



ARTICLE

DDF: A Dual-Domain Fusion Network for Sparse-View CT Reconstruction

Ziqiao Pan^{1,2} , Yuxi Chen², Bo Yang^{1,*}, Legend Zhang³, Renee Zeng³, Xiaoran Ma⁴, Zhixin Liu⁵, Wei Wei⁶ and Guangyu Xu⁷

¹School of Automation Engineering, University of Electronic Science and Technology of China, Chengdu, China

²Glasgow College, University of Electronic Science and Technology of China, Chengdu, China

³Future Tech Institute, Guangzhou Huashang University, Guangzhou, China

⁴School of Life Sciences, Peking University, Beijing, China

⁵School of Urban and Environmental Sciences, Shaoxing Institute of Technology, Shaoxing, China

⁶School of Intelligent Manufacturing, Xi'an University, Xi'an, China

⁷School of the Environment, The University of Queensland, Brisbane St Lucia, QLD, Australia

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

Received: 22 July 2026; Accepted: 18 September 2026; Published: 21 October 2026

ABSTRACT: Sparse-view computed tomography (SVCT) is a critical technique for reducing radiation exposure, yet it inherently suffers from severe reconstruction artifacts. While dual-domain deep learning methods have advanced SVCT restoration, prevailing cascaded and parallel architectures are limited by one-way information flow or shallow interaction between the two domains. To address this limitation, this work proposes a Dual-Domain Fusion (DDF) framework that establishes a physics-guided, closed-loop feedback mechanism. Unlike conventional pipelines, DDF explicitly maps the intermediate image-domain prediction back to the projection domain via forward projection. Through tailored Sine Fusion and computed tomography (CT) Fusion modules, DDF adaptively integrates structurally rich image feedback with features extracted from the measured sinogram, enabling complementary information exchange and mitigating sinogram information loss. The framework is instantiated using ViewTrans and the Nonlinear Activation Free Network (NAFNet) as the projection- and image-domain backbones, respectively, within a modular formulation that facilitates flexible backbone substitution. Experiments on the 2016 Low-Dose CT Grand Challenge dataset demonstrate strong performance under severe sparsity. At a sparsity factor of 12, DDF achieves a peak signal-to-noise ratio (PSNR) of 36.71 dB and outperforms the matched cascaded architecture by 5.91 dB on the fixed 200-slice test set.

KEYWORDS: Sparse-view CT; CT reconstruction; dual-domain fusion; deep learning

1 Introduction

X-ray computed tomography (CT) is widely used in clinical diagnosis and treatment planning. However, it exposes patients to ionizing radiation, which may increase potential health risks, especially for radiation-sensitive populations [1]. Reducing radiation dose has therefore become an important topic. Two commonly used strategies are lowering the tube current or voltage and reducing the number of projection views. The latter strategy, usually referred to as sparse-view CT (SVCT), directly decreases the angular sampling density during scanning and can also construct paired supervised data by downsampling full-view sinograms. However, insufficient angular measurements make the reconstruction problem challenging because classical

filtered back-projection (FBP), which relies on sufficiently sampled projection data, often produces streak artifacts and structural distortion under sparse-view conditions [2,3].

Deep learning has been widely explored for sparse-view CT reconstruction [4]. According to the data domain in which the network operates, existing methods can be broadly divided into sinogram-domain (projection-domain), image-domain, and dual-domain approaches. Sinogram-domain methods aim to restore missing or degraded projection data before reconstruction, while image-domain methods take sparse-view FBP images as contaminated inputs and learn to improve image quality [5–8]. Dual-domain methods further combine projection-domain correction and image-domain refinement, and have shown stronger reconstruction ability by using information from both domains [9–12]. Compared with single-domain methods, dual-domain reconstruction is more consistent with the physical imaging process because it considers both measured projection data and reconstructed CT images.

Despite their effectiveness, existing dual-domain methods typically treat cross-domain interaction as a loosely coupled or linear process, falling into either cascaded or parallel paradigms. Conventional cascaded architectures exhibit a strictly one-way information flow (from sinogram to image), which isolates the final refinement from raw projection constraints and often yields over-smoothed results. Conversely, parallel architectures process domains concurrently but rely on naive, late-stage feature concatenation. Unlike generic multimodal features, the sinogram and image domains in CT are connected by known projection operators. This relationship motivates feeding the intermediate image prediction back into the sinogram branch rather than allowing information to flow only from the sinogram domain to the image domain.

For this reason, the Dual-Domain Fusion (DDF) framework is proposed. Its “Feedback and Fusion” strategy transforms the conventional open-loop pipeline into a dynamically coupled, closed-loop system. Specifically, after initial processing, the structurally enhanced image prediction is explicitly mapped back to the sinogram space via a deterministic forward projection operator. This formulation enables the projection-domain restoration to leverage anatomical information for geometry-aware compensation. Rather than performing a trivial residual addition, a dedicated Sine Fusion module adaptively integrates this structurally rich feedback with projection features estimated from the measurements. Finally, a CT Fusion module reconciles the harmonized sinogram reconstruction with the intermediate image predictions. By embedding fixed physical operators within this feedback loop, DDF enables information learned in each domain to inform the other before the final reconstruction.

The main contribution of this work is the DDF framework, an architecture for sparse-view CT reconstruction that addresses the limitations of traditional one-way cascades by introducing explicit forward projection feedback combined with tailored Sine Fusion and CT Fusion modules. This design establishes a geometry-guided closed loop, enabling the final reconstruction to leverage complementary information from both domains. An effective instantiation uses ViewTrans [13] in the projection domain to capture view-wise dependencies and NAFNet [14] in the image domain to suppress residual artifacts, while maintaining a modular, operator-based design that allows flexible substitution of these domain-specific networks. Experiments on the 2016 Low-Dose CT Grand Challenge dataset demonstrate strong performance under severe sparsity conditions; notably, at a sparsity factor of 12, DDF achieves a PSNR of 36.71 dB and a structural similarity index measure (SSIM) of 0.931, outperforming the matched cascaded pipeline by 5.91 dB.

2 Related Work

Sparse-view CT reconstruction has been studied through analytical, iterative, and learning-based methods. Classical analytical methods, especially filtered back projection (FBP), are widely used because of their efficiency and simple implementation [2,3]. Their weakness becomes obvious in the sparse-view setting.

Once many angular measurements are absent, FBP has no mechanism to infer the missing projection information, so streak artifacts and structural distortion are directly propagated into the reconstructed image.

Model-based iterative methods improve sparse-view reconstruction by introducing prior knowledge into the optimization process. Typical methods, such as total variation minimization and compressed sensing, exploit image sparsity in gradient or transform domains and combine data fidelity with regularization terms [15,16]. They are still important because they keep the reconstruction tied to the imaging model. In practice, however, their repeated projection operations are expensive, and the reconstruction quality depends heavily on the chosen prior and regularization parameters.

Early deep learning approaches primarily operated within a single data domain. Image-domain networks, such as the Residual Encoder-Decoder Convolutional Neural Network (RED-CNN) [7] and the post-processing framework by Jin et al. [8], focus on eliminating artifacts from FBP reconstructions. While computationally efficient, these models process already-degraded inputs; consequently, they tend to over-smooth weak anatomical boundaries when sparse sampling causes an irreversible loss of high-frequency projection cues. Alternatively, sinogram-domain methods process the data prior to reconstruction. For instance, Lee et al. [5] synthesize missing projection views, the Sinogram Interpolation and Image Denoising (SIPID) framework [6] integrates sinogram interpolation with image denoising, and ViewTrans [13] models view-wise dependencies directly within the projection space. Although these studies demonstrate the critical value of projection-domain information, relying solely on upstream sinogram completion remains insufficient to eliminate the complex residual artifacts introduced during the subsequent FBP process. Broader reviews of deep learning in medical imaging also identify data efficiency, model generalization, and robustness across acquisition conditions as persistent challenges [17]; these considerations are directly relevant to the translation of learned CT reconstruction models.

Dual-domain methods emerged to bridge this gap by exploiting the complementary nature of projection and image representations. These approaches predominantly follow either cascaded or parallel paradigms. Cascaded architectures typically perform sequential restoration from the sinogram to the image domain [9], with models like DuDoTrans [10] incorporating Transformer blocks to capture long-range dependencies. In contrast, parallel networks such as DDPNet [11] process both domains simultaneously, while subsequent frameworks [12] introduce controllable modulation to adapt to varying degradation levels. However, a fundamental limitation across both paradigms remains the restricted depth of cross-domain interaction. Cascaded methods suffer from an inherently unidirectional information flow, preventing downstream image refinements from correcting upstream projection representations. Conversely, parallel designs typically optimize each branch independently for the majority of the network, relying only on shallow, late-stage feature aggregation. Consequently, existing architectures lack a dynamic feedback mechanism that enables image-domain structural corrections to actively re-enter and optimize the sinogram domain prior to the final reconstruction.

Recent work has pushed this area in three useful directions. First, Transformer and multi-domain models try to strengthen long-range modeling. The dual-domain Transformer reconstruction method of Shi et al. [18] treats sinogram information as a sequence-like signal, Multi-Domain Sparse-View CT Reconstruction (MDST) [19] combines convolution with Swin Transformer blocks, and Tri-Domain Sparse-View CT Reconstruction based on Sparse Transformer (TD-STrans) [20] adds the frequency domain to recover high-frequency details under very sparse sampling. Second, optimization-inspired methods try to make dual-domain learning more stable. The Multi-Sparse-View and Multi-Scale Refinement-Correction Network (MVMS-RCN) [21] refines projection errors and image-domain corrections in a unified unfolding framework, while the stable dual-domain reconstruction method of Ding et al. [22] formulates sparse-view CT as learned alternating minimization with regularizers in both data and image domains. Third, newer

prior-guided and generative methods have started to revisit data consistency. The method of Wu et al. [23] combines dual-domain deep priors, multi-scale fusion attention, and physics-informed consistency; the Dual-Domain Diffusion Model (DDDM) [24] uses a sinogram upgrading module followed by an image refining diffusion module. These recent studies are valuable for the present work because they confirm that stronger reconstruction requires more than a better image denoiser. The projection domain must remain active during reconstruction.

