PNMC: Four-dimensional conebeam CT reconstruction combining prior network and motion compensation

Zhengwei Ou, Jiayi Xie, Ze Teng, Xianghong Wang, Peng Jin, Jichen Du, Mingchao Ding, Huihui Li, Yang Chen, Tianye Niu

PNMC: Four-Dimensional Conebeam CT Reconstruction Combining Prior Network and Motion Compensation

Zhengwei Ou[a,b], Jiayi Xie[c,d], Ze Teng[c,e], Xianghong Wang[b], Peng Jin[b], Jichen Du[f], Mingchao Ding[f], HuiHui Li[b], Yang Chen[a,g,*], Tianye Niu[b,f,**]

[a]*School of Computer Science and Engineering, Southeast University, Nanjing, China*
[b]*Shenzhen Bay Laboratory, Shenzhen, China*
[c]*Beijing Key Laboratory of Magnetic Resonance Imaging Devices and Technology, Peking University Third Hospital, Beijing, China*
[d]*Department of Automation, Tsinghua University, Beijing, China*
[e]*Department of Radiology, National Cancer Center/National Clinical Research Center for Cancer/Cancer Hospital, Chinese Academy of Medical Sciences and Peking Union Medical College*
[f]*Peking University Aerospace School of Clinical Medicine, Aerospace Center Hospital, Beijing, China*
[g]*Centre de Recherche en Information Biomedical SinoFrancais, Rennes, France*

ARTICLE INFO

Article history:

Keywords: 4D conebeam CT, deep learning, sparse-view CT reconstruction, prior image, motion-compensated.

ABSTRACT

Four-dimensional conebeam computed tomography (4D CBCT) is an efficient technique to overcome motion artifacts caused by organ motion during breathing. 4D CBCT reconstruction in a single scan usually divides projections into different groups of sparsely sampled data based on the respiratory phases. The reconstructed images within each group present poor image quality due to the limited number of projections. To improve the image quality of 4D CBCT in a single scan, we propose a novel reconstruction scheme that combines prior knowledge with motion compensation. We apply the reconstructed images of the full projections within a single routine as prior knowledge, providing structural information for the network to enhance the restoration structure. The prior network (PN-Net) is proposed to extract features of prior knowledge and fuse them with the sparsely sampled data using an attention mechanism. The prior knowledge guides the reconstruction process to restore the approximate organ structure and alleviates severe streaking artifacts. The deformation vector field (DVF) extracted using deformable image registration among different phases is then applied in the motion-compensated ordered-subset simultaneous algebraic reconstruction algorithm to generate 4D CBCT images. Proposed method has been evaluated using simulated and clinical datasets and has shown promising results by comparative experiment. Compared with previous methods, our approach exhibits significant improvements across various evaluation metrics.

1. Introduction

Conebeam Computed Tomography (CBCT) is widely used in image-guided radiation therapy (IGRT) and surgery. CBCT enables precise treatment using real-time monitoring and adjustment of patient positioning, minimizing the impact of patient motion during the treatment. This improves treatment effec-

*Corresponding author.
**Corresponding author.
e-mail: niuty@szbl.ac.cn (Tianye Niu), chenyang.list@seu.edu.cn (Yang Chen)

tiveness and maximizes the safety of surrounding healthy tissues [10, 13]. The scanning duration, typically lasting around one minute per rotation and covering approximately 10-20 respiratory cycles, often leads to significant motion distortions, especially in the chest and upper abdomen [20]. These distortions can produce artifacts that may introduce biases within treatment, leading to high radiation exposure to healthy tissues and a decrease in the intended dose to the targeted tumor site [6]. Hence, there is an urgent need to mitigate motion artifacts in CBCT imaging.

Four-dimensional (4D) CBCT is an advanced imaging technique to mitigate motion artifacts [7]. It involves respiratory motion monitoring, projection acquisition, projection sorting, reconstruction, and motion artifact correction. By utilizing the patient's respiratory profile, multiple groups of projections are divided into distinct respiratory phases. Reconstructing images from these groups produces 3D dynamic sequence images [8]. The widespread adoption of 4D CBCT is restricted by the long scanning time and complex instrument design using existing 4D CT equipment with respiratory gating devices. Using a single regular 3D CBCT scan for 4D reconstruction is thus appealing. Nevertheless, each group of projections exhibits non-uniform and sparse angular distribution, resulting in insufficient data condition and compromised accuracy of image reconstruction.

To tackle this challenging problem, numerous 4D CBCT reconstruction algorithms have been proposed. These methods can be divided into respiratory-correlated and motion-compensated methods according to the various use of breathing signals. Respiratory-correlated methods reconstruct each phase sequence image using the acquired sparse projections [28]. This strategy explores the substantial temporal correlation among various phases by incorporating a spatiotemporal framework. Specifically, an iterative approach for 4D CBCT reconstruction is proposed to incorporate the temporal non-local means (TNLM) regularization term [21]. This method enables the simultaneous reconstruction of all phase images, yielding improved performance compared with the utilization of total variation (TV). An extension of spatial TV to the spatial-temporal domain is also employed for 4D cardiac imaging [19] to suppress motion artifacts. Nevertheless, excessive regularization may result in visual distortions and over-smoothing, degrading the quality of reconstruction results [5].

Deformable image registration is commonly used for motion-compensated 4D CBCT reconstruction. This type of method investigates the correlation among distinct phases by extracting deformation vector fields (DVFs) from CBCT images of various phases to account for motion artifacts [5]. The artifact model-based cyclic motion-compensation algorithm (acMoCo) acquires inaccurate DVF disturbed by multitudinous artifacts [3]. Simultaneous motion estimation and image reconstruction method (SMEIR) [23] includes the motion model in the iterative reconstruction process to improve the precision of motion model estimation and considerably enhance the quality of image reconstruction. Motion-compensated methods usually assume a regular breathing pattern at each respiratory phase [29]. When the amplitude and period of breathing are strongly irregular, the DVF estimation accuracy may be degraded, resulting in the decline of reconstruction quality.

Deep learning methods are extensively employed in medical imaging to optimize and enhance medical images using either large datasets or the principles of transfer learning. These techniques have been combined with traditional methods to enhance efficiency and precision [15]. In 4D CBCT reconstruction, neural networks have shown their ability to learn from extensive datasets, extracting crucial image features and filling in missing information from sparsely sampled views. For example, U-Net-based interpolation has improved image quality by filling in missing data in sparse-view sinograms, reducing artifacts [18]. Additionally, iCT-net [19] is a novel approach that seamlessly integrates the reconstruction process within the network architecture, optimizing the entire reconstruction pipeline. Moreover, researchers have explored incorporating prior images into CNN-based methods to boost reconstruction performance. CycN-Net [32], for instance, introduces an innovative approach that encodes both the degraded and prior images, leveraging their combined information during the decoding step. This integration of prior knowledge offers new opportunities for enhancing the quality and accuracy of medical image reconstructions, benefiting diagnostic and clinical applications.

In recent times, there has been a proliferation of methodologies that amalgamate deep learning with motion compensation strategies. For instance, the 3D U-net has been employed to enhance image quality in conjunction with motion compensation techniques, as demonstrated in [27]. In our prior research, we leveraged this approach by utilizing a multi-scale adversarial network to improve sparse-view reconstruction, as outlined in [26]. Nonetheless, these deep learning networks exhibit limitations in terms of their generalization capacity and robustness. Their performance on test datasets may not consistently meet the requirements for medical treatment. Consequently, the robustness of the network becomes crucial, particularly given the discrepancies in data distribution between the training and test datasets.

In addition to 4D reconstruction algorithms, respiratory profile extraction is also critical for accurate projections phasing. Phase measurement techniques are employed to acquire respiratory profiles and these techniques can be divided into two types based on the invasion style. Invasive methods involve the insertion of fiducial markers including metal particles into the human body to track respiratory movements. Nevertheless, these methods require medical surgery to place the markers and may pose additional risks to the patient. Noninvasive methods generally utilize various instruments or algorithms to monitor respiratory motion, including the diaphragmatic localization monitoring. The diaphragm position in upper abdominal CBCT imaging can be obtained directly from the projection images without additional sensors [4].

