Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

DSPAN: Dual-domain structure-aware adversarial learning with physics-inspired consistency for low-dose CT denoising

Dayu Xiao

xiaody@bmie.neu.edu.cn

Jiaying University

Wenyong Feng

Jiaying University

Xiaozeng Wang

Guangdong Hilly Agriculture Intelligent Equipment Engineering Technology Research Center

Yan Kang

Shenzhen Technology University

Research Article

Keywords: low-dose CT denoising, generative adversarial network, dual-domain attention, physics-inspired consistency, structural preservation

Posted Date: November 1st, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11173720/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License. Read Full License

Additional Declarations: No competing interests reported.

Abstract

Background

Low-dose computed tomography (LDCT) reduces radiation exposure but increases noise and artifacts that can obscure low-contrast anatomy. We developed the Dual-domain Structure-Preserving Adversarial Network (DSPAN) to improve LDCT image fidelity.

Methods

DSPAN combines a ResNet-based generator with dual-domain structure-aware attention and coupled gradient- and frequency-consistency learning. We used five-fold patient-level cross-validation on all 10 patients in the 2016 NIH-AAPM-Mayo Clinic Low Dose CT Grand Challenge dataset. Each fold included seven training patients, one validation patient for checkpoint selection, and two test patients. Every patient was tested once. Slice-level measurements were averaged within each patient, and performance was summarized descriptively across the 10 patients. No null-hypothesis tests were performed.

Results

DSPAN achieved a mean peak signal-to-noise ratio of 40.2105 ± 1.6183 dB, structural similarity of 0.9424 ± 0.0220, and root mean squared error of 2.5359 ± 0.5303 in 8-bit-equivalent units. It had the best mean value for all six evaluated image-quality metrics. Compared with DU-GAN, mean peak signal-to-noise ratio increased by 0.1505 dB and root mean squared error decreased by 0.0410. Module ablation showed a 1.0662 dB mean peak signal-to-noise ratio gain over ResNetGAN alone.

Conclusions

Combining structure-aware attention with gradient and frequency consistency improved internal image-quality performance in this 10-patient dataset. External validation and diagnostic reader studies are needed to establish generalizability and clinical relevance.

Background

Computed tomography (CT) supports clinical diagnosis through high spatial resolution, rapid acquisition, and broad anatomical coverage. Its use nevertheless exposes patients to ionizing radiation. Under the "as low as reasonably achievable" (ALARA) principle, dose reduction lowers photon counts and increases image noise. The resulting degradation can weaken diagnostic confidence and affect clinical decisions [1–3]. Low-dose CT (LDCT) limits radiation exposure, but its images often contain quantum noise, streak artifacts, and structural degradation. These effects are particularly evident in low-contrast regions.

Early LDCT restoration methods used conventional image processing and statistical reconstruction. Iterative reconstruction and regularization methods, including total variation (TV) minimization, can suppress noise while preserving major anatomical structures[1]. Non-local means (NLM) and block-matching filters use self-similarity priors within CT images to reduce noise[2–4]. These approaches depend on manually designed priors and iterative optimization, which can be computationally expensive and difficult to adapt to different noise distributions. Excessive regularization can also oversmooth subtle anatomical details.

Deep learning enables data-driven feature representation and nonlinear mapping between noisy and clean medical images. Convolutional neural networks (CNNs) learn image priors from data and have been applied to several inverse imaging problems[5, 6]. Residual architectures are commonly used for LDCT denoising. Chen et al. proposed a residual encoder-decoder convolutional neural network (REDCNN) that learned noise-removal mappings while retaining image detail[7]. Residual dense networks subsequently used hierarchical feature aggregation to improve feature reuse and restoration[8]. CNN-based methods can improve objective image-quality metrics, but pixel-wise objectives such as mean squared error (MSE) or L1 loss may produce smooth outputs with limited recovery of high-frequency anatomy. Complementary CNN-based LDCT studies have investigated three-dimensional transfer learning, wavelet-framelet priors, structure-sensitive multi-scale learning, and edge-enhanced dense connectivity [9–12].

Generative adversarial networks (GANs) have been applied to medical image restoration, where a generator and discriminator compete to improve perceptual realism [13]. Yang et al. combined Wasserstein distance with perceptual loss for LDCT denoising and reported better visual quality and structural restoration than conventional CNN models [14]. Later studies examined different adversarial strategies, perceptual constraints, and architectures [15]. In medical images, however, perceptual realism does not guarantee pixel-level fidelity. A GAN may introduce artificial textures or hallucinated structures that alter clinical interpretation, which makes explicit structural constraints

important. Previous LDCT models have used adversarial noise-distribution matching, sharpness-aware learning, dual-domain discrimination, artifact/detail attention, and high-frequency-sensitive objectives [16–20].

Transformer and state-space models extend LDCT denoising beyond local convolutional features. Swin Transformer denoisers can represent long-range dependencies [21], whereas hybrid convolutional state-space models such as CT-Mamba efficiently model global context [22]. Domain adaptation and self-supervised learning have also been explored when paired data are limited [23–25]. Most existing methods still emphasize spatial features and make less explicit use of gradients and frequency spectra. Joint modeling of these signals may help preserve anatomical boundaries under frequency-dependent noise. Recent work includes token-based and hybrid transformer architectures [26–29], as well as CoreDiff and general image-to-image diffusion models [30, 31]. Benchmarking studies and systematic reviews also call for standardized evaluation and perceptual-quality assessment [32, 33].

Anatomical structure appears in image gradients, edges, and high-frequency components. Gradients describe local boundaries and intensity transitions, whereas frequency representations summarize global texture and noise distributions. Spatial and frequency information have previously been used as complementary constraints in image restoration [34, 35]. Their explicit interaction remains less studied in LDCT denoising, particularly within structure-aware attention. Matching the gradients and spectra of restored and reference images may also discourage unrealistic structures during optimization. We describe these constraints as physics-inspired because they reflect image structure, not because they model CT acquisition physics. They do not represent the forward operator, projection statistics, or data-acquisition process. Channel and spatial attention [36, 37] and frequency-aware reconstruction objectives [38] provide the methodological basis for this combination.

On this basis, we developed DSPAN for LDCT denoising. Its DSAM component incorporates spatial-gradient and frequency-domain descriptors into feature attention. SPPCL adds corresponding gradient- and frequency-consistency terms to the generator objective. Together, the components aim to suppress noise while retaining anatomical structure.

This study makes four contributions:

1. We developed DSPAN, a dual-domain adversarial framework that combines structure-aware feature modulation with coupled gradient and frequency consistency constraints.
2. We designed DSAM to fuse a fixed Sobel-gradient descriptor and a normalized Fourier-magnitude descriptor with the original intermediate feature map. Channel-wise attention then weights structure-relevant features while retaining a residual feature path.
3. We formulated SPPCL as a coupled regularizer that measures first-order image-gradient and Fourier-magnitude differences between denoised and normal-dose images.
4. We compared DSPAN with a classical regularizer and three learned denoisers, then used module- and component-level ablations to examine the contributions of DSAM, SPPCL, and their subcomponents.

Methods

Overview of the DSPAN framework

DSPAN uses an adversarial architecture in which a ResNet-based generator denoises LDCT images and a discriminator compares its outputs with normal-dose references (Fig. 1). DSAM operates on intermediate generator features. SPPCL contributes only to training and does not add inference-time parameters.

Given an LDCT image x_{ld}, the generator $G(\bullet)$ aims to produce the denoised CT image:

$$\widehat{x} = G\left(x_{ld}\right)$$

1

where $\widehat{x}$ represents the restored image. The discriminator $D(\bullet)$ distinguishes between the generated image and the corresponding normal-dose CT image x_{nd}, encouraging the generator to synthesize perceptually realistic structures.

Two components extend the underlying ResNetGAN. DSAM constructs channel-wise attention from the original feature tensor, spatial gradients, and normalized Fourier magnitude. SPPCL measures first-order gradient and Fourier-magnitude differences between generated and normal-dose images.

The complete optimization objective is formulated as:

$$\mathscr{L}_{total} = \mathscr{L}_{content} + \lambda_{adv}\mathscr{L}_{adv} + \lambda_{phy}\mathscr{L}_{SPPCL}$$

2

where $\mathscr{L}_{content}$ represents the reconstruction fidelity loss, $\mathscr{L}_{adv}$ denotes adversarial loss, and L

$\mathscr{L}_{SPPCL}$ represents the proposed physics-inspired consistency constraint.

ResNet-based generator and discriminator

DSPAN uses a residual convolutional generator. Residual connections ease optimization and retain low-level image information during restoration. The input LDCT image is first mapped into a high-dimensional feature space using an initial convolutional layer:

$$F_0 = \phi_0 (x_{ld})$$

3

where $\phi_0(\bullet)$ denotes the initial convolution operation.

The feature representation is subsequently processed by a series of residual blocks:

$$F_i = R_i (F_{i-1}), \ i = 1, \ldots, N$$

4

where R_i denotes the $i\text{-}th$ residual block and N represents the number of residual units.

The generator contained 16 residual blocks with 64 feature channels. Each block used two 3 × 3 convolutions, with batch normalization and PReLU after the first convolution, followed by an identity shortcut. The first and final convolutions used 9 × 9 kernels. The network processed single-channel 2D CT slices at scaling factor 1 and therefore performed denoising without spatial upsampling.

After residual feature extraction, DSAM is inserted into the intermediate feature space:

$$F_{att} = DSAM (F_i)$$

5

The enhanced feature is then reconstructed through convolutional layers to obtain the final denoised CT image:

$$\widehat{x} = G (F_{att})$$

6

The final convolutional layers preserve the CT intensity range while reconstructing structural detail.

Dual-domain structure-aware attention module

DSAM constructs an attention tensor from an intermediate feature map and two single-channel descriptors (Fig. 2). Both descriptors are derived from the channel-wise mean feature. One represents local gradient magnitude, and the other represents normalized Fourier magnitude.

Motivation

Standard convolutions capture local spatial features but do not explicitly encode anatomical boundaries or spectral content. Because CT structure is expressed in both gradients and frequency distributions, DSAM uses both descriptors as complementary inputs.

