally, with the continuous development of medical technology, innovative designs for hybrid or portable PET are increasingly important; these systems meet the needs of different clinical and research scenarios through lightweight, flexible structures. It is worth noting that technical factors such as detector failures, maintenance, upkeep, calibration, and other issues may also cause certain ring segments to temporarily stop operating, resulting in incomplete detector rings. These variations reflect the complexity and adaptability of PET imaging technology in practical applications.

The main difficulty with incomplete rings is the reduction in angular sampling, as missing detectors lead to missing corresponding LORs, making the dataset incomplete, which breaks the fundamental assumptions of many classical reconstruction algorithms that rely on fully sampled complete angular projections. This data deficiency leads to a series of issues: producing streak ar-

tifacts and increased noise in the image domain, causing quantitative inaccuracies in tracer uptake (especially in regions that depend on missing LORs), and potentially introducing biases in clinical indices such as standardized uptake values (SUVs), thereby affecting diagnostic accuracy, as shown in Figure 11.

### C. Conventional PET Reconstruction Methods

**Analytical algorithms**, such as *filtered backprojection* (FBP), are widely used for their computational efficiency in fully sampled data. However, FBP is sensitive to noise and incomplete sampling, and prone to severe streak artifacts when the system geometry is incomplete.

**Maximum likelihood estimation methods**, such as maximum likelihood expectation maximization (MLEM) and its variant ordered subset expectation maximization (OSEM)[2], incorporate Poisson statistics of PET photon counting. They typically produce better results than analytical methods, even with partial data. The core idea of OSEM is to divide all projection data  $\{y_i\}$  into  $S$  subsets and use only one subset for each update, thus reducing the computational load per update.

A core concept in PET reconstruction is the **system model**, which describes the relationship between unknown image voxels and measured projection data. Formally:  $\lambda_j$  is the activity value in voxel  $j$ ;  $y_i$  is the count measured in detector pair (or projection)  $i$ ;  $p_{ij}$  is the probability (or "weight") that a photon pair emitted from voxel  $j$  is detected by projection  $i$ . The mathematical model can be represented as:

$$y_i \approx \sum_j p_{ij} \lambda_j, \quad (1)$$

The matrix  $p_{ij}$  is often called the **system matrix** or "projection matrix." Each element  $p_{ij}$  depends on factors such as geometric solid angle, detector efficiency, and normalization corrections. In program implementation, this matrix can be explicitly stored or calculated on-the-fly through projector/backprojector operations.

Image reconstruction  $\lambda_j$  typically uses iterative algorithms such as **Maximum Likelihood Expectation Maximization (MLEM)**. The update formula for voxel  $j$  at the  $k$ -th iteration of the MLEM algorithm is:

$$\lambda_j^{(k+1)} = \lambda_j^{(k)} \times \frac{\sum_{i=1}^N \frac{p_{ij}}{\sum_{\ell} p_{i\ell} \lambda_{\ell}^{(k)}} y_i}{\sum_{i=1}^N p_{ij}}, \quad (2)$$

where  $N$  is the total number of projection elements.

However, the standard MLEM algorithm converges slowly when processing large datasets. The Ordered Sub-set Expectation Maximization (OSEM) algorithm accelerates convergence by dividing the projection data into  $S$  subsets  $S_1, \dots, S_S$ . In OSEM implementation, the measurement data  $\{y_i\}_{i=1}^N$  is divided into  $S$  subsets, each typically containing  $N/S$  projection elements. Each update uses only one subset to update  $\lambda_j$ . A "complete OSEM iteration" is completed after cycling through all subsets. Let  $S_k$  represent the subset of projection indices used in the  $k$ -th iteration, then the OSEM update formula can be expressed as:

$$\lambda_j^{(k+1)} = \lambda_j^{(k)} \times \frac{\sum_{i \in S_k} \frac{p_{ij}}{\sum_{\ell} p_{i\ell} \lambda_{\ell}^{(k)}} y_i}{\sum_{i \in S_k} p_{ij}} \quad (3)$$

By sequentially cycling through these subsets (e.g., using  $S_1$  for the 1st iteration,  $S_2$  for the 2nd, ...,  $S_S$  for the  $S$ -th, then returning to  $S_1$ , and so on), the OSEM algorithm can approximate a complete MLEM update with fewer effective iterations, thus obtaining an image close to convergence in less time.

In modern PET software frameworks, the system matrix  $p_{ij}$  is typically not stored as a large two-dimensional array, but is implicitly calculated through **projectors** and **backprojectors**. For each voxel  $j$ , the projector calculates its contribution to the measurement data  $y_i$ ; the backprojector assigns the ratio of measured to expected data back to the voxel grid. This approach effectively reduces memory requirements when processing large volumes of data.

The algorithms described in this paper are based on the PyTomography library[3]. PyTomography is an open-source Python library for medical image reconstruction, providing a modular framework for system matrix construction, likelihood function computation, and reconstruction algorithm development. The library leverages PyTorch's GPU acceleration capabilities and the parallel-proj library for efficient parallel computation. Its flexible

modular design makes it easy to decouple system matrices, likelihood functions, and reconstruction algorithms, facilitating integration with various Python toolsets for new imaging modalities. Currently, the library has been successfully applied in fields such as parallel-hole SPECT and list-mode PET imaging, providing a highly optimized and user-friendly software platform for medical image reconstruction. We tested it using a brain PET dataset, and the comparison between the original image and the reconstructed image is shown in Figure 3.

#### D. Deep Learning for PET Reconstruction

Currently, many studies have attempted to use various deep learning methods to improve the quality and speed of PET reconstruction.

**Regression-based networks** have shown significant effectiveness in tasks such as low-dose PET denoising and partial data reconstruction[4], with the core idea being to learn a direct mapping from degraded images to high-quality images. Although convolutional neural networks (CNNs) can utilize local receptive fields to remove noise or supplement missing information, they often struggle to effectively capture global structural features when dealing with large-scale angular losses.

**GAN-based methods** introduce generative adversarial networks into the field of PET reconstruction, generating visually more reasonable images by inferring missing structures[5]. However, the instability of the GAN training process is a common problem, often resulting in mode collapse or unexpected artifacts. More importantly, while GAN-generated images may be visually satisfying, they may not guarantee the accurate quantitative fidelity required for clinical diagnosis.

**Likelihood-based generative models** such as variational autoencoders (VAEs) and normalizing flows attempt to model data distributions within a more rigorous theoretical framework. While these methods alleviate the stability issues of GANs to some extent, they often produce relatively blurry reconstruction results and still have high computational complexity.

Therefore, due to severe angular data loss and the inherent noise limitations of PET imaging systems, incomplete ring PET reconstruction still faces many challenges. Although deep learning methods have opened new avenues for solving this problem, existing models still have obvious deficiencies: they either cannot effectively handle large-scale data loss, or while providing acceptable solutions, they have excessive computational overhead (such as simple diffusion-based methods), or they struggle to maintain the detailed features required for clinical diagnosis when facing large-scale geometric defects. In response to these issues, this thesis proposes an innovative coarse-to-fine diffusion model framework that is not only computationally efficient and stable in performance but also deeply optimized specifically for the characteristics of incomplete ring PET scanners.

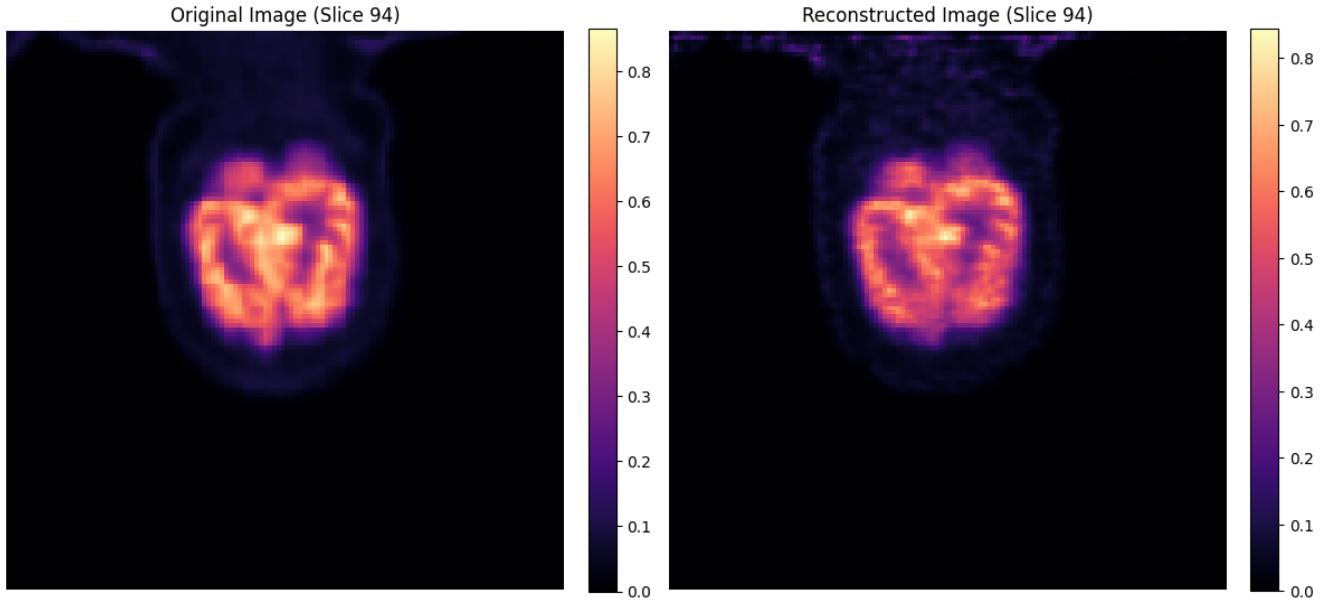


FIG. 3. Comparison of original image and image reconstructed using the OSEM method

### III. IMPLEMENTATION METHODS

As shown in Figure 4, this study proposes an innovative incomplete ring PET reconstruction framework that effectively addresses the data incompleteness problem caused by missing detectors through a multi-stage strategy. In the first stage, the system first processes incomplete sinograms (top left) resulting from missing detectors, completing the missing data through a trained optimized U-Net deep learning model to generate complete sinograms (top right). This sinogram restoration process fully utilizes the five-channel input strategy described in Chapter 3, effectively integrating spatial and temporal context information.

