Journal of Nondestructive Evaluation (2025) 44:101
https://doi.org/10.1007/s10921-025-01244-3

SwinIR-based Dual-Domain Reconstruction for Sparse-View Computed Tomography

Jonas Van der Rauwelaert[1] · Caroline Bossuyt[2] · Stijn E. Verleden[1] · Jan Sijbers[2]

Received: 15 May 2025 / Accepted: 12 July 2025 / Published online: 18 August 2025

Abstract

Sparse-view computed tomography (CT) remains a significant challenge due to undersampling artifacts and loss of structural detail in the reconstructed images. In this work, we introduce DDSwinIR, a dual-domain reconstruction framework that leverages Swin Transformer-based architectures to recover high-quality CT images from severely undersampled sinograms. DDSwinIR operates in three stages: sinogram upsampling, deep learning-based reconstruction, and a residual refinement module that addresses domain-specific inconsistencies. While previous dual-domain deep learning (DD-DL) approaches improve reconstruction quality, they lack a systematic analysis of component contributions and do not generalize to unseen number of projections. DDSwinIR addresses these gaps through a modular and transparent design, allowing quantification of each network's module. Our results highlight that early application of data consistency, especially after initial sinogram reconstruction, yields the most substantial and reliable improvements, particularly under extreme sparsity. We also introduce sparse-view concatenation, which enhances performance by improving feature propagation in highly undersampled settings. Extensive evaluation across varying numbers of projections reveal strong generalization when trained on sparse data and tested on denser configurations, but not vice versa, underscoring the importance of low-sparsity training. Compared to conventional reconstruction methods, DDSwinIR achieves superior artifact suppression and detail preservation. This work establishes DDSwinIR as an interpretable and generalizable solution for sparse-view CT, responding to the need for DD-DL reconstruction frameworks for practical applicability.

Keywords Sparse-view CT · Swin transformer · Dual-domain · Residual refinement

1 Introduction

Computed Tomography (CT) is a widely used, non-invasive medical and industrial imaging technique, which allows to generate cross-sectional or volumetric images of internal structures by acquiring X-ray projections from multiple angles. However, since X-rays involve ionizing radiation, prolonged or excessive exposure can pose health risks [1, 2]. To mitigate these risks, sparse-view CT can be employed to reduce the number of projection angles, thereby minimizing radiation dose and acquisition time. However, reducing the number of projections may lead to image artifacts, such as streaking and blurring, due to insufficient sampling. Additionally, it may degrade the quality of reconstructed images, making it more challenging to accurately visualize fine details or detect anomalies [2].

Various processing methods have been proposed to mitigate undersampling artefacts in sparse-view CT, either operating in the projection or image domain. Projection-domain pre-processing techniques generally estimate missing sinogram data through interpolation or inpainting [1, 3, 4]. While sinogram-based methods have been shown to reduce sparse-view artefacts by efficiently exploit projec-

Jonas Van der Rauwelaert and Caroline Bossuyt contributed equally to this work

✉ Jonas Van der Rauwelaert
Jonas.VanderRauwelaert@UAntwerpen.be

Caroline Bossuyt
Caroline.Bossuyt@UAntwerpen.be

Stijn E. Verleden
Stijn.Verleden@UAntwerpen.be

Jan Sijbers
Jan.Sijbers@UAntwerpen.be

[1] ASTARC, University of Antwerp, Universiteitsplein 1, Wilrijk 2610, Belgium

[2] imec-Vision Lab, Department of Physics, University of Antwerp, Universiteitsplein 1, Wilrijk 2610, Belgium

tion domain relations (e.g., Helgason-Ludwig consistency criteria [5]), they do not leverage prior knowledge from the image domain (e.g., spatial correlations). Purely image domain post-processing techniques, on the other hand, have also shown to reduce noise and suppress artifacts, but the resulting images are not data consistent in that the restored image not necessarily adheres to the measured projection data [6, 7]. Iterative reconstruction algorithms enhance image quality by repeatedly refining estimates based on the available data, preserving structural details more effectively [8, 9]. However, these methods involve repeated forward and backward projections, leading to high computational costs [10, 11].

To address the limitations of the aforementioned reconstruction approaches, Deep Learning (DL) approaches have merged as powerful alternatives, leveraging data-driven priors to improve reconstruction quality in data-limited scenarios such as sparse-view CT. While most DL approaches rely on supervised learning with large datasets, reconstruction methods based on untrained neural networks, such as Deep Image Prior (DIP) [12–14], offer a fundamentally different approach by using the structure of the network itself as an implicit prior. Rather than relying on training data, DIP learns directly from the observed data, making it particularly well-suited for scenarios where data is sparse or not available [15]. Extensions such as SDIP [16] and RBP-DIP [17] build on this approach and achieve improved reconstruction performance under increasingly ill-posed conditions.

Early DL models for sparse-view CT primarily relied on Convolutional Neural Networks (CNNs), a subset of Artificial Neural Networks (ANNs). While CNNs and their variants, such as U-Net-based architectures [18–21], excel at extracting local features, they fall short when it comes to modeling long-range dependencies [22]. This limitation is especially relevant in sparse-view settings where global context is crucial. In contrast, the Transformer model [23] uses self attention to capture complex long-range relations between different elements of the input. Although originally designed for natural language processing, transformers have been successfully adapted for vision tasks [22]. Notable variants like the Swin Transformer [24] improve efficiency through a shifted window approach. By applying self-attention within local windows and subsequently shifting these windows, the model reduces computational cost of global self-attention while capturing cross-window interactions.

Many DL approaches mirror traditional frameworks by operating either in the sinogram domain (by predicting missing projections) or the image domain (by suppressing post-reconstruction artifacts) [25–29]. Although architectural innovations improve modeling capabilities, methods restricted to a single domain tend to inherit the limitations of their respective domain. This motivates interest in Dual-Domain (DD) methods that take advantage of the complementary strengths of both domains. By jointly processing information from both domains, these models often outperform their single-domain counterparts [30]. Effectively coordinating the two representations, however, remains a significant challenge - particularly when aiming to preserve data consistency, i.e., ensuring that the reconstructed image accurately corresponds to its measured projection data. One common strategy to address this is to incorporate the sparse projection data as an auxiliary input, assisting the network in aligning the output with the measured data [31–33]. However, when projection data is severely limited, the effectiveness of this strategy can be low. As an alternative, some DD methods incorporate residual correction modules [34–36] to enhance performance. By minimizing the residuals between intermediate predictions and ground truth data in both domains, these modules aim to correct fine details and improve the data consistency.

While previous DD-DL methods have demonstrated improvements in overall reconstruction quality, they often lack a systematic evaluation of how each component of the method contributes to the overall performance. Moreover, these methods tend to neglect the quality assessment of the restored sinogram, a critical factor in ensuring that the reconstructed images truthfully represent the underlying projection data. Prior work [37] has suggested that omitting the image-domain residual refinement module, specifically the associated neural network, can streamline training and reduce computational costs, with minimal impact on the resulting reconstruction quality in terms of PSNR and SSIM. This highlights the need for a more modular and transparant analysis of DD-DL frameworks to better understand and justify the function of each component. Moreover, these methods are often trained for a fixed number of projections and may struggle to generalize to unseen numbers of projections, limiting their practical applicability.

In this work, we propose a SwinIR-based Dual Domain reconstruction framework, dubbed DDSwinIR, an enhanced DD-DL framework tailored to real-world CT systems. To this end, we extend the architecture proposed in [37] in three aspects: 1) data consistency is enforced to ensure the final solution adheres to the measured data, by minimizing the difference between the forward projection of the solution and the measured data, thereby reducing the risk of hallucinated features that can arise from the reliance on learned priors, 2) the geometry is extended to support fan-beam acquisition, and 3) to further improve reconstruction quality, sparse-view image concatenations, which inject additional image-domain information into the network, enriching context and facilitating better feature propagation. To address the aforementioned gap, our focus is on quantitatively analyzing the contribution of individual components within a modular dual-domain framework, using SwinIR as a proven

and robust network. To validate the impact of each architectural component, we conduct a systematic evaluation of both intermediate image reconstructions and their corresponding sinograms. This modular analysis provides deeper insight into the individual contributions of each module within the DDSwinIR framework. Furthermore, we evaluate the framework's ability to generalize to numbers of projections that differ from the training configuration, highlighting its adaptability.

2 Method

The proposed reconstruction framework, referred to as dual-domain Shifted Window Image Restoration network (DDSwinIR), aims to enhance image quality of reconstructions obtained from sparse CT data by leveraging a three-stage framework: sinogram upsampling (green), initial reconstruction (blue), and residual refinement (orange), as illustrated in Fig. 1.

The initial reconstruction and residual refinement steps utilize SwinIR networks [38], denoted as $\Omega_1, \Omega_2, \Omega_3$, and Ω_4. SwinIR is a DL model that was designed for image restoration in various vision tasks [38]. As illustrated in Fig. 2, it comprises three key components: shallow feature extraction, deep feature extraction, and image reconstruction. First, a convolutional layer extracts low-level (shallow) features, generating an initial feature map with enhanced abstraction. Next, deep feature extraction, focusing on high frequency information, is performed using a series of Residual Swin Transformer Blocks (RSTBs) alongside an additional single convolutional layer. The RSTB block contains multiple Swin Transformer Layers (STLs), with residual connections, which are designed to capture relationships across different regions of the input. Each STL consists of two residual connections, each containing a LayerNorm [39] and either Multi-head Self Attention (MSA) or a Multi-Layer Perceptron (MLP). The MSA module includes multiple attention heads, each independently computing self-attention over the input feature map. This allows the model to capture both local and global dependencies. By operating in parallel, the multiple heads enable the network to learn diverse relational patterns, thereby enhancing its representational power [23]. The MLP is used to further refine the feature representations after the attention mechanism by applying pointwise transformations, allowing the model to learn more complexity. Consecutive STLs alternate between Window MSA (W-MSA) and Shifted Window MSA (SW-MSA). The Window MSA efficiently captures short-range dependencies by dividing the feature map into non-overlapping (local) windows, where local self-attention is applied within each window. The Shifted window MSA shifts the windows before performing MSA within the windows, allowing information to flow across regions and facilitating effective long-range feature learning while maintaining computational efficiency.

A residual connection merges the shallow and deep features, improving gradient flow and feature integration. The combined features are then passed through a final convolutional layer, which, together with the original input via an additional residual connection, is mapped to the final reconstructed image.

