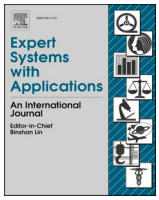
Expert Systems With Applications 294 (2025) 128817

Contents lists available at ScienceDirect

Expert Systems With Applications

journal homepage: www.elsevier.com/locate/eswa



Multi-scale adaptive residual cold diffusion model for Low-Dose CT denoising

Ju Zhang [a,1], Guangyu Liu [a,1], Jikang Chen [a], Yun Cheng [b,*]

[a] School of Information Science and Technology, Hangzhou Normal University, Hangzhou, China
[b] Department of Medical Imaging, Zhejiang Hospital, Hangzhou, China

ARTICLE INFO

Keywords:
Low-Dose CT
Denoising
Cold Diffusion
Multi-scale Adaptive Residual

ABSTRACT

Low-Dose Computed Tomography (LDCT) images are typically severely affected by noise and artifacts, and it is crucial to enhance CT image quality while preserving diagnostic accuracy. In recent years, with the extension of classical diffusion models to Cold diffusion, the fast generation of high-quality images has been achieved through deterministic sampling capabilities. This paper proposes a Multi-scale Adaptive Residual Cold Diffusion Model (MarCoDiff) for LDCT image denoising tasks. Firstly, MarCoDiff employs a cosine mean-preserving degradation (CMPD) operator to effectively simulate the physical degradation process of CT images and accelerate sampling steps. Additionally, a cosine scheduling strategy is integrated to enhance learning at intermediate time steps, and improve the stability of reverse process sampling. Secondly, MarCoDiff constructs a multi-scale adaptive residual denoising network. In this network, we develop an adaptive frequency adjustment module (AFAM) to adaptively adjust the fusion of different frequency components between input features and time step embedding features, effectively preventing over-smoothing or noise in denoised images. We propose a multi-scale residual adaptive block (MRAB) to reconstruct features at each scale from coarse to fine, adapt to texture variations, restore intricate details and edges, and dynamically eliminate anisotropic noise. We design a hybrid-dilated-convolution context block (HCB) that employs mixed symmetric dilation rates to form dilated convolution chains, capturing multi-scale contextual information without further resolution reduction. The model implements a two-stage denoising strategy to efficiently generate higher-quality denoised CT images by alleviating noise and artifact effects. Extensive comparative experiments are conducted on three LDCT datasets. Experimental results demonstrate that the proposed method outperforms other approaches while exhibiting enhanced generalization capability under varying noise levels. We have made our code publicly available at https://github.com/levvvis/MarCoDiff.

1. Introduction

Computed Tomography (CT), as a widely used medical imaging technology in diagnosis and treatment, plays an indispensable role in clinical diagnostics. Nevertheless, the relatively high radiation dose associated with conventional CT scanning may pose cancer risks to patients (Goldman, 2007). Therefore, reducing radiation dosage to mitigate this potential risk has become particularly urgent. However, Low-Dose Computed Tomography (LDCT) images are often accompanied by increased noise and artifacts, which undoubtedly impose additional challenges on physicians' diagnostic work. In this context, enhancing LDCT image quality while reducing patient radiation exposure has emerged as a critical factor in improving diagnostic accuracy and reliability.

LDCT image denoising methods primarily fall into three categories: sinogram filtering, iterative reconstruction, and image post-processing. Sinogram filtering is a denoising method performed in the projection domain, reconstructing CT images through filtered back-projection (Pan et al., 2009) techniques (Wang et al., 2006; Manduca et al., 2009). Representative approaches include structure-adaptive filtering (Kaur et al., 2018) and bilateral filtering (Manduca et al., 2009). Iterative reconstruction constructs objective functions based on noise statistical

* Corresponding author at: Department of Medical Imaging, Zhejiang Hospital, Hangzhou, China.
E-mail addresses: juzhang@hznu.edu.cn (J. Zhang), 2023112011059@stu.hznu.edu.cn (G. Liu), 2024112011026@stu.hznu.edu.cn (J. Chen), 962110508@qq.com (Y. Cheng).
[1] These authors contribute equally.

https://doi.org/10.1016/j.eswa.2025.128817
Received 8 April 2025; Received in revised form 13 June 2025; Accepted 26 June 2025
Available online 29 June 2025

models of projection data and image prior distributions, obtaining reconstructed images through optimization methods. Relevant methodologies encompass dictionary learning (Xu et al., 2012), total variation (Zheng et al., 2018), and non-local mean priors (Liu et al., 2012; Ma et al., 2012). Both sinogram domain filtering and iterative reconstruction methods depend on projection data, yet most manufacturers do not provide such data, thereby increasing the difficulty of algorithm development. Compared to the first two categories, image post-processing methods operate directly on CT images without requiring original projection data. Notable examples include non-local mean filtering (Li et al., 2014) and block-matching algorithms (Sheng et al., 2014). Although these approaches achieve faster reconstruction speeds, the non-uniform nature of noise makes statistical calculations challenging, potentially leading to blurred structural information in CT images.

In recent years, deep learning technology has achieved breakthrough progress and extensive application in image denoising. Convolutional Neural Networks (CNN), leveraging their exceptional feature extraction capabilities, can effectively remove noise by learning noise patterns and image characteristics. The RED-CNN proposed by Chen et al. (Chen et al., 2017) employs a fully convolutional encoder-decoder network with skip connections to enhance training efficiency, achieving remarkable denoising performance. EDCNN (Liang et al., 2020) utilizes an edge enhancement module with Sobel convolution and a densely connected strategy to extract edge features more accurately, thereby improving denoised image quality. UNeXt (Mazandarani, Babyn and Alirezaie, 2023) decomposes the denoising process into a series of stages, where its improved residual blocks extract multi-scale features and employ image reconstruction blocks to progressively merge convolutional information, thereby bridging the gaps between features. However, traditional convolutional operations focus solely on extracting local features while neglecting global and contextual information in images. Moreover, applying identical convolution filters to all pixel locations disregards content differences across spatial positions (Zhou et al, 2021). Concurrently, traditional convolution demonstrates insufficient sensitivity to capture geometric deformations in fine structures, resulting in detail loss.

Residual networks offer a novel approach to LDCT image denoising. RAD-UNet (Qiao and Du, 2022) integrates residual connections, attention mechanisms, and dense connections into the network, enhancing nonlinear fitting capacity and improving performance in suppressing streak artifacts. Wu et al. (Wu et al., 2024) incorporated adaptive edge priors into residual networks to enhance model interpretability and increase reliability in clinical diagnostics. Through cross-layer connections, residual networks enable direct information transmission and preservation, allowing the stacking of network layers to improve feature representation capabilities. However, as residual networks progressively downsample images, excessively reduced image resolution leads to the loss of texture and structural information (Lu et al., 2023).

Recently, diffusion models have emerged as a novel class of generative models, demonstrating exceptional performance in tasks including image synthesis (Rombach et al., 2022:), image restoration (Xia et al., 2023:), and image super-resolution (Saharia et al., 2022). The diffusion model gradually introduces noise into the data and trains the model to remove this noise, ultimately restoring the original data. This approach has gained significant attention in medical imaging, being applied to medical image segmentation (Shuai et al., 2404), MRI reconstruction (Güngör et al., 2023), and LDCT image denoising (Xia et al., 2209). Although diffusion models show potential in these domains, their sampling speed remains relatively slow. Typically, Denoising Diffusion Probabilistic Model (DDPM) (Ho et al., 2020) require nearly a thousand denoising steps to generate a single image. To address this limitation, the Improved Denoising Diffusion Probabilistic Models (IDDPM) (Nichol and Dhariwal, 2021) introduced learnable variance terms, enabling image generation with fewer reverse sampling steps. However, such methods relying on random noise exhibit constraints, particularly when processing LDCT images containing mixed noise patterns, where traditional Gaussian noise assumptions may leave residual artifacts after denoising. Recent advancements reveal that Cold diffusion (Bansal et al., 2024) eliminates randomness through deterministic image transformations (e.g., blurring, masking), enabling finer control over the sampling process and enhancing image quality.