DDF follows this trend but makes a different design choice. Instead of adding another domain, using a heavy iterative prior, or relying on diffusion sampling, it introduces a deterministic feedback path from the refined CT image back to the sinogram domain. The forward-projected image prediction is not simply added to the restored sinogram. It is fused through the proposed Sine Fusion module, so image-domain structural cues can guide projection-domain compensation while the measured-projection representation is still preserved. The final CT Fusion module then reconciles the image-domain prediction with the feedback-enhanced reconstruction. In this sense, DDF is most closely related to dual-domain and recent physics-aware methods, but its core contribution is the closed-loop exchange between the two domains before producing the final CT image.

3 Methods

3.1 CT Physical Model

The physical basis of CT imaging is the attenuation of X-rays when they pass through an object [25]. Let I_0 and I denote the incident and transmitted X-ray intensities, respectively. For a homogeneous material, the attenuation process can be described by the Beer–Lambert law:

$$I = I_0 \exp(-\mu L), \quad (1)$$

where μ is the linear attenuation coefficient and L is the path length. In practical CT imaging, the scanned object is usually composed of tissues which are not uniform, and the detector measurement is determined by the accumulated attenuation along the X-ray path. After logarithmic transformation, the projection measurement can be written as

$$p = -\ln\left(\frac{I}{I_0}\right) = \int_L f(x, y) dl, \quad (2)$$

where $f(x, y)$ denotes the two-dimensional attenuation coefficient distribution of the scanned slice, and L denotes the corresponding ray path.

For clarity, the projection process can be written in the parallel-beam form as the Radon transform:

$$p(\theta, s) = \int \int f(x, y) \delta(x \cos \theta + y \sin \theta - s) dx dy, \quad (3)$$

where θ is the projection angle, s is the detector position, and $\delta(\cdot)$ denotes the Dirac delta function. Although the experiments in this work use fan-beam simulated projections, the same physical principle holds: each detector measurement records an integral of the attenuation distribution along an X-ray path, and the sinogram represents these measurements over different scanning views.

Let $Z \in \mathbb{R}^{M \times N}$ denote the full-view sinogram, where M is the number of detector bins and N is the number of full-view projection views. The measurement vector at the j -th view is denoted as $z_j \in \mathbb{R}^M$, and the full-view sinogram can be written as

$$Z = \{z_0, z_1, \dots, z_{N-1}\}. \quad (4)$$

For sparse-view CT, only a subset of projection views is measured. Given a sparse-view factor S , the sparse-view sinogram is defined as

$$Z_s = \{z_{kS}\}_{k=0}^{K-1}, \quad K = \frac{N}{S}, \quad (5)$$

where $Z_s \in \mathbb{R}^{M \times K}$. The reconstruction task is to estimate a high-quality CT image $\hat{X}$ from the incomplete projection data Z_s , with the full-view CT image X used as the reference target.

In this work, the projection domain and the image domain are connected by two fixed physical operators: filtered back-projection (FBP) and forward projection (FP). FBP is denoted as

$$X = \mathcal{B}(Z), \quad (6)$$

where $\mathcal{B}(\cdot)$ maps a sinogram into the image domain. FP is denoted as

$$Z = \mathcal{P}(X), \quad (7)$$

where $\mathcal{P}(\cdot)$ maps an image back into the projection domain.

Both operators are discretized using the same fan-beam geometry as that used to generate the training data. Specifically, the image grid is 256×256 , the source-to-rotation-center distance is 605 pixels, and the curved detector spans $[-17.8^\circ, 17.8^\circ]$ with an angular detector spacing of 0.1° . The MATLAB arc-detector parameterization defines rays using the source-to-rotation-center distance and detector fan angles; therefore, an independent source-to-detector distance is not introduced in this numerical setup. A full scan contains $N_\theta = 360$ uniformly sampled views and $N_s = 357$ detector bins. Let $A \in \mathbb{R}^{HW \times N_\theta N_s}$ denote the precomputed sparse geometric back-projection matrix. The implemented operators can be expressed as

$$\mathcal{B}(Z) = \text{reshape}(A \mathcal{H}(C \text{vec}(Z))), \quad (8)$$

$$\mathcal{P}(X) = \text{reshape}(A^T \text{vec}(X)), \quad (9)$$

where C denotes fan-beam cosine weighting and $\mathcal{H}$ denotes ramp filtering. Thus, FP is the transpose of the geometric back-projection matrix A , rather than the transpose or inverse of the complete filtered operator $\mathcal{B}$. During construction of A , a ray sample located between pixel centers distributes its contribution to the four neighboring pixels using bilinear interpolation weights. FP and FBP use the same precomputed coefficients, and their fan-beam geometry and coordinate convention match those used for projection simulation, thereby avoiding a systematic geometry mismatch in the feedback path. Nevertheless, $\mathcal{P}$ remains a discrete approximation to the continuous ray transform, so interpolation and grid discretization can introduce small residual errors. DDF therefore uses its output as a geometry-aware feedback representation that is adaptively weighted by Sine Fusion, rather than imposing it as an exact data-consistency equality.

The matrices in FP and FBP are fixed and contain no trainable parameters. Sparse matrix multiplication in FP is differentiable with respect to X , so the image-domain network receives gradients through the feedback path during end-to-end optimization; no gradient is used to update the projection matrix itself. The FBP operator provides the reconstruction path from sinogram to CT image, while FP provides a geometry-aware feedback path from image to sinogram. These two operators form the physical basis of the proposed dual-domain feedback fusion framework.

3.2 Motivation: Feedback-Based Dual-Domain Fusion

A common dual-domain reconstruction pipeline first restores the sparse sinogram in the projection domain, then reconstructs an intermediate CT image through FBP, and finally refines the image using an

image-domain network. This design follows the physical reconstruction process and is more effective than using either domain alone. However, its information flow is essentially unidirectional, from the projection domain to the image domain. Once the projection-domain output is transformed into the image domain, the final reconstruction is largely determined by the image domain refinement network. Consequently, part of the projection-domain correction may be weakened after subsequent image-domain processing.

Another limitation is the insufficient interaction between the restored sinogram and the refined CT image. The sinogram domain preserves the measurement structure of CT acquisition, while the image domain contains direct anatomical and structural representations. These two domains are complementary, but a simple cascaded pipeline does not explicitly project the image-domain prediction back to the projection domain. Therefore, the restored sinogram cannot benefit from the structural correction learned in the image domain, and the consistency between the two representations is not fully exploited.

Based on this observation, a Dual-Domain Fusion (DDF) framework with a feedback path from the image domain to the sinogram domain is proposed. The refined CT image is forward projected into the sinogram domain using the fixed operator $\mathcal{P}$ defined above and fused with the projection-domain prediction. The fused sinogram is then reconstructed again and combined with the image-domain prediction to obtain the final CT image. This design allows the projection and image domains to complement each other through explicit feedback and fusion.

3.3 Basic Reconstruction Pipeline

The proposed framework is built on a cascaded dual-domain backbone followed by a feedback path and two fusion modules. For clarity, the main notations used in the proposed framework are summarized in Section Nomenclature before describing the reconstruction process. In this paper, Z denotes sinogram-domain representations, while X denotes image-domain CT representations.

Given the sparse-view sinogram Z_s , a projection-domain restoration network first estimates a restored full-view sinogram:

$$Z_p = F_p(Z_s), \quad (10)$$

where $F_p(\cdot)$ denotes the projection-domain model.

The restored sinogram is transformed into the image domain by FBP:

$$X_p = FBP(Z_p). \quad (11)$$

Here, X_p is an intermediate reconstruction obtained from the projection-domain prediction. An image-domain restoration network then refines this image:

$$X_i = F_i(X_p), \quad (12)$$

where $F_i(\cdot)$ denotes the image-domain model.

A conventional cascaded dual-domain method would directly take X_i as the final reconstruction. In contrast, the proposed DDF framework further uses X_i to generate feedback information in the projection domain. The final reconstruction is obtained by

$$\hat{X} = F_{DDF}(Z_p, X_i), \quad (13)$$

where $F_{DDF}(\cdot)$ denotes the proposed dual-domain feedback fusion module.

3.4 Backbone Instantiation

The formulation above describes the reconstruction process with two restoration operators, namely the projection-domain model F_p and the image-domain model F_i . In the main implementation of this paper, these two operators are not left abstract. They are instantiated as ViewTrans and NAFNet, respectively, because the two networks match the characteristics of the projection domain and the image domain. The two backbone instantiations are summarized separately in Figs. 1 and 2.

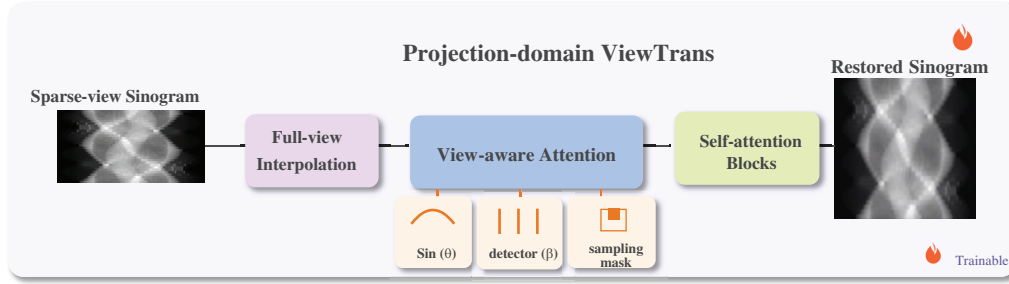


Figure 1: Projection-domain backbone. ViewTrans restores the sparse-view sinogram with angular, detector-position, and sampling-mask encodings.