2.1 Upsampling

A key challenge in NN-based sinogram upsampling is that an NN would require retraining for each level of sparsity. To

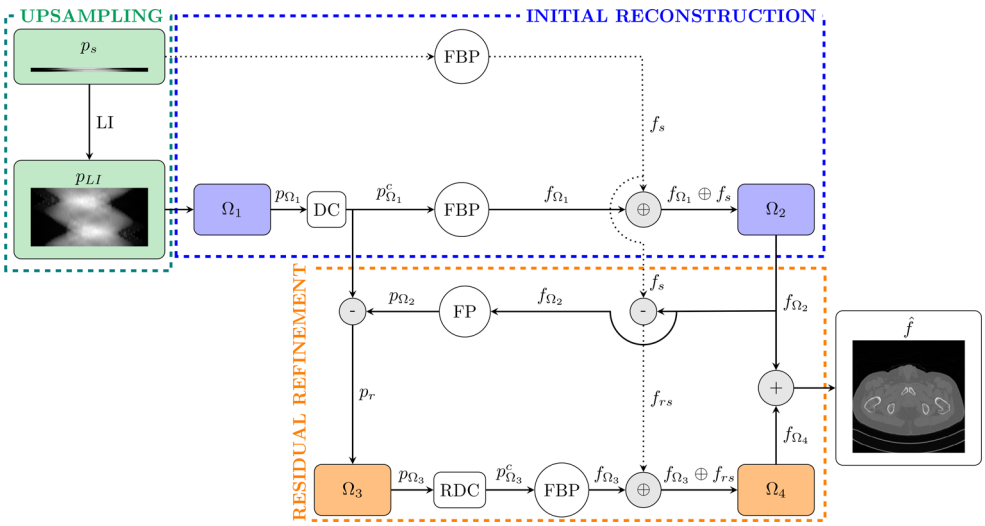


Fig. 1 Proposed DDSwinIR framework consisting of sinogram upsampling (green), initial reconstruction (blue), and residual refinement (orange). Concatenation steps are denoted with $\oplus$

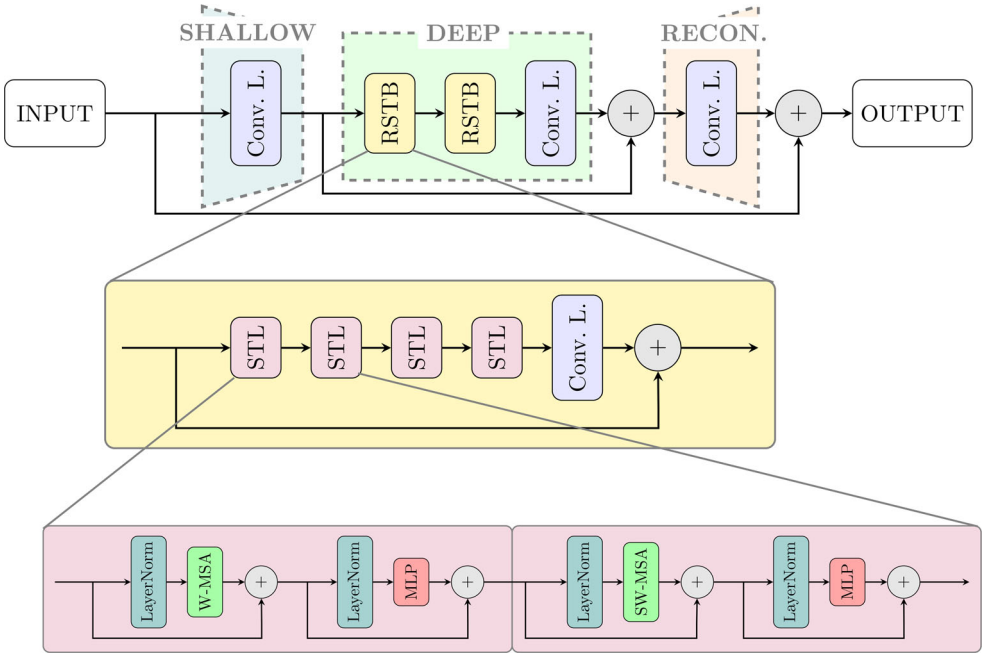


Fig. 2 Architecture of the SwinIR network, comprising shallow and deep feature extraction stages followed by an image reconstruction step

avoid this, we linearly interpolate (LI) the acquired sparse sinogram p_s, to obtain an upsampled sinogram p_{LI} that has the same dimensions as the full-view sinogram p. While interpolation introduces artifacts, these are mitigated through the initial reconstruction and residual refinement steps.

2.2 Initial Reconstruction

First, a SwinIR network, denoted Ω_1, is applied to p_{LI}, resulting in a refined sinogram

$$p_{\Omega_1} = \Omega_1(p_{\text{LI}})\,. \tag{1}$$

To enforce data consistency (i.e., minimizing the difference between the estimated sinogram p_{Ω_1} and the measured sparse sinogram p_s, mitigating errors from hallucinated features introduced by learned priors), a data consistency operation ϕ_{DC} is applied, producing a corrected sinogram

$$p^c_{\Omega_1} = \phi_{\text{DC}}(p_{\Omega_1})\,, \tag{2}$$

in which the projection values of p_{Ω_1} at the coordinates of p_s are replaced with those of p_s. If the sparse angles in p_s do not directly align with those in p_{Ω_1}, linear interpolation is first applied to p_{Ω_1} to match the angles of p_s. The resulting consistent sinogram $p^c_{\Omega_1}$ is then used to compute an initial reconstruction f_{Ω_1} using filtered back projection (FBP) [40], denoted by ϕ_{FBP}, such that

$$f_{\Omega_1} = \phi_{\text{FBP}}(p^c_{\Omega_1})\,. \tag{3}$$

Finally, f_{Ω_1} is concatenated with the sparse reconstruction

$$f_s = \phi_{\text{FBP}}(p_s)\,, \tag{4}$$

which is also obtained using FBP. This concatenation results in a two-channel input tensor, where both images are stacked along the channel dimension. This stacked input is then passed through a second SwinIR network, Ω_2, to generate the enhanced reconstruction

$$f_{\Omega_2} = \Omega_2\left(f_{\Omega_1} \oplus f_s\right)\,. \tag{5}$$

2.3 Residual Refinement

To further reduce reconstruction errors, a residual refinement step is performed. First, f_{Ω_2} is forward projected, producing the sinogram

$$p_{\Omega_2} = \phi_{\text{FP}}(f_{\Omega_2})\,, \tag{6}$$

with ϕ_{FP} denoting the forward projection operator. The residual sinogram is then computed as the difference between the initial estimate and the forward projection of the enhanced image:

$$p_r = p_{\Omega_1} - p_{\Omega_2}\,. \tag{7}$$

Next, a third SwinIR network, Ω_3, predicts a correction to this residual by estimating its deviation from the full-view

sinogram p, resulting in a refined estimate

$$p_{\Omega_3} = \Omega_3(p_r)\,. \tag{8}$$

To improve consistency with the measured data, a residual consistency operation ϕ_{RDC}, analogous to ϕ_{DC}, is applied. This step replaces values in p_{Ω_3} at the measured projection angles with the residuals between p_{Ω_2} and the sparse sinogram p_s, yielding a data-consistent residual sinogram

$$p^c_{\Omega_3} = \phi_{\text{DCR}}(p_{\Omega_3})\,. \tag{9}$$

From the sinogram $p^c_{\Omega_3}$, an image f_{Ω_3} is then reconstructed using FBP:

$$f_{\Omega_3} = \phi_{\text{FBP}}(p^c_{\Omega_3})\,, \tag{10}$$

and subsequently concatenated with the residual image

$$f_{rs} = f_s - f_{\Omega_2}\,, \tag{11}$$

which captures the discrepancy between the sparse reconstruction and the enhanced output. The concatenated result is passed through a fourth SwinIR network, Ω_4, to estimate a final refinement:

$$f_{\Omega_4} = \Omega_4\left(f_{\Omega_3} \oplus f_{rs}\right)\,. \tag{12}$$

This concatenation provides Ω_4 with enriched context, allowing it to more accurately estimate the residual between f_{Ω_2} and the ground truth image f. The final reconstruction $\hat{f}$ is obtained by adding the learned residual to f_{Ω_2}:

$$\hat{f} = f_{\Omega_2} + f_{\Omega_4}\,. \tag{13}$$

This multi-stage approach enables high-quality image reconstruction from sparse-view sinograms, while retaining flexibility by eliminating the need to retrain for varying levels of sparsity.

3 Experiments

To test the performance of the proposed DDSwinIR reconstruction framework, simulation experiments were performed. First, we examined the effect of sinogram-domain data consistency in both the sinogram and image domains by analyzing the outputs of the networks Ω_1 and Ω_3, both with and without the application of the respective consistency mechanisms, ϕ_{DC} and ϕ_{RDC}. Second, we assessed the impact of sparse-view concatenation in both the sinogram and image domains by comparing the outputs of networks Ω_2 and Ω_4 with and without concatenation applied to their inputs. Third, we quantified the contributions of each SwinIR network by evaluating performance improvements between stages, revealing how each network incrementally contributes to the progressive enhancement of reconstruction quality. Finally, we evaluated the generalizability of the framework by testing it on numbers of projections not seen during training, assessing its robustness to changes in sampling density and content. In all experiments, reconstruction performance was measured using the Peak Signal-to-Noise Ratio (PSNR) and Structural Similarity Index Metric (SSIM).

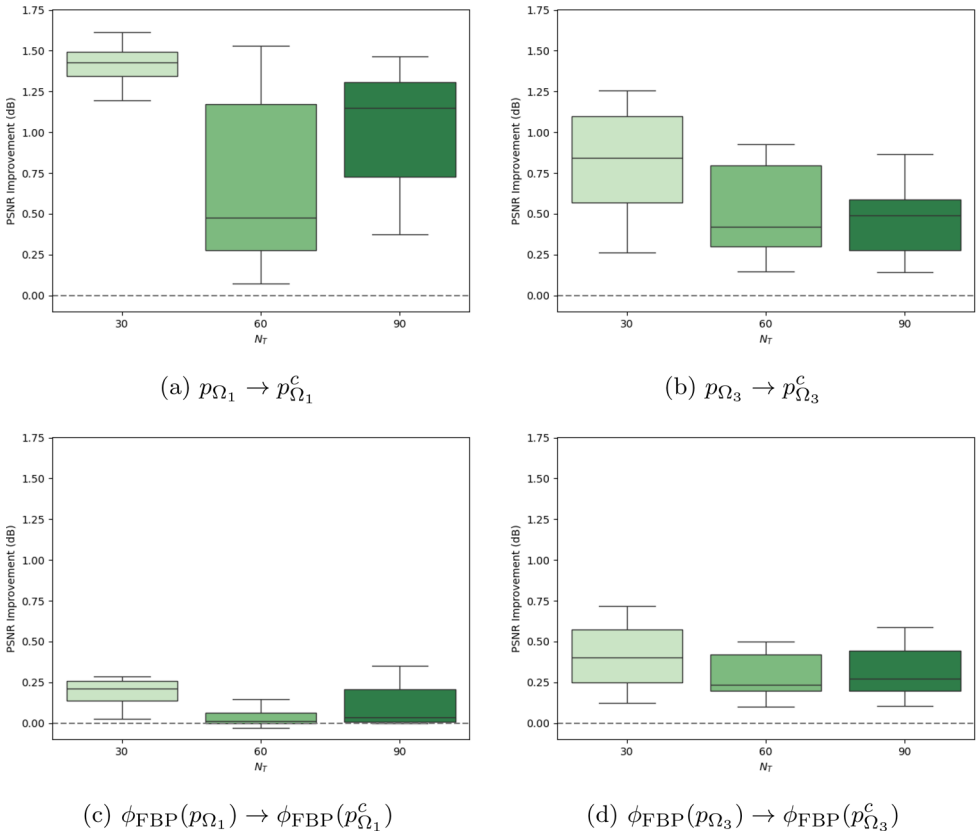
All experiments were conducted using a fan-beam CT setup, consisting of an X-ray source and detector rotating along a circular trajectory around the object. The source-to-object distance (SOD) was set to 70 cm, and the object-to-detector distance (ODD) was 35 cm. The detector spanned a total length of 80 cm and was composed of 800 equidistant square detector elements, each with a width of 1 mm. The scanned object was discretized on a 512 × 512 voxel grid, covering a 38 × 38 cm field of view centered at the rotation axis. This acquisition geometry was simulated using the ASTRA Toolbox [41] and was designed to resemble a clinical CT setup.

