Computer Modeling in Engineering & Sciences

Tech Science Press

https://doi.org/10.32604/cmes.2026.085787

REVIEW

Sparse-View CT Reconstruction with Deep Learning: A Comprehensive Survey

Shuaiqi Cheng[1,2], Yuxi Chen[2], Bo Yang[1,*], Legend Zhang[3], Junmin Lyu[3], Guangyu Xu[4] and Chao Liu[5]

[1]School of Automation Engineering, University of Electronic Science and Technology of China, Chengdu, China
[2]Glasgow College, University of Electronic Science and Technology of China, Chengdu, China
[3]Future Tech Institute, Guangzhou Huashang University, Guangzhou, China
[4]School of the Environment, The University of Queensland, Brisbane St Lucia, QLD, Australia
[5]LIRMM, University of Montpellier-CNRS, Montpellier, France

*Corresponding Author: Bo Yang. Email: boyang@uestc.edu.cn

Received: 18 May 2026; Accepted: 28 July 2026; Published: 28 September 2026

ABSTRACT: Computed tomography (CT) is an essential medical imaging technique that produces high-resolution cross-sectional images, but delivers substantial radiation dose to patients. Sparse-view CT (SVCT) reduces radiation dose by decreasing projection views, but causes streak artifacts that degrade image quality. Deep learning has emerged as a powerful tool to address this challenge. Although prospective paired acquisitions for low-current/voltage CT may be constrained by radiation-dose management and clinical workflow considerations, SVCT provides a practical way to construct paired training data through retrospective angular downsampling of full-view projections. This survey provides a systematic review of deep learning–based SVCT, analyzing nearly 70 articles published since 2017. We describe CT imaging fundamentals with differentiable back-projection, propose a taxonomy of reconstruction frameworks (image-domain, sinogram-domain, dual-domain, direct mapping, and enhanced iterative), and review neural network architectures, loss functions and datasets. We further provide open-source implementations of fan-beam forward projection and filtered back projection for dual-domain training. Finally, we discuss future directions and key remaining challenges.

KEYWORDS: Sparse-view CT; deep learning; CT reconstruction; low dose CT; sinogram

1 Introduction

Computed tomography (CT) is an indispensable medical imaging technique that acquires tomographic projections to construct high-resolution cross-sectional images [1,2]. Despite its widespread clinical applications—ranging from trauma assessment to cancer screening—CT examinations deliver substantially higher ionizing radiation doses than conventional radiography, increasing the lifetime risk of cancer [3,4]. To mitigate these risks, three primary low-dose scanning strategies have been developed: reducing tube current/voltage, limited-angle scanning, and sparse-view sampling.

While reducing tube output is straightforward, it severely degrades the signal-to-noise ratio (SNR) due to increased quantum noise [5,6]. The other two strategies reduce the total number of projection views. For a typical fan-beam system, a short-scan (≈200°) trajectory equipped with Parker weighting can achieve artifact-free reconstruction within the object support, though a full 360° scan remains the standard practice in most clinical CT systems [7]. Limited-angle CT restricts this angular range (e.g., to 90°) [8], which often causes severe anatomical distortions. In contrast, sparse-view CT (SVCT) preserves the full angular coverage

but uniformly increases the angular sampling interval. This approach effectively reduces radiation exposure while maintaining global structural context, making it a highly promising low-dose alternative.

SVCT is particularly well-suited for deep learning techniques because paired training data can be constructed by retrospectively downsampling full-view projections. This strategy is related to, but distinct from, the paired normal-dose and simulated low-dose data available in public datasets such as the AAPM LDCT Grand Challenge [9,10]. In SVCT studies, angularly subsampled projections can be generated from full-view projection data while preserving the corresponding full-view reconstruction as a reference. Consequently, SVCT provides a practical framework for supervised learning, although retrospectively generated sparse-view data may not fully reproduce the noise, scatter, detector-response, and motion characteristics of prospective sparse-view acquisitions. However, three critical domain-gap issues must be considered when interpreting models trained on retrospectively constructed paired data. First, the noise characteristics, physical calibration, and detector response (e.g., X-ray scatter, beam hardening) of actual sparse-view acquisitions differ systematically from simulated downsampling. Second, many public datasets (e.g., LIDC-IDRI) only provide reconstructed images without raw projections; deriving sparse-view inputs via an additional FBP step introduces system-specific errors that widen the Sim2Real gap. Third, supervised models may implicitly overfit to the aliasing artifacts of a specific downsampling operator rather than learning a generalizable inverse mapping. These challenges necessitate dedicated domain adaptation and unpaired learning strategies, which are discussed in Section 4.3 alongside recent dose-level prompted [11] and cycle-consistent approaches.

Historically, conventional algorithms have struggled with SVCT. Analytical methods like filtered back projection (FBP) [12] fail under missing measurements, producing severe streak artifacts that obscure low-contrast lesions, as illustrated in Fig. 1.

To address this, iterative methods [13,14] incorporating sparsity priors—such as Total Variation (TV) [15] and dictionary learning [16,17]—were developed to mitigate underdetermination. However, despite improving image quality, iterative methods suffer from prohibitive computational costs and lengthy reconstruction times [18,19].

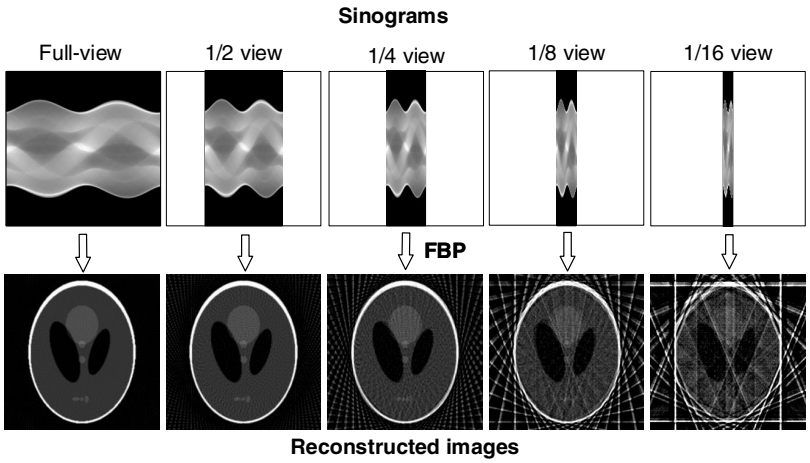


Figure 1: Reconstruction samples of the classic FBP at different view sparsity levels.

Consequently, deep learning has emerged as one of the most widely used and effective approaches for SVCT. By learning complex sample distributions in high-dimensional spaces, deep neural networks can capture useful prior information for resolving ill-posed inverse problems [20].

To ensure a comprehensive and unbiased review of the literature, we conducted a systematic literature search following established guidelines. We queried databases including IEEE Xplore, PubMed, Web of Science, and Google Scholar using keywords such as "sparse-view CT reconstruction", "deep learning", "sinogram interpolation", and "CT inverse problems". We focused on papers published between 2017 and 2026, with particular emphasis on advances from 2020 to 2026. After title/abstract screening and full-text reviews of relevant articles reporting quantitative evaluations, we selected approximately 70 core articles to form the basis of this survey.

While several surveys have reviewed deep learning for CT reconstruction [21,22], they often cover broad topics (e.g., metal artifacts, interior tomography) without a dedicated, systematic focus on SVCT. This work differentiates itself through the following contributions: (1) it introduces a unified, mutually exclusive taxonomy based on the location of deep learning components (image-domain, sinogram-domain, dual-domain, direct-mapping, enhanced-iterative); (2) it provides an SVCT-specific critical analysis of how architectures (e.g., Transformers, diffusion models, unrolled networks) handle artifact modeling and physics fidelity; (3) it open-sources Python implementations of differentiable fan-beam forward projection and FBP to facilitate dual-domain model training; and (4) it covers the latest milestones up to 2026, offering an up-to-date roadmap for researchers.

The remainder of this paper is organized as follows. Section 2 briefly reviews CT imaging principles. Section 3 categorizes the reconstruction frameworks used in recent studies. Section 4 critically discusses key deep learning architectures, loss functions, and datasets. Finally, the main challenges and future directions are summarized in Section 5.

2 Fundamentals of CT Imaging

2.1 Lambert–Beer Law

In CT scanning, the detector-measured X-ray intensity follows the Lambert–Beer attenuation law. For a homogeneous medium of thickness d (Fig. 2a), the attenuation is

$$I = I_0 e^{-\mu d}, \tag{1}$$

where μ is the linear attenuation coefficient and d is the path length.

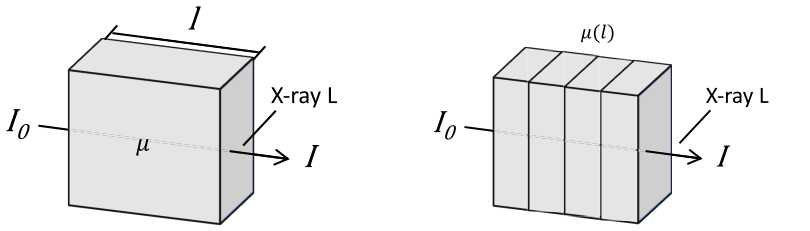
For a non-homogeneous medium (Fig. 2b), the attenuation becomes

$$I = I_0 \exp\left(-\int_L \mu(\ell)\, d\ell\right), \tag{2}$$

where $\mu(\ell)$ is the spatially varying attenuation coefficient along ray path L. Equivalently,

$$-\log\left(\frac{I}{I_0}\right) = \int_L \mu(\ell)\, d\ell, \tag{3}$$

where I/I_0 is transmittance and $\log(I_0/I)$ is absorbance.



a Homogeneous medium of thickness d **b** Non-homogeneous medium with path L

Figure 2: Illustration of the Lambert–Beer law. (**a**) Uniform attenuation with constant μ; (**b**) Spatially varying attenuation described by a line integral. Note that (**a**) corresponds to a special case of (**b**).

2.2 Radon Transform and its Inverse

Let $L_{\theta,s}$ denote a line in 2D space parameterized by (θ, s),

$$L_{\theta,s} = \{(x, y) \mid x\cos\theta + y\sin\theta = s\}, \tag{4}$$

where θ is the projection angle and s is the signed distance of the line to the origin. The line integral of an object $f(x, y)$ along $L_{\theta,s}$ is

$$p(\theta, s) = \iint f(x, y)\,\delta(x\cos\theta + y\sin\theta - s)\,dx\,dy, \tag{5}$$

Here, $p(\theta, s)$ is the sinogram; $p_\theta(s)$ denotes its 1D slice at angle θ. The inverse Radon transform recovers $f(x, y)$ from $p(\theta, s)$. By the Fourier slice theorem, filtered back-projection (FBP) reconstructs $f(x, y)$ as

$$f(x, y) = \int_0^\pi p'_\theta(x\cos\theta + y\sin\theta)\,d\theta, \tag{6}$$

where $p'_\theta(s)$ is obtained by convolving $p_\theta(s)$ with the ramp filter.

2.3 Forward Projection and FBP for Fan Beam Geometry

Fan beam scanning (Fig. 3), standard in practical CT systems, adds geometric complexity to projection and reconstruction. The fan-beam sinogram is $p(\beta, \gamma)$, with coordinate relation

$$s = D\sin\gamma, \quad \theta = \beta - \gamma, \tag{7}$$

where D is the distance from the ray source to the rotation center (usually also the origin of the object's coordinate system).

The rebinning method converts $p(\beta, \gamma)$ to $p(\theta, s)$ via Eq. (7), then applies parallel-beam FBP. However, this interpolation introduces errors [23], particularly severe for sparse-view sinograms.

We now present an algebraic reconstruction method for fan beam geometry without rebinning, expressed as linear matrix equations suitable for embedding in deep learning networks.

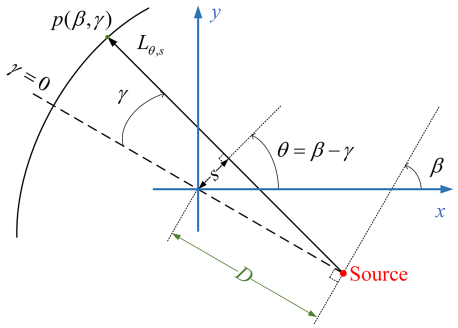


Figure 3: Fan-beam CT scan geometry.

2.3.1 Forward Projection

Let $\mathbf{f}$ be an N-dimensional (column) vector reshaped from all N pixels in image $f(x, y)$, and $\mathbf{p}$ be an M-dimensional vector reshaped from all M measurements in the sinogram $p(\beta, \gamma)$. The forward projection can be expressed as

$$\mathbf{p} = \mathbf{A}\mathbf{f}, \tag{8}$$

where $\mathbf{A}$ is an $M \times N$ system matrix.

Each row of Eq. (8) corresponds to one X-ray. For the i-th ray, $i = 1, 2, \ldots, M$,

$$p_i = \mathbf{a}_i\mathbf{f} = \sum_{j=1}^{N} a_{ij} f_j, \tag{9}$$

where p_i is the i-th measurement of $\mathbf{p}$, $\mathbf{a}_i$ is the i-th row of $\mathbf{A}$, and a_{ij} and f_j denote the j-th elements of $\mathbf{a}_i$ and $\mathbf{f}$, respectively. The elements in $\mathbf{a}_i$ are sparse because each ray intersects only a few pixels.

As shown in Fig. 4, the weight a_{ij} is calculated by

$$a_{ij} = \begin{cases} \dfrac{d_i - d_{ij}}{d_i^2}, & d_{ij} < d_i \\ 0, & \text{otherwise} \end{cases}, \tag{10}$$

where $d_i = \max(|\cos\theta|, |\sin\theta|)$ is the critical distance at which a pixel is hit by the ray, and d_{ij} is the distance from the pixel f_j to the ray. For ray $p_i = L_{\beta,\gamma}$, the angle θ in Eq. (10) follows from Eq. (7), and the distance d_{ij} is

$$d_{ij} = |x' \cos\theta + y' \sin\theta|, \tag{11}$$

where $x' = x - D\sin\beta$ and $y' = y + D\cos\beta$ are the pixel coordinates of f_j after translating the image origin to the ray source. All coordinates and distances in this section are pixel-normalized, i.e., the unit length represents the length of one pixel.

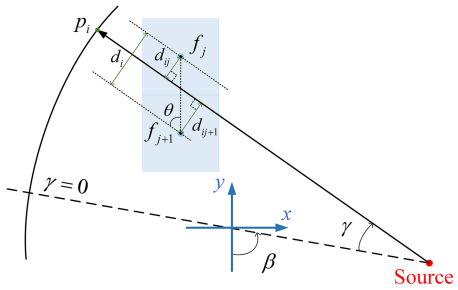


Figure 4: Calculation of the weights (contributions) of pixels f_j, f_{j+1} to projection measurement p_i.

2.3.2 Filtered Back-Projection

Inserting Eq. (7) into Eq. (6) and extending the scan view to $[0, 2\pi)$ gives the fan-beam reconstruction formula,

$$f(x,y) = \frac{1}{2}\int_0^{2\pi} \frac{D}{d'^2} p'_\beta(\gamma')d\beta, \tag{12}$$

where $d'^2 = x'^2 + y'^2$ is the squared distance from the pixel (x, y) to the ray source at view β, and γ' is the detector angle corresponding to the ray passing through (x, y) on the current fan beam that satisfies

$$\sin\gamma' = \frac{x'\cos\beta + y'\sin\beta}{d'}, \ \gamma' \in \left[-\frac{\pi}{2}, \frac{\pi}{2}\right]. \tag{13}$$

The filtered projection signal is

$$p'_\beta(\gamma) = \left(p_\beta(\gamma)\cos\gamma\right) * h'(\gamma), \tag{14}$$

where the cosine weighting $p_\beta(\gamma)\cos\gamma$ is convolved with the fan-beam ramp filter,

$$h'(\gamma) = \frac{\gamma^2}{\sin^2\gamma} h(\gamma). \tag{15}$$

Discretizing Eq. (12): let sampling intervals be $\Delta\beta$ and $\Delta\gamma$, with U and $2V+1$ samples. Denote $p_\beta(\gamma)$ as $p_u(v)$ and $p'_\beta(\gamma)$ as $p'_u(v)$, where

$$\begin{aligned} &\beta = u\Delta\beta, \ u = 0, 1, \ldots, U-1, \\ &\gamma = v\Delta\gamma, \ v = -V, \ldots, 0, \ldots, V. \end{aligned} \tag{16}$$

The discretized reconstruction at (x, y) is

$$\hat{f}_{x,y} = \frac{1}{2} D\Delta\beta\Delta\gamma \sum_{u=0}^{U-1} \frac{p'_u(v')}{d'^2}, \tag{17}$$

where $p'_u(v')$ with $v' = \gamma'/\Delta\gamma$ can be obtained by linear interpolation of the discrete filtered projection $p'_u(v)$. By linear interpolation (Fig. 5),

$$p'_u(v') = (1-\alpha)p'_u(\lfloor v' \rfloor) + \alpha p'_u(\lfloor v' \rfloor + 1), \tag{18}$$

where $\lfloor \cdot \rfloor$ is the floor function and $\alpha = v' - \lfloor v' \rfloor$ is the interpolation weight.

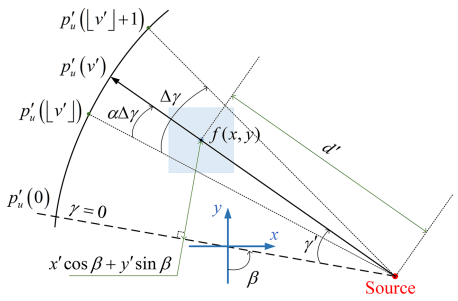


Figure 5: Calculation of $p'_u(v')$ by interpolation.

To embed fan-beam reconstruction into deep learning frameworks, the FBP steps are expressed as matrix-vector operations:

Step 1 γ-wise weighting for sinogram: Each view of the sinogram is element-wise weighted by a pre-computed vector **w**, incorporating the geometry factor.