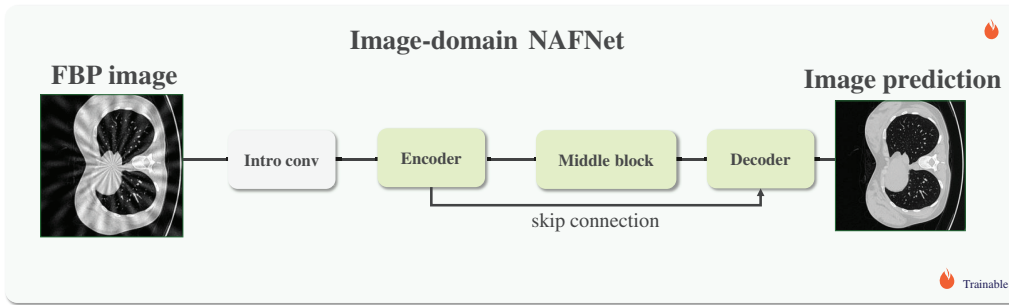


Figure 2: Image-domain backbone. NAFNet refines the FBP image through an encoder-decoder restoration branch with skip connections.

3.4.1 Projection-Domain Backbone: ViewTrans with Angular Encoding

For the projection-domain branch, F_p is implemented as ViewTrans, a view-aware sinogram restoration backbone. In the code implementation, ViewTrans corresponds to the `sin_angle` module. The input sparse-view sinogram is first interpolated to the full angular size, and ViewTrans then learns to compensate the missing angular information in the sinogram representation. Since sparse-view artifacts mainly come from missing projection angles, this branch focuses on modeling the relationship among views before FBP reconstruction.

Let the interpolated sinogram feature be denoted as $S \in \mathbb{R}^{N_\theta \times N_s}$, where N_θ and N_s are the numbers of projection views and detector bins, respectively. Before the first attention operation, ViewTrans attaches three simple physical cues to S : the angle code $\sin \theta$, the detector-position code β , and the sparse-view mask M . Layer normalization (LN) is applied to the sinogram feature before projection. Thus, the network can know which view it is processing, which detector bin is involved, and whether the view is originally sampled. These cues are concatenated with the normalized sinogram feature to generate the query and key:

$$[Q, K] = \phi_q ([\text{LN}(S), \sin \theta, \beta, M]), \quad (14)$$

where $\phi_q(\cdot)$ is a linear projection. The value feature is generated from $\text{LN}(S)$. After this view-aware attention block, four self-attention blocks further refine the sinogram along the view dimension. The output of this branch is the restored sinogram Z_p , which is further used by both the cascaded reconstruction path and the feedback fusion path.

3.4.2 Image-Domain Backbone: NAFNet

For the image-domain branch, F_i is implemented using NAFNet [14]. In the implementation used here, NAFNet is configured as a single-channel image restoration network with width 32, numbers of encoder blocks [1, 1, 1, 28], one middle block, and numbers of decoder blocks [1, 1, 1, 1]. This branch takes the FBP reconstruction X_p from the restored sinogram as input and learns to suppress streak artifacts and recover local image structures. Its encoder-decoder structure captures both global context and local details, and the skip connections help preserve image structures. The image-domain prediction X_i is not only an intermediate reconstruction result, but also the source of feedback information for the projection domain through FP. Therefore, ViewTrans and NAFNet are not used as two independent post-processing modules; instead, their outputs are reconnected by the proposed feedback and fusion design.

Although the main model uses ViewTrans and NAFNet, the mathematical interface of DDF is still written with F_p and F_i . This notation is useful for later ablation, where the projection-domain or image-domain backbone can be replaced while keeping the same FP feedback, Sine Fusion, and CT Fusion pipeline. However, the main method evaluated in this paper should be understood as the DDF architecture implemented on the ViewTrans and NAFNet dual-domain backbone.

3.5 Dual-Domain Fusion Framework

The structure of the proposed DDF framework is shown in Fig. 3. DDF starts from the two outputs of the cascaded backbone: the restored sinogram Z_p and the image-domain prediction X_i . The key idea is to project X_i back to the sinogram domain, fuse it with Z_p , and then return the fused information to the image domain for final reconstruction.

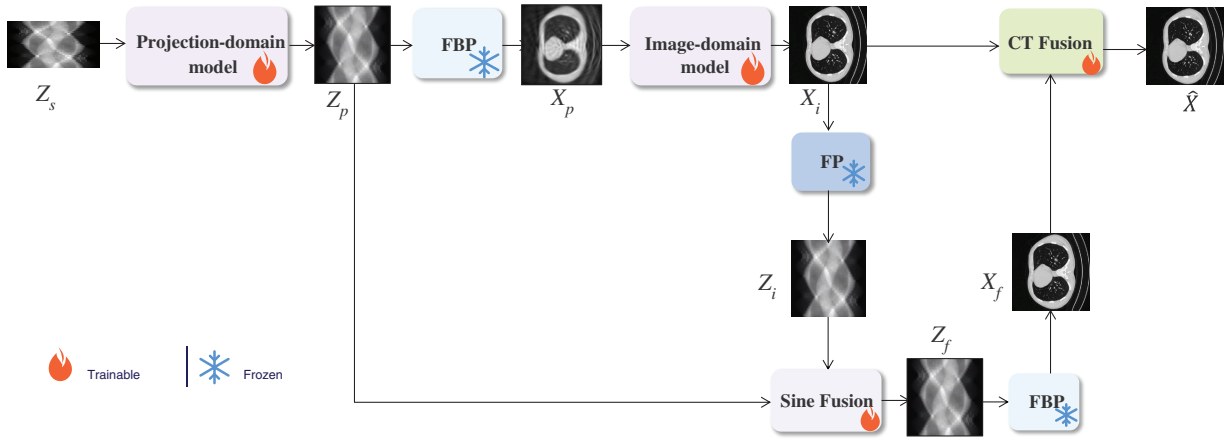


Figure 3: Overall structure of the Dual-Domain fusion framework.

3.5.1 Image-to-Sinogram Feedback

The image-domain prediction X_i contains structural information learned by the image-domain restoration network. To reuse this information in the projection domain, DDF applies the fixed, differentiable forward-projection operator to X_i :

$$Z_i = \mathcal{P}(X_i) = \text{reshape}(A^T \text{vec}(X_i)). \quad (15)$$

The obtained Z_i can be regarded as the sinogram-domain representation of the image-domain prediction. Different from Z_p , which is directly estimated from sparse-view measurements, Z_i carries image-domain structural correction after artifact removal. Therefore, Z_p and Z_i provide two complementary sinogram representations. Although $\mathcal{P}$ has no trainable parameters, gradients from Sine Fusion propagate through this sparse matrix multiplication to X_i and hence to the image-domain backbone.

3.5.2 Sine Fusion

Sine Fusion is designed to fuse two sinogram-domain representations: the restored sinogram Z_p produced by the projection-domain model and the feedback sinogram Z_i obtained by forward projecting the image-domain prediction. The structure of the Sine Fusion module is shown in Fig. 4.

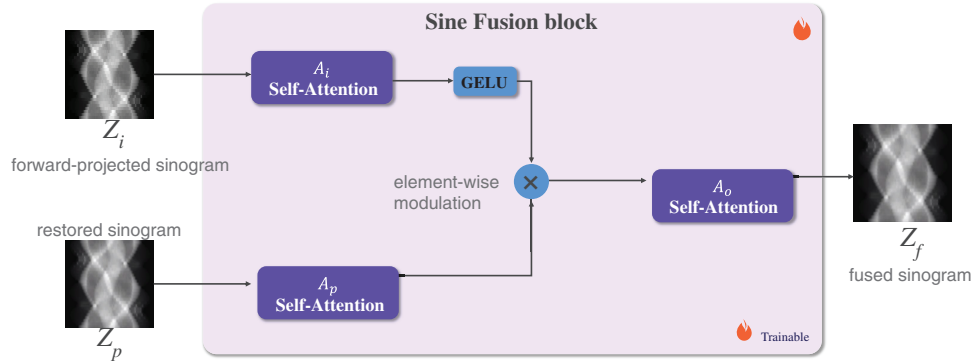


Figure 4: Structure of the proposed sine fusion module. Three single-head self-attention blocks operate along the projection-view dimension. The feedback-derived modulation signal is multiplied by the restored-sinogram feature, and the modulated representation is refined by an output attention block to produce Z_f .

Both inputs have the shape

$$Z_p, Z_i \in \mathbb{R}^{B \times N_\theta \times N_s}, \quad (16)$$

where B denotes the batch size, $N_\theta = 360$ is the number of projection views in the full-view representation, and $N_s = 357$ is the number of detector bins.

The module contains three independent Transformer-style self-attention blocks, denoted by $A_i(\cdot)$, $A_p(\cdot)$, and $A_o(\cdot)$. Each block uses one attention head and preserves the input dimensions. The projection views are treated as tokens, while the detector-bin dimension forms the feature embedding. Thus, the attention operations model dependencies among different projection views rather than spatial dependencies in the reconstructed CT image.

For an input $Z \in \mathbb{R}^{B \times N_\theta \times N_s}$, each attention block first applies layer normalization and constructs query, key, and value representations:

$$Q = ZW_Q, \quad K = ZW_K, \quad V = ZW_V. \quad (17)$$

Single-head scaled dot-product self-attention is then computed along the projection-view dimension:

$$\text{Attn}(Z) = \text{Softmax}\left(\frac{QK^T}{\sqrt{N_s}}\right)V. \quad (18)$$

The attention output is followed by a linear projection and a residual connection. A second layer-normalization operation and a feed-forward network with Gaussian Error Linear Unit (GELU) activation are then applied with another residual connection. Together, these operations constitute each $A_k(\cdot)$, where $k \in \{i, p, o\}$.

The feedback sinogram is first transformed into a modulation signal:

$$G = \text{GELU}(A_i(Z_i)). \quad (19)$$

In parallel, the restored sinogram is processed to obtain a projection-domain feature:

$$R_p = A_p(Z_p). \quad (20)$$

The two representations are combined through element-wise modulation and subsequently refined by the output attention block:

$$Z_f = A_o(G \odot R_p), \quad (21)$$

where $\odot$ denotes element-wise multiplication. The first branch introduces structural information derived from the image-domain prediction, whereas the second branch retains information learned directly from the projection measurements. The output attention block further models view-wise dependencies after modulation and produces the fused sinogram used by the subsequent FBP operation.