The SwinIR networks were trained and evaluated using the 2016 NIH-AAPM-Mayo Clinic Low Dose CT Grand Challenge dataset [42], which consists of high-resolution (512×512) CT images of the torso from 10 patients. For training, validation, and testing, 1200, 200, and 400 slices were randomly selected from 6, 2, and 2 (distinct) patients, respectively. Projection data were generated from the CT images using the forward projection operator in the ASTRA toolbox, following the previously described acquisition setup. Full-view sinograms were generated using 720 projection angles, while sparse-view sinograms were trained using $N_T = 30$, 60, or 90 projections, equiangularly selected in the range of $[0, 2\pi)$. Testing (evaluation) was conducted using $N_E = 20$ to 100 projections, in steps of 10. For $N_E = 50, 70,$ and 100, which did not align with subsets of the full-view set, interpolation between available angles was used to simulate the projections. Each reconstruction block (Ω) was trained independently using the Adam optimizer [43], with an initial learning rate of 10^{-4} and a batch size of 1. The architecture of each block included two RSTBs and four STLs. Within each STL, the W-MSA and SW-MSA used 4 attention heads and a window size of 8, while convolutional layers had an embedding dimension of 60. Training was conducted for a maximum of 200 epochs, with early stopping applied if the validation loss did not improve over three consecutive epochs relative to the lowest recorded value.

4 Results

4.1 Data Consistency

Figures 3 and 11 show PSNR and SSIM improvements, respectively, due to enforcing data consistency for models



(a) $p_{\Omega_1} \rightarrow p^c_{\Omega_1}$ (b) $p_{\Omega_3} \rightarrow p^c_{\Omega_3}$

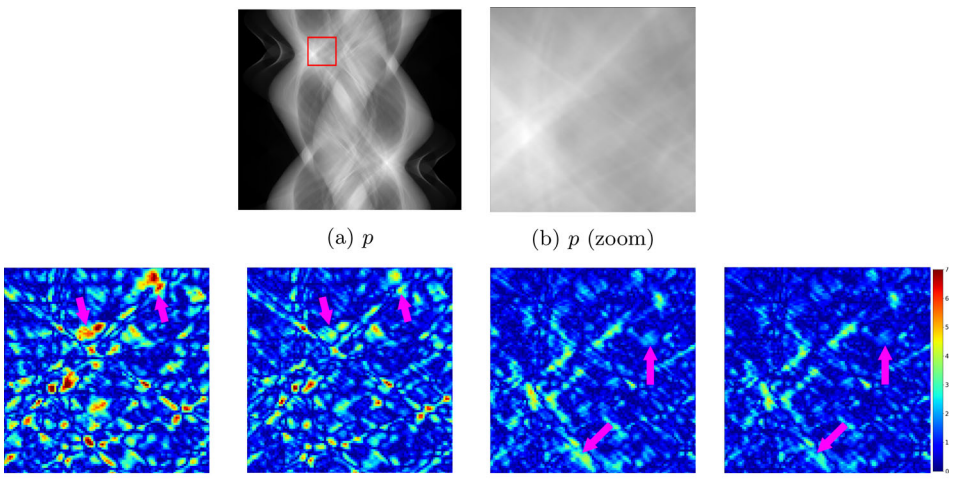
(c) $\phi_{\text{FBP}}(p_{\Omega_1}) \rightarrow \phi_{\text{FBP}}(p^c_{\Omega_1})$ (d) $\phi_{\text{FBP}}(p_{\Omega_3}) \rightarrow \phi_{\text{FBP}}(p^c_{\Omega_3})$

Fig. 3 Sinogram domain: PSNR improvements resulting from data consistency, shown for the output of Ω_1 (a) and Ω_3 (b). Image domain: Corresponding improvements observed from Ω_1 (c) and Ω_3 (d)

trained with $N_T = 30$, 60, and 90 projections, evaluated in the sinogram (Figs. 3a, 3b, 11a, and 11b) and image domains (Figs. 3c, 3d, 11c, and 11d). Each subplot presents boxplots for the three values of N_T, aggregated over all values of N_E (20 to 100 views).

In the sinogram domain, we compared p_{Ω_1} Eq. 1 with its data-consistent counterpart $p^c_{\Omega_1}$ Eq. 2 in terms of PSNR and SSIM, calculated against the full-view sinogram p, as shown in Figs. 3a and 11a, respectively. A similar comparison was made for p_{Ω_3} Eq. 8 and $p^c_{\Omega_3}$ Eq. 9, shown in Figs. 3b and 11b. In the image domain, we evaluated the FBP reconstructions $\phi_{\text{FBP}}(p_{\Omega_1})$ and $\phi_{\text{FBP}}(p^c_{\Omega_1})$ Eq. 3 in Figs. 3c and 11c, as well as $\phi_{\text{FBP}}(p_{\Omega_3})$ and $\phi_{\text{FBP}}(p^c_{\Omega_3})$ Eq. 10 in Figs. 3d and 11d.

Figure 4 illustrates the impact of data consistency in the sinogram domain. Specifically, Fig. 4a displays the sinogram p, with a highlighted Region Of Interest (ROI). A magnified view of this region is provided in Fig. 4b. The corresponding absolute error maps between p and the reconstructions for



(a) p (b) p (zoom)

(c) $|p_{\Omega_1} - p|$ (d) $|p^c_{\Omega_1} - p|$ (e) $|p_{\Omega_3} - p|$ (f) $|p^c_{\Omega_3} - p|$

Fig. 4 Ground truth (a) with zoomed region (b) indicated by a red square. Absolute error maps of the zoomed region relative to that of p without (c, e) and with (d, f) data consistency for $N_T = 30$ and $N_E = 80$

p_{Ω_1} and $p^c_{\Omega_1}$ are shown in Figs. 4c and 4d, respectively, for $N_T = 30$ and $N_E = 60$. Similarly, Figs. 4e and 4f present the absolute error maps for p_{Ω_3} and $p^c_{\Omega_3}$, respectively.

4.2 Data Concatenation

Figures 5 and 12 show PSNR and SSIM improvements, respectively, from data concatenation for models trained with $N_T = 30, 60$, and 90 projections, evaluated in the image (Figs. 5a, 5b, 12a, and 12b) and sinogram domains (Fig. 5c, 5d, 12c, and 12d). Each subplot presents boxplots for the three values of N_T, aggregated over all values of N_E.

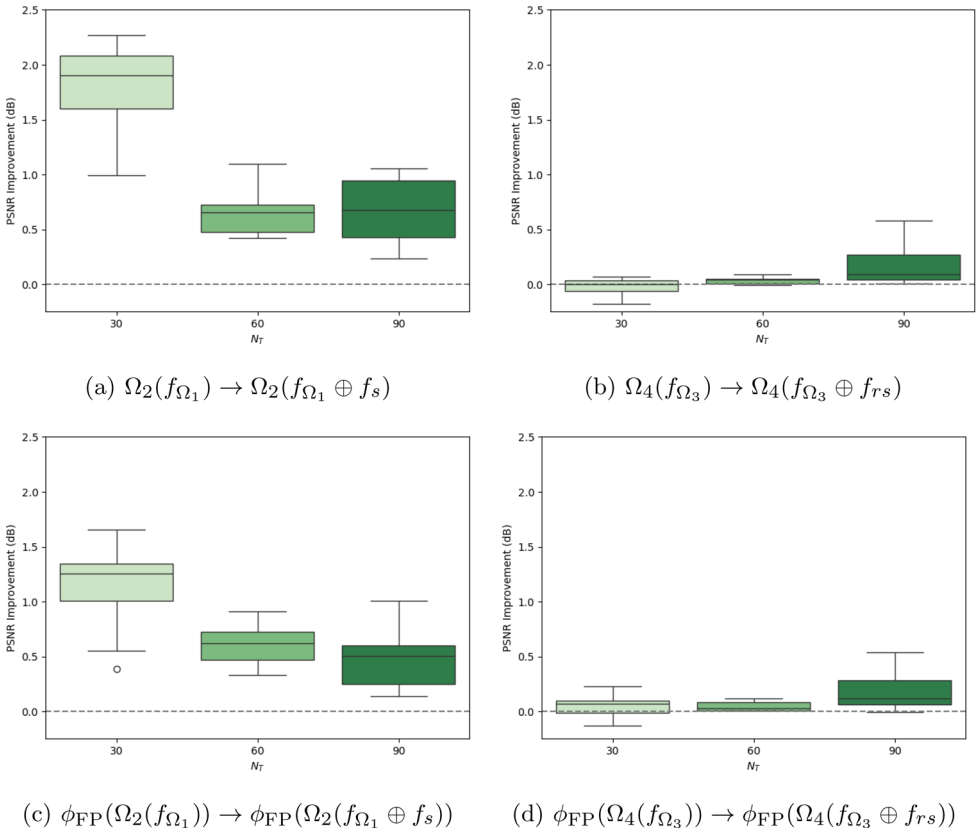
In the image domain, we evaluated the impact of concatenating the sparse reconstruction f_s during the transition from Ω_1 to Ω_2, comparing $\Omega_2(f_{\Omega_1})$ with $\Omega_2(f_{\Omega_1} \oplus f_s)$ Eq. 5, using PSNR and SSIM calculated against the full-view image f, as shown in Figs. 5a and 12a, respectively. A similar evaluation was made for the transition from Ω_3 to Ω_4, comparing $\Omega_4(f_{\Omega_3})$ and $\Omega_4(f_{\Omega_3} \oplus f_{rs})$ Eq. 12, as shown in Figs. 5b and 12b. In the sinogram domain, we assessed the effect of concatenation by comparing $\phi_{\text{FP}}(\Omega_2(f_{\Omega_1}))$ with $\phi_{\text{FP}}(\Omega_2(f_{\Omega_1} \oplus f_s))$ Eq. 6 in Figs. 5c and 12c and $\phi_{\text{FP}}(\Omega_4(f_{\Omega_3}))$ with $\phi_{\text{FP}}(\Omega_4(f_{\Omega_3} \oplus f_{rs}))$, as shown in Fig. 5d and 12d.

