



First-order primal–dual algorithm for sparse-view neutron computed tomography-based three-dimensional image reconstruction

Yang Liu^{1,2} · Teng-Fei Zhu¹ · Zhi Luo¹ · Xiao-Ping Ouyang^{1,3}

Received: 14 December 2022 / Revised: 9 April 2023 / Accepted: 18 April 2023 / Published online: 21 August 2023

© The Author(s), under exclusive licence to China Science Publishing & Media Ltd. (Science Press), Shanghai Institute of Applied Physics, the Chinese Academy of Sciences, Chinese Nuclear Society 2023

Abstract

Neutron computed tomography (NCT) is widely used as a noninvasive measurement technique in nuclear engineering, thermal hydraulics, and cultural heritage. The neutron source intensity of NCT is usually low and the scan time is long, resulting in a projection image containing severe noise. To reduce the scanning time and increase the image reconstruction quality, an effective reconstruction algorithm must be selected. In CT image reconstruction, the reconstruction algorithms can be divided into three categories: analytical algorithms, iterative algorithms, and deep learning. Because the analytical algorithm requires complete projection data, it is not suitable for reconstruction in harsh environments, such as strong radiation, high temperature, and high pressure. Deep learning requires large amounts of data and complex models, which cannot be easily deployed, as well as has a high computational complexity and poor interpretability. Therefore, this paper proposes the OS-SART-PDTV iterative algorithm, which uses the ordered subset simultaneous algebraic reconstruction technique (OS-SART) algorithm to reconstruct the image and the first-order primal–dual algorithm to solve the total variation (PDTV), for sparse-view NCT three-dimensional reconstruction. The novel algorithm was compared with other algorithms (FBP, OS-SART-TV, OS-SART-AwTV, and OS-SART-FGPTV) by simulating the experimental data and actual neutron projection experiments. The reconstruction results demonstrate that the proposed algorithm outperforms the FBP, OS-SART-TV, OS-SART-AwTV, and OS-SART-FGPTV algorithms in terms of preserving edge structure, denoising, and suppressing artifacts.

Keywords NCT · First-order primal–dual algorithm · OS-SART · Total variation · Sparse-view

This work was supported by the National Key Research and Development Program of China (No. 2022YFB1902700), the Joint Fund of Ministry of Education for Equipment Pre-research (No. 8091B042203), the National Natural Science Foundation of China (No. 11875129), the Fund of the State Key Laboratory of Intense Pulsed Radiation Simulation and Effect (No. SKLIPR1810), the Fund of Innovation Center of Radiation Application (No. KFZC2020020402), the Fund of the State Key Laboratory of Nuclear Physics and Technology, Peking University (No. NPT2023KFY06), the Joint Innovation Fund of China National Uranium Co., Ltd., State Key Laboratory of Nuclear Resources and Environment, East China University of Technology (No. 2022NRE-LH-02), and the Fundamental Research Funds for the Central Universities (No. 2023JG001).

✉ Yang Liu
yliu@ncepu.edu.cn

¹ School of Nuclear Science and Engineering, North China Electric Power University, Beijing 102206, China

² State Key Laboratory of Nuclear Resources and Environment, East China University of Technology, Nanchang 330013, China

³ State Key Laboratory of Intense Pulsed Radiation Simulation and Effect, Xi'an 710024, China

1 Introduction

Neutron tomography is a noninvasive measurement method that differs from X-ray imaging, and muon tomography has been applied in various fields with a high accuracy and reliability [1, 2]. The strong transmission of neutrons in metallic elements makes neutron tomography extremely useful for examining hydrogen in heavy elements under harsh experimental conditions such as strong radiation, high-temperature, and high-pressure two-phase flow systems. In addition, neutrons are used in boron neutron therapy owing to their ability to detect and reconstruct the three-dimensional (3D) distribution of boron concentration [3, 4]. Neutron tomography systems consist of two parts: a neutron photographic system and a computed tomography (CT) system capable of measuring and displaying the 3D void rate in the boiling flow in a heated rod beam. They have also been applied to simulate the cores of advanced nuclear [5–8]. Asano and Takenaka [9] used neutron tomography to determine

the void-rate distribution in two-phase air and water flows. Kureta et al. visualized the toroidal streaming in a heated fuel rod bundle using 3D neutron tomography [10]. Trem sin et al. used neutron imaging to assess the attachment, deposition, and structural integrity of fuel core blocks [11]. Andersson et al. developed a portable fast NCT to obtain the void rate distribution in fuel rod bundles in a boiling water reactor [12]. Although portable NCT systems can perform measurements at any altitude, their neutron generators exhibit extremely low neutron yields. Therefore, promoting the rapid development of neutron tomography for improving the efficiency and safety of nuclear fuels is an important research direction.

The most crucial part of a neutron tomography system is the 3D reconstruction algorithm, which is directly related to the image quality and measurement reliability. Because of the long scanning time of neutron CT (ranging from 30 min to several hours) and the complexity of practical application environments, it is difficult to obtain complete projection data. Consequently, reconstruction algorithms that can efficiently utilize sparse-view projection data are required to obtain accurate measurement results. Furthermore, the reconstruction of sparse-view projection data is helpful for investigations in harsh environments such as strong radiation. Currently, algorithms are used for the reconstruction of sparse views and highly under-sampled projection data. Therefore, this study was focused on a 3D reconstruction algorithm for sparse-view NCT.

Sparse-view reconstruction is a highly complex mathematical problem and is a significant research direction in the image reconstruction field. Because prior information can be embedded in the iterative reconstruction algorithm, it is considered the sparse-view reconstruction algorithm most likely to replace the FBP algorithm. With the development of deep learning, it has been applied to the field of tomographic reconstruction. Many researchers have worked on sparse-view tomography and achieved significant results. As proposed in [13], the DEAR model, which adds a priori information and data in the image domain to the compressed sensing-based variation model, can not only eliminate the artifacts of the reconstructed image but also improve the quality of the reconstructed image. A multiple adversarial learning angiography image reconstruction framework has been proposed in the literature [14], which solves the challenge of low-intensity aortic reconstruction by introducing dual-correlation constrained adversarial learning; its application in clinical data shows its feasibility and effectiveness. However, because of practical limitations, this study focused on neutron CT 3D reconstruction, and it was not possible to obtain a large number of neutron projection images for supervised training. Therefore, this study focuses on an iterative algorithm for neutron CT 3D reconstruction. Currently, mainstream iterative reconstruction algorithms include ART [15], SIRT [16], and SART [17].

Although all these iterative algorithms can reconstruct high-quality images, when the projection data are extremely sparse and no a priori information is introduced, the reconstructed images show significant artifacts. Donoho [18] proposed the compressive sensing (CS) theory, which has promoted the rapid development of sparse-view image reconstruction. According to the mathematical implications of the CS theory, an image may be reconstructed exactly if it is sparse or can be sparsely represented by a spatial transformation. The evolution of the CS theory has facilitated the application of total variational algorithms in sparse-view image reconstruction. Rudin proposed the ROF denoising model, which reduces noise while preserving the image edge and detailed structural information using the total variance as a constraint for image denoising [19]. Several TV-based techniques and variations to improve the performance of the algorithm have been put forward, including directional TV [20], weighted TV [21], edge-preserving TV [22], and weight TV [23]. The total variation model is a typical regularization model. Total variance minimization is particularly important in CT imaging because of its ability to resolve acute discontinuities. This is critical for many imaging problems, because the sample edges contain critical structural information about the object. However, because of the non-smooth character of the total variance, it is difficult to achieve total variance regularization minimization. Furthermore, one drawback of these models is that they assume that the image is piecewise constant, which destroys significant information in the image. The total-variational image-denoising problem is a class of convex optimization problems. To solve convex optimization problems, the second-order method exhibits good convergence and requires only a small number of iterations. However, each iteration of the second-order method is extremely complicated, making it difficult to apply to large-scale problems. In contrast, the first-order method involves only function values and gradient information, and the process of each iteration is relatively simple. In the case of gradient magnitude sparsity, the total variance of the intermediate images is minimized, resulting in high-quality reconstructed images. Therefore, this study employs a first-order method to address the total-variance image noise reduction problem, and proposes an algorithm applicable to the 3D reconstruction of NCT in sparse views.

The goal of this study was to find an efficient 3D reconstruction algorithm for sparse-view NCT. To address artifacts and noise in sparse-view projection-reconstructed images, we propose the ordered subset simultaneous algebraic reconstruction technique (OS-SART-PDTV) algorithm, which uses the ordered subset simultaneous algebraic reconstruction technique (OS-SART) to reconstruct the image and the first-order primal–dual algorithm to solve the total variation (PDTV). This novel algorithm achieves a blurred version of a piecewise constant object through phenomenological modeling, and reconstructs high-quality 3D

images by minimizing the total variance of the intermediate images through gradient-amplitude sparsity, which allows the reconstructed image interiors to change quickly and smoothly. Next, we present the proposed method, including the proposed algorithm and NCT fundamentals, in Sect. 2. In Sect. 3, the simulation experiments and a real neutron projection experiment are discussed. Section 4 discusses the research process and presents conclusions. Finally, a short conclusion is presented in Sect. 5.

2 Principles and algorithms Of NCT

2.1 NCT system

This section describes the fundamentals of CT image reconstruction. In general, CT image reconstruction generates a 3D volume from a 2D collection of projected images using mathematical algorithms. This requires the development of a measurement model using mathematical methods to relate the measured data to the desired physical properties. Image reconstruction consists of two processes: forward and back projection. Forward projection enables the calculation of model measurements corresponding to the actual measurements that match the physical properties. Model measurements corresponding to the actual measurements were calculated using forward projection. Forward projection refers to the mapping of an objective onto the projection domain. Back projection is the opposite of forward projection, which determines the physical properties of a measurement. Therefore, these two operations largely determine

the image reconstruction precision. The change in forward projection depends on the radiation beam geometry. Based on the selected system, source of radiation, and available data acquisition systems, the appropriate geometry of the beam was selected so that reliable measurements could be obtained. The NCT system primarily includes a collimator and shield, rotating sample stage, and neutron imaging system. The neutron imaging system included a 2D neutron image detector, conversion screen, and 3D reconstruction software. The algorithm used for 3D reconstruction directly determines the imaging quality and measurement reliability of the system, and is the core of the system. A schematic of the NCT scanning system is shown in Fig. 1.

In NCT systems, the mass of the neutron source and noise generated by the electronics can cause the projection data to contain noise. Therefore, Poisson noise and Gaussian noise are added to the numerical projection data reconstruction to validate the denoising ability of the algorithm while testing its reliability [24]. Therefore, these two noise components must be considered during NCT. The equation is as follows:

$$\hat{I} = \text{Poisson}(I) + \text{Gaussian}(m_{ic} + \delta_{ic}^2), \quad (1)$$

where $\hat{I}$ denotes noisy transmission data, I represents the mean photon number, m_{ic} indicates the average value of the electronic noise, and δ_{ic}^2 is the variance of the electronic noise. System calibration typically has a value from m_{ic} to zero. The variance δ_{ic}^2 was estimated via dark current measurements [25].

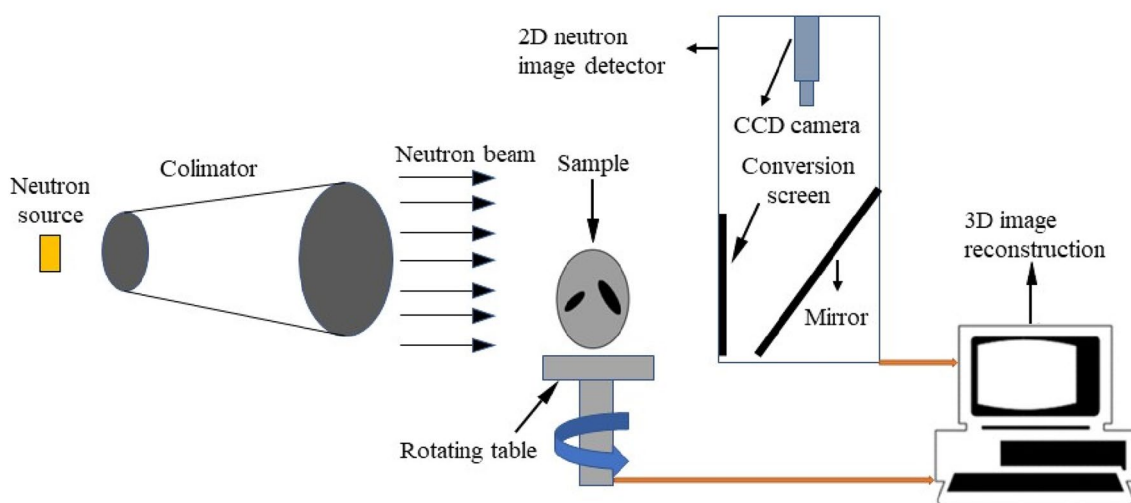


Fig. 1 (Color online) Schematic of the NCT system

2.2 Principle of CT

2.2.1 CT projection acquisition

In NCT, the geometry of the neutron beam is primarily parallel. Among the mathematical concepts, the parallel geometric beam is the simplest. An object can be represented by a binary function $R(x, y)$, which represents the intensity function on the x – y plane. The projection represents the decay integral of the neutron beam as it passes through the object from the neutron source, which is considered to be the sum of the line integral and ray. The expression of the projection is given by the equation

$$h(s, \beta) = \int_L R(x, y) ds, \quad (2)$$

where L represents a neutron beam passing straight through an object $R(x, y)$, s represents the detector coordinates in the projection domain, and β denotes the counterclockwise rotation angle of the x -axis in the projection domain. According to the Delta functions in Eq. (2) can be rewritten as follows:

$$h(s, \beta) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} R(x, y) \delta(x \cos \beta + y \sin \beta - s) dx dy. \quad (3)$$

Equation (3) represents the Radon transform of the object $R(x, y)$, which transforms the space data into the frequency domain. The projection consisted of a group of integrals of the neutron attenuation coefficients of the sample measured at the detector for every given angle.