3.5.3 CT Fusion

After Sine Fusion, the fused sinogram Z_f is reconstructed into the image domain through FBP:

$$X_f = \text{FBP}(Z_f). \quad (22)$$

The final reconstruction is obtained by fusing X_f with the image-domain prediction X_i . CT Fusion adopts the Cross Gating Block (CGB) introduced in Multi-Axis Multilayer Perceptron for Image Processing (MAXIM) [26], whose structure is shown in Fig. 5. Unlike scalar or channel-only gating, CGB performs complementary global-grid and local-block spatial mixing before exchanging information between the two image-domain branches.

Let $X_i, X_f \in \mathbb{R}^{B \times C_{in} \times H \times W}$. Two trainable 1×1 convolutions first align both inputs to C channels:

$$X_1 = P_x(X_i), \quad Y_1 = P_y(X_f), \quad (23)$$

where $X_1, Y_1 \in \mathbb{R}^{B \times H \times W \times C}$ after conversion to the channel-last layout used by the gating multilayer perceptron (MLP). Each branch is then normalized, linearly projected along the channel dimension, and activated:

$$X_2 = \text{GELU}(W_1 \text{LN}(X_1)), \quad Y_2 = \text{GELU}(W_2 \text{LN}(Y_1)). \quad (24)$$

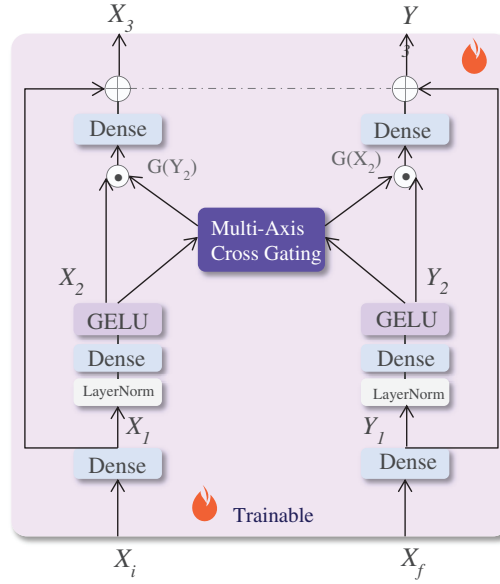


Figure 5: Structure of the cross gating block used in CT Fusion. The multi-axis gating operator combines global-grid and local-block spatial mixing, and the two branches modulate one another through element-wise multiplication followed by trainable output projections and residual aggregation.

The multi-axis gating operator $G(\cdot)$ acts on the spatial dimensions while preserving the tensor shape. For $T \in \mathbb{R}^{B \times H \times W \times C}$, a channel projection first expands the representation to $2C$ channels and divides it into two equal parts:

$$[U, V] = \text{Split}(\text{GELU}(W_{\text{in}} \text{LN}(T))), \quad (25)$$

where $U, V \in \mathbb{R}^{B \times H \times W \times C}$. The first half is partitioned into a fixed $g_h \times g_w$ grid. At every within-cell location, a grid MLP mixes the $g_h g_w$ spatial cells, thereby capturing long-range dependencies. The second half is partitioned into non-overlapping $f_h \times f_w$ blocks, and a block MLP mixes the $f_h f_w$ positions within each block to capture local structure. Denoting the corresponding reshape-MLP-inverse-reshape operations by $\mathcal{R}_g$ and $\mathcal{R}_b$, respectively, the gating response is

$$G(T) = W_{\text{out}}[\mathcal{R}_g(U) \parallel \mathcal{R}_b(V)], \quad (26)$$

where $\parallel$ denotes channel-wise concatenation. In the implementation used here, $g_h = g_w = f_h = f_w = 2$. Thus, the grid branch mixes globally distributed cells and the block branch mixes neighboring pixels; the operator is spatial rather than channel-only gating.

The two independently parameterized gating operators generate cross-branch modulation signals, $G_x(X_2)$ and $G_y(Y_2)$. Consistent with the implementation, the feedback-enhanced branch is updated first, and its updated representation also contributes to the final image branch:

$$Y_3 = Y_1 + \mathcal{D} \left(W_y^o (Y_2 \odot G_x(X_2)) \right), \quad (27)$$

$$X_3 = X_1 + Y_3 + \mathcal{D} \left(W_x^o (X_2 \odot G_y(Y_2)) \right), \quad \hat{X} = X_3, \quad (28)$$

where W_x^o and W_y^o are trainable output projections, $\mathcal{D}$ denotes dropout, and $\odot$ is element-wise multiplication. This ordered residual update matches the implementation: X_3 aggregates its own gated feature, the updated Y branch, and the original X shortcut.

No custom weight initialization is imposed. The convolutional and linear projections use the default PyTorch initialization, while the scale and bias of each layer-normalization operation are initialized to one and zero, respectively. All CGB parameters are optimized jointly with the remaining DDF network by the end-to-end image reconstruction objective. For one gating operator, the channel projections require $\mathcal{O}(BHW C^2)$ operations, whereas grid and block mixing require $\mathcal{O}(BHW C(g_h g_w + f_h f_w))$. Therefore, with fixed grid and block sizes, its computational cost grows linearly with the image area HW , avoiding the quadratic spatial cost of full self-attention.

3.6 Training Objective

The DDF is trained using an image-domain reconstruction objective. DDF does not include a sinogram-domain consistency loss or measured-view replacement. Forward projection is therefore used to provide geometry-aware feedback rather than to enforce exact data consistency. Projection-domain supervision is not used in the final training stage because the target of the task is the reconstructed CT image and a lower discrete sinogram-domain error does not necessarily produce a better image after filtering and back-projection. In addition, directly constraining the feedback representation to equal the simulated full-view sinogram would treat the discretized FP operator as exact, despite the interpolation and grid-approximation errors discussed above. A perceptual loss is also omitted because feature-space similarity can favor visually plausible textures without guaranteeing attenuation-value fidelity. These choices keep the objective aligned with pixel-wise reconstruction accuracy and isolate the effect of the proposed feedback architecture without introducing an additional loss weight. Let X denote the ground truth image and $\hat{X}$ denote the final output of DDF. According to the forward process described above, the final reconstruction is

$$\hat{X} = F_{CF}(X_i, X_f), \quad (29)$$

where X_i is the image-domain prediction and $X_f = FBP(Z_f)$ is the feedback-enhanced reconstruction obtained from the fused sinogram.

The training loss is defined as the mean squared error between the final reconstruction and the reference CT image:

$$\mathcal{L}_{\text{img}} = \frac{1}{|\Omega|} \|\hat{X} - X\|_2^2, \quad (30)$$

where Ω denotes the number of pixels.

A measured-view fidelity term could nevertheless be incorporated as

$$\mathcal{L}_{\text{meas}} = \|M \odot (\mathcal{P}(\hat{X}) - Z_s^\dagger)\|_2^2, \quad (31)$$

where M selects the acquired views and $Z_s^\dagger$ denotes the sparse measurements placed on the full-view angular grid. Optimizing $\mathcal{L}_{\text{img}} + \lambda_{\text{meas}} \mathcal{L}_{\text{meas}}$ may further strengthen data consistency, particularly under extreme sparsity, but it requires a controlled study of λ_{meas} and robustness to operator mismatch. This extension is therefore left for future work rather than being introduced without an ablation in the present study.

4 Experiments

4.1 Dataset, Training Strategy, and Experiment Setting

Dataset and projection simulation: The proposed DDF framework is evaluated on the public 2016 National Institutes of Health–American Association of Physicists in Medicine (NIH–AAPM)–Mayo Clinic Low Dose CT Grand Challenge dataset. In this study, 2000 CT slices from 10 patients are used. A patient-wise partition is adopted: 1600 slices from eight patients are used for training, 200 slices from a ninth patient are used for validation, and the remaining 200 slices from the held-out tenth patient form the test set. No patient contributes slices to more than one subset. The validation set is used exclusively for model selection, whereas the test set remains unseen until the final evaluation. The original CT images are stored in Digital Imaging and Communications in Medicine (DICOM) format with a spatial resolution of 512×512 . To reduce computational cost, all CT images are resized to 256×256 before projection simulation and network training.

Full-view projection data are generated by numerical forward projection. Specifically, the fanbeam function in MATLAB is used to simulate a fan-beam CT scanning geometry. The distance between the X-ray source and the slice center is set to 605 pixels. The detector covers an angular range from -17.8° to $+17.8^\circ$ with an interval of 0.1° . After one full circular scan, each full-view sinogram contains 360 projection views and 357 detector bins. The same geometry and coordinate convention are used to construct the fixed sparse matrices employed by FP and FBP in DDF.

Training setting: The DDF is trained end to end under the supervision of ground-truth CT images. The FP and FBP operators are frozen throughout training, while gradients through FP are propagated to the preceding image-domain network. The image-domain mean squared error loss defined in the training objective is adopted for optimization, so that the training objective is directly consistent with the final reconstruction quality.

Adam is used as the optimizer with an initial learning rate of 1×10^{-4} . The batch size is set to 3, and the model is trained for 500 epochs. A step learning-rate scheduler is adopted, where the learning rate is multiplied by 0.33 every 10 epochs. The checkpoint with the best validation PSNR is selected, and its performance is then evaluated once on the held-out test set.

To examine whether this schedule causes premature optimization stagnation, a controlled scheduler-sensitivity experiment is conducted at $S = 12$. StepLR, cosine annealing, and ReduceLROnPlateau are compared under the same fixed 100-epoch budget, 1600/200 training/validation split, Adam optimizer, batch size, random seed, and shared initial weights. Fig. 6 shows the validation trajectories and learning-rate schedules, and Table 1 reports the validation-selected checkpoints. These validation-patient values are not directly comparable to the held-out test-patient values in Table 2.