To address the above challenges, we propose a Multi-scale Adaptive Residual Cold Diffusion Model (MarCoDiff) for LDCT image denoising. First, building upon the Cold diffusion, MarCoDiff introduces a novel cosine mean-preserving degradation (CMPD) operator. This operator replaces pure noise images with LDCT images as the terminal point of the forward process (degradation process) and the initial point of the reverse process (sampling process). The cosine scheduling strategy embedded in this approach enables stage-wise learning of image distributions during denoising, thereby enhancing the sampling stability of the reverse process. Second, we design a two-stage multi-scale adaptive residual network as the denoising implementation for the Cold diffusion. This architecture adapts to geometric deformations and noise distributions in images, while incorporating time step embedding features to adjust the fusion of different frequency components information. MarCoDiff's capability ensures effective denoising while preserving image integrity and maintaining clear structural details. Comprehensive experiments conducted on both public dataset and our proprietary synthetic datasets demonstrate that MarCoDiff achieves superior performance compared to existing state-of-the-art methods.

The main contributions of this work are as follows:

- We propose a multi-scale adaptive residual cold diffusion model for LDCT denoising. We propose a cosine mean-preserving degradation (CMPD) operator which employs a cosine scheduling strategy to improve the degradation process, reducing sampling time while balancing denoising performance and structural fidelity, thereby enhancing sampling stability.
- We design an adaptive frequency adjustment module (AFAM). By leveraging combined spatial and frequency domain information, AFAM adaptively adjusts the fusion of different frequency components from input features and time step embedding features. This effectively prevents over-smoothing or noise in the denoised results.
- We propose a multi-scale residual adaptive block (MRAB). During upsampling, MRAB reconstructs features at each scale in a coarse-to-fine manner by integrating multi-scale deformable convolution and adaptive dynamic convolution. This design enables adaptive handling of texture variations, restoration of intricate details and edges, and dynamic elimination of anisotropic noise.
- We design a hybrid-dilated-convolution context block (HCB). At bottleneck layers, HCB utilizes hybrid symmetric dilation rates to construct dilated convolution chains. This expands the receptive field without further resolution reduction, captures multi-scale contextual information, and enhances feature fusion through inter-group skip connections.

The rest of this paper is as follows. In Section 2, we present related work. Section 3 introduces the proposed Cold Diffusion Model for LDCT Denoising. Experimental studies are presented in Section 4. We conclude our work in Section 5.

2. Related work

2.1. LDCT image denoising

Deep learning models have been widely applied to image denoising tasks due to their advantages in extracting image features. DnCNN (Zhang et al., 2017) is a deep convolutional neural network-based denoising method that directly learns the mapping relationship between noisy and noise-free data to achieve denoising. CBDNet (Guo et al., 2019:), composed of a noise estimation subnetwork and a denoising subnetwork, simulates various types of noise that may occur

in realistic images, enabling efficient denoising without relying on prior noise knowledge. Building on this, Anwar et al. (Anwar and Barnes, 2019) proposed RIDNet, a more practical and flexible single-stage blind denoising network, which employs residual structures to facilitate the flow of low-frequency information and utilizes feature attention to enhance inter-channel dependencies. A notable feature of NAFNet (Chen et al., 2022) is its complete elimination of nonlinear activation functions, replacing them with simple multiplications or removing them entirely, thereby simplifying the model architecture while maintaining superior denoising performance.

With the rapid development of deep learning, many researchers have applied these techniques to LDCT image denoising tasks. RED-CNN (Chen et al., 2017) uses residual connections to enhance network performance and adopts an encoder-decoder architecture to minimize structural information distortion during processing, ensuring effective denoising. EDCNN (Liang et al., 2020) is an end-to-end denoiser designed as a fully convolutional network that integrates edge enhancement modules, dense connections, and composite losses, enabling outstanding performance in preserving details and suppressing noise. Universal Anatomy-Initialized Noise Distribution Learning Framework (UNAD) (Gu, Deng and Wang, 2024), building on RED-CNN, optimizes feature extraction through additional convolutional layers and enhances model prior knowledge via noise modeling and inter-slice CT correlations during pre-training, achieving denoising results. CTformer (Wang et al., 2023) employs an overlapping window-based Transformer and Token2Token dilation strategy within an encoder-decoder architecture for LDCT image denoising, overcoming the limitations of convolutional layers restricted to local feature perception and promoting long-range feature interactions. Recently, Yang et al. (Yang et al., 2024) proposed All-In-One Medical Image Restoration (AMIR), a single model capable of handling multiple medical image restoration tasks. Based on Restormer (Transformer-based UNet), AMIR incorporates routing modules to enhance denoising performance in both spatial and channel dimensions. Additionally, some researchers have utilized generative adversarial networks (GAN) (Goodfellow et al., 2020) for LDCT image denoising. WGAN-VGG (Yang et al., 2018) improves training stability and generation quality by introducing Wasserstein distance as the loss function and employs a VGG network as a feature extractor to guide WGAN training, thereby enhancing denoising outcomes. DU-GAN (Huang et al., 2021) extracts content and texture information by constructing image and gradient pathways in the generator and employs a U-shaped discriminator for global and local evaluations to maintain visual consistency between fused and source images.

2.2. Diffusion models

As one of the most powerful generative models in recent years, diffusion models have demonstrated superior performance over GAN (Goodfellow et al., 2020) in tasks such as image generation, image super-resolution, and image restoration, while also exhibiting greater training stability. DDPM frames data generation as a stochastic process of gradually adding noise until data is fully corrupted. The model then incrementally removes noise during the reverse process to recover the original data. DDPM restore images through multi-step iterative refinement, enabling more accurate preservation of high-frequency details. This approach effectively mitigates the over-smoothing phenomenon commonly observed in Transformer-based methods, which stems from their relatively weaker capacity for local texture modeling. However, DDPM typically requires numerous iterations during sampling, resulting in relatively slow speeds. DDIM (Song et al., 2010), a variant of DDPM, predicts results after multiple denoisings within a single denoising iteration, thereby reducing sampling steps and significantly accelerating sampling. SR3 (Saharia et al., 2022) introduces a conditional diffusion model that incorporates a low-resolution image at each diffusion step to provide explicit guidance for iterative refinement of high-resolution images. DiffIR (Xia et al., 2023:) leverages the strong mapping capabilities of diffusion models to estimate a compact image restoration prior for guiding image recovery.

In the field of image denoising, diffusion models also show immense potential and broad applicability. Zhu et al. (Zhu and Chen, 2024) proposed a novel blind image denoising diffusion method, treating noisy images as intermediate states in a Gaussian diffusion process and decomposing image noise into multiple sub-level noises for sequential elimination. Chen et al. (Chen and Shen, 2024) developed a Swin Transformer-based denoising network for DDPM, integrating a k-Nearest Neighbor (kNN) based memory attention module and a hierarchical time stream embedding scheme to enhance model proficiency. For LDCT image denoising, CoCoDiff (Gao and Shan, 2022) trains a noise estimation network using residual CT images as input, with LDCT images serving as conditional guidance. It iteratively denoises from random Gaussian noise via a reverse Markov chain. However, the Gaussian noise introduced in these methods introduces uncertainty, and such randomness is unnecessary for image denoising tasks, especially given that LDCT image noise does not strictly follow a Gaussian distribution. The introduction of Cold diffusion (Bansal et al., 2024) allows for more flexible degradation processes. Inspired by this, Gao et al. (Gao et al., 2023) proposed CoreDiff, a cold diffusion model for LDCT image denoising. CoreDiff incorporates a mean-preserving degradation operator to simulate the physical CT imaging process, avoiding CT value drift and achieving superior denoising performance. Du et al. (Du et al., 2025) proposed Combination of Edge Enhancement and Cold Diffusion Model (CECDM), a LDCT image denoising method based on edge enhancement and cold diffusion models. This approach employs a composite loss function for network training, building upon an improved sobel operator and convolutional block attention module.

3. Multi-scale adaptive residual Cold diffusion model for Low-Dose CT denoising

In this section, we present the overall framework of the proposed multi-scale adaptive residual cold diffusion model, MarCoDiff, which consists of a cosine mean-preserving degradation operator-based Cold diffusion and a multi-scale adaptive residual denoising network. The model employs a two-stage denoising strategy to achieve more acceptable edges and textures while fusing multi-scale information for denoising. Subsequently, we elaborate on the three main components of the network: the adaptive frequency adjustment module (AFAM), multi-scale residual adaptive block (MRAB) and hybrid-dilated-convolution context block (HCB).