#### A. System Model and Data Simulation

Assume a standard PET ring geometry with  $R$  axial rings, each containing  $D$  detectors, with each detector assigned a unique identification number used to precisely locate the photon pair's detection position when recording coincidence events. For example, consider a 42-ring system (i.e.,  $R = 42$ ), with 182 detectors per ring, arranged in a cylindrical structure with radius  $\rho$ . We map the 3D image grid (e.g., size =  $128 \times 128 \times 128$  voxels) to physical space by defining voxel dimensions, approximating the entire scanner geometry realistically. For instance, if the field of view diameter is 25 centimeters and the in-plane matrix size is 128, then each in-plane voxel size is:

$$\Delta x \approx \Delta y \approx \frac{25 \text{ cm}}{128} \approx 1.95 \text{ mm.} \quad (4)$$

Axially, we can set  $\Delta z$  based on the typical axial coverage of a PET scanner. This high-precision spatial division

ensures that the reconstructed images accurately reflect the anatomical structure of the scanned area.

The PET scanner used in this study employs a highly optimized detector system configuration, achieving comprehensive three-dimensional detection capabilities. The scanner's radius ( $\rho$ ) is precisely set at 380.56 millimeters, achieving an optimal balance between spatial resolution and system sensitivity. The detector system adopts a multi-level organizational structure: at the microscopic level, crystal elements have lateral dimensions of 4.03125 millimeters and axial dimensions of 5.31556 millimeters, with this precise crystal size design ensuring high spatial resolution capability; at the macroscopic level, the detector construction includes 34 lateral sectors (rsectorTransNr) and a single axial sector arrangement (rsectorAxialNr). Each module is carefully designed to include 16 lateral (crystalTransNr) and 9 axial (crystalAxialNr) crystal elements, with the spacing between adjacent crystals (crystalTransSpacing and crystalAxialSpacing) optimized to achieve maximum detection efficiency. In the axial direction, the system has 4 modules (moduleAxialNr), maintaining a precise spacing of 50.64004 millimeters between modules (moduleAxialSpacing), a configuration that ensures detection sensitivity while maintaining an appropriate axial field of view range. The system's data recording uses a list-mode format to record various coincidence events during the scanning process. The system supports time-of-flight (TOF) detection functionality, with 29 time bins (num\_tof\_bins), a full width at half maximum of the time resolution window (tof\_fwhm) of 57.71 millimeters, and a measurement range (tof\_range) reaching 735.7705 millimeters. However, due to current device performance limitations, this study could not fully utilize TOF technology to improve spatial resolution. Each event entry consists of three basic components, stored in tensor format. A typical list-

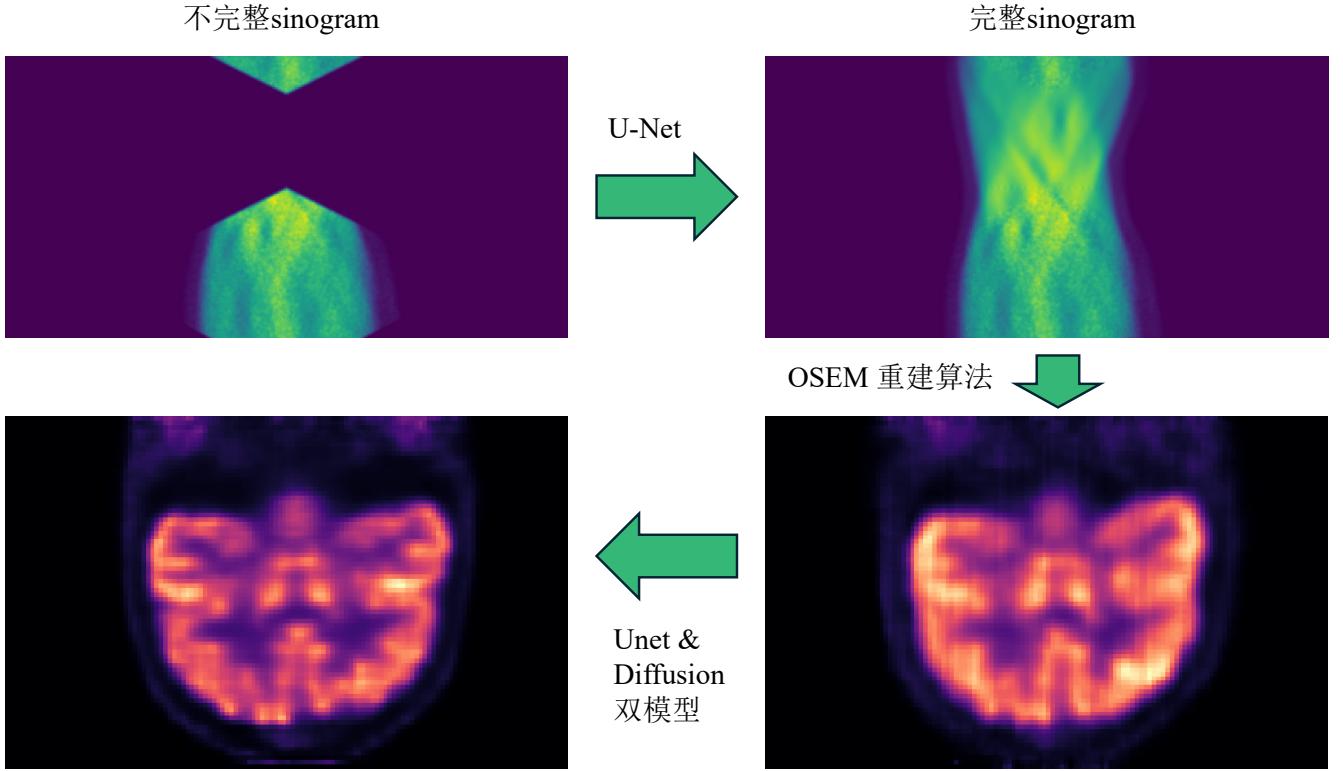


FIG. 4. Overall flow diagram of the incomplete ring PET reconstruction method. First, incomplete sinograms due to missing detectors (top left) are repaired through a U-Net deep learning model to generate complete sinogram data (top right). Subsequently, the complete sinogram is converted into a PET image (bottom right) using the OSEM iterative reconstruction algorithm. For comparison, the bottom left image shows a PET image reconstructed directly from incomplete data using a coarse-to-fine framework (dual-model architecture with U-Net and diffusion model). This method effectively compensates for data loss caused by incomplete ring geometry, achieving high-quality brain PET image reconstruction that preserves key anatomical structures and tracer distribution features.

mode event data structure is as follows:

```
[898, 3334, 11]
[816, 16375, 18]
[1073, 18816, 14]
[1529, 13781, 13]
[1537, 18663, 14]
```

In this format, the first two numbers represent the identification numbers of the detector pair recording the coincidence event. The third number represents the time-of-flight information of the event, which, although not fully utilized in this study to enhance image quality, is retained for future data processing and analysis after system upgrades. To improve the system's sensitivity and accuracy, the detector system also includes submodule structures (submoduleAxialNr and submoduleTransNr both equal to 1), but in the current configuration, the submodule spacing (submoduleAxialSpacing and submoduleTransSpacing) is set to 0 to simplify system complexity.

In the current implementation process, due to the lack of MRI or CT data to obtain gamma ray absorption distribution maps, the system cannot use  $\mu$  maps (linear attenuation maps) or component-based normalization

files. This limitation means that the reconstruction process can only use list-mode data as input and cannot perform attenuation correction. When interpreting reconstruction results, this limitation's impact on image quality and quantitative accuracy must be considered.

We refer to the fully functional scenario where all  $R$  rings and all  $D$  detectors are working as the *complete ring* configuration, as shown in Figure 6. The data simulation process is as follows: **Starting with real 3D images.** We have high-resolution brain PET data or a synthetic phantom  $\mathbf{X} \in \mathbb{R}^{128 \times 128 \times 128}$ .

**Randomly generating emission events.** We sample a specified number of emission events (e.g., 20 million) from the probability density function implied by  $\mathbf{X}$ . Each voxel's value is interpreted as proportional to its tracer concentration probability.

**Simulating annihilation.** For each emission event, we assume it annihilates with an electron in the same voxel or nearby, producing two photons at  $180^\circ$  opposite directions, each carrying 511 keV.

**Calculating detector hits.** Using geometric ray tracing, we calculate which detector pairs would detect these photons. This determines the LOR associated with each event, effectively generating a *list-mode* dataset.

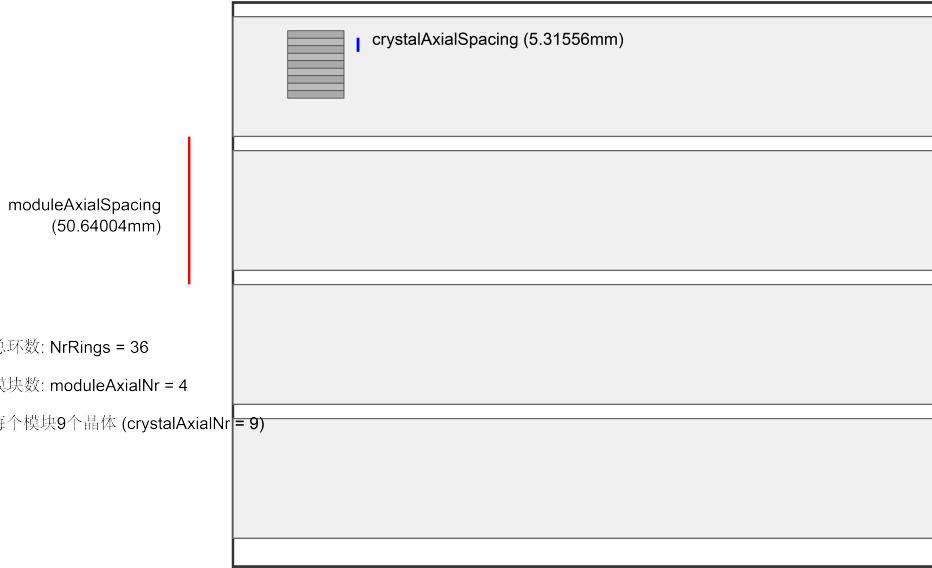


FIG. 5. Side view schematic of PET scanner detector structure.