Figure 6 illustrates the impact of data concatenation in the image domain. Specifically, Fig. 6a displays f, with a highlighted ROI. A magnified view of this region is provided in Fig. 6b. The corresponding reconstructions for $\Omega_2(f_{\Omega_1})$ and $\Omega_2(f_{\Omega_1} \oplus f_s)$ are shown in Figs. 6c and 6d, respectively, for $N_T = 30$ and $N_E = 80$. Their respective absolute error maps, relative to f, are shown in Fig. 6g (for Figs. 6c) and 6h (for Fig. 6d). Similarly, Figs. 6e and 6f present the reconstructions for $\Omega_4(f_{\Omega_3})$ and $\Omega_4(f_{\Omega_3} \oplus f_{rs})$, respectively. The corresponding absolute error maps, relative to f, are shown in Fig. 6i (for Figs. 6e) and 6j (for Fig. 6f).

4.3 Contribution of SwinIR Networks

Figures 7 and 13 show PSNR and SSIM improvements, respectively, in the sinogram (left) and image (right) domain for models trained with $N_T = 30, 60$, and 90 projections (shown across rows). The boxplots in Figs. 7a, 7c, and 7e summarize sinogram-domain PSNR gains for $N_T = 30, 60$, and 90, respectively, across all values of N_E, covering four transitions: from linear interpolation to the output of Ω_1 ($p_{\text{LI}} \rightarrow p^c_{\Omega_1}$) Eq. 2, from Ω_1 to Ω_2 ($p^c_{\Omega_1} \rightarrow p_{\Omega_2}$) Eqs. 2 and 6, the addition of residual correction via Ω_3 ($p_{\Omega_2} \rightarrow p_{\Omega_2} + p^c_{\Omega_3}$) Eqs. 6 and 9, and final refinement with Ω_4 ($p_{\Omega_2} + p^c_{\Omega_3} \rightarrow p_{\Omega_2} + p_{\Omega_4}$) Eqs. 6 and 9, where p_{Ω_4} is defined as:

$$p_{\Omega_4} = \phi_{\text{FP}}(f_{\Omega_4}) \,. \tag{14}$$



(a) $\Omega_2(f_{\Omega_1}) \rightarrow \Omega_2(f_{\Omega_1} \oplus f_s)$ (b) $\Omega_4(f_{\Omega_3}) \rightarrow \Omega_4(f_{\Omega_3} \oplus f_{rs})$

(c) $\phi_{\text{FP}}(\Omega_2(f_{\Omega_1})) \rightarrow \phi_{\text{FP}}(\Omega_2(f_{\Omega_1} \oplus f_s))$ (d) $\phi_{\text{FP}}(\Omega_4(f_{\Omega_3})) \rightarrow \phi_{\text{FP}}(\Omega_4(f_{\Omega_3} \oplus f_{rs}))$

Fig. 5 Image domain: PSNR improvements resulting from sparse-view data concatenation, shown for transitions from Ω_1 to Ω_2 (a) and Ω_3 to Ω_4 (b). Sinogram domain: Corresponding improvements observed from Ω_1 to Ω_2 (c) and Ω_3 to Ω_4 (d)

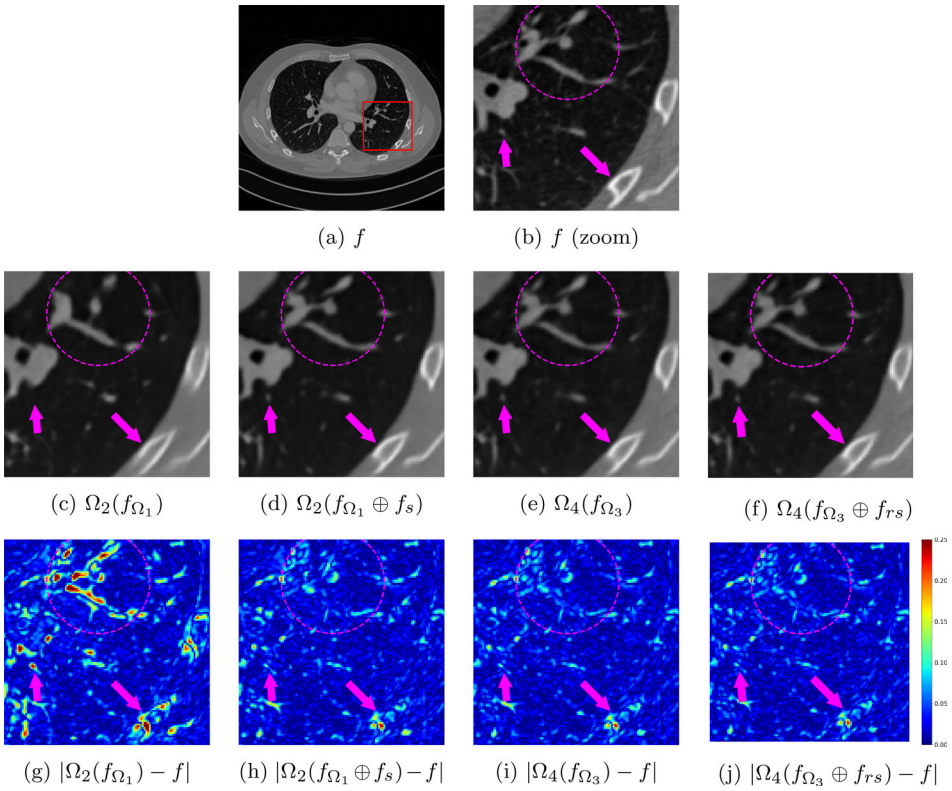


Fig. 6 Ground truth (a) with zoomed region (b) indicated by a red square. Reconstructions and corresponding absolute error maps of the zoomed region without (c, e, g, i) and with (d, f, h, j) data concatenation for $N_T = 30$ and $N_E = 60$

In the same way, the boxplots in Figs. 13a, 13c, and 13e summarize sinogram-domain SSIM gains. The scatterplots in Figs. 7b, 7d, and 7f illustrate image-domain cumulative PSNR gains for $N_T = 30, 60$, and 90, respectively, covering the corresponding transitions: Ω_1 ($f_{\text{LI}} \rightarrow f_{\Omega_1}$) Eq. 3, Ω_2 ($f_{\Omega_1} \rightarrow f_{\Omega_2}$) Eq. 3, 5, Ω_3 ($f_{\Omega_2} \rightarrow f_{\Omega_2} + f_{\Omega_3}$) Eq. 5, 10, and Ω_4 ($f_{\Omega_2} + f_{\Omega_3} \rightarrow f_{\Omega_2} + f_{\Omega_4}$) Eqs. 5, 10, 13. In the same way, the scatterplots in Figs. 13b, 13d, and 13f illustrate image-domain cumulative SSIM gains. Each scatter point represents performance at a specific value of N_E (for a fixed N_T), evaluated relative to the interpolated input. Complementing this analysis, Fig. 8 presents the absolute difference between the outputs of successive networks, trained and tested for 60 views, in both the sinogram (upper panel in Fig. 8) and image (lower panel in Fig. 8) domains. These maps highlight localized changes after each transition, providing further insight into the incremental contributions of each SwinIR network in the reconstruction framework.

4.4 Generalization

Figure 9 presents the quantitative absolute error maps with respect to the ground truth image f (Fig. 8f). These results correspond to reconstructions obtained using FBP, SIRT, an alternative implementation of the proposed framework based on U-Nets (cfr. DDUNet) [18], and DDSwinIR. All models were trained with $N_T = 30, 60$, and 90 views, and evaluated for $N_E = 20, ..., 100$ views. Fig. 10 shows the quantitative absolute error maps in the sinogram domain, relative to the reference p (Fig. 8a), for linear interpolation, DDUNet, and DDSwinIR models trained on $N_T = 30, 60$, and 90 views, again evaluated across $N_E = 20, ..., 100$ views.

5 Discussion

5.1 Data Consistency

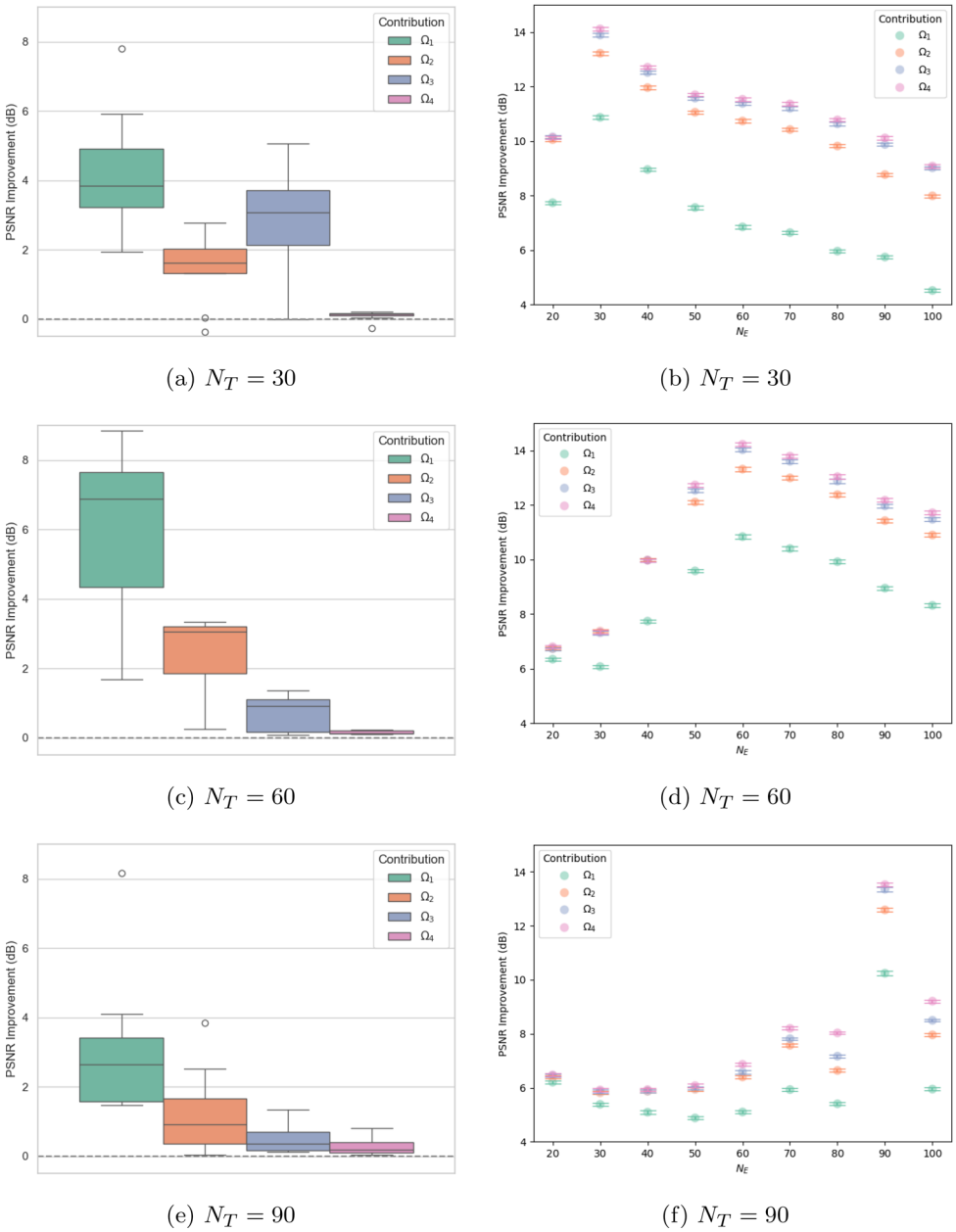
In the sinogram domain, enforcing data consistency after Ω_1 results in notable PSNR gains, particularly at lower N_T. For $N_T = 90$, the median PSNR improvement is approximately 1.2 dB, while at $N_T = 30$, the PSNR gain amounts to 1.4 dB. At $N_T = 60$, the median PSNR improvement is more modest (~ 0.5 dB), but the broader spread suggests sensitivity to specific values of N_E. Some of this variability may reflect the influence of interpolated projections used for $N_E = 50, 70$, and 90, particularly near $N_T = 60$ and $N_T = 90$. When data consistency is applied after Ω_3, a comparable trend is observed. The same patterns for Ω_1 and Ω_3 also appear in the image domain, albeit with generally smaller improvements. This indicates that the primary impact of the consistency step lies in correcting the data in projection space. Nonetheless, the observed improvements in the image domain support enforcement of data consistency at multiple stages (after Ω_1 and Ω_3) within the reconstruction framework.