3.1. Cold diffusion

Diffusion models are generative models inspired by non-equilibrium thermodynamics. Through T steps Markov chain, these models progressively add Gaussian noise to the original image x_0. Let $\beta_t \in (0,1)$ denote the variance of Gaussian noise added at each step, with $\alpha_t = 1 - \beta_t$ and $\overline{\alpha}_t = \prod_{i=0}^{t} \alpha_i$. The forward process is defined as:

$$q(x_t|x_0) = \mathcal{N}(x_t; \sqrt{\overline{\alpha}_t}x_0, (1-\overline{\alpha}_t)\mathbf{I}) \tag{1}$$

where $\mathcal{N}(\bullet)$ represents the normal distribution and $\mathbf{I}$ denotes the identity matrix, with $t \in [0, T]$. The reverse process involves recovering the source image from Gaussian noise, where the model gradually denoises noise distribution to map it back to the original data distribution for image generation. The formula is expressed as:

$$p_\theta(x_{t-1}|x_t) = \mathcal{N}\left(x_{t-1}; \mu_\theta(x_t, t), \sum\nolimits_\theta (x_t, t)\right) \tag{2}$$

where θ represents the model parameters. This reverse process is implemented using a learnable deep neural network.

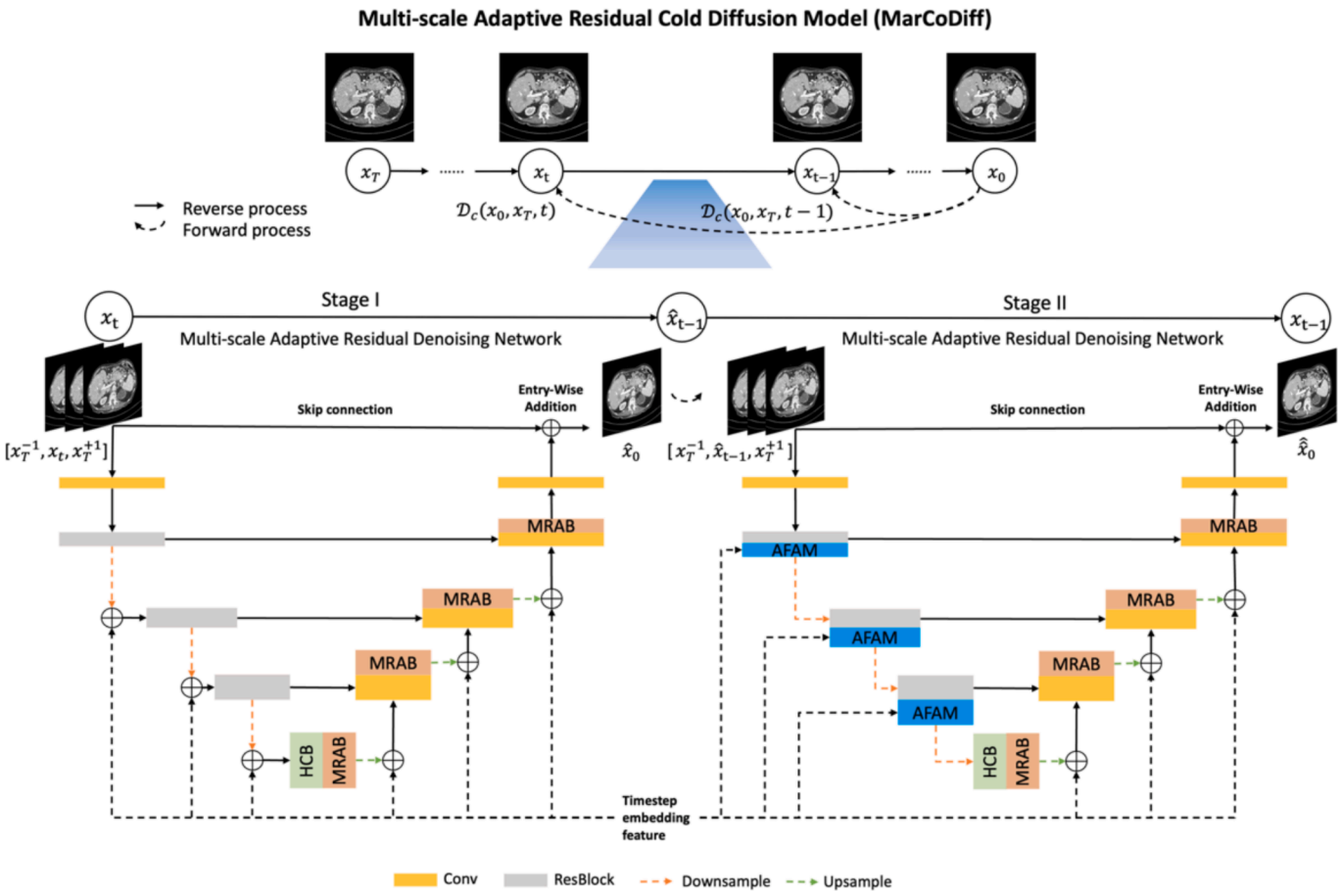


Fig. 1. Overall framework of Multi-scale Adaptive Residual Cold Diffusion Model.

Cold diffusion represents a generalized form of diffusion models that construct forward diffusion processes through deterministic transformations rather than relying on Gaussian noise. Specifically, given an image x_0, a predefined degradation operator $\mathscr{D}(\bullet)$ is applied to continuously degrade x_0 over t steps, defined as:

$$x_t = \mathscr{D}(x_0, t) \tag{3}$$

where $t \in [1, T]$, and T denotes the total number of diffusion time steps. The degradation process must satisfy:

$$\mathscr{D}(x_0, 0) = x_0 \tag{4}$$

Similar to the reverse process described above, Cold diffusion employs a restoration operator $\mathscr{R}(\bullet)$ to invert the degradation process for image recovery and generation. These models utilize a restoration-redegradation sampling algorithm to iteratively generate images through T sampling steps, expressed as:

$$\hat{x}_0 = \mathscr{R}_\theta(x_t, t) \approx x_0 \tag{5}$$

$$x_{t-1} = \mathscr{D}(\hat{x}_0, t-1), t = T, T-1, \ldots, 1 \tag{6}$$

The restoration operator $\mathscr{R}$ is implemented through a deep neural network that samples images from degraded inputs, parameterized by θ. The network can be optimized by the following objective function:

$$\min_\theta \mathbb{E}\|\mathscr{R}_\theta(\mathscr{D}(x, t), t) - x\| \tag{7}$$

When the restoration operator can perfectly restore the original image, this implies that the relation $\mathscr{R}(x_t, t) = x_0$ holds for all t. However, since the restoration operator is generally an approximate inverse of degradation, deviations between restoration operations and original degradation processes may lead to inaccurate sampled images. To address this, Cold diffusion proposes an improved sampling strategy. Instead of directly sampling x_t from x_0, they sample through intermediate variables:

$$x_{t-1} = x_t - \mathscr{D}(\hat{x}_0, t) + \mathscr{D}(\hat{x}_0, t-1) \tag{8}$$

This approach estimates the previous step's image by computing the difference between the current degraded image and two distinct degradation levels, thereby reducing error accumulation caused by imperfect restoration networks. This restoration method proves more stable and effective for deterministic degradation, enabling higher-quality image generation.

3.2. *The proposed model MarCoDiff*

The forward process of diffusion models involves artificial degradation through the addition of Gaussian noise. However, the noise in LDCT image is fundamentally mixed noise. This mismatch hinders the model's ability to accurately learn the reverse process. Inspired by Cold diffusion and CoreDiff (Gao et al., 2023) mean-preserving degradation operator, MarCoDiff employs a cosine mean-preserving degradation (CMPD) operator, as illustrated by the dashed line representing the forward process in the upper part of Fig. 1. Specifically, we gradually degrade Normal-Dose Computed Tomography (NDCT) images into corresponding LDCT images, where the LDCT image serves as the sampling starting point while retaining low signal-to-noise ratio yet structural information. The degraded image x_t obtained at each time step precisely matches the noise statistical characteristics of LDCT images, the formulation is defined as:

$$x_t = \mathscr{D}(x_0, x_T, t) = \alpha_t x_0 + (1 - \alpha_t) x_T \tag{9}$$

where α_t satisfies $\alpha_t < \alpha_{t-1}$. As t increases, the proportion of the NDCT image x_0 in the degraded image x_t diminishes, leaving less effective information in x_t– aligning with the objective of classical diffusion models' forward process. However, distinctively, the pixel mean of images during processing remains consistent with the NDCT image throughout the forward process. This design simulates the physical process of CT dose reduction and prevents CT value shifts caused by noise accumulation in conventional diffusion models. x_T represents the LDCT image, the final state achieved through the forward process.