Step 2 Filtering weighted sinogram: The weighted sinogram is filtered by 1D convolution with a pre-computed kernel $\mathbf{h}'$ that implements the fan-beam ramp filter.

Step 3 Back projection: Each pixel is reconstructed by weighted accumulation of filtered projections, expressed as

$$\hat{\mathbf{f}} = \mathbf{B}\mathbf{p}', \tag{19}$$

where $\hat{\mathbf{f}}$ denotes the reconstructed image, $\mathbf{p}'$ is the filtered sinogram, and **B** is the back-projection matrix.

Using Eqs. (8) and (19), data can be efficiently converted between projection and image domains, enabling physics-informed deep learning reconstruction.

3 Reconstruction Frameworks

In recent years, the advancement of deep learning technologies has increasingly attracted researchers to explore their application in SVCT reconstruction. We classified deep learning-based SVCT methods into five categories corresponding to five reconstruction frameworks: image domain, sinogram domain, dual domain, direct mapping, and enhanced iterative frameworks. The classification criteria for these five frameworks are based on **the location and role of the deep learning component within the overall reconstruction pipeline**, as follows:

- **Image-domain methods**: The deep learning model operates exclusively in the image domain, processing CT images reconstructed by a classical algorithm (e.g., FBP) to remove artifacts and noise. The deep learning model does not interact with raw projection data.
- **Sinogram-domain methods**: The deep learning model operates exclusively in the sinogram (projection) domain, recovering full-view sinograms from sparse-view inputs before reconstruction. The deep learning model does not process CT images.
- **Dual-domain methods**: Deep learning models are deployed in both the sinogram and image domains simultaneously, with explicit data consistency or fusion mechanisms linking the two domains. The pipeline explicitly includes both domain-specific processing stages.
- **Direct-mapping methods**: A single deep learning model directly maps the sparse-view sinogram to the full-view CT image in an end-to-end manner, without explicitly decomposing the process into separate domain-specific stages or iterative optimization steps.

- **Enhanced-iterative methods**: The deep learning model is embedded within an iterative optimization loop that enforces explicit data fidelity to the raw measurements, typically combining a physics-based iterative scheme with a learned regularization prior.

These five categories are mutually exclusive by design: image-domain methods exclude sinogram processing; sinogram-domain methods exclude image processing; dual-domain methods require *both* domain-specific deep learning modules with explicit cross-domain linkage; direct-mapping methods use a single end-to-end model without intermediate domain decomposition or iterative steps; and enhanced-iterative methods embed the network within an iterative optimization that enforces explicit measurement fidelity. In Tables 1 and 2, the label "Enhanced iterative" denotes this same enhanced-iterative framework. Table 1 outlines the representative methods, datasets, and supervision paradigms for each of the five frameworks. To facilitate literature-level comparison, Table 2 reports the specific network models and their performance metrics (PSNR and SSIM) as provided in the original studies. Because the studies use different datasets and numbers of projection views, these values should not be interpreted as a strictly controlled quantitative comparison.

Table 1: Overview of SVCT methods: framework, dataset availability, and supervision type.

Framework	Representative Methods	Dataset Support	Supervision
Image domain	U-net [24–26], CNN [27,28], RNN [29], Transformer [30], Diffusion [31,32]	AAPM, TCIA	Supervised
Sinogram domain	U-net [33,34], Diffusion [35], RNN [36], CNN [37], Transformer [38]	AAPM, TCIA	Sup./Unsup.
Dual domain	U-net [39,40], GAN [41–43], Transformer [44,45], Diffusion [46]	AAPM, TCIA, LDCT-P	Supervised
Direct mapping	AUTOMAP [47], CNN [48], NeRF [49], U-net [50]	AAPM, clinical CT	Supervised
Enhanced iterative	U-net [51,52], CNN [53,54], Transformer [55], Diffusion [56,57]	AAPM, LDCT-P	Iterative/PnP

Table 2: Classification of literature according to the reconstruction framework used. Performance metrics (PSNR and SSIM) are provided for methods with available quantitative results, using the datasets and numbers of projection views reported in the original studies. The entries are intended for literature-level reference rather than strictly controlled quantitative comparison.

Framework	Work	Year	Model	Dataset	Views	PSNR	SSIM
Image domain	[24]	2017	U-net	—	—	—	—
	[25]	2018	U-net	TCIA	—	—	—
	[27]	2018	CNN	—	—	—	—
	[26]	2018	U-net	AAPM	60	38.92	0.927
	[29]	2019	RNN	AAPM	60	38.07	0.971
	[28]	2019	CNN	LIDC-IDRI	120	43.17	0.994
	[58]	2020	U-net	TCIA	—	—	—
	[59]	2021	U-net+RNN	ALERT Task Order 4	—	—	—
	[60]	2021	CNN	—	—	—	—
	[61]	2021	U-net+Attention	TCIA	60	35.82	0.937

(Continued)

Table 2 (continued)

Framework	Work	Year	Model	Dataset	Views	PSNR	SSIM
	[62]	2023	U-net+Attention	TCIA	50	39.37	0.969
	[30]	2023	CNN+Transformer	TCIA	—	—	—
	[31]	2024	Diffusion	AAPM	36	27.23	0.900
	[32]	2024	Diffusion+Transformer	LDCT-P	60	33.41	0.968
	[63]	2024	GAN+Transformer	AAPM	90	39.47	0.900
	[64]	2025	U-net	AAPM, LIDC-IDRI	60	37.018	0.927
	[65]	2026	CNN+Attention	AAPM, C4KC-KiTS [66], NBIA	60	37.77	0.959
	[67]	2025	U-net+Diffusion	AAPM	36	41.78	0.971
Sinogram domain	[33]	2019	U-net	TCIA	—	—	—
	[34]	2019	U-net	—	—	—	—
	[35]	2022	Diffusion	AAPM	92	38.69	0.913
	[36]	2022	RNN	—	—	—	—
	[37]	2023	CNN	LIDC-IDRI, AAPM	24	20.74	0.621
	[38]	2026	Transformer	AAPM	30	29.09	0.777
Dual domain	[41]	2018	GAN	Data Science Bowl 2017	—	—	—
	[68]	2020	CNN	AAPM	—	37.09	0.92
	[42]	2020	U-net+GAN	LDCT-Pset	23	34.90	0.877
	[39]	2021	U-net	AAPM	120	39.789	0.973
	[44]	2022	Transformer	LDCT-P	64	29.739	0.718
	[45]	2022	Transformer	AAPM	24	27.55	0.843
	[43]	2022	U-net+GAN	LIDC-IDRI	10	34.55	0.919
	[40]	2022	U-net+Transformer	AAPM	48	42.67	0.982
	[69]	2022	Transformer	LIDC-IDRI	30	35.41	0.864
	[70]	2022	U-net+RNN	DeepLesion	—	—	—
	[71]	2023	U-net+Attention	AAPM	128	38.82	0.926
	[72]	2023	U-net	COVID-19 pneumonia	128	37.10	0.958
	[73]	2023	Transformer	TCIA	30	41.058	0.963
	[74]	2023	U-net	AAPM	18	35.03	0.918
	[75]	2024	U-net+Transformer	LDCT-P	32	31.61	0.840
	[76]	2024	U-net+GAN+Attention	TCIA	30	45.60	0.980
	[77]	2024	RNN	LDCT-P	32	31.95	0.844
	[78]	2024	U-net+Attention	MIDRC-RICORD (TCIA)	18	30.75	0.820
	[79]	2024	Transformer	AAPM	60	40.11	0.945
	[80]	2024	U-net+Diffusion	NBIA database	—	—	—
	[81]	2024	U-net	—	—	—	—
	[46]	2025	U-net+Diffusion	LDCT-P, TCIA	60	40.61	0.952
	[82]	2025	Transformer	TCIA	15	38.06	0.938
	[83]	2026	U-net+Attention	AAPM, self-constructed lung abnormality CT dataset	72	37.17	0.931
Direct mapping	[47]	2018	CNN	—	—	—	—
	[48]	2019	CNN	Coronary artery/abdomen CT	—	—	—
	[49]	2022	MLP-based NeRF	AAPM	64	36.76	0.872
	[50]	2023	U-net	CIFAR-10	—	—	—

(Continued)

Table 2 (continued)

Framework	Work	Year	Model	Dataset	Views	PSNR	SSIM
Enhanced iterative	[84]	2018	CNN	AAPM	64	38.97	0.949
	[51]	2021	U-net	AAPM	60	37.91	0.939
	[52]	2021	U-net	TCIA	—	—	—
	[85]	2021	U-net	AAPM	30	28.308	0.866
	[53]	2022	CNN	LIDC-IDRI	—	—	—
	[86]	2022	U-net	AAPM	32	40.84	0.967
	[54]	2022	CNN	AAPM	64	41.20	0.954
	[56]	2022	Diffusion	LIDC-IDRI, LDCT-P	23	35.24	0.905
	[87]	2023	U-net+GAN	TCIA	—	—	—
	[88]	2023	U-net+Attention	AAPM	32	36.68	0.944
	[57]	2023	Diffusion	AAPM	60	36.10	0.970
	[89]	2024	U-net	AAPM	60	43.01	0.981
	[90]	2024	CNN	LDCT-P	16	38.98	0.936
	[91]	2024	U-net+Diffusion	AAPM	—	—	—
	[92]	2024	Diffusion	AAPM	60	38.49	0.978
	[55]	2023	Transformer	AAPM, Ellipses Dataset	64	34.05	0.941
	[93]	2024	U-net	AAPM, LDCT-P	32	26.61	0.839
	[94]	2024	Transformer	—	—	—	—
	[95]	2024	Diffusion	AAPM, LIDC-IDRI	90	37.20	0.978
	[96]	2024	Diffusion	AAPM	30	39.60	0.931
	[97]	2024	Convolutional dictionary	LDCT-P	60	35.13	0.918
	[98]	2025	CNN	CHAOS [99], Lung-PET-CT-Dx [100]	—	—	—
	[101]	2025	Diffusion	AAPM	18	37.05	0.941
	[102]	2025	Transformer	AAPM	32	42.036	0.979
	[103]	2025	MLP (Implicit neural representation)	AAPM, COVID-19 [104], CMB-CRC [105]	—	—	—
	[106]	2025	U-net+Attention	AAPM	64	42.53	0.965
	[107]	2026	CNN	—	—	—	—
	[108]	2026	Diffusion+Transformer	AAPM, COCA [109], XCAT [110], etc.	55	40.19	0.963

The conference and journal reports of RegFormer describe the same method and closely related experimental settings. To avoid counting the same method twice, this review retains only the journal article [55] in the quantitative summary and bibliography.

3.1 Image-Domain Refinement

Image-domain methods, also known as image post-processing methods, improve the quality of reconstructed images by removing artifacts and noise from the initial CT images reconstructed from the raw sparse-view sinograms. Although traditional image filtering methods have been widely used for CT reconstruction to suppress noise [111,112], they often fall short in effectively suppressing the globally distributed, complex streak artifacts inherent in SVCT images (as shown in Fig. 1). Deep learning models, with their strong representation learning capability, can map noisy sparse-view images to clear full-view images end-to-end, significantly improving the reconstruction quality when trained on large, high-quality

datasets. Compared to traditional methods, deep learning-based image-domain approaches benefit from advances in the computer vision community, allowing the direct application of state-of-the-art visual task models. Image-domain methods also eliminate the need for projection data processing or specific scanning equipment, making them widely applicable.

The basic framework for image-domain reconstruction is shown in Fig. 6, which starts with the application of a classical reconstruction algorithm, e.g., FBP, to obtain a noisy initial CT image from the sparse-view sinogram. The initial CT image is then refined using a neural network to obtain the final CT image. Various deep learning models have been applied to image-domain refinement. Zhang et al. [25] proposed a U-shaped convolutional network to refine initial CT images, which combines DenseNet and deconvolution techniques, earning it the name DD-Net. Han and Ye [26], inspired by the deep convolutional framelet theory, proposed a tight frame U-net (TF U-net), which integrates an orthogonal wavelet frame into the U-net architecture to improve the recovery of high-frequency information in CT images. In [60], Xie and Yang proposed a feature fusion residual network (FFRN) to remove artifacts from SVCT images based on residual blocks and residual skip dense blocks (RSDB), a variant of residual dense block.

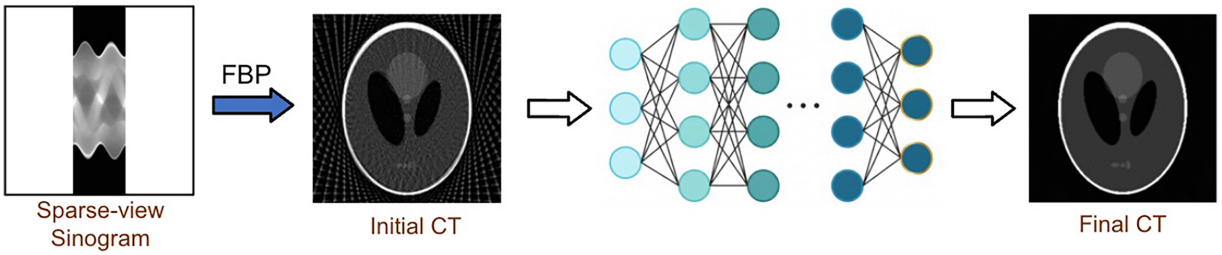


Figure 6: Image-domain reconstruction framework.

In addition to the above methods, there is a special class of image-domain methods that first back project each view vector of the sparse-view sinogram separately to produce a stack of single-view back-projection 'images' (see Fig. 7) and then feed them into an image-domain network for final CT reconstruction. The stacked single-view images can be regarded as a set of multi-channel feature maps of the initial CT image, containing directional projection features and image-domain spatial information, and are particularly suitable as input to image-domain networks for further inference. In [27], Ye et al. proposed a deep back-projection model, which uses a 20-layer convolutional neural network to process the stacked back-projection maps to recover a CT image. To improve inference efficiency, Li et al. [59] proposed a recurrent stacked back-projection model (RSBP), which uses a convolutional long short-term memory (LSTM) to process the back-projection maps sequentially before a U-net post-processing module. In [30], Liu and Sajda further proposed an unsupervised back-projection framework, which uses a spatial transformation projector to re-project the reconstructed image from back-projection maps into the sinogram domain for comparison with the input sinogram, thus achieving self-supervised training.

Image-domain methods can use the representational power of deep learning to model artifact distributions, preserving image detail while reducing artifacts and noise. However, image-domain methods do not work directly with raw projection data and are highly dependent on the initial CT images as input, which are usually reconstructed by the FBP from a sparse-view sinogram and contain very complex artifacts. This post-processing mode tends to be inefficient, especially in highly sparse cases. In addition, projecting the CT image output from an image-domain model back into the sinogram domain may produce a sinogram that is inconsistent with the raw measurements. This lack of projection-domain constraints may affect the reconstruction accuracy.

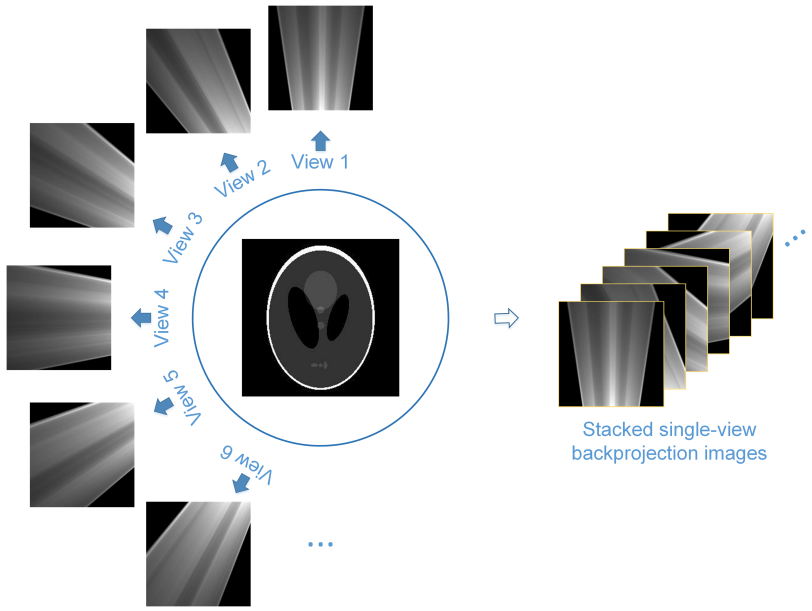


Figure 7: Stacked single-view back projection.

3.2 Sinogram-Domain Recovery

Sparse view sampling in the projection domain is a direct cause of CT image degradation. Accordingly, the most natural solution is to recover full-view sinograms before reconstruction, known as the sinogram-domain pre-processing methods. Traditional interpolation techniques, including linear interpolation and principal component analysis (PCA), have been used to fill in missing projection views. However, their performance is limited, particularly with highly sparse protocols.

To improve the quality of SVCT reconstruction, deep learning techniques were also introduced into the projection domain to recover full-view sinograms. As shown in Fig. 8, sinogram-domain models process projection data directly (predicting missing views and suppressing measurement noise), thus blocking error propagation into the image domain. By learning manifold distribution of sinograms and incorporating tailored network architectures and loss functions, deep learning models can and do excel at using small amounts of projected data to infer missing data. When properly designed, sinogram-domain models can be aligned with the physical principles of CT imaging, inferring full-view sinograms that are consistent with both the acquired measurements and known physical priors. New image generation techniques, such as diffusion models, can further improve the inference of full-view sinograms, even in ultra-sparse cases.