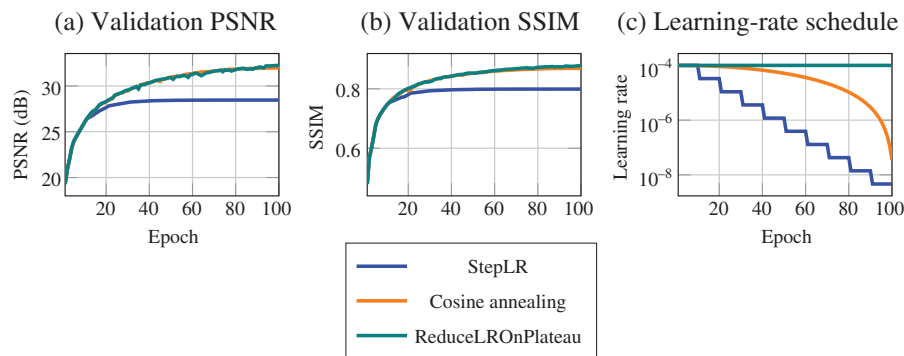


Figure 6: Controlled scheduler-sensitivity analysis at $S = 12$. All runs use the same data split, initialization, optimizer, and 100-epoch budget. Panels show validation PSNR, validation SSIM, and the applied learning rate.

Table 1: Validation results of the scheduler-sensitivity experiment at $S = 12$. The reported PSNR and SSIM correspond to the checkpoint selected by the highest validation PSNR.

Scheduler	Best Epoch	PSNR (dB)	SSIM	Final LR
StepLR	98	28.459	0.7992	4.64×10^{-9}
Cosine annealing	100	31.963	0.8691	3.47×10^{-8}
ReduceLROnPlateau	99	32.247	0.8775	1.00×10^{-4}

Table 2: Quantitative comparison under different sparse factors. Bold values indicate the best performance among all compared methods.

Method	$S = 2$		$S = 4$		$S = 8$		$S = 12$	
	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM
Sparse FBP	24.85	0.793	23.37	0.717	21.36	0.619	19.97	0.567
RED-CNN	24.95	0.798	23.50	0.728	21.41	0.628	20.00	0.574
DuDoTrans	37.55	0.933	35.69	0.919	32.41	0.886	22.73	0.675
Cascade	33.51	0.893	33.10	0.882	31.48	0.854	30.80	0.841
DDF (Ours)	39.30	0.954	37.25	0.934	37.44	0.938	36.71	0.931

Note: All methods are evaluated on the same 200 test slices at matched view counts.

StepLR reaches a stable plateau within the comparison budget: its validation PSNR changes by only 0.00015 dB over the last ten epochs. Cosine annealing approaches a plateau, whereas ReduceLROnPlateau continues to improve at epoch 100. This sensitivity analysis indicates that validation-driven scheduling can further improve optimization. Nevertheless, StepLR is retained for the main DDF, Cascade, and controlled ablation comparisons because all of these models were trained with the same pre-specified optimization protocol. Retrospectively applying a more favorable scheduler only to DDF would confound the architectural comparison. Consequently, the reported DDF–Cascade and module-ablation differences isolate the feedback-fusion design rather than scheduler-specific tuning; the alternative schedules identify additional optimization headroom but are not used to replace the main reported test results.

Experiment setting: The evaluation benchmarks DDF against conventional analytic Sparse FBP, the image-domain RED-CNN [7], and the dual-domain DuDoTrans [10]. To ensure a consistent evaluation, all methods are tested on the same ordered set of 200 held-out slices at matched view counts. DDF and Cascade are trained on the same 1600-slice training set and selected using the same 200-slice validation set. DuDoTrans operates on projection data and therefore requires view-matched models. Because the available checkpoints did not cover all evaluated view counts, the 180-, 90-, and 45-view models are reproduced using the same patient-wise training and validation subsets, while the original 30-view checkpoint is used directly. RED-CNN is retained with its existing pretrained checkpoint to evaluate its direct transfer performance under the same sparse-view test protocol, rather than a task-specific retrained variant. The test set is not used for training or model selection. Sparse FBP requires no training. The Cascade baseline is introduced to isolate the contribution of the proposed feedback fusion mechanism. Specifically, Cascade employs the same projection-domain (F_p , instantiated as ViewTrans) and image-domain (F_i , instantiated as NAFNet) restoration networks as DDF, but omits the image-to-sinogram feedback, Sine Fusion, and CT Fusion modules. By controlling the backbone capacity and training protocol, this direct ablation attributes the performance difference to the feedback-based architecture. ViewTrans remains the projection backbone of Cascade and DDF, but a standalone ViewTrans accuracy row is not reported because an independently trained checkpoint

matching all four acquisition settings was unavailable. Alternative backbone configurations are examined in a separate ablation study.

All methods are evaluated under four sparse-view factors, i.e., $S = 2, 4, 8, 12$ corresponding to 180, 90, 45, and 30 measured projection views, respectively. These settings cover both mild and severe angular undersampling conditions. Every reported test result is computed at the matched view count on the same ordered set of 200 held-out test slices. Model selection and reproduction training use only the training and validation subsets described above; the held-out test set is not used for fitting or checkpoint selection. Quantitative evaluation of the reconstructed CT image quality is performed using Peak Signal-to-Noise Ratio (PSNR), Structural Similarity Index Measure (SSIM), Root Mean Squared Error (RMSE), and Mean Absolute Error (MAE). All four metrics are computed between the reconstructed CT image and the ground truth. PSNR and SSIM measure signal fidelity and structural similarity, respectively, with higher values indicating better reconstruction quality. RMSE and MAE directly quantify normalized intensity errors and are defined for each slice as

$$\text{RMSE} = \sqrt{\frac{1}{|\Omega|} \|\hat{X} - X\|_2^2}, \quad \text{MAE} = \frac{1}{|\Omega|} \|\hat{X} - X\|_1. \quad (32)$$

Predictions are clipped to $[0, 1]$ before evaluation, and the per-slice values are averaged over the fixed 200-slice test set. Lower RMSE and MAE indicate better reconstruction quality. In addition to quantitative comparison, qualitative visual comparisons are provided to evaluate artifact suppression and structural preservation. Computational cost is also reported in terms of trainable parameters and floating-point operations (FLOPs) to analyze the trade-off between accuracy and complexity among the different methods.

Furthermore, module-level ablation studies are conducted to verify the effectiveness of the key components in DDF. The ablation experiments mainly investigate the contribution of Sine Fusion and CT Fusion, as well as the overall image-to-sinogram feedback path. The module-level variants are evaluated at all four sparse-view factors. The additional $S = 2, 4, 8$ variants use the same optimizer and learning-rate schedule as DDF; training is stopped after the validation PSNR changes by less than 0.01 dB over ten consecutive epochs once the scheduled learning rate has decayed below 1.3×10^{-7} , and the checkpoint with the best validation PSNR is retained. All experiments are implemented within the PyTorch framework and executed on an NVIDIA GeForce RTX 4090 graphics processing unit (GPU).

4.2 Comparison Experiments

4.2.1 Qualitative Evaluation

For clarity, qualitative comparisons are presented separately under three representative sparse-view factors, namely $S = 2, S = 8$, and $S = 12$. For each setting, Ground Truth, Sparse FBP, RED-CNN, Cascade, and DDF are compared. Representative qualitative examples are selected from the same 200-slice test set used for quantitative evaluation. The first row in each figure shows the reconstructed images with a local zoomed region, while the second row shows the corresponding error maps with respect to the ground truth.

Fig. 7 shows the qualitative results under the relatively mild sparse-view setting of $S = 2$. Although Sparse FBP already introduces visible streak artifacts, the main anatomical structures are still recognizable. RED-CNN reduces part of these artifacts, while Cascade and DDF further improve visual quality. In this setting, the differences among the advanced reconstruction methods are not very large, but DDF still exhibits cleaner local details and slightly lower residual error in the enlarged region.

Fig. 8 presents the comparison under $S = 8$, where the reconstruction problem becomes more challenging. Sparse FBP suffers from much stronger artifact contamination. The learning-based baselines alleviate

these degradations to different extents, but residual blur is still visible, especially around boundaries and fine structures. Compared with RED-CNN and Cascade, DDF produces sharper local structures and a more faithful reconstruction, which is further supported by the reduced intensity of the error map.

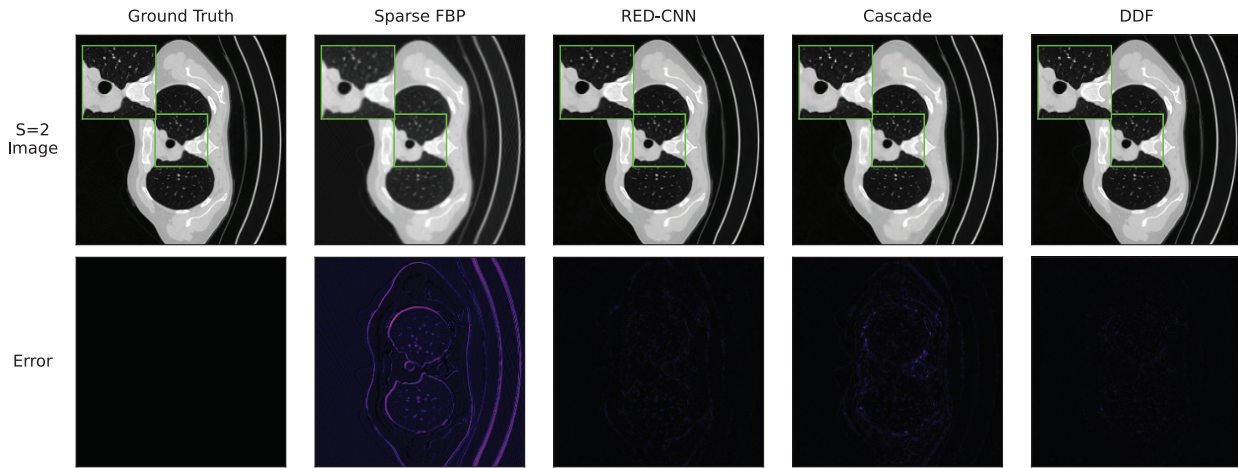


Figure 7: Qualitative comparison under sparse-view factor $S = 2$.