Unlike CoreDiff's linear scheduling strategy, its degeneration weight ratio decreases uniformly over time and demonstrates uniform sensitivity to time steps. The degradation weight α_t in the CMPD operator adopts a cosine scheduling strategy, formulated as follows:

$$\alpha_t = \frac{f(t)}{f(0)}, f(t) = cos\left(\frac{t/T + s}{1 + s} \bullet \frac{\pi}{2}\right)^2 \tag{10}$$

This cosine scheduling ensures slow degradation during the beginning and end of the phase. This indicates that the cosine scheduling strategy places greater emphasis on the features of NDCT images during the beginning of training phase. To prevent the network predictions from being adversely affected by an insufficient degradation level at the beginning, the offset s is set to 0.008. At the end of phase, it mitigates overfitting to LDCT image x_T. The intensification of degradation in intermediate phase compels the network to concentrate its learning efforts on stage-specific denoising proficiency. Compared to linear scheduling, our cosine strategy prioritizes high-frequency detail preservation and edge sharpness through gradual initial degradation. It also enhances stability in the restoration-redegradation sampling algorithm of Cold diffusion by preventing abrupt transitions from uniform time step degradation. The reverse process is ultimately expressed as:

$$x_t = \mathscr{D}_c(x_0, x_T, t) \tag{11}$$

To mitigate structural distortion and preserve detail during reverse sampling, spatial neighboring slices are incorporated as contextual information into the denoising network (Shan, 2018). Specifically, given a slice x with adjacent upper and lower slices x^{-1} and x^{+1}, the contextual version x^c is formed by concatenating x^{-1}, x, and x^{+1} along the channel dimension. At any time step t, contextual information always references the initial sampling point x_T. For time step t, this process can be formulated as:

$$x_t^c = Concat\left(x_T^{-1}, x_t, x_T^{+1}\right) \tag{12}$$

The $\mathscr{R}_\theta$ corresponds to our proposed MarCoDiff, which iteratively generates images through a two-stage recomposition-degradation sampling algorithm based on Cold diffusion. The formulation is as follows:

$$\widehat{x}_0 = \mathscr{R}_\theta\left(x_t^c, t\right) \approx x_0 \tag{13}$$

$$x_{t-1} = x_t - \mathscr{D}_c(\widehat{x}_0, x_T, \mathrm{t}) + \mathscr{D}_c(\widehat{x}_0, x_T, \mathrm{t} - 1) \tag{14}$$

This enables iterative sampling to obtain the final predicted image x_0. The ultimate training objective function of the MarCoDiff two-stage algorithm is defined as:

$$\min_\theta \mathbb{E}\left[\left\|\mathscr{R}_\theta\left(x_t^c, t\right) - x_0\right\|^2 + \left\|\widehat{\widehat{x}}_0 - x_0\right\|^2\right] \tag{15}$$

where $\widehat{\widehat{x}}_0$ represents $\widehat{\widehat{x}}_0 = \mathscr{R}_\theta\left(x_{t-1}^c, t-1\right)$ at step t-1 in the second stage.

The CMPD operator utilizes LDCT image as the starting point for the reverse process, restoring images while preserving original structures. This approach reduces sampling steps and accelerates the reverse process. The cosine scheduling strategy enhances learning at intermediate time steps, improving sampling stability during the reverse process. These mechanisms collectively guide image sampling more effectively, generating high-quality denoised images.

3.3. Multi-scale adaptive residual network

As shown in the lower part of Fig. 1, our network employs a 3 × 3 convolution for initial shallow feature extraction from input LDCT images. The features are then processed through a multi-scale encoder-decoder architecture. In the encoder, ResBlock (He et al., 2016) extracts multi-scale features while filtering noise. Different from standard ResBlock, we remove batch normalization layers and adopt LeakyReLU (Maas et al., 2013) as the activation function. To avoid compromising the image structure, we restrict the number of downsampling operations. Specifically, our denoising network contains four scales with channel dimensions of 32, 64, 128, and 256 respectively. Residual connections between encoder and decoder are implemented at all scales except the lowest one to ensure effective information transmission. When images undergo progressive downsampling, their spatial resolution becomes excessively low, leading to information loss. Therefore, at the bottleneck layer, we implement a HCB. This module employs four groups dilated convolutions of symmetric and hybrid dilation rates to extract contextual information. By maintaining resolution while expanding receptive fields, it captures multi-scale features without additional downsampling. Skip connections between dilation groups further enhance contextual information fusion. In the decoder, MRAB reconstruct multi-scale features from coarse to fine, integrating a multi-scale deformable convolution (MDC) and an adaptive dynamic convolution (ADC). MDC adapts to complex texture variations and edge patterns at different scales, enhancing detail recovery capabilities. ADC incorporates channel-wise and spatial interaction information into dynamic kernel generation, enabling denoising through dynamic convolutions. The model employs a dual-phase strategy to mitigate noise and artifacts. Using preliminary outputs from the first stage as guidance, the second stage incorporates an AFAM. This module adaptively fuses spatial and frequency domain features while adjusting different frequency components through time step embedding features integration, preventing over-smoothing or noise in denoised images. For downsampling, we use 2 × 2 convolution kernel with stride 2. Instead of transposed convolution, bilinear upsampling combined with 1 × 1 convolution is adopted for resolution recovery, effectively reducing checkerboard artifacts commonly caused by transposed convolution operations.

3.4. Adaptive frequency adjustment module (AFAM)

In diffusion models, time steps represent the progression of the diffusion process. The embedding of each time step t helps the model understand the current stage of the diffusion process and provides high-dimensional feature representations for each time step. In the first stage of the denoising network, we initially coarsely add time step embedding features to input features at each scale. Guided by the first-stage output, we then introduce the AFAM during the downsampling of the second stage, which combines spatial and frequency domain information while adaptively adjusting feature fusion based on different time steps.

In LDCT images, high-frequency components correspond to edge textures, with most noise concentrated in high-frequency regions. The overall structural organization primarily resides in low-frequency information. Denoising methods typically suppress high-frequency noise through filtering, but this inevitably leads to loss of critical edge details in CT images. Unlike approaches that solely replenish high-frequency information (Wang et al., 2405), our method adaptively preserves image details while maintaining structural integrity and stability, thereby avoiding over-smoothing or noise in denoised images. To achieve this, we adjust the spectrum of time step embedding features at each downsampling scale in fine granularity, enabling the model to focus on stage-specific characteristics. The application of Fast Fourier Transform (FFT) facilitates conversion from spatial domain to frequency domain,

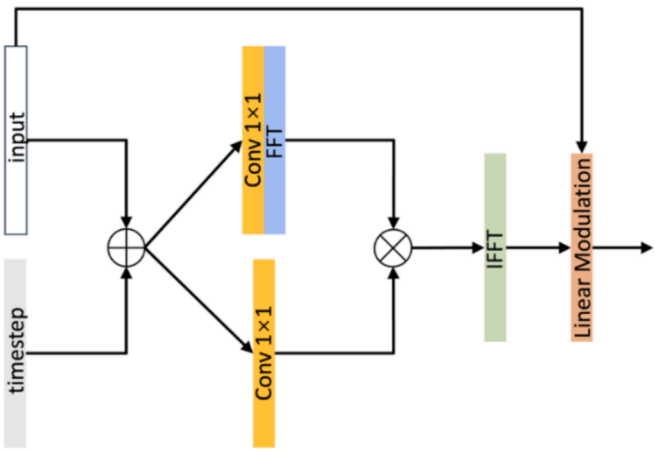


Fig. 2. The structure of adaptive frequency adjustment module (AFAM).

capturing global frequency distribution patterns and compensating for the local receptive field limitations of traditional convolutions. This empowers the model to directly enhance or suppress specific frequency components while progressively adjusting high and low-frequency information.

As illustrated in Fig. 2, after embedding time step into input features F_{in} to obtain F_1, one branch processes spatial domain features through 1 × 1 convolution to capture local structures, while the other applies 1 × 1 convolution followed by FFT $\mathscr{F}$ projection to frequency domain for global noise distribution analysis. The multiplied outputs from both branches yield combined spatial and frequency domain information representations that enrich feature informativeness and enhance targeted processing. Specifically, global noise suppression operates in the frequency domain, while spatial features compensate for authentic structural edges in denoised outputs. An Inverse FFT (IFFT) $\mathscr{IF}$ transforms features back to spatial domain. Through linear feature modulation (Perez et al, 2018), two learnable parameters α and β dynamically determine fusion ratios of spatial and frequency domain information based on time step embedding feature F_{te}, enabling adaptive adjustment of frequency components. The formulation is expressed as:

$$F_1 = F_{in} + F_{te} \tag{16}$$

$$F_2 = \mathscr{IF}(f^{1\times1}(F_1) \otimes \mathscr{F}(f^{1\times1}(F_1))) \tag{17}$$

$$F_{out} = F_2 \times \alpha + F_{in} \times \beta \tag{18}$$

This enhancement of time step embedding feature representations enables more effective time step dependency modeling, allowing the model to adaptively modulate different frequency components. Consequently, it prevents common denoising artifacts such as over-smoothing or noise in denoised image while maintaining structural fidelity.