The SSIM results complement the PSNR findings and underscore the value of enforcing data consistency throughout the reconstruction framework, particularly in the image domain.

These improvements are visually supported by Fig. 4. Enforcing data consistency after Ω_1 reduces the difference between the reconstructed and ground-truth sinogram, as highlighted by the pink arrows when comparing results without (Fig. 4c) and with (Fig. 4d) data consistency. A similar, though weaker, improvement is seen after Ω_3, in Figs. 4e and 4f, respectively. It is noteworthy that these improvements come with only a marginal computational cost. Specifically, enforcing data consistency after Ω_1 increases computation time by just (3.7 ± 0.8) ms, and after Ω_3 by (11.3 ± 1.2) ms, corresponding to just 0.01% and 0.17% of the total runtime, respectively.

5.2 Data Concatenation

The results in Fig. 5 demonstrate the effectiveness of sparse-view data concatenation, particularly during the transition from Ω_1 to Ω_2. Integrating sparse reconstructions early into the framework improves PSNR and SSIM in both the image and sinogram domains. The improvement in PSNR is most pronounced under high sparsity conditions (e.g., $N_T = 30$), where additional image-domain information from f_s Eq.



(a) $N_T = 30$ (b) $N_T = 30$

(c) $N_T = 60$ (d) $N_T = 60$

(e) $N_T = 90$ (f) $N_T = 90$

Fig. 7 Each row corresponds to a specific value of N_T (30, 60, or 90 views, in order). Sinogram domain: (a, c, e) Boxplots show PSNR improvements for Ω_1 (green), Ω_2 (orange), Ω_3 (purple), and Ω_4 (pink), over all values of N_E. Image domain: (b, d, f) Scatter plots show cumulative PSNR improvements of the networks relative to the interpolation baseline, evaluated at each value of N_E, along with the corresponding standard errors

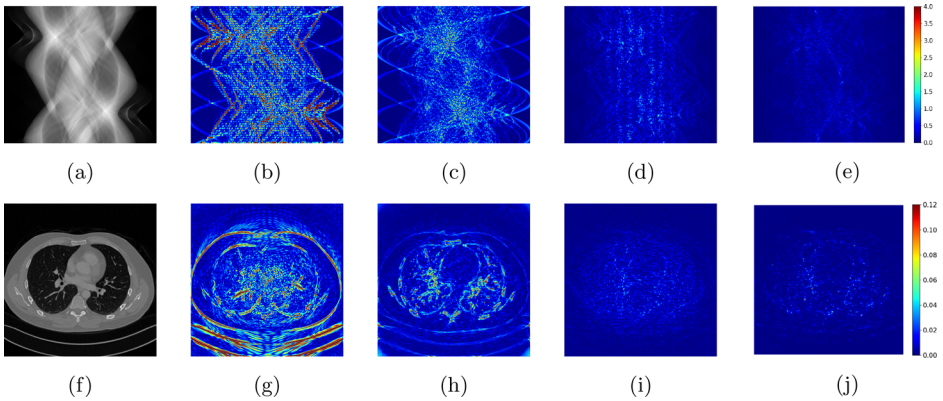


Fig. 8 Ground truth (a, f) and absolute differences (b-e, g-j) between successive intermediate reconstructions for the model with $N_T = N_E = 60$ projections. The upper panel shows the sinogram domain: (a) p, (b) $p_{\text{LI}} \rightarrow p^c_{\Omega_1}$, (c) $p^c_{\Omega_1} \rightarrow p_{\Omega_2}$, (d) $p_{\Omega_2} \rightarrow p_{\Omega_2} + p^c_{\Omega_3}$, and (e) $p_{\Omega_2} + p^c_{\Omega_3} \rightarrow p_{\Omega_2} + p_{\Omega_4}$. The lower pannel shows the image domain: (f) f, (g) $f_{\text{LI}} \rightarrow f_{\Omega_1}$, (h) $f_{\Omega_1} \rightarrow f_{\Omega_2}$, (i) $f_{\Omega_2} \rightarrow f_{\Omega_2} + f_{\Omega_3}$, and (j) $f_{\Omega_2} + f_{\Omega_3} \rightarrow f_{\Omega_2} + f_{\Omega_4}$

4 or $f_r s$ Eq. 11 significantly enhances performance gains. Pixel fidelity wise, this emphasizes the value of early feature fusion in addressing data insufficiency and enhancing reconstruction quality. As N_T increases to 60 or 90 projections, the benefits of concatenation diminish, indicating reduced dependence on auxiliary sparse-view features when more projection data is available. While PSNR improvements are largest at $N_T = 30$, SSIM gains in the image domain are smaller at this number of projections trained on and become more noticeable at $N_T = 60$ and 90. This suggests early concatenation mainly boosts pixel-wise accuracy under extreme sparsity, whereas structural similarity benefits more with moderate to lower sparsity. In the refinement stage ($\Omega_3 \rightarrow \Omega_4$), a slight PSNR and SSIM drop is observed for $N_T = 30$, suggesting that late-stage concatenation may degrade performance when projection data is sparse. For $N_T = 60$ and 90, PSNR modestly improves, suggesting that residual concatenation becomes more beneficial when the network is less constrained by data. At $N_T = 90$, SSIM in the image domain also slightly decreases, even though PSNR rises, suggesting that over-refinement may affect structural quality when data is abundant. In the sinogram domain, the PSNR and SSIM gains are less significant but mirror the overall trend, implying that image-domain concatenation contributes indirectly to sinogram reconstruction through improved forward projections.

The benefits of data concatenation are also evident in Fig. 6. Without concatenation, key anatomical details, such as the structural features and white specks highlighted by the arrows and circle, are poorly recovered (Figs. 6c and 6g), whereas with concatenation, these features reappear (Figs. 6d and 6h), indicating substantial improvements in reconstruction quality. After Ω_3, however, the differences are minimal, with only minor details being better preserved (Figs. 6e and 6i vs. 6f and 6j). These improvements are achieved with only moderate additional cost: data concatenation for $\Omega_2(f_{\Omega_1} \oplus f_s)$ Eq. 5 and $\Omega_4(f_{\Omega_3} \oplus f_{rs})$ Eq. 12 increases computational cost by (158.2 ± 13.2) ms and (190.4 ± 13.6) ms, respectively, corresponding to only 2.4% and 2.9% of the total runtime. The final data concatenation step leads to only a marginal gain in PSNR. Hence, although its computational cost is marginal as well, it could be omitted.

5.3 Contribution of SwinIR Networks

Across all values of N_T, the Ω_1 module consistently accounts for the largest PSNR and SSIM gains. For the PSNR, this effect is most pronounced at $N_T = 60$, likely reflecting a trade-off between data sufficiency and the potential for further improvement. PSNR responds strongly to pixel-level corrections, which are effectively introduced at this stage, whereas SSIM, which captures structural similarity, tends to improve more gradually. Enforcing data consistency at this stage (see Fig. 3a) further amplifies the effectiveness of Ω_1. Subsequent networks (Ω_2, Ω_3, and Ω_4) account for progressively smaller gains. Nonetheless, these later stages remain valuable. In the model trained with $N_T = 30$, a notable deviation is observed for the PSNR: Ω_3 contributes more than Ω_2 in the sinogram domain. This likely reflects the difficulty of Ω_2, operating in the image domain, to extract reliable features from severely limited input data. This findings underscore the importance of residual correction in low-data regimes and support the inclusion of all networks, even when incremental gains diminish.