In this study, we propose a novel method to generate 4D CBCT images combining prior knowledge with motion compensation from a single routine scan. The respiratory phase information is derived directly from the diaphragm position of the original projections without additional sensors. The proposed PN-Net is applied to improve the initial phase-sorted image

quality and the estimation accuracy of DVF due to its enhancement capability. Specifically, we propose a dual encoder structure network based on a conditional U-Net structure to integrate prior knowledge. The prior image is reconstructed using all projections belonging to the respiratory phases. To extract global information, we add an attention mechanism to the decoder of the network. The DVFs from enhanced images are applied in the 4D CBCT reconstruction using a standard ordered-subset simultaneous algebraic reconstruction (OS-SART). Compared with previous works, we address the non-equiangular distribution of projections caused by irregular breathing and achieve high-quality 4D CBCT images from a conventional 3D CBCT scan.

2. Methodology

The proposed workflow illustrated in Fig. 1 consists of the following steps. The CBCT projections acquired from a routine 3D scan are processed to pinpoint the position of the diaphragm and extract the respiratory curve. Initially, the projections sorted based on the respiratory curve are reconstructed to generate 4D sparse-view CBCT images, characterized by pronounced streaking artifacts. A prior network, incorporating the prior image reconstructed using all the projections, is trained to enhance the quality of these sparse-view CBCT images. The network effectively mitigates streaking artifacts and restores anatomical detail, although some fine structures remain absent. Deformable image registration is applied to derive the DVFs among optimized images to compensate for the anatomical detail loss. The motion-compensated iterative reconstruction utilizes both the DVFs and the measured projection as inputs, enabling the reconstruction of high-quality 4D CBCT images.

2.1. Projection Phasing

The diagram detailing the proposed phase-sorting method is displayed in Fig. 2. We utilize logarithmic operations and first-order differentiation to heighten the contrast between the lung and liver in each projection. These contrast-enhanced projections are then converted into a gray-scale image and summed along the lateral detector direction to eradicate the fluctuation of the diaphragm locations within the projections. Each summed 1D vector is sequentially arranged within a 2D matrix, with the matrix coordinates corresponding to the rotation angle and the detector's spatial position. The respiratory signal is obtained by aligning each vector within the matrix. A low-pass filter, incorporating a five-point window moving average, is implemented to smooth the motion signal. The final motion curve is derived by normalizing the filtered respiratory signal.

2.2. Network Inputs Generating

In the preprocessing step, the original projections are denoised using the penalty-weighted least-square (PWLS) algorithm and are mapped into the line integration domain [24]. Processed projections are reconstructed using a standard Feldkamp-Davis-Kress(FDK) algorithm [8] and is written as follows:

$$X_k = FDK(Y_k), \tag{1}$$

where Y_k represents the projection of the k-th phase, X_k represents the sparse-view reconstruction image containing characteristics of the k-th respiratory phase along with significant artifacts and noise. Y_k is divided by the respiratory curve, causing biases in the quantity and distribution of projections for each phase due to the patient's irregular breathing patterns. The variation in the duration and amplitude of each breath results in a partial quality gap among the sparse-view reconstruction images. When employing projections from all phases in FDK reconstruction, the resultant image serves as the integrated prior image. Despite encompassing information from all phases and consequently displaying motion artifacts, it maintains distinct organ structures.

2.3. PN-Net Enhancement

As shown in Fig. 3, we introduced a novel network, named PN-Net, to address severe streak artifacts and the loss of structural information induced by sparse-view projection and non-equiangular distribution. Drawing on the U-net architecture, the PN-Net adopts an encoder-decoder structure, further augmented with a skip connection. In contrast to CycN-Net [29], which incorporates prior knowledge in the decoder, PN-Net opts to fuse prior knowledge within the additional encoder to fully integrate this information. Both encoders employ four down-sampling operations to generate feature maps of varying spatial resolutions, with these feature maps being directly added to the subsequent block [25]. Our results demonstrate that this prior feature fusion module contributes to a more effective static structure.

The network encoders use dense blocks, which consist of multiple sublayers to extract multilayer features. Compared to traditional convolutional layers, dense blocks interconnect the feature maps produced by all previous layers, thus integrating low-level and high-level information. As demonstrated in Fig. 3(b), the inputs undergo a 2×2 max pooling operation and sublayers using 2D convolution blocks (BN+ReLU+Conv) with dense connectivity [11]. This dense block structure simplifies network training and ensures compactness, resolving issues associated with gradient disappearance and explosion [12].

Due to the localized action of the convolution kernel's receptive field, features of distinct regions may become correlated after multiple convolution layers, thereby limiting network efficiency and performance. To widen the receptive field, we revised the decoder phase to include a self-attention technique after the residual block. Consequently, up-sampling of features can leverage global context information to restore spatial structure [22, 25]. As illustrated in Fig. 4, the non-local block generates three types of matrices through a convolution operation, corresponding to the query, key, and value in the self-attention mechanism. The central concept posits that the response at a pixel is the summation of feature weights across all other pixels. The equation for this block can be expressed as follows:

$$y_i = \frac{1}{c(x)} \sum_{\forall j} f(x_i, x_j) g(x_j), \tag{2}$$

where x represents the input feature image, i represents the response at the current location, j represents the global response, The function f calculates the correlation among i-th position and all other positions; The function g can map a point to a vector;

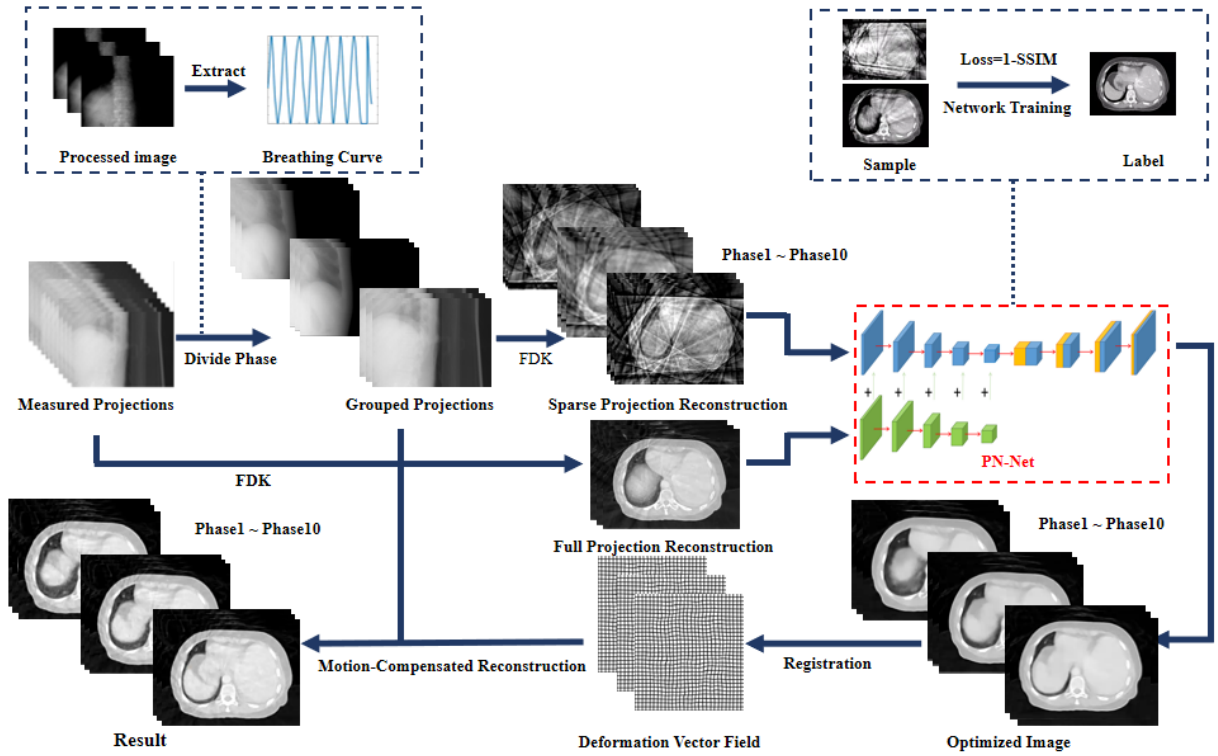

Fig. 1. The workflow of the proposed PNMC method is presented. The projections of a routine 3D CBCT scan are initially divided into distinct phases according to the position of the diaphragm. The sparse-view images reconstructed by grouped projections are then enhanced using the pre-trained PN-Net. The PN-Net fuses the full projection reconstruction as the prior knowledge. The enhanced images are used to estimate DVFs among different phases. The DVFs and phase-sorted projections are applied to the motion-compensated iterative reconstruction.