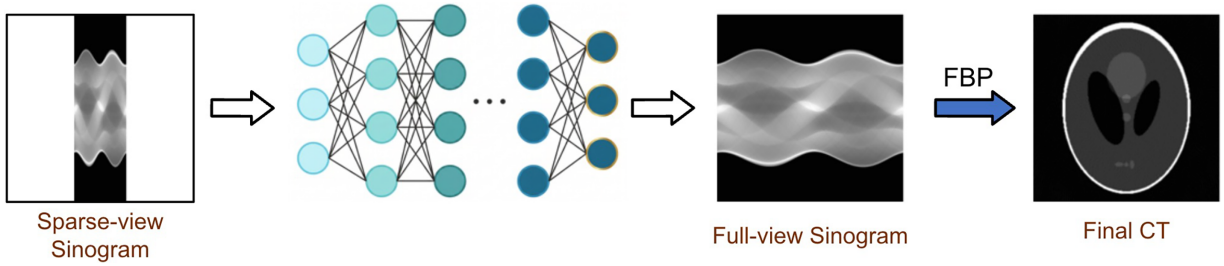


Figure 8: Sinogram domain reconstruction framework.

Depending on the form of the input data, sinogram-domain models can be categorized into interpolation and refinement types. The interpolation type directly feeds the observed views into a deep-learning model to estimate missing views. The refinement type first obtains a pseudo-full sinogram using traditional interpolation or reprojection techniques and then refines it using a deep learning model.

Lee et al. [33] proposed a direct interpolation network to synthesize missing projection data, which was trained on sinogram patches to reduce memory and computational costs. The synthesized data were reconstructed using FBP to obtain the final CT images. In [37], Guo et al. proposed a lightweight interpolation network that can be directly appended with the FBP layer to perform the reconstruction task or incorporated into a dual-domain reconstruction framework as a sinogram-domain preprocessing module. Because the size of the input and output data can be minimized, interpolation-type networks allow for minimalist structures, and training and inference efficiency is usually high, but cannot be adapted to SVCT tasks with different sparse levels.

Refinement-type networks, on the other hand, abandon the obsession with minimalism, working on *full-size* sinograms but adapting to any sparse level. In [35], Xia et al. introduced a patch-based denoising diffusion probabilistic model (DDPM) for SVCT that was trained in an unsupervised mode based on patches extracted from full-view sinograms, eliminating the need for paired full-view and sparse-view sinograms. In the inference stage, the Radon transform is first applied to the initial CT images reconstructed by the FBP to generate pseudo-full-view sinograms, which are segmented into patches for reverse diffusion processing. The processed patches are finally recombined into complete sinograms and reconstructed using the FBP. This patch-based framework allows for parallel processing of large datasets, making it well suited for complex clinical reconstruction tasks. In [57], Guan et al. also proposed a sinogram-domain generative model, which introduces a multi-channel strategy into the score-based generative model (SGM) to optimize the learning of prior distributions for full-view sinograms, and embedded projection data consistency terms into the inverse diffusion process to ensure that the iteratively recovered sinograms are always consistent with the original measurements.

Compared to image-domain methods, sinogram-domain methods are typically more efficient because they can process raw measurements directly. However, since they do not act on the CT image, sinogram-domain models are dependent on the device and its scanning configuration, and their inference biases on the final CT image are complex and difficult to manage. Accordingly, sinogram-domain models are less commonly used alone and are often used in conjunction with image-domain models in a dual-domain reconstruction framework.

3.3 Dual-Domain Reconstruction

Researchers have proposed dual-domain methods that integrate both the image and the projection domains. By simultaneously applying fidelity constraints in the sinogram and image domains, dual-domain methods effectively utilize both geometric information from the projection data and detailed features from the initially reconstructed images. Compared to single-domain methods, dual-domain methods typically show superior performance in recovering fine details and reducing artifacts. Dual-domain methods often use two architectures, serial and parallel.

3.3.1 Serial Architecture

As shown in Fig. 9, the serial architecture cascades the sinogram- and image-domain models. A full-view sinogram is first recovered from the sparse-view input using a sinogram-domain model, then

reconstructed into CT images using algorithms such as FBP or FDK, and finally refined by the image-domain model to remove artifacts and improve quality.

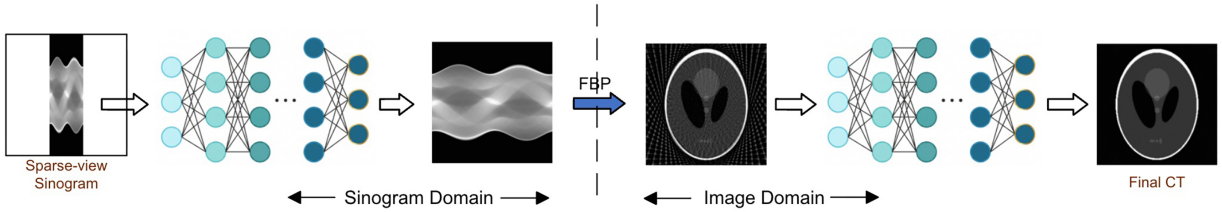


Figure 9: Dual-domain serial architecture.

In [68], Wang et al. proposed an end-to-end dual-domain network that progressively addresses SVCT reconstruction through three modules: projection data denoising and interpolation, sinogram-to-CT conversion, and CT image refinement. In the sinogram-to-CT conversion module, the filtering and back-projection steps of the FBP algorithm are implemented in the form of differentiable one-dimensional convolution and sparse matrix operations, which allow back propagation of reconstruction losses from the image domain to the sinogram domain. Cheslerean-Boghiu et al. [72] proposed WNet, which enhances the dual-domain serial reconstruction framework by introducing a trainable reconstruction layer. Both the sinogram and image-domain models are built on the U-net architecture, and an FBP module with trainable filter coefficients is sandwiched between them. In [81], Sun et al. proposed an extended dual-domain serial model that includes three stages. In stage 1, the initial image reconstructed by FBP is denoised in the image domain using the classical DDNet [25]; in stage 2, the denoised image is transformed into the sinogram domain by forward projection, and an efficient CNN (Sino-Net) is used for sinogram-domain enhancement; and in stage 3, the enhanced sinogram is reconstructed by FBP, and the reconstructed result is concatenated with the output of stage 1 and then fed into a lightweight image-domain network for final refinement.

3.3.2 Parallel Architecture

The dual-domain parallel architecture typically first reconstructs an initial CT image from the sparse-view sinogram using a classical algorithm such as FBP, and then simultaneously processes both the sparse-view sinogram and the initial CT image in the sinogram and image domains, respectively, using two parallel pipelines. The features in the two pipelines can be shared and integrated through a specially designed interaction mechanism, and then fused in the image domain to obtain the final reconstructed image. A typical structure of the parallel architecture is shown in Fig. 10. The dashed arrows in the figure indicate an optional feature fusion scheme, where the CT image output from the image-domain model can also be re-projected into the sinogram domain to be fused with the output of the sinogram-domain model.

Gao et al. [71] proposed an attention-based dual-branch network architecture (ADB-Net) with two parallel branches to extract features from sinograms and CT images. The sinogram branch focuses on capturing global features, while the CT branch extracts multi-scale features from the reconstructed image. These features are then integrated using an attention-based fusion module to optimize the quality of the reconstruction. This approach effectively reduces artifacts while preserving fine image details. In [82], Li et al. proposed a tri-domain SVCT model, called TD-Strans, which adds an additional Fourier-domain filling branch to the parallel sinogram- and image-domain branches to infer unobserved Fourier coefficients. Shao et al. [83] proposed MDPRNet, a parallel dual-domain network with a multi-stage progressive architecture: early stages extract multi-scale features using encoder-decoders, while the final stage preserves fine spatial details via a single-scale subnetwork. A cross-stage feature adapter (CFA) with learnable attention gates

facilitates feature fusion across stages. Importantly, a multi-view synergistic training strategy (MSTS) groups sparse-view data into ultra-sparse and sparse subsets, enabling a single unified model to adapt to diverse sparsity levels without separate training.

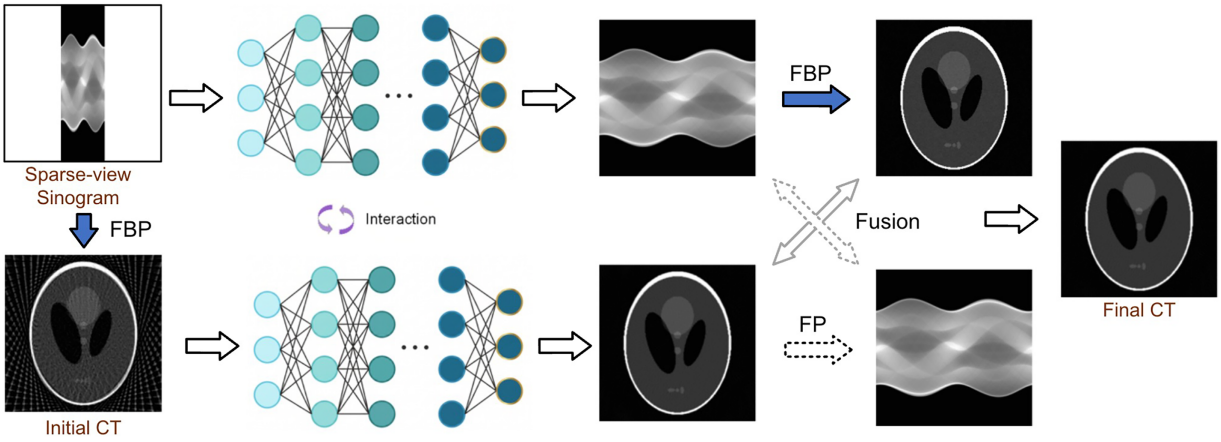


Figure 10: Dual-domain parallel architecture.

In general, dual-domain methods provide better results in SVCT reconstruction than single-domain methods because they can combine the advantages of sinogram and image-domain processing to remove artifacts and noise and restore image detail. In addition, dual-domain methods can improve inter-domain consistency through joint loss constraints and cross-domain feature interaction, avoiding the inter-domain bias that often occurs in a single-domain method. However, the higher computational cost remains a challenge for dual-domain methods, highlighting the need for more efficient training strategies to reduce computation time in future research.

3.4 Direct Mapping

CT image reconstruction is particularly challenging under constrained acquisition conditions, such as SVCT, where limited measurements in the projection domain result in an underdetermined system. As a result, the inverse problem becomes ill-posed, making accurate CT reconstruction difficult using traditional mathematical methods. Deep learning, as a powerful end-to-end high-dimensional information inference tool, is able to directly map from projection measurements to image-domain pixels, thus bridging the gap caused by the ill-posedness. Deep learning-based direct mapping methods have received much attention in recent years. Fig. 11 illustrates the direct mapping architecture for SVCT.

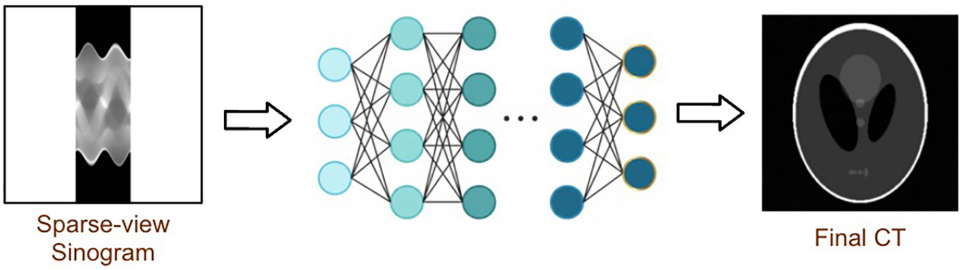


Figure 11: Direct mapping architecture.

In [47], Zhu et al. proposed a generalized applicable image reconstruction framework called AUTOMAP, which for the first time establishes end-to-end cross-domain mappings from sensor data to image data in a purely data-driven manner, demonstrating the great potential of deep learning models in medical image reconstruction. The first three layers of AUTOMAP are fully-connected layers that enable domain transformation, followed by several convolutional layers to recover image detail and structure. AUTOMAP is extensible to reconstruction tasks across multiple imaging modalities and demonstrates superior noise immunity and artifact removal capabilities. However, it has a large number of parameters, resulting in long training and inference times. To improve computational efficiency, Mizusawa et al. [50] introduced stacked U-nets into the direct mapping framework, where the sinogram is first upsampled to the target size and then directly mapped to the CT image through six cascaded U-net modules. The efficiency of cross-domain mapping is improved by replacing the fully-connected layers with fully convolutional layers.

Some of the deep neural networks in the direct mapping framework are inspired by classical reconstruction pipelines in their structural design. In [48], Li et al. proposed an iCT-Net consisting of four cascading components. Although all components are mainly composed of learnable convolutional layers, they can functionally correspond to the cascaded steps of the FBP-based CT imaging pipeline. Some methods learn mappings to the image domain from spaces other than the sinogram domain. Kim et al. [49] proposed a coordinate-based neural representation method to learn patient-specific mappings from the spatial coordinate space to the image domain. This mapping generates a patient-specific prior image that can be reconstructed by a sparse-view forward projection and then a back projection to simulate streak artifact map. Subtracting the artifact map from the real image reconstructed by FBP from the sparse-view sinogram results in a higher quality CT image.

Deep learning-based direct mapping methods tightly integrate data-driven strategies with the SVCT task. Most of these methods do not rely on complex scan geometries and classical imaging knowledge, which simplifies the CT reconstruction process. Supported by a large amount of training data, the end-to-end direct mapping framework is able to learn complex mappings between the projection and image domains and achieve more accurate results than classical physics-driven reconstruction methods. However, as mentioned in [113], the computational and data costs of direct mapping methods are high. Future research needs to focus on reducing algorithmic complexity and training costs to improve their efficiency and feasibility in practical applications.

3.5 Enhanced Iterative Reconstruction

Iterative reconstruction methods treat CT image reconstruction as an optimization problem for unknown variables. By repeatedly switching between forward and backward projections, the CT image is continuously adjusted to minimize the error with respect to the projection data. A general iterative model is formulated as

$$\min_{\mathbf{f}} \frac{1}{2}\|\mathbf{A}\mathbf{f}-\mathbf{p}\|_2^2 + \lambda R(\mathbf{f}) \tag{20}$$

where $\mathbf{A}$ is the forward-projection matrix as in Eq. (8), $\mathbf{f}$ is the CT image to be reconstructed, and $\mathbf{p}$ is the projection data. The first term in Eq. (20) is the data fidelity term, which measures the consistency between the reconstructed image and projection measurements, and the second term $R(\mathbf{f})$ is the regularization term with weight λ, which provides additional prior constraints for SVCT reconstruction to deal with insufficient measurements. The gradient descent algorithm can be used to iteratively update the CT image,

$$\mathbf{f}^{k+1} = \mathbf{f}^{k} - \alpha\mathbf{A}^{T}(\mathbf{A}\mathbf{f}^{k}-\mathbf{p}) - \alpha\lambda\Delta_{\mathbf{f}}R(\mathbf{f}^{k}), \tag{21}$$

where $\mathbf{A}^T$ is usually replaced by the FBP module for better convergence efficiency [55], and $\Delta_{\mathbf{f}} R(\mathbf{f}^k)$ is the gradient of $R(\mathbf{f}^k)$ with respect to $\mathbf{f}$.

Traditional iterative algorithms rely heavily on prior knowledge and parameter tuning to design an appropriate regularization term, which limits their performance in complex cases. To overcome these problems, deep learning techniques have been introduced to replace or complement classical (usually sparse theory-based) regularization tactics, improving not only convergence speed but also generalization. Deep learning-enhanced iterative schemes typically retain the update component corresponding to the data fidelity term, $-\alpha \mathbf{A}^T(\mathbf{A}\mathbf{f}^k - \mathbf{p})$, and replace the update component of the regularization term, $\alpha\lambda\Delta_{\mathbf{f}} R(\mathbf{f}^k)$, with deep neural network inference, as shown in Fig. 12.

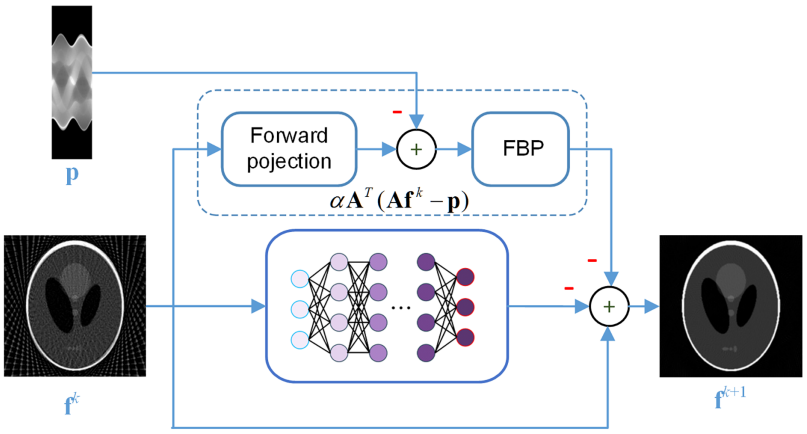


Figure 12: General update framework for iterative models (at step k).

The LEARN framework proposed by Chen et al. [84] computes the update component of the regularization term using multilayer CNNs. Given a predetermined number of iterations, the iterative model can be unfolded into a deep CNN and trained end-to-end with supervision. Xia et al. [55] further improved the iterative model by splitting it into two blocks, one consisting of convolutional layers for local regularization and the other introducing the Swin Transformer for non-local regularization. The self-attention mechanism in the Swin Transformer allows the iterative model to better capture long-range dependencies in CT images. Wang et al. [86] proposed ADMM-SVNet, which combines deep learning with the alternating direction method of multipliers (ADMM) to iteratively solve Eq. (20). ADMM is an optimization framework based on the augmented Lagrangian and alternating minimization. Unlike classical gradient descent, it decomposes the original problem into subproblems involving primal and auxiliary variables, together with dual-variable updates. The traditional sparse transform and inverse transform modules in the ADMM reconstruction framework are replaced by two U-nets in ADMM-SVNet. Kang et al. [97] proposed DCDL-GS, which embeds an interpretable convolutional dictionary learning network with a nonlocal group sparse prior into the iterative reconstruction framework, enhancing both feature representation and model transparency. In [93], Cheng et al. proposed LIR-Net, which further improves deep learning-enhanced iterative schemes by introducing learnable forward and backward projection operators and dual-domain joint optimization and iteration units.