Given an intermediate feature map:

$$F \in R^{C \times H \times W}$$

7

$$F_m = \frac{1}{C} \sum_{c=1}^{C} F_c$$

8

the channel-wise average feature is first calculated.

The resulting feature map is used to generate two structural descriptors, gradient-domain representation and frequency-domain representation.

Gradient Structure Branch

The gradient branch used fixed 3 × 3 Sobel kernels to extract horizontal and vertical responses from the channel-wise mean feature (Fig. 2).

$$G_x = S_x * F_m$$

9

$$G_y = S_y * F_m$$

10

The gradient magnitude is calculated as:

$$G = \sqrt{G_x^2 + G_y^2 + \epsilon}$$

11

where S_x and S_y represent Sobel kernels.

The Sobel operators are fixed convolution kernels. Horizontal and vertical responses are combined to form the gradient map. The gradient descriptor gives the attention mechanism explicit information about boundaries, which may help limit oversmoothing.

Frequency-domain Branch

The frequency branch used a two-dimensional fast Fourier transform to describe global spectral content (Fig. 2).

The Fourier transform of the feature representation is computed as:

$$F_f = |\mathscr{F}(F_m)|$$

12

where $\mathscr{F}(\bullet)$ denotes the two-dimensional Fourier transform.

The normalized frequency magnitude is obtained by:

$$\widehat{F}_f = \frac{F_f}{max\left(F_f\right) + \epsilon}$$

13

The normalized spectrum gives the network information that may help distinguish noise-related frequencies from anatomically relevant high-frequency content. Before feature fusion, the Fourier-magnitude spectrum was normalized by dividing it by its maximum value plus ϵ, where $\epsilon = 10^{-6}$ was added to prevent division by zero.

Attention Generation and Feature Enhancement

The original feature tensor, gradient-magnitude map, and normalized Fourier-magnitude map were concatenated along the channel dimension. A 3 × 3 convolution, ReLU, 1 × 1 convolution, and sigmoid activation produced a channel-wise attention tensor.

$$F_{cat} = \left[F, G, \widetilde{F}_f\right]$$

14

An attention map is generated through convolutional transformation:

$$A = \sigma\ (Conv(\ F_{cat}\ ))$$

15

where $\sigma\ (\bullet\ )$ represents the sigmoid activation function.

Finally, the refined feature is obtained through residual attention modulation:

$$F_{out} = F + A \bigodot F$$

16

where $\bigodot$ denotes element-wise multiplication.

The DSAM implementation follows this residual attention formulation, producing enhanced features through:

Structure-preserving physics-inspired consistency learning

SPPCL added gradient and frequency consistency terms to the generator objective (Fig. 3). These terms were evaluated only during training.

Motivation

Adversarial restoration can improve perceptual quality while introducing structures that are plausible but inconsistent with the paired CT reference. Medical denoising therefore requires explicit constraints on anatomical fidelity.

SPPCL imposes two complementary constraints gradient consistency loss and frequency consistency loss.

The overall physics-inspired consistency loss is defined as:

$$\mathscr{L}_{SPPCL} = \lambda_{\ g}\mathscr{L}_{grad} + \lambda_{\ f}\mathscr{L}_{freq}$$

18

Gradient Consistency Constraint

We converted the generated and normal-dose images into horizontal and vertical first-order difference maps (Fig. 3). The gradient term sums the L1 discrepancies between the corresponding maps.

For generated image $\widehat{x}$:

$$\nabla_{\ x}\widehat{x} = \widehat{x}(i+1, j) - \widehat{x}(i, j)$$

19

$$\nabla_{\ y}\widehat{x} = \widehat{x}(i, j+1) - \widehat{x}(i, j)$$

20

Similarly, gradients are calculated for reference CT:

$$\nabla\ x_{nd}$$

21

The gradient loss is defined as:

$$\mathscr{L}_{grad} = \|\nabla\ \widehat{x} - \nabla\ x_{nd}\|_1$$

22

In the implementation, the gradient loss was the summed L1 discrepancy between corresponding horizontal and vertical difference maps. This term penalizes boundary loss and edge distortion.

Frequency Consistency Constraint

We applied a two-dimensional Fourier transform to the generated and normal-dose images (Fig. 3). The frequency term compares their unnormalized Fourier-magnitude spectra using an L1 loss.

The Fourier magnitude representations are:

$$M_{\hat{x}} = |\mathscr{F}(\hat{x})|$$

23

$$M_{nd} = |\mathscr{F}(x_{nd})|$$

24

The frequency loss is formulated as:

$$\mathscr{L}_{freq} = \|M_{\hat{x}} - M_{nd}\|_1$$

25

For the spectral constraint, the implementation computes the L1 discrepancy between the unnormalized Fourier-magnitude spectra of generated and reference images.

Optimization objective

The final generator objective combines content reconstruction, adversarial learning, and SPPCL:

$$\mathscr{L}_G = \alpha\, \mathscr{L}_{content} + \beta\, \mathscr{L}_{adv} + \gamma\, \mathscr{L}_{SPPCL}$$

26

where α controls reconstruction fidelity, β balances adversarial learning, and γ controls the combined structure-preserving consistency constraint.

In our experiments, the gradient and frequency terms within SPPCL were weighted by α = 1.0 and β = 0.2, respectively, and the combined consistency loss was weighted by γ = 0.05 in the generator objective.

The adversarial loss follows the standard GAN formulation:

$$\mathscr{L}_{adv} = -E\left[log\left(D(\,G\left(x_{ld}\right)\,)\right)\right]$$

27

The discriminator used eight 3 × 3 convolutional blocks with alternating stride-1 and stride-2 operations, LeakyReLU activations, and batch normalization except in the first block. Adaptive average pooling to 6 × 6 was followed by a 1,024-unit fully connected layer and a scalar logit. Binary cross-entropy with logits was used for adversarial training.

Experimental setup

Dataset and study design

Experiments ran on Ubuntu with Python 3.11, PyTorch 2.6.0, torchvision 0.21.0, and the CUDA 12.4 runtime reported by PyTorch. Training was conducted on one NVIDIA RTX 5880 Ada GPU with 48 GB of memory.

We used the de-identified 2016 NIH-AAPM-Mayo Clinic Low Dose CT Grand Challenge dataset, which contains paired normal-dose and quarter-dose CT images from 10 patients at 1-mm and 3-mm slice thicknesses. Only the 1-mm reconstructions were included. We used a fixed five-fold patient-level design with seven training patients, one validation patient, and two test patients per fold. The assignments were:

fold 0, test L067/L096 and validation L109; fold 1, test L109/L143 and validation L192; fold 2, test L192/L286 and validation L291; fold 3, test L291/L310 and validation L333; and fold 4, test L333/L506 and validation L067. The remaining seven patients formed the training set. Thus, every patient served as a test patient exactly once. Training samples were random 64 × 64 paired patches; candidate patches with mean normalized intensity below − 0.9 in either image were rejected as background. No patient or slice appeared in more than one partition within a fold. The dataset and its paired image/projection resources have been described in detail by Moen et al.[39].

This study was conducted in accordance with the Declaration of Helsinki (World Medical Association, 1964, as amended) and complied with the national and institutional ethical standards governing life-science and medical research involving human data in China, namely the Measures for Ethical Review of Life Sciences and Medical Research Involving Humans (Guowei Kejiao Fa [2023] No. 4) and the Charter of the Science and Technology Ethics Committee of Jiaying University (Jia Yuan [2023] No. 75). This study was reviewed and approved by the Medical Research Ethics Committee of Jiaying University (approval no. JYYXLL2026-24). The source dataset was publicly released in de-identified form. This study was a secondary computational analysis of existing data and involved no participant recruitment, intervention, direct participant contact, or access to identifiable private information. Under the ethics review documentation approved by the Medical Research Ethics Committee of Jiaying University, additional informed consent to participate was not required for this secondary computational analysis of publicly available, de-identified data. Accordingly, no additional informed consent to participate was obtained for the present study. The original data collection and public release were governed by the ethical and data-sharing procedures of the source dataset.

DSPAN training and implementation

The generator used 16 residual blocks and 64 feature channels. A separate model was trained from scratch for each fold using seed 392 plus the fold index. SRResNet pretraining ran for up to 130 epochs and initialized adversarial training; DSPAN then ran for up to 300 epochs. Both stages were evaluated on the fold-specific validation patient every 10 epochs, and the checkpoint with the highest mean validation-patient PSNR was retained without consulting the test patients. Training used Adam optimizers, batch size 8, and an initial learning rate of 1 × 10 − 4; the adversarial-stage learning rate was reduced tenfold halfway through training. The generator objective combined VGG19 feature-space mean-squared error (weight 1), binary cross-entropy adversarial loss (weight 0.01), and SPPCL (weight 0.05). Within SPPCL, gradient and frequency terms were weighted 1.0 and 0.2. Input arrays were normalized to [− 1, 1]. The feature-space and adversarial terms followed established perceptual-loss and super-resolution GAN formulations [40, 41].

Comparison methods

We compared DSPAN with the quarter-dose input, total variation (TV) denoising, REDCNN, EDCNN, and DU-GAN; the paired normal-dose image served as the reference. All learned methods used exactly the same five patient-level folds, preprocessing, patch sampling rule, and evaluation window as DSPAN. REDCNN, EDCNN, and DU-GAN were retrained independently on the seven training patients in every fold, and their best checkpoints were selected by the same validation-patient PSNR rule. REDCNN used the original 96-channel residual encoder-decoder, MSE loss, Adam, a 1 × 10 − 5 initial learning rate, and 300 epochs. EDCNN used its edge-enhanced architecture and compound objective for 300 epochs. DU-GAN used image- and gradient-domain discriminators and exponential moving-average generator weights for 80 epochs. TV required no training and was applied to the same held-out test images. Baseline implementation choices followed the original reports and contemporary LDCT-denoising literature [7, 12, 18, 32, 33].