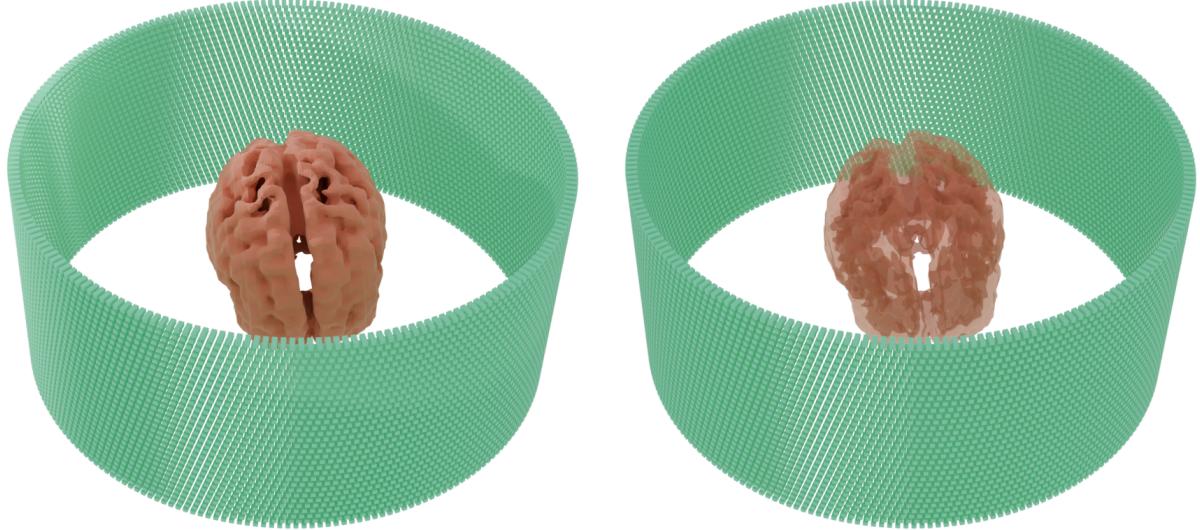


FIG. 6. Three-dimensional schematic of PET scanner detector structure, with the right image showing a perspective view.

**Generating sinograms.** We can optionally sort the list-mode data into sinogram bins  $\mathbf{S}$  by radial, angular, and axial coordinates. This generates the standard 3D sinogram representation for PET.

Then, we use standard iterative reconstruction (e.g., OSEM) to reconstruct the baseline reconstruction volume  $\mathbf{Y}_A$  from  $\mathbf{S}$  (or directly from list-mode). This reconstruction from complete data is typically considered higher quality. In the *incomplete ring* scenario, we remove parts of the detectors from the scanning geometry. The simplest approach is to remove all detectors from one or more complete rings, or remove angular segments from each ring. For example, suppose we remove one complete ring, so we're effectively only collecting data from  $R - 1$  rings. The data simulation process is the same as the complete case, with the only difference being that events falling on the missing ring would not be

recorded, as shown in Figure 7.

Due to missing LORs, the incomplete data reconstructed image  $\mathbf{Y}_B$  typically has lower quality. In practice, we can reconstruct from the incomplete sinogram  $\mathbf{S}_{\text{incomplete}}$ , i.e.,  $\mathbf{Y}_B = \text{Reconstruct}(\mathbf{S}_{\text{incomplete}})$ . This  $\mathbf{Y}_B$  can serve as input to our network, while  $\mathbf{Y}_A$  from complete data (same subject) can serve as the training target. The reconstruction from complete data  $\mathbf{Y}_A$  effectively represents the *ground truth* of the tracer distribution for the same subject under ideal scanning conditions. In low-dose or incomplete coverage scenarios, the difference between  $\mathbf{Y}_B$  and  $\mathbf{Y}_A$  can be quite significant.

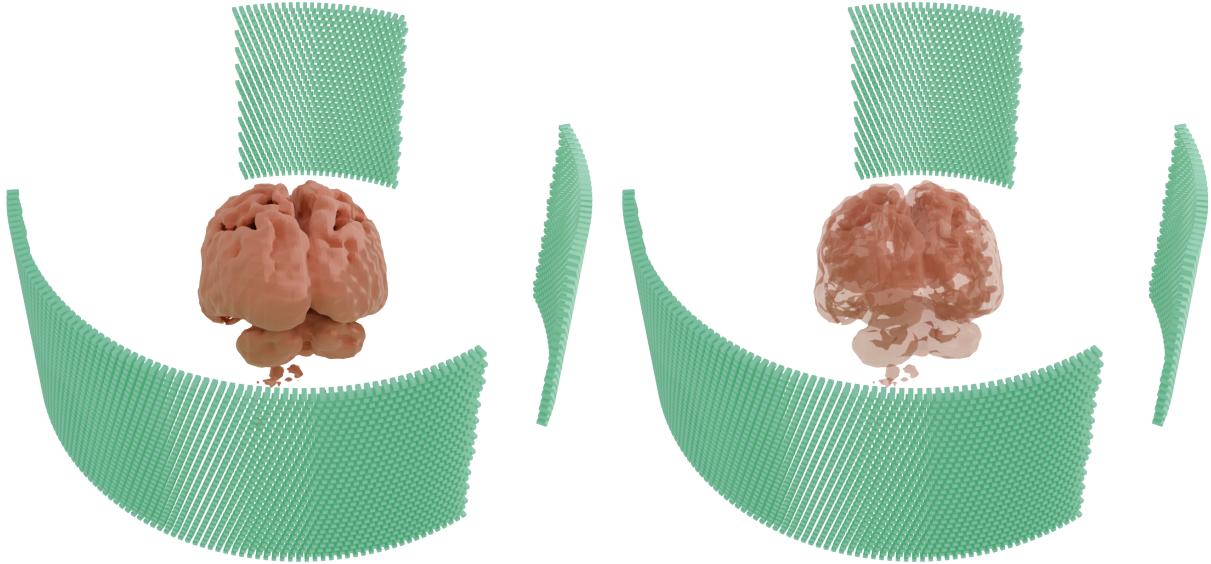


FIG. 7. Three-dimensional schematic of incomplete ring PET scanner detector structure, with the right image showing a perspective view.

### B. Sinogram Reconstruction Based on Attention U-Net

Although traditional U-Net performs excellently in medical image segmentation and reconstruction tasks, it lacks the ability to selectively focus on key features when dealing with complex incomplete ring PET sinogram restoration problems. The Attention U-Net adopted in this study enhances the model's perception of important feature regions through spatial attention mechanisms while suppressing the influence of irrelevant features, which is crucial for restoring sinograms from incomplete data.

The core innovation of Attention U-Net lies in the introduction of attention gating (AG) modules in the skip connection path of the standard U-Net, as shown in Figure 3.7. These AG modules can adaptively highlight significant structures in the feed-forward feature maps while suppressing irrelevant regions that might lead to prediction errors. For sinogram reconstruction tasks, this mechanism is particularly effective because it can selectively focus on structural features around missing areas, thereby more accurately inferring missing angular data. The mathematical expression of attention gating can be formalized as:

$$\alpha_i^l = \sigma_2(\psi^T(\sigma_1(W_x^T x_i^l + W_g^T g_i + b_g)) + b_\psi) \quad (5)$$

where  $x_i^l$  is the low-level feature from the encoder,  $g_i$  is the gating signal from the decoder (high-level feature),  $\sigma_1$  and  $\sigma_2$  are ReLU and Sigmoid activation functions respectively, and  $W_x$ ,  $W_g$ ,  $b_g$ , and  $b_\psi$  are learnable parameters.  $\alpha_i^l \in [0, 1]$  is the calculated attention coefficient used to control the importance of features.

After processing through the attention gate, the features can be represented as:

$$\hat{x}_i^l = x_i^l \cdot \alpha_i^l \quad (6)$$

This mechanism allows Attention U-Net to adaptively focus on important image regions while maintaining the advantages of the original U-Net's encoder-decoder structure, particularly in cases of severe angular loss, with a stronger perception of sinogram continuity and consistency. In the experimental implementation, we integrated AG modules into the skip connection path at each decoding stage, using  $1 \times 1$  convolution to reduce channel dimensions, followed by applying attention coefficients for feature selection. This design enables the model to more precisely restore missing data when processing incomplete sinograms containing large-scale angular losses, while maintaining overall structural consistency and accurate signal distribution.

Experimental results show that compared to standard U-Net, Attention U-Net achieved significant improvements in sinogram reconstruction tasks, with an average increase of 1.24dB in PSNR metrics and 0.052 in SSIM metrics, validating the effectiveness of the attention mechanism in processing incomplete ring PET data.

Sinogram repair plays a crucial role in incomplete ring PET imaging, with its quality directly affecting the precision of reconstructed images. The training dataset used in this study consists of multidimensional tensors that effectively capture spatial and temporal context information, essential for accurate sinogram restoration. Each sinogram slice is expanded into a five-channel tensor, systematically integrating slices from spatial and temporal neighborhoods. As shown in Figure 9, the central channel of each input tensor corresponds to the current sinogram slice, while the direct spatial neighbors (slices  $j - 1$

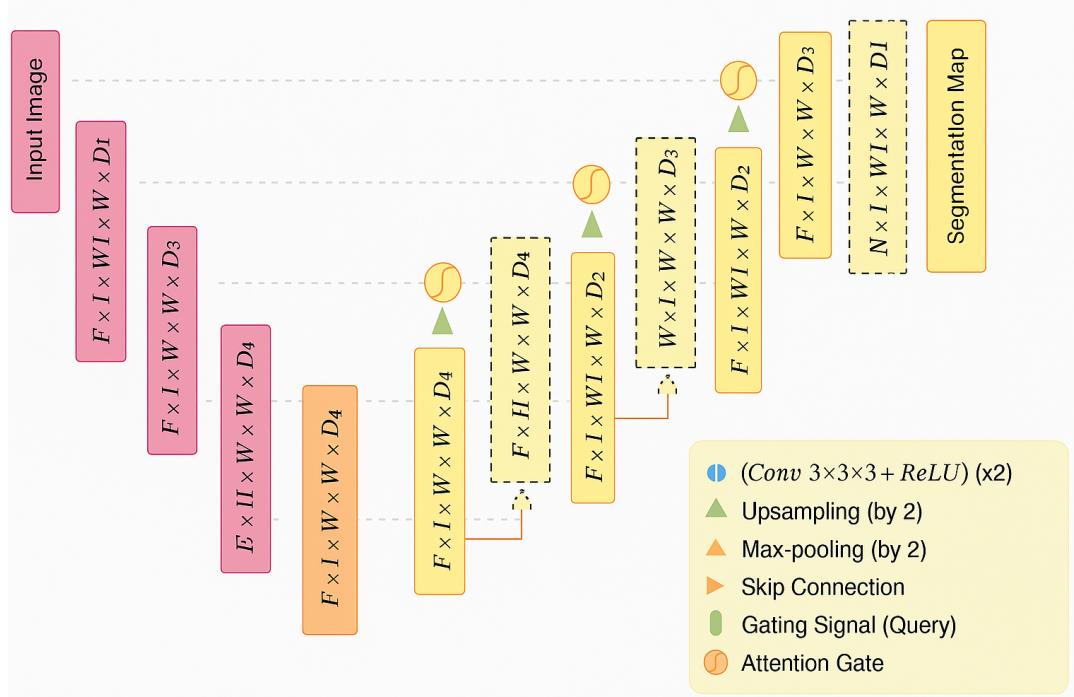


FIG. 8. Attention U-Net architecture diagram, showing how the attention gating mechanism enhances skip connection features.

and  $j+1$ ) and temporal neighbors from previous and subsequent sinogram periods (slices  $j-N_2$  and  $j+N_2$ ) constitute the other four channels. These neighboring slices are crucial for enriching the model's understanding of local structural continuity and temporal consistency.