In [52], Wu et al. provided another way to combine deep learning with the traditional iterative reconstruction framework, called the DRONE architecture. It consists of three cascaded components: embedding, refinement, and awareness. The embedding and refinement components use dual-domain deep neural networks for initial reconstruction and residual-based refinement, respectively, to obtain a pair of

data-image priors. The awareness component finally integrates the priors into an iterative optimization framework based on compressed sensing by constructing additional prior constraint terms in Eq. (20).

Deep learning models combined with traditional iterative frameworks can reduce or even eliminate the dependence on training data. Zhang et al. [51] proposed an unsupervised learning method for robust and iterative reconstruction, which introduces a robust and enhanced denoising autoencoding prior (REDAEP) into the iterative framework in Eq. (20) as a regularization term. The learning of REDAEP is self-supervised, requiring only clean and Gaussian noised image pairs, and is enhanced by using a virtual variable augmentation technique and the L_p-norm loss. In [53], Shu et al. proposed an iterative reconstruction method that uses an untrained convolutional neural network as a CT image generator and combines TV regularization to iteratively optimize both the weights and the input latent vector of the generator. Although the generator is not trained, as indicated in [114], the structure of a deep convolutional network itself is sufficient to capture a large number of low-level image statistics priors, known as deep image priors (DIP). By using DIP, Shu's method is completely independent of training data. Xu et al. [98] extended the DIP to the projection domain and proposed the Deep Radon Prior (DRP) framework, which integrates an untrained neural network as an implicit prior into the iterative reconstruction loop and enables gradient feedback between the image and Radon domains across multiple optimization stages. Tian et al. [103] proposed Spener, an unsupervised iterative framework that embeds self-prior features from previous iterations into an implicit neural representation (INR) to constrain the solution space, achieving strong out-of-domain generalization and robustness to noisy sinograms.

Wu et al. [106] proposed a DPMA model that integrates three key innovations: a residual regularization strategy for prior-guided optimization, a multi-scale attention mechanism for joint global-local feature extraction, and a physics-informed consistency module based on range-null space decomposition to enforce projection data fidelity. Dou et al. [107] proposed a coarse-to-fine hierarchical framework that bridges projection and image domains through a FISTA ([115])-based iterative reconstruction stage, where the improved sinogram serves as the data-consistency constraint while learned network parameters guide artifact suppression and convergence, achieving effective reconstruction with as few as six projection views.

Deep learning-enhanced iterative methods combine the nonlinear representations of deep neural networks with the global optimization of traditional iterative schemes to provide an effective solution for SVCT reconstruction, which can more effectively exploit the data sparsity and prior information of the dataset for dual-domain co-optimization, removing artifacts and noise without destroying data consistency. The recently popular generative diffusion techniques bring new power to the 'iterative theory + deep learning' framework. The iterative denoising mechanism of diffusion models is easier to combine with the iterative reconstruction framework (see Section 4.1.4 for more discussion). In the future, with the continuous improvement of deep learning techniques and computational power, deep learning-enhanced iterative methods are expected to improve reconstruction quality while reducing computational time, providing more powerful tools for low-dose CT reconstruction and promoting their wide application in clinical scenarios.

3.6 Emerging Paradigms

Beyond the five core frameworks discussed above, emerging paradigms from the broader computer vision community, such as differentiable rendering and continuous volumetric representations, are influencing SVCT as well. Li et al. [116] introduced a 3D Gaussian representation [117] for SVCT, combining FBP-guided initialization with a differentiable CT projector. This approach achieves greater accuracy and faster convergence than implicit neural representation methods. Subsequently, An and Zhang [118] extended Gaussian splatting to dynamic CT using a self-supervised prior transfer framework for 4D CT reconstruction, and Zhang et al. [119] proposed using compact-support cardinal B-splines instead of Gaussian kernels

to avoid truncation errors and reduce high-frequency noise. Yang et al. [120] took a different approach to SVCT, focusing on acquisition rather than reconstruction. They proposed a multi-task framework that learns task-specific, sparse-view sampling strategies tailored to different scan types and downstream clinical tasks.

4 Deep Learning Techniques, Loss Functions and Datasets

With the widespread adoption of deep learning in medical imaging, various architectures, learning paradigms and supporting techniques, many of which originated in computer vision or natural language processing (NLP), have been adapted for the SVCT task. These technical components operate at different conceptual levels and are often combined within a single reconstruction framework. For instance, a diffusion model (a learning paradigm) may employ either a convolutional neural network (CNN) or a transformer as its backbone (an architectural model). This section therefore reviews the key deep learning techniques (Section 4.1), loss functions (Section 4.2), and datasets (Section 4.3) that underpin current SVCT research. By examining these components in a structured yet flexible manner, we aim to highlight the design choices that influence reconstruction performance.

4.1 Key Deep Learning Techniques

Deep learning–based SVCT methods rely on various network architectures and learning paradigms, most of which have been extensively validated for conventional vision tasks. These elements are not mutually exclusive: a network may have a specific architecture—such as a convolutional neural network (CNN) or a transformer—and be trained using a particular learning paradigm—such as a generative adversarial network (GAN) or a diffusion model. Consequently, a single work may blend multiple techniques and appear in more than one category below. This subsection reviews five central technical threads of SVCT reconstruction: CNNs, Transformers, generative adversarial networks (GANs), diffusion models, and RNNs. For each, we discuss the underlying principles, common variants, and applications for overcoming challenges in sparse-view CT.

4.1.1 CNN-Based Methods

CNNs are perhaps the most widely used network architectures in various image and vision tasks. By stacking multiple convolutional and pooling layers, CNNs are able to extract multi-scale image features from input images, enabling efficient hierarchical processing similar to the human visual system. Its core computational unit, the convolutional layer, dramatically reduces the number of parameters and computational complexity through local receptive fields and weight sharing, making CNNs particularly suitable for handling high-resolution image data. Since both sinograms and CT slices can be considered as 2D 'images', CNNs are widely used in both domains of SVCT to recover missing projection views or to remove image artifacts and noise.

Specifically, CNNs are particularly well-suited for SVCT reconstruction due to their ability to model local texture patterns and multi-scale features, which directly correspond to the spatially localized streak artifacts and anatomical structures in CT images. However, their inherent local receptive fields pose a challenge: SVCT artifacts are globally distributed (as shown in Fig. 1), so purely local processing may not capture long-range artifact correlations. To address this, SVCT-specific CNN designs incorporate residual learning, multi-scale architectures, and attention mechanisms to extend the effective receptive field and prioritize artifact-relevant features. The residual learning scheme, in particular, has proven highly effective for SVCT because the artifact patterns (the residual to be predicted) are structurally simpler than the underlying anatomical content, allowing the network to focus its capacity on artifact suppression. Fig. 13 shows the general U-net architecture used for image-domain CT reconstruction.

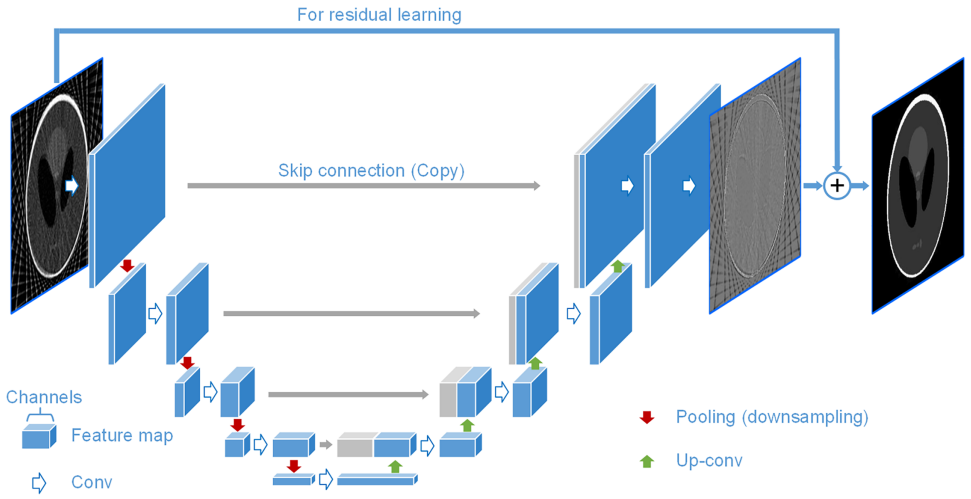


Figure 13: General U-net architecture for (image-domain) CT reconstruction.

The fully convolutional network (FCN) architecture is often adopted by CNN-based SVCT models, allowing them to be fed with inputs of arbitrary size. U-net [121], as a typical FCN, is often used as the backbone of SVCT models, consisting of two primary components: an encoder path that performs multi-scale feature extraction through cascaded convolutional and pooling (downsampling) operations, and a decoder path that progressively restores the spatial resolution of feature maps through upsampling and convolutional operations. Skip connections are used between the encoder and decoder paths to allow more efficient use of multi-scale features (especially low-level features) from the encoder in the decoding phase. This U-net-based architecture can operate in the sinogram and/or image domain. The architecture typically uses a residual learning scheme, where the U-net actually predicts a noise map (the difference between the input and the ground truth), and it is summed with the input to obtain the final CT image (or sinogram). Jin et al. [24] showed that the residual learning scheme significantly improves the reconstruction performance of the U-net model.

Many studies in the literature extend the U-net-based CT reconstruction architecture by integrating more advanced modules, such as residual blocks and attention mechanisms, to improve performance. In [62], Chan et al. proposed an attention-gated U-net (Att-U-net) for image-domain artifact correction by incorporating the pre-trained ResNet50 and gated attention into the U-net architecture. The pre-trained ResNet50 has strong feature representation capabilities and can speed up network convergence, and the residual connection used in each internal block of ResNet can effectively solve the degradation problem of deep models. The use of attention gates can improve the compatibility of local and global features in the U-net pipeline, thus optimizing the learning of salient features relevant to specific tasks. In [74], Ma et al. proposed a frequency-band-aware and self-guided network, called FreeSeed, for SVCT reconstruction, which uses two subnetworks in the image domain, FreeNet and SeedNet, for artifact removal and detail refinement, respectively. FreeNet introduces fast Fourier convolution [122] into the U-shaped denoising architecture to learn globally distributed artifacts. SeedNet consists of residual Fourier convolution blocks that can provide supervision signals to help FreeNet refine the image details contaminated by the artifacts. Unlike conventional convolution blocks that have local receptive fields, the Fourier convolution blocks have global receptive fields that help to learn globally distributed streak artifacts in the SVCT images.

Other FCNs have been designed for SVCT, often eschewing the U-shape for a simpler structure and more efficient computation. The sinogram-domain interpolation network proposed in [37] consists of only

four convolutional layers without pooling to achieve minimal structure and extreme inference speed. The FFRN [60] mentioned in Section 3.1 also uses a non-U-shaped FCN architecture, whose backbone is built on residual blocks and their variants.

Similarly to other visual tasks, CNNs are currently the most widely used deep learning networks in SVCT, especially for the image-domain reconstruction. Generalized visual task architectures such as U-net are widely used, and various enhancement techniques such as residual blocks, attention mechanisms, and hybrid with other network structures (e.g., Transformers or RNNs) are integrated. Such SVCT models built on generic architectures are usually easier to train for convergence and can even be well initialized by transfer or cross-domain learning; however, generic structures are usually more complex and redundant in parameters and computational units, which can reduce the efficiency of SVCT reconstruction models and their generalization on small medical image datasets. Therefore, the development of elegantly structured networks that are highly adapted to the SVCT task remains a worthwhile pursuit.

4.1.2 Transformer-Based Methods

Transformers address a fundamental limitation of CNNs in SVCT: the globally distributed streak artifacts caused by angular undersampling cannot be effectively modeled by local convolutions alone. The self-attention mechanism enables Transformers to capture long-range dependencies across the entire CT image or sinogram, which is essential for modeling the global structure of artifacts that span large spatial distances. However, applying Transformers to SVCT introduces unique challenges. In the sinogram domain, tokenizing the projection data into patches disrupts the physically meaningful angular ordering of views; ViewTrans addresses this by treating each projection view as a token. In the image domain, the high spatial resolution of CT images (typically 512 × 512 or larger) makes the quadratic complexity of self-attention prohibitively expensive, motivating the use of window-based architectures like Swin-T. Despite these challenges, Transformers' ability to model global context makes them particularly promising for SVCT reconstruction, especially for preserving fine anatomical details that local methods may blur.

Recognizing the limitations of local receptive fields of CNNs in capturing long-range dependencies, many studies have built the backbone of SVCT models with attention-based networks represented by Transformer [123]. Transformer, as a novel architecture based on the self-attention mechanism, was originally designed for sequence-to-sequence tasks in NLP. In recent years, it has been extended to vision tasks, spawning many variants, such as Vision Transformer (ViT) [124], Swin Transformer (Swin-T) [125], and Pale Transformer [126], and bringing a major impact on the vision community, where CNNs originally dominated. This impact has spilled over into the SVCT task, and researchers have sought to exploit the ability of Transformers to capture long-range features and global context to deal with the global artifacts and noise caused by sparse-view scanning. Fig. 14 shows the architecture of DDPTransformer.

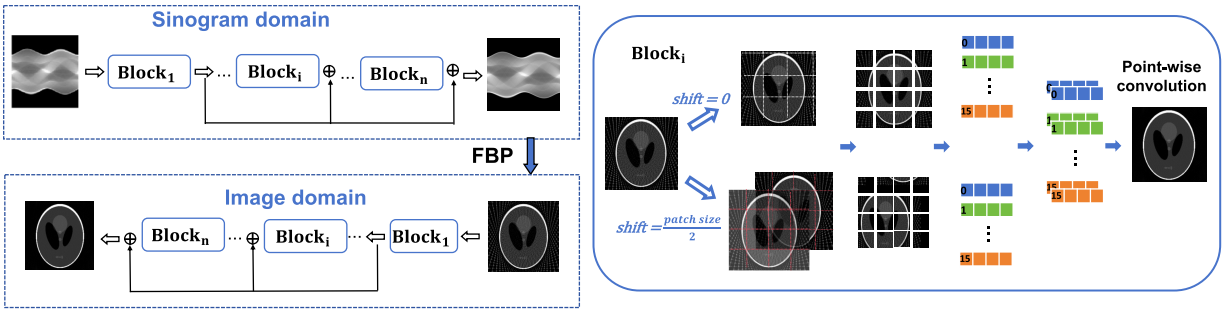


Figure 14: The architecture of DDPTransformer.

In [44], Li et al. proposed DDPTransformer, which constructs sinogram- and image-domain subnetworks by stacking patch-based Transformer blocks in the dual-domain framework shown in Fig. 14. Within each Transformer block, two parallel self-attention branches are formed based on two patching strategies, the multilayer perceptron (MLP) operations are replaced by the newly designed layer-conv-layer (LCL) module, and the global features from the two branches are fused with a point-wise convolution layer. In addition to the dual-domain framework, Transformers have also been incorporated into the iterative framework. The RegFormer [55] discussed in Section 3.5 uses non-local iterative blocks based on the Swin-T along with local iterative blocks based on the CNN for a more comprehensive joint regularization of CT images. In [94], Wang et al. proposed a hybrid-domain integrative transformer iterative network (HITI-Net) for sparse-view spectral CT reconstruction, where Swin-T blocks are used to remove global streak artifacts in the material domain and improve the accuracy of material decomposition. Furthermore, Chen et al. [38] introduced ViewTrans, a vanilla physics-informed Transformer that treats each projection view as a token instead of dividing the sinogram data into patches.

Transformers based on self-attention outperform CNNs in global feature extraction and long-range dependency modeling, which facilitates detail preservation and global artifact removal in the SVCT task. However, Transformers used in the sinogram domain mostly tokenize the input sinogram by patching, which does not naturally match the spatial arrangement of the projection data. In addition, the high computational cost and training data requirements of multi-head self-attention pose challenges to its application in SVCT.

4.1.3 GAN-Enhanced Methods

GANs offer a unique advantage for SVCT: the discriminator provides a high-level (semantic) evaluation of reconstruction quality that goes beyond pixel-wise losses like MSE. This is particularly valuable for SVCT because the primary goal is not pixel-perfect accuracy but rather the perceptual fidelity of anatomical structures and the suppression of visually distracting streak artifacts. The adversarial loss acts as a learned perceptual metric that naturally complements the data fidelity term, helping the generator preserve fine structural details that MSE-based losses tend to smooth out. However, GANs introduce specific challenges for SVCT: (1) medical image datasets are typically small and anatomical variations are limited, making the discriminator prone to overfitting and mode collapse; (2) the high dynamic range and soft-tissue contrast requirements of CT images demand fine-grained adversarial training that is sensitive to hyperparameters; and (3) the ill-posed nature of SVCT means that the generator may 'hallucinate' anatomically implausible structures that still fool the discriminator. These challenges make training stability a critical concern when applying GANs to SVCT.

The GAN architecture consists of a generator G and a discriminator D, where D is trained to discriminate between real and generated samples, while G learns to generate as realistic samples as possible that can fool D. Let $\hat{\mathbf{x}} = G(\mathbf{z})$ denote the sample generated by G, the generative-adversarial game is formulated as

$$\min_G \max_D \mathbb{E}_{\mathbf{x}\sim P_{data}}[\log D(\mathbf{x})] + \mathbb{E}_{\hat{\mathbf{x}}\sim P_G}[\log(1 - D(\hat{\mathbf{x}}))], \tag{22}$$

where $D(a)$ is the probability that a is true, as determined by the discriminator.

The main value of GAN techniques in SVCT is that the discriminators can provide additional adversarial constraints on the estimated CT images, thus avoiding over-smoothing caused by over-reliance on conventional losses such as mean squared error (MSE) and preserving as much image detail as possible. Assuming that G is an SVCT model and D is a trained discriminator, then the adversarial loss

$$L_{Adv} = \log(1 - D(\hat{\mathbf{x}})) \tag{23}$$

can be interpreted as a high-level (semantic) evaluation of the inferred results of G.