Evaluation metrics and statistical reporting

Image fidelity was summarized using peak signal-to-noise ratio (PSNR), structural similarity index (SSIM), root mean squared error (RMSE), feature similarity index (FSIM), visual information fidelity (VIF), and noise quality measure (NQM). For all six metrics, reconstructed and reference images were clipped to the prespecified [− 1000, 1000] HU interval and linearly mapped to [0, 255]. PSNR used a fixed data range of 255; RMSE is therefore reported in 8-bit-equivalent units. SSIM was computed in MATLAB R2025a with DynamicRange set to 255 [42]. FSIM used the phase-congruency and gradient-magnitude formulation; VIF used the wavelet-domain vifvec implementation; and NQM used a viewing distance of 3.5 image heights (16.3702° visual angle) according to their original formulations [43–45]. Slice-level metrics were averaged within each patient and method before group summaries; the patient was the independent unit (n = 10). Table 1 reports the mean and standard deviation of the 10 patient-level means. No slice-level or patient-level null-hypothesis tests were performed. Between-method comparisons are reported as descriptive differences in patient-level means.

Convergence and Fourier-magnitude analyses

Training loss and validation PSNR were recorded for each learned method. Validation PSNR was evaluated every 10 epochs and was used only to select the best checkpoint within each fold. Figure 7 displays representative training-loss trajectories for the four learned comparison methods. Because the objectives and numerical scales differed, the raw loss magnitudes were not compared across algorithms.

A descriptive Fourier-magnitude analysis was conducted on the four representative L506 slices shown in Figs. 4 and 5. Each reconstruction and its paired normal-dose reference were multiplied by the same two-dimensional Hann window before an unnormalized two-dimensional Fourier transform. Absolute Fourier-magnitude differences were radially averaged from 0 to 0.5 cycles per pixel. Relative spectral errors were calculated in low (0–0.15), mid (0.15–0.30), and high (0.30–0.50 cycles per pixel) bands and averaged across the four selected slices.

Use of generative AI tools

ChatGPT (OpenAI) was used solely for English-language polishing, grammar checking, and related improvements to readability. The authors reviewed and approved all suggested language edits. ChatGPT was not used to generate or analyze data, produce figures, perform statistical analyses, or formulate the scientific conclusions. The authors take full responsibility for the accuracy and integrity of the manuscript.

Results

Quantitative comparison

Table 1 reports the patient-level comparison after each of the 10 patients had served once in the test set. DSPAN achieved 40.2105 ± 1.6183 dB PSNR, 0.9424 ± 0.0220 SSIM, and 2.5359 ± 0.5303 RMSE. Its FSIM, VIF, and NQM values were 0.9871 ± 0.0040, 0.6410 ± 0.0525, and 35.3497 ± 1.5460 dB, respectively. These means were the best among the evaluated methods for all six metrics. Against DU-GAN, the closest comparator for PSNR, RMSE, FSIM, and NQM, DSPAN increased PSNR by 0.1505 dB, FSIM by 0.0009, and NQM by 0.4797 dB. It also reduced mean RMSE by 0.0410. EDCNN had the highest non-DSPAN mean SSIM and VIF; DSPAN exceeded those values by 0.0011 and 0.0083, respectively. All differences are descriptive comparisons of patient-level means, and no inferential p values are reported.

Table 1
Patient-level image-quality comparison across 10 patients.

Method	PSNR (dB) ↑	SSIM ↑	RMSE ↓	FSIM ↑	VIF ↑	NQM (dB) ↑
Quarter-dose CT	**36.3942 ± 1.8632**	**0.8601 ± 0.0495**	**3.9603 ± 0.9727**	**0.9712 ± 0.0097**	**0.6246 ± 0.0609**	**33.7328 ± 1.9419**
TV	38.8479 ± 1.9257	0.9208 ± 0.0396	2.9914 ± 0.7690	0.9809 ± 0.0072	0.5864 ± 0.0482	33.7207 ± 1.8300
REDCNN	39.2585 ± 1.3856	0.9342 ± 0.0229	2.8148 ± 0.4909	0.9827 ± 0.0045	0.5686 ± 0.0454	34.0063 ± 1.6458
EDCNN	40.0381 ± 1.5918	0.9413 ± 0.0247	2.5851 ± 0.5307	0.9860 ± 0.0044	0.6327 ± 0.0514	34.7036 ± 1.5518
DU-GAN	40.0600 ± 1.5635	0.9412 ± 0.0232	2.5769 ± 0.5193	0.9862 ± 0.0043	0.6305 ± 0.0525	34.8700 ± 1.6336
DSPAN	40.2105 ± 1.6183	0.9424 ± 0.0220	2.5359 ± 0.5303	0.9871 ± 0.0040	0.6410 ± 0.0525	35.3497 ± 1.5460

Five-fold patient-level quantitative comparison across 10 patients (mean ± standard deviation of patient-level means). All images were clipped to [− 1000, 1000] HU and mapped to [0, 255]; RMSE is reported in 8-bit-equivalent units. Each patient served as a test patient once.

Table 2 reports four illustrative slices from patient L506, which was held out in fold 4. DSPAN had the highest PSNR and lowest RMSE for all four slices. Its SSIM was highest or tied for highest for slices 148 and 329, whereas EDCNN had slightly higher SSIM for slices 178 and 401. These correlated slice-level examples are descriptive and were not treated as independent observations for statistical inference.

Table 2
Image-quality metrics for four representative L506 test slices.

	Slice 1			Slice 2			Slice 3			Slice 4		
	PSNR	**SSIM**	**RMSE**	**PSNR**	**SSIM**	**RMSE**	**PSNR**	**SSIM**	**RMSE**	**PSNR**	**SSIM**	**RMSE**
Quarter dose	36.8729	0.8918	3.6551	36.8956	0.8874	3.6455	31.9571	0.7396	6.4370	38.0640	0.8901	3.1867
TV	39.1416	0.9416	2.8149	39.0861	0.9395	2.8329	34.1390	0.8189	5.0072	40.4972	0.9423	2.4081
REDCNN	39.3357	0.9439	2.7527	38.8073	0.9410	2.9253	35.8915	0.8746	4.0923	40.2596	0.9468	2.4749
EDCNN	40.2302	0.9536	2.4833	39.9656	0.9494	2.5601	36.2864	0.8783	3.9104	41.4897	0.9580	2.1481
DU-GAN	40.2369	0.9526	2.4814	39.9595	0.9482	2.5619	36.3313	0.8814	3.8902	41.4998	0.9570	2.1456
DSPAN	40.4193	0.9536	2.4298	40.0946	0.9487	2.5224	36.3838	0.8856	3.8668	41.6868	0.9578	2.0999

Descriptive results for four representative L506 test slices generated by the fold-4 checkpoints under the same standardized evaluation protocol as Table 1.

Qualitative comparison

Figures 4 and 5 present fold-4 outputs for L506 slices 148, 178, 329, and 401. From top to bottom, the rows show quarter-dose CT, TV, REDCNN, EDCNN, DU-GAN, DSPAN, and normal-dose CT. The red boxes in Fig. 4 identify the regions enlarged in Fig. 5.

Across the four selected slices, all denoising methods reduced visible granular noise relative to the quarter-dose inputs. TV and REDCNN produced smoother textures, whereas EDCNN, DU-GAN, and DSPAN retained sharper small structures and tissue boundaries. DSPAN outputs were visually close to the normal-dose references in the marked regions, although residual texture and local intensity differences remained.

The enlarged regions show that DSPAN suppressed granular noise while retaining the marked low-contrast boundaries and small high-contrast structures. Some residual smoothing and local intensity differences remained relative to normal-dose CT.

Module-level ablation

Table 3 summarizes the patient-level module ablation under the same five-fold assignments as the main experiment. Relative to ResNetGAN, adding DSAM increased mean PSNR from 39.1443 to 39.8540 dB, increased SSIM from 0.9322 to 0.9386, and reduced RMSE from 2.8555 to 2.6359. Adding SPPCL increased mean PSNR to 39.8107 dB and SSIM to 0.9363 and reduced RMSE to 2.6500. The complete DSPAN model achieved 40.2105 dB PSNR, 0.9424 SSIM, and 2.5359 RMSE. Thus, DSPAN improved mean PSNR by 1.0662 dB over ResNetGAN, by 0.3565 dB over ResNetGAN + DSAM, and by 0.3998 dB over ResNetGAN + SPPCL. Figure 6 complements these quantitative comparisons with a qualitative module-level ablation, showing corresponding quarter-dose CT, ResNetGAN, ResNetGAN + DSAM, ResNetGAN + SPPCL, DSPAN, and normal-dose CT images from one prespecified held-out fold.

Table 3
Patient-level module ablation of DSAM and SPPCL.

Model	DSAM	SPPCL	PSNR	SSIM	RMSE
ResNetGAN	**—**	**—**	**39.1443**	**0.9322**	**2.8555**
ResNetGAN + DSAM	✓	—	39.8540	0.9386	2.6359
ResNetGAN + SPPCL	—	✓	39.8107	0.9363	2.6500
DSPAN	✓	✓	40.2105	0.9424	2.5359

Five-fold patient-level module ablation of DSAM and SPPCL across 10 test patients.

Component-level ablation

Table 4 evaluates the spatial-gradient and frequency branches of DSAM with SPPCL enabled in every variant. The no-branch model achieved 39.8107 dB PSNR, 0.9363 SSIM, and 2.6500 RMSE. The gradient-only variant improved these means to 39.9572 dB, 0.9402, and 2.6058, respectively. The frequency-only variant increased PSNR to 39.9044 dB and reduced RMSE to 2.6216, but its SSIM decreased to 0.9334. Using both branches produced the strongest aggregate results: 40.2105 dB PSNR, 0.9424 SSIM, and 2.5359 RMSE.

Table 4
Patient-level component ablation of the DSAM branches.

Model	Gradient branch	Frequency branch	PSNR	SSIM	RMSE
ResNetGAN + SPPCL	**—**	**—**	**39.8107**	**0.9363**	**2.6500**
+ Gradient	✓	—	39.9572	0.9402	2.6058
+ Frequency	—	✓	39.9044	0.9334	2.6216
DSPAN	✓	✓	40.2105	0.9424	2.5359

Five-fold patient-level component ablation of the DSAM branches.

Table 5 evaluates the gradient and frequency constraints of SPPCL with DSAM enabled in every variant. The no-constraint model achieved 39.8540 dB PSNR, 0.9386 SSIM, and 2.6359 RMSE. The gradient-only variant decreased PSNR and SSIM to 38.7952 dB and 0.9302 and increased RMSE to 2.9604. The frequency-only variant produced similar but slightly lower aggregate performance than the no-constraint

model, with 39.8473 dB PSNR, 0.9356 SSIM, and 2.6396 RMSE. In contrast, the combined constraints achieved 40.2105 dB PSNR, 0.9424 SSIM, and 2.5359 RMSE, the strongest mean values in this ablation.