2.2.2 Central slice theorem

The Central slice theorem establishes a connection between the 2D projection of an object and its 1D projection in the Fourier domain. The 1D Fourier transform function $K(\sigma)$ of the projection function $h(s, \beta)$ of the density function $R(x, y)$ in a certain direction is the value of the 2D Fourier transform of the density function $R(x, y)$ on the x – y plane along the same direction on a straight line past the origin. Equation (4) represents the defined 1D Fourier variation of the objective function in polar coordinates.

$$K(\sigma, \beta) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} R(x, y) e^{-j2\pi\sigma(x \cos \beta + y \sin \beta)} dx dy \quad (4)$$

Equation (4) represents the 2D Fourier transform of the spatial frequency $\sigma_x = \sigma \cos \beta$, $\sigma_y = \sigma \sin \beta$ in the Fourier space. Therefore, the following equation was derived:

$$R(x, y) = \int_0^{2\pi} \int_0^{\infty} Q(\sigma_x, \sigma_y) e^{j2\pi\sigma(x \cos \beta + y \sin \beta)} d\sigma d\beta. \quad (5)$$

Several problems arise when the central slice theorem is used for the Fourier transform reconstruction. First, the theorem produces data in the Fourier space that are inconsistent with the Cartesian space. The data were interpolated into Cartesian coordinates. However, interpolation in the Fourier space before the anti-Fourier transform can have a significant influence on the reconstruction. The second problem is the difficulty in performing targeted reconstruction and identifying fine structures within small areas.

2.3 Algorithms

2.3.1 TV algorithm

In the NCT, the measurement time can be drastically reduced by reducing the number of projection views; however, this leads to sparse sampling. Based on the CS theory, a signal can be reconstructed exactly if it is sparse or can be represented sparsely. Under ideal conditions, NCT reconstruction can be converted to solve the problem of a linear system; thus, the discrete model of CS theory may be represented.

$$\mathbf{L}g = b, \quad (6)$$

where $\mathbf{L} = \{l_{ij}\}$ stands for system matrix b denotes the projection data obtained at each angle and g represents the image to be reconstructed. Mathematically, solving Eq. (6) usually converts it into a least square problem:

$$\arg \min_g \|\mathbf{L}g - b\|_2^2. \quad (7)$$

However, in sparse-view image reconstruction, the incompleteness of the projection data leads to reconstructed images containing noise and artifacts. Therefore, various regularization models have been proposed to obtain approximate solutions, which are defined by the following equations:

$$\arg \min_g \|\mathbf{L}g - b\|_2^2 + \lambda Z(g), \quad (8)$$

where $\arg \min_g \|\mathbf{L}g - b\|_2^2$ represents the data fidelity term, $\lambda Z(g)$ denotes the regular term, and λ is the regularization parameter used to balance the regular and data terms. The equation that defines TV as regularization is given in Eq. (9)

$$\|g\|_{TV} = \sum \sqrt{|\nabla_x g|^2 + |\nabla_y g|^2 + |\nabla_z g|^2}, \quad (9)$$

where g denotes the 3D image, $\nabla_i g$ represents the gradient along i direction, and $|\cdot|$ denotes the complex modulus.

2.3.2 First-order primal–dual algorithm

Owing to the segmental smoothness of the regularization term of the variational model, the total variational model can effectively preserve image edges during image recovery and achieve improved recovery results. Therefore, the total variational model, particularly the total variational regularization model proposed by Rudin, Osher, and Fatemi, has gained popularity. According to the ROF model, the discrete total variational model of the 3D images can be formulated as follows [26]:

$$\min_g \frac{1}{2} \|Hg - u\|_2^2 + \alpha CTV(g), \quad (10)$$

where $\|\cdot\|_2$ stands for the Euclidean parametrization, $\alpha > 0$ is the regularization argument, and $CTV(g)$ is the regular term of the discrete total variational model. To obtain a stable solution to the problem of minimization, ∇ , which denotes the difference algorithm on the space $\mathbb{R}^{3n^2}$, is defined as shown in Eq. (11) below:

$$(\nabla g)_{e,u,q} = \left((\nabla g)_{e,u,q}^h, (\nabla g)_{e,u,q}^v, (\nabla g)_{e,u,q}^w \right), \quad (11)$$

where

$$\begin{aligned} (\nabla g)_{e,u,q}^h &= \begin{cases} g_{e+1,u,q} - g_{e,u,q} & e < N \\ 0, & e = N \end{cases} \quad (e = 0, 1, 2, \dots, N) \\ (\nabla g)_{e,u,q}^v &= \begin{cases} g_{e,u+1,q} - g_{e,u,q} & u < M \\ 0, & u = M \end{cases} \quad (u = 0, 1, 2, \dots, M) \\ (\nabla g)_{e,u,q}^w &= \begin{cases} g_{e,u,q+1} - g_{e,u,q} & q < Z \\ 0, & q = Z \end{cases} \quad (q = 0, 1, 2, \dots, Z) \end{aligned}$$

$g_{e,u,q}$ denotes the $((N * M) * q + e + N * u)$ -th element of g , whose position in the image is (e, u, q) ; $(\nabla g)_{e,u,q}^h$, $(\nabla g)_{e,u,q}^v$, and $(\nabla g)_{e,u,q}^w$ are the different operators in the x , y , and z directions at a pixel $((N * M) * q + e + N * u)$. The expression of the regular terms of the total variational model is shown below.

$$\begin{aligned} CTV(g) &= \|\nabla g\|_1 \\ &= \sum_{0 \leq e \leq N, 0 \leq u \leq M, 0 \leq q \leq Z} \sqrt{\sum \left| (\nabla g)_{e,u,q}^h \right|^2 + \left| (\nabla g)_{e,u,q}^v \right|^2 + \left| (\nabla g)_{e,u,q}^w \right|^2} \end{aligned} \quad (12)$$

In this study, the above model was improved using the finite difference matrix $\mathbf{R}$ instead of the difference operator ∇ .

$$CTV(g) = \|\mathbf{R}g\|_1 \quad (13)$$

Thus, the expression of the model investigated in the present study can be obtained as follows:

$$\min_g \frac{1}{2} \|\mathbf{H}g - u\|_2^2 + \alpha \|\mathbf{R}g\|_1. \quad (14)$$

Although the total variational model efficiently preserves image edges, it is difficult to obtain the minimum value of the total variational model by solving Eq. (14) directly because of the non-smooth nature of the total variational model [27]. Because the first-order primal–dual algorithm is an efficient method for solving non-smooth convex optimizations of images, we employed it to solve Eq. (14).

Definition 1: Assume that the binary function $J(\chi, \nu)$ ($\chi \in N^n, \nu \in N^n$) is convex for χ , and concave for ν . If, for any $(\chi \in N^n, \nu \in N^n)$, there exists an (χ^*, ν^*) such that Eq. (15) holds, (χ^*, ν^*) is the saddle point of the function $J(\chi, \nu)$.

$$J(\chi^*, \nu) \leq J(\chi^*, \nu^*) \leq J(\chi, \nu^*) \quad (15)$$

Chambolle and Pock suggested a first-order primal–dual algorithm to solve the following model based on the properties of the saddle points [28]:

$$\min_{\chi \in N^n} \max_{\nu \in N^n} \mathcal{L}(\chi, \nu) \equiv o(\chi) + \langle K\chi, \nu \rangle - \vartheta(\nu), \quad (16)$$

where o and ϑ are lower semi-continuous convex functions, and K denotes the line operator. The first-order primal–dual iterations were as follows [29, 30].

$$\chi^{(d+1)} = \arg \min_{\chi \in N^n} o(\chi) + \langle K\chi, \nu \rangle - \frac{1}{2s} \left\| \chi - \chi^{(d)} \right\|_2^2, \quad (17)$$

$$\bar{\chi}^{(d+1)} = \chi^{(d+1)} + \theta (\chi^{(d+1)} - \chi^{(d)}), \quad (18)$$

$$\nu^{(d+1)} = \arg \min_{\nu \in N^n} \langle K\bar{\chi}^{(d+1)}, \nu \rangle - \vartheta(\nu) - \frac{1}{2t} \left\| \nu - \nu^{(d)} \right\|_2^2, \quad (19)$$

where the parameter $s, t > 0$ is the iteration steps for the primal and dual variables, separately; θ denotes the combinational parameter, which is used to ensure the algorithm convergence.

Definition 2: Assuming that function $o(g) : N^n \rightarrow N$, we call the formula in Eq. (20) as a conjugate function of the function $o(g)$ [31].

$$o^*(\xi) = \sup_{\chi \in N^n} \{ \xi^T g - o(g) \} \quad (20)$$

where ξ represents the dual vector of g , N denotes the real number space, N^n represents the k -dimensional real space, and T denotes the transpose.

According to Definition 2, the $\|\mathbf{R}g\|_1$ can be expressed as $\max_{\|\xi\|_1 \leq 1} \xi^T \mathbf{R}g$. Then, Eq. (14) can be converted into a mini-extreme problem [32] as follows:

$$\min_g \max_{\|\xi\|_\infty \leq 1} \frac{1}{2} \|\mathbf{H}g - u\|_2^2 + \alpha \xi^T \mathbf{R}g. \quad (21)$$

The iterative formula in Eq. (21) can be obtained by using a first-order primal–dual algorithm.

$$g^{(d+1)} = \arg \max_{g \in \mathbb{N}^{3n^2}} \frac{1}{2} \|\mathbf{H}g - u\|_2^2 + \alpha g^T \mathbf{R}^T \xi^{(d)} + \frac{1}{2s} \|g - g^{(d)}\|_2^2 \quad (22)$$

$$\bar{g}^{(d+1)} = g^{(d+1)} + \theta(g^{(d+1)} - g^{(d)}) \quad (23)$$

$$\xi^{(d+1)} = \arg \max_{\|\xi\|_\infty \leq 1} \alpha \xi^T \mathbf{R} \bar{g}^{(d+1)} - \frac{1}{2t} \|\xi - \xi^{(d)}\|_2^2 \quad (24)$$

The minimization problem for the subproblem g can be converted into solving the solution of the linear Eq. (25). $\mathbf{H}$ is a 3×3 chunking matrix.

$$(\mathbf{H}^T \mathbf{H} + \frac{1}{s} \mathbf{I})^{-1} g^{(d+1)} = \mathbf{H}^T u + \frac{1}{s} g^{(d)} - \alpha \mathbf{R}^T \xi^{(d)} \quad (25)$$

where $\mathbf{H}^T$ is the transpose of $\mathbf{H}$ and $\mathbf{R}^T$ is the transpose of $\mathbf{R}$.

$$g^{(d+1)} = \left(\mathbf{H}^T \mathbf{H} + \frac{1}{s} \mathbf{I} \right)^{-1} \left(\mathbf{H}^T u + \frac{1}{s} g^{(d)} - \alpha \mathbf{R}^T \xi^{(d)} \right) \quad (26)$$

For subproblem ξ , the maximum problem of ξ is first transformed into a minimization problem.

$$\begin{aligned} \xi^{(d+1)} &= \arg \max_{\|\xi\|_\infty \leq 1} \alpha \xi^T \mathbf{R} \bar{g}^{(d+1)} - \frac{1}{2t} \|\xi - \xi^{(d)}\|_2^2 \\ &= \arg \min_{\|\xi\|_\infty \leq 1} \frac{1}{2} \|\xi - (\xi^{(d)} - t\alpha \mathbf{R} \bar{g}^{(d+1)})\|_2^2 \end{aligned} \quad (27)$$

Then, the Eq. (27) is the $\xi^{(d+1)} = P_A(\xi^{(d)} - t\alpha \mathbf{R} \bar{g}^{(d+1)})$, which denotes the projection of $\xi^{(d)} - t\alpha \mathbf{R} \bar{g}^{(d+1)}$ onto set A .