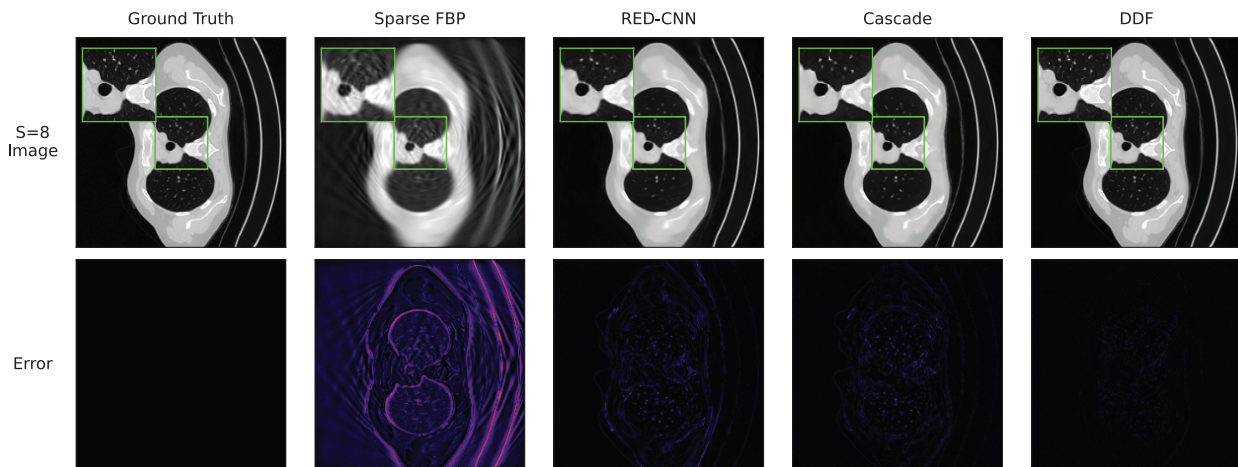


Figure 8: Qualitative comparison under sparse-view factor $S = 8$.

[Fig. 9](#) shows the results under the highest sparse condition of $S = 12$. In this case, Sparse FBP is heavily degraded, with severe streak artifacts. Although the other baselines improve the reconstruction quality, noticeable residual artifacts and structural smoothing remain. Among all compared methods, DDF achieves the best visual consistency with the ground truth, particularly in preserving boundary sharpness and recovering local anatomical details. Its error map is also more concentrated and less pronounced than those of the competing methods.

Comparing [Figs. 7–9](#) together, a consistent trend can be observed: as the sparse-view factor increases, all methods experience gradual performance degradation, but that of DDF is less severe. This indicates that the proposed dual-domain feedback and fusion strategy improves robustness under increasingly undersampled conditions.

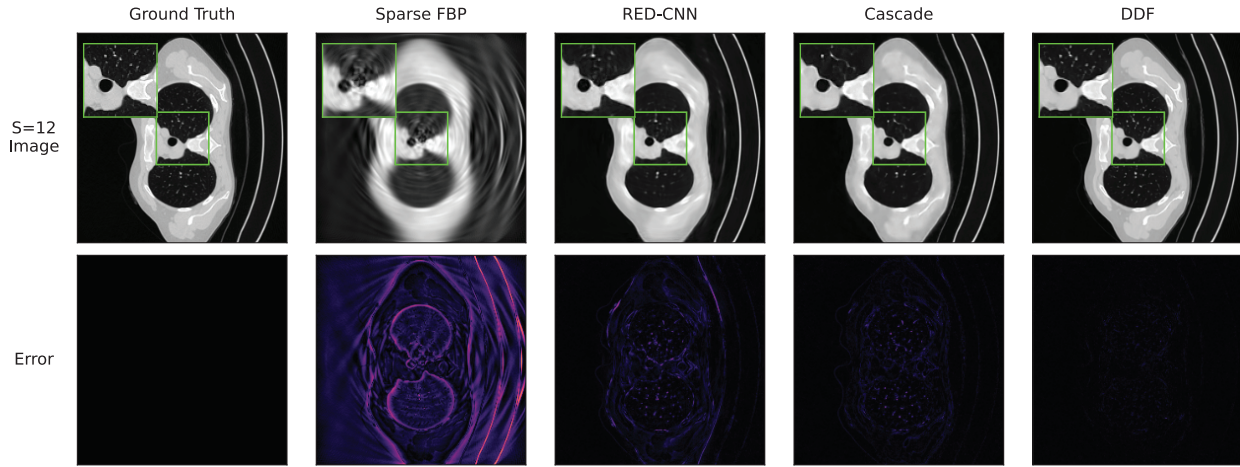


Figure 9: Qualitative comparison under sparse-view factor $S = 12$.

4.2.2 Quantitative Evaluation

Table 2 reports the results on the fixed 200-slice test set. Sparse FBP decreases from 24.85 dB and 0.793 at $S = 2$ to 19.97 dB and 0.567 at $S = 12$. RED-CNN provides only a small improvement over Sparse FBP under this sparse-view protocol, whereas the matched Cascade baseline is substantially stronger. DuDoTrans achieves 37.55/0.933, 35.69/0.919, 32.41/0.886, and 22.73/0.675 at $S = 2$, $S = 4$, $S = 8$, and $S = 12$, respectively. DDF achieves the highest PSNR and SSIM in every setting.

The efficacy of the feedback fusion strategy is evaluated by comparing DDF with the Cascade baseline. While Cascade substantially improves upon the image-domain and analytical baselines, DDF yields PSNR improvements of 5.79, 4.15, 5.96, and 5.91 dB for $S = 2, 4, 8$, and 12 , respectively. This consistent advantage indicates that reconnecting image-domain predictions with projection-domain representations through the proposed FP feedback path, Sine Fusion, and CT Fusion recovers structural information beyond the standard dual-domain pipeline.

Table 3 reports the complementary RMSE and MAE results under the same evaluation protocol. DDF achieves the lowest error for every sparse-view factor. Its RMSE remains between 0.0106 and 0.0142, whereas the matched Cascade ranges from 0.0204 to 0.0278. At $S = 12$, DDF reduces RMSE from 0.0278 to 0.0142 and MAE from 0.0176 to 0.0103 relative to Cascade. The consistent reductions agree with the PSNR and SSIM findings and show that the feedback pathway improves both structural quality and normalized intensity accuracy.

Table 3: RMSE/MAE comparison on the fixed 200-slice test set. Lower values are better. Bold values indicate the best performance among all compared methods.

Method	$S = 2$	$S = 4$	$S = 8$	$S = 12$
Sparse FBP	0.0553/0.0280	0.0656/0.0350	0.0828/0.0473	0.0972/0.0577
RED-CNN	0.0546/0.0278	0.0647/0.0343	0.0823/0.0468	0.0968/0.0575
DuDoTrans	0.0133/0.0094	0.0165/0.0109	0.0241/0.0150	0.0732/0.0478
Cascade	0.0204/0.0139	0.0214/0.0144	0.0257/0.0166	0.0278/0.0176
DDF (Ours)	0.0106/0.0081	0.0134/0.0098	0.0131/0.0096	0.0142/0.0103

4.2.3 Noise Robustness

Robustness to measurement noise is evaluated at the most severe sparse-view setting, $S = 12$. Because the stored numerical projections are not expressed as calibrated physical attenuation integrals, they are first linearly scaled so that the maximum line integral over the fixed test set equals 5.0. For a scaled noiseless line integral p , photon counts are sampled according to the Beer–Lambert model

$$y \sim \text{Poisson}(I_0 \exp(-p)), \quad \tilde{p} = -\log\left(\frac{\max(y, 1)}{I_0}\right), \quad (33)$$

after which $\tilde{p}$ is mapped back to the original numerical projection scale. Noise is added to the full-view projection before the original uniform angular downsampling and linear interpolation. Two incident photon levels, $I_0 = 10^5$ and $I_0 = 10^4$, are tested using fixed random seeds on the same 200 slices.

Table 4 shows the expected progressive degradation as the photon count decreases. Nevertheless, DDF remains the best-performing method for every metric and noise level. At $I_0 = 10^5$, DDF exceeds Cascade by 4.99 dB in PSNR and 0.084 in SSIM; at $I_0 = 10^4$, the corresponding margins remain 2.74 dB and 0.065. DDF also reduces RMSE from 0.0391 to 0.0286 and MAE from 0.0228 to 0.0174 relative to Cascade at the lower photon level. These results demonstrate robustness to an unseen measurement-noise distribution.

Table 4: Noise robustness at $S = 12$ on the fixed 200-slice test set. Bold values indicate the best performance among all compared methods.

Condition	Method	PSNR	SSIM	RMSE	MAE
Noise-free	Sparse FBP	19.97	0.567	0.0972	0.0577
	Cascade	30.80	0.841	0.0278	0.0176
	DDF (Ours)	36.71	0.931	0.0142	0.0103
$I_0 = 10^5$	Sparse FBP	19.87	0.530	0.0982	0.0592
	Cascade	30.39	0.837	0.0291	0.0182
	DDF (Ours)	35.38	0.921	0.0165	0.0115
$I_0 = 10^4$	Sparse FBP	19.16	0.381	0.1061	0.0687
	Cascade	27.84	0.794	0.0391	0.0228
	DDF (Ours)	30.59	0.859	0.0286	0.0174

To assess whether the improvements are consistent across test images, two-sided paired Wilcoxon signed-rank tests were applied to the per-slice PSNR and SSIM values. Holm correction was performed separately within each sparse-factor and metric family, and 95% confidence intervals (CIs) for the mean paired differences were estimated using 10,000 paired bootstrap resamples. All 32 DDF-to-baseline comparisons remained significant after correction ($p_{\text{Holm}} < 2 \times 10^{-24}$). Relative to DuDoTrans, DDF improves mean PSNR/SSIM by 1.74 dB/0.0213, 1.56 dB/0.0153, 5.04 dB/0.0528, and 13.98 dB/0.2559 at $S = 2$, $S = 4$, $S = 8$, and $S = 12$, respectively. At $S = 2$, the corresponding 95% CIs are [1.56, 1.92] dB and [0.0195, 0.0230], with Holm-adjusted p -values below 1.50×10^{-28} ; DDF is better on 177 and 186 of the 200 slices for PSNR and SSIM, respectively. Across $S = 4$, $S = 8$, and $S = 12$, the previously reported CIs and significance conclusions remain unchanged. Table 5 summarizes the central controlled comparison against Cascade. DDF outperformed Cascade on every one of the 200 test slices in all four settings. Because the test slices belong to one held-out patient, these tests quantify slice-level consistency within that subject rather than population-level variation across patients.