Table 5
Patient-level component ablation of the SPPCL constraints.

Model	Gradient loss	Frequency loss	PSNR	SSIM	RMSE
ResNetGAN + DSAM	**—**	**—**	**39.8540**	**0.9386**	**2.6359**
+ Gradient	✓	—	38.7952	0.9302	2.9604
+ Frequency	—	✓	39.8473	0.9356	2.6396
DSPAN	✓	✓	40.2105	0.9424	2.5359

Five-fold patient-level component ablation of the SPPCL constraints.

Model complexity

Table 6 reports inference-time model size and computational complexity, which do not vary across cross-validation folds. DSAM increased the generator parameter count from 1.233 M to 1.276 M (3.5%) and FLOPs from 5.064 G to 5.237 G (3.4%). SPPCL is used only during training and therefore added no inference-time parameters or FLOPs. Relative to ResNetGAN, ResNetGAN + DSAM increased mean PSNR by 0.7097 dB, ResNetGAN + SPPCL increased it by 0.6664 dB, and the complete DSPAN model increased it by 1.0662 dB.

Table 6
Model complexity and patient-level performance.

Method	Parameters (M)	FLOPs (G)	PSNR (dB)
ResNetGAN	**1.233**	**5.064**	**39.1443**
ResNetGAN + DSAM	1.276	5.237	39.8540
ResNetGAN + SPPCL	1.233	5.064	39.8107
DSPAN	1.276	5.237	40.2105

Model complexity and five-fold patient-level performance of the module-level variants.

Training-loss convergence

Figure 7 presents representative training-loss trajectories for REDCNN, EDCNN, DU-GAN, and DSPAN. REDCNN, EDCNN, and DU-GAN showed pronounced downward trends over the displayed runs, whereas the DSPAN objective retained greater short-term fluctuation. These curves describe optimization behavior only. Because the methods optimize different objectives on different numerical scales, their raw loss values cannot be used to rank denoising performance. Model checkpoints were selected according to validation-patient PSNR.

Fourier-magnitude analysis

Figure 8 summarizes the descriptive Fourier-magnitude analysis of the four representative L506 slices. DSPAN produced mean relative spectral errors of 6.81%, 28.37%, and 47.82% in the low-, mid-, and high-frequency bands, respectively. DU-GAN had slightly lower low- and mid-frequency errors (6.75% and 28.26%), whereas DSPAN had the lowest high-frequency error among the evaluated methods.

Discussion

Across the patient-level analysis, DSPAN ranked first on all six image-quality metrics. It yielded the highest PSNR, SSIM, FSIM, VIF, and NQM and the lowest RMSE. Agreement across these measures reduces the likelihood that the ranking arose from the emphasis of a single metric. Yet the differences from the strongest comparators were small for PSNR, SSIM, RMSE, and FSIM. NQM and VIF showed larger separations, which may indicate better perceptually weighted fidelity. Both remain image-based surrogates, however, so the study establishes improved internal reconstruction quality rather than clinical benefit.

How structural information enters the network distinguishes DSPAN from related LDCT denoisers. EDCNN strengthens edge features, and DU-GAN uses separate image- and gradient-domain discriminators [12, 18]. Other adversarial models apply artifact/detail attention or high-frequency-sensitive objectives [19, 20]. DSPAN instead uses fixed gradient and Fourier descriptors to modulate intermediate features, then constrains reconstructed images with corresponding consistency terms. Spatial and frequency information therefore affect both feature

selection and output regularization. Matched training and evaluation show how this design behaves, but the experiments do not establish superiority over every possible dual-domain implementation.

The DSAM ablation indicates that its gradient and Fourier descriptors were most useful when combined. With SPPCL held constant, the dual-branch model outperformed both single-branch variants in PSNR, SSIM, and RMSE. The frequency-only branch improved PSNR and RMSE but reduced SSIM relative to the no-branch model. One plausible explanation is that spectral information reduces global error but cannot localize anatomical boundaries as precisely as a gradient descriptor. Combining both branches may provide global frequency context together with spatially localized edge information. This explanation remains tentative because the study did not compare alternative descriptors or capacity-matched controls, and the ablation reports aggregate means.

SPPCL showed a different dependence between its two constraints. Neither isolated constraint improved PSNR, SSIM, and RMSE simultaneously relative to the no-constraint model. The coupled loss increased mean PSNR by 0.3565 dB and SSIM by 0.0038 while reducing RMSE by 0.1000. This result is compatible with complementary signals: gradient loss acts on local intensity transitions, whereas frequency loss constrains global spectral content. The pattern could also arise from the selected weights or from competition among reconstruction, adversarial, and consistency losses. The ablation supports the tested configuration, but it cannot identify an interaction mechanism without factorial or weight-sensitivity experiments.

Optimization curves should be interpreted separately from held-out performance. Figure 7 shows decreasing losses for REDCNN, EDCNN, and DU-GAN, while the DSPAN objective exhibits greater short-term fluctuation. Because these methods optimize different objectives on incompatible scales, their curves cannot rank denoising performance. Stable optimization alone therefore does not demonstrate robust image reconstruction.

The spectral analysis offers a narrower view of the structures retained by each method. DSPAN had the lowest high-frequency error in the four displayed slices, whereas DU-GAN performed slightly better in the low- and mid-frequency bands. Together with the enlarged images, these results are compatible with stronger preservation of fine boundaries, but they do not show uniform spectral superiority. Fourier magnitude lacks spatial location and phase information, so it cannot distinguish anatomical detail from structured noise on its own. PSNR, SSIM, FSIM, VIF, and NQM have a related limitation: correlated full-reference metrics do not measure lesion visibility or diagnostic decisions. Previous benchmarking and perceptual-quality reviews likewise distinguish numerical fidelity from perceived or task-based image quality [32, 33].

The cross-validation design provides broader internal testing than a single train-test split. Every patient contributed one out-of-fold test result, and no patient crossed partitions within a fold. Patient-level averaging also avoids treating correlated slices as independent replicates. The five folds are not independent cohorts, however, because their training sets overlap and the effective sample remains 10 patients. The precision of the small between-method differences is therefore uncertain without paired intervals or prespecified inferential tests. Checkpoint selection was based on PSNR. Agreement with the other metrics is informative, but it does not eliminate a possible selection preference.

Measured architectural overhead was small, although deployment efficiency was not directly tested. DSAM increased generator parameters by 3.5% and FLOPs by 3.4%, while SPPCL added no inference-time operations. These increments may be acceptable in pipelines that already support a ResNetGAN-scale generator. Parameter counts and FLOPs do not capture wall-clock latency, peak memory use, preprocessing, or hardware-dependent throughput. The present evidence therefore supports limited architectural overhead, but not real-time or resource-efficient deployment.

Several limitations restrict the scope of these findings. The experiments included 10 patients from one public dataset, only 1-mm reconstructions, one seed per fold, and paired low- and normal-dose targets. Consequently, scanner, reconstruction-kernel, anatomical, and dose diversity were limited. Independent 2D processing also omits through-plane context and may allow inter-slice inconsistency. The study did not include an external cohort, dose-shift stress testing, repeated seeds, loss-weight sensitivity analysis, radiologist assessment, or a diagnostic task. The next priority is multi-center and cross-protocol validation with paired patient-level uncertainty estimates, repeated training, and systematic failure analysis. Subsequent comparisons with 2.5D or 3D models and blinded task-based reader studies should test whether the measured fidelity gains have diagnostic relevance.

Conclusions

DSPAN combines gradient- and frequency-informed feature attention with coupled consistency learning for adversarial LDCT denoising. In five-fold patient-level cross-validation, it achieved the best mean value for each of the six image-quality metrics. Ablation results supported the combined DSAM branches and coupled SPPCL terms under the tested configuration. These findings indicate improved reconstruction fidelity with modest additional inference complexity in this 10-patient dataset. They do not establish diagnostic benefit or robustness in other

populations or acquisition settings. External multi-center validation, repeated training, sensitivity analysis, and reader- or task-based assessment are needed before clinical relevance can be determined.

Abbreviations

ALARA
as low as reasonably achievable
CNN
convolutional neural network
CT
computed tomography
DSAM
dual-domain structure-aware attention module
DSPAN
Dual-domain Structure-Preserving Adversarial Network
DU-GAN
dual-domain generative adversarial network
EDCNN
edge enhancement-based densely connected network
FLOPs
floating-point operations
FSIM
feature similarity index
GAN
generative adversarial network
HU
Hounsfield units
LDCT
low-dose computed tomography
MSE
mean squared error
NLM
non-local means
NQM
noise quality measure
PReLU
parametric rectified linear unit
PSNR
peak signal-to-noise ratio
REDCNN
residual encoder-decoder convolutional neural network
ReLU
rectified linear unit
RMSE
root mean squared error
SPPCL
structure-preserving physics-inspired consistency learning
SSIM
structural similarity index
TCIA
The Cancer Imaging Archive
TV
total variation
VIF

visual information fidelity.

Declarations

Ethics approval and consent to participate

This study was conducted in accordance with the Declaration of Helsinki (World Medical Association, 1964, as amended) and complied with the national and institutional ethical standards governing life-science and medical research involving human data in China, namely the Measures for Ethical Review of Life Sciences and Medical Research Involving Humans (Guowei Kejiao Fa [2023] No. 4) and the Charter of the Science and Technology Ethics Committee of Jiaying University (Jia Yuan [2023] No. 75). This study was reviewed and approved by the Medical Research Ethics Committee of Jiaying University (approval no. JYYXLL2026-24). The source dataset was publicly released in de-identified form. This study was a secondary computational analysis of existing data and involved no participant recruitment, intervention, direct participant contact, or access to identifiable private information. Under the ethics review documentation approved by the Medical Research Ethics Committee of Jiaying University, additional informed consent to participate was not required for this secondary computational analysis of publicly available, de-identified data. Accordingly, no additional informed consent to participate was obtained for the present study. The original data collection and public release were governed by the ethical and data-sharing procedures of the source dataset.