This interpretation is reinforced by the visualizations in Fig. 8. Absolute error maps between output of two successive networks confirm the observed trends: the most significant structural changes occur after applying Ω_1, with subsequent stages introducing more subtle refinements. These results indicate that while the bulk of the reconstruction

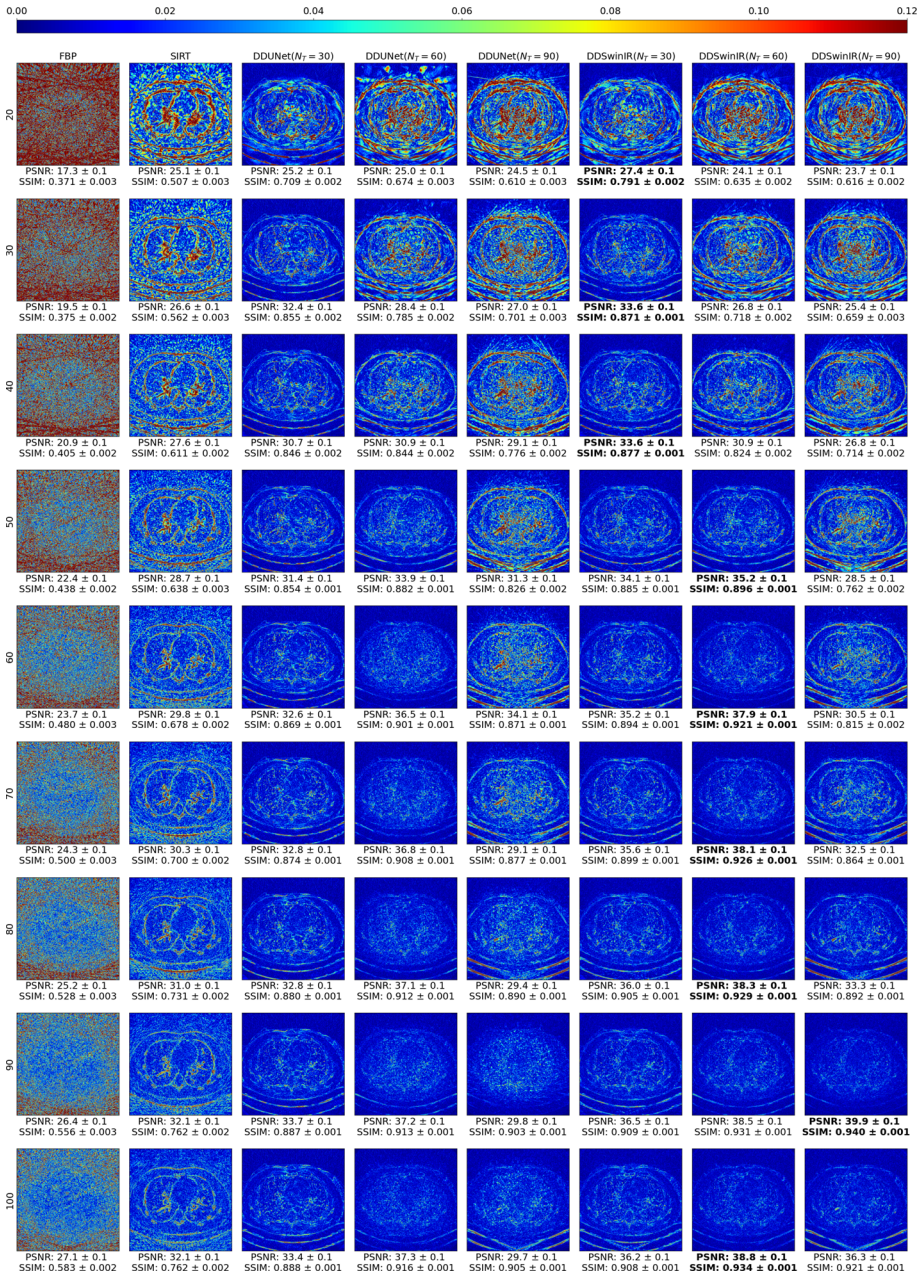


Fig. 9 Absolute error maps between f and reconstructions obtained using FBP, SIRT, DDUNet, and DDSwinIR models trained on $N_T = 30, 60$, and 90 views. Columns represent methods; rows correspond to the values of N_E. Results with the highest PSNR and SSIM are marked in bold

improvement is achieved during the initial reconstruction, the subsequent networks are necessary to attain optimal image quality.

The scatter plots in Fig. 7 reveal a clear asymmetry in model generalization: performance deteriorates significantly when $N_E < N_T$, whereas degradation is far less pronounced when $N_E > N_T$. This suggests that Ω_1, in particular, learns to exploit the spatial layout of the interpolated sinogram. A mismatch in angular coverage, where $N_E < N_T$, disrupts this learned representation, limiting the network's ability to leverage measured data. In contrast, when additional projections are present during testing (i.e., $N_E > N_T$), the interpolated sinogram incorporates a greater proportion of accurate data.

5.4 Generalization

In the image domain, sparse-view FBP reconstructions are dominated by streak artifacts that intensify with increasing sinogram sparsity, resulting in significant degradation in both PSNR and SSIM. SIRT mitigates these artifacts through iterative refinement, albeit at the expense of image sharpness and the loss of fine structural detail. DDUNet reconstructs images with improved quality in terms of PSNR and SSIM compared to FBP and SIRT. However, our proposed DDSwinIR achieves the most effective suppression of undersampling artifacts while preserving high-frequency features, yielding superior PSNR and SSIM, as illustrated by the absolute error

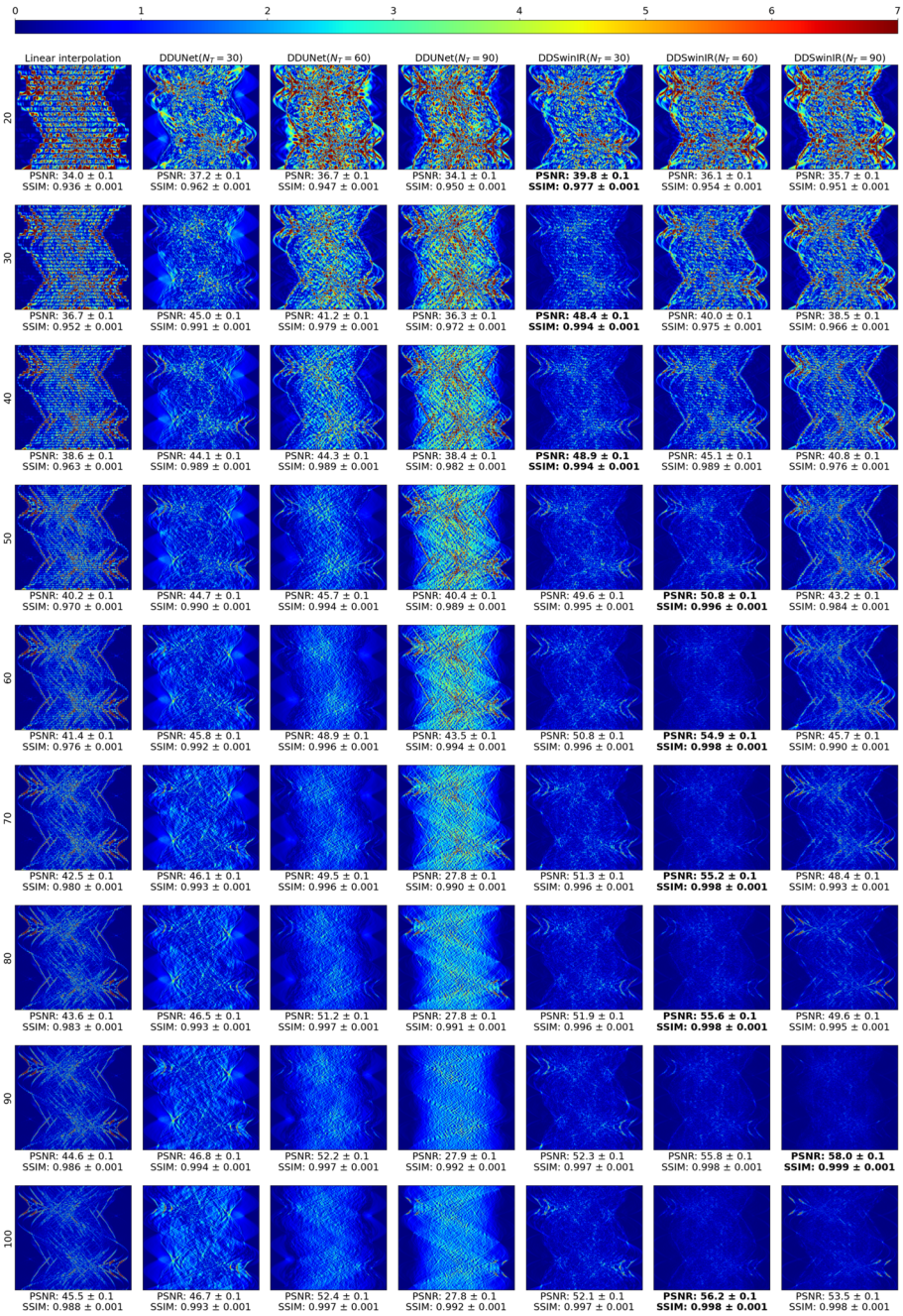


Fig. 10 Absolute error maps between p and either the linearly interpolated sinogram at N_E, or the forward projections of DDUNet and DDSwinIR reconstructions trained on $N_T = 30$, 60, and 90 views. Columns represent methods; rows correspond to the values of N_E. Results with the highest PSNR and SSIM are marked in bold

maps in Fig. 9, particularly when $N_T = N_E$. A notable exception occurs when DDSwinIR is trained with $N_T = 90$ and evaluated at $N_E < 60$. Under these conditions, performance deteriorates significantly. This performance gap reflects a pronounced sensitivity to a mismatch in projection numbers N_T and N_E. Specifically, models trained on densely sampled data exhibit limited adaptability when confronted with severe undersampling. For example, using a model trained with $N_T = 90$ and tested at $N_E = 30$ results in a substantial performance drop, 8.2 dB in PSNR and 0.212 in SSIM, compared to a model trained and tested at $N_T = N_E = 30$. In contrast, when training at $N_T = 30$ and testing at $N_E = 90$, the degradation is smaller, only 3.3 dB in PSNR and 0.031 in SSIM, relative to the $N_T = N_E = 90$ baseline. This asymmetry strongly supports the strategy of training at the lower end of the intended projection range, as it improves the network's ability to generalize across a broader set of unseen numbers of projections.

The sinogram-domain results in Fig. 10 reinforce these conclusions. Linear interpolation of the sinogram introduces structured and repetitive artifacts into the upsampled sinogram. DDUNet reduces these artifacts to some extent but

shows significant performance degradation at $N_T = 90$. In contrast, the DDSwinIR framework effectively reduces artifacts and reconstructs images with consistently higher PSNR and SSIM across all N_E, demonstrating the ability of the framework to effectively restore projection data.