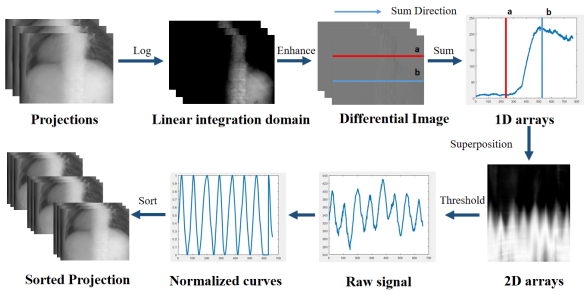

Fig. 2. Diagram illustrates the proposed method for extracting respiratory curves. The half-fan projections are modified to increase edge contrast. The enhanced images are summed in the extension direction and stacked chronologically. After a filter, the respiratory signal is extracted as the foundation for projection phasing.

The function c is the normalized function; The block calculates the correlation based on the position and generates a new feature map with the same size. In Fig. 4, there is the operation of the residual connection adding the input to the output as the result, which can maintain network stability.

The loss function of the network is computed with structure similarity index measure(SSIM):

$$\mathrm{L} = 1 - \mathrm{SSIM} = 1 - \frac{(2\mu_x\mu_y + c_1)(\sigma_{xy} + c_2)}{(\mu_x^2 + \mu_y^2 + c_1)(\sigma_x^2 + \sigma_y^2 + c_2)}, \quad (3)$$

where u_x and u_y represent the mean value in the image x and image y, σ_x and σ_y represent the standard deviance between the image x and image y, and σ_{xy} represents the covariance. c_1 and c_2 represent constants to avoid system errors caused by a denominator of zero. SSIM is a perception-based computational model, it can focus on the fuzzy changes of the structural information in human perception. The value of SSIM close to 1 means the correlation is strong, so we subtract its value from 1 as a loss function. Compared with other loss functions, it enables a more accurate image quality assessment.

2.4. Motion Compensated Iterative Registration

The image optimization achieved using the PN-Net demonstrates a noteworthy reduction in artifacts and an enhancement in structural clarity. Nonetheless, owing to inherent limitations in information availability, certain structural details may be compromised. In order to recapture these missing details, a deformation vector field (DVF) is computed through deformable registration between each pair of phases. Subsequently, this DVF is utilized to compensate for motion during both forward and backward projection in the iterative reconstruction process, resulting in the recovery of the lost details [1]. Fig. 5 illustrates the overall flow of this process:

$$\mu_t = D_{k\to t}\mu_k, \quad (4)$$

where μ_t and μ_k denote images in the t-th phase and k-th phase respectively, while $D_{k\to t}$ signifies deformation fields registered from the k-th phase to the t-th phase. Consequently, the transformation among different phase images can be achieved by deformation fields, inducing the following formula:

$$p_t = AD_{k\to t}\mu_k, \quad (5)$$

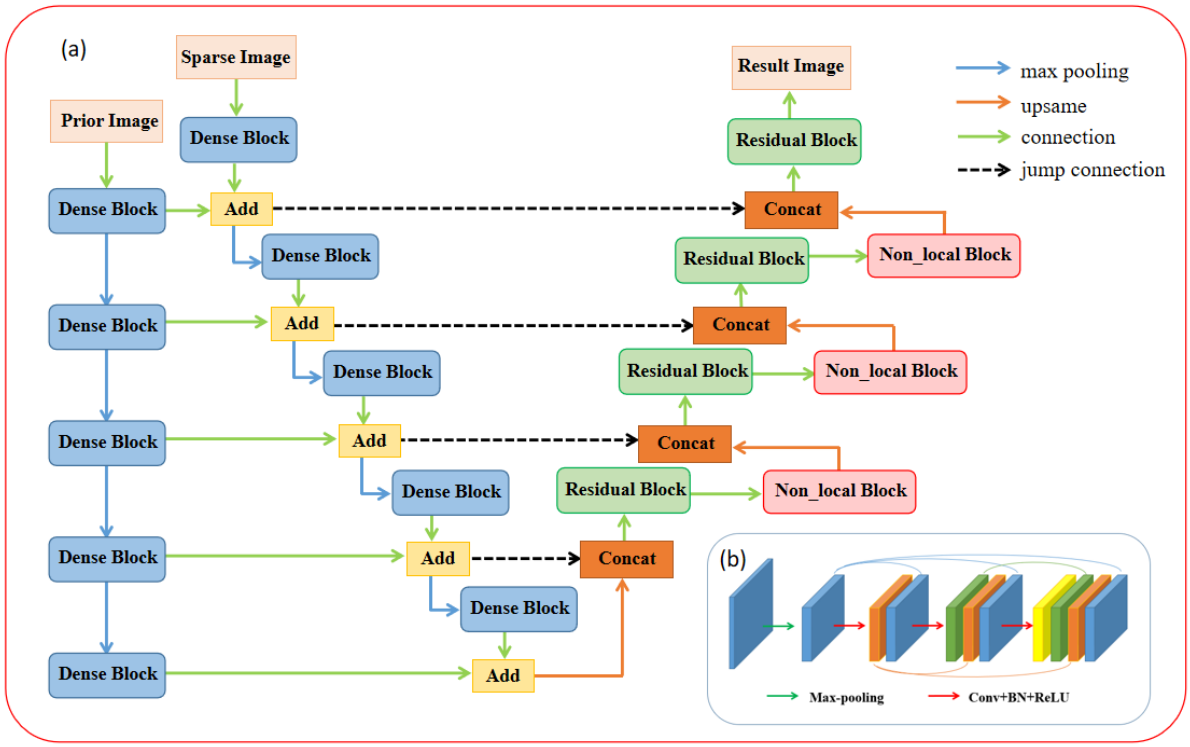

Fig. 3. (a) The architecture of the proposed PN-Net. PN-Net incorporates an encoder path to fusion the prior image, the outputs of both encoders are summed in each down-sampling step. The non-local module is added in the decoder to recover image features based on global features. This design can greatly utilize input information. (b) The framework of the dense block. The maps are directly overlaid after convolution layers, sharing information from multiple levels.