Consent for publication

Not applicable. This study used publicly available, de-identified CT data. No identifiable personal information or identifiable patient images are included in this manuscript.

Availability of data and materials

The de-identified 2016 NIH-AAPM-Mayo Clinic Low Dose CT Grand Challenge dataset is available from The Cancer Imaging Archive (TCIA), https://doi.org/10.7937/9NPB-2637, subject to the repository's access conditions.

Acknowledgements

The authors thank colleagues at Jiaying University for discussions and technical assistance during the experiments. The 2016 NIH-AAPM-Mayo Clinic Low Dose CT Grand Challenge dataset was provided by the Mayo Clinic and managed by NIH, AAPM, and the Mayo Clinic.

Funding

This work was supported by the Talent Scientific Research Initiation Project of Jiaying University (2024RC17).

The funder provided financial support only and had no role in study design, data collection, analysis, interpretation, or manuscript preparation. These activities were performed by the authors.

Submission declaration

This manuscript is not under consideration by any other journal.

Competing interests

The authors declare that they have no competing interests.

Authors' contributions

DX: Conceptualization, Methodology, Software, Validation, Formal analysis, Investigation, Data curation, Writing—Original draft, Visualization.

WF: Methodology, Software, Validation.

XW: Resources, Supervision, Project administration, Funding acquisition.

YK: Supervision.

All authors read, revised, and approved the final manuscript.

References

1. Tian Z, Jia X, Yuan K, Pan T, Jiang SB. Low-dose CT reconstruction via edge-preserving total variation regularization. Phys Med Biol. 2011;56:5949–67.
2. Zhang H, Zeng D, Zhang H, Wang J, Liang Z, Ma J. Applications of nonlocal means algorithm in low-dose X-ray CT image processing and reconstruction: A review. Med Phys. 2017;44:1168–85.
3. Dabov K, Foi A, Katkovnik V, Egiazarian K. Image denoising by sparse 3D transform-domain collaborative filtering. IEEE Trans Image Process. 2007;16:2080–95.
4. Buades A, Coll B, Morel JM. A non-local algorithm for image denoising. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). 2005;2:60–5.
5. Li Y, Sixou B, Peyrin F. A review of deep learning methods for MRI reconstruction. IEEE Trans Med Imag. 2019;38:1355–67.
6. Jin KH, McCann MT, Froustey E, Unser M. Deep convolutional neural network for inverse problems in imaging. IEEE Trans Image Process. 2017;26:4509–22.
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 Imag. 2017;36:2524–35. https://doi.org/10.1109/TMI.2017.2715284.
8. Zhang Y, Tian Y, Kong Y, Zhong B, Fu Y. Residual dense network for image restoration. IEEE Trans Pattern Anal Mach Intell. 2018;41:2750–63.
9. Shan H, Zhang Y, Yang Q, Kruger U, Kalra MK, Sun L, et al. 3-D convolutional encoder-decoder network for low-dose CT via transfer learning from a 2-D trained network. IEEE Trans Med Imaging. 2018;37:1522–34. https://doi.org/10.1109/TMI.2018.2832217.
10. Kang E, Chang W, Yoo J, Ye JC. Deep convolutional framelet denosing for low-dose CT via wavelet residual network. IEEE Trans Med Imaging. 2018;37:1358–69. https://doi.org/10.1109/TMI.2018.2823756.
11. You C, Cong W, Wang G, Yang Q, Shan H, Gjesteby L, et al. Structurally-sensitive multi-scale deep neural network for low-dose CT denoising. IEEE Access. 2018;6:41839–55. https://doi.org/10.1109/ACCESS.2018.2858196.
12. Liang T, Jin Y, Li Y, Wang T, EDCNN. Edge enhancement-based Densely Connected Network with Compound Loss for Low-Dose CT Denoising. 2020. pp. 193–8. https://doi.org/10.1109/ICSP48669.2020.9320928
13. Goodfellow I, Pouget-Abadie J, Mirza M, Xu B, Warde-Farley D, Ozair S et al. Generative adversarial nets. Adv Neural Inform Process Syst (NeurIPS). 2014;:2672–80.
14. Yang Q, Yan P, Zhang Y, Yu H, Shi Y, Mou X, et al. Low-dose CT image denoising using a generative adversarial network with wasserstein distance and perceptual loss. IEEE Trans Med Imag. 2018;37:1348–57.
15. Sarkar K, Bag S, Tripathi PC. Noise aware content-noise complementary GAN with local and global discrimination for low-dose CT denoising. Neurocomputing. 2024;582:127473. https://doi.org/10.1016/j.neucom.2024.127473.
16. Wolterink JM, Leiner T, Viergever MA, Isgum I. Generative adversarial networks for noise reduction in low-dose CT. IEEE Trans Med Imaging. 2017;36:2536–45. https://doi.org/10.1109/TMI.2017.2708987.
17. Yi X, Babyn P. Sharpness-aware low-dose CT denoising using conditional generative adversarial network. J Digit Imaging. 2018;31:655–69. https://doi.org/10.1007/s10278-018-0056-0.
18. Huang Z, Zhang J, Zhang Y, Shan H. DU-GAN: Generative Adversarial Networks With Dual-Domain U-Net-Based Discriminators for Low-Dose CT Denoising. IEEE Trans Instrum Meas. 2022;71:1–12. https://doi.org/10.1109/TIM.2021.3128703.
19. Zhang X, Han Z, Shangguan H, Han X, Cui X, Wang A. Artifact and Detail Attention Generative Adversarial Networks for Low-Dose CT Denoising. IEEE Trans Med Imaging. 2021;40:3901–18. https://doi.org/10.1109/TMI.2021.3101616.
20. Yang L, Shangguan H, Zhang X, Wang A, Han Z. High-Frequency Sensitive Generative Adversarial Network for Low-Dose CT Image Denoising. IEEE Access. 2020;8:930–43. https://doi.org/10.1109/ACCESS.2019.2961983.
21. Zhang B, Zhang Y, Wang B, He X, Zhang F, Zhang X. Denoising swin transformer and perceptual peak signal-to-noise ratio for low-dose CT image denoising. Measurement. 2024;227:114303. https://doi.org/10.1016/j.measurement.2024.114303.
22. Li L, Wei W, Yang L, Zhang W, Dong J, Liu Y, et al. CT-mamba: A hybrid convolutional state space model for low-dose CT denoising. Comput Med Imaging Graphics. 2025;124:102595. https://doi.org/10.1016/j.compmedimag.2025.102595.
23. Wang J, Tang Y, Wu Z, Tsui BMW, Chen W, Yang X, et al. Domain-adaptive denoising network for low-dose CT via noise estimation and transfer learning. Med Phys. 2023;50:74–88. https://doi.org/10.1002/mp.15952.
24. Su B, Hu X, Zha Y, Wu Z, Ma Y, Xu J, et al. Zero-shot low-dose CT image denoising via patch-based content-guided diffusion models. IEEE Trans Instrum Meas. 2025;74:1–12. https://doi.org/10.1109/TIM.2025.3547474.
25. Chu Y, Zhou L, Luo G, Kang K, Dong S, Han Z, et al. HorusEye: A self-supervised foundation model for generalizable X-ray tomography restoration. Nat Comput Sci. 2026;6:372–87. https://doi.org/10.1038/s43588-026-00973-3.

26. Wang D, Fan F, Wu Z, Liu R, Wang F, Yu H. CTformer: convolution-free Token2Token dilated vision transformer for low-dose CT denoising. Phys Med Biol. 2023;68:065012. https://doi.org/10.1088/1361-6560/acc000.
27. Yang L, Li Z, Ge R, Zhao J, Si H, Zhang D. Low-Dose CT Denoising via Sinogram Inner-Structure Transformer. IEEE Trans Med Imaging. 2023;42:910–21. https://doi.org/10.1109/TMI.2022.3219856.
28. Li H, Yang X, Yang S, Wang D, Jeon G. Transformer With Double Enhancement for Low-Dose CT Denoising. IEEE J Biomedical Health Inf. 2023;27:4660–71. https://doi.org/10.1109/JBHI.2022.3216887.
29. Zhang S-Y, Wang Z-X, Yang H-B, Chen Y-L, Li Y, Pan Q, et al. Hformer: highly efficient vision transformer for low-dose CT denoising. Nucl Sci Tech. 2023;34:63. https://doi.org/10.1007/s41365-023-01208-0.
30. Gao Q, Li Z, Zhang J, Zhang Y, Shan H. CoreDiff: Contextual Error-Modulated Generalized Diffusion Model for Low-Dose CT Denoising and Generalization. IEEE Trans Med Imaging. 2024;43:745–59. https://doi.org/10.1109/TMI.2023.3320812.
31. Saharia C, Chan W, Chang H, Lee C, Ho J, Salimans T et al. Palette: Image-to-image diffusion models. 2022. pp. 1–10. https://doi.org/10.1145/3528233.3530757
32. Eulig E, Ommer B, Kachelrieß M. Benchmarking deep learning-based low-dose CT image denoising algorithms. Med Phys. 2024;51:8776–88. https://doi.org/10.1002/mp.17379.
33. Kim W, Jeon S-Y, Byun G, Yoo H, Choi J-H. A systematic review of deep learning-based denoising for low-dose computed tomography from a perceptual quality perspective. Biomed Eng Lett. 2024;14:1153–73. https://doi.org/10.1007/s13534-024-00419-7.
34. Kaji S, Kida S. Overview of image-to-image translation by use of deep neural networks: denoising, super-resolution, modality conversion, and reconstruction in medical imaging. Radiol Phys Technol. 2019;12:235–48. https://doi.org/10.1007/s12194-019-00520-y.
35. Wang Z, Liu D, Yang J, Han W, Huang T. Deep networks for image super-resolution with sparse prior. Proceedings of The IEEE International Conference On Computer Vision (iccv). 2015;:370–8.
36. Woo S, Park J, Lee J-Y, Kweon IS. CBAM: Convolutional Block Attention Module. In: Computer Vision – ECCV 2018. 2018. pp. 3–19. https://doi.org/10.1007/978-3-030-01234-2_1
37. Hu J, Shen L, Sun G, Squeeze-, -Excitation, Networks. 2018. pp. 7132–41. https://doi.org/10.1109/CVPR.2018.00745
38. Jiang L, Dai B, Wu W, Loy CC. Focal Frequency Loss for Image Reconstruction and Synthesis. 2021. pp. 13899–909. https://doi.org/10.1109/ICCV48922.2021.01366
39. Moen TR, Chen B, Holmes DR, Duan X, Yu Z, Yu L, et al. Low-dose CT image and projection dataset. Med Phys. 2021;48:902–11. https://doi.org/10.1002/mp.14594.
40. Johnson J, Alahi A, Fei-Fei L. Perceptual losses for real-time style transfer and super-resolution. European Conference on Computer Vision (ECCV). 2016;:694–711.
41. Ledig C, Theis L, Huszár F, Caballero J, Cunningham A, Acosta A et al. Photo-realistic single image super-resolution using a generative adversarial network. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). 2017;:4681–90.
42. Wang Z, Bovik AC, Sheikh HR, Simoncelli EP. Image quality assessment: from error visibility to structural similarity. IEEE Trans Image Process. 2004;13:600–12. https://doi.org/10.1109/TIP.2003.819861.
43. Zhang L, Zhang L, Mou X, Zhang D. FSIM: A feature similarity index for image quality assessment. IEEE Trans Image Process. 2011;20:2378–86.
44. Sheikh HR, Bovik AC. Image information and visual quality. IEEE Trans Image Process. 2006;15:430–44.
45. Damera-Venkata N, Kite TD, Geisler WS, Evans BL, Bovik AC. Image quality assessment based on a degradation model. IEEE Trans Image Process. 2000;9:636–50. https://doi.org/10.1109/83.841940.