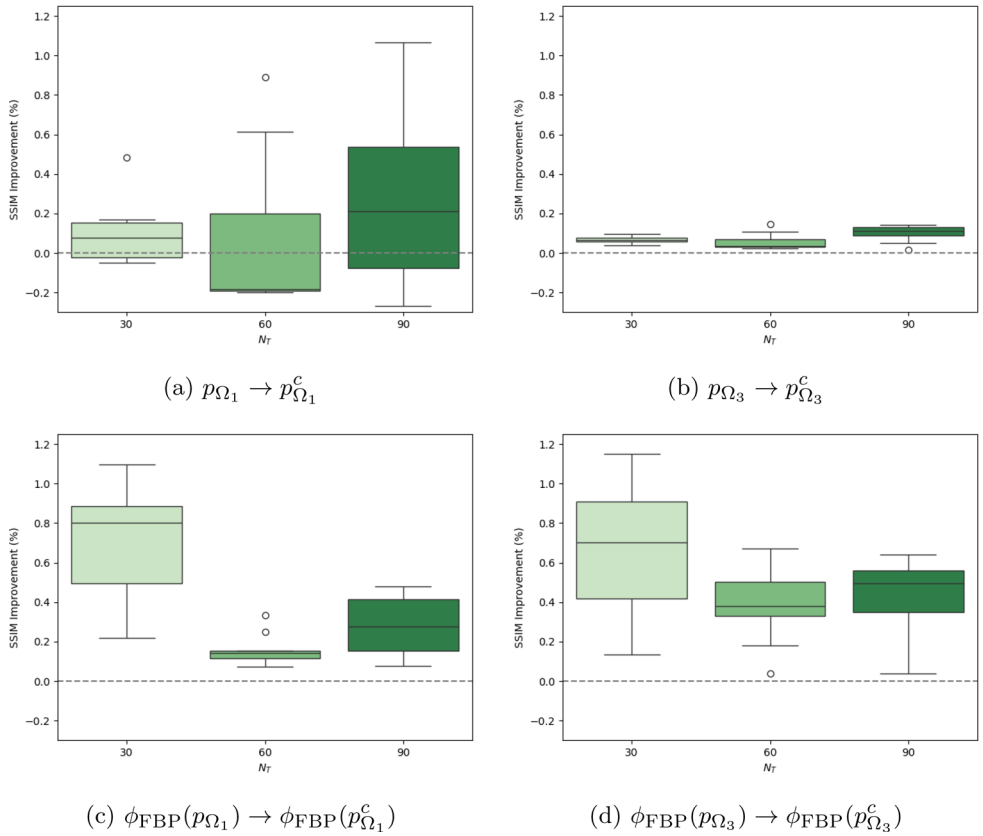
6 Conclusion

This work introduces DDSwinIR, a modular deep learning framework for sparse-view CT reconstruction, built on a dual-domain design. Rather than prioritizing performance over existing deep learning models, DDSwinIR is purpose-built to disentangle and quantify the contributions of each reconstruction stage, offering interpretability. Using the proven SwinIR architecture, the framework integrates sinogram upsampling, initial reconstruction, and residual refinement into a cohesive pipeline. The results show that, while the majority of structural corrections are achieved during sinogram upsampling and the initial reconstruction, dual-domain residual refinement is crucial for eliminating remaining errors and improving overall quality.

Notably, data consistency and sparse-view concatenation within the initial reconstruction module are key drivers of the most substantial and reliable gains, particularly under extreme sparsity, while maintaining low computational cost. Furthermore, our results reveal a clear asymmetry in model generalization: networks trained on sparse data generalize more effectively to denser views than vise versa. This suggests that models trained on the lower end of the projection range are better equipped for real-world sparse-view scenarios, where acquisition conditions are often variable.

Looking ahead, the framework can be further enhanced by replacing linear interpolation with more advanced sinogram upsampling techniques and incorporating sinogram-domain concatenation that explicitly encodes the locations of measured views. In addition, a single unified model trained on mixed sparsity levels may improve generalization by helping the model better identify and leverage the positions of measured views. Finally, while this work focuses on 2D fan-beam geometry, DDSwinIR is compatible with 3D cone-beam CT, as demonstrated by the successful extension of Swin Transformers to 3D in recent works [44, 45].

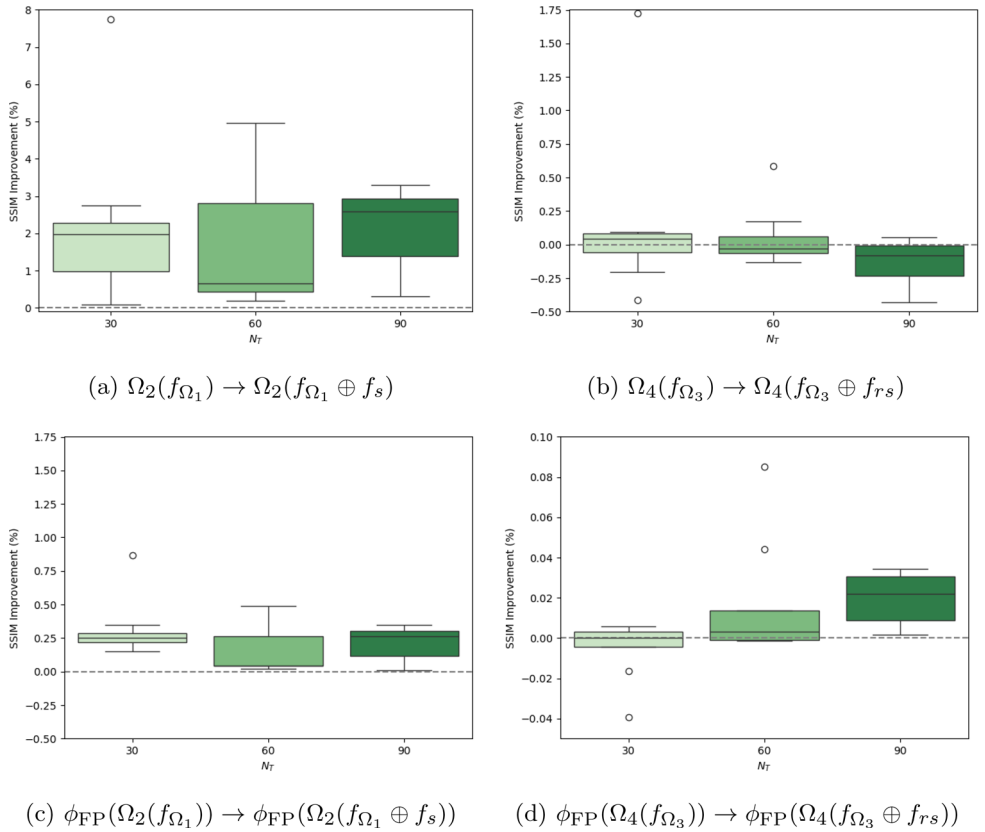
Appendix



(a) $p_{\Omega_1} \rightarrow p^c_{\Omega_1}$ (b) $p_{\Omega_3} \rightarrow p^c_{\Omega_3}$

(c) $\phi_{\mathrm{FBP}}(p_{\Omega_1}) \rightarrow \phi_{\mathrm{FBP}}(p^c_{\Omega_1})$ (d) $\phi_{\mathrm{FBP}}(p_{\Omega_3}) \rightarrow \phi_{\mathrm{FBP}}(p^c_{\Omega_3})$

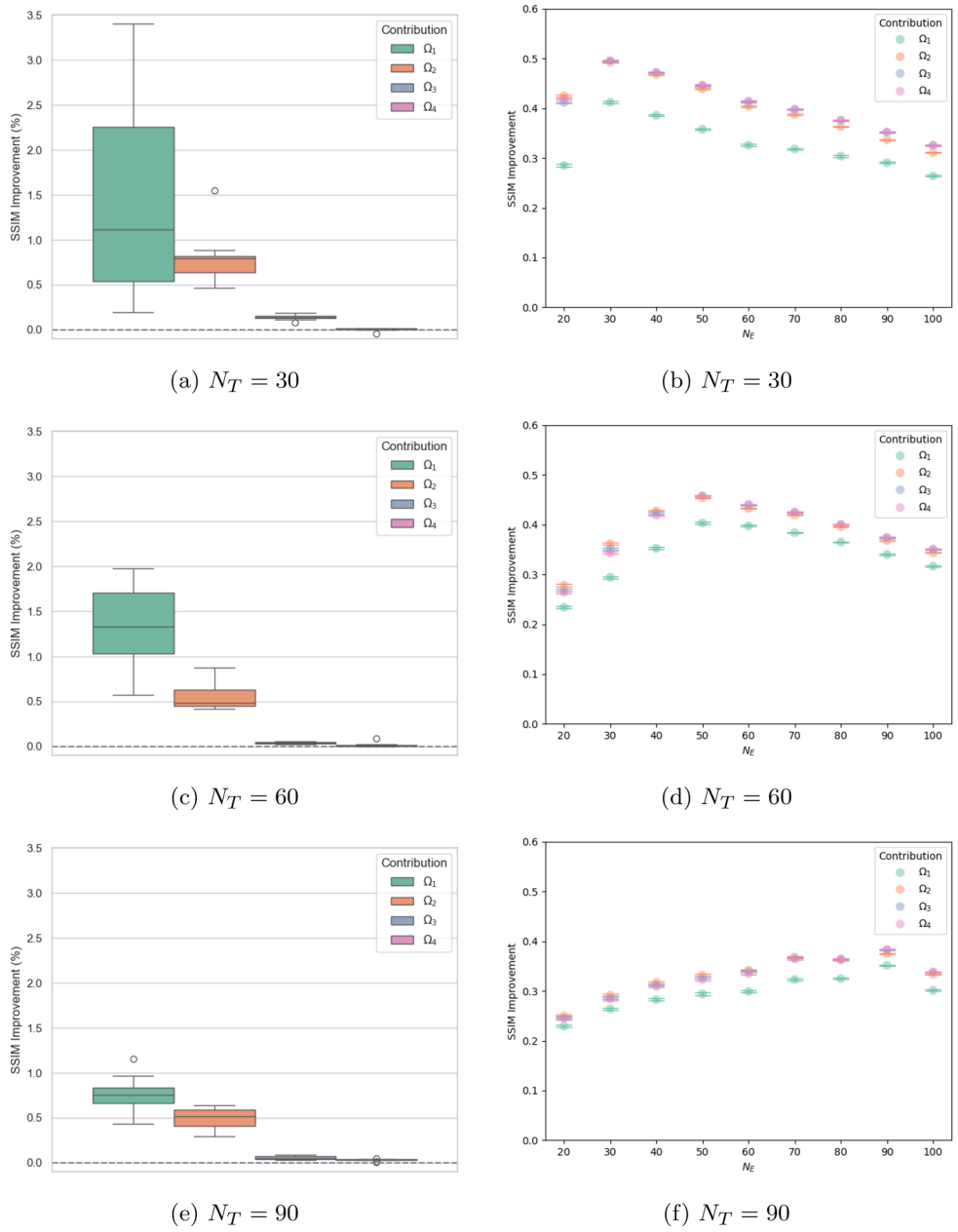
Fig. 11 Sinogram domain: SSIM improvements resulting from data consistency, shown for the output of Ω_1 (a) and Ω_3 (b). Image domain: Corresponding improvements observed from Ω_1 (c) and Ω_3 (d)



(a) $\Omega_2(f_{\Omega_1}) \rightarrow \Omega_2(f_{\Omega_1} \oplus f_s)$ (b) $\Omega_4(f_{\Omega_3}) \rightarrow \Omega_4(f_{\Omega_3} \oplus f_{rs})$

(c) $\phi_{\mathrm{FP}}(\Omega_2(f_{\Omega_1})) \rightarrow \phi_{\mathrm{FP}}(\Omega_2(f_{\Omega_1} \oplus f_s))$ (d) $\phi_{\mathrm{FP}}(\Omega_4(f_{\Omega_3})) \rightarrow \phi_{\mathrm{FP}}(\Omega_4(f_{\Omega_3} \oplus f_{rs}))$

Fig. 12 Image domain: SSIM improvements resulting from sparse-view data concatenation, shown for transitions from Ω_1 to Ω_2 (a) and Ω_3 to Ω_4 (b). Sinogram domain: Corresponding improvements observed from Ω_1 to Ω_2 (c) and Ω_3 to Ω_4 (d)



(a) $N_T = 30$ (b) $N_T = 30$

(c) $N_T = 60$ (d) $N_T = 60$

(e) $N_T = 90$ (f) $N_T = 90$

Fig. 13 Each row corresponds to a specific value of N_T (30, 60, or 90 views, in order). Sinogram domain: (a, c, e) Boxplots show SSIM improvements for Ω_1 (green), Ω_2 (orange), Ω_3 (purple), and Ω_4 (pink), over all values of N_E. Image domain: (b, d, f) Scatter plots show cumulative SSIM improvements of the networks relative to the interpolation baseline, evaluated at each value of N_E, along with the corresponding standard errors

Acknowledgements The authors have received funding from Research Foundation Flanders (G090020N), the Flemish Government under *Onderzoeksprogramma Artificiële Intelligentie (AI) Vlaanderen*, and from the University of Antwerp (50954).