3.5. Multi-scale residual adaptive block (MRAB)

The tissue edges in LDCT images exhibit the non-uniform characteristic. Traditional convolution kernels with fixed geometric shapes struggle to adapt to regional geometric deformations, where fixed sampling points tend to smooth textures. Compared with traditional convolutions, modulated deformable convolution enhances pixel correlation by adaptively adjusting sampling positions of convolution kernels, enabling convolution kernels to conform to target structures. It learns spatial offsets for each sampling point and modulates weights to accommodate structural variations in CT images.

The distributions of noise, artifacts, and high-frequency components of structures in CT images demonstrate similarity (Zhang et al., 2021), while the noise exhibits anisotropic characteristic. Traditional convolutions with fixed kernels fail to adapt to complex image features and

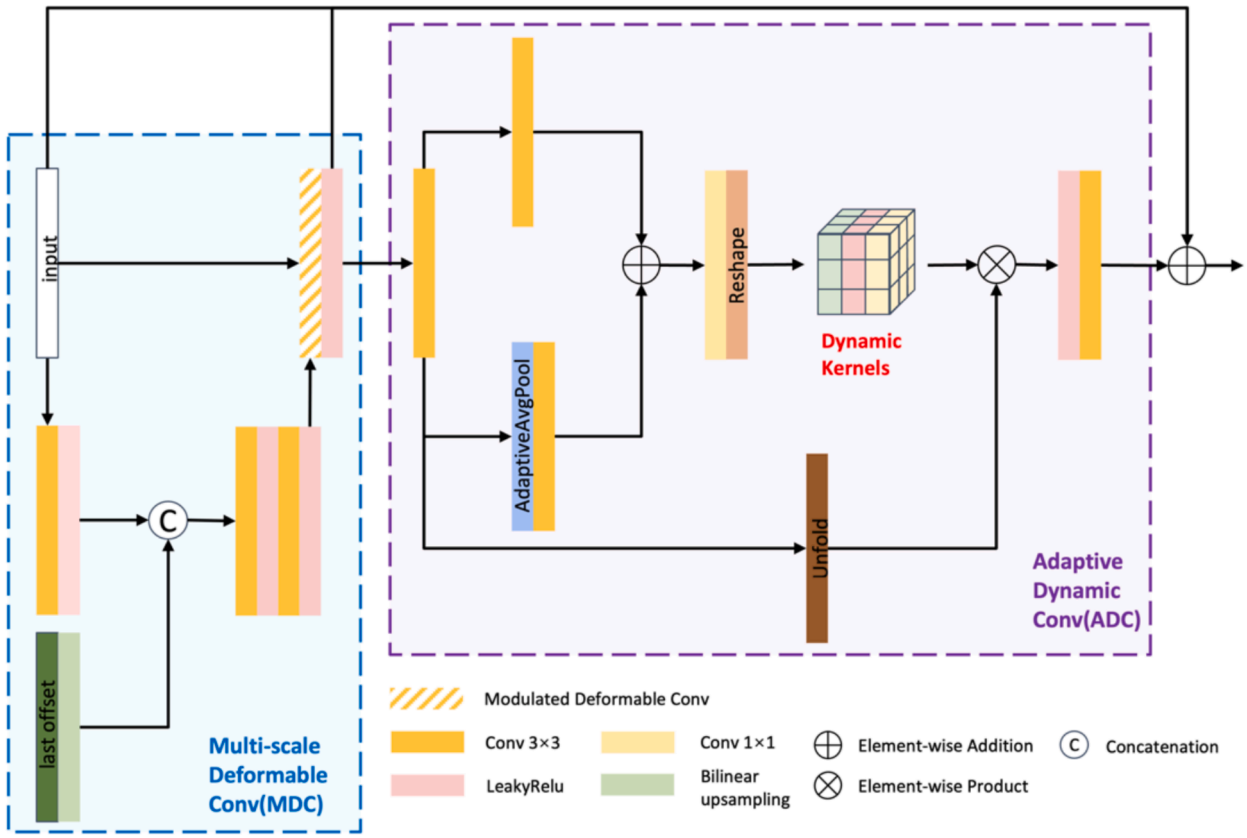


Fig. 3. The structure of multi-scale residual adaptive block (MRAB).

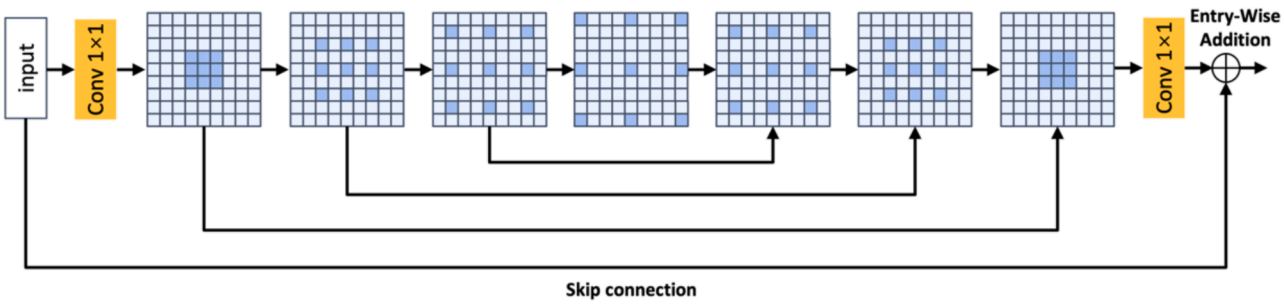


Fig. 4. The structure of hybrid-dilated-convolution context block (HCB).

ignores regional heterogeneity, leading to loss of texture and structural information. In contrast, dynamic convolutions adaptively adjust convolution kernels according to spatial anisotropic noise and textural complexity, endowing networks with dynamic noise suppression and detail restoration capabilities.

Our proposed MRAB, as illustrated in Fig. 3, achieves multi-scale feature reconstruction from coarse to fine during upsampling. The framework integrates a multi-scale deformable convolution (MDC) and an adaptive dynamic convolution (ADC). Adopting a ResBlock-like structure, we replace its first convolutional layer with MDC. This MDC enhances adaptability to structural deformations in LDCT images, enabling more effective extraction of complex edge-texture features, using LeakyReLU as activation function. Subsequently, ADC incorporates channel-wise information interaction into kernel generation based on spatial context, guiding dynamic convolution to enhance noise adaptation and texture preservation. Local residual learning is employed a in the MRAB output to strengthen information flow and improve representational capacity.

As shown in the blue region of Fig. 3, the MDC architecture utilizes previous scale offsets as references, integrating extracted features with upsampling offsets from the last scale for multi-scale reconstruction from coarse to fine. We employ bilinear interpolation f^{up} to upsample the last scale offsets F_{offset}^{s-1} to current scale. Then F_{offset}^{s-1} and input features F_{in} jointly estimate F_{offset}^{s}. This strategy enables precise localization of correlated features across scales through small-scale reference offsets. The formulation is expressed as:

$$F_1^s = f^{up}\left(F_{offset}^{s-1}\right) \tag{19}$$

$$F_2^s = \tau\left(f^{3\times3}(F_{in})\right) \tag{20}$$

$$F_{offset}^s = \tau(f^{3\times3}(\tau(f^{3\times3}(Concat(F_1^s, F_2^s))))) \tag{21}$$

where $f^{3\times3}$ denotes 3 × 3 convolution and τ represents LeakyReLU activation. Let F_{in} denote input features at position p. The offset Δp is obtained via 3 × 3 convolution, while learnable modulation scalar Δm is computed through activation. The modulated deformable convolution output F_{out} can be expressed as:

$$\{\Delta p, \Delta m\} = f^{3\times3}\left(F_{offset}\right) \tag{22}$$

$$\Delta m = \sigma(\Delta m) \tag{23}$$

$$F_{out}(p) = \sum_{p_i \in N(p)} w_i \bullet F_{in}(p_i + \Delta p_i) \bullet \Delta m_i \tag{24}$$

where $N(p)$ denotes the neighborhood of position p, w_i represents kernel weights of position p, and p_i indicates positions within $N(p)$. Thus, MDC adjusts feature magnitudes and spatial receptive fields to enhance capability of complex structure catpture.