Table1 First-order primal-dual algorithm

Set $s > 0, t > 0, \theta \in [0, 1]$, input, $u, \alpha, \mathbf{R}$

Initialize $\xi^{(0)}, g^{(0)}$

Repeat $k \leftarrow k + 1$

$$g^{(d+1)} = \left(\mathbf{H}^T \mathbf{H} + \frac{1}{s} \mathbf{I} \right)^{-1} \left(\mathbf{H}^T u + \frac{1}{s} g^{(d)} - \alpha \mathbf{R}^T \xi^{(d)} \right)$$

$$\bar{g}^{(d+1)} = g^{(d+1)} + \theta(g^{(d+1)} - g^{(d)})$$

$$\xi^{(d+1)} = P_A(\xi^{(d)} - t\alpha \mathbf{R} \bar{g}^{(d+1)})$$

$$\text{Calculate the deviation value } RD = \frac{\|g^{(i+1)} - g^{(i)}\|_2}{\|g^{(i+1)}\|_2}$$

Until $RD \leq 10^{-4}$, or k reaches the maximum number of iterations

Output recovery image $g^{(i+1)}$

where s is the iteration step of the original variable, t is the iteration step of the dyadic variable, and k is the number of iterations.

$$A \equiv \left\{ \xi \in \mathbb{R}^{3n^2} \times \mathbb{R}^{3n^2} : \|\xi\|_\infty \leq 1 \right\} \quad (28)$$

The defining equations for the projection operator of any vector q on set A are given by Eq. (29)(table 1)

$$P_A(q) \equiv \arg \min_{\xi \in A} \|\xi - q\|_2^2, \forall q \quad (29)$$

Then, there is $P_A(q) = \frac{q}{\max(1, \|q\|_2^2)}$.

2.3.3 Adaptive weighted total variation algorithm

The conventional TV term is based on the assumption that the reconstructed image is piecewise constantly distributed, and this approach can cause excessive smoothing of the reconstructed image edges. To alleviate the problem of over-smoothing of traditional TV edges, many researchers have studied weighted adaptive total variation [33]; therefore, in this study, an adaptive weighted total variation (AwTV) minimization image model was used for comparison with the proposed algorithm.

$$\min_{\mu \geq 0} \|\mu\|_{\text{AwTV}} \text{ subject to } |p - \mathbf{A}\mu| \leq \varepsilon \quad (30)$$

where AwTV denotes the adaptive weight of the total variance of the reconstructed image. The $\|\mu\|_{\text{AwTV}}$ is defined as follows:

$$\begin{aligned} \|\mu\|_{\text{AwTV}} &= \sum_{x,y,z} \sqrt{w_x(\mu_{x,y,z} - \mu_{x-1,y,z})^2 + w_y(\mu_{x,y,z} - \mu_{x,y-1,z})^2 + w_z(\mu_{x,y,z} - \mu_{x,y,z-1})^2} \end{aligned} \quad (31)$$

$$\begin{aligned} w_x &= \exp \left[-\left(\frac{\mu_{x,y,z} - \mu_{x-1,y,z}}{\tau} \right)^2 \right] \\ w_y &= \exp \left[-\left(\frac{\mu_{x,y,z} - \mu_{x,y-1,z}}{\tau} \right)^2 \right] \\ w_z &= \exp \left[-\left(\frac{\mu_{x,y,z} - \mu_{x,y,z-1}}{\tau} \right)^2 \right] \end{aligned} \quad (32)$$

where w_i denotes the weight factor, and τ is a scaling factor in the weights used to control the diffusion intensity of each iteration.

2.3.4 Fast Gradient Projection Algorithm

Beck and Teboulle proposed a fast gradient projection algorithm [34], which is derived as follows. The model expression to solve the TV-based denoising problem is given by Eq. 33.

$$\min_{x \in C} x - b_F^2 + 2\lambda \text{TV}(x) \quad (33)$$

TV is a nonsmooth regularization function that can be an isotropic TV_I or anisotropic TV_{I_1} , and C belongs to a closed convex subset of $E \equiv \mathbb{R}^{m \times n}$. The nonsmoothness of the TV function leads to significant difficulties in solving Eq. 33. To address this challenge, Chambolle proposed a gradient-based algorithm for solving pairwise problems using a pairwise approach under unconstrained conditions. Thus, a constrained problem pair was constructed according to the proposed method. The following symbols are required.

- $\mathcal{P}$ denotes the set of matrix-pairs (p^1, p^2) which meet:

$$(p_{i,j}^1)^2 + (p_{i,j}^2)^2 \leq 1, (0 \leq i \leq m, 0 \leq j \leq n)$$

$$|p_{i,m+1}^1| \leq 1, 0 \leq i \leq m$$

$$|p_{m+1,j}^2| \leq 1, 0 \leq j \leq n$$

- $\mathcal{L} : \mathbb{R}^{(m-1) \times n} \times \mathbb{R}^{m \times (n-1)} \rightarrow \mathbb{R}^{m \times n}$ is linear operation, which is defined according to the following equation.

$$\mathcal{L}(p^1, p^2)_{i,j} = (p_{i,j}^1 - p_{i-1,j}^1) + (p_{i,j}^2 - p_{i,j-1}^2)$$

$$1 \leq i \leq m, 1 \leq j \leq n$$

- The operator $\mathcal{L}^T : \mathbb{R}^{m \times n} \rightarrow \mathbb{R}^{(m-1) \times n} \times \mathbb{R}^{m \times (n-1)}$ adjacent to $\mathcal{L}$ is calculated according to the following formula.

$$\mathcal{L}^T(x) = (p^1, p^2)$$

where $p \in \mathbb{R}^{m \times n}$ is the matrix, whose defining equations are

$$p_{i,j}^1 = x_{i,j} - x_{i+1,j}, i = 1, \dots, m-1, j = 1, \dots, n$$

$$p_{i,j}^2 = x_{i,j} - x_{i,j+1}, i = 1, \dots, m, j = 1, \dots, n-1$$

P_C belongs to an orthogonal projection operator in set C . Therefore, when $C = B_{l,u}$, the $P_{B_{l,u}}$ can be defined as:

$$P_{B_{l,u}}(X)_{i,j} = \begin{cases} lx_{ij} < l \\ x_{ij}l \leq x_{ij} \leq u \\ ux_{ij} > u \end{cases}$$

Based on the notation above, we obtain the dual problem in Eq. 33 and show the relationship between the primal and dual optimal solutions. To analyze this relationship, we propose the following hypothesis: We assume that $((p^1, p^2)) \in \mathcal{P}$ is the optimal solution to this problem.

$$\min_{(p^1, p^2) \in \mathcal{P}} \left\{ h(p^1, p^2) \equiv -\|H_C(b - \lambda \mathcal{L}(p^1, p^2))\|_F^2 + \left\| b - \lambda \mathcal{L}(p^1, p^2) \right\|_F^2 \right\} \quad (34)$$

The $H_C(x)$ defined as:

$$H_C(x) = x - P_C(x), x \in \mathbb{R}^{m \times n} \quad (35)$$

Because objective function h in Eqs. (34) is continuously differentiable and its gradient is defined as

$$\nabla h(p^1, p^2) = -2\lambda \mathcal{L}^T P_C(b - \lambda \mathcal{L}(p^1, p^2)), \quad (36)$$

$$T(x, p^1, p^2) = \text{Tr}(\mathcal{L}(p^1, p^2)^T x). \quad (37)$$

Therefore, problem Eq. 34 can be converted into the following equation.

$$\min_{x \in C} \max_{(p^1, p^2) \in \mathcal{P}} \left\{ X - b_F^2 + 2\lambda \text{Tr}(\mathcal{L}(p^1, p^2)^T x) \right\} \quad (38)$$

Because the objective function is concave in p^1, p^2 and convex in x , the maximum and minimum orders can be exchanged. Thus, the following equation is obtained:

$$\max_{(p^1, p^2) \in \mathcal{P}} \min_{x \in C} \left\{ x - b_F^2 + 2\lambda \text{Tr}(\mathcal{L}(p^1, p^2)^T x) \right\}. \quad (39)$$

This can be rewritten as follows:

$$\max_{(p^1, p^2) \in \mathcal{P}} \min_{x \in C} \left\{ X - (b - \lambda \mathcal{L}(p^1, p^2))_F^2 - b - \lambda \mathcal{L}(p^1, p^2)_F^2 + b_F^2 \right\}, \quad (40)$$

where the optimal solution of the minimum problem is

$$x = P_C(b - \lambda \mathcal{L}(p^1, p^2)). \quad (41)$$

Subsequently, by substituting Eq. 41 into Eq. 40 and omitting the constant term, we obtain the following dual problem:

$$\max_{(p^1, p^2) \in \mathcal{P}} \left\{ P_C(b - \lambda \mathcal{L}(p^1, p^2)) - (b - \lambda \mathcal{L}(p^1, p^2))_F^2 - b - \lambda \mathcal{L}(p^1, p^2)_F^2 \right\}. \quad (42)$$

Our goal is to solve the dual problem in Eq. 34, whose gradient is expressed by Eq. 36. Therefore, the gradient projection algorithm, which is an effective method for solving the denoising problem, is presented. Because the norm of $x \in \mathbb{R}^{m \times n}$ is Frobenius norm. For $(p^1, p^2) \in \mathbb{R}^{(m-1) \times n} \times \mathbb{R}^{m \times (n-1)}$, the norm is expressed as follows:

$$\|p^1, p^2\| = \sqrt{\|p^1\|_F^2 + \|p^2\|_F^2}, \quad (43)$$

$$Lh \leq \lambda^2. \quad (44)$$

The projection onto set $\mathcal{P}$ can be computed simply: For (p^1, p^2) , projection $P_{\mathcal{P}}(p^1, p^2)$ is given by

$P_p(p^1, p^2) = (r^1, r^2)$. The equations of $r \in \mathbb{R}^{m \times n}$ are shown below:

$$\begin{aligned} r_{ij}^1 &= \begin{cases} \frac{p_{ij}^1}{\max \left\{ 1, \sqrt{(p_{ij}^1)^2 + (p_{ij}^2)^2} \right\}} \\ \frac{p_{ij}^1}{\max \left\{ 1, |p_{ij}^1| + |p_{ij}^2| \right\}} \end{cases} \\ r_{ij}^2 &= \begin{cases} \frac{p_{ij}^2}{\max \left\{ 1, \sqrt{(p_{ij}^1)^2 + (p_{ij}^2)^2} \right\}} \\ \frac{p_{ij}^2}{\max \left\{ 1, |p_{ij}^1| + |p_{ij}^2| \right\}} \end{cases} \end{aligned} \quad (45)$$

Substituting the objective function, gradient equation, and maximum Lipschitz constant into Eqs. 34 into the gradient projection algorithm introduced above, we obtain the denoising algorithm.

2.3.5 Os-sart algorithm

In 1984, Anderson and Kak proposed an improved SART. The SART improves on the ART algorithms by calculating the error of all rays passing through a pixel at the same projection angle. The ART uses only one ray per iteration, whereas the SART corrects all the rays at the same angle. This is similar to lowering the noise introduced by the ART algorithm. Its formula is as follows:

$$x_j^{(k+1)} = x_j^k + \frac{\sum_{p_i \in P_\varphi} \left\{ \lambda_k \frac{p_i - \sum_{m=0}^{M-1} w_{im} x_{im}^{(k)}}{\sum_{m=0}^{M-1} w_{im}} w_{ij} \right\}}{\sum_{p_i \in P_\varphi} w_{ij}}. \quad (46)$$

The OS-SART is a combination of ordered subsets and the SART algorithm, and its formula is as follows:

$$x_j^{(k+1)} = x_j^k + \frac{\sum_{i \in S_t} \left\{ \lambda_k \frac{p_i - \sum_{m=0}^{M-1} w_{im} x_{im}^{(k)}}{\sum_{m=0}^{M-1} w_{im}} w_{ij} \right\}}{\sum_{i \in S_t} w_{ij}}, \quad (47)$$

where i denotes the i -th ray at a certain projection angle, j denotes the pixel value in the i -th ray, k denotes the number of iterations, and λ is the relaxation factor.