Figures

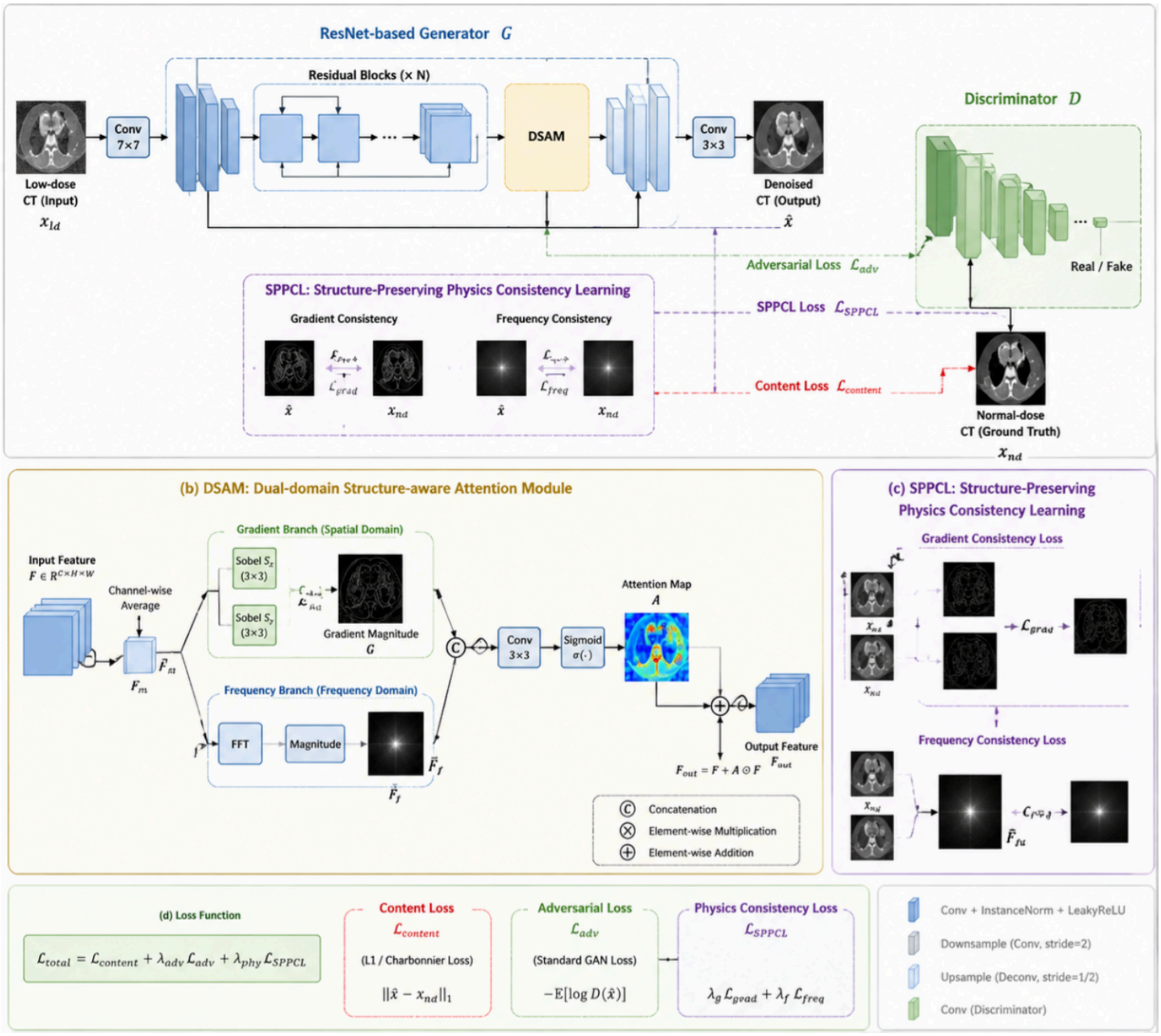


Figure 1

DSPAN architecture and optimization framework. (a) The ResNet-based generator, DSAM, discriminator, and loss paths. (b) DSAM combines gradient and frequency descriptors with the original feature tensor to produce residual attention. (c) SPPCL applies gradient and Fourier-magnitude consistency during training. (d) The generator objective combines content, adversarial, and SPPCL terms.

$$feature \times attention + feature \tag{17}$$

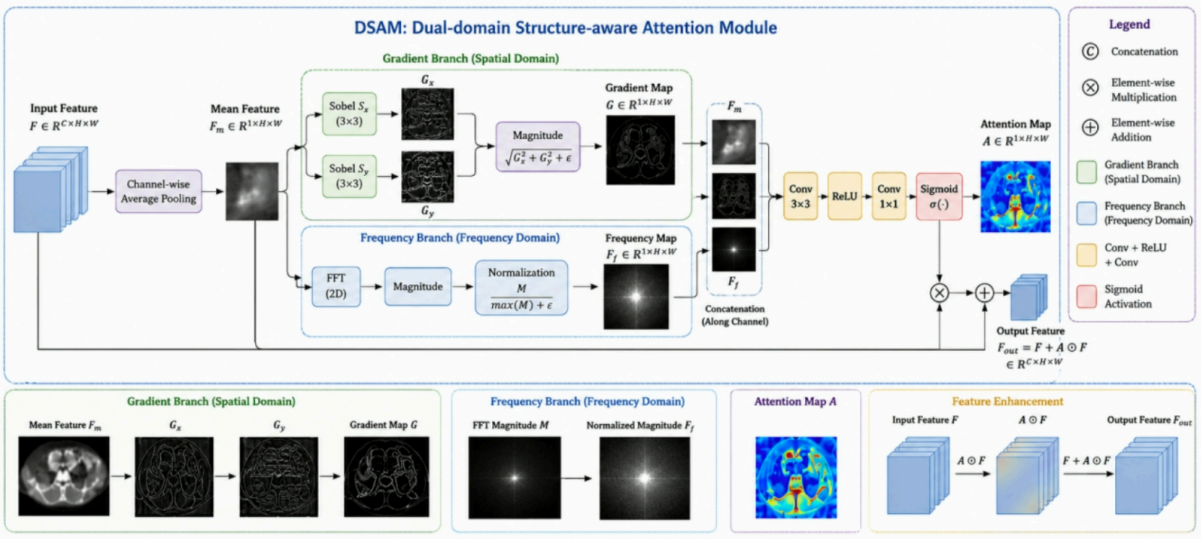


Figure 2

Dual-domain structure-aware attention module. The channel-wise mean feature is transformed with fixed Sobel filters and a two-dimensional fast Fourier transform. The original feature tensor and both descriptors are concatenated, transformed into a channel-wise attention tensor, and applied through residual modulation.