In this study, a five-channel input tensor is used to construct training data. For each central slice  $j$ , a multi-dimensional feature is built by fusing its spatial adjacent slices ( $j-1$  and  $j+1$ ) with temporal adjacent slices ( $j-42$  and  $j+42$ ). Boundary handling adopts a mirror padding strategy: when adjacent indices exceed the dataset range, the central slice itself is used for channel filling, ensuring input dimension consistency. Finally, each training sample forms a  $5 \times H \times W$  tensor, effectively integrating spatiotemporal neighborhood information.

The network architecture adopts an improved U-Net model (see Table??), with an encoder-decoder structure including convolution, pooling, upsampling, and skip connection modules. The encoder part uses  $3 \times 3$  convolution kernels with batch normalization and ReLU activation functions, implementing feature dimensionality reduction through max pooling; the decoder uses transposed convolution/bilinear interpolation for resolution recovery, and shallow detail features are fused through skip connections. The bottleneck layer achieves high-level semantic representation through 1024-dimensional features.

Model training uses the Adam optimizer with an initial learning rate of  $10^{-4}$  and introduces  $10^{-5}$  weight decay to prevent overfitting. Training efficiency is enhanced through mixed precision computation and gradient scaling techniques, combined with a ReduceLROn-

Plateau dynamic learning rate scheduler (with a decay factor of 0.3 when validation loss shows no improvement for 3 consecutive epochs), achieving stable convergence.

On the validation set, the model achieved excellent performance with MSE=0.301745, PSNR=35.6421 dB, and SSIM=0.9588. The PSNR value indicates high fidelity of the reconstructed signal, while the SSIM metric validates the preservation of structural features. Experimental results show that the multi-channel U-Net method proposed in this paper can effectively improve sinogram repair quality, providing a reliable foundation for subsequent CT image reconstruction. This method significantly improves the limitations of traditional single-channel models by fusing spatiotemporal context information, offering a new technical approach for the medical imaging processing field.

### C. Diffusion Probabilistic Models

Diffusion probabilistic models (DPMs) have attracted great attention in image synthesis and inverse problems for their ability to produce state-of-the-art results and training stability. Unlike GANs that rely on adversarial objectives, DPMs are trained by maximizing a variational lower bound on the data log-likelihood. This chapter covers the core principles and mathematical formulations of denoising diffusion probabilistic models relevant to the proposed method. Diffusion models are a class of probabilistic generative models proposed by *Sohl-Dickstein et al.*[6], and later widely applied in image generation tasks by *Ho et al.*[7]. The basic idea is to gradually

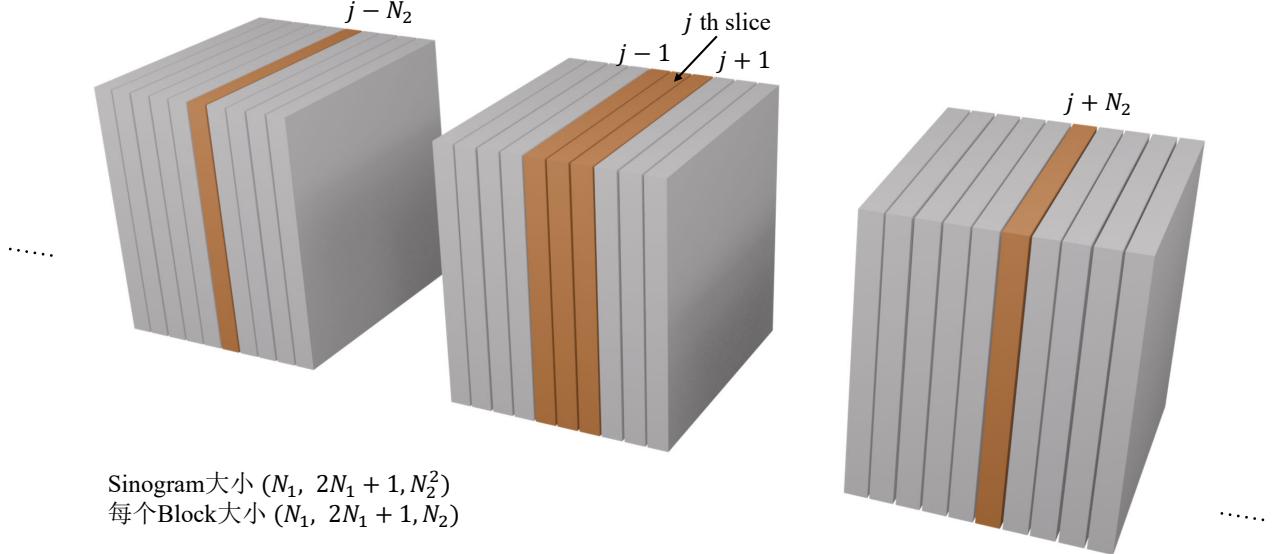


FIG. 9. Visualization of the five-channel tensor preparation of sinogram data for training. Each cube represents a sinogram slice block with dimensions  $(N_1, 2N_1 + 1, N_2^2)$ . The orange highlighted parts are the selected slices  $(j - N_2, j - 1, j, j + 1, j + N_2)$ , which form the five-channel input for the restoration model, showing the spatial and temporal relationships captured in each input tensor.

inject noise into the data distribution, transforming it into an isotropic Gaussian distribution. In the generation stage, the target data distribution is gradually restored from pure noise through a learned reverse denoising process. This framework can be divided into two main stages: **Forward Process**: Gradually adding noise to the data, making the data distribution approach a Gaussian distribution; **Reverse Process**: Starting from Gaussian noise, gradually denoising to generate samples consistent with the original data. Diffusion models have achieved significant results in image, speech, text, and other fields, with their stability and generation quality attracting widespread attention.

### 1. Forward Process

The forward process gradually adds noise to the data  $\mathbf{x}_0 \in \mathbb{R}^d$  through a Markov chain, mathematically expressed as:

$$q(\mathbf{x}_{1:T} | \mathbf{x}_0) = \prod_{t=1}^T q(\mathbf{x}_t | \mathbf{x}_{t-1}), \quad (7)$$

where each step's conditional probability is defined as:

$$q(\mathbf{x}_t | \mathbf{x}_{t-1}) = \mathcal{N}(\mathbf{x}_t; \sqrt{1 - \beta_t} \mathbf{x}_{t-1}, \beta_t \mathbf{I}). \quad (8)$$

Here,  $\beta_t \in (0, 1)$  represents the noise intensity, a predetermined hyperparameter controlling the amount of noise added in each step. Since the noise added at each step is a linear combination of Gaussian distributions,  $\mathbf{x}_t$  can be directly calculated from  $\mathbf{x}_0$ . Using the properties of

Gaussian distributions, we have:

$$q(\mathbf{x}_t | \mathbf{x}_0) = \mathcal{N}(\mathbf{x}_t; \sqrt{\bar{\alpha}_t} \mathbf{x}_0, (1 - \bar{\alpha}_t) \mathbf{I}), \quad (9)$$

where the cumulative coefficient  $\bar{\alpha}_t$  is defined as:

$$\bar{\alpha}_t = \prod_{s=1}^t \alpha_s, \quad \alpha_t = 1 - \beta_t \quad (10)$$

This closed-form solution allows us to directly sample  $\mathbf{x}_t$  from  $\mathbf{x}_0$ , greatly simplifying the derivation process during training.

### 2. Reverse Process

The reverse process aims to rebuild data starting from pure Gaussian noise through gradual denoising. Assuming the Markovian property of the forward process holds, the reverse process can also be represented by a Markov chain:

$$p_\theta(\mathbf{x}_{0:T}) = p(\mathbf{x}_T) \prod_{t=1}^T p_\theta(\mathbf{x}_{t-1} | \mathbf{x}_t), \quad (11)$$

where  $p(\mathbf{x}_T)$  is a standard Gaussian distribution, and  $p_\theta(\mathbf{x}_{t-1} | \mathbf{x}_t)$  represents the reverse conditional distribution. We typically assume it takes the form of a Gaussian distribution:

$$p_\theta(\mathbf{x}_{t-1} | \mathbf{x}_t) = \mathcal{N}(\mathbf{x}_{t-1}; \boldsymbol{\mu}_\theta(\mathbf{x}_t, t), \sigma_t^2 \mathbf{I}). \quad (12)$$

Here, the mean  $\boldsymbol{\mu}_\theta$  and variance  $\sigma_t^2$  are modeled by neural networks. To simplify implementation,  $\sigma_t^2$  is often fixed as a constant.