2.3.6 Proposed algorithm

Based on the OS-SART iterative algorithm and first-order primal-dual algorithm with total variation denoising, we propose a new algorithm for sparse-view NCT 3D image reconstruction (OS-SART-PDTV). The iterative process of

Table 2 OS-SART-PDTV algorithm

1. While the stop criterion is not met
2. Step1: OS-SART reconstruct the image
3. Initialization: g_0 , $\lambda_{OS-SART}$, $N_{OS-SART}$ and N_{PDTV}
4. For $n = 1$ to $N_{OS-SART}$ do; (n denotes the number of iterations)
5. $g_j^{(k+1)} = g_j^k + \frac{\sum_{i \in S_t} \left\{ \lambda_k \frac{p_i - \sum_{m=0}^{M-1} w_{im} x_{im}^{(k)}}{\sum_{m=0}^{M-1} w_{im}} w_{ij} \right\}}{\sum_{i \in S_t} w_{ij}}$
6. Non-negativity constraint; If $g_{OS-SART}^{(n)} < 0$, $g_{OS-SART}^{(n)} = 0$
7. Step2: PDTV
8. Set $s > 0$, $t > 0$, $\theta \in [0, 1]$, input, u , α , $\mathbf{R}$
9. Initialize $\xi^{(0)}$, $g^{(0)} = g_{OS-SART}$
10. While the stop criterion is not met
11. $g^{(d+1)} = \left(\mathbf{H}^T \mathbf{H} + \frac{1}{s} \mathbf{I} \right)^{-1} \left(\mathbf{H}^T u + \frac{1}{s} g^{(d)} - \alpha \mathbf{R}^T \xi^{(d)} \right)$
12. $\bar{g}^{(d+1)} = g^{(d+1)} + \theta (g^{(d+1)} - g^{(d)})$
13. $\xi^{(d+1)} = P_A(\xi^{(d)} - t\alpha \mathbf{R} \bar{g}^{(d+1)})$
14. Calculate the deviation value RD , $RD = \frac{\ g^{(d+1)} - g^{(d)}\ _2}{\ g^{(d+1)}\ _2}$
15. Until $RD \leq 10^{-4}$, or d reaches the maximum number of iterations
16. Output recovery image $g^{(d+1)}$
17. end if;
18. Until the stopping criteria are reached
19. Get the final reconstructed image $g_{OS-SART-PDTV}$

this algorithm contains two loops: the outer and inner loops are OS-SART for image reconstruction and PDTV for image denoising (Table 2).

3 Experiment

3.1 Quantitative evaluation index

In this study, we used four image evaluation metrics to quantitatively analyze the mass of the reconstructed images for each algorithm. The accuracy of the image reconstruction was quantitatively analyzed using the correlation coefficient (CC) metric, which is defined as follows:

$$CC = \frac{\sum_{k=1}^M (\varpi(y) - \overline{\varpi})(\varpi_{true}(y) - \overline{\varpi_{true}})}{\left\{ \sum_{k=1}^M (\varpi(y) - \overline{\varpi})^2 \sum_{y=1}^M (\varpi_{true}(y) - \overline{\varpi_{true}})^2 \right\}^{1/2}}, \quad (48)$$

where $\overline{\varpi_{true}}$ is the image to be reconstructed, M denotes the voxel number, and $\varpi(y)$ represents the reconstructed value at voxel y . When the reconstructed image was the same as the image to be reconstructed, the CC value was 1.

The ((MSSIM) provides a comprehensive evaluation of the reconstructed image by comparing the differences between the image to be reconstructed and the reconstructed image in terms of brightness, contrast, and structure.

$$\text{Brightness information: } l(\varepsilon, \eta) = \frac{2u_\varepsilon u_\eta + c_1}{u_\varepsilon^2 + u_\eta^2 + c_1}.$$

$$\text{Contrast information: } c(\varepsilon, \eta) = \frac{2\sigma_\varepsilon \sigma_\eta + c_2}{\sigma_\varepsilon^2 + \sigma_\eta^2 + c_2}.$$

$$\text{Structure Information: } s(\varepsilon, \eta) = \frac{\varsigma_{\varepsilon\eta} + c_3}{\varsigma_\varepsilon + \varsigma_\eta + c_3}.$$

where ε, η denote the standard image and the image to be evaluated, respectively; u_ε and u_η are the mean values of the image; σ_ε and σ_η are the standard deviations of the image; $\varsigma_{\varepsilon\eta}$ is the covariance of the image; C_1, C_2 and C_3 are constants. The mean structural similarity of an image is defined as

$$\text{MSSIM} = \frac{1}{M} \sum_{t=1}^M [l(\varepsilon_t, \eta_t)]^\alpha \times [c(\varepsilon_t, \eta_t)]^\beta \times [s(\varepsilon_t, \eta_t)]^\gamma. \quad (49)$$

RMSE is defined by the following equations:

$$\text{RMSE} = \sqrt{\frac{\sum_{n=1}^N (\hat{p}_n - p_n)^2}{N}}, \quad (50)$$

$\hat{p}_n$ represents a voxel in the image to be reconstructed, p_n represents a voxel in the reconstructed sample, and N indicates the number of voxels. A higher RMSE value indicates a larger error between the reconstructed images.

Universal quality image (UQI) is a widely used image evaluation index, which is defined as follows:

$$\text{UQI} = \frac{2\text{cov}(\hat{\mu}, \mu)}{\hat{\rho}^2 + \rho^2} \cdot \frac{2\hat{\mu}\mu}{\hat{\mu}^2 + \mu^2}, \quad (51)$$

where cov denotes the covariance function, $\hat{\mu}, \mu$ are the means, and $\hat{\rho}^2, \rho^2$ are the variances of the reconstructed

image and the image to be reconstructed, respectively. The UQI value ranges from 0 to 1 and increases with similarity.

3.2 Digital 3D shepp–logan model experiment

In the digital simulation experiment, we compared and analyzed the performance of five algorithms: FBP, OS-SART-TV, OS-SART-AwTV, OS-SART-FGPTV, and OS-SART-PDTV by reconstructing 3D Shepp–Logan model projection data. The dimensions of the 3D Shepp–Logan model were $256 \times 256 \times 256$. An iterative algorithm must set reasonable iterative parameters to obtain an ideally reconstructed image. The following parameters were used to compare the three iterative algorithms: OS-SART-TV $\lambda = 2.7$, $\text{Niter}_{\text{OS-SART}} = 80$, $\lambda_{\text{TV}} = 200$, and $\text{Niter}_{\text{TV}} = 300$. For the OS-SART-AwTV algorithm $\lambda = 2.5$, $\text{Niter}_{\text{OS-SART}} = 80$, $\lambda_{\text{AwTV}} = 100$, and $\text{Niter}_{\text{TV}} = 250$. For the OS-SART-FGPTV algorithm, $\lambda = 2.2$, $\text{Niter}_{\text{OS-SART}} = 80$, $\lambda_{\text{PDTV}} = 0.006$, $\text{Niter}_{\text{FGPTV}} = 150$. For the OS-SART-PDTV algorithm, $\lambda = 2.2$, $\text{Niter}_{\text{OS-SART}} = 80$, $\lambda_{\text{PDTV}} = 0.0032$, $\text{Niter}_{\text{PDTV}} = 300$. To verify the denoising capability of the algorithm, the following parameters were set according to the noise model of the NCT system: These parameters include the photon incident flux 1×10^5 , $m_{\text{ic}} = 0$, and $\delta_{\text{ic}}^2 = 10$.

Figure 2 shows the relationship between the RMSE of the reconstructed images using the OS-SART-PDTV algorithm and the number of iterations for different sparse views (28 and 38 projection views), showing that the OS-SART-PDTV algorithm can converge to a stable solution after a certain number of iterations.

Figure 3 shows the reconstruction of images with different numbers of projected views using the five algorithms. As illustrated in Fig. 3, the smoothness and sharpness of

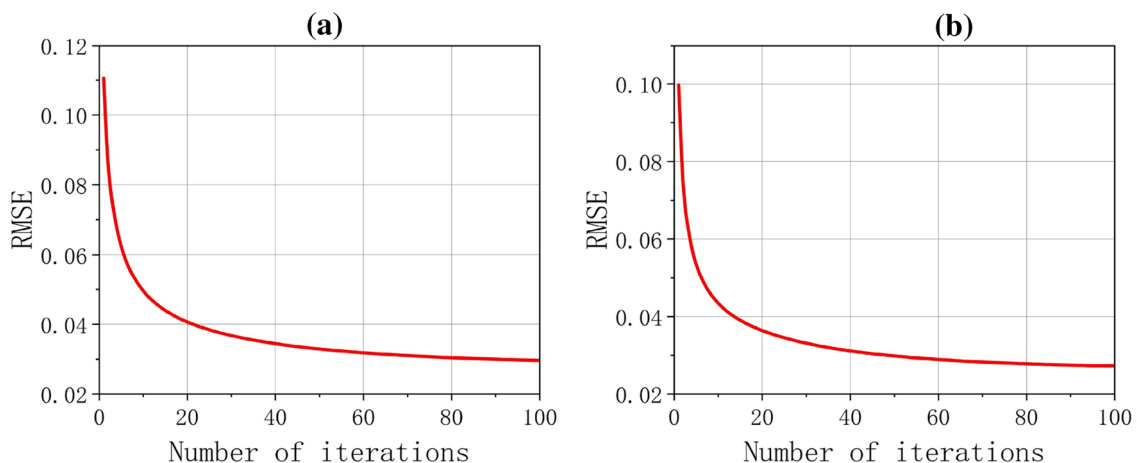
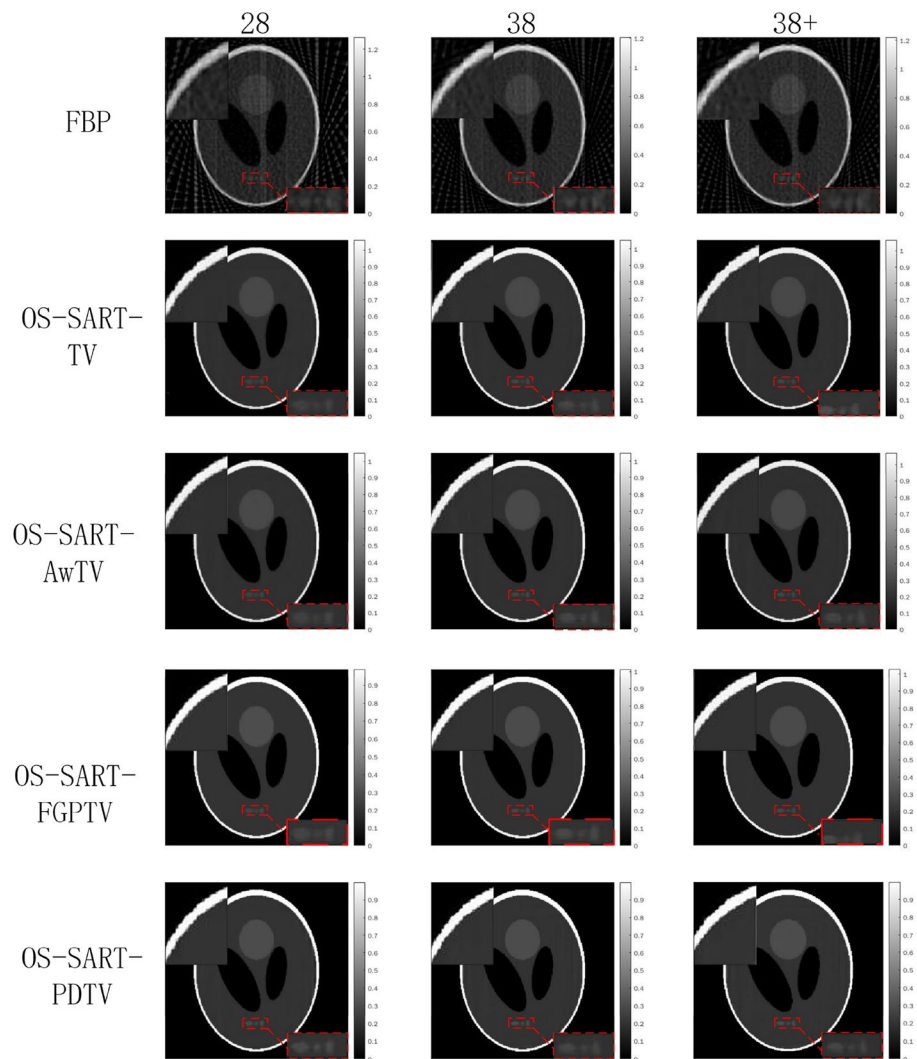


Fig. 2 RMSE versus iteration steps for the OS-SART-PDTV algorithm from different sparse views: **a** 28; **b** 38

Fig. 3 Sliced images at $z = 130$ for four algorithms reconstructing reconstructed images with different numbers of projection views



the images reconstructed using the five algorithms improved rapidly as the number of projection views increased. Compared to the other four iterative algorithms, the reconstructed images of the FBP algorithm showed severe bar artifacts. The reconstructed images from the four iterative algorithms are smoother and have sharper boundaries than those from the FBP algorithm. As shown in the enlarged ROI in Fig. 3, the OS-SART-TV, OS-SART-AwTV, and OS-SART-FGPTV algorithms can effectively reduce artifacts; however, part of the edge structure is not exactly reconstructed. In contrast, the novel algorithm can not only maintain good reconstruction performance but also reconstruct finer structures.