In [41], Zhao et al. added a discriminator at the end of the dual-domain serial framework (see Fig. 9) to compute adversarial loss and improve the training of the image-domain refinement model. The dual-domain iterative model, DRONE [52], discussed in Section 3.5, uses GAN with Wasserstein distance in the image domain to provide adversarial loss as a complement to MSE to preserve details and features in the inferred image priors.

In [42], Wei et al. proposed a 2-step SVCT model, SIN-4c-PRN, which uses GANs in both the sinogram and image domains, as shown in Fig. 15. In the sinogram domain, two patch discriminators are used to focus on global and local descriptions of the generated sinograms, respectively. Perceptual losses computed from the multi-scale features extracted by the discriminators are introduced to enhance the stability of the generative adversarial training.

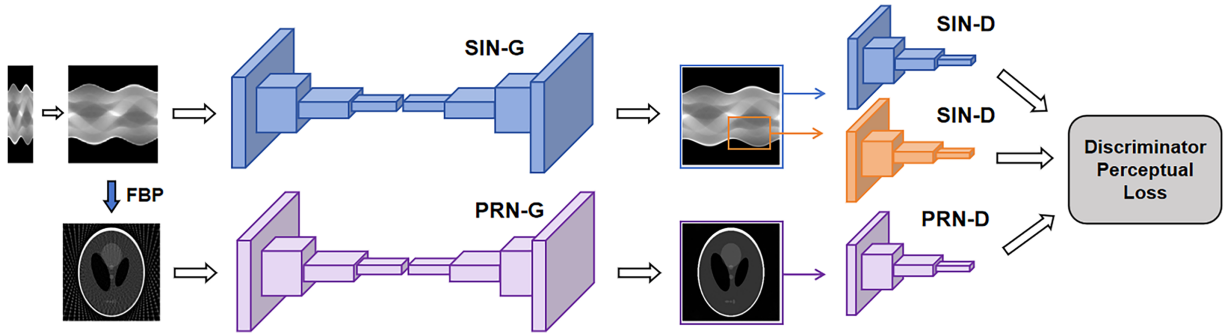


Figure 15: Architecture SIN-4c-PRN model.

In general, introducing adversarial loss into SVCT can optimize the training of the reconstruction model to obtain finer outputs, but GANs suffer from training stability problems, especially on medical image datasets, which are typically smaller than those used in conventional vision tasks. Accordingly, how to regularize and terminate the generative-adversarial game in time to obtain an accurate semantic loss metric for the estimated CT images is a concern when applying GANs to improve SVCT reconstruction.

In addition to using adversarial loss as a semantic regularization term, GANs provide a robust solution to ill-posed inverse problems, including those encountered in SVCT. A well-trained generator can approximate the manifold distribution of full-view sinograms or CT images, effectively serving as a low-dimensional model of the target data representation. Within the constraints of data consistency, the latent space (i.e., the generator's input space) can be explored to identify a low-dimensional latent code that produces a reconstructed image or sinogram that is aligned with the learned manifold and sparse-view measurements. Such latent-space optimization could provide a systematic way to address underdetermination in SVCT, but its application to SVCT reconstruction remains under-explored.

4.1.4 Diffusion-Based Methods

Diffusion models bring a fundamentally different capability to SVCT: the ability to learn the manifold distribution of normal CT images from large datasets, and then generate fine-detail reconstructions through iterative denoising. This is particularly powerful for SVCT because it provides a strong image prior that can compensate for the severe information loss caused by angular undersampling. The iterative denoising process naturally aligns with the iterative nature of classical reconstruction, making diffusion models easy to integrate with physics-based components (e.g., data consistency layers, FBP operators). However, the standard Gaussian diffusion process is a poor physical match for SVCT: the forward diffusion adds noise in an image-space manner, whereas SVCT degradation is a deterministic, physics-driven angular subsampling

in the Radon domain. Several works have addressed this mismatch by redesigning the forward process as sinogram-domain degradation (e.g., CT-SDM using cold diffusion, CvG-Diff using artifact degradation operators), making the model better aligned with the actual physical acquisition. A key remaining challenge is the tension between the generative nature of diffusion models and the need for strict measurement fidelity: during reverse diffusion, the model may generate anatomically plausible but physically incorrect details that violate the sparse-view measurements. Ensuring data consistency without sacrificing the quality gains from the learned prior remains an open research question. Additionally, the iterative sampling process of diffusion models is computationally expensive, motivating research into accelerated sampling strategies.

Diffusion models, which have recently gained wide attention in various vision tasks, have excellent image generation and denoising capabilities. The pipeline of diffusion models consists of a pre-defined forward diffusion process and a trainable reverse diffusion process. In the forward process, noise is progressively added to the image, pushing the data toward a prior noise distribution. In the reverse process, the model learns to gradually recover the original image from the noise, also called denoising or sampling. With their excellent performance in routine vision tasks, diffusion models have been transferred to medical imaging.

Both DDPM and SGM, the two types of diffusion models (although they can be unified by a general stochastic differential equation (SDE) framework [127]), have been used for SVCT reconstruction. In the SVCT task, diffusion models can learn the prior distributions of normal images and iteratively *generate* fine-detail CT images through reverse diffusion. However, how to incorporate data fidelity constraints into the reverse process to prevent the generative model from fantasizing image details that match the manifold distribution of normal CT but are inconsistent with the raw projected measurements is a key concern of diffusion-based reconstruction methods.

In [56], Song et al. proposed a generalized SGM framework to solve inverse problems in medical imaging, including SVCT and under-sampled MRI. In the framework, an unconditional SGM is first trained on normal medical image datasets to capture their prior distribution, and an iterative sampling algorithm is provided to generate reconstructed images, which can be viewed as a fusion of the conventional iterative technique (see Eq. (21)) with the iterative rule of diffusion models. By incorporating the given physical measurement model into the unconditional SGM sampling process, Song's algorithm makes the generated images consistent with both the observed measurements and the priors learned from normal images. Following Song's work, Wu et al. [91] proposed a multi-channel optimization generative model (MOGM) for stable SVCT reconstruction. To improve the efficiency of the data fidelity module and provide more accurate guidance for the predictor-corrector sampling process [127], a multi-channel fusion scheme is designed that uses the same trained single-channel SGM to construct multiple parallel inference streams starting from independently sampled noise and fuses their inferred samples with weights at each iteration step. Similar works that integrate the traditional iterative reconstruction technique into the predictor-corrector sampling framework include SWORD [92], DCDS [96], and DPER [95].

For zero-shot SVCT, He et al. [101] proposed variational score solver (VSS), which distills a latent diffusion model into a variational score prior and integrates it with data consistency in an iterative solver, performing competitively with supervised methods without paired training data. More recently, Li et al. [108] proposed the cross-distribution diffusion priors-driven iterative reconstruction (CDPIR) method to address the out-of-distribution (OOD) generalization challenge that often degrades the performance of learning-based iterative methods. CDPIR incorporates a cross-distribution diffusion prior derived from a scalable interpolant Transformer (SiT), which is trained on multiple datasets, into a model-based iterative reconstruction loop. The method alternates between data-consistency updates and sampling from a unified stochastic interpolant to achieve state-of-the-art reconstruction quality. CDPIR also demonstrates robustness under domain shifts caused by scanner, protocol, or anatomical variations.

All of the above works follow the Gaussian noise-based diffusion framework that dominates conventional vision tasks, but this forward diffusion setup does not physically align with the deterministic measurement undersampling that causes degradation in SVCT. To this end, Yang et al. [80] redefined forward diffusion as an image degradation process caused by continuous downsampling of full-view sinograms and proposed a dual-domain diffusion model (DDDM) for SVCT reconstruction. In the sinogram domain, a sinogram upgrading module (SUM) based on an attention U-net is trained to upgrade the sparse-view sinogram step by step, and in the image domain, an improved conditional DDPM along with an accelerated sampling technique is used to further refine the reconstructed image. Similarly, Yang et al. [46] proposed a sampling diffusion model called CT-SDM, which redesigns the degradation and recovery operators in the sinogram domain within the framework of the cold diffusion model [128] by replacing the Gaussian noise diffusion with projected view resampling to adapt to the physical process of SVCT scanning, and achieves self-adaptation to various sparse sampling rates. A group-random sampling strategy was developed that can dynamically adjust the sampling rate and select projection views during training, achieving effective data augmentation. The transformation agnostic cold sampling (TACoS) algorithm [128] was used for reverse diffusion in the sinogram domain, restoring the full-view sinogram, and in the image domain, a single ResNet block was used for light-weight image refinement. Chen et al. [67] proposed CvG-Diff, which reformulates SVCT reconstruction as a generalized diffusion process in the image domain by replacing Gaussian noise with a deterministic artifact degradation operator that captures angular subsampling effects across different sparsity levels. Two novel strategies, error-propagating composite training (EPCT) and semantic-prioritized dual-phase sampling (SPDPS), are introduced to suppress artifact propagation and improve sampling efficiency, achieving high-quality reconstruction (38.34 dB PSNR for 18-view CT) in only 10 sampling steps.

In [65], Zhou et al. proposed a conditional embedding fusion diffusion model (CEF-DM) in the image domain, which employs a U-shaped FourierNet to generate an initial reconstruction, and then uses it as a conditioning input to guide the reverse diffusion process toward detail-rich outputs. A conditional attention embedding module (CAEM) is designed to comprehensively fuse the conditional information with time-step embeddings throughout the denoising process, and the final reconstruction is obtained by summing the initial estimate with the generated residual details. Diffusion priors have also been extended to joint sparse-view and metal artifact reduction. Hyun et al. [129] combined a DDPM with implicit neural representation in a self-supervised framework for SVMAR, alternately performing MAR inpainting guided by diffusion priors and SVCT data fidelity refinement.

An alternative approach to leveraging diffusion for SVCT was presented in [89]. Instead of training a generative model for reverse diffusion, Wu et al. introduced only the forward diffusion process into the DIP framework (which is also used in [53], see Section 3.5). By adding multi-level Gaussian noise to the initial FBP reconstructed images used as input to the DIP network, the diversity and dynamics of the input training data can be increased to provide richer prior information.

Diffusion models can learn the manifold distribution of normal CT images and thus can iteratively *generate* visually appealing CT images from sparse projection data without the need for paired training data. However, it is possible for the generative models to generate false details that appear plausible during the iterative sampling process, especially in ultra-sparse configurations (corresponding to weak data fidelity constraints). Consequently, ensuring that the *generated* CT images are real and not just look real (conforming to the manifold distribution of normal CT) is the first concern when applying the generative technique to medical imaging. In addition, the iterative sampling mode of diffusion models typically results in high computational costs, and improving the efficiency of training and sampling is also a concern.

4.1.5 RNN-Based Methods

RNNs, and in particular LSTM and GRU, offer a unique inductive bias for SVCT: they naturally model sequential dependencies, which aligns with the angular ordering of projection views in a CT scan. Each view is acquired at a specific rotation angle, and adjacent views are physically related through the geometry of the Radon transform. This makes RNNs a compelling choice for sinogram-domain processing, where they can capture the angular continuity of projection data. In the image domain, RNNs have been applied to multi-stage iterative refinement, where the recurrent structure naturally models the progressive removal of streak artifacts across stages. However, RNNs face challenges in SVCT: training instability due to vanishing gradients can be problematic when processing long sequences of views (e.g., 360 views in a full scan); and the inherently 2D nature of CT images and sinograms means that 1D sequential models may miss important spatial correlations that 2D convolutions capture more effectively. Hybrid architectures that combine CNNs with RNNs (e.g., convLSTM in RSBP, convGRU in R^2-Net) have emerged as a pragmatic solution, using convolutions to capture spatial features and recurrence to model angular dependencies.

RNNs have significant advantages when dealing with sequential data. The projection views observed in a CT scan are actually a sequence arranged in time (angle). LSTM and GRU are the most commonly used RNN variants that are better at learning long-term dependencies within a sequence. Compared to convolutional neural networks with limited receptive fields, LSTM and GRU are superior in capturing long-range or global features during SVCT reconstruction. In the context of SVCT, the sequential nature of projection data similarly benefits from the capabilities of RNN-based architectures. Fig. 16 illustrates the RSBP-CNN architecture, in which recurrent processing is applied to the stacked back-projection features before U-net refinement.

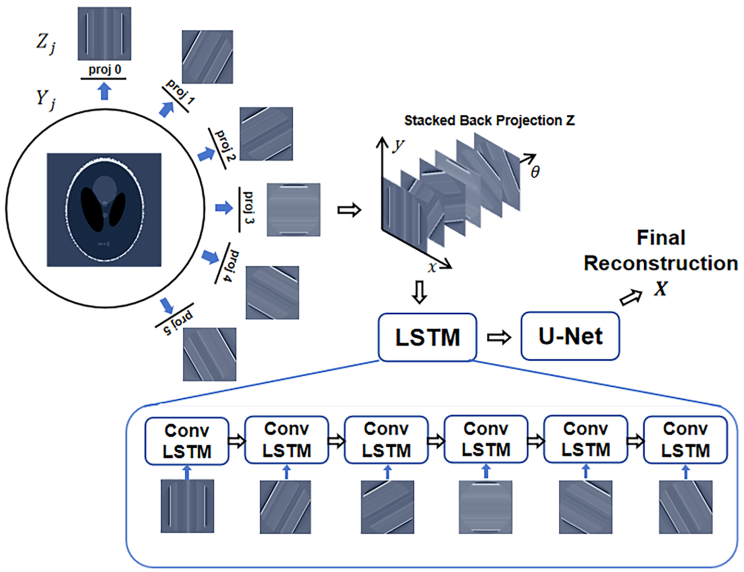


Figure 16: RSBP-CNN architecture diagram.

To improve artifact removal in the image domain, Shen et al. [29] proposed a multi-stage model, R^2-Net, which recursively removes streak artifacts by introducing a recurrent mechanism based on convolutional GRU (convGRU). As mentioned in Section 3.1, a convolutional LSTM module was also used in RSBP [59] to sequentially process the stacked back projection images, thus providing a richer input of long-range features to the subsequent U-net. In [70], Zhou et al. proposed a dual-domain data-consistent recurrent network, DuDoDR-Net, for simultaneous sparse-view reconstruction and metal artifact reduction, which uses two

U-shaped networks with convGRU as bottlenecks to recurrently recover data between the image domain and the sinogram domain.

In summary, most of the deep learning models used in SVCT are derived from conventional vision tasks and are usually implemented using a data-driven manner. Each network or module has its own advantages, and it is critical to select and integrate them appropriately to make them suitable for the SVCT scenario, achieving a balance between reconstruction accuracy, computational complexity, and clinical applicability. The choice of technique should be guided by the specific SVCT challenge at hand: CNNs excel at local texture refinement but struggle with global artifacts; Transformers capture long-range dependencies but incur high computational cost; diffusion models provide strong image priors but require careful integration with physics-based components; GANs offer perceptual fidelity but suffer from training instability; and RNNs naturally model angular dependencies in the sinogram but are less effective for spatial feature extraction. Hybrid architectures that strategically combine multiple techniques—such as CNN-Transformer hybrids for local-global feature fusion, or diffusion-integrated iterative frameworks for physics-data-driven reconstruction—are increasingly recognized as the most promising direction for advancing SVCT reconstruction performance. Table 3 organizes the literature reviewed in this survey according to the networks and modules used, some of which may use hybrid architectures and therefore appear in multiple rows. Although there are some attempts to combine deep learning with SVCT expertise in the literature, the design of frameworks, networks, or modules that are physically adapted to the SVCT task and in accordance with the CT imaging principles to enable both data- and physics-driven and interpretable reconstruction is still a challenge and deserves in-depth research in the future.

Table 3: Classification of literature according to the key deep learning (DL) techniques used.

DL Techniques	Features	Works (Year)
CNNs	U-net	[24] (2017) [25,26] (2018) [33,34] (2019) [42,58] (2020) [51,59] (2021) [39,52,61,85] (2021) [40,43,70,86] (2022) [50,62,71,72,74,87,88] (2023) [75,76,78,80,81,89,91,93] (2024) [46,64,67,106] (2025) [83,129] (2026)
	Non-U-shaped	[27,41,47,84] (2018) [28,48] (2019) [68] (2020) [60] (2021) [53,54] (2022) [30,37] (2023) [90] (2024) [98] (2025) [65,107] (2026)

(Continued)

Table 3 (continued)

DL Techniques	Features	Works (Year)
Transformers	Attention	[61] (2021) [43] (2022) [62,71,88] (2023) [76,78] (2024) [106] (2025) [65,83] (2026)
	Swin-T	[40,45] (2022) [55,73] (2023) [79,94] (2024) [82] (2025)
	Others	[44,69] (2022) [32,63,75] (2024) [102] (2025) [38,108] (2026)
Diffusion	—	[35,56] (2022) [57] (2023) [31,32,80,91,92,95,96] (2024) [46,67,101] (2025) [65,108,129] (2026)
GANs	—	[41] (2018) [42] (2020) [52] (2021) [43] (2022) [87] (2023) [63,76] (2024)
RNNs	—	[29] (2019) [59] (2021) [36,70] (2022) [77] (2024)

4.2 Loss Functions

Various loss functions popular in conventional vision tasks have also been used for SVCT reconstruction, such as the mean squared error (MSE), the mean absolute error (MAE), the structural similarity index measure (SSIM), and the adversarial loss mentioned in Section 4.1.3. These loss functions play different roles in balancing pixel-level accuracy, structural fidelity, and perceptual quality in CT reconstruction.