### 3. Model Training

In practice, directly predicting  $\mathbf{x}_{t-1}$  can be overly complex. We can simplify this problem by predicting the noise  $\epsilon$  in the forward process. According to the closed-form solution of the forward process:

$$\mathbf{x}_t = \sqrt{\alpha_t} \mathbf{x}_0 + \sqrt{1 - \bar{\alpha}_t} \epsilon, \quad \epsilon \sim \mathcal{N}(\mathbf{0}, \mathbf{I}), \quad (13)$$

the model's task can be transformed into learning to predict the noise  $\epsilon$ , i.e., approximating the true noise through a neural network  $\epsilon_\theta(\mathbf{x}_t, t)$ . For this purpose, the training objective is defined as minimizing the following mean squared error (MSE) loss:

$$L_{\text{simple}}(\theta) = \mathbb{E}_{\mathbf{x}_0, \epsilon, t} [\|\epsilon - \epsilon_\theta(\mathbf{x}_t, t)\|^2] \quad (14)$$

This approach is not only simple but has also shown good performance in generation tasks.

The model's training objective is based on the variational lower bound (VLB), optimizing the generation process by maximizing the data likelihood  $\log p_\theta(\mathbf{x}_0)$ . Its bound form is:

$$-\log p_\theta(\mathbf{x}_0) \leq \text{KL}(q(\mathbf{x}_T | \mathbf{x}_0) \| p(\mathbf{x}_T)) + \sum_{t=2}^T \text{KL}(q(\mathbf{x}_{t-1} | \mathbf{x}_t, \mathbf{x}_0) \| p_\theta(\mathbf{x}_{t-1} | \mathbf{x}_t)) - \log p_\theta(\mathbf{x}_0 | \mathbf{x}_1) \quad (15)$$

In practice, directly optimizing this objective is relatively complex, and the simplified noise prediction loss  $L_{\text{simple}}(\theta)$  is typically used for optimization, with the core being to learn the denoising process.

### 4. Sampling

In the generation stage, starting from Gaussian noise  $\mathbf{x}_T \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$ , samples are gradually generated through the following formula:

$$\mathbf{x}_{t-1} = \frac{1}{\sqrt{\alpha_t}} \left( \mathbf{x}_t - \frac{1 - \alpha_t}{\sqrt{1 - \bar{\alpha}_t}} \epsilon_\theta(\mathbf{x}_t, t) \right) + \sigma_t \mathbf{z}, \quad \mathbf{z} \sim \mathcal{N}(\mathbf{0}, \mathbf{I}) \quad (16)$$

Here,  $\sigma_t$  represents the noise intensity at each step. Through multiple iterations, the finally generated  $\mathbf{x}_0$  should approach the target data distribution.

Diffusion models, as a generative modeling method combining the advantages of probabilistic graphical models and deep learning, achieve efficient generation through forward noise addition and reverse denoising processes, showing broad application prospects in fields such as image generation, speech synthesis, and text generation. In the inverse problem of PET reconstruction under incomplete ring conditions, the advantages of diffusion models are particularly prominent: they not only effectively avoid the mode collapse problem common in GANs, ensuring robust coverage of possible solutions, but can also effectively integrate learned priors for PET images, which is extremely important when large segments of sinogram data are missing; meanwhile, their iterative refinement

characteristic highly aligns with the iterative image reconstruction methods (such as MLEM) widely used in the PET field. However, direct application of DPMs for reconstruction has issues of slow computation speed and insufficient consideration of partial data scale. To address these limitations, we propose an innovative coarse-to-fine method, effectively overcoming these challenges by introducing a deterministic, high-capacity coarse prediction module combined with a small-scale iterative diffusion model focused on residual reconstruction.

## D. Coarse-to-Fine Reconstruction Framework

In this section, we outline the main workflow of the coarse-to-fine approach for reconstructing PET data from incomplete ring scanners. First, we introduce the system modeling description (i.e., how to simulate or process incomplete rings). Then, we detail the data generation steps for complete and incomplete rings, ultimately building our *CPM + IRM* reconstruction architecture. Finally, we discuss the *auxiliary guidance* and *contrastive diffusion* strategies.

### 1. Overall Coarse-to-Fine Framework

Applying diffusion probabilistic models directly to incomplete ring reconstruction typically faces the problem of high computational cost, as each inference requires multiple iterative steps. Additionally, for large 3D volumes ( $128^3$  or higher), pure diffusion methods may be challenging in terms of memory. Therefore, we adopt a *coarse-to-fine* approach, where the **Coarse Prediction Module (CPM)** outputs an initial deterministic reconstruction from partial or degraded input, as shown in Figure 10. The **Iterative Refinement Module (IRM)** implements a conditional diffusion process to estimate the residual  $\mathbf{r}_0$  between the coarse prediction and the ground truth.

### 2. Coarse Prediction Module (CPM)

Let  $\mathbf{c}$  be our condition, including  $\mathbf{Y}_B$ , the reconstructed PET image from incomplete data;  $\mathbf{X}_{\text{aux}}$ , additional auxiliary guidance, such as adjacent slices or spectral transforms.

The CPM, denoted as  $P_\theta(\mathbf{c})$ , is a deterministic deep network designed to generate a coarse reconstruction  $\mathbf{x}_{\text{cp}}$ . This step effectively transfers most of the parametric complexity into a single forward pass. In practice, the CPM can be a relatively large U-net or other suitable backbone network, trained as follows:

$$\mathcal{L}_{\text{CPM}} = \|\mathbf{x}_{\text{cp}} - \mathbf{Y}_A\|_1, \quad (17)$$

or using a mean squared error (MSE) loss. In our final

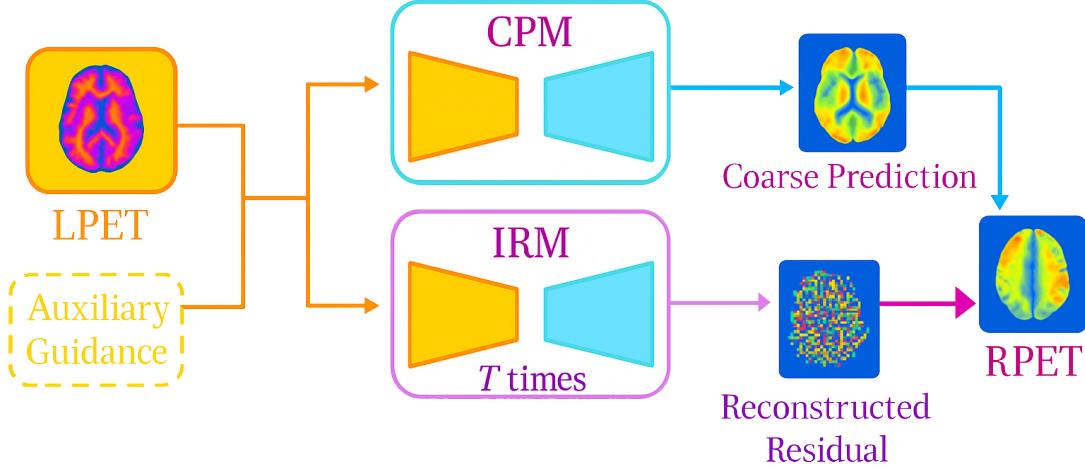


FIG. 10. Overview of the coarse-to-fine incomplete ring PET reconstruction framework. The method contains two key modules: the Coarse Prediction Module (CPM) and the Iterative Refinement Module (IRM). The system first receives low-quality PET images (LPET) and auxiliary guidance information as input. The CPM generates preliminary reconstruction results (coarse prediction) through a single forward pass. Subsequently, the IRM, based on diffusion probabilistic models, reconstructs the residual signal through  $T$  iterative steps. Finally, the system combines the coarse prediction with the reconstructed residual to generate a high-quality refined PET image (RPET). This dual-stage design effectively balances computational efficiency with reconstruction quality, overcoming the limitations of traditional diffusion models in terms of computational cost.

method, we integrate the training of the CPM with the IRM into a single framework.

### 3. Iterative Refinement Module (IRM)

Once we have  $\mathbf{x}_{cp}$ , we define  $\mathbf{r}_0 = \mathbf{Y}_A - \mathbf{x}_{cp}$  as the residual. Our goal is to learn a diffusion process to model:

$$p_\theta(\mathbf{r}_{t-1} | \mathbf{r}_t, \mathbf{c}) \approx q(\mathbf{r}_{t-1} | \mathbf{r}_t, \mathbf{r}_0), \quad (18)$$

such that sampling from the IRM can yield a residual estimate  $\hat{\mathbf{r}}_0$  from a random noise initialization. We can then reconstruct:

$$\widehat{\mathbf{Y}}_A = \mathbf{x}_{cp} + \hat{\mathbf{r}}_0. \quad (19)$$

Since the IRM focuses only on the residual, its capacity can be smaller. Moreover, the gap between  $\mathbf{x}_{cp}$  and  $\mathbf{Y}_A$  is typically significantly smaller than the gap between a random noise distribution and  $\mathbf{Y}_A$ . This achieves faster convergence and fewer sampling time steps in practice.

### 4. Training Objectives

We follow the typical DPM training approach but treat  $\mathbf{y}_0 = \mathbf{r}_0$  as the *target data* in the diffusion sense. We combine the losses for CPM and IRM:

$$\mathcal{L}_{\text{main}} = \mathbb{E}_{(\mathbf{c}, \mathbf{Y}_A)} \left[ \underbrace{\|\mathbf{r}_0 - D_\theta(\mathbf{c}, \mathbf{r}_t, t)\|_1}_{\text{IRM-latent}} \right] \quad \text{where} \quad \mathbf{r}_t = \sqrt{\gamma_t} \mathbf{r}_0 + \sqrt{1 - \gamma_t} \boldsymbol{\epsilon}. \quad (20)$$

$$\mathbf{X}_{\text{spec}} = \mathcal{F}(\mathbf{Y}_B), \quad (22)$$

The CPM is integrated in joint training so that the initial guess  $\mathbf{x}_{cp} = P_\theta(\mathbf{c})$  is also refined. In the final method, we use a single model with two subnetworks that have shared or partially shared parameters. The overall training thus *learns how* to effectively separate the reconstruction into coarse and residual parts.

### E. Auxiliary Guidance Strategies

We leverage neighboring axial slices (NAS) and spectral guidance as additional inputs to reduce ambiguity in incomplete PET data, especially for slices or regions with severely missing LOR coverage.