In addition, we compared and analyzed the error images of the reconstructed images using different algorithms that reflected the difference between the pixel values of the reconstructed and reference images. According to the error image in Fig. 4, the image reconstructed using the FBP algorithm lost many features and produced many artifacts, indicating that the algorithm is not suitable for sparse-view

NCT 3D reconstruction. The other four iterative algorithms performed better at suppressing artifacts and reconstructing detailed structural information. The OS-SART-PDTV algorithm has much less detail loss in the error image than the other algorithms.

We plotted the profile of the reconstructed images for each algorithm in Fig. 5 to further evaluate the performance of each algorithm. The number of projection views from left to right was 28, 38, and 38+, respectively. Here, 38+ indicates that the noise model was added to the 3D Shepp-Logan model. As shown in Fig. 5, there was a large fluctuation in the FBP profile line with the largest deviation from the true pixel value. Although the OS-SART-TV, OS-SART-AwTV, and OS-SART-FGPTV algorithms performed better than the FBP algorithm, there was still a gap between their amplitudes and true pixel values. The profile of OS-SART-PDTV matches the reference value best and is closest to the true value. This also indicates that

Fig. 4 Slice of error images at $z=130$

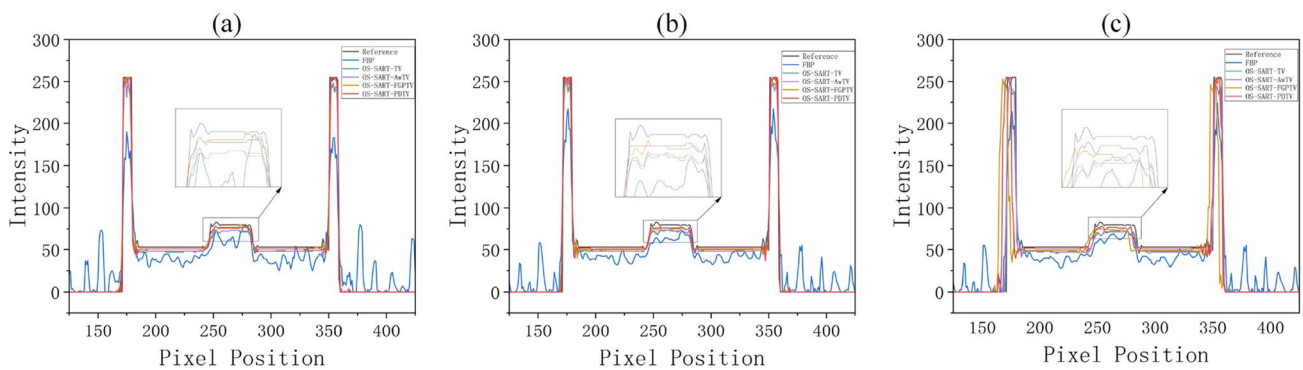
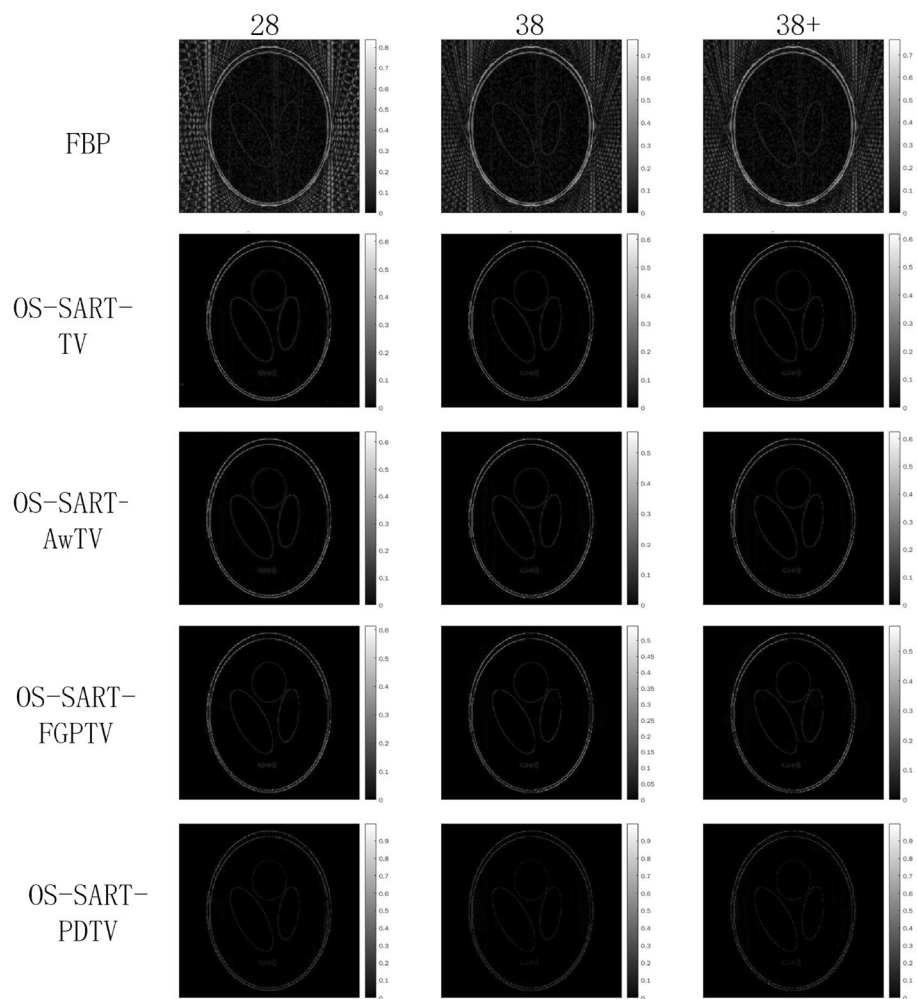


Fig. 5 (Color online) Horizontal profiles of the 3D Shepp–Logan model at $z=130^{\text{th}}$ slices. **a** 28; **b** 38; **c** 38+

the OS-SART-PDTV algorithm is well-suited for sparse-view NCT 3D reconstruction.

As shown in Fig. 6. We used four image evaluation metrics to quantitatively analyze the reconstructed images from each algorithm. First, the evaluation indices of the images reconstructed using noiseless projection data

were analyzed. From the CC, MSSIM, and UQI histograms shown in Fig. 6, the values of the three metrics for the reconstructed images from each reconstruction algorithm gradually increased as the number of projection views increased. These three metrics exhibit similar characteristics.

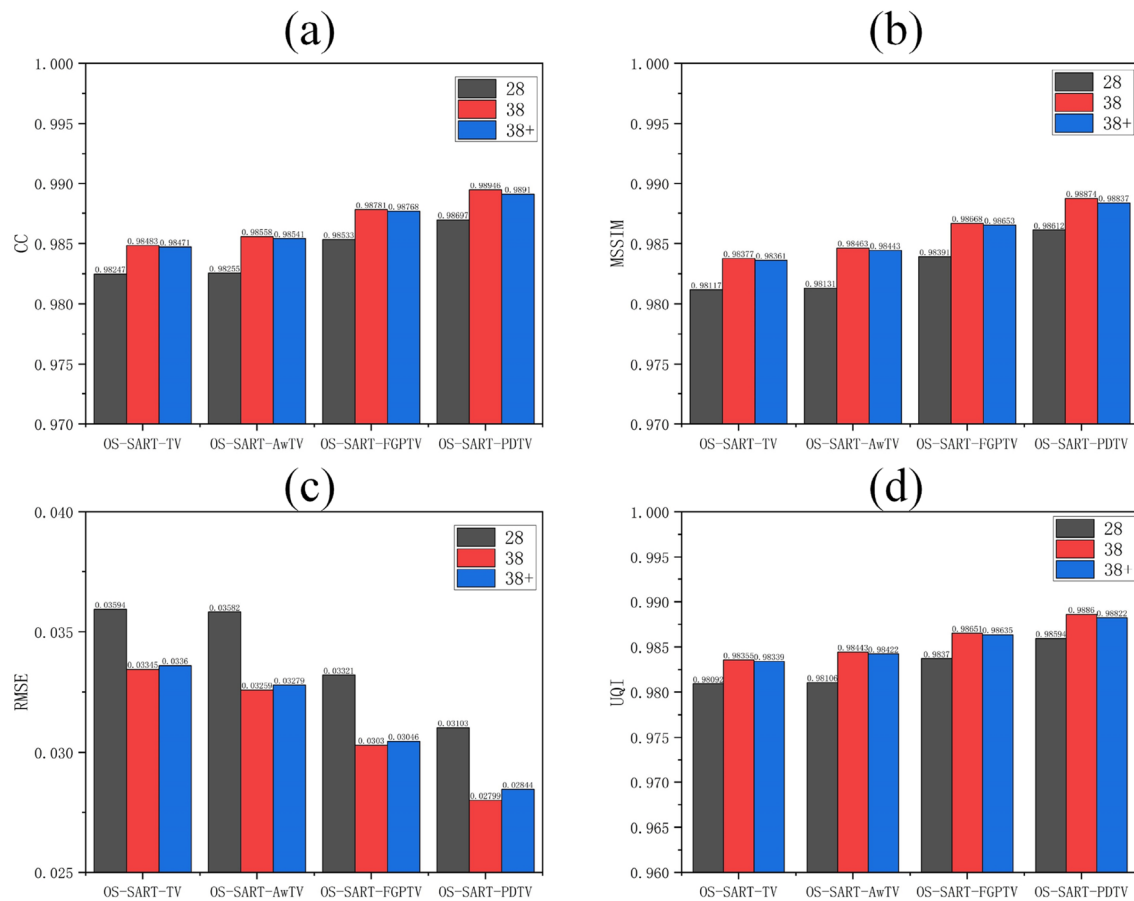


Fig. 6 (Color online) Histogram of reconstructed image evaluation metrics for the 3D Shepp-Logan model. **a** CC; **b** MSSIM; **c** RMSE; **d** UQI

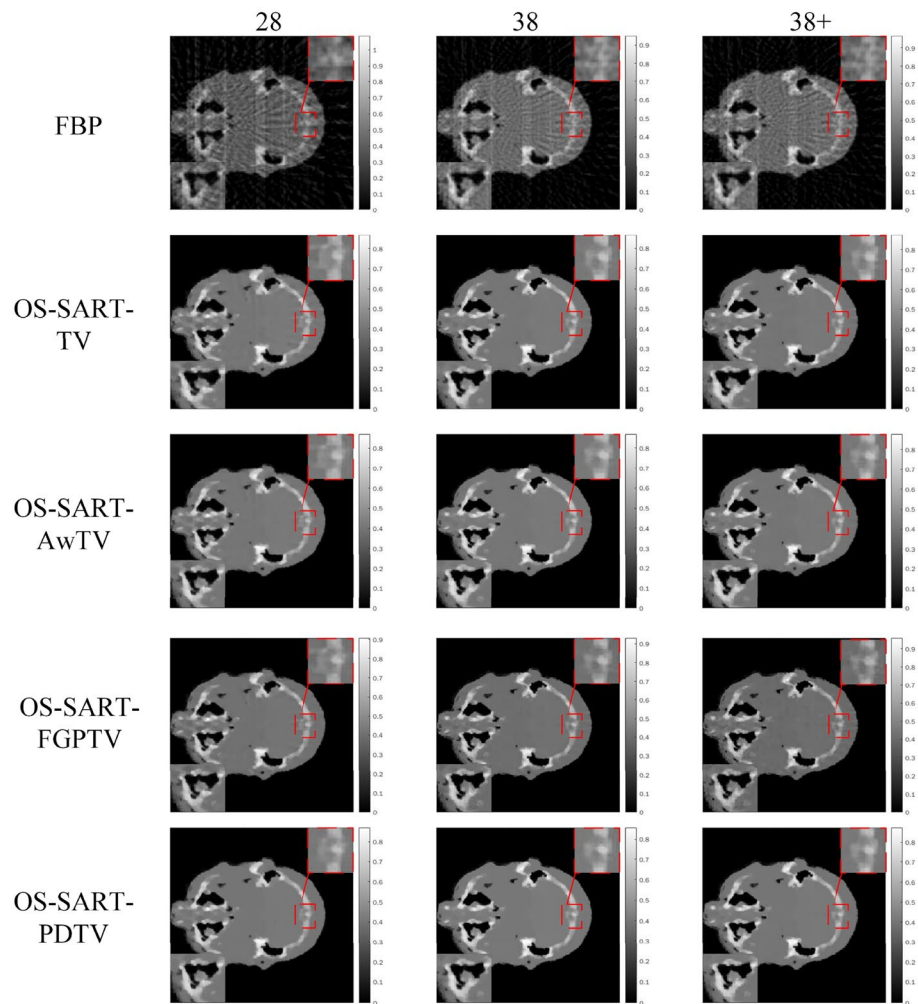
According to the RMSE histogram in Fig. 6c, the reconstructed images of OS-SART-PDTV had the smallest RMSE values compared to the other algorithms for the same projection views. The RMSE of the reconstructed image for each algorithm gradually decreased as the number of projection views increased. The RMSE value of the images reconstructed using the OS-SART-PDTV algorithm at 28 projection views was 0.03103, indicating that the reconstructed image was closest to the real image and that the quality of the reconstructed images was higher. As shown in the UQI histogram in Fig. 6d, the UQI values of the reconstructed images for each algorithm gradually increased as the number of projected views increased. The trend of the UQI was the opposite to that of the RMSE. The reconstructed images from the OS-SART-PDTV algorithm had the highest UQI values for the same number of projection views, indicating that it outperformed the other algorithms. The UQI value of the image reconstructed using the OS-SART-PDTV algorithm was 0.98594 for the 28 projection views. We then quantitatively analyzed four evaluation metrics for the images reconstructed using data containing noisy projections. From the CC, MSSIM,