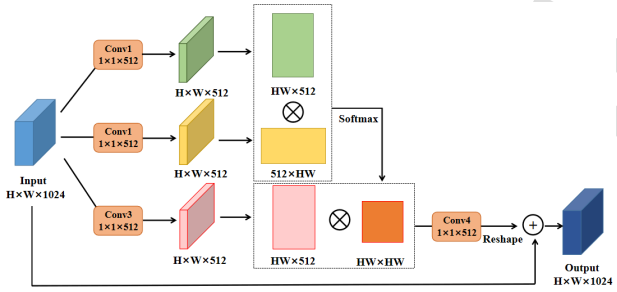

Fig. 4. Specific process of the non-local block. Three characteristic images are obtained by convolution, and the correlation among image positions is calculated by dimension reduction. The feature map obtained through the self-attention mechanism is then transformed into the same size as the input and added to the input as the final result.

where A represents the projection matrix and p_t corresponds to the projected images of the t-th phase. Hence, the projection image of any phase can be obtained by using the deformation field. The projection of each respiratory phase serves as an ordered subset, and the Ordered-Subset Simultaneous Algebraic Reconstruction Technique (OS-SART) used to perform these ordered subsets is computationally efficient.

$$x_t^{i+1} = x_t^i + \lambda_n V_k^{-1} D_{k\to t}(A_k^T(W_k(p^k - A_k D_{t\to k}(x_t^i))), \quad (6)$$

where x_t^i represents the reconstructed image of the t-th phase at the i-th iteration. λ_n is employed to update the relaxation parameters, beginning at 1 and decreasing by 10e-2 at each iteration. V_k^{-1} signifies the diagonal weighted matrix of the forward projection, while W_k denotes the weighting matrix for backward projection [16].

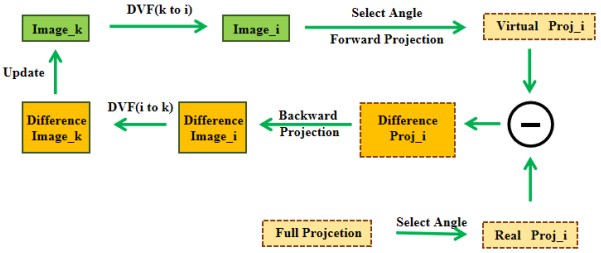

Fig. 5. Detailed workflow of motion-compensated iterative reconstruction of the image in the k-th phase. The k-th phase image is deformable registered to the i-th phase image and then subtracts the i-th phase measured projections. Finally, the k-th phase image is updated with the difference image. Iterating this process may compensate for missing information.

2.5. *Implementation Details*

In the network training phase, the PN-Net utilizes Adam algorithms for optimization, with set β_1=0.9, β_2=0.999. The learning rate is initially set at 10e-4, which is adjusted to one-tenth of its original value every 20 iterations, finally decreasing to 10e-6. The batch size is established at 2. The training sequence concludes after 50 iterations, taking approximately twenty hours. All the processes outlined in this article are run

Table 1. Parameters of simulated and clinical data acquisition

Data Sources	**Simulated Phantom**	**Hospital Clinical Trial**	**SPARE**
SID/SAD	1500/1000	1500/1000	1500/1000
Pixel Size(mm^2)	0.5x0.5	0.388x0.388	0.388x0.388
Detector Size	1024x768	1024x768	1008x752
Angle Range	0°-360°	0°-360°	0°-360°
Voxel Size(mm^3)	0.75x0.75x2	0.75x0.75x2	1x1x1
Dimension	512x512x120	512x512x120	450x450x220
Projection Number	600	660	680

on a PC with the following specifications: Intel(R) Core(TM) i7-6900K CPU@3.20GHz CPU, NVIDIA GeForce GTX 1080 Ti GPU, and 128GB memory.

3. EXPERIMENTS

3.1. Dataset

In 4D CBCT reconstruction, obtaining high-quality images for deep learning across all respiratory phases significantly complicates data acquisition. To overcome this challenge, we devised a simulation method involving the projection of 4D CT images from 13 patient cases, each comprising ten volumetric images representing ten breathing phases. By incorporating the respiratory signal into high-quality 4D CT images, we were able to simulate corresponding sparse-view images and prior images. The slice images of ten patients were assembled to create the training dataset, which consisted of approximately 12,000 CBCT images after data augmentation. The remaining three patients constituted the testing dataset, comprising around 1,800 CBCT images. Throughout the training process, we performed scaling normalization on all images, wherein pixel-wise or voxel-wise intensities were normalized to fall within the data range of [-1, 1]. We assessed the proposed strategy on two types of clinical patient datasets to verify the method's reliability. The first is a set of CBCT projections from patients with liver cancer who underwent routine CBCT scans and SBRT at our institution. The second is the public dataset from the SPARE Challenge (https://image-x.sydney.edu.au/spare-challenge/). Both of these clinical datasets exhibit substantial differences from the training set, thereby serving as a rigorous test to evaluate the robustness of our workflow. The collection parameters of the dataset are presented in Table 1.

3.2. Evaluation and Comparison

3.2.1. Comparison Methods

The proposed methodology is compared with existing methods under identical datasets to ensure an accurate performance assessment. This comparison is bifurcated into two stages:

In the first stage, we conducted a comprehensive comparative analysis among three distinct reconstruction approaches: Cycn-Net, MaMO, and the conventional FISTA-TV method, as elaborated in Table 2. Cycn-Net, aiming to our proposed approach, incorporates prior knowledge to enhance image quality. However, it relies on sparse-view images from adjacent phases as its prior knowledge source and is exclusively dependent on neural network optimization, lacking integration with conventional

Table 2. Description of different methods for comparison

Method	**Description**
PNMC	The optimization of CBCT image is achieved through the implementation of PN-Net enhancement combined with motion-compensated reconstruction techniques. This approach leverages prior knowledge to accurately extract deformation fields, enhancing the precision of the process.
MaMO	The optimization of CBCT image is achieved through the implementation of MSD-GAN enhancement combined with motion-compensated reconstruction techniques. MSD-GAN utilizes multiple discriminators to optimize images to obtain better deformation fields for motion compensation.
CycN-Net	The optimization of CBCT image is achieved through the utilization of an enhanced prior image network. This approach involves incorporating information from multiple adjacent phases as prior knowledge into the network, aimed at enhancing the network's capabilities and compensating for motion-related information.
FISTA-TV	The optimization of CBCT image is achieved through a reconstruction approach that involves the phase-sorted projections being processed using the Fast Iterative Shrinkage Thresholding Algorithm (FISTA) with Total Variation (TV) regularization. This method reduces noise by encouraging small changes in the gradient magnitude of the image.
FDK	The optimization of CBCT image is achieved through a direct reconstruction using phase-sorted projections and the standard Feldkamp-Davis-Kress (FDK) algorithm.

medical reconstruction algorithms. In contrast, the FISTA-TV method focuses on reconstructing the original signal by minimizing a loss function with sparse prior information. This method is emblematic of traditional reconstruction algorithms and is renowned for its capacity to generate high-quality images from incomplete, noisy, or blurred source images. Meanwhile, MaMO shares similarities with our approach in terms of neural network utilization and motion compensation techniques but does not integrate prior knowledge, boasting a more streamlined structural design. Each of these three approaches is representative in its own right, collectively providing a comprehensive evaluation framework that underscores the strengths of our proposed approach.

In the second stage, we conducted a comparative evaluation of various networks for sparse image reconstruction to highlight the superior performance of the modified network architecture. The networks under consideration encompassed the conventional U-net, the Multi-Scale Generator Adversarial Network (MSG-GAN), and the Prior-Net. Detailed comparison specifics are documented in Table 3.

3.2.2. Non-equiangular Projection Comparison

During free breathing, respiratory curve fluctuations are irregular, leading to an uneven distribution of projection numbers in each respiratory phase. This variability may result in substantial discrepancies in sparse-view reconstruction quality, potentially degrading the 4D CBCT images. Notwithstanding, numerous methods still divide the projection evenly based on a uniform angle, a scenario that seldom occurs in treatment processes. To ascertain if the proposed method can address this issue, we compare the results under both ideal and free breath-

Table 3. Description of different networks for comparison

Network	Description
PN-Net	A network modeled on the U-Net structure includes an additional encoder to fuse prior images and a self-attention module is integrated into the decoder. Dense blocks replace traditional convolution enhancement for feature extraction.
Prior-Net	A network, based on the U-net model, includes a prior feature fusion module. This module stacks prior features with sparse view image features across channels to fuse the information effectively.
MSD-GAN	A network built on the Generative Adversarial Network (GAN) model comprises one generator and three discriminators. The generated image is sent to three discriminators for comparison at different scales, and the evaluation results from the three are combined to update the generator.
U-Net	A network with an encoder-decoder structure includes four up-sampling and down-sampling operations. The generated feature graph is amalgamated via a skip connection.

ing conditions.