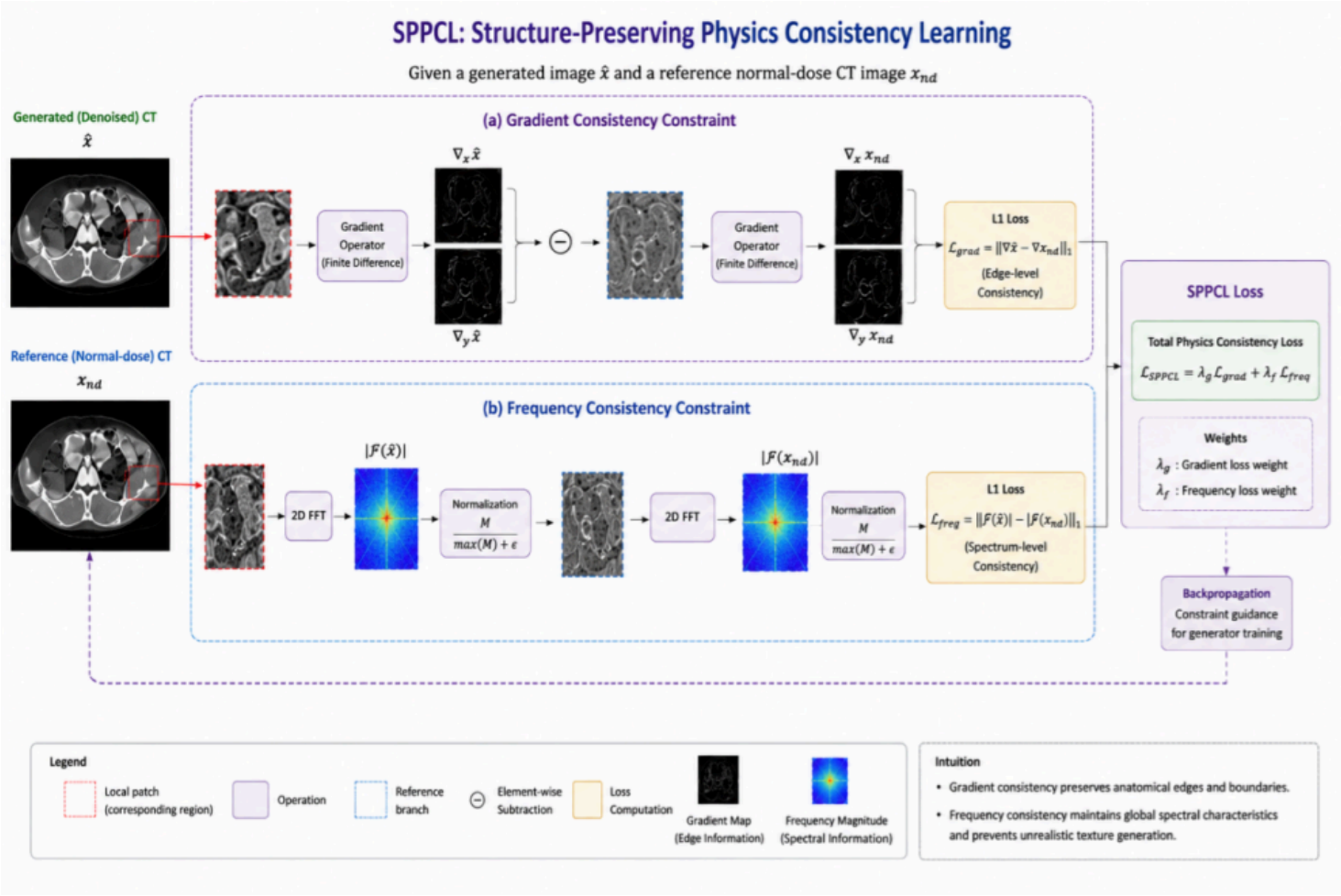


Figure 3

Structure-preserving physics consistency learning. (a) The gradient term minimizes L1 differences between horizontal and vertical first-order image differences. (b) The frequency term minimizes the L1 difference between unnormalized Fourier-magnitude spectra. The weighted terms are combined in SPPCL.

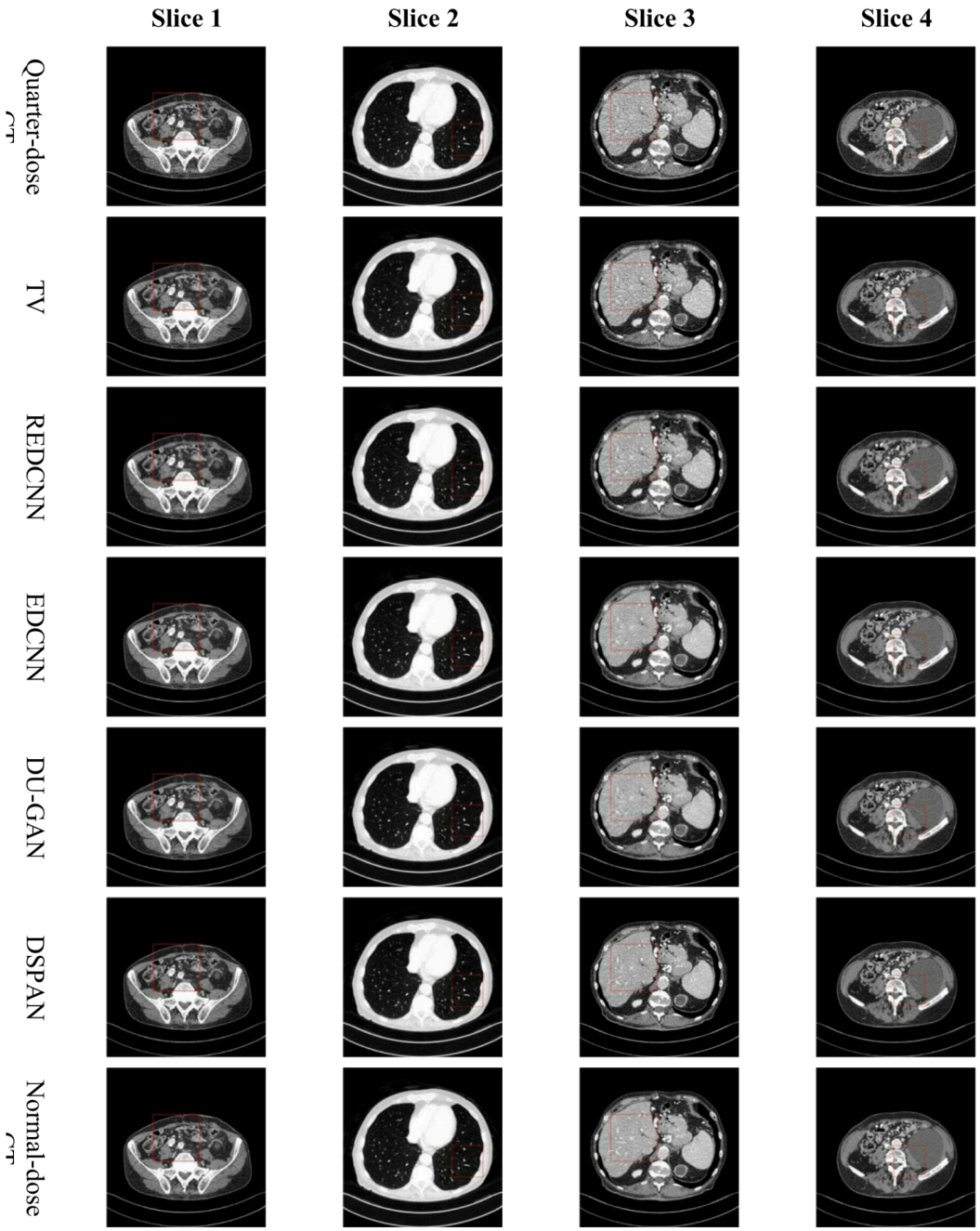


Fig. 4. Qualitative comparison of four representative L506 test slices generated by the fold-4 validation-selected checkpoints. Rows show quarter-dose CT, TV, REDCNN, EDCNN, DU-GAN, DSPAN, and normal-dose CT; columns show slices 148, 178, 329, and 401.

Figure 4

Qualitative comparison of four representative L506 test slices generated by the fold-4 validation-selected checkpoints. Rows show quarter-dose CT, TV, REDCNN, EDCNN, DU-GAN, DSPAN, and normal-dose CT; columns show slices 148, 178, 329, and 401.

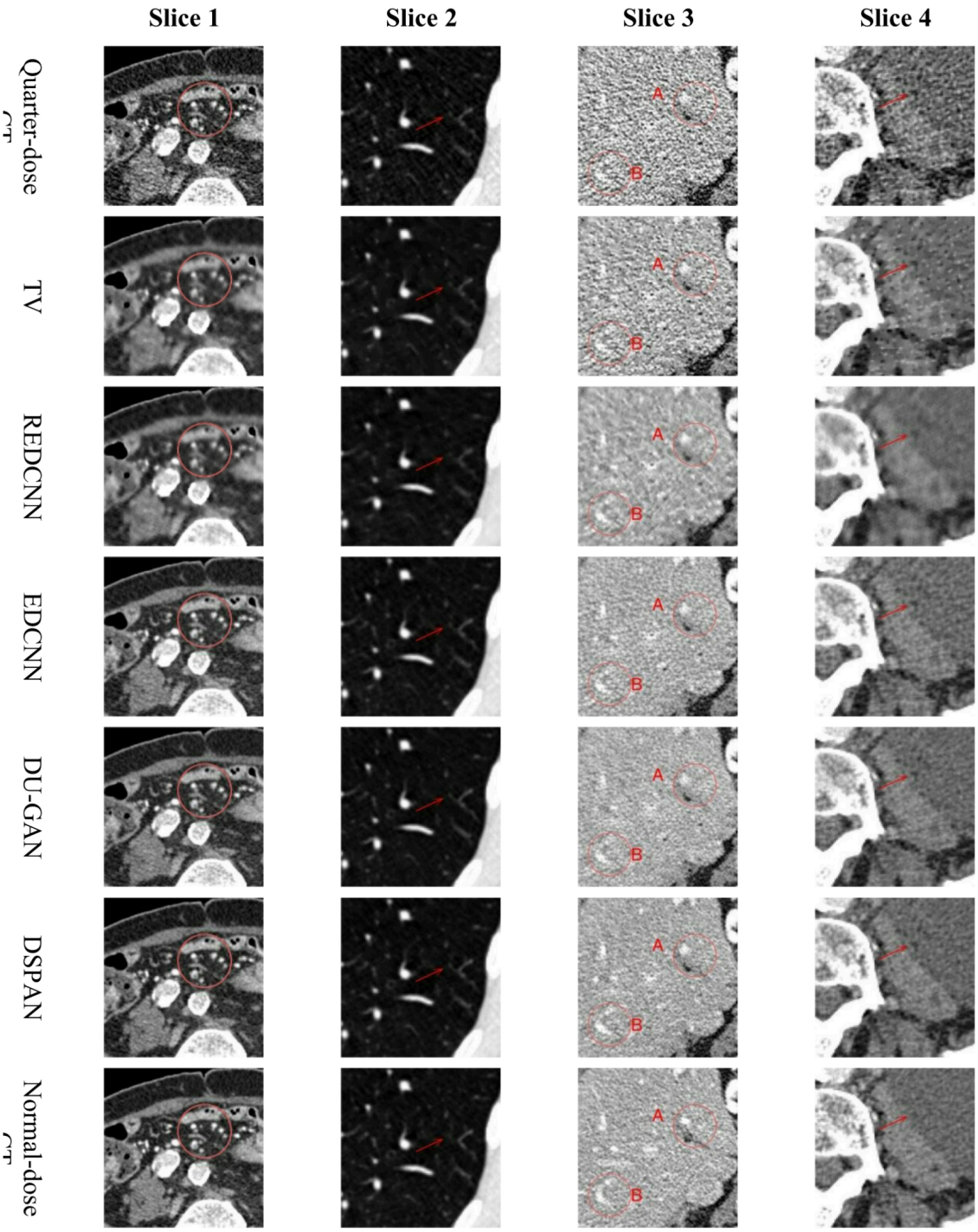


Figure 5

Enlarged regions from the fold-4 qualitative comparison. Red circles, arrows, and labels identify low-contrast boundaries and small structures used for descriptive visual comparison.