and UQI histograms, it can be seen that the OS-SART-PDTV algorithm has the largest evaluation metrics for these three reconstructed images when the same number of noise-containing projections are reconstructed. According to the RMSE histogram, the reconstructed image from the OS-SART-PDTV algorithm had the lowest RMSE value. Therefore, based on the quantitative and visual analyses of each algorithm, the OS-SART-PDTV algorithm has the highest-quality reconstructed images with the same number of projection views.

3.3 Digital head model experiment

We further analyzed the performance of these five algorithms by reconstructing a digital head model. The reconstructed images from the five algorithms used for reconstructing different projection view numbers are shown in Fig. 7. Based on the reconstructed images, the smoothness and sharpness of the reconstructed images of all five algorithms improve significantly as the number of projected views increases. Compared to the other four algorithms, the reconstructed images of the FBP algorithm show severe bar

Fig. 7 (Color online) Sliced images at $z = 70$ for four algorithms reconstructing reconstructed images with different number of projection views



artifacts. The reconstructed images from these four iterative algorithms were smoother and had cleaner image boundaries than those from the FBP algorithm. As shown in the enlarged ROI in Fig. 7, the OS-SART-TV, OS-SART-AwTV, and OS-SART-FGPTV algorithms are effective in reducing artifacts, but the edge structure of the image is not accurately

reconstructed. In contrast, the proposed OS-SART-PDTV algorithm can not only reduce artifacts, but also reconstruct more detailed structural information.

To further evaluate the performance of the algorithm, the contours of the reconstructed images for each algorithm are plotted in Fig. 8. As shown in Fig. 8, the numbers of

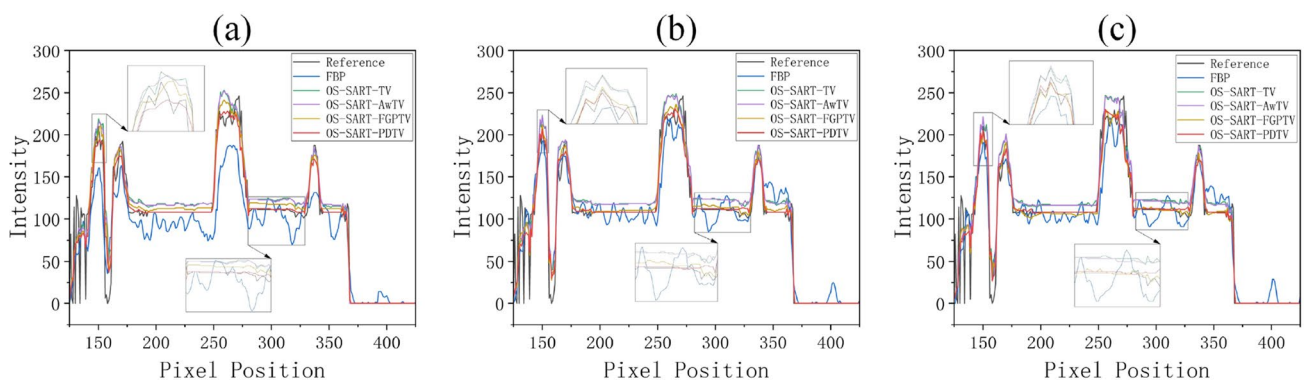


Fig. 8 (Color online) Horizontal profiles of the digital head model at $z = 250$ th slices. **a** 28; **b** 38; **c** 38 +

projected views from left to right are 28, 38, and 38 +, where 38 + indicates that noise has been added to the projection data. As shown in Fig. 8, there was a large fluctuation in the FBP profile, which had the largest gap from the true pixel values. Although the OS-SART-TV, OS-SART-AwTV, and OS-SART-FGPTV algorithms performed better than the FBP algorithm, there was still a large deviation between the actual and true pixel values. The profile of OS-SART-PDTV matches the reference value best and is closest to the true value. This also indicates that the OS-SART-PDTV algorithm is well-suited for sparse-view NCT 3D reconstruction.

According to the RMSE histogram of the reconstructed images from each algorithm, the OS-SART-PDTV algorithm had the smallest RMSE of the reconstructed images. As the number of projected views gradually increased, the RMSE of the reconstructed image for each algorithm gradually decreased. When the number of projected views is 28, the RMSE value of the OS-SART-PDTV algorithm-reconstructed image is 0.02404, which indicates that the reconstructed image of this algorithm is the closest to the real image, and the quality of the reconstructed image is higher.

The UQI histogram in Fig. 9d shows that the UQI values of the reconstructed images of each algorithm increased as the number of projected views increased. The UQI trend was opposite to that of the RMSE. The OS-SART-PDTV algorithm reconstructs the image with the largest UQI value for the same number of projected views, indicating that this algorithm outperforms the other three iterative algorithms. When the number of projected views was 28, the UQI of the image reconstructed by PDTV algorithm was 0.9927. We then quantitatively evaluated images containing noisy projections reconstructed using the four algorithms. The OS-SART-PDTV algorithm reconstructs the image with the largest CC, MSSIM, and UQI values for the same number of projected views. According to the RMSE histogram, the RMSE value of the reconstructed image obtained using OS-SART-PDTV was the smallest. Therefore, by quantitative and visual analyses of each algorithm, OS-SART-PDTV exhibited superior performance in sparse view reconstruction.

To analyze the relationship between the OS-SART-PDTV algorithm and regularization parameters, we

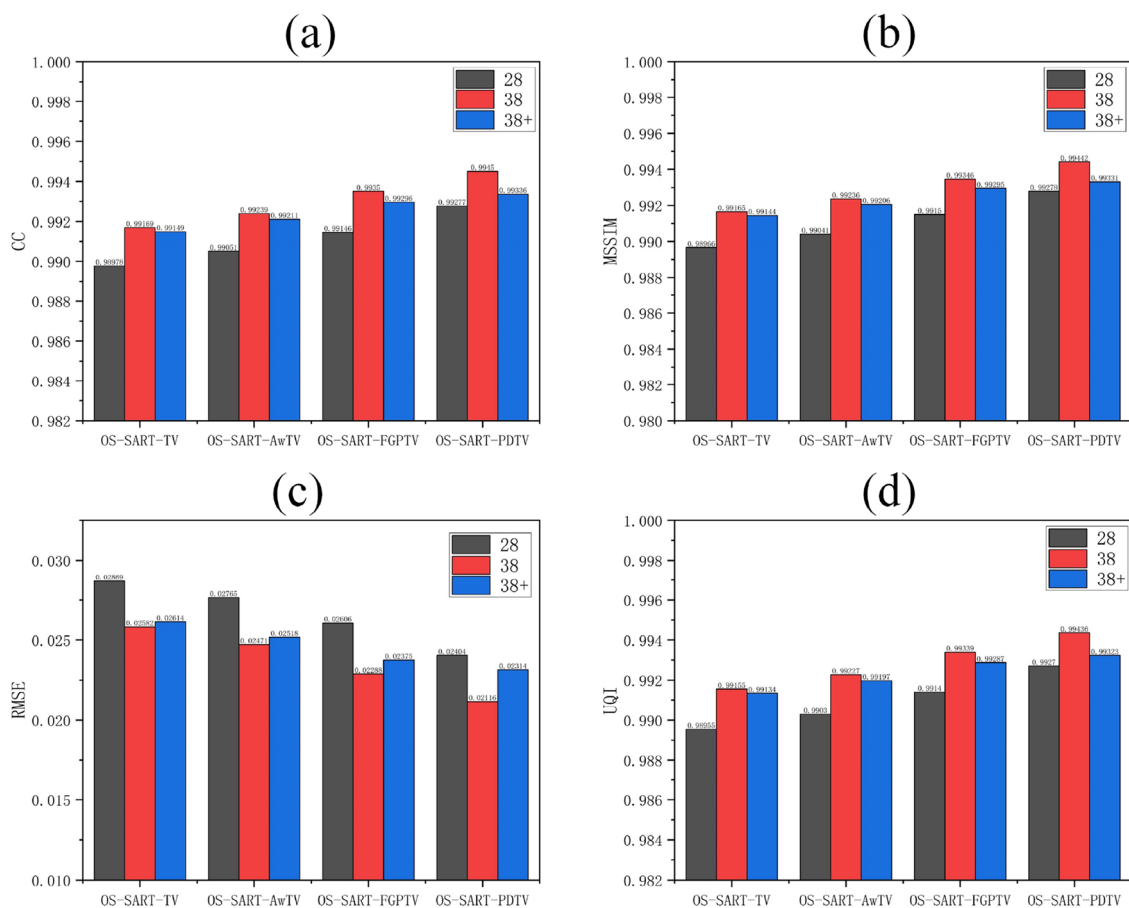


Fig. 9 (Color online) Histogram of reconstructed image evaluation metrics for the digital head model. **a** CC; **b** MSSIM; **c** RMSE; **d** UQI

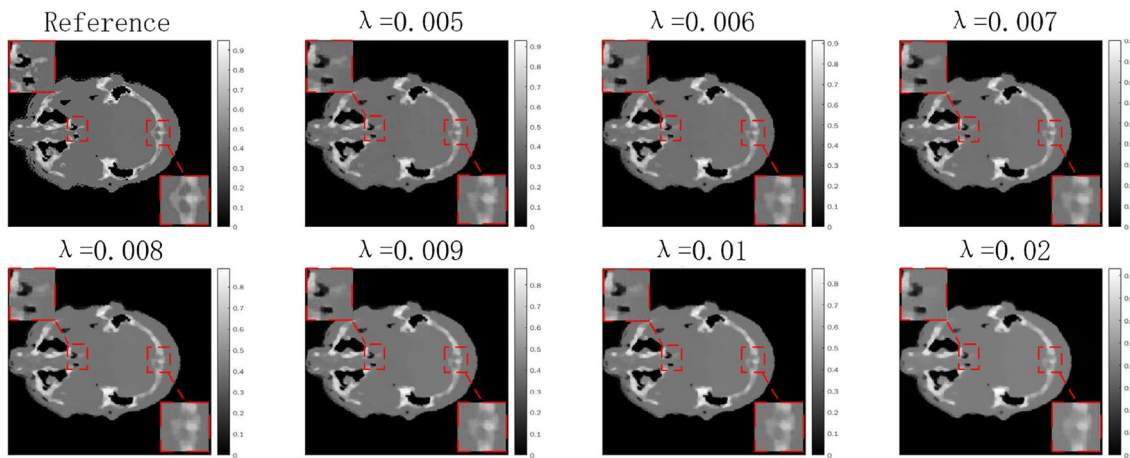


Fig. 10 (Color online) Reconstructed images with different regularization parameters for the OS-SART-PDTV algorithm 38+



Fig. 11 (Color online) Clock model

present the reconstructed images with different regularization parameters in Fig. 10. Based on the reconstructed images, an optimal image was obtained by adjusting the regularization parameters when reconstructing the head model. According to the reconstructed images with different regularization parameters, it can be seen that the reconstructed image quality reaches the best when the regularization parameters reach a certain value as the reconstruction parameters gradually increase. As the regularization parameter continued to increase, the quality of the reconstructed image decreased.