3.2.3. Evaluation Metrics

To furnish an accurate assessment of the results, we employ a range of criteria to gauge the regions of interest (ROI), including L1-Error, SSIM, Root Mean Square Error (RMSE), and Peak Signal-to-Noise Ratio (PSNR). The formulations for RMSE and PSNR are as follows:

$$\mathrm{RMSE} = \sqrt{\frac{1}{MN}\sum_{i=1}^{M}\sum_{j=1}^{N}(x_{ij} - \hat{x}_{ij})^2}, \tag{7}$$

$$\mathrm{PSNR} = 20\log_{10}\frac{\max(x)}{\mathrm{RMSE}}, \tag{8}$$

where M and N represent the size of the image, i and j symbolize the coordinate position, x and $\hat{x}$ represent real voxel and predicted voxel. The PSNR indicates the ratio of the peak signal energy to the average noise energy, and it is inversely proportional to the RMSE.

4. RESULTS

4.1. Simulation Data

As depicted in Fig. 6 and Fig. 7, we evaluate several 4D CBCT reconstruction methodologies utilizing images reconstructed from distinct respiratory phases. Transverse displays are presented in Fig. 6, while coronal views are demonstrated in Fig. 7. The comparison includes several iterative and deep-learning-based algorithms, the abbreviations of which are listed in Table 2. Compared to FDK reconstruction, all methods exhibit substantial enhancements in image quality. The iterative method employing FISTA-TV reconstructs only the general outline, but it exhibits partially inaccurate boundaries. To alleviate the severe streaking artifacts, strong regularization was enacted, which resulted in over-smoothing structures. The CycN-Net algorithm's results present a superior visualization of the overall structure, but the boundaries remain blurred due to the severely unequal angular distribution. The MaMO algorithm exhibits an acceptable performance with a clear organ structure. However, it diverges from the ground truth in detail and retains more streak artifacts owing to the lack of prior image information. Compared to other methods, the proposed approach displays notable improvements, showcasing sharp edges, effectively suppressed streak artifacts, and high resolution in the reconstructed images. As evident from Table 4, the proposed method outshines the others, achieving the lowest CT RMSE value and the highest global SSIM/PSNR. Overall, our approach excels in reconstructing both contours and organizational structure.

Table 4. Images quality metrics using different methods in five phases(RMSE:HU)

	Metric	FDK	FISTA-TV	Cyc-Net	MaMO	**PNMC**
Phase 1	SSIM	0.1178	0.6101	0.6490	0.7928	**0.9150**
	PSNR	14.71	23.13	24.72	28.36	**35.49**
	RMSE	349.14	139.53	105.88	76.43	**31.90**
Phase 3	SSIM	0.0861	0.6114	0.6776	0.8087	**0.9148**
	PSNR	12.22	24.20	25.21	28.58	**35.42**
	RMSE	475.06	122.21	102.71	73.90	**32.88**
Phase 5	SSIM	0.1381	0.6297	0.6811	0.8012	**0.8994**
	PSNR	16.11	25.06	25.56	29.11	**34.71**
	RMSE	302.13	113.45	90.34	71.06	**35.47**
Phase 7	SSIM	0.1372	0.6156	0.6790	0.8179	**0.9004**
	PSNR	15.06	23.80	24.07	29.09	**35.05**
	RMSE	338.32	129.20	88.05	70.20	**33.82**
Phase 9	SSIM	0.1430	0.6467	0.6798	0.7897	**0.8897**
	PSNR	14.06	24.45	25.82	28.29	**33.78**
	RMSE	376.50	118.72	89.57	76.26	**41.63**

4.2. Clinical Data

The 3D-FDK reconstruction images, derived from full projections, serve as a reference for comparative analysis. The FISTA-TV yields a visually blurred image with significant smoothing, delivering scant useful information. While MaMO enhances image resolution and diminishes textural structure in spinal sections, the over-smoothed areas and streak artifacts obscure bone and organ structures. In contrast, the proposed method accurately restores distinct margins and intricate tissue details with superior resolution, outperforming reference images, thereby underscoring its therapeutic potential. As depicted in Fig. 9, we compare 1D profiles according to the results in Fig. 8 (a). Remarkably, the proposed method's line shape nearly mirrors that of the reference image.

4.3. Network Comparison

To assess the optimization ability of the PN-Net, we compare it with three types of networks: U-Net, Prior-Net, and MSD-GAN using the simulated dataset. As seen in Fig. 10, the U-Net results reveal the most considerable deviation from reality, displaying a distorted overall form and a disorganized internal structure. The MSD-GAN, employing three discriminators to discern feature information, performs marginally better than the U-Net, exhibiting more uniform voxel values and a roughly complete shape.

When paired with prior knowledge, the Prior-Net demonstrates a significant improvement in result quality, yielding enhanced outcomes with clear boundaries and balanced light-dark

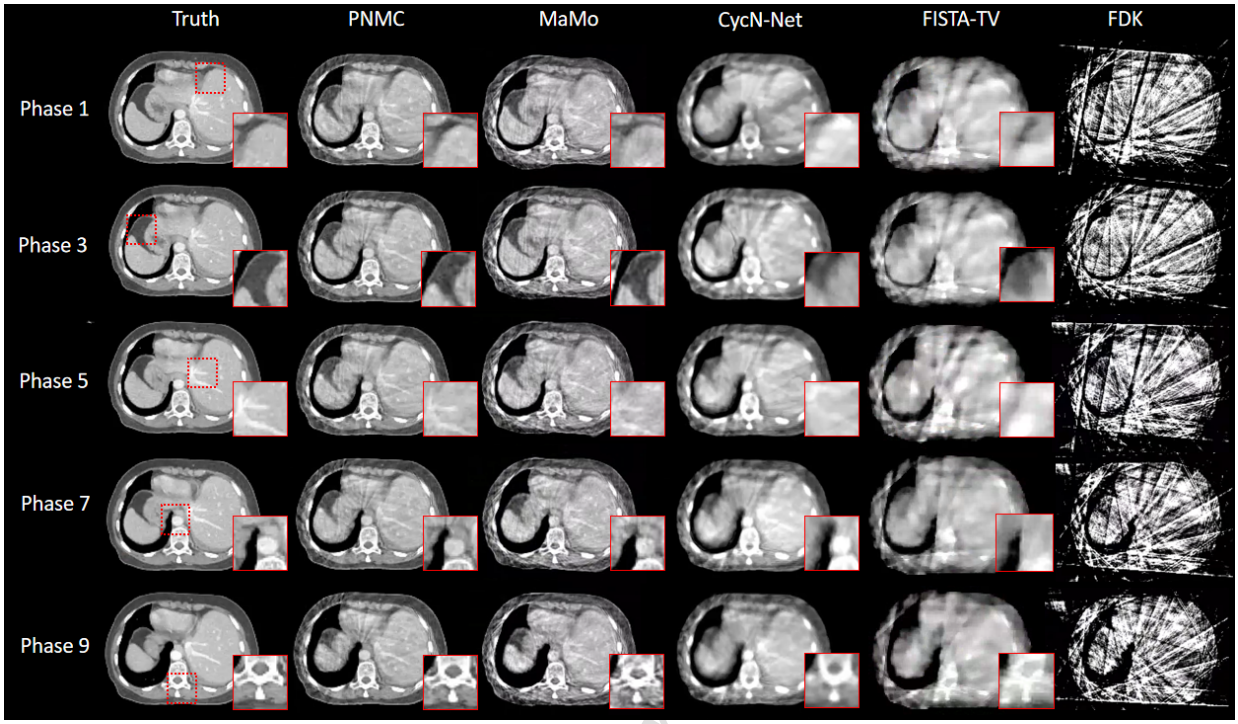

Fig. 6. Test results of the simulated dataset in the transverse view. Sequentially, PNMC, MaMO, FISTA-TV, CycN-Net and FDK algorithms are compared to reconstruct images in five phases. The fluctuation of breathing is mainly reflected on the left side of the reconstructed image. Regions of interest like liver, artery, and muscle regions are indicated by red dashed boxes. The display window is [-200,300] HU for all.