The MSE is the most widely used index for quantifying pixel-wise differences between an estimated image and its ground truth. Let $\hat{\mathbf{x}} \in \mathbb{R}^N$ be the estimated image containing N pixels and $\mathbf{x}$ be the ground

truth, then the MSE loss can be defined by the L_2-norm as

$$L_{MSE} = \frac{1}{N}\|\mathbf{x} - \hat{\mathbf{x}}\|_2^2 \tag{24}$$

The MSE can effectively measure the pixel-level discrepancy, but it also often leads to over-smoothing, especially in anatomically rich regions of CT images. The MAE based on the L_1-norm is often used instead of the MSE to mitigate over-smoothing,

$$L_{MAE} = \frac{1}{N}\|\mathbf{x} - \hat{\mathbf{x}}\|_1 \tag{25}$$

Compared to the MSE, the MAE is less sensitive to outliers and therefore more effective at preserving high-frequency detail. However, it may lead to slower convergence during training. The Charbonnier loss has also been chosen as a composite of the MSE and the MAE, e.g., in [32,44],

$$L_{Cha} = \frac{1}{N}\sum_{i=1}^{N}\sqrt{(x_i - \hat{x}_i)^2 + \varepsilon^2} \tag{26}$$

where x_i and $\hat{x}_i$ are the elements of $\mathbf{x}$ and $\hat{\mathbf{x}}$, respectively, and ε is a small constant. When the error $x_i - \hat{x}_i$ is much larger than ε, the Charbonnier loss approaches the MAE to avoid over-penalizing large errors, and when the error is much smaller than ε, it is close to the MSE to fine-fit small errors. Another similar composite loss is the Huber loss, used in [72].

Reconstructed CT images are typically provided to physicians for review, but pixel-level losses such as MSE, MAE, Charbonnier, and Huber losses may be inconsistent with human visual perception. To address this issue, the SSIM is often introduced to evaluate the structural similarity between images by taking into account brightness, contrast, and structural information,

$$SSIM = \frac{(2\mu_x\mu_{\hat{x}} + C_1)(2\sigma_{x\hat{x}} + C_2)}{(\mu_x^2 + \mu_{\hat{x}}^2 + C_1)(\sigma_x^2 + \sigma_{\hat{x}}^2 + C_2)} \tag{27}$$

where μ_a and σ_a^2 are the mean and variance of the pixels within the region of interest, respectively, $\sigma_{x\hat{x}}$ is the covariance between the ground truth and estimates, and $C_{1/2}$ are small constants that prevent division by zero. In practice, the SSIM typically uses a sliding window to compute the mean structural similarity of all (assuming M) local windows, and the corresponding loss function is defined as,

$$L_{SSIM} = 1 - \frac{1}{M}\sum_{i=1}^{M} SSIM_i \tag{28}$$

where $SSIM_i$ denotes the SSIM of the i-th window. As it is closer to human visual perception, the SSIM loss is often used in combination with the MSE or MAE to improve the quality of reconstructed images in the SVCT task [43,62,71,78,85,88,90]. In [25,75,80], multi-scale SSIM (MS-SSIM), a variant of SSIM, has also been used to train SVCT models.

Higher-level semantic similarity metrics, such as the adversarial loss mentioned in Section 4.1.3 and the perceptual loss [41,42,63], have also been used in SVCT to improve the visual perception of reconstructed images. Adversarial loss can provide a semantic-level judgment of whether a reconstructed image belongs to the normal image set or not, based on the learned manifold distribution of normal images. Perceptual loss uses pre-trained classical networks, such as VGG and ResNet, to extract feature representations of input images, and measures the differences between images in terms of distances (L_2 or L_1 norm) in feature space.

The perceptual loss is more in line with the human visual system, focusing more on the semantic content of the reconstructed image than on the pixel accuracy. The introduction of adversarial and perceptual losses is beneficial for recovering image details and improving visual plausibility, but in the SVCT task, these high-level semantic losses need to be coupled with low-level pixel losses to ensure that the content of the reconstructed image is realistic and believable.

In fact, the evaluation criteria for results inferred in the sinogram domain and the image domain are not the same. The sinogram domain outputs signal-level projection data, and it is reasonable to use lower-level losses such as the MSE and MAE, while the image domain may introduce higher-level or subsequent task-related losses, depending on the specific purpose of the output CT image. The inconsistency of the losses between the sinogram domain and the image domain should also be noted. Taking MSE loss as an example, a lower MSE in the sinogram domain does not always guarantee a lower MSE in the image domain after applying FBP. Fig. 17 illustrates the inconsistencies of MSE in the sinogram and image domains. For the same sinogram, different types of noise (Gaussian noise, constant bias, and spatially correlated noise) result in different levels of error in the image domain. In the sinogram domain, the MSE introduced by constant bias is nearly four times that caused by Gaussian noise. However, after FBP, the constant bias is effectively removed by the ramp filter, resulting in an MSE that is less than 2% of that caused by Gaussian noise in the image domain. Furthermore, in Fig. 17c, a spatially correlated noise is carefully designed in the sinogram domain. Despite its large amplitude, it cancels out during the FBP reconstruction, resulting in a significantly lower MSE in the image domain compared to the Gaussian noise.

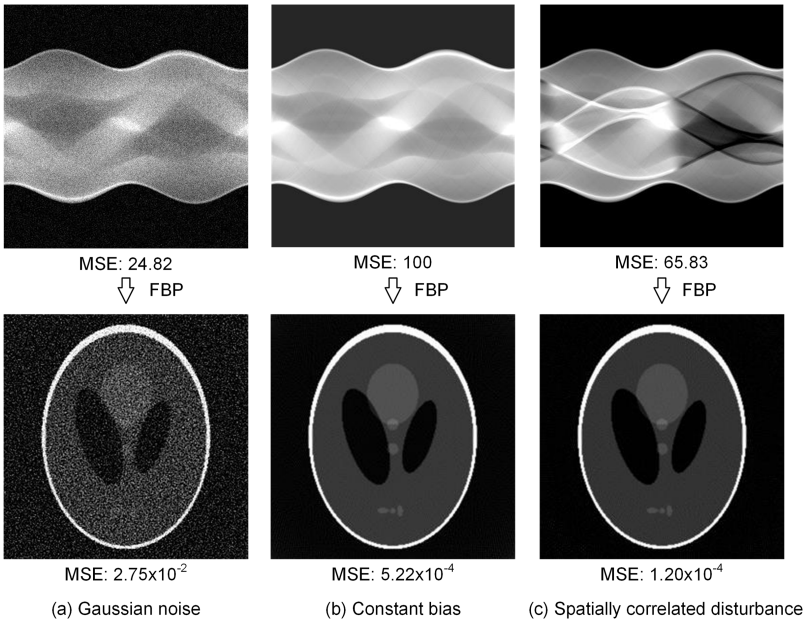


Figure 17: Inconsistency of deviation metrics between the sinogram and image domains.

The inconsistency between sinogram- and image-domain metrics is due to the complex spatial and frequency domain relationships between sinograms and reconstructed images, which highlights the limitations of relying solely on sinogram-domain loss for model training. Since sinograms themselves are not directly available to physicians, training a model with sinogram-domain loss does not ensure that a high-quality CT image will ultimately be obtained [37]. It is more reasonable for sinogram-domain models to append an FBP layer and define the training loss in the image domain.

4.2.1 Task-Oriented Evaluation Metrics

The pixel-level and structural metrics discussed above—PSNR, SSIM, and their variants—were originally designed for natural images and quantify only low-level signal fidelity. In the medical imaging context, they have been repeatedly shown to correlate poorly with diagnostic accuracy: an image can achieve a high PSNR yet hide a clinically significant lesion, or a perceptually convincing reconstruction may distort subtle low-contrast features that radiologists rely on [130]. Motivated by this gap, recent SVCT research has shifted towards *task-oriented evaluation* that explicitly links reconstruction quality to clinical utility.

The first layer of task-oriented evaluation is the *Medical Image Quality Assessment (MIQA)* framework, which formalizes quality assessment along three dimensions: (i) technical (fidelity to ground truth), (ii) perceptual (alignment with radiologist judgment), and (iii) task-based (impact on downstream diagnostic performance) [131]. Building on this taxonomy, the perceptual dimension is most directly evaluated through *radiologist visual quality assessment (VQA)*, also known as reader studies. In such studies, board-certified radiologists grade reconstructed images on a Likert-type scale (typically 1–5) across clinically meaningful criteria such as noise, artifact severity, contrast at anatomical boundaries, and overall diagnostic confidence. Two or more readers score each case independently, often blinded to the reconstruction method, and inter-rater agreement is quantified by Cohen's or Fleiss' κ. Reader studies remain the de-facto clinical gold standard and are increasingly required by journals and regulatory bodies for prospective validation of low-dose CT techniques [132].

The second, more objective, layer of task-oriented evaluation quantifies the impact of reconstruction on downstream clinical tasks. The two most widely used families are: (i) *detection metrics*, particularly the area under the ROC curve (AUC), sensitivity, specificity, and the figure of merit (FOM) for localization tasks such as lung-nodule or micro-calcification detection; and (ii) *segmentation metrics*, dominated by the Dice similarity coefficient (DSC), the 95th-percentile Hausdorff distance (HD95), and the average symmetric surface distance (ASSD) for organ and lesion delineation. These metrics can be computed automatically using pre-trained task networks (e.g., a lung-nodule detector or a multi-organ segmentor), which makes them attractive for large-scale benchmarking where reader studies are impractical. Recent work has shown that training or fine-tuning SVCT networks with task-loss components explicitly tied to such downstream metrics can substantially improve diagnostic performance even when pixel-level metrics remain comparable [133].

In summary, the SVCT community is moving from generic image-quality metrics to a multi-level evaluation protocol that combines (a) technical fidelity (PSNR/SSIM, still useful for ablation), (b) perceptual quality (radiologist VQA, the clinical gold standard), and (c) task-driven quantification (detection AUC, segmentation Dice). We therefore recommend that future SVCT studies report, at minimum, one metric from each category, and complement retrospective downsampling-based experiments with reader studies and task-based evaluation whenever clinical translation is claimed.

4.3 Datasets

Datasets are the infrastructure for deep learning models. The section introduces the publicly available datasets commonly used for SVCT.

LDCT Grand Challenge Dataset: The 2016 Low-Dose CT Grand Challenge [9], organized by the NIBIB, AAPM, and Mayo Clinic, provided abdominal CT data from 30 patient cases, 10 for training and 20 for testing. This dataset is commonly referred to as the AAPM dataset in the literature. Due to the high demand for this dataset in academia and industry, the AAPM and Mayo Clinic have publicly released it as a standard dataset for low-dose CT reconstruction. All 30 cases have been updated in 2021 to ensure that the data format is consistent with The Cancer Imaging Archive (TCIA) [134] database and can be downloaded via

a link provided by the AAPM [10]. The dataset has been used in many peer-reviewed studies. According to our statistics, of the nearly 70 SVCT papers published in the last five years, about half have used it as a benchmark dataset.

LDCT Image and Projection Dataset: The low-dose CT image and projection dataset (LDCT-P) [135,136] was created by the Mayo Clinic with NIBIB funding and consists of head, chest, and abdominal CT scan data from 299 patient cases, 150 cases from Siemens scanners and 149 cases from GE scanners. All cases have normal-dose and simulated low-dose projection data, full-dose image, and clinical data identifying all pathology. The dataset can be viewed as an extension of the 2016 Grand Challenge dataset: the standalone Grand Challenge release contains 30 cases (10 for training and 20 for testing), while LDCT-P includes a 13-case subset from that cohort. The LDCT-P dataset is publicly available through the TCIA.

LIDC-IDRI Dataset: The Lung Image Database Consortium (LIDC) and Image Database Resource Initiative (IDRI) have provided a publicly available dataset of 1018 chest CT cases with XML-based annotations from experienced radiologists to stimulate the development of CAD methods for lung nodule detection, classification, and quantitative assessment. Although not specifically designed for CT reconstruction, several SVCT studies have used LIDC-IDRI as a benchmark for experimental validation. The LIDC-IDRI dataset was established through the LIDC/IDRI program and is currently distributed through the TCIA [137].

Other CT collections from TCIA and NBIA such as LungCT-Diagnosis [33], LUNGx challenge [58], QIN-LungCT-Seg [30], MIDRC-RICORD [78,138], and TCGA-KIRC [46] were also used for SVCT validation. In addition, a larger CT dataset, DeepLesion [139], consisting of 32,120 CT slices from 10,594 CT scans of 4427 unique patients, publicly released by NIH Clinical Center [140], also provides broader data support for the SVCT task [70].

Publicly available datasets are critical for research in deep learning-based CT reconstruction, as they are the data fuel necessary for training CT models and the cornerstone for objective and fair performance evaluation. The above-mentioned datasets have greatly contributed to the development of low-dose CT imaging techniques, and it is hoped that more large-scale, high-quality datasets will be released and shared in the future, so that low-dose CT technology can benefit patients even more.

Despite their contributions, the existing datasets have notable limitations that should be considered when interpreting results. First, the AAPM dataset and LDCT-P are both derived from a limited number of patient cases (30 and 299, respectively), which may not adequately capture the anatomical diversity and pathological variations encountered in clinical practice. Second, many commonly used datasets are dominated by abdominal or chest CT, while musculoskeletal and other anatomical regions remain comparatively underrepresented, limiting the generalizability of models trained solely on these data. Third, in the AAPM and LDCT-P datasets, the sparse-view data are typically generated by retrospective downsampling of full-view projections, which may not fully replicate the physical characteristics of actual sparse-view acquisition (e.g., detector noise, scatter, and motion artifacts). Fourth, many datasets are derived from scanners from a limited number of vendors, primarily Siemens and GE, which may introduce vendor-specific biases. These limitations underscore the need for more diverse, multi-vendor, and prospective datasets to better validate SVCT methods in real-world clinical settings.

A particularly important source of Sim2Real domain gap comes from the mismatch between simulated downsampling and actually acquired sparse-view data. The noise characteristics, physical calibration, beam-hardening correction, gain map, and detector response of real clinical sparse-view projections are determined by vendor-specific hardware and acquisition protocols, and cannot be faithfully reproduced by simple uniform subsampling of full-view projections. As a result, models trained on simulated pairs often overfit to the aliasing patterns of the specific downsampling operator and degrade noticeably when deployed on

prospective data. Fifth, many widely used datasets, such as LIDC-IDRI, only provide reconstructed CT images without the original raw projections. In such cases, sparse-view inputs must be synthesized by re-projecting the available images through an additional FBP step (and a software forward projector), which introduces system-specific reconstruction errors, non-standard ray geometry, and re-projection inconsistencies that further widen the gap between the training distribution and prospective clinical sparse-view scans. These Sim2Real limitations motivate the use of unpaired or dose-level prompted learning strategies [11], transfer learning from simulated to real CT data [141], and cycle-consistent adversarial frameworks [142], and the acquisition of dedicated prospective sparse-view datasets with raw projection data for more realistic training and evaluation.

5 Conclusion

SVCT is a classic prediction problem involving the reconstruction of high-resolution CT images from limited projection measurements. Given its effectiveness in complex, high-dimensional tasks, deep learning is well suited to addressing this challenge. The synergy between deep learning and SVCT is set to grow significantly. This growth is partly driven by the availability of retrospective downsampling for constructing paired training data, which has enabled rapid development of supervised learning methods. However, this strategy is inherently limited by domain shift: downsampled data do not fully reflect the noise characteristics, scatter, and motion artifacts of actual prospective sparse-view acquisition, and many public datasets lack raw projection data altogether. Addressing these limitations is critical for moving from retrospective evaluation to real-world clinical deployment.

However, current approaches often treat SVCT as a standard vision task by applying generic denoising or interpolation techniques. Influenced by purely data-driven models from the computer vision community, many of these approaches fail to utilize fully the specific imaging physics and prior knowledge inherent in CT scans. This results in inefficient, non-interpretable, suboptimal solutions.

The ongoing goal of CT imaging is to achieve lower radiation doses, faster scanning speeds, and clearer images. Despite this progress, deep learning-based SVCT still faces several challenges:

- **High computational cost:** Deep models are computationally expensive, hindering deployment on CT scanners and use in time-critical scenarios.
- **Generalization and robustness:** Performance is sensitive to variations in imaging conditions (e.g., noise, scan geometry) and anatomical differences, and varies significantly across datasets.
- **Lack of interpretability:** The "black-box" nature of data-driven models limits trust and clinical adoption.
- **Clinical validation gap:** Most studies rely on retrospective evaluation; prospective and multi-reader studies are scarce.
- **Ultra-low-dose reconstruction:** Combining sparse views with other dose-reduction strategies (e.g., low tube voltage) remains an open problem.

Looking ahead, future research should focus on several key areas. Firstly, **task-oriented evaluation metrics** (see Section 4.2.1) that are aligned with clinical tasks such as disease detection and organ segmentation need to be developed. This is because conventional metrics, such as PSNR and SSIM, have been shown to be misaligned with diagnostic accuracy. Moreover, image quality does not automatically translate to diagnostic reliability; task-based evaluation involving clinical readers and disease-specific performance metrics is essential. Secondly, **physics-based dual-engine models** that integrate CT imaging mechanisms and prior knowledge with data-driven learning should be designed to improve robustness, generalizability, and interpretability. This includes leveraging differentiable projectors and domain-specific regularizers.

Thirdly, the focus should gradually shift towards holistic **3D reconstruction** using multi-slice or helical projection data, moving beyond slice-by-slice 2D approaches to achieve more comprehensive volumetric imaging. Fourthly, **real-world clinical validation** is urgently needed. Most current methods are evaluated on retrospectively downsampled data from existing datasets; prospective studies on actual sparse-view scanners and multi-vendor data are essential to assess real-world performance. Fifthly, **computational efficiency and deployment** should be prioritized. Lightweight models, quantization, and hardware-aware design can facilitate integration into clinical CT pipelines with real-time constraints. Sixthly, **foundation models and self-supervised pre-training** for CT reconstruction hold promise. Pre-training on large-scale unlabeled CT data followed by fine-tuning on sparse-view tasks may reduce data dependence and improve generalization. Seventhly, addressing **ultra-low-dose reconstruction** by combining sparse views with other dose-reduction strategies (e.g., low tube voltage, iterative reconstruction) remains an open challenge that requires joint optimization of acquisition and reconstruction.

Addressing these issues could lead to more accurate, efficient, and clinically reliable SVCT solutions.