3.4 Real neutron projection experiment

To further validate the performance of this novel algorithm in practical applications, the performance of the algorithm was verified using projection data from an NCT-based clock model (Fig. 11) provided by Burkhard Schillinger of the Technical University of Munich, Germany [35]. Schillinger et al. provided 201 neutron projection images captured at equal angles in the range of 0° to 180° . They also provided two dark field and two open field background images. The background images were obtained using the CCD camera with the neutron beam turned off, which contained read-out noise as well as electron noise caused by dark currents. Open-field images were obtained using a CDD camera with neutrons turned on and without the samples.

As shown in Fig. 12, the artifacts of the reconstructed images by the FBP algorithm gradually increase as the number of projected views decreases. When the number of projected views is 45, the four iterative algorithms outperform the FBP algorithm in suppressing artifacts and reconstructing fine structure information. Although the OS-SART-TV, reconstructs images with fewer artifacts, the edges of the reconstructed images are too smooth, resulting in the loss of detailed structural information of the reconstruction images. Despite OS-SART-AwTV and OS-SART-FGPTV algorithms can also reconstruct a lot of image details, their reconstruction results are poor compared with the OS-SART-PDTV algorithm that reconstructs the same number of sparse views. Visually, the OS-SART-PDTV algorithm outperforms other algorithms in terms of noise removal, artifact suppression, and retention of detailed structural information.

The performances of the three iterative algorithms were quantitatively analyzed using four image-evaluation metrics, as shown in Fig. 13. In Fig. 13, the RMSE of the

Fig. 12 (Color online) Sliced images at $z=201$ for four algorithms reconstructing reconstructed images with different number of projection views

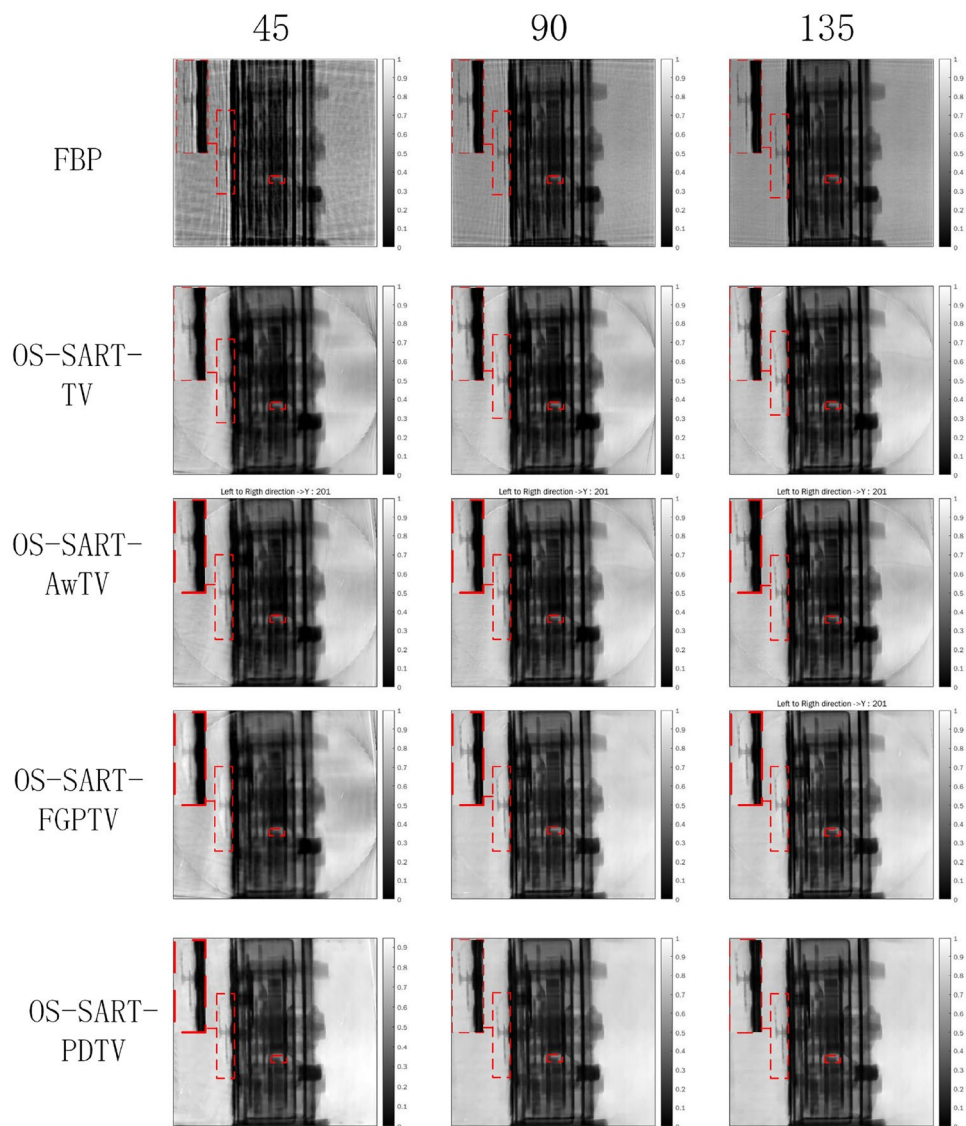


image reconstructed by the OS-SART-PDTV algorithm is the smallest, indicating the minimal error between the reference image and the reconstructed image, whereas the image reconstructed by the OS-SART-PDTV algorithm has the highest CC, MSSIM, and UQI values, indicating that it is most similar to the reference image. In summary, the OS-SART-PDTV algorithm performed well in denoising, suppressing artifacts, and reconstructing fine structures when reconstructing the projection data of sparse NCT scans, indicating that this novel algorithm can be applied to the 3D reconstruction of sparse-view NCT.

Figure 14 presents the profiles of the reconstructed images. According to the profile, the OS-SART-PDTV algorithm has an advantage over the other algorithms, and its profile is the closest to the reference.

4 Discussion

Herein, we proposed a 3D reconstruction algorithm for sparse-view NCT. Because the analytical algorithm (FBP) has inherent defects that make it unsuitable for sparse-view reconstruction, we used an iterative algorithm applied to the 3D reconstruction of sparse-view NCT. Furthermore, the iterative algorithm can embed a priori information to regularize the image. Therefore, the TV-based iterative algorithm is widely used in sparse-view CT image reconstruction because it is superior to other algorithms in artifact suppression and denoising. By reconstructing the projection data of the 3D Shepp-Logan, head, and clock models, the OS-SART-PDTV algorithm outperformed the other algorithms in both

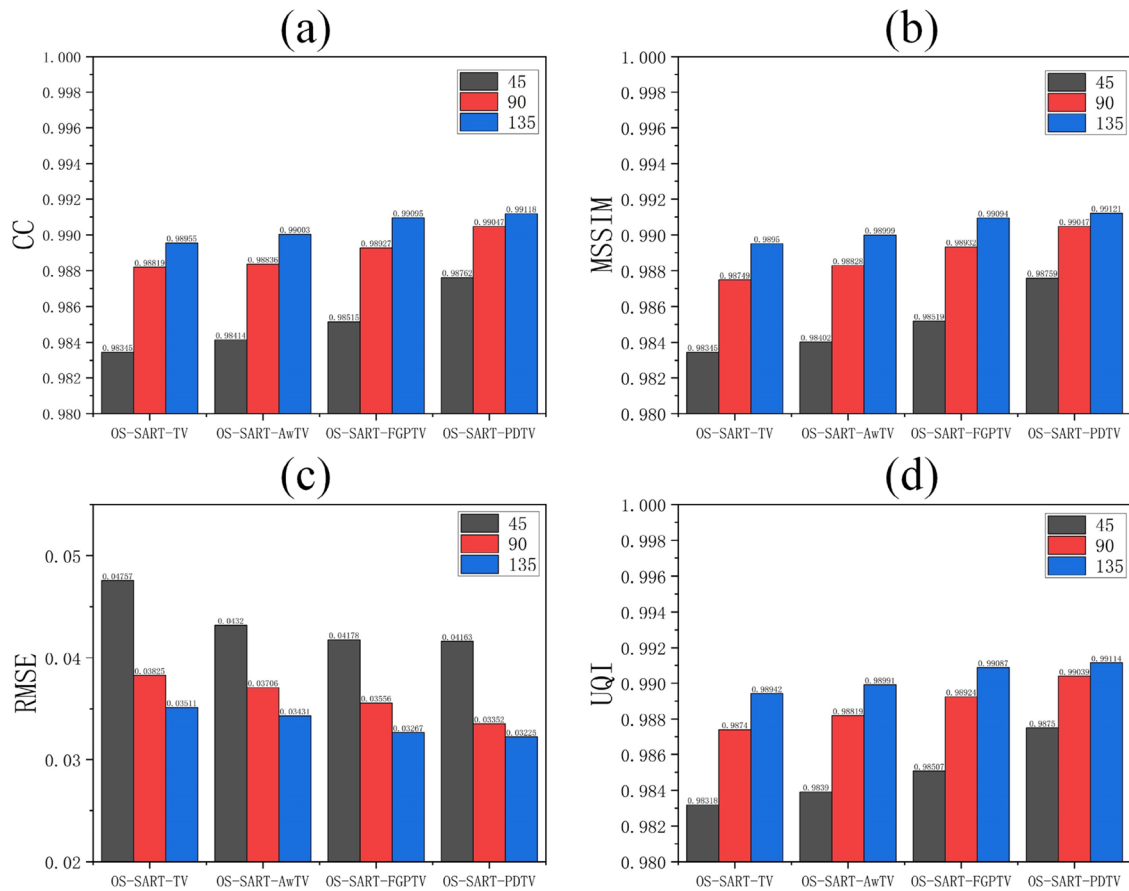


Fig. 13 (Color online) Histogram of reconstructed image evaluation metrics for the clock model. **a** CC; **b** MSSIM; **c** RMSE; **d** UQI

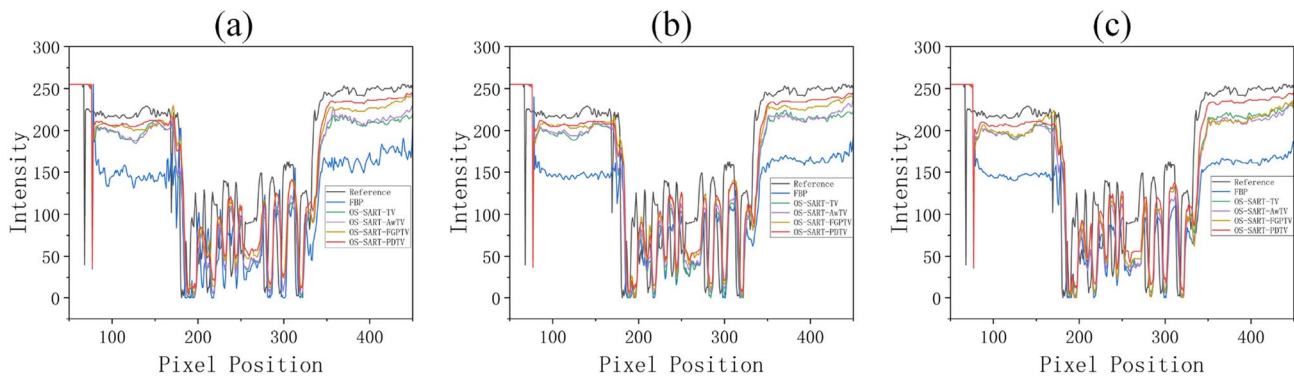


Fig. 14 (Color online) The profile of reconstructed images by FBP, OS-SART-TV, OS-SART-AwTV, OS-SART-FGPTV, and OS-SART-PDTV at $z=201$. The number of projection views is a 45; b 90; c 135

quantitative and visual analyses. Therefore, OS-SART-PDTV is suitable for sparse-view NCT. In addition, deep learning, an important research direction in the field of image reconstruction, has led to many significant research results in sparse-view reconstruction, such as in [36], using short C-arm CT. scans. This study proposed

the DRONE model, which uses a codec network in the embedding module to extract depth features in the data and image domains, combined with the Wasserstein distance generation adversarial network to maintain details and features in the image domain. It combines data residual and image residual networks in the refinement module

to restore fine structure features from the output of the embedding module. According to the CS iterative reconstruction model, the depth prior of the data and image domains is regularized, thus ensuring the robustness of the DRONE network and making the DRONE model good at retaining image edge information, recovering image features, and reconstructing accurate image information in practical applications.