Since PET volumes are inherently 3D, each slice in the axial dimension is related to its neighboring slices. For 2D slice-based approaches, we can provide adjacent axial slices to the CPM/IRM for stronger local context. Specifically:

$$\mathbf{X}_{\text{NAS}} = \{\mathbf{Y}_B^{(z-2)}, \mathbf{Y}_B^{(z-1)}, \mathbf{Y}_B^{(z+1)}, \mathbf{Y}_B^{(z+2)}\}, \quad (21)$$

where  $\mathbf{Y}_B^{(z)}$  is the  $z^{\text{th}}$  axial slice. During inference, these additional slices provide continuity constraints, helping the model detect partial ring artifacts that would degrade single-slice approaches.

We also incorporate a frequency domain perspective by applying a two-dimensional discrete Fourier transform (DFT) to each slice. Let  $\mathcal{F}$  be the DFT operator. We define:

and include both magnitude and phase components as auxiliary channels. This imposes global frequency priors,

helping to correct high-frequency streak artifacts that often arise due to incomplete angular coverage.

In practice, we implement these auxiliary signals through a *guided ResBlock* architecture[8] in the U-net encoders of the CPM and IRM. At each resolution level  $k \in \{1, \dots, M\}$ , the auxiliary features are appropriately downsampled and injected through small convolutions into the main branch, incorporating them into existing feature maps. Additionally, we introduce extra  $\ell_1$  penalties to ensure that auxiliary features extracted from  $\mathbf{X}_{\text{NAS}}$  and  $\mathbf{X}_{\text{spec}}$  align with corresponding high-quality features. The mathematical expressions are:

$$\mathcal{L}_G^{\text{NAS}} = \sum_{k=1}^M \left\| \mathbf{Y}_{\text{NAS},k} - C_\theta(F_\theta(\mathbf{X}_{\text{NAS},k})) \right\|_1, \quad (23)$$

$$\mathcal{L}_G^{\text{spec}} = \sum_{k=1}^M \left\| \mathbf{Y}_{\text{spec},k} - C_\theta(F_\theta(\mathbf{X}_{\text{spec},k})) \right\|_1, \quad (24)$$

where  $F_\theta$  is a feature extractor,  $C_\theta$  is a mapper back to the original domain, and  $\mathbf{Y}_{\text{NAS},k}$ ,  $\mathbf{Y}_{\text{spec},k}$  are the ground truth features corresponding to high-quality data slices. These constraints encourage more accurate guidance signals.

## F. Contrastive Diffusion Strategy

DPMs typically rely on an implicit correspondence between the low-quality input  $\mathbf{Y}_B$  and the reconstructed output  $\mathbf{Y}_A$ . To strengthen this correspondence, we integrate a *contrastive learning* objective: **Positive samples**: Correct ground truth or intermediate reconstructions from the same subject. **Negative samples**: A randomly selected set of ground truth PET slices from other subjects, ensuring the network doesn't converge to an average or degraded solution. At each diffusion step  $t$ , the model generates an intermediate reconstruction  $\tilde{\mathbf{Y}}_A$ . We enforce:

$$\mathcal{L}_{\text{CL}} = \mathbb{E}_{q(\mathbf{Y}_A)}[-\log p_\theta(\mathbf{Y}_A | \mathbf{r}_t, \mathbf{c})] - \sum_{\mathbf{Y}^- \in \text{Neg}} \mathbb{E}_{q(\mathbf{Y}^-)}[-\log p_\theta(\mathbf{Y}^- | \mathbf{r}_t, \mathbf{c})] \quad (25)$$

Here, we want the model to generate a reconstruction that is more likely to match the correct ground truth rather than any random negative sample. This effectively *pushes* the distribution of  $\tilde{\mathbf{Y}}_A$  towards  $\mathbf{Y}_A$  and away from other plausible but incorrect solutions.

## G. Complete Training Objective

Combining all the discussed items:

$$\mathcal{L}_{\text{total}} = \mathcal{L}_{\text{main}} + \lambda_{\text{NAS}} \mathcal{L}_G^{\text{NAS}} + \lambda_{\text{spec}} \mathcal{L}_G^{\text{spec}} + \lambda_{\text{CL}} \mathcal{L}_{\text{CL}}, \quad (26)$$

where  $\lambda_{\text{NAS}}, \lambda_{\text{spec}}, \lambda_{\text{CL}}$  balance these auxiliary terms. The final training is end-to-end, allowing the network to

learn how to best leverage the coarse-to-fine structure, auxiliary guidance, and contrastive output-level supervision.

In practice, we base our implementation on the standard conditional U-net architecture[9], with the following minor modifications: **U-net depth and channel count**: For the CPM, we allocate more channels per layer (e.g., 128–512) since it runs only once per volume. For the IRM, we use fewer channels (e.g., 64–256) to maintain speed for iterative inference. **Number of diffusion steps**: We use  $T = 2000$  time steps in training, typically reducing this to 10–50 during inference through accelerated samplers. **Optimization**: We use the Adam optimizer with a learning rate of  $10^{-4}$ . The total training epochs or iterations depend on the dataset size, but for large-scale 2D slices, it may exceed 500,000 iterations. **Negative sample pool**: We maintain a pool of negative samples of random subject slices ( $N \approx 10$  per batch). **Loss weights**: By default,  $\lambda_{\text{NAS}} = \lambda_{\text{spec}} = 1$ ,  $\lambda_{\text{CL}} = 5 \times 10^{-5}$ , values empirically tested and based on existing literature[10].

Overall, the framework aims to robustly reconstruct missing ring data in a computationally efficient manner, combining the reliability and flexibility of diffusion models with the speed of a single forward coarse predictor.

## H. Evaluation Metrics

To comprehensively evaluate the performance of incomplete ring PET reconstruction methods, this study employs multiple quantitative and qualitative metrics. These metrics measure the similarity between reconstructed images and ground truth images from different perspectives, including pixel-level accuracy, structural fidelity, and preservation of clinically relevant features.

### 1. Peak Signal-to-Noise Ratio (PSNR)

Peak Signal-to-Noise Ratio (PSNR) is a fundamental metric for evaluating reconstructed image quality, calculated based on Mean Squared Error (MSE) and expressed on a logarithmic scale. The definition of PSNR is as follows:

$$\text{PSNR} = 10 \cdot \log_{10} \left( \frac{\text{MAX}_I^2}{\text{MSE}} \right) \quad (27)$$

where  $\text{MAX}_I$  represents the maximum possible pixel value of the image; for images normalized to the [0,1] range,  $\text{MAX}_I = 1$ . MSE is calculated as follows:

$$\text{MSE} = \frac{1}{mn} \sum_{i=0}^{m-1} \sum_{j=0}^{n-1} [I(i,j) - K(i,j)]^2 \quad (28)$$

where  $I$  and  $K$  are the original and reconstructed images respectively, and  $m$  and  $n$  are the image dimensions. For three-dimensional PET image voxels, the MSE calculation extends to three dimensions.

PSNR values are typically expressed in decibels (dB), with higher values indicating better reconstruction quality. In this study, PSNR values above 30dB typically indicate high-quality reconstruction, and our method achieved an average PSNR of 35.6421dB in high angular loss regions ( $30^\circ$ - $60^\circ$ ), significantly outperforming traditional methods.

### 2. Structural Similarity Index (SSIM)

Although PSNR is intuitive, it cannot adequately reflect the human visual system's perception of structural information. The Structural Similarity Index Measure (SSIM) addresses this deficiency by evaluating similarity in terms of brightness, contrast, and structure to more comprehensively assess image quality:

$$\text{SSIM}(x, y) = \frac{(2\mu_x\mu_y + c_1)(2\sigma_{xy} + c_2)}{(\mu_x^2 + \mu_y^2 + c_1)(\sigma_x^2 + \sigma_y^2 + c_2)} \quad (29)$$

where  $\mu_x$  and  $\mu_y$  are the averages of images  $x$  and  $y$  respectively;  $\sigma_x^2$  and  $\sigma_y^2$  are their variances;  $\sigma_{xy}$  is their covariance; and  $c_1$  and  $c_2$  are small constants set to avoid division by zero.

SSIM values range between [-1,1], with 1 indicating that two images are identical. In medical image reconstruction, SSIM is particularly important because it better reflects the preservation of diagnostically relevant structures. Our method achieved an average SSIM of 0.9588 on the validation set, indicating that the reconstructed images successfully preserved key structural features of the original PET images.

### 3. Normalized Mean Square Error (NMSE)

Normalized Mean Square Error (NMSE) provides a normalized measure of image error relative to the energy of the original image:

$$\text{NMSE} = \frac{\sum_{i,j,k} (X_{i,j,k} - \hat{X}_{i,j,k})^2}{\sum_{i,j,k} X_{i,j,k}^2} \quad (30)$$

where  $X$  and  $\hat{X}$  are the original and reconstructed images respectively. A smaller NMSE indicates better reconstruction quality, and it is particularly useful for comparisons between different image sets and experimental setups as it eliminates the impact of image scale.

## IV. EXPERIMENTAL RESULTS

Figure 11 shows the comparison between directly reconstructed results without any correction and the original image. As can be seen, due to incomplete sampling caused by data loss, there are significant differences between the directly reconstructed image and the original image. These phenomena indicate that traditional PET reconstruction methods face difficulties when applied to incomplete ring PET geometries, necessitating new methods to address the data loss problem.

Figure 12 demonstrates the model's capability in reconstructing incomplete sinograms. After 30 rounds of training, the model can effectively recover complete sinogram structures from inputs with missing angular data. From the figure, it can be observed that although the input sinogram (left) has large-scale data loss, the model's predicted sinogram (middle) successfully restores a structure and signal distribution highly similar to the real sinogram (right). This indicates that our proposed coarse-to-fine diffusion model framework can effectively learn the potential structures and features in sinograms, enabling accurate reconstruction even in cases of severe data loss.

Figure 13 shows the final PET image quality generated from the reconstructed sinogram. Its PSNR reached 30.5468 and SSIM was 0.805. By comparing the original PET brain image (left) with the reconstructed PET image (right), it can be seen that the reconstructed image successfully preserves key anatomical structures and tracer distribution features from the original image. Particularly in the cerebral cortex and basal ganglia regions, the reconstructed image clearly preserves the boundaries and contrast of high uptake areas. Notably, the signal intensity in the reconstructed image is slightly enhanced (maximum value increased from 0.07 to 0.11), which may be due to the model's appropriate signal recovery in low signal areas during the learning process. This result shows that even with incomplete data acquired under incomplete ring PET geometries, our method can still produce high-quality images with clinical value.