Previous approaches typically discard spatial information and fail to effectively aggregate channel-specific features. We propose an efficient method for guiding dynamic filter generation. Since dynamic convolutions and deformable convolutions produce complementary representations, we integrate ADC after MDC to enhance anisotropic noise adaptation. As shown in the red region of Fig. 3, our ADC design incorporates a spatial context extractor branch and a channel information interaction branch to generate dynamic convolution kernels. Compared with traditional dynamic attention mechanisms, this architecture not only reduces parameter quantity but also alleviates joint optimization challenges (Li et al., 2103). This enables the network to dynamically adjust according to input features at different hierarchical levels, thereby enhancing the model's ability to suppress noise in different regions and of different intensities. The spatial context extractor branch is implemented using 3 × 3 convolution. The channel information interaction branch consists of a global average pooling followed by 3 × 3 convolution for information fusion. Spatial and channel information are integrated through element-wise addition operations. Finally, a 1 × 1 convolution is employed to project the fused features into convolution kernel shapes, producing output dimensions of (k × k × c) × h × w. These are reshaped into a series of per-pixel kernels $W_{i,j} \in \mathbb{R}^{k\times k\times c}$, where $i \in \{1, 2, \ldots, h\} j \in \{1, 2, \ldots, w\}$. The overall formulation can be expressed as:

$$F_1 = f^{3\times3}(F_{in}) \tag{25}$$

$$F_2 = f_{groups}^{3\times3}(F_1) \oplus f^{3\times3}(GAP(F_1)) \tag{26}$$

$$W_{i,j} = Reshape\left(f^{1\times1}(F_2)\right) \tag{27}$$

$$F_{out} = f^{3\times3}\left(\tau\left(Unfold(F_1) \otimes W_{i,j}\right)\right) \oplus F_{in} \tag{28}$$

where $f_{groups}^{3\times3}$ denotes grouped convolution, ⊕ represents element-wise addition, and ⊗ denotes element-wise product. ADC leverages content-adaptive properties and spatial context to guide local kernels generation, achieving superior local feature processing. Thus, the MRAB can be formally expressed as:

$$F_{MRAB} = F_{ADC}(\tau(F_{MDC}(F_{in}))) \oplus F_{in} \tag{29}$$

where F_{MDC} and F_{ADC} denote outputs from respective modules. The combination of MDC and ADC enables MRAB to adaptively handle textural variations, restore complex details and edges, and dynamically suppress noise in different regions and of different intensities while preserving complete structures.

3.6. Hybrid-dilated-convolution context block (HCB)

Downsampling can expand the receptive field of networks to capture more image contextual information. However, continuous downsampling reduces image resolution, leading to information loss and detail blurring. Particularly in CT images, this may cause excessive edge smoothing and loss of textural details. In previous studies, addressing this issue solely with single-dilation-rate dilated convolutions resulted in

pixels retrieving information in a checkerboard pattern while still losing substantial information. The underlying reason is that the convolutional results of the current layer remain independent from those of the previous layer, lacking correlation and causing local information loss – a phenomenon termed the gridding effect.

To resolve these issues, we introduce a HCB in the network's bottleneck layer. Inspired by a U-shaped architecture, HCB employs a hybrid symmetric dilation rate chain of dilated convolutions. This approach expands the receptive field without altering image resolution, effectively eliminating the gridding phenomenon while preserving image contextual information.

As illustrated in Fig. 4, the process begins with a 1×1 convolution to compress feature channels, simplifying subsequent operations. HCB adopts four groups of symmetric and hybrid dilation rates (1, 2, 3, 4, 3, 2, 1) to construct a comprehensive information capture mechanism. During the dilation rate increasing phase, the receptive field progressively enlarges to capture information from local to global features. In the decreasing phase, progressive skip connections between dilated convolutions with identical dilation rates integrate contextual features while preserving image details. A subsequent 1×1 convolution restores the output channels to match the original input features. Skip connections between input and output features are implemented to prevent information blockage.

4. Experimental studies

This section aims to evaluate the performance of our proposed model in image denoising tasks through a series of experiments. We will elaborate on the experimental settings, covering the following aspects: datasets, comparative methods, evaluation metrics, training details, and loss function. Through comparative analysis of the experimental results, we will demonstrate the model's performance under various conditions. Additionally, we conduct ablation experiments to investigate the roles of individual modules in the model, thereby revealing their specific contributions to overall performance. We have made our code publicly available at https://github.com/levvvis/MarCoDiff.

4.1. Datasets

To comprehensively evaluate the performance of LDCT image denoising algorithms, this study employed three distinct datasets, including one public dataset and two synthetic datasets. These datasets tested the algorithms under different conditions, ensuring the reliability of experimental results. During the preprocessing, we performed data normalization and, in accordance with the methodology described in the preceding Methods section, concatenated spatially adjacent slices of the current image along the channel dimension.

(1) Public LDCT Dataset.

The NIH-AAPM-Mayo Clinic LDCT Grand Challenge in 2016 dataset was utilized for training and testing. This dataset contains 5,936 CT slices of size 512×512 with a 1 mm thickness from 10 patients. Data acquired at 120 kV and 200 mAs are defined as full-dose data, while those acquired at 120 kV and 50 mAs are designated as quarter-dose data. Case L506, comprising 526 image pairs, was selected as the test set, with the remaining 9 patients' data used for training.

(2) Constructed LDCT Datasets from NDCT.

We developed our own LDCT datasets based on the Qin_LUNG_CT dataset. The Qin_LUNG_CT dataset comprises 3,841 NDCT images of size 512×512 with a 5 mm thickness from 46 patients. We developed LDCT datasets both in the sinogram domain and in the image domain respectively with different noise levels.

In the Sinogram Domain

Projection data were generated through simulated CT scans, with noise added to simulate low-dose conditions. Images were then reconstructed from the projection data.

The geometric configuration for simulating LDCT imaging and reconstruction was implemented using the toolbox for iterative CT reconstruction (TIGRE) library. Key parameters are listed in Table 1.

Table 1
Geometric configuration of CT imaging.

Parameters	Values
Distance source detector	1536 mm
Distance source origin	1000 mm
Number of voxels	512×512
Size of the image	$512mm \times 512mm$
Size of each voxel	$1mm \times 1mm$
Number of detector pixels	1024
Size of each pixel on the detector	0.8 mm
Geometry mode	Parallel beam

(a) Projection Data Generation

In practical CT scanning, X-rays pass through the scanned object from various angles, with detectors recording the attenuation intensity after radiation penetration at each angle. To simulate this process, we first load the CT image I from the target dataset. Using the Ax function from the TIGRE library, we generate projection data based on the geometric configuration specified in Table 1 and 360 equally spaced angles (0 to 2π). This simulation mimics the radiation attenuation data acquired by CT scanning equipment from multiple angles.

The mathematical representation is as follows:

$$P = \alpha \bullet \mathrm{I} \tag{30}$$

where α denotes the projection operator, and P represents the generated projection data with dimensions [360, 1024]. This indicates 1024 projection values recorded by the detector at each of the 360 projection angles.

(b) Lose Dose/Noise Simulation

To simulate LDCT image noise characteristics, Poisson noise and Gaussian noise were added to the projection data P. Poisson noise intensities were set to $\lambda = 1 \times 10^4, 1 \times 10^5, 1 \times 10^6$, with the probability distribution:

$$P(x;\lambda) = \frac{\lambda^x e^{-\lambda}}{x!} \tag{31}$$

Gaussian noise parameters are mean $\mu = 0$ and standard deviation $\sigma = 10$, with the probability density function:

$$\mathcal{N}(x;\mu,\sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{\frac{(x-\mu)^2}{2\sigma^2}} \tag{32}$$

The final noisy projection data P' is expressed as:

$$P' = P + \mathcal{N} \tag{33}$$

(c) Image Reconstruction

The ordered-subset simultaneous algebraic reconstruction techniques (OS-SART) was employed for iterative reconstruction of noisy projection data. The OS-SART iterative formula is:

$$I^{k+1} = I^k + \lambda \sum_{j \in S} \frac{\alpha_j^T (P_j - \alpha_j I^k)}{\sum_{j \in S} \alpha_j^T \alpha_j} \tag{34}$$

where I^k is the image at the k-th iteration, λ is the relaxation factor, and S denotes the current subset. A total of 100 iterations ensured convergence and image quality.

Three Poisson noise levels were applied in the sinogram domain, corresponding to high-dose, low-dose, and ultra-low-dose CT datasets, as shown in Fig. 5.