Generally, in practical NCT scanning systems, the projected image is affected by photon statistical noise and electrical noise. Therefore, several iterative reconstruction algorithms cannot achieve satisfactory results in NCT scanning systems. To further test the performance of our algorithm, we reconstructed the projection data obtained from the NCT scan-clock model using the OS-SART-PDTV algorithm. Based on the results of the reconstructed neutron projection data, OS-SART-PDTV exhibited a good performance in denoising, suppressing artifacts, and reconstructing fine structures. According to the results of the reconstruction of the 3D Shepp-Logan model, the OS-SART-PDTV algorithm converged monotonically to a stable solution (Fig. 3).

Iterative reconstruction algorithms generally need to optimize several parameters to obtain a stable solution. This is a common problem encountered by all iterative reconstruction algorithms. The OS-SART-PDTV algorithm must optimize the following parameters: $\lambda_{\text{OS-SART}}$, $N_{\text{OS-SART}}$, λ_{PDTV} and N_{PDTV} . We found that the quality of the reconstructed images was significantly improved when $\lambda_{\text{OS-SART}}$ varied within a small range. In other words, the OS-SART-PDTV algorithm was more sensitive to this parameter. $N_{\text{OS-SART}}$ is usually set in the range of 50–100 because the improvement rate of the image quality gradually decreases as the number of iterations increases, and after reaching a certain number of iterations, the image quality no longer improves. According to our experimental results, the OS-SART-PDTV algorithm obtained a better reconstructed image after 50 iterations. In addition, the two parameters λ_{PDTV} and N_{PDTV} must be tuned several times to obtain high-quality reconstructed images. When reconstructing the projection data of a new sample, the two parameters λ_{PDTV} and N_{PDTV} must be readjusted.

5 Conclusion

NCT has proven to be a highly effective noninvasive measurement method and has been widely applied in nuclear engineering, thermodynamics, and other fields. However, because of the complexity of the NCT application environment, obtaining complete projection data is challenging.

The selection of an efficient sparse-view reconstruction algorithm is directly related to the system imaging quality and reliability of the measurement results. To this end, we propose the OS-SART-PDTV algorithm for sparse-view NCT 3D. The OS-SART-PDTV algorithm is divided into two steps: OS-SART for image reconstruction, and PDTV for solving the total variance using a first-order primal–dual algorithm. Based on the reconstruction results of the projection data obtained from the NCT scan clock model, the OS-SART-PDTV algorithm has a significant advantage over the other algorithms in preserving edge structure, denoising, and suppressing artifacts. Therefore, the OS-SART-PDTV algorithm has great potential for application in NCT 3D reconstruction.

Author contributions All authors contributed to the study conception and design. Material preparation, data collection and analysis were performed by Yang Liu, Teng-Fei Zhu, Zhi Luo and Xiaoping Ouyang. The first draft of the manuscript was written by Yang Liu and Teng-Fei Zhu, and all authors commented on previous versions of the manuscript. All authors read and approved the final manuscript.

Data availability The data that support the findings of this study are openly available in Science Data Bank at <https://doi.org/10.57760/sciencedb.j00186.00110> and <https://cstr.cn/31253.11.sciencedb.j00186.00110>.

Declarations

Conflict of interest The authors declare that they have no competing interests.

References

1. F. Tian, C.R. Geng, X.B. Tang et al., Analysis of influencing factors on the method for determining boron concentration and dose through dual prompt gamma detection. *Nucl. Sci. Tech.* **32**(4), 35 (2021). <https://doi.org/10.1007/s41365-021-00873-3>
2. S.C. Huang, H. Zhang, K. Bai et al., Monte Carlo study of the neutron ambient dose equivalent at the heavy ion medical machine in Wuwei. *Nucl. Sci. Tech.* **33**(9), 119 (2022). <https://doi.org/10.1007/s41365-022-01093-z>
3. Z.W. Lü, G.X. Wei, H.Q. Wang et al., New flexible CsPbBr₃-based scintillator for X-ray tomography. *Nucl. Sci. Tech.* **33**(8), 98 (2022). <https://doi.org/10.1007/s41365-022-01085-z>
4. S.Y. Luo, Y.H. Huang, X.T. Ji et al., Hybrid model for muon tomography and quantitative analysis of image quality. *Nucl. Sci. Tech.* **33**(7), 81 (2022). <https://doi.org/10.1007/s41365-022-01070-6>
5. L. Steinbock, Transmission tomography of nuclear fuel pins and bundles with an electronic line camera system. *J. Nucl. Mater.* **178**(2–3), 277–283 (2008). <https://doi.org/10.1016/j.advwatres.2008.01.022>
6. H. Umekawa, S. Furui, M. Ozawa et al., Application of CT-processing to neutron radiography imaging of a fluidized-bed. *Part. Part. Syst. Char.* **23**(3–4), 272–278 (2006). <https://doi.org/10.1002/ppsc.200601066>
7. L.Y. Liu, X.P. Ouyang, R.L. Gao et al., Latest developments in room-temperature semiconductor neutron detectors: prospects and

- challenges. *Sci. China Phys. Mech. Astron.* **66**, 232001 (2023). <https://doi.org/10.1007/s11433-022-2021-6>
8. P. Andersson, E. Andersson-Sunden, H. Sjöstrand et al., Neutron tomography of axially symmetric objects using 14 MeV neutrons from a portable neutron generator. *Rev. Sci. Instrum.* **85**(8), 085109 (2014). <https://doi.org/10.1063/1.4890662>
 9. N. Takenaka, H. Asano, T. Fujii et al., Application of fast neutron radiography to three-dimensional visualization of steady two-phase flow in a rod bundle. *Nucl. Instrum. Meth. Phys. Res. Sect. A.* **424**(1), 73–76 (1999). [https://doi.org/10.1016/S0168-9002\(98\)01322-9](https://doi.org/10.1016/S0168-9002(98)01322-9)
 10. M. Kureta, H. Tamai, H. Yoshida et al., Development of design technology on thermal-hydraulic performance in tight-lattice rod bundles: v-estimation of void fraction. *J. Power Energy Syst.* **2**(1), 271–282 (2008). [https://doi.org/10.1016/S0168-9002\(98\)01322-9](https://doi.org/10.1016/S0168-9002(98)01322-9)
 11. A.S. Tremsin, S.C. Vogel, M. Mocko et al., Feller, Non-destructive studies of fuel pellets by neutron resonance absorption radiography and thermal neutron radiography. *J. Nucl. Mater.* **440**(1–3), 633–646 (2013). <https://doi.org/10.1016/j.jnucmat.2013.06.007>
 12. P. Andersson, T. Bjelkenstedt, E.A. Sundén et al., Neutron tomography using mobile neutron generators for assessment of void distributions in thermal hydraulic test loops. *Phys. Proc.* **69**, 202–209 (2015). <https://doi.org/10.1016/j.phpro.2015.07.029>
 13. W. Wu, X. Guo, Y. Chen et al., Deep embedding-attention-refinement for sparse-view CT reconstruction. *IEEE T. Instrum. Meas.* **7**, 1–11 (2023). <https://doi.org/10.1109/TIM.2022.3221136>
 14. W. Zhang, Z. Zhou, Z.F. Gao et al., Multiple adversarial learning based angiography reconstruction for ultra-low-dose contrast medium CT. *IEEE J. Biomed. Health Inform.* **27**(1), 409–420 (2023). <https://doi.org/10.1109/JBHI.2022.3213595>
 15. R. Gordon, R. Bender, G.T. Herman, Algebraic reconstruction techniques (ART) for three-dimensional electron microscopy and X-ray photography. *J. Theor. Biol.* **29**(3), 471–481 (1970). [https://doi.org/10.1016/0022-5193\(70\)90109-8](https://doi.org/10.1016/0022-5193(70)90109-8)
 16. P. Gilbert, Iterative methods for the three-dimensional reconstruction of an object from projections. *J. Theor. Biol.* **36**(1), 105–117 (1972). [https://doi.org/10.1016/0022-5193\(72\)90180-4](https://doi.org/10.1016/0022-5193(72)90180-4)
 17. A.H. Andersen, A.C. Kak, Simultaneous algebraic reconstruction technique (SART): a superior implementation of the ART algorithm. *Ultrason. Imaging* **6**(1), 81–94 (1984). [https://doi.org/10.1016/0161-7346\(84\)90008-7](https://doi.org/10.1016/0161-7346(84)90008-7)
 18. E.J. Candès, J. Romberg, T. Tao, Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information. *IEEE T. Inf. Theory* **52**(2), 489–509 (2006). <https://doi.org/10.1109/TIT.2005.862083>
 19. L.I. Rudin, S. Osher, E. Fatemi, Nonlinear total variation based noise removal algorithms. *Physica D.* **60**(1–4), 259–268 (1992). [https://doi.org/10.1016/0167-2789\(92\)90242-F](https://doi.org/10.1016/0167-2789(92)90242-F)
 20. A. Cai, L. Wang, H. Zhang et al., Edge guided image reconstruction in linear scan CT by weighted alternating direction TV minimization. *J. X-ray Sci. Technol.* **22**(3), 335–349 (2014). <https://doi.org/10.3233/XST-140429>
 21. Y. Guo, L. Zeng, J. Wang et al., Image reconstruction method for exterior circular cone-beam CT based on weighted directional total variation in cylindrical coordinates. *J. Inverse-III pose Problem* **28**(2), 155–172 (2020). <https://doi.org/10.1515/jiip-2019-0012>
 22. Y. Wang, Z. Qi, A new adaptive-weighted total variation sparse-view computed tomography image reconstruction with local improved gradient information. *J. X-ray Sci. Technol.* **26**(6), 957–975 (2018). <https://doi.org/10.3233/XST-180412>
 23. H.L. Qi, Z.J. Chen, L.H. Zhou, CT image reconstruction from sparse projections using adaptive TpV regularization. *Comput. Math. Methods Med.* **2015**, 354869 (2015)
 24. D. Zeng, J. Huang, Z. Bian et al., A simple low-dose x-ray CT simulation from high-dose scan. *IEEE T. Nucl. Sci.* **62**(5), 2226–2233 (2015)
 25. H. Jiang, *Computed Tomography: Principles, Design, Artifacts, And Recent Advances* (SPIE, Bellingham, Washington USA, 2009). <https://doi.org/10.1117/3.817303>
 26. P. Blomgren, T.F. Chan, Color tv: total variation methods for restoration of vector-valued images. *IEEE T. Image Process.* **7**(3), 304–309 (1998). <https://doi.org/10.1109/83.661180>
 27. J. Pryce, R. Tyrell Rockafellar, *Convex Analysis* (Princeton University Press, 1970), xviii 451 pp. *Proceedings of the Edinburgh Mathematical Society*, **18**(4), 339–339 (1973). <https://doi.org/10.1017/S0013091500010142>
 28. B. Hunt, O. Kubler, Karhunen-Loeve multispectral image restoration, part I: theory. *IEEE T. Acoustics Speech and Signal Process.* **32**(3), 592–600 (1984). <https://doi.org/10.1109/TASSP.1984.1164363>
 29. E. Esser, X. Zhang, T.F. Chan, A general framework for a class of first order primal-dual algorithms for convex optimization in imaging science. *SIAM J. Imaging Sci.* **3**(4), 1015–1046 (2010). <https://doi.org/10.1137/09076934X>
 30. B. He, X. Yuan, Convergence analysis of primal-dual algorithms for a saddle-point problem: from contraction perspective. *SIAM J. Imaging Sci.* **5**(1), 119–149 (2012). <https://doi.org/10.1137/100814494>
 31. S. Boyd, L. Vandenberghe, *Convex Optimization*. Cambridge University Press. (2004)
 32. A. Chambolle, T. Pock, A first-order primal-dual algorithm for convex problems with applications to imaging. *J. Math. Imaging Vis.* **40**(1), 120–145 (2011). <https://doi.org/10.1007/s10851-010-0251-1>
 33. W. Wu, D. Hu, K. An et al., A high-quality photon-counting CT technique based on weight adaptive total-variation and image-spectral tensor factorization for small animals imaging. *IEEE T. Instrum. Meas.* **70**, 1–14 (2021). <https://doi.org/10.1109/TIM.2020.3026804>
 34. A. Beck, M. Teboulle, Fast gradient-based algorithms for constrained total variation image denoising and deblurring problems. *IEEE T. Image Process.* **18**(11), 2419–2434 (2009). <https://doi.org/10.1109/TIP.2009.2028250>
 35. B. Schillinger, A.E. Craft, A freeware path to neutron computed tomography. *Phys. Procedia* **88**, 348–353 (2017). <https://doi.org/10.1016/j.phpro.2017.06.047>
 36. W. Wu, D. Hu, C. Niu et al., DRONE: dual-domain residual-based optimization network for sparse-view CT reconstruction. *IEEE T. Med. Imaging* **40**(11), 3002–3014 (2021). <https://doi.org/10.1109/TMI.2021.3078067>

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.