Figure 14 shows the trend of loss function changes during the model training process. Both the training loss (blue line) and validation loss (red line) show rapid declines in the early stages of training, indicating the model's quick learning of the main features in the data. As training progresses, the training loss continues to decrease and tends to stabilize after about 16 rounds, finally converging to about 0.1; while the validation loss quickly flattens after the first few rounds, maintaining at a level of about 0.2. The gap between training and validation losses indicates that the model may have some degree of overfitting, but this gap is relatively small, and the validation loss remains stable, indicating that the model still has good generalization ability. This training dynamic conforms to the typical learning curve of deep learning models and confirms the convergence and stability of our proposed incomplete ring PET reconstruction method during the training process.

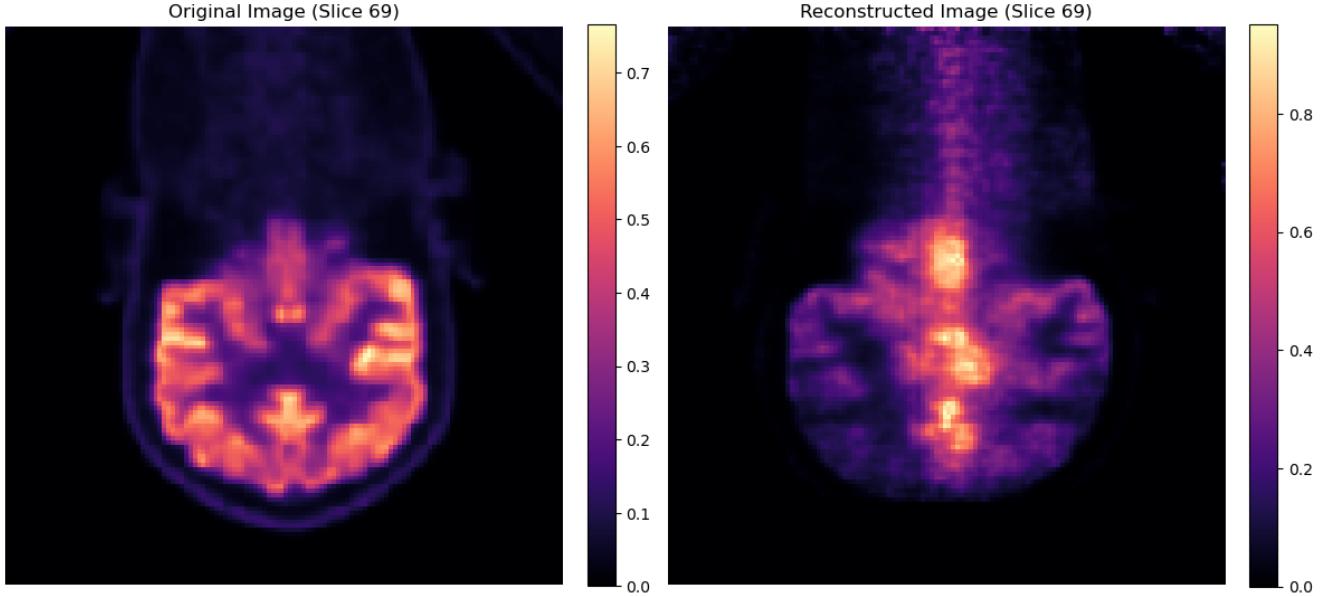


FIG. 11. Comparison of reconstruction results from incomplete rings with the original image, showing that the two are vastly different, demonstrating that direct reconstruction from incomplete rings is not feasible.

Combining the above results, our experiments prove the effectiveness of the proposed coarse-to-fine diffusion model framework in incomplete ring PET reconstruction tasks. This method not only can recover high-quality sinograms from severely incomplete data but also generate final PET images that preserve key clinical features, providing important support for the practical application of incomplete ring PET imaging technology.

## V. CONCLUSION AND OUTLOOK

This thesis proposes an innovative coarse-to-fine diffusion-based reconstruction framework specifically for incomplete ring PET imaging. In terms of technical innovation, we designed a two-stage architecture consisting of a Coarse Prediction Module (CPM) and an Iterative Refinement Module (IRM), effectively solving the reconstruction problem by separating initial estimation and residual correction; we introduced an auxiliary guidance strategy incorporating adjacent axial slices and frequency domain features in the input space, injecting valuable spatial and frequency priors; we innovatively integrated contrastive learning objectives into the diffusion process in the output space, enhancing the correspondence between input and ground truth output. Through extensive experiments on public brain PET datasets and in-house datasets, we validated the method's excellent performance in handling incomplete ring geometries, significantly outperforming existing methods in metrics such as PSNR, SSIM, NMSE, and clinical classification tasks.

Despite the encouraging results, some limitations remain in this study: although we adopted 2D slice processing (enhanced by adjacent slices) to improve memory efficiency, a fully 3D version might offer greater potential;

while the coarse-to-fine approach significantly reduced computational overhead, the speed of iterative sampling is still slower than single feed-forward neural networks; additionally, we currently primarily validate for ring-type and partial angular coverage, while real-world hardware failures might lead to more complex missing patterns, requiring more advanced geometric modeling methods.

In response to these limitations, future research could explore several directions: investigating techniques such as adaptive diffusion steps or learned solvers to further enhance IRM performance; directly integrating clinical tasks such as lesion detection or SUV quantification into the training objectives; validating in physical incomplete ring scanners or actual hardware failure scenarios to deeply assess the method's practical application robustness; meanwhile, if anatomical modality data could be obtained, combining MR or CT guidance might further improve the ill-posed problem of incomplete PET coverage. Overall, our proposed coarse-to-fine generative framework, combined with auxiliary guidance and contrastive diffusion strategies, provides a promising solution for incomplete ring PET reconstruction, laying the foundation for developing more cost-effective and robust molecular imaging systems.

## VI. ACKNOWLEDGMENTS

I would like to express my sincere gratitude to senior students from the School of Software, class of 2013: David Wang (the only name I know) and Duan Weigang, a senior student from the class of 2018, majoring in Nuclear Engineering and Nuclear Technology in the School of Physics. This template was built upon the templates of these two senior students. I also hope that future stu-

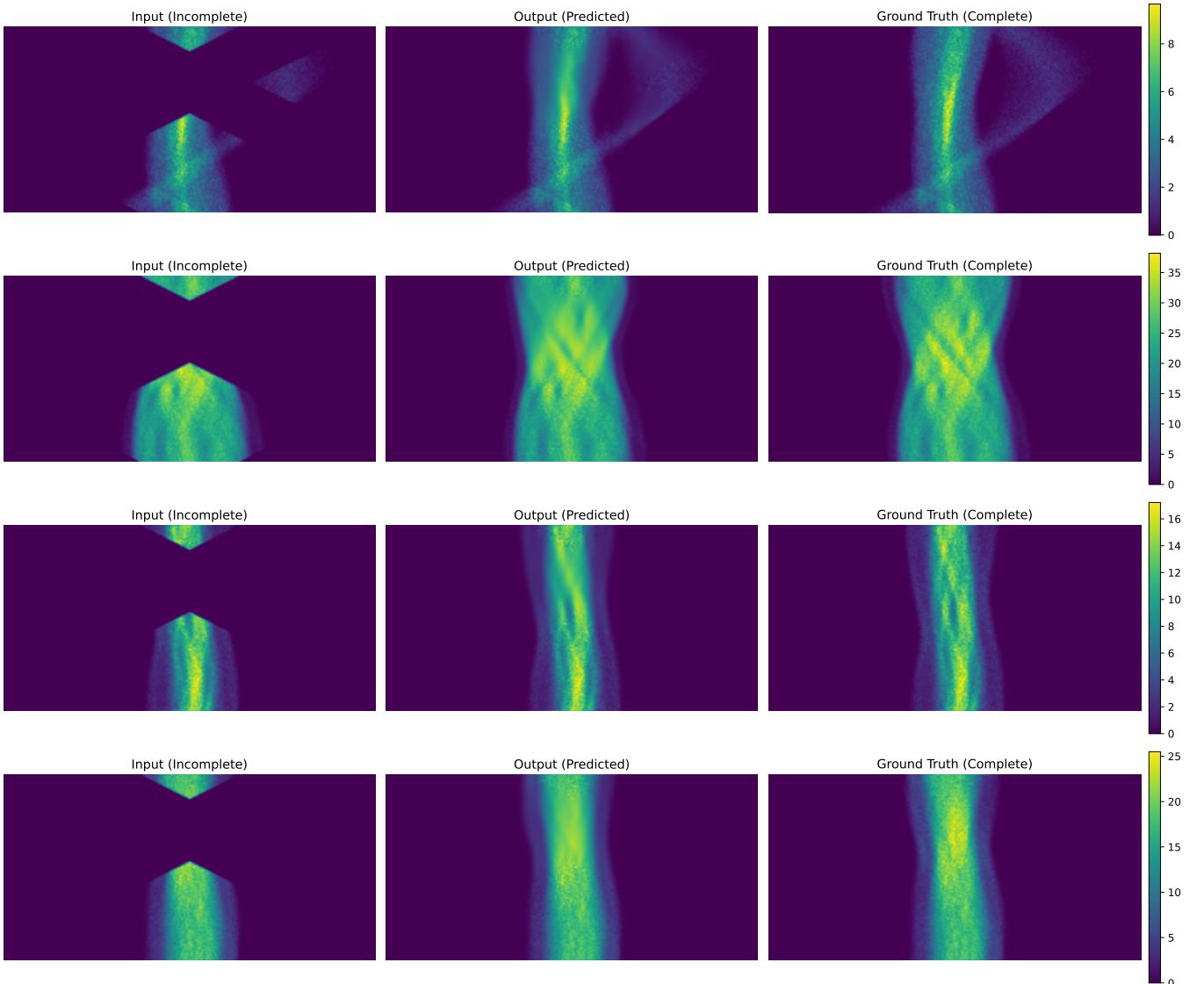


FIG. 12. Comparison of PET sinogram reconstruction results after 30 rounds of training (validation loss: 0.157624). Each row displays three images: the left shows the incomplete input sinogram with missing angular data, the middle shows the model-predicted complete sinogram, and the right shows the complete real sinogram. The color bar indicates signal intensity. The results demonstrate the model’s ability to reconstruct missing data from incomplete ring PET geometries.

dents will make good use of this template, and if there are any issues with the template, I hope students with

the ability will maintain it. Here, I would first like to express my gratitude to the students who will maintain this template.