Table 5: Paired slice-level comparison of DDF against Cascade on the 200-slice test set. Values are mean paired improvements with 95% bootstrap CIs.

Setting	Δ PSNR (dB)	Δ SSIM	Holm-Adjusted p
$S = 2$	5.79 [5.64, 5.93]	0.0615 [0.0590, 0.0640]	$<10^{-33}$
$S = 4$	4.15 [4.01, 4.29]	0.0518 [0.0503, 0.0534]	$<10^{-33}$
$S = 8$	5.96 [5.80, 6.12]	0.0839 [0.0818, 0.0858]	$<10^{-33}$
$S = 12$	5.91 [5.77, 6.05]	0.0895 [0.0876, 0.0913]	$<10^{-33}$

4.2.4 Computational Cost

Table 6 reports both theoretical complexity and measured inference efficiency. Runtime and peak allocated graphics processing unit (GPU) memory were measured at the most challenging setting, $S = 12$, on an NVIDIA GeForce RTX 4090 with a batch size of one. Each method was executed for 20 warm-up iterations followed by 100 timed iterations. Compute Unified Device Architecture (CUDA) synchronization was applied before and after every timed inference, and data loading and host-to-device transfer were excluded. The table reports the mean and standard deviation of the resulting end-to-end latency measurements. Sparse FBP is analytical and therefore has no trainable network parameters.

Table 6: Computational complexity and measured inference efficiency at $S = 12$. Latency is reported as mean $\pm$ standard deviation over 100 runs.

Method	Param. (M)	FLOPs (G)	Latency (ms)	Peak Memory (MB)
Sparse FBP	0	–	48.07 ± 0.17	2881.98
Cascade	24.26	36.57	56.70 ± 0.28	2984.71
RED-CNN	1.85	220.36	53.24 ± 0.02	2897.16
DuDoTrans	0.62	109.86	505.06 ± 66.92	769.18
ViewTrans	7.15	2.21	49.49 ± 0.06	2920.05
DDF (Ours)	28.10	39.52	154.67 ± 2.11	3903.21

The compared methods achieve different levels of reconstruction accuracy while requiring different computational costs. RED-CNN has 1.85M parameters but requires 220.36 G FLOPs, indicating that a smaller parameter count does not necessarily correspond to lower computational cost. DuDoTrans and ViewTrans introduce Transformer-based or view-aware modeling, with 0.62M and 7.15M parameters, respectively. The ViewTrans entry in Table 6 is an architecture-only timing of the projection branch followed by FBP; timing does not depend on the availability of a standalone accuracy checkpoint.

Compared with the Cascade baseline, DDF introduces a moderate increase in computational cost. The number of trainable parameters increases from 24.26 to 28.10M, and the FLOPs increase from 36.57 to 39.52 G. This additional cost mainly comes from the image-to-sinogram feedback path, Sine Fusion, and CT Fusion. Unlike simply enlarging the backbone network, these additional components are designed to build explicit interaction between the projection domain and the image domain.

The measured mean latency of DDF is 154.67 ms per slice, compared with 56.70 ms for Cascade. This increase is expected because DDF performs an additional forward projection and FBP operation and executes the two feedback-fusion modules. Nevertheless, DDF remains substantially faster than DuDoTrans (505.06 ms), while its peak allocated GPU memory is 3903.21 MB. Considering the consistent PSNR and SSIM improvements reported in Table 2, these measurements quantify the accuracy–efficiency trade-off of

the feedback pathway rather than attributing the performance gain solely to model size. Historical training runs were not instrumented under a common timing protocol, so retrospective training-time values are not reported. In addition, DDF is a two-dimensional model and the reported latency is per slice; no volume-level or three-dimensional reconstruction latency is claimed.

4.3 Ablation Studies

Two groups of ablation experiments separately investigate the contributions of the proposed feedback-fusion architecture and the selected domain-specific backbones. The first group evaluates the image-to-sinogram feedback pathway, Sine Fusion, and CT Fusion across all four sparse-view factors. The second group evaluates alternative projection-domain and image-domain backbone configurations.

4.3.1 Module-Level Ablation

The first analysis evaluates the contributions of the feedback pathway and the two proposed fusion modules while keeping ViewTrans and NAFNet as the projection-domain and image-domain backbones, respectively.

Because Sine Fusion and CT Fusion are responsible for combining two intermediate representations, directly deleting either module would disconnect the reconstruction pipeline. Each proposed module is therefore replaced with a parameter-free operation that preserves the two inputs while removing learned adaptive fusion. For the variant without the proposed Sine Fusion, the restored sinogram Z_p and feedback sinogram Z_i are combined by element-wise summation. For the variant without the proposed CT Fusion, the image-domain prediction X_i and feedback-enhanced reconstruction X_f are combined by their arithmetic mean instead of the Cross Gating Block.

The Cascade baseline is used to evaluate the complete FP-driven feedback pathway. It retains the same ViewTrans and NAFNet backbones but removes the image-to-sinogram feedback branch, including FP feedback, Sine Fusion, the second FBP operation, and CT Fusion. Its output is therefore given directly by

$$X_i = F_i \left(FBP \left(F_p(Z_s) \right) \right). \quad (34)$$

Because the feedback sinogram $Z_i = P(X_i)$ and the subsequent fusion stages depend on the FP operation, removing FP necessarily disables the complete downstream feedback branch. Thus, the comparison between Cascade and full DDF evaluates the contribution of the complete feedback-fusion pathway.

As shown in Table 7, both parameter-free replacements consistently underperform full DDF. Replacing Sine Fusion with element-wise summation reduces PSNR by 9.24, 8.01, 8.97, and 16.38 dB at $S = 2, 4, 8$, and 12, respectively. Although this variant still receives both Z_p and Z_i , a fixed sum cannot adapt their relative contributions or reconcile domain-specific discrepancies. This consistent degradation supports the learned gated Sine Fusion mechanism.

Replacing the Cross Gating Block with arithmetic averaging reduces PSNR by 6.41, 5.97, 7.76, and 11.51 dB across the same four settings. The corresponding SSIM reductions are 0.068, 0.083, 0.125, and 0.198. Thus, simply averaging X_i and X_f cannot fully exploit their complementary structural information, particularly as angular undersampling becomes more severe.

The Cascade comparison further isolates the complete FP-driven feedback-fusion pathway because Cascade and DDF use the same ViewTrans and NAFNet backbones. DDF improves upon Cascade by 5.79, 4.15, 5.96, and 5.91 dB at $S = 2, 4, 8$, and 12, respectively. Together, the multi-sparsity results demonstrate that the feedback path and both learned fusion mechanisms contribute consistently rather than only under the 30-view setting.

Table 7: Module-level ablation study under different sparse-view factors. Each entry reports PSNR (dB)/SSIM on the fixed 200-slice test set. “Sum” and “Avg.” are parameter-free replacements for Sine Fusion and CT Fusion, respectively. Bold values indicate the best performance among all compared methods.

Variant	$S = 2$	$S = 4$	$S = 8$	$S = 12$
w/o Sine Fusion (Sum)	30.06/0.836	29.24/0.815	28.47/0.795	20.33/0.535
w/o CT Fusion (Avg.)	32.89/0.886	31.28/0.851	29.68/0.813	25.20/0.733
Cascade	33.51/0.893	33.10/0.882	31.48/0.854	30.80/0.841
Full DDF (Ours)	39.30/0.954	37.25/0.934	37.44/0.938	36.71/0.931

4.3.2 Backbone Ablation

The influence of the projection-domain and image-domain backbones is further investigated. In this group, the complete DDF pipeline, including FP feedback, Sine Fusion, the second FBP operation, and CT Fusion, is kept unchanged. Only the projection-domain network F_p , the image-domain network F_i , or both are replaced. Table 8 reports the results across all four sparse-view factors. The $S = 2$, $S = 4$, and $S = 8$ variants are selected using the separate validation subset and evaluated once on the same fixed 200-slice test set; the matched $S = 12$ results are retained from the original backbone study.

Table 8: Backbone-substitution ablation under different sparse-view factors. Each entry is PSNR (dB)/SSIM. Bold values indicate the best performance among all compared methods.

Variant	$S = 2$	$S = 4$	$S = 8$	$S = 12$
P-CNN	38.95/0.951	37.23/0.933	36.36/0.926	35.40/0.910
P-Swin	38.31/0.948	36.46/0.929	35.67/0.920	31.87/0.865
I-CNN	36.19/0.922	31.05/0.851	31.29/0.852	29.02/0.791
I-Restor	36.01/0.922	33.28/0.886	34.35/0.896	32.46/0.864
Mixed	38.06/0.943	36.05/0.921	32.92/0.878	26.87/0.729
DDF (Ours)	39.30/0.954	37.25/0.934	37.44/0.938	36.71/0.931

Note: P-CNN and P-Swin replace ViewTrans with ResUNet and SwinIR, respectively; I-CNN and I-Restor replace NAFNet with RED-CNN and Restormer, respectively; Mixed uses SwinIR and RED-CNN.

Across the four settings, full DDF provides the highest PSNR and SSIM. P-CNN is the strongest substitution, approaching DDF at $S = 2$ and $S = 4$ but showing a larger deficit as angular undersampling becomes more severe. P-Swin follows a similar pattern, while the image-domain substitutions generally incur larger losses. The Mixed configuration remains competitive at $S = 2$ and $S = 4$ but degrades markedly at $S = 8$ and $S = 12$, indicating that simultaneously replacing both domain-specific networks compounds representation and cross-domain compatibility limitations.