Acknowledgement: Not applicable.

Funding Statement: This work was supported by Chengdu Science and Technology Program (2026-YF08-00034-GX).

Author Contributions: The authors confirm contribution to the paper as follows: Conceptualization, Shuaiqi Cheng and Bo Yang; methodology, Shuaiqi Cheng, Yuxi Chen, and Bo Yang; software, Shuaiqi Cheng; validation, Yuxi Chen; formal analysis, Shuaiqi Cheng and Yuxi Chen; investigation, Shuaiqi Cheng and Yuxi Chen; data curation, Shuaiqi Cheng and Yuxi Chen; writing—original draft preparation, Shuaiqi Cheng; writing—review and editing, Yuxi Chen, Bo Yang, Legend Zhang, Junmin Lyu, Guangyu Xu, and Chao Liu; visualization, Shuaiqi Cheng and Yuxi Chen; supervision, Bo Yang, Legend Zhang, Junmin Lyu, Guangyu Xu, and Chao Liu; project administration, Bo Yang. All authors reviewed and approved the final version of the manuscript.

Availability of Data and Materials: This study did not generate a new dataset. The datasets discussed in this review are publicly available from the repositories cited in the manuscript. The open-source code developed in this study is publicly available at https://github.com/githCS2/FP_FBP.

Ethics Approval: Not applicable.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Buzug TM. Computed tomography. In: Springer handbook of medical technology. Berlin/Heidelberg, Germany: Springer; 2011. p. 311–42.
2. Withers PJ, Bouman C, Carmignato S, Cnudde V, Grimaldi D, Hagen CK, et al. X-ray computed tomography. Nat Rev Methods Primers. 2021;1(1):18. doi:10.1038/s43586-021-00015-4.
3. Kalender WA. Computed tomography: Fundamentals, system technology, image quality, applications. Hoboken, NJ, USA: John Wiley & Sons, Inc.; 2011.
4. Brenner DJ, Hall EJ. Computed tomography—an increasing source of radiation exposure. New Engl J Med. 2007;357(22):2277–84. doi:10.1056/nejmra072149.
5. Kalra MK, Maher MM, Toth TL, Hamberg LM, Blake MA, Shepard JA, et al. Strategies for CT radiation dose optimization. Radiology. 2004;230(3):619–28. doi:10.1148/radiol.2303021726.
6. Goldman LW. Principles of CT: Radiation dose and image quality. J Nucl Med Technol. 2007;35(4):213–25.
7. Noo F, Defrise M, Clackdoyle R, Kudo H. Image reconstruction from fan-beam projections on less than a short scan. Phys Med Biol. 2002;47(14):2525. doi:10.1088/0031-9155/47/14/311.

8. Chen Z, Jin X, Li L, Wang G. A limited-angle CT reconstruction method based on anisotropic TV minimization. Phys Med Biol. 2013;58(7):2119. doi:10.1088/0031-9155/58/7/2119.
9. McCollough C. TU-FG-207A-04: overview of the low dose CT grand challenge. Med Phys. 2016;43 (6Part35):3759–60.
10. AAPM, Clinic M. AAPM 2016 low-dose CT grand challenge dataset. American association of physicists in medicine; 2016 [cited 2025 Apr 10]. Available from: https://aapm.app.box.com/s/eaw4jddb53keg1bptavvvd1sf4x3pe9h.
11. Zhang Y, Zhang W, Chen G, Wang P, Li D, Ma J, et al. DLPP-UL: dose-level parameter prompted unpaired learning for low-dose CT image restoration. Pattern Recognit. 2026;171:112272.
12. Shepp LA, Logan BF. The Fourier reconstruction of a head section. IEEE Trans Nucl Sci. 1974;21(3):21–43. doi:10.1109/tns.1974.6499235.
13. Beister M, Kolditz D, Kalender WA. Iterative reconstruction methods in X-ray CT. Phys Medica. 2012;28(2):94–108. doi:10.1016/j.ejmp.2012.01.003.
14. Geyer LL, Schoepf UJ, Meinel FG, Nance JW Jr, Bastarrika G, Leipsic JA, et al. State of the art: Iterative CT reconstruction techniques. Radiology. 2015;276(2):339–57.
15. Liu Y, Liang Z, Ma J, Lu H, Wang K, Zhang H, et al. Total variation-stokes strategy for sparse-view X-ray CT image reconstruction. IEEE Trans Med Imaging. 2013;33(3):749–63. doi:10.1109/tmi.2013.2295738.
16. Xu Q, Yu H, Mou X, Zhang L, Hsieh J, Wang G. Low-dose X-ray CT reconstruction via dictionary learning. IEEE Trans Med Imaging. 2012;31(9):1682–97. doi:10.1109/tmi.2012.2195669.
17. Zhang Y, Mou X, Wang G, Yu H. Tensor-based dictionary learning for spectral CT reconstruction. IEEE Trans Med Imaging. 2016;36(1):142–54. doi:10.1109/TMI.2016.2600249.
18. Klink T, Obmann V, Heverhagen J, Stork A, Adam G, Begemann P. Reducing CT radiation dose with iterative reconstruction algorithms: the influence of scan and reconstruction parameters on image quality and CTDIvol. Eur J Radiol. 2014;83(9):1645–54. doi:10.1016/j.ejrad.2014.05.033.
19. Willemink MJ, de Jong PA, Leiner T, de Heer LM, Nievelstein RA, Budde RP, et al. Iterative reconstruction techniques for computed tomography Part 1: Technical principles. Eur Radiol. 2013;23(6):1623–31. doi:10.1007/s00330-012-2765-y.
20. Hyun CM, Seo JK. In: Seo JK, editor. Deep learning for Ill posed inverse problems in medical imaging. Singapore: Springer Nature Singapore; 2023. p. 319–39.
21. Di J, Lin J, Zhong L, Qian K, Qin Y. Review of sparse-view or limited-angle CT reconstruction based on deep learning. Laser Optoelectron Prog. 2023;60(8):0811002.
22. Szczykutowicz TP, Toia GV, Dhanantwari A, Nett B. A review of deep learning CT reconstruction: concepts, limitations, and promise in clinical practice. Curr Radiol Rep. 2022;10(9):101–15.
23. Fessler JA. Fundamentals of CT reconstruction in 2D and 3D. In: Brahme A, editor. Comprehensive biomedical physics. Oxford, UK: Elsevier; 2014. p. 263–95.
24. Jin KH, McCann MT, Froustey E, Unser M. Deep convolutional neural network for inverse problems in imaging. IEEE Trans Image Process. 2017;26(9):4509–22. doi:10.1109/tip.2017.2713099.
25. Zhang Z, Liang X, Dong X, Xie Y, Cao G. A sparse-view CT reconstruction method based on combination of DenseNet and deconvolution. IEEE Trans Med Imaging. 2018;37(6):1407–17. doi:10.1109/tmi.2018.2823338.
26. Han Y, Ye JC. Framing U-Net via deep convolutional framelets: application to sparse-view CT. IEEE Trans Med Imaging. 2018;37(6):1418–29. doi:10.1109/tmi.2018.2823768.
27. Ye DH, Buzzard GT, Ruby M, Bouman CA. Deep back projection for sparse-view CT reconstruction. In: Proceedings of the 2018 IEEE Global Conference on Signal and Information Processing (GlobalSIP); 2018 Nov 26–28; Anaheim, CA, USA. p. 1–5.
28. Qian Y, Xie S, Zhuang W, Li H. Sparse-view CT reconstruction based on improved re-sidual network. In: Okada H, Atluri SN, editor. Computational and experimental simulations in engineering. Berlin/Heidelberg, Germany: Springer; 2020. p. 1069–80.
29. Shen T, Li X, Zhong Z, Wu J, Lin Z. R2-Net: recurrent and recursive network for sparse-view CT artifacts removal. In: Proceedings of the MICCAI 2019: 22nd International Conference; 2019 Oct 13–17; Shenzhen, China. p. 319–27.

30. Liu X, Sajda P. Unsupervised sparse-view backprojection via convolutional and spatial transformer networks. In: Liu F, Zhang Y, Kuai H, Stephen EP, Wang H, editor. Brain Informatics. Berlin/Heidelberg, Germany: Springer; 2023. p. 308–17.
31. Zhao F, Zhao J, Liu M. Image domain ultra-sparse view CT artifact removal via conditional denoising diffusion probability model. In: Proceedings of the 2023 12th International Conference on Computing and Pattern Recognition. New York, NY, USA: Association for Computing Machinery; 2024. p. 416–21.
32. Liu P, Fang C, Qiao Z. A dense and U-shaped transformer with dual-domain multi-loss function for sparse-view CT reconstruction. J X-ray Sci Technol. 2024;32(2):207–28. doi:10.3233/xst-230184.
33. Lee H, Lee J, Kim H, Cho B, Cho S. Deep-neural-network-based sinogram synthesis for sparse-view CT image reconstruction. IEEE Trans Radiat Plasma Med Sci. 2019;3(2):109–19. doi:10.1109/trpms.2018.2867611.
34. Cao G, Vekhande S, Dong X. Sinogram interpolation for sparse-view micro-CT with deep learning neural network. In: Proceedings of the Medical Imaging 2019: Physics of Medical Imaging; 2019 Feb 16–21; San Diego, CA, USA. 109482O p.
35. Xia W, Cong W, Wang G. Patch-based denoising diffusion probabilistic model for sparse-view CT reconstruction. arXiv:2211.10388. 2022.
36. Li S, Ye W, Li F. LU-Net: combining LSTM and U-Net for sinogram synthesis in sparse-view SPECT reconstruction. Math Biosci Eng. 2022;19(4):4320–40.
37. Guo F, Yang B, Feng H, Zheng W, Yin L, Yin Z, et al. An efficient sinogram domain fully convolutional interpolation network for sparse-view computed tomography reconstruction. Appl Sci. 2023;13(20):11264. doi:10.3390/app132011264.
38. Chen Y, Cheng S, Yang B, Liu C, Zhang L, Zheng W. ViewTrans: a physics-informed transformer for sparse-view CT sinogram restoration. Digit Signal Process. 2026;177:106113.
39. Hu D, Liu J, Lv T, Zhao Q, Zhang Y, Quan G, et al. Hybrid-domain neural network processing for sparse-view CT reconstruction. IEEE Trans Radiat Plasma Med Sci. 2021;5(1):88–98. doi:10.1109/trpms.2020.3011413.
40. Pan J, Zhang H, Wu W, Gao Z, Wu W. Multi-domain integrative Swin transformer network for sparse-view tomographic reconstruction. Patterns. 2022;3(6):100498. doi:10.2139/ssrn.3991087.
41. Zhao Z, Sun Y, Cong P. Sparse-view CT reconstruction via generative adversarial networks. In: Proceedings of the 2018 IEEE Nuclear Science Symposium and Medical Imaging Conference Proceedings (NSS/MIC); 2018 Nov 10–17; Sydney, Australia. p. 1–5.
42. Wei H, Schiffers F, Würfl T, Shen D, Kim D, Katsaggelos AK, et al. 2-step sparse-view CT reconstruction with a domain-specific perceptual network. arXiv:2012.04743. 2020.
43. Sun C, Deng K, Liu Y, Yang H. A lightweight dual-domain attention framework for sparse-view CT reconstruction. In: Proceedings of the 2022 IEEE 8th International Conference on Computer and Communications (ICCC); 2022 Dec 9–12; Virtual. p. 2102–6. doi:10.1109/iccc56324.2022.10065958.
44. Li R, Li Q, Wang H, Li S, Zhao J, Yan Q, et al. DDPTransformer: dual-domain with parallel transformer network for sparse view CT image reconstruction. IEEE Trans Comput Imaging. 2022;8:1101–16.
45. Wang C, Shang K, Zhang H, Li Q, Zhou SK. DuDoTrans: dual-domain transformer for sparse-view CT reconstruction. In: Haq N, Johnson P, Maier A, Qin C, Würfl T, Yoo J, editors. Machine learning for medical image reconstruction. Berlin/Heidelberg, Germany: Springer; 2022. p. 84–94.
46. Yang L, Huang J, Yang G, Zhang D. CT-SDM: A sampling diffusion model for sparse-view CT reconstruction across various sampling rates. IEEE Trans Med Imaging. 2025;44(6):2581–93.
47. Zhu B, Liu JZ, Cauley SF, Rosen BR, Rosen MS. Image reconstruction by domain-transform manifold learning. Nature. 2018;555(7697):487–92. doi:10.1038/nature25988.
48. Li Y, Li K, Zhang C, Montoya J, Chen GH. Learning to reconstruct computed tomography images directly from sinogram data under a variety of data acquisition conditions. IEEE Trans Med Imaging. 2019;38(10):2469–81. doi:10.1109/tmi.2019.2910760.
49. Kim B, Shim H, Baek J. A streak artifact reduction algorithm in sparse-view CT using a self-supervised neural representation. Med Phys. 2022;08:49.

50. Mizusawa S, Sei Y, Orihara R, Ohsuga A. Computed tomography image reconstruction using stacked U-Net. Comput Med Imaging Graph. 2021;90:101920. doi:10.1016/j.compmedimag.2021.101920.
51. Zhang F, Zhang M, Qin B, Zhang Y, Xu Z, Liang D, et al. REDAEP: robust and enhanced denoising autoencoding prior for sparse-view CT reconstruction. IEEE Trans Radiat Plasma Med Sci. 2021;5(1):108–19.
52. Wu W, Hu D, Niu C, Yu H, Vardhanabhuti V, Wang G. DRONE: dual-domain residual-based optimization network for sparse-view CT reconstruction. IEEE Trans Med Imaging. 2021;40(11):3002–14.
53. Shu Z, Entezari A. Sparse-view and limited-angle CT reconstruction with untrained networks and deep image prior. Comput Methods Programs Biomed. 2022;226(7697):107167. doi:10.1016/j.cmpb.2022.107167.
54. Su T, Cui Z, Yang J, Zhang Y, Liu J, Zhu J, et al. Generalized deep iterative reconstruction for sparse-view CT imaging. Phys Med Biol. 2022;67(2):025005. doi:10.1088/1361-6560/ac3eae.
55. Xia W, Yang Z, Lu Z, Wang Z, Zhang Y. RegFormer: a local-nonlocal regularization-based model for sparse-view CT reconstruction. IEEE Trans Radiat Plasma Med Sci. 2023;8(2):184–94.
56. Song Y, Shen L, Xing L, Ermon S. Solving inverse problems in medical imaging with score-based generative models. arXiv:2111.08005. 2021.
57. Guan B, Yang C, Zhang L, Niu S, Zhang M, Wang Y, et al. Generative modeling in sinogram domain for sparse-view CT reconstruction. IEEE Trans Radiat Plasma Med Sci. 2024;8(2):195–207. doi:10.1109/trpms.2023.3309474.
58. Nakai H, Nishio M, Yamashita R, Ono A, Nakao KK, Fujimoto K, et al. Quantitative and qualitative evaluation of convolutional neural networks with a deeper U-Net for sparse-view computed tomography reconstruction. Acad Radiol. 2020;27(4):563–74. doi:10.1016/j.acra.2019.05.016.
59. Li W, Buzzard GT, Bouman CA. Sparse-view CT reconstruction using recurrent stacked back projection. In: Proceedings of the 2021 55th Asilomar Conference on Signals, Systems, and Computers; 2021 Oct 31–Nov 3; Virtual. p. 862–6. doi:10.1109/ieeeconf53345.2021.9723242.
60. Xie S, Yang T. Artifact removal in sparse-angle CT based on feature fusion residual network. IEEE Trans Radiat Plasma Med Sci. 2021;5(2):261–71. doi:10.1109/trpms.2020.3000789.
61. Sun C, Liu Y, Yang H. Degradation-aware deep learning framework for sparse-view CT reconstruction. Tomography. 2021;7(4):932–49. doi:10.3390/tomography7040077.
62. Chan Y, Liu X, Wang T, Dai J, Xie Y, Liang X. An attention-based deep convolutional neural network for ultra-sparse-view CT reconstruction. Comput Biol Med. 2023;161:106888. doi:10.1016/j.compbiomed.2023.106888.
63. Zhang T, Liu J, Wu F, Wang K, Huang S, Zhang Y. Artifact suppression for sparse view CT via transformer-based generative adversarial network. Biomed Signal Process Control. 2024;95(9840):106297. doi:10.1016/j.bspc.2024.106297.
64. Lv L, Li C, Wei W, Sun S, Ren X, Pan X, et al. Optimization of sparse-view CT reconstruction based on convolutional neural network. Med Phys. 2025;52(4):2089–105.
65. Zhou C, Sun Y, Liang J, Liu J, Liu Q. Sparse-view CT image reconstruction using conditional embedding fusion diffusion model. Neurocomputing. 2026;659(2):131748. doi:10.1016/j.neucom.2025.131748.
66. Heller N, Sathianathen NJ, Kalapara AA, Walczak E, Moore K, Kaluzniak H, et al. The KiTS19 challenge data: 300 kidney tumor cases with clinical context, CT semantic segmentations, and surgical outcomes. arXiv:1904.00445. 2019.
67. Chen J, Lin Y, Qin Y, Wang H, Li X. Cross-view generalized diffusion model for sparse-view CT reconstruction. In: Proceedings of Medical Image Computing and Computer Assisted Intervention—MICCAI 2025; 2025 Sep 23–27; Daejeon, Republic of Korea. p. 140–50.
68. Wang W, Xia XG, He C, Ren Z, Lu J, Wang T, et al. An end-to-end deep network for reconstructing CT images directly from sparse sinograms. IEEE Trans Comput Imaging. 2020;6:1548–60. doi:10.1109/tci.2020.3039385.
69. Shi C, Xiao Y, Chen Z. Dual-domain sparse-view CT reconstruction with Transformers. Phys Medica. 2022;101:1–7. doi:10.1016/j.ejmp.2022.07.001.
70. Zhou B, Chen X, Zhou SK, Duncan JS, Liu C. DuDoDR-Net: dual-domain data consistent recurrent network for simultaneous sparse view and metal artifact reduction in computed tomography. Med Image Anal. 2022;75:102289. doi:10.1016/j.media.2021.102289.