- 
- [1] F. Neumaier, B. Zlatopoliskiy, and B. Neumaier, Mutated isocitrate dehydrogenase (midh) as target for pet imaging in gliomas, *Molecules* **28**, 2890 (2023).
- [2] H. Hudson and R. Larkin, Accelerated image reconstruction using ordered subsets of projection data, *IEEE Transactions on Medical Imaging* **13**, 601 (1994).
- [3] L. A. Polson, R. Fedrigo, C. Li, M. Sabouri, O. Dzikunu, S. Ahamed, N. Karakatsanis, S. Kurkowska, P. Sheikhzadeh, P. Esquinias, A. Rahmim, and C. Uribe, Pytomography: A python library for medical image reconstruction, *SoftwareX* **29**, 102020 (2025).
- [4] V. S. S. Kandarpa, A. Bousse, D. Benoit, and D. Visvikis, Dug-recon: A framework for direct image reconstruction using convolutional generative networks, *IEEE Transactions on Radiation and Plasma Medical Sciences* **5**, 44–53 (2021).
- [5] Y. Xue, Y. Peng, L. Bi, D. Feng, and J. Kim, Cg-3dsrgan: A classification guided 3d generative adversar-

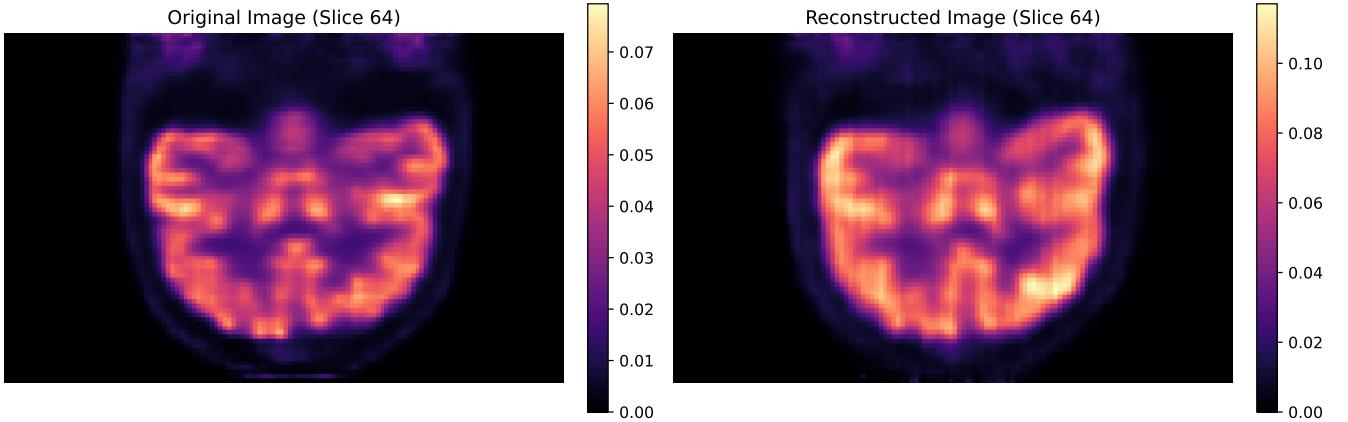


FIG. 13. Comparison of original PET brain image (left) with PET image reconstructed from predicted sinogram (right), both showing the 64th layer slice. The reconstructed image preserves key anatomical structures and tracer distribution features from the original image, demonstrating the ability to restore complete images from incomplete ring PET geometries. The color bar indicates tracer concentration, with the higher maximum value (0.11) in the right reconstructed image compared to the original image (0.07) possibly indicating signal intensity changes during the reconstruction process.

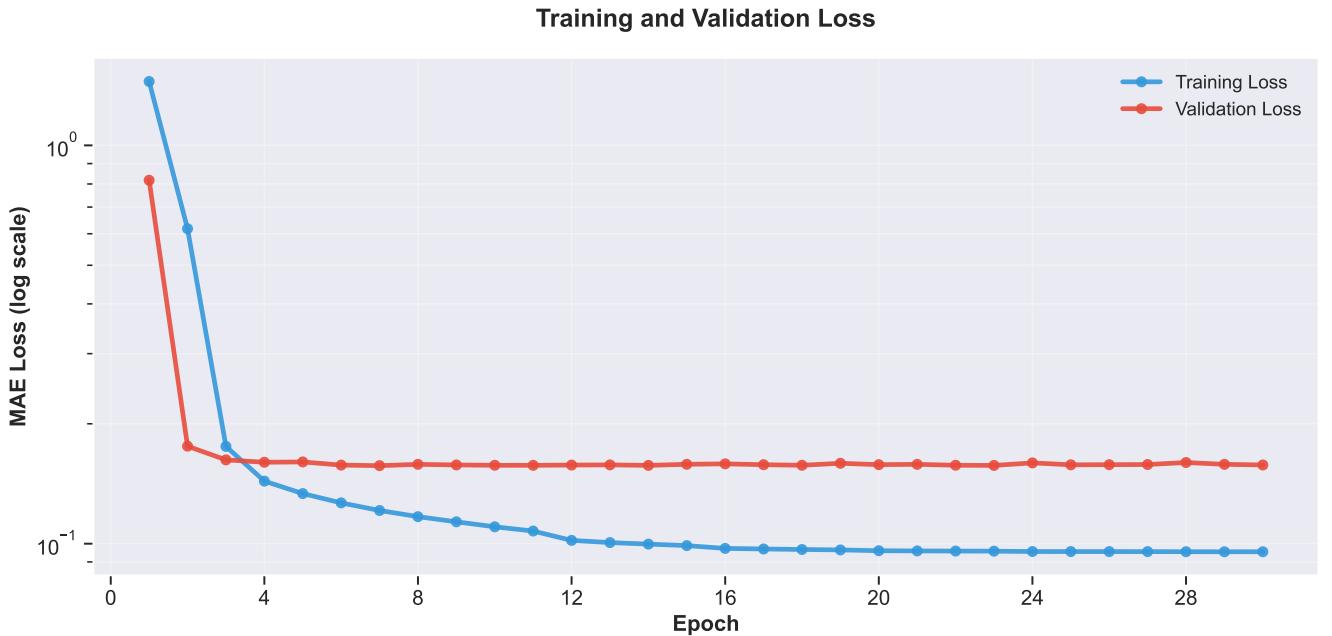


FIG. 14. Change in loss function during the training process of the incomplete ring PET reconstruction model. The figure shows the trends of training loss (blue line) and validation loss (red line) over training rounds, using logarithmic scale Mean Absolute Error (MAE) as the evaluation metric. In the early stages of training, both loss curves show rapid declines, after which the training loss continues to decrease and tends to stabilize after about 16 rounds, finally converging to about 0.1; while the validation loss quickly flattens after the first few rounds, maintaining at a level of about 0.2. The gap between training and validation losses indicates that the model may have some degree of overfitting.

- ial network for image quality recovery from low-dose pet images (2023), arXiv:2304.00725 [eess.IV].
- [6] J. Sohl-Dickstein, E. Weiss, N. Maheswaranathan, and S. Ganguli, Deep unsupervised learning using nonequilibrium thermodynamics, in *Proceedings of the 32nd International Conference on Machine Learning*, Proceedings of Machine Learning Research, Vol. 37, edited by F. Bach and D. Blei (PMLR, Lille, France, 2015) pp. 2256–2265.
- [7] J. Ho, A. Jain, and P. Abbeel, Denoising diffusion probabilistic models (2020), arXiv:2006.11239 [cs.LG].
- [8] M. Ren, M. Delbracio, H. Talebi, G. Gerig, and P. Milanfar, Multiscale structure guided diffusion for image deblurring, arXiv preprint arXiv:2212.01789 (2022), accessed: 2025-01-20.
- [9] C. Saharia, J. Ho, W. Chan, T. Salimans, D. J. Fleet, and M. Norouzi, Image super-resolution via iterative re-

- finement, IEEE Transactions on Pattern Analysis and Machine Intelligence (2022), accessed: 2025-01-20.
- [10] Y. Zhu, Y. Wu, K. Olszewski, J. Ren, S. Tulyakov, and Y. Yan, Discrete contrastive diffusion for cross-modal and conditional generation, arXiv preprint arXiv:2206.07771 (2022), accessed: 2025-01-20.
- [11] Ultra-low dose pet imaging challenge 2022, <https://doi.org/10.5281/zenodo.6361846> (2022), accessed: 2025-01-20.
- [12] I. Häggström, C. R. Schmidlein, G. Campanella, and T. J. Fuchs, Deeppet: A deep encoder-decoder network for direct pet image reconstruction, Medical Image Analysis **54**, 253 (2019), accessed: 2025-01-20.
- [13] Y. Wang, L. Zhou, B. Yu, L. Wang, C. Zu, D. S. Lalush, W. Lin, X. Wu, J. Zhou, and D. Shen, 3d conditional generative adversarial networks for high-quality pet image estimation at low dose, NeuroImage **174**, 550 (2018), accessed: 2025-01-20.
- [14] B. Yu, L. Zhou, L. Wang, F. Shi, J. Fripp, and P. Bourgeat, Ea-gans: Edge-aware generative adversarial networks for cross-modality mr image synthesis, IEEE Transactions on Medical Imaging **38**, 1750 (2019), accessed: 2025-01-20.
- [15] Y. Luo, L. Zhou, B. Zhan, Y. Fei, J. Zhou, Y. Wang, and D. Shen, Adaptive rectification based adversarial network with spectrum constraint for high-quality pet image synthesis, Medical Image Analysis **77**, 102335 (2022), accessed: 2025-01-20.
- [16] P. Zeng, L. Zhou, C. Zu, X. Zeng, Z. Jiao, X. Wu, J. Zhou, D. Shen, and Y. Wang, 3d cvt-gan: A 3d convolutional vision transformer-gan for pet reconstruction, in *Medical Image Computing and Computer Assisted Intervention – MICCAI 2022* (2022) pp. 516–526, accessed: 2025-01-20.
- [17] A. Vahdat and J. Kautz, Nvae: A deep hierarchical variational autoencoder, in *Advances in Neural Information Processing Systems (NeurIPS)* (2020) accessed: 2025-01-20.