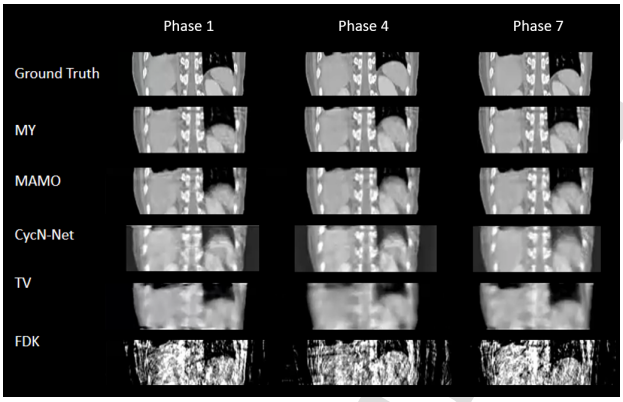

Fig. 7. Test results for the simulated dataset in the coronal view, Sequentially, PNMC, MaMO, CycN-Net FISTA-TV and FDK algorithms are compared to reconstruct images in three phases. The display window is [-200,300] HU for all.

variations. Furthermore, the PN-Net, with its more complex downsampling path and the integration of an attention mechanism module, offers substantial advantages over other networks in terms of detail and precision in image recovery, particularly in liver, arteries, and muscle areas. As illustrated in Table 5, we calculate SSIM, PSNR, and RMSE for the network-optimized images' areas of interest presented in Fig. 10. The proposed method exhibits the highest SSIM/PSNR and the lowest RMSE among the three regions.

Table 5. IMAGE QUALITY METRICS USING DIFFERENT NETWORKS IN THE FIVE PHASES(RMSE:HU)

	Metric	U-Net	MSD-GAN	Prior-Net	**PN-Net**
ROI1	SSIM	0.4351	0.4494	0.8025	**0.8237**
	PSNR	26.70	27.72	30.67	**35.97**
	RMSE	165.040	151.37	62.72	**50.66**
ROI2	SSIM	0.4299	0.5216	0.8419	**0.8737**
	PSNR	19.16	20.44	30.86	**31.13**
	RMSE	183.99	163.96	44.17	**39.56**
ROI3	SSIM	0.4055	0.5076	0.8812	**0.9063**
	PSNR	16.67	18.59	30.99	**32.41**
	RMSE	315.49	188.05	32.96	**32.46**

4.4. SPARE Challenge

The SPARE challenge datasets comprise both simulated and clinical datasets, with the Varian clinical dataset selected for testing. Importantly, we refrain from retraining the network using the challenge dataset, ensuring the training and testing data of the network exhibit different distributions. Consequently, we can test the network's generalization. Fig. 11 depicts the results. After PN-Net enhancement, artifacts disappear, and organ contours are restored, with missing structures compensated post-motion compensation. From three different slices, the proposed method's results reveal clear edges and precise tissue details. Compared to our method, FISTA-TV continues to produce over-smoothed regions, while MaMO reduces image resolution and slightly blurs edges. Notably, all methods successfully reconstruct tumor motion in this dataset, with the

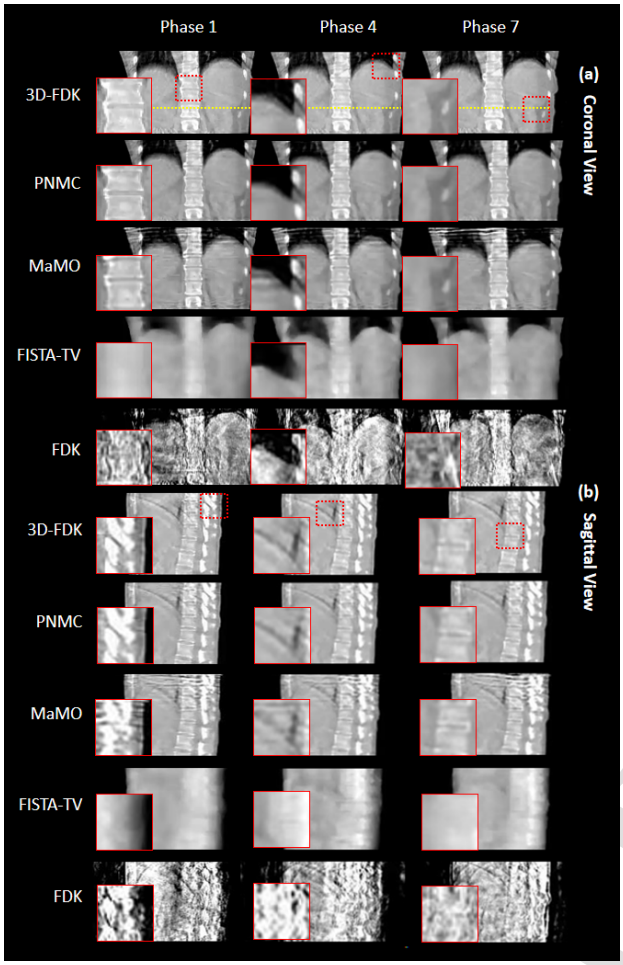

Fig. 8. Reconstruction results using different methods in the selected phases of patient studies: (a) coronal view, (b) sagittal view. Comparison details focused on bones and marginal tissue are shown in the red box. The yellow lines drawn at one-third of the ordinate are used to compare in the following sections. The window is shown as [-350,600] HU for all.

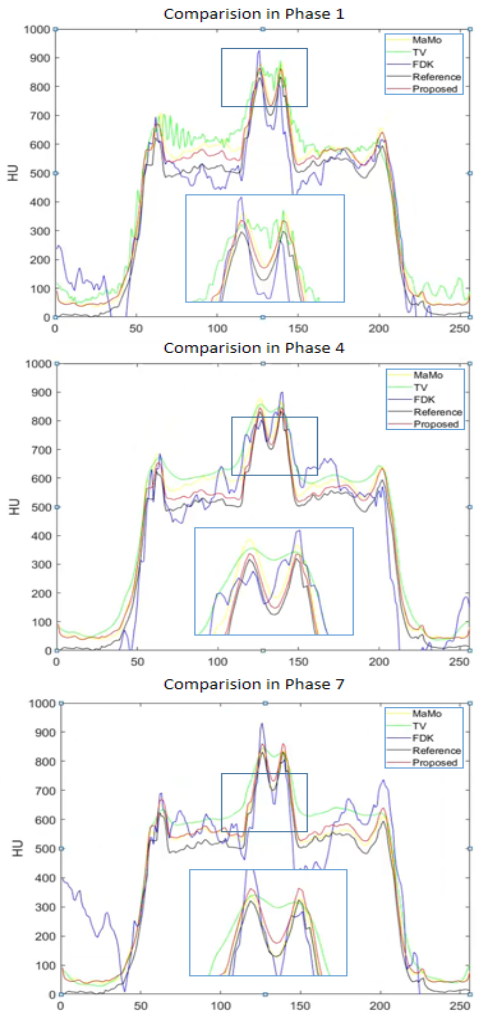

Fig. 9. The 1-D profiles of the CBCT images using different methods along the yellow line in Fig. 8. Part of the center areas are enlarged for a clearer comparison. It shows all curves are similar in general trend but quite different in the details. The 3D-FDK image curve is used as a reference, and the results that differ too much from it are regarded as inaccurate.

proposed method significantly improving the clarity of the tumor contour.