71. Gao X, Su T, Zhang Y, Zhu J, Tan Y, Cui H, et al. Attention-based dual-branch deep network for sparse-view computed tomography image reconstruction. Quant Imaging Med Surg. 2023;3:13.
72. Cheslerean-Boghiu T, Hofmann FC, SchultheißM, Pfeiffer F, Pfeiffer D, Lasser T. WNet: a data-driven dual-domain denoising model for sparse-view computed tomography with a trainable reconstruction layer. IEEE Trans Comput Imaging. 2023;9:120–32.
73. Li Y, Sun X, Wang S, Li X, Qin Y, Pan J, et al. MDST: multi-domain sparse-view CT reconstruction based on convolution and swin transformer. Phys Med Biol. 2023;68(9):095019.
74. Ma C, Li Z, Zhang J, Zhang Y, Shan H. FreeSeed: frequency-band-aware and self-guided network for sparse-view CT reconstruction. In: Proceedings of the Medical Image Computing and Computer Assisted Intervention—MICCAI 2023 Oct 8–12; Vancouver, BC, Canada. p. 250–9.
75. Lin J, Li J, Dou J, Zhong L, Di J, Qin Y. DdeNet: a dual-domain end-to-end network combining Pale-Transformer and Laplacian convolution for sparse view CT reconstruction. Biomed Signal Process Control. 2024;96:106593.
76. Li G, Deng Z, Ge Y, Luo S. HEAL: high-frequency enhanced and attention-guided learning network for sparse-view CT reconstruction. Bioengineering. 2024;11(7):646. doi:10.3390/bioengineering11070646.
77. Lin J, Li J, Jiazhen D, Zhong L, Di J, Qin Y. Dual-domain reconstruction network incorporating multi-level wavelet transform and recurrent convolution for sparse view computed tomography imaging. Tomography. 2024;10(1):133–58. doi:10.3390/tomography10010011.
78. Cheng CC. Sparse-view tomographic reconstruction using residual U-Net with attention gates. SPIE. 2024;12926:351–8. doi:10.1117/12.2688209.
79. He T, Jiang X, Wu J, Wang W, Zhang H, Li Z. Dual-domain image reconstruction network integrating residual attention for sparse view computed tomography. In: Proceedings of the 2024 IEEE International Conference on Medical Artificial Intelligence (MedAI); 2024 Nov 15–17; Chongqing, China. p. 282–7.
80. Yang C, Sheng D, Yang B, Zheng W, Liu C. A dual-domain diffusion model for sparse-view CT reconstruction. IEEE Signal Process Lett. 2024;31:1279–83. doi:10.1109/lsp.2024.3392690.
81. Sun C, Salimi Y, Angeliki N, Boudabbous S, Zaidi H. An efficient dual-domain deep learning network for sparse-view CT reconstruction. Comput Methods Programs Biomed. 2024;256:108376. doi:10.1109/nss/mic/rtsd57108.2024.10655255.
82. Li Y, Sun X, Wang S, Guo L, Qin Y, Pan J, et al. TD-STrans: tri-domain sparse-view CT reconstruction based on sparse transformer. Comput Methods Programs Biomed. 2025;260:108575.
83. Shao J, Chen H, Li Q, Huang X, Shu J, Liu L, et al. Multi-stage dual-domain progressive network with synergistic training for sparse-view CT reconstruction. Neural Netw. 2026;195(12):108221. doi:10.1016/j.neunet.2025.108221.
84. Chen H, Zhang Y, Chen Y, Zhang J, Zhang W, Sun H, et al. LEARN: learned experts' assessment-based reconstruction network for sparse-data CT. IEEE Trans Med Imaging. 2018;37(6):1333–47.
85. Zhang H, Liu B, Yu H, Dong B. MetaInv-Net: meta inversion network for sparse view CT image reconstruction. IEEE Trans Med Imaging. 2021;40(2):621–34.
86. Wang S, Li X, Chen P. ADMM-SVNet: an ADMM-based sparse-view CT reconstruction network. Photonics. 2022;9(3):186.
87. Wu W, Guo X, Chen Y, Wang S, Chen J. Deep embedding-attention-refinement for sparse-view CT reconstruction. IEEE Trans Instrum Meas. 2023;72(1):1–11. doi:10.1109/tim.2022.3221136.
88. Cheng W, He J, Liu Y, Zhang H, Wang X, Liu Y, et al. CAIR: combining integrated attention with iterative optimization learning for sparse-view CT reconstruction. Comput Biol Med. 2023;163:107161.
89. Wu J, Jiang X, Zhong L, Zheng W, Li X, Lin J, et al. Linear diffusion noise boosted deep image prior for unsupervised sparse-view CT reconstruction. Phys Med Biol. 2024;69(16):165029. doi:10.1088/1361-6560/ad69f7.
90. Fan X, Chen K, Yi H, Yang Y, Zhang J. MVMS-RCN: a dual-domain unified CT reconstruction with multi-sparse-view and multi-scale refinement-correction. IEEE Trans Comput Imaging. 2024;10:1749–62.
91. Wu W, Pan J, Wang Y, Wang S, Zhang J. Multi-channel optimization generative model for stable ultra-sparse-view CT reconstruction. IEEE Trans Med Imaging. 2024;43(10):3461–75. doi:10.1109/tmi.2024.3376414.
92. Xu K, Lu S, Huang B, Wu W, Liu Q. Stage-by-stage wavelet optimization refinement diffusion model for sparse-view CT reconstruction. IEEE Trans Med Imaging. 2024;43(10):3412–24. doi:10.1109/tmi.2024.3355455.

93. Cheng Y, Li Q, Li R, Wang T, Zhao J, Yan Q, et al. LIR-Net: learnable iterative reconstruction network for fan beam CT sparse-view reconstruction. IEEE Trans Med Imaging. 2024;10:181–95.
94. Wang Y, Ren J, Cai A, Wang S, Liang N, Li L, et al. Hybrid-domain integrative transformer iterative network for spectral CT imaging. IEEE Trans Instrum Meas. 2024;73:1–13. doi:10.1109/tim.2024.3379388.
95. Du C, Lin X, Wu Q, Tian X, Su Y, Luo Z, et al. DPER: diffusion prior driven neural representation for limited angle and sparse view CT reconstruction. arXiv:2404.17890. 2024.
96. Li Z, Chang D, Zhang Z, Luo F, Liu Q, Zhang J, et al. Dual-domain collaborative diffusion sampling for multi-source stationary computed tomography reconstruction. IEEE Trans Med Imaging. 2024;43(10):3398–411. doi:10.1109/tmi.2024.3420411.
97. Kang Y, Liu J, Wu F, Wang K, Qiang J, Hu D, et al. Deep convolutional dictionary learning network for sparse view CT reconstruction with a group sparse prior. Comput Methods Programs Biomed. 2024;244(9840):108010. doi:10.1016/j.cmpb.2024.108010.
98. Xu S, Fu J, Sun Y, Cong P, Xiang X. Deep radon prior: a fully unsupervised framework for sparse-view CT reconstruction. Comput Biol Med. 2025;189:109853.
99. Kavur AE, Gezer NS, Barış M, Aslan S, Conze PH, Groza V, et al. CHAOS challenge-combined (CT-MR) healthy abdominal organ segmentation. Med Image Anal. 2021;69(4):101950. doi:10.1016/j.media.2020.101950.
100. Li P, Wang S, Li T, Lu J, Huangfu Y, Wang D. A large-scale CT and PET/CT dataset for lung cancer diagnosis. Little Rock, AR, USA: The Cancer Imaging Archive; 2020. doi:10.7937/TCIA.2020.NNC2-0461.
101. He L, Du W, Liao P, Fan F, Chen H, Yang H, et al. Solving zero-shot sparse-view CT reconstruction with variational score solver. IEEE Trans Med Imaging. 2025;44(9):3586–99. doi:10.1109/tmi.2024.3475516.
102. Wang S, Sun X, Li Y, Wei Z, Guo L, Li Y, et al. ADMM-TransNet: ADMM-based sparse-view CT reconstruction method combining convolution and transformer network. Tomography. 2025;11(3):23.
103. Tian X, Chen L, Wu Q, Du C, Shi J, Wei H, et al. Unsupervised self-prior embedding neural representation for iterative sparse-view CT reconstruction. Proc AAAI Conf Artif Intell. 2025;39(7):7383–91. doi:10.1609/aaai.v39i7.32794.
104. Shakouri S, Bakhshali MA, Layegh P, Kiani B, Masoumi F, Ataei Nakhaei S, et al. COVID19-CT-dataset: an open-access chest CT image repository of 1000+ patients with confirmed COVID-19 diagnosis. BMC Res Notes. 2021;14(1):178. doi:10.1186/s13104-021-05592-x.
105. National Cancer Institute Cancer Moonshot Biobank. Radiology data from the cancer moonshot biobank-colorectal cancer (CMB-CRC) collection. Little Rock, AR, USA: The Cancer Imaging Archive; 2021. doi:10.7937/DJG7-GZ87.
106. Wu J, Lin J, Jiang X, Zheng W, Zhong L, Pang Y, et al. Dual-domain deep prior guided sparse-view CT reconstruction with multi-scale fusion attention. Sci Rep. 2025;15(1):16894. doi:10.1038/s41598-025-02133-5.
107. Dou J, Fang J, Tang J, Zhong L, Di J, Qin Y. Extremely sparse-view CT reconstruction under joint physical and sparse constraints. Opt Lasers Eng. 2026;203(2):109802. doi:10.1016/j.optlaseng.2026.109802.
108. Li H, Han S, Mao H, Shi Y, Fang C, Zhang J, et al. Cross-distribution diffusion priors-driven iterative reconstruction for sparse-view CT. IEEE Trans Med Imaging. 2026;45(7):3878–94. doi:10.1109/tmi.2026.3687173.
109. AIMI S. Coca-coronary calcium and chest CT's dataset. Stanford AIMI; 2022 [cited 2026 Jan 1]. Available from: https://stanfordaimi.azurewebsites.net/datasets/e8ca74dc-8dd4-4340-815a-60b41f6cb2aa.
110. Segars WP, Sturgeon G, Mendonca S, Grimes J, Tsui BMW. 4D XCAT phantom for multimodality imaging research. Med Phys. 2010;37(9):4902–15. doi:10.1118/1.3480985.
111. Borsdorf A, Raupach R, Flohr T, Hornegger J. Wavelet based noise reduction in CT-images using correlation analysis. IEEE Trans Med Imaging. 2008;27(12):1685–703. doi:10.1109/tmi.2008.923983.
112. Sheng K, Gou S, Wu J, Qi SX. Denoised and texture enhanced MVCT to improve soft tissue conspicuity. Med Phys. 2014;41(10):101916. doi:10.1118/1.4894714.
113. Ravishankar S, Ye JC, Fessler JA. Image reconstruction: from sparsity to data-adaptive methods and machine learning. Proc IEEE. 2020;108(1):86–109.
114. Lempitsky V, Vedaldi A, Ulyanov D. Deep image prior. In: Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition; 2018 Jun 18–23; Salt Lake City, UT, USA. p. 9446–54.

115. Beck A, Teboulle M. A fast iterative shrinkage-thresholding algorithm for linear inverse problems. SIAM J Imaging Sci. 2009;2(1):183–202. doi:10.1137/080716542.
116. Li Y, Fu X, Li H, Zhao S, Jin R, Zhou SK. 3DGR-CT: sparse-view CT reconstruction with a 3D gaussian representation. Med Image Anal. 2025;103:103585.
117. Kerbl B, Kopanas G, Leimkuehler T, Drettakis G. 3D gaussian splatting for real-time radiance field rendering. ACM Trans Graph. 2023;42(4):3592433. doi:10.1145/3592433.
118. An J, Zhang X. DSV-CTGS: dynamic sparse-View CT reconstruction based on gaussian splatting and prior transfer. In: Proceedings of the ICASSP 2026—2026 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP); 2026 May 4–8; Barcelona, Spain. p. 11552–6.
119. Zhang Y, Chen W, Zhang G. CBR: a cardinal B-spline representation for sparse-view CT reconstruction. In: Proceedings of the ICASSP 2026—2026 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP); 2026 May 4–8; Barcelona, Spain. p. 7311–5.
120. Yang L, Huang J, Fang Y, Aviles-Rivero AI, Schönlieb CB, Zhang D, et al. Learning task-specific sampling strategy for sparse-view CT reconstruction. IEEE Trans Instrum Meas. 2025;74(2):1–11. doi:10.1109/tim.2025.3554318.
121. Ronneberger O, Fischer P, Brox T. U-Net: convolutional networks for biomedical image segmentation. In: Proceedings of the Medical Image Computing and Computer-Assisted Intervention—MICCAI 2015; 2015 Oct 5–9; Munich, Germany. p. 234–41.
122. Chi L, Jiang B, Mu Y. Fast fourier convolution. In: Proceedings of the 34th International Conference on Neural Information Processing Systems. NIPS'20; 2020 Dec 6–12; Vancouver BC Canada.
123. Vaswani A. Attention is all you need. In: Proceedings of the 31st Conference on Neural Information Processing Systems (NIPS 2017); 2017 Dec 4–9; Long Beach, CA, USA.
124. Dosovitskiy A, Beyer L, Kolesnikov A, Weissenborn D, Zhai X, Unterthiner T, et al. An image is worth 16 × 16 words: transformers for image recognition at scale. arXiv:2010.11929. 2010.
125. Liu Z, Lin Y, Cao Y, Hu H, Wei Y, Zhang Z, et al. Swin transformer: hierarchical vision transformer using shifted windows. In: Proceedings of the IEEE/CVF International Conference on Computer Vision; 2021 Oct 10–17; Montreal, QC, Canada. p. 10012–22.
126. Wu S, Wu T, Tan H, Guo G. Pale transformer: a general vision transformer backbone with pale-shaped attention. arXiv:2112.14000. 2021.
127. Song Y, Sohl-Dickstein J, Kingma DP, Kumar A, Ermon S, Poole B. Score-based generative modeling through stochastic differential equations. arXiv:2011.13456. 2020.
128. Bansal A, Borgnia E, Chu HM, Li JS, Kazemi H, Huang F, et al. Cold diffusion: inverting arbitrary image transforms without noise. In: Proceedings of the Thirty-seventh Conference on Neural Information Processing Systems; 2023 Dec 10–16; Orleans, LA, USA.
129. Hyun S, Yun S, Lee S, Choi Di, Cho S. Diffusion prior-guided implicit neural representation for metal artifact reduction in sparse-view CT reconstruction. Phys Med Biol. 2026;71(10):105007. doi:10.1088/1361-6560/ae6a5a.
130. Breger A, Karner C, Selby I, Gröhl J, Dittmer S, Lilley E, et al. A study on the adequacy of common IQA measures for medical images. In: International Conference on Medical Imaging and Computer-Aided Diagnosis. Berlin/Heidelberg, Germany: Springer; 2024. p. 451–62.
131. Lee W, Wagner F, Galdran A, Shi Y, Xia W, Wang G, et al. Low-dose computed tomography perceptual image quality assessment. Med Image Anal. 2025;99(3):103343. doi:10.1016/j.media.2024.103343.
132. Ries A, Dorosti T, Thalhammer J, Sasse D, Sauter A, Meurer F, et al. Improving image quality of sparse-view lung tumor CT images with U-Net. Eur Radiol Exp. 2024;8(1):54. doi:10.1186/s41747-024-00450-4.
133. Sefercioglu N, Unal MO, Ertas M, Yildirim I. Task-adaptive low-dose CT reconstruction. arXiv:2511.07094. 2025.
134. Clark K, Vendt B, Smith K, Freymann J, Kirby J, Koppel P, et al. The cancer imaging archive (TCIA): maintaining and operating a public information repository. J Digit Imaging. 2013;26(6):1045–57.
135. McCollough C, Chen B, Holmes DR III, Duan X, Yu Z, Yu L, et al. Low dose CT image and projection data (LDCT-and-projection-data) (Version 6) [Data set]. The Cancer Imaging Archive; 2020 [cited 2025 Apr 10]. doi:10.7937/9NPB-263710.7937/9npb-2637.

136. Moen TR, Chen B, Holmes DR III, Duan X, Yu X, Yu Z, et al. Low-dose CT image and projection dataset. Med Phys. 2021;48(2):902–11. doi:10.1002/mp.14594.
137. Armato SG III, McLennan G, Bidaut L, McNitt-Gray MF, Meyer CR, Reeves AP, et al. Data from LIDC-IDRI (Version 4) [Data set]. Little Rock, AR, USA: The Cancer Imaging Archive; 2015. doi:10.7937/K9/TCIA.2015.LO9QL9SX.
138. Tsai EB, Simpson S, Lungren MP, Hershman M, Roshkovan L, Colak E, et al. The RSNA international COVID-19 open radiology database (RICORD). Radiology. 2021;299(1):E204–13. doi:10.1148/radiol.2021203957.
139. Yan K, Wang X, Lu L, Summers RM. DeepLesion: automated mining of large-scale lesion annotations and universal lesion detection with deep learning. J Med Imaging. 2018;5(3):036501. doi:10.1117/1.JMI.5.3.036501.
140. Yan K, Wang X, Lu L, Summers RM. DeepLesion: a large-scale and diverse CT lesion dataset; 2018 [cited 2026 Apr 10]. Available from: https://nihcc.app.box.com/v/DeepLesion.
141. Barbano R, Kereta ž., Hauptmann A, Arridge SR, Jin B. Unsupervised knowledge-transfer for learned image reconstruction. Inverse Probl. 2022;38(10):104004. doi:10.1088/1361-6420/ac8a91.
142. Hu Y, Zhou H, Cao N, Li C, Hu C. Synthetic CT generation based on CBCT using improved vision transformer CycleGAN. Sci Rep. 2024;14(1):11455. doi:10.1038/s41598-024-61492-7.