In the Image Domain

Gaussian noise with standard deviations of 15, 30, 45, and 60 was added to NDCT images, generating four paired subsets with different noise levels, as shown in Fig. 6. The constructed datasets not only simulated diverse noise conditions in clinical LDCT imaging but also provided multiple testing scenarios for evaluating denoising algorithms.

Fig. 5. LDCT images with different noise levels in the sinogram domain.

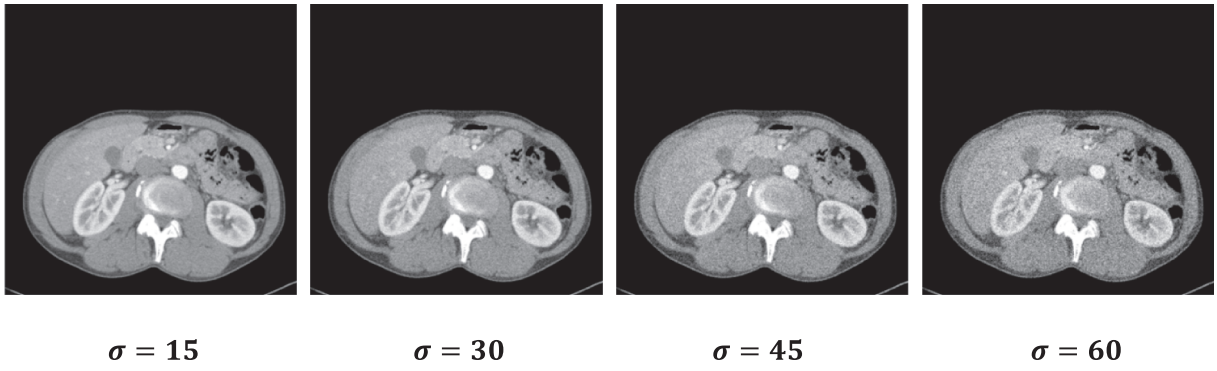


Fig. 6. LDCT images with different noise levels in the image domain.

The training set primarily consisted of CT images from 45 patients, while 74 CT image groups from patient R0274 were reserved for testing.

4.2. Comparative methods

This paper compares the proposed MarCoDiff with four categories of LDCT image denoising methods, including: 1) CNN-based models: RED-CNN (Chen et al., 2017) and UNAD (Gu, Deng and Wang, 2024); 2) Transformer-based models: CTformer (Wang et al., 2023) and AMIR (Yang et al., 2024); 3) GAN-based model: DU-GAN (Huang et al., 2021); and 4) diffusion-based models: IDDPM (Nichol and Dhariwal, 2021) and CoreDiff (Gao et al., 2023). The training parameters of these comparative methods were configured according to their original publications. All models were retrained and validated on the same test set to ensure a fair and comprehensive comparative analysis using multiple evaluation metrics.

4.3. Evaluation metrics

To objectively assess the performance of each denoising network, three widely adopted metrics were selected: peak signal-to-noise ratio (PSNR), structural similarity index (SSIM), and root mean squared error (RMSE).

PSNR quantifies image quality by measuring the similarity between images. Its formula is defined as:

$$PSNR = 10 \bullet log_{10}\frac{MAX^2}{MSE} \tag{35}$$

where MAX denotes the maximum possible pixel value of the image. MSE calculates the mean squared error between images I and K:

$$MSE = \frac{1}{mn}\sum_{i=0}^{m-1}\sum_{j=0}^{n-1}(I(i,j) - K(i,j))^2 \tag{36}$$

where m and n represent the height and width of the image. PSNR is expressed in decibels (dB), with higher values indicating better image quality.

SSIM evaluates image similarity by considering luminance, contrast, and structural features. Its formula is:

$$SSIM(x,y) = \frac{\left(2\mu_x\mu_y + C_1\right)\left(2\sigma_{xy} + C_2\right)}{\left(\mu_x^2 + \mu_y^2 + C_1\right)\left(\sigma_x^2 + \sigma_y^2 + C_2\right)} \tag{37}$$

where μ_x and μ_y are the mean values of images, σ_x and σ_y denote their variances, σ_{xy} is their covariance, and C_1, C_2 are constants to stabilize division. SSIM ranges from 0 to 1, with values closer to 1 indicating higher similarity and better preservation of structural details.

RMSE measures prediction accuracy for continuous data by computing the square root of the average squared differences between predicted and ground truth values:

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \widehat{y}_i)^2} \tag{38}$$

where n is the number of samples, y_i is the true value, and $\widehat{y}_i$ is the predicted value. A lower RMSE signifies smaller prediction errors and

Table 2
Comparison of loss weights on the AAPM-Mayo dataset.

Loss Weights	PSNR	SSIM	RMSE
$\alpha_1 = 0.7, \alpha_2 = 0.3$	29.5359	0.8750	13.6610
$\alpha_1 = 0.6, \alpha_2 = 0.4$	29.5479	0.8751	13.6412
$\alpha_1 = 0.5, \alpha_2 = 0.5$	29.5561	0.8751	13.6268
$\alpha_1 = 0.3, \alpha_2 = 0.7$	29.5725	0.8753	13.6016
$\alpha_1 = 0.4, \alpha_2 = 0.6$	**29.5782**	**0.8754**	**13.5936**

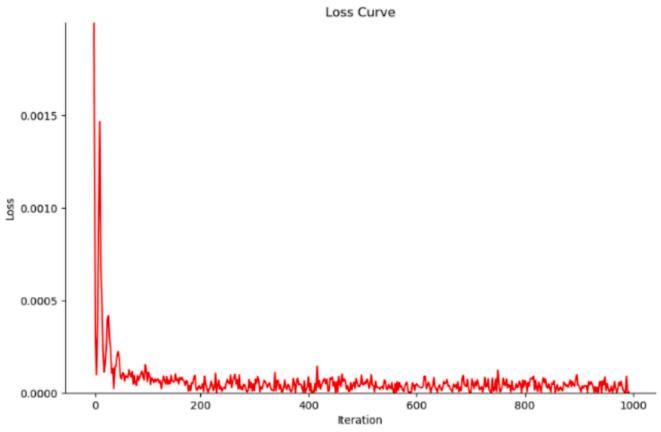


Fig. 7. Training loss curve up to 1000 iterations.

Table 3
Quantitative comparison of denoising methods on the AAPM-Mayo dataset.

Methods	Params	PSNR	SSIM	RMSE
LDCT	/	24.4688 ± 5.2925	0.8246 ± 0.1103	24.6370 ± 9.3343
RED-CNN (2017)	1.9×10^6	28.2974 ± 4.2872	0.8438 ± 0.0859	15.6672 ± 4.8471
DU-GAN (2021)	G:2×10^5 D:1.1×10^8	28.2718 ± 4.6869	0.8444 ± 0.0960	15.4576 ± 5.5962
CTformer (2023)	1.5×10^6	28.7993 ± 4.7875	0.8600 ± 0.0975	14.8606 ± 4.8695
UNAD (2024)	2.1×10^6	29.3488 ± 4.8468	0.8689 ± 0.0969	13.9611 ± 4.7273
AMIR (2024)	2.4×10^7	29.4365 ± 4.7299	0.8676 ± 0.0998	13.8183 ± 4.7711
IDDPM (2021)	1.1×10^8	29.2297 ± 4.7050	0.8651 ± 0.1013	14.1404 ± 4.6463
CoreDiff (2023)	4.8×10^6	29.5106 ± 4.7354	0.8729 ± 0.0957	13.6993 ± 4.7701
Our method	7.5×10^6	**29.5782 ± 4.7279**	**0.8754 ± 0.0963**	**13.5936 ± 4.7862**

higher model accuracy.

4.4. Training details

The model input size is $3 \times 512 \times 512$, containing LDCT images and their spatially adjacent slices. During training, the Adam optimizer was employed with a fixed learning rate of 2×10^{-4}, and the model was trained for 1×10^5 iterations. The total sampling steps T were set to 10. MarCoDiff was implemented and trained using PyTorch with a batch size of 4. The model was constructed on the following hardware configuration: an Intel(R) Xeon(R) Platinum 8352 V CPU @ 2.10 GHz with 128 GB RAM and NVIDIA RTX 4090 GPU (24 GB).