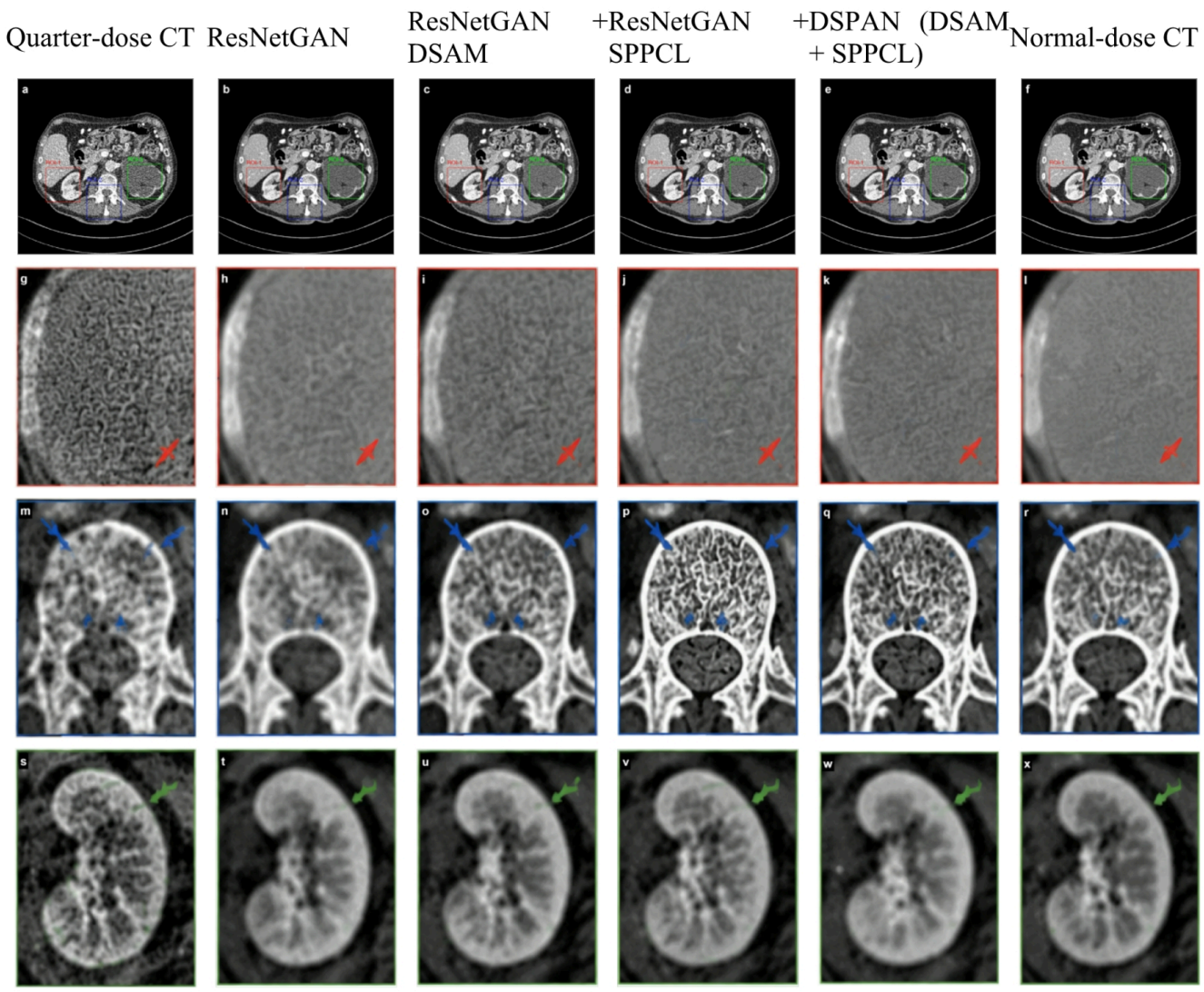


Figure 6

Qualitative module-level ablation generated from a prespecified held-out fold. Panels a–f show quarter-dose CT, ResNetGAN, ResNetGAN + DSAM, ResNetGAN + SPPCL, DSPAN, and normal-dose CT images, respectively. The enlarged rows illustrate differences in residual noise, smoothing, and boundary continuity among the variants. In these selected regions, DSPAN showed closer visual correspondence to the normal-dose reference than the ResNetGAN variants. This single example is descriptive and was not used for statistical inference.

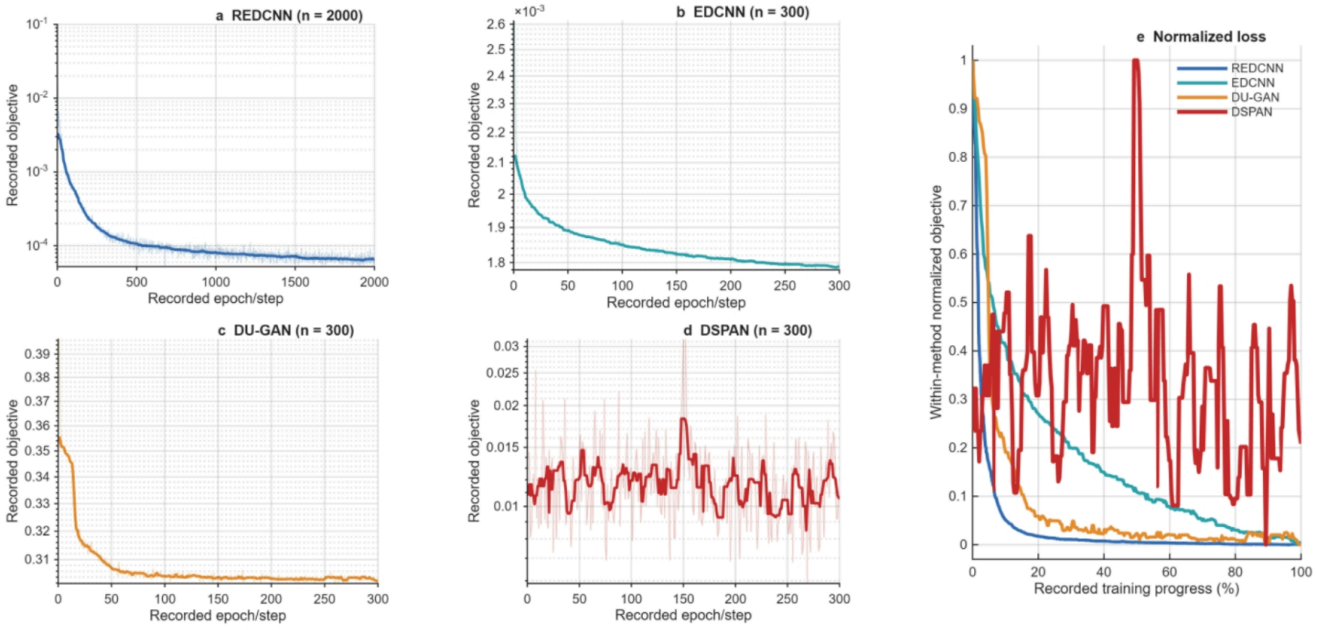


Figure 7

Representative training-loss trajectories. Panels a–d show the recorded objective values for REDCNN, EDCNN, DU-GAN, and DSPAN, respectively; panel e compares within-method normalized trajectories over recorded training progress. The curves are descriptive because the objective definitions and scales differ among methods.

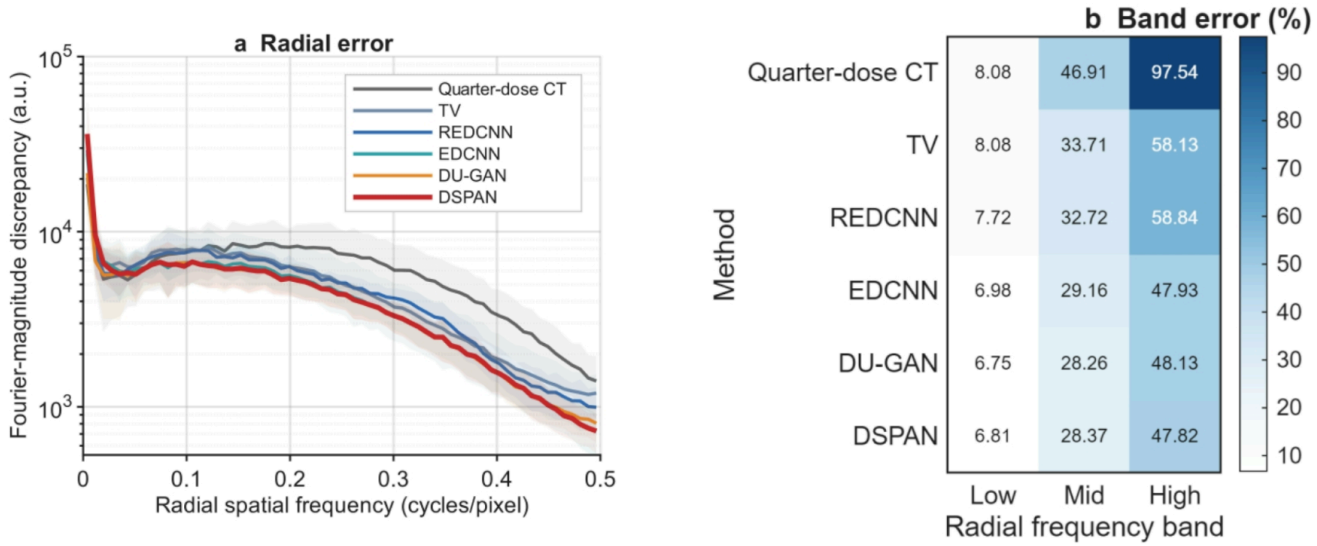


Figure 8

Fourier-magnitude fidelity relative to normal-dose CT. Four representative L506 test slices from fold 4 were evaluated. (a) Radially averaged absolute Fourier-magnitude discrepancy. (b) Mean relative spectral error in the low-, mid-, and high-frequency bands. The analysis is descriptive and based on selected slices.