4.5. Non-equiangular Comparison

We reconstructed images using different projection distributions and selected the region greatly affected by respiration for comparison. The projection numbers for the non-equiangular and equiangular distributions are 52 and 66, respectively, out of a total of 660 projections. As shown in the Fig. 12, the sparse-view reconstruction images from the non-equiangular test exhibit more cluttered artifacts with noticeable white stripes. The PN-Net significantly reduces the gap, with image (a1) missing only partial details due to the occlusion of the white artifact compared to image (a2). Following motion-compensated iterative reconstruction, the discrepancy between the two results decreases further, with both achieving high quality. These results demonstrate that our proposed method can effectively mitigate the effects of irregular breathing.

5. DISCUSSION

In this study, we enhance 4D CBCT reconstruction, widely used to minimize blurring from respiratory motion, by integrating deep learning models, prior knowledge, and motion compensation within a single routine CBCT scan. Unlike traditional methods that rely on previous reconstructions, our approach derives prior knowledge from all projections in a single scan, providing essential structural information for the network's recovery process. We introduce PN-Net, designed to fuse multi-level features in the encoder, using network-optimized images to obtain accurate deformation fields for motion-compensated iterative reconstruction. This method not only promises high-quality 4D CBCT image reconstruction, as demonstrated by evaluations with simulated and clinical datasets but also enhances image quality in medical applications. It relies solely on a single 3D CBCT scan, reducing the need for additional instrumentation and patient radiation exposure. Our simulations

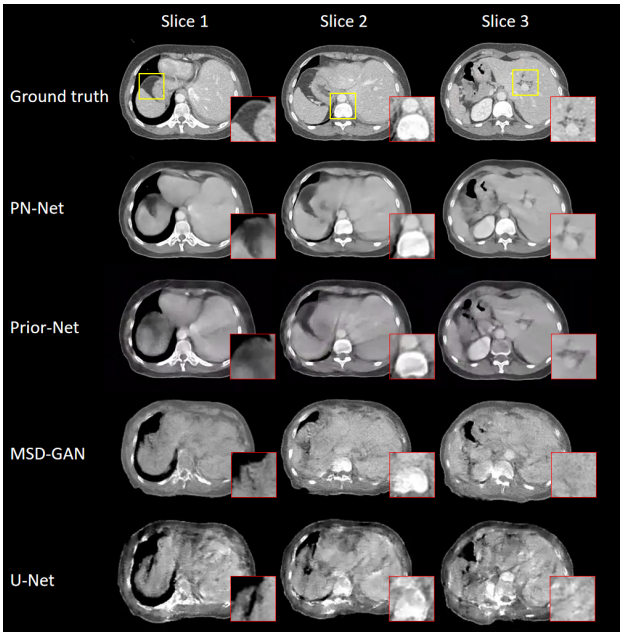

Fig. 10. The network-optimized images from sparse-view reconstruction using different networks. From top to bottom are the results enhanced by PN-Net, Prior-Net, MSD-GAN and U-Net. Regions of interest like liver, artery, and muscle regions are indicated by red dashed boxes. The display window is [-200,300] HU for all.

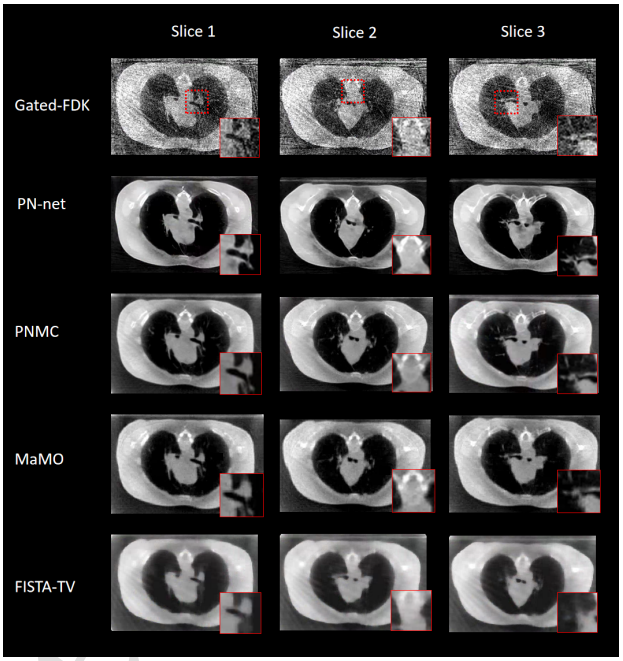

Fig. 11. Reconstruction results on SPARE challenge dataset using different methods. We compare the results of PN-Net, PMNC, MaMO, and FISTA-TV in three slices. All of the methods have significantly restored the image quality, but the other methods have lost additional details compared with the proposed method. The display window is [-200,300] HU for all.

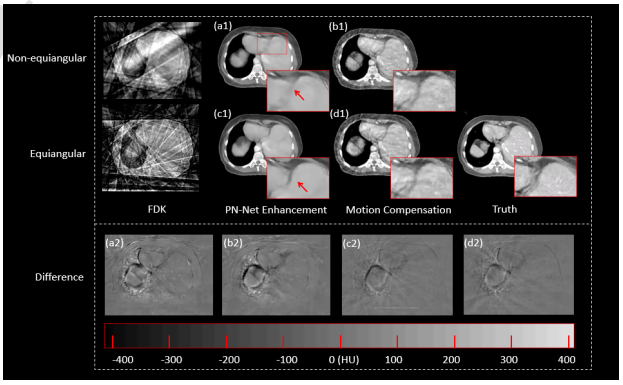

Fig. 12. Test results on the projection of equiangular and non-equiangular distribution. images (a2)-(d2) show the difference between the true value and images (a1)-(d1). The color change at the bottom is used to measure the difference, and approaching black or white indicates a large difference. The organs affected by respiratory movements in the figure have significant differences.

show stable reconstruction results under free-breathing conditions, despite uneven projection distributions, highlighting the method's robustness. This approach is set to improve diagnostic and treatment methods, raising the standard of patient care. However, further research and comprehensive clinical validation are necessary to fully assess its implications and practical applications.

Deep learning has markedly advanced medical imaging due to its robust non-linear capability. The network can decipher complex and intricate patterns, producing superior-quality results compared to traditional methods. However, disparities in distributions between training and test data can result in significant degradation of network results. This issue is particularly prominent in 4D CBCT reconstruction, where simulated data used for network training greatly differs from clinical data. Consequently, it's critical to maintain the network's robustness to data with varying distributions. As demonstrated in our experimental results, integrating prior knowledge effectively mitigates this issue. We directly applied the proposed method to the SPARE dataset, yielding exceptionally promising results.

In the proposed method, a certain level of robustness is observed regarding distribution disparities between the training and test datasets. Nonetheless, when the test dataset markedly deviates from the training dataset, the outcomes become unpredictable and challenging to manage. Furthermore, the chosen registration method imposes constraints on the registration of high-resolution images due to the substantial resource and time demands it entails. Based on this observation, there is potential for further refinement in this workflow. In future research, we aspire to further enhance the proposed method's generalization and efficiency. Unsupervised learning may be a promising approach, with techniques such as patch-based multi-layer unsupervised learning previously employed to enhance the spatial resolution and microstructure of pulmonary arteries [2]. Another strategy involves the use of transfer learning. We could refine models trained on large-scale datasets to reduce training costs and boost generalization performance [15]. Furthermore, the fusion of the metaverse and medical diagnostics represents a highly promising direction for future development[14]. It offers the opportunity to explore the integration of digital twin

technologies[17, 18], harnessing the full potential of modeling for simulating 4D CBCT reconstructions.

The motion-compensated method represents another facet ripe for optimization. Improving the registration method's performance is crucial since the accuracy of DVF directly impacts the reconstruction quality. We could consider using deep learning-based deformable registration techniques [9]. If the motion-compensated reconstruction process can concurrently estimate motion and provide high-quality reconstruction, it will undoubtedly yield superior results. We could refine the DVFs by utilizing the image with motion compensation during both the forward and backward projection processes. This could enhance the image quality of 4D CBCT and the precision of motion estimation.