The substituted backbones do not exhibit perfectly monotonic changes between every adjacent sparse factor. In particular, I-CNN and I-Restor produce slightly higher test scores at $S = 8$ than at $S = 4$. Each entry was independently selected on the validation subset and evaluated using the checkpoint and acquisition geometry corresponding to that sparse factor. These small reversals therefore reflect configuration-specific optimization and test-set variation rather than a change in the definition of the sparse factors. The overall trend from $S = 2$ to $S = 12$ remains downward for every variant.

These results support both the effectiveness of the feedback-fusion topology and the importance of selecting suitable domain-specific backbones. DDF consequently provides a replaceable operator-level interface, but it does not imply performance independence from backbone capacity or domain compatibility.

5 Discussion

DDF is designed to change the interaction pattern between the projection and image domains, rather than simply increasing network depth. In a conventional cascade, the restored sinogram is transformed into an image and the final reconstruction is then dominated by the image-domain network. This one-way flow may weaken measurement-related information after FBP. By contrast, DDF forward-projects the image-domain prediction back into the sinogram domain, where structural information learned from the image can be fused with projection-domain features before the final reconstruction. This feedback path makes the two domains complementary in a more explicit and physically meaningful way.

The experimental results and ablation studies support this design. Compared with sparse-view FBP and the cascaded baseline, DDF consistently improves PSNR and SSIM, especially under severe sparsity. Because the Cascade baseline uses the same ViewTrans and NAFNet backbones but removes FP feedback, Sine Fusion, and CT Fusion, the performance gain mainly comes from the proposed feedback fusion mechanism. The performance degradation observed when replacing the proposed Sine Fusion and CT Fusion modules with parameter-free summation and averaging further demonstrates the importance of adaptive fusion in both domains.

Several limitations remain. The controlled experiments on the public 2016 NIH–AAPM–Mayo Clinic benchmark demonstrate the effectiveness of DDF under the simulated fan-beam setting considered in this study. Evaluation on external patient cohorts and real raw projection data would further assess its clinical generalization and is left for future work. The 256×256 resolution provides a common and computationally feasible setting for DDF, Cascade, and the ablation variants, so it does not confound the controlled architectural comparisons. Increasing the resolution to 512×512 would quadruple the number of image elements, require a new projector, and substantially increase memory and training cost. Native-resolution reconstruction therefore remains future work for assessing the preservation of very fine structures in clinical applications. Although the projection-domain Poisson experiments demonstrate robustness across two unseen photon-count levels, electronic noise, alternative CT geometries, and unseen nonuniform or irregular sparse-view patterns have not been evaluated. DDF also introduces additional FP/FBP operations and fusion modules. Extending the fixed two-dimensional sparse operators to higher-resolution images or three-dimensional volumes would substantially increase operator storage and computation; lightweight fusion blocks and more efficient projectors are therefore important directions for practical deployment. Beyond sparse-view reconstruction, the same feedback idea may be useful for low-dose noise, metal artifacts, limited-angle artifacts, and other cases where projection-domain measurements and image-domain structures are inconsistent.

6 Conclusion

This paper proposes DDF, a feedback-based dual-domain framework for sparse-view CT reconstruction. Unlike conventional cascaded pipelines, DDF projects the image-domain prediction back to the sinogram domain and uses Sine Fusion and CT Fusion to adaptively combine projection-domain and image-domain information. This closed-loop design strengthens cross-domain interaction before the final reconstruction.

Experiments on the 2016 NIH–AAPM–Mayo Clinic Low Dose CT Grand Challenge dataset show that DDF outperforms sparse-view FBP and representative learning-based baselines under $S = 2, 4, 8$, and 12. Qualitative comparisons and ablation studies further verify the effectiveness of the feedback path and the two fusion modules. While feedback fusion improves agreement between the two domains, the present framework does not explicitly impose a projection-domain consistency constraint. Future work will investigate such constraints to further improve reconstruction performance, together with validation on

real raw projection data, higher-resolution and 3D reconstruction, and more efficient implementations for practical use.

Acknowledgement: We would like to express our gratitude to Hao Feng for his support and assistance in data organization, data preprocessing, experiments, and inspections.

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

Author Contributions: Ziqiao Pan: Investigation, Software, Validation, Formal Analysis, Visualization, and Writing—Original Draft. Yuxi Chen: Writing—Review and Editing. Bo Yang: Methodology, Supervision, and Writing—Review and Editing. Legend Zhang, Renee Zeng, Xiaoran Ma, Zhixin Liu, Wei Wei, and Guangyu Xu: Writing—Review and Editing. All authors reviewed and approved the final version of the manuscript.

Availability of Data and Materials: The data used in this study are derived from the public 2016 NIH–AAPM–Mayo Clinic Low Dose CT Grand Challenge dataset. Dataset information and access instructions are available from the official AAPM Grand Challenge page: <https://www.aapm.org/GrandChallenge/LowDoseCT/>.

Ethics Approval: Not applicable. This study used only publicly available, fully anonymized data from the 2016 NIH–AAPM–Mayo Clinic Low Dose CT Grand Challenge and involved no new participant recruitment or collection of identifiable information. The original patient data were collected with approval from the institutional review board of the host institution and were anonymized before public release.

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

Nomenclature

Notation	Description
Z_s	Sparse-view input sinogram
Z_p	Restored sinogram produced by the projection-domain model
Z_i	Feedback sinogram obtained by forward projecting the image-domain prediction
Z_f	Fused sinogram after Sine Fusion
X_p	Intermediate CT image reconstructed from Z_p
X_i	Image-domain prediction produced by the image-domain model
X_f	Feedback-enhanced CT image reconstructed from Z_f
X	Ground-truth CT image
$\hat{X}$	Final reconstructed CT image

References

1. Brenner DJ, Hall EJ. Computed tomography: an increasing source of radiation exposure. *N Engl J Med*. 2007;357(22):2277–84. doi:10.1056/nejmra072149.
2. Ramachandran GN, Lakshminarayanan AV. Three-dimensional reconstruction from radiographs and electron micrographs: application of convolutions instead of Fourier transforms. *Proc Natl Acad Sci U S A*. 1971; 68(9):2236–40. doi:10.1073/pnas.68.9.2236.
3. Shepp LA, Logan BF. The Fourier reconstruction of a head section. *IEEE Trans Nucl Sci*. 1974;21(3):21–43.
4. Cheng S, Chen Y, Yang B, Liu C, Zhang L, Yin L, et al. Deep learning for sparse-view CT reconstruction: a survey. *SSRN J*. 2025;1(1):311. doi:10.2139/ssrn.5366340.
5. 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.

6. Yuan H, Jia J, Zhu Z. SIPID: a deep learning framework for sinogram interpolation and image denoising in low-dose CT reconstruction. In: Proceedings of the IEEE International Symposium on Biomedical Imaging; 2018 Apr 4–7; Washington, DC, USA. p. 1521–4.
7. Chen H, Zhang Y, Kalra MK, Lin F, Chen Y, Liao P, et al. Low-dose CT with a residual encoder-decoder convolutional neural network. *IEEE Trans Med Imaging*. 2017;36(12):2524–35. doi:10.1109/tmi.2017.2715284.
8. 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.
9. 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.
10. Wang C, Shang K, Zhang H, Li Q, Hui Y, Zhou SK. DuDoTrans: dual-domain transformer for sparse-view CT reconstruction. In: Proceedings of the Medical Image Computing and Computer Assisted Intervention, MICCAI 2022; 2022 Sep 18–22; Singapore. Cham, Switzerland: Springer; 2022. p. 84–94.
11. Ge R, He Y, Xia C, Sun H, Zhang Y, Hu D, et al. DDPNet: a novel dual-domain parallel network for low-dose CT reconstruction. In: Proceedings of the Medical Image Computing and Computer Assisted Intervention, MICCAI 2022; 2022 Sep 18–22; Singapore. Cham, Switzerland: Springer; 2022. p. 748–57.
12. Ye X, Sun Z, Xu R, Wang Z, Li H. Low-dose CT reconstruction via dual-domain learning and controllable modulation. In: Proceedings of the Medical Image Computing and Computer Assisted Intervention, MICCAI 2022; 2022 Sep 18–22; Singapore. Cham, Switzerland: Springer; 2022. p. 549–59.
13. 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(1):106113. doi:10.1016/j.dsp.2026.106113.
14. Chen L, Chu X, Zhang X, Sun J. Simple baselines for image restoration. In: Proceedings of the European Conference on Computer Vision; 2022 Oct 23–27; Tel Aviv, Israel. Cham, Switzerland: Springer; 2022. p. 17–33.
15. Sidky EY, Pan X. Image reconstruction in circular cone-beam computed tomography by constrained, total-variation minimization. *Phys Med Biol*. 2008;53(17):4777–807. doi:10.1088/0031-9155/53/17/021.
16. Chen GH, Tang J, Leng S. Prior image constrained compressed sensing (PICCS): a method to accurately reconstruct dynamic CT images from highly undersampled projection data sets. *Med Phys*. 2008;35(2):660–3. doi:10.1118/1.2836423.
17. Mahmood T, Rehman A, Saba T, Nadeem L, Bahaj SAO. Recent advancements and future prospects in active deep learning for medical image segmentation and classification. *IEEE Access*. 2023;11:113623–52. doi:10.1109/access.2023.3313977.
18. 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.
19. Li Y, Sun X, Wang S, Li X, Qin Y, Pan J. MDST: multi-domain sparse-view CT reconstruction based on convolution and swin transformer. *Phys Med Biol*. 2023;68(9):095019. doi:10.1088/1361-6560/acc2ab.
20. 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. doi:10.1016/j.cmpb.2024.108575.
21. 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. doi:10.1109/tci.2024.3507645.
22. Ding C, Zhang Q, Wang G, Ye X, Chen Y. LAMA: stable dual-domain deep reconstruction for sparse-view CT. *arXiv:2410.21111*. 2024.
23. 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.
24. 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.
25. Buzug TM. Computed tomography: from photon statistics to modern cone-beam CT. Berlin/Heidelberg, Germany: Springer; 2008.
26. Tu Z, Talebi H, Zhang H, Yang F, Milanfar P, Bovik AC, et al. MAXIM: multi-axis MLP for image processing. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition; 2022 Jun 19–24; New Orleans, LA, USA. p. 5769–80