Author Contributions J.V.d.R.: Investigation, Methodology, Software, Validation, Visualization, Writing– original draft, Writing– review &; editing. C.B.: Methodology, Validation, Visualization, Writing– original draft, Writing– review &; editing. S.E.V.: Supervision, Review &; editing. J.S.: Conceptualization, Supervision, Writing– review &; editing.

Data Availability No datasets were generated or analysed during the current study.

Declarations

Competing interests The authors declare no competing interests.

References

1. Brenner, D.J., Hall, E.J.: Computed tomography-An increasing source of radiation exposure. N. Engl. J. Med. **357**(22), 2277–2284 (2007)
2. Di, J., Lin, J., Zhong, L., Qian, K., Qin, Y.: Review of sparse-view or limited-angle CT reconstruction based on deep learning. Laser & Optoelectronics Progress **60**(8), 0811002 (2023)
3. Kalke, M., Siltanen, S.: Sinogram interpolation method for sparse-angle tomography. Appl. Math. **5**(3), 423–441 (2014)
4. Kim, H.-G., Yoo, H.: Image enhancement for computed tomography using directional interpolation for sparsely-sampled sinogram. Optik **166**, 227–235 (2018)
5. Van Gompel, G., Defrise, M., Van Dyck, D.: Elliptical extrapolation of truncated 2D CT projections using Helgason-Ludwig consistency conditions. In: Medical Imaging 2006: Physics of Medical Imaging, vol. 6142, pp. 1408–1417 (2006)
6. Geraldo, R.J., Cura, L.M., Cruvinel, P.E., Mascarenhas, N.D.: Low dose CT filtering in the image domain using MAP algorithms. IEEE Trans. Rad. Plasma Med. Sci. **1**(1), 56–67 (2016)
7. Zeng, G.L.: An attempt of directly filtering the sparse-view CT images by BM3D. In: 7th International Conference on Image Formation in X-Ray Computed Tomography, vol. 12304, pp. 568–575 (2022)
8. Trampert, J., Leveque, J.-J.: Simultaneous iterative reconstruction technique: Physical interpretation based on the generalized least squares solution. J. Geophys. Res. Solid Earth **95**(B8), 12553–12559 (1990)
9. Huang, K., Gao, Z., Yang, F., Zhang, H., Zhang, D.: An improved discrete algebraic reconstruction technique for limited-view based on gray mean value guidance. J. Nondestr. Eval. **42**(1), 6 (2023)
10. Lv, L., Li, C., Wei, W., Sun, S., Ren, X., Pan, X., Li, G.: Optimization of sparse-view CT reconstruction based on convolutional neural network. Med. Phys. **52**(4), 2089–2105 (2025)
11. Lee, M., Kim, H., Kim, H.-J.: Sparse-view CT reconstruction based on multi-level wavelet convolution neural network. Physica Med. **80**, 352–362 (2020)
12. Ulyanov, D., Vedaldi, A., Lempitsky, V.: Deep image prior. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 9446–9454 (2018)
13. Baguer, D.O., Leuschner, J., Schmidt, M.: Computed tomography reconstruction using deep image prior and learned reconstruction methods. Inverse Prob. **36**(9), 094004 (2020)
14. Shu, Z., Entezari, A.: Sparse-view and limited-angle CT reconstruction with untrained networks and deep image prior. Comput. Methods Programs Biomed. **226**, 107167 (2022)
15. Bossuyt, C., De Beenhouwer, J., Sijbers, J., Iuso, D., Costin, M., Escoda, J., Le Hoang, T.V., Dekker, A.J.: Deep image prior for sparse-view reconstruction in static, rectangular multi-source X-ray CT systems for cargo scanning. In: Developments in X-Ray Tomography XV, vol. 13152, p. 131520 (2024). SPIE
16. Shu, Z., Pan, Z.: SDIP: Self-reinforcement deep image prior framework for image processing (2024). arXiv preprint arXiv:2404.12142
17. Shu, Z., Entezari, A.: RBP-DIP: Residual back projection with deep image prior for ill-posed CT reconstruction. Neural Netw. **180**, 106740 (2024)
18. Ronneberger, O., Fischer, P., Brox, T.: U-Net: Convolutional networks for biomedical image segmentation. In: Medical Image Computing and Computer-assisted Intervention–MICCAI 2015: 18th International Conference, Munich, Germany, October 5-9, 2015, Proceedings, Part III 18, pp. 234–241 (2015). Springer
19. Han, Y., Ye, J.C.: Framing U-Net via deep convolutional framelets: Application to sparse-view CT. IEEE Trans. Med. Imaging **37**(6), 1418–1429 (2018)
20. Liu, P., Zhang, H., Lian, W., Zuo, W.: Multi-level wavelet convolutional neural networks. IEEE Access **7**, 74973–74985 (2019)
21. Gupta, H., Jin, K.H., Nguyen, H.Q., McCann, M.T., Unser, M.: CNN-based projected gradient descent for consistent CT image reconstruction. IEEE Trans. Med. Imaging **37**(6), 1440–1453 (2018)
22. Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., Dehghani, M., Minderer, M., Heigold, G., Gelly, S., et al.: An image is worth 16x16 words: Transformers for image recognition at scale (2020). arXiv preprint arXiv:2010.11929
23. Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, Ł., Polosukhin, I.: Attention is all you need. Adv. Neural Inform. Process. Syst. **30** (2017)
24. Liu, Z., Lin, Y., Cao, Y., Hu, H., Wei, Y., Zhang, Z., Lin, S., Guo, B.: Swin transformer: Hierarchical vision transformer using shifted windows. In: Proceedings of the IEEE/CVF International Conference on Computer Vision, pp. 10012–10022 (2021)
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 **37**(6), 1407–1417 (2018)
26. Li, Z., Cai, A., Wang, L., Zhang, W., Tang, C., Li, L., Liang, N., Yan, B.: Promising generative adversarial network based sinogram inpainting method for ultra-limited-angle computed tomography imaging. Sensors **19**(18), 3941 (2019)
27. Dong, X., Vekhande, S., Cao, G.: Sinogram interpolation for sparse-view micro-CT with deep learning neural network. In: Med-

ical Imaging 2019: Physics of Medical Imaging, vol. 10948, pp. 692–698 (2019)
28. Guan, B., Yang, C., Zhang, L., Niu, S., Zhang, M., Wang, Y., Wu, W., Liu, Q.: Generative modeling in sinogram domain for sparse-view CT reconstruction. IEEE Trans. Rad. Plasma Med. Sci. **8**(2), 195–207 (2023)
29. Kuo, C., Wei, T.-T., Chen, J.-J., Tseng, Y.-C.: Circular LSTM for low-dose sinograms inpainting. IEEE Access **11**, 78480–78488 (2023)
30. Lee, D., Choi, S., Kim, H.-J.: High quality imaging from sparsely sampled computed tomography data with deep learning and wavelet transform in various domains. Med. Phys. **46**(1), 104–115 (2019)
31. Wang, C., Shang, K., Zhang, H., Li, Q., Zhou, S.K.: DuDoTrans: dual-domain transformer for sparse-view CT reconstruction. In: International Workshop on Machine Learning for Medical Image Reconstruction, pp. 84–94 (2022)
32. Yuan, H., Jia, J., Zhu, Z.: Sipid: A deep learning framework for sinogram interpolation and image denoising in low-dose CT reconstruction. In: 2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018), pp. 1521–1524 (2018)
33. 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)
34. Pan, J., Zhang, H., Wu, W., Gao, Z., Wu, W.: Multi-domain integrative Swin transformer network for sparse-view tomographic reconstruction. Patterns **3**(6) (2022)
35. 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 **40**(11), 3002–3014 (2021)
36. Li, Y., Sun, X., Wang, S., Li, X., Qin, Y., Pan, J., Chen, P.: MDST: multi-domain sparse-view CT reconstruction based on convolution and swin transformer. Phys. Med. Biol. **68**(9), 095019 (2023)
37. Van der Rauwelaert, J., Bossuyt, C., Sijbers, J.: Dual domain Swin transformer based reconstruction method for sparse-view computed tomography. e-Journal of Nondestructive Testing **30**(2) (2025). Presented at the 14th Conference on Industrial Computed Tomography (iCT), 4–7 February 2025, Antwerp, Belgium
38. Liang, J., Cao, J., Sun, G., Zhang, K., Van Gool, L., Timofte, R.: SwinIR: Image restoration using swin transformer. In: Proceedings of the IEEE/CVF International Conference on Computer Vision, pp. 1833–1844 (2021)
39. Brody, S., Alon, U., Yahav, E.: On the expressivity role of LayerNorm in transformers' attention. arXiv preprint arXiv:2305.02582 (2023)
40. Feldkamp, L.A., Davis, L.C., Kress, J.W.: Practical cone-beam algorithm. J. Opt. Soc. Am. A **1**(6), 612–619 (1984)
41. Van Aarle, W., Palenstijn, W.J., Cant, J., Janssens, E., Bleichrodt, F., Dabravolski, A., De Beenhouwer, J., Batenburg, K.J., Sijbers, J.: Fast and flexible X-ray tomography using the ASTRA toolbox. Opt. Express **24**(22), 25129–25147 (2016)
42. McCollough, C.: 2016 low-dose CT grand challenge (2022). https://doi.org/10.21227/4yqw-2364
43. Kingma, D., Ba, J.: Adam: A method for stochastic optimization. In: International Conference on Learning Representations, San Diego, USA (2015)
44. Cao, N., Li, Q., Sun, K., Zhang, H., Ding, J., Wang, Z., Chen, W., Gao, L., Sun, J., Xie, K., Ni, X.: MBST-driven 4D-CBCT reconstruction: Leveraging Swin transformer and masking for robust performance. Comput. Methods Programs Biomed. 108637 (2025)
45. Tang, Y., Yang, D., Li, W., Roth, H.R., Landman, B., Xu, D., Nath, V., Hatamizadeh, A.: Self-supervised pre-training of Swin transformers for 3D medical image analysis. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 20730–20740 (2022)