6. CONCLUSION

We propose an efficient method for reconstructing 4D CBCT images from a single routine scan to tackle current challenges in clinical applications. The method comprises three stages. First, the respiratory curve derived from the original projection segments the projections into ten respiratory phases. Next, PN-Net is employed to enhance the sparse-view reconstruction image. Owing to its ability to fully extract the structural information from the prior image, PN-Net surpasses other networks in edge and organ recovery capabilities. Finally, iterative reconstruction is used to compensate for motion information in each phase, enhancing the reconstruction's accuracy. The proposed method exhibits superior performance on both simulated and clinical data, significantly improving image quality without altering the scanning setup or duration.

Acknowledgments

This work is supported in part by Natural Science Foundation of China (General Grant 82372041, Scientific Research Instrument Development Project 61827808), Beijing Natural Science Foundation (Z210008), National Key Research and Development Program of China under Grant 2022YFE0116700.

References

[1] Aganj, I., Iglesias, J.E., Reuter, M., Sabuncu, M.R., Fischl, B., 2017. Mid-space-independent deformable image registration. Neuroimage 152, 158–170.

[2] Barlow, H.B., 1989. Unsupervised learning. Neural computation 1, 295–311.

[3] Brehm, M., Paysan, P., Oelhafen, M., Kachelrieß, M., 2013. Artifact-resistant motion estimation with a patient-specific artifact model for motion-compensated cone-beam ct. Medical physics 40, 101913.

[4] Chao, M., Wei, J., Li, T., Yuan, Y., Rosenzweig, K.E., Lo, Y.C., 2016. Robust breathing signal extraction from cone beam ct projections based on adaptive and global optimization techniques. Physics in Medicine & Biology 61, 3109.

[5] Chee, G., O'Connell, D., Yang, Y., Singhrao, K., Low, D., Lewis, J., 2019. Mcsart: an iterative model-based, motion-compensated sart algorithm for cbct reconstruction. Physics in Medicine & Biology 64, 095013.

[6] Crawford, C.R., King, K.F., Ritchie, C.J., Godwin, J.D., 1996. Respiratory compensation in projection imaging using a magnification and displacement model. IEEE transactions on medical imaging 15, 327–332.

[7] Dietrich, L., Jetter, S., Tücking, T., Nill, S., Oelfke, U., 2006. Linac-integrated 4d cone beam ct: first experimental results. Physics in Medicine & Biology 51, 2939.

[8] Feldkamp, L.A., Davis, L.C., Kress, J.W., 1984. Practical cone-beam algorithm. Josa a 1, 612–619.

[9] Fu, Y., Lei, Y., Wang, T., Curran, W.J., Liu, T., Yang, X., 2020. Deep learning in medical image registration: a review. Physics in Medicine & Biology 65, 20TR01.

[10] Goyal, S., Kataria, T., 2014. Image guidance in radiation therapy: techniques and applications. Radiology research and practice 2014.

[11] Hauser, J., Zeligman, A., Averbuch, A., Zheludev, V.A., Nathan, M., 2020. Dd-net: spectral imaging from a monochromatic dispersed and diffused snapshot. Applied Optics 59, 11196–11208.

[12] Huang, G., Liu, Z., Van Der Maaten, L., Weinberger, K.Q., 2017. Densely connected convolutional networks, in: Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 4700–4708.

[13] Jaffray, D.A., Siewerdsen, J.H., Wong, J.W., Martinez, A.A., 2002. Flat-panel cone-beam computed tomography for image-guided radiation therapy. International Journal of Radiation Oncology* Biology* Physics 53, 1337–1349.

[14] Jamshidi, M., Dehghaniyan Serej, A., Jamshidi, A., Moztarzadeh, O., 2023a. The meta-metaverse: ideation and future directions. Future Internet 15, 252.

[15] Jamshidi, M.B., Sargolzaei, S., Foorginezhad, S., Moztarzadeh, O., 2023b. Metaverse and microorganism digital twins: A deep transfer learning approach. Applied Soft Computing 147, 110798.

[16] Kak, A.C., Slaney, M., 2001. Principles of computerized tomographic imaging. SIAM.

[17] Moztarzadeh, O., Jamshidi, M., Sargolzaei, S., Jamshidi, A., Baghalipour, N., Malekzadeh Moghani, M., Hauer, L., 2023a. Metaverse and healthcare: Machine learning-enabled digital twins of cancer. Bioengineering 10, 455.

[18] Moztarzadeh, O., Jamshidi, M., Sargolzaei, S., Keikhaee, F., Jamshidi, A., Shadroo, S., Hauer, L., 2023b. Metaverse and medical diagnosis: A blockchain-based digital twinning approach based on mobilenetv2 algorithm for cervical vertebral maturation. Diagnostics 13, 1485.

[19] Ritschl, L., Sawall, S., Knaup, M., Hess, A., Kachelrieß, M., 2012. Iterative 4d cardiac micro-ct image reconstruction using an adaptive spatio-temporal sparsity prior. Physics in Medicine and Biology 57, 1517–1525. URL: http://dx.doi.org/10.1088/0031-9155/57/6/1517, doi:10.1088/0031-9155/57/6/1517.

[20] Shieh, C.C., Gonzalez, Y., Li, B., Jia, X., Rit, S., Mory, C., Riblett, M., Hugo, G., Zhang, Y., Jiang, Z., et al., 2019. Spare: Sparse-view reconstruction challenge for 4d cone-beam ct from a 1-min scan. Medical physics 46, 3799–3811.

[21] Tian, Z., Jia, X., Dong, B., Lou, Y., Jiang, S.B., 2011. Low-dose 4dct reconstruction via temporal nonlocal means. Medical physics 38, 1359–1365.

[22] Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, Ł., Polosukhin, I., 2017. Attention is all you need. Advances in neural information processing systems 30.

[23] Wang, J., Gu, X., 2013. Simultaneous motion estimation and image reconstruction (smeir) for 4d cone-beam ct. Medical physics 40, 101912.

[24] Wang, J., Li, T., Lu, H., Liang, Z., 2006. Penalized weighted least-squares approach to sinogram noise reduction and image reconstruction for low-dose x-ray computed tomography. IEEE transactions on medical imaging 25, 1272–1283.

[25] Wang, Z., Zou, N., Shen, D., Ji, S., 2020. Non-local u-nets for biomedical image segmentation, in: Proceedings of the AAAI conference on artificial intelligence, pp. 6315–6322.

[26] Yang, P., Ge, X., Tsui, T., Liang, X., Xie, Y., Hu, Z., Niu, T., 2023. Four-dimensional cone beam ct imaging using a single routine scan via deep learning. IEEE Transactions on Medical Imaging 42, 1495–1508. doi:10.1109/TMI.2022.3231461.

[27] Zhang, Z., Liu, J., Yang, D., Kamilov, U.S., Hugo, G.D., 2023. Deep learning-based motion compensation for four-dimensional cone-beam computed tomography (4d-cbct) reconstruction. Medical Physics 50, 808–820. URL: http://dx.doi.org/10.1002/mp.16103, doi:10.1002/mp.16103.

[28] Zhi, S., Kachelrieß, M., Mou, X., 2020. High-quality initial image-guided 4d cbct reconstruction. Medical physics 47, 2099–2115.

[29] Zhi, S., Kachelrieß, M., Pan, F., Mou, X., 2021. Cycn-net: A convolutional neural network specialized for 4d cbct images refinement. IEEE Transactions on Medical Imaging 40, 3054–3064.

We propose an efficient method for reconstructing 4D CBCT images from a single routine scan to tackle current challenges in clinical applications. The method comprises three stages. First, the respiratory curve derived from the original projection segments the projections into ten respiratory phases. Next, PN-Net is employed to enhance the sparse-view reconstruction image. Owing to its ability to fully extract the structural information from the prior image, PN-Net surpasses other networks in edge and organ recovery capabilities. Finally, iterative reconstruction is used to compensate for motion information in each phase, enhancing the reconstruction's accuracy.

All authors disclosed no relevant relationships.