4.5. Loss function

We adopted the mean squared error (MSE) as the loss function for optimization. The objective function mentioned above measures the discrepancy between the model-predicted images and the ground-truth images. The MSE loss directly quantifies this difference, ensuring straightforward implementation and optimization during training. For MarCoDiff's two-stage model, we assigned weights to the loss function to modulate the importance of different loss terms in the overall loss calculation. The formula is defined as follows:

$$L = \alpha_1 \frac{1}{n} \sum_{i=1}^{n} (x_i - \hat{x}_i)^2 + \alpha_2 \frac{1}{n} \sum_{i=1}^{n} (x_i - \hat{\hat{x}}_i)^2 \tag{39}$$

where n is the number of samples, x_i is the NDCT image of the i-th sample, $\hat{x}_i$ is the first-stage predicted image, and $\hat{\hat{x}}_i$ is the second-stage predicted image. Here, α_1 and α_2 correspond to the loss weights for the first and second stages, respectively. Table 2 demonstrate that the model achieves optimal performance when $\alpha_1 = 0.4$ and $\alpha_2 = 0.6$. The training loss curve is shown in Fig. 7.

Table 4
Quantitative comparison of denoising methods under varying Gaussian noise levels (σ = 15,30,45,60) in the image domain of the QIN_LUNG_CT dataset.

Methods	σ = 15		
	PSNR	SSIM	RMSE
LDCT	33.7572 ± 2.1741	0.9285 ± 0.0489	8.3032 ± 2.2384
RED-CNN (2017)	37.7136 ± 2.3097	0.9618 ± 0.0298	5.2627 ± 0.8993
DU-GAN (2021)	36.7960 ± 2.8957	0.9596 ± 0.0547	6.2830 ± 1.1675
CTformer (2023)	38.4959 ± 2.7195	0.9643 ± 0.0328	4.8324 ± 0.9673
UNAD (2024)	38.9834 ± 2.9715	0.9666 ± 0.0317	4.5867 ± 0.9905
AMIR (2024)	38.9928 ± 2.9670	0.9666 ± 0.0309	4.5822 ± 1.0025
IDDPM (2021)	36.5873 ± 3.3286	0.9539 ± 0.0336	5.9990 ± 1.0691
CoreDiff (2023)	38.8808 ± 2.9713	0.9662 ± 0.0316	4.6377 ± 0.9799
Our method	**39.2208 ± 3.0068**	**0.9701 ± 0.0332**	**4.5268 ± 0.9823**
Methods	σ = 30		
	PSNR	SSIM	RMSE
LDCT	27.7579 ± 2.2137	0.8434 ± 0.1061	16.5674 ± 4.5602
RED-CNN (2017)	34.7125 ± 2.3883	0.9295 ± 0.0618	7.4344 ± 1.2904
DU-GAN (2021)	34.4530 ± 2.9479	0.9296 ± 0.0832	8.1991 ± 1.6003
CTformer (2023)	35.0791 ± 2.5125	0.9291 ± 0.0681	7.1399 ± 1.3610
UNAD (2024)	35.9786 ± 2.9330	0.9361 ± 0.0661	6.4736 ± 1.3881
AMIR (2024)	35.6647 ± 2.8689	0.9332 ± 0.0664	6.7068 ± 1.3947
IDDPM (2021)	33.5885 ± 2.4299	0.9184 ± 0.0634	8.4714 ± 1.4516
CoreDiff (2023)	35.8918 ± 2.9675	0.9361 ± 0.0630	6.5409 ± 1.4052
Our method	**36.2234 ± 3.0346**	**0.9384 ± 0.0665**	**6.3009 ± 1.3647**
Methods	σ = 45		
	PSNR	SSIM	RMSE
LDCT	24.2855 ± 2.2022	0.7900 ± 0.1349	24.7148 ± 6.7549
RED-CNN(2017)	33.1159 ± 2.3473	0.9073 ± 0.0829	8.9258 ± 1.5634
DU-GAN(2021)	32.9265 ± 2.7993	0.9102 ± 0.0975	9.7343 ± 1.8771
CTformer(2023)	33.2244 ± 2.1390	0.9043 ± 0.0854	8.8085 ± 1.6423
UNAD(2024)	34.4803 ± 2.7056	0.9172 ± 0.0846	7.6740 ± 1.5765
AMIR(2024)	34.5097 ± 2.6991	0.9175 ± 0.0838	7.6484 ± 1.5986
IDDPM(2021)	32.4929 ± 2.4362	0.9002 ± 0.0820	9.6070 ± 1.7764
CoreDiff(2023)	34.6027 ± 2.7768	0.9196 ± 0.0809	7.5771 ± 1.6189
Our method	**34.8128 ± 2.8358**	**0.9220 ± 0.0805**	**7.3973 ± 1.5985**
Methods	σ = 60		
	PSNR	SSIM	RMSE
LDCT	21.8669 ± 2.2499	0.7559 ± 0.1515	32.6547 ± 9.1488
RED-CNN (2017)	32.1778 ± 2.1699	0.8927 ± 0.0965	9.9345 ± 1.6214
DU-GAN (2021)	31.7187 ± 2.5102	0.8926 ± 0.1153	11.1445 ± 1.9378
CTformer (2023)	31.9632 ± 1.9279	0.8876 ± 0.1005	10.1733 ± 1.7198
UNAD (2024)	33.4121 ± 2.4285	0.9036 ± 0.0956	8.6574 ± 1.5950
AMIR (2024)	33.5731 ± 2.4733	0.9064 ± 0.0917	8.5077 ± 1.6621
IDDPM (2021)	32.0032 ± 2.3480	0.8901 ± 0.0940	10.1569 ± 1.7559
CoreDiff (2023)	33.7224 ± 2.5617	0.9080 ± 0.0920	8.3717 ± 1.6974
Our method	**33.9491 ± 2.6073**	**0.9108 ± 0.0919**	**8.1602 ± 1.6981**

Table 5
Quantitative comparison of denoising methods under varying Poisson levels $(\lambda = 1e^4, 1e^5, 1e^6)$ in the sinogram domain of the QIN_LUNG_CT dataset with constant-level Gaussian noise ($\sigma = 10$).

Methods	$\lambda = 1e^4$		
	PSNR	SSIM	RMSE
LDCT	28.5991 ± 3.0722	0.8509 ± 0.1211	15.2992 ± 4.8211
RED-CNN (2017)	35.4323 ± 3.0040	0.9347 ± 0.0681	6.9182 ± 1.8661
DU-GAN (2021)	34.6524 ± 3.5109	0.9315 ± 0.0939	9.1034 ± 2.3374
CTformer (2023)	35.8062 ± 3.1067	0.9352 ± 0.0720	6.6353 ± 1.8387
UNAD (2024)	36.7946 ± 3.3111	0.9445 ± 0.0622	5.9336 ± 1.6703
AMIR (2024)	37.5671 ± 3.1568	0.9523 ± 0.0504	5.4081 ± 1.4290
IDDPM (2021)	35.3665 ± 3.3511	0.9327 ± 0.0710	7.0111 ± 2.0541
CoreDiff (2023)	38.2704 ± 3.0754	0.9647 ± 0.0364	4.9998 ± 1.3761
Our method	**38.9421 ± 3.2007**	**0.9686 ± 0.0342**	**4.6287 ± 1.2895**
Methods	$\lambda = 1e^5$		
	PSNR	SSIM	RMSE
LDCT	18.8082 ± 2.9803	0.7180 ± 0.1706	47.1601 ± 14.4450
RED-CNN (2017)	31.8666 ± 2.5207	0.8859 ± 0.1091	10.3585 ± 2.5416
DU-GAN (2021)	31.3017 ± 3.1436	0.8789 ± 0.1330	11.7950 ± 3.1761
CTformer (2023)	31.4092 ± 2.4688	0.8763 ± 0.1131	10.9192 ± 2.6681
UNAD (2024)	31.9230 ± 3.3851	0.8798 ± 0.1232	10.5076 ± 3.6027
AMIR (2024)	34.7135 ± 2.7162	0.9350 ± 0.0449	7.3835 ± 1.3499
IDDPM (2021)	32.1754 ± 3.0533	0.8872 ± 0.1154	10.0386 ± 2.4044
CoreDiff (2023)	34.6120 ± 3.5733	0.9367 ± 0.0670	7.5663 ± 1.7694
Our method	**35.8311 ± 3.4510**	**0.9391 ± 0.0210**	**6.2918 ± 1.0874**
Methods	$\lambda = 1e^6$		
	PSNR	SSIM	RMSE
LDCT	12.6303 ± 2.1202	0.6174 ± 0.1561	94.6382 ± 20.2890
RED-CNN (2017)	27.3868 ± 2.0791	0.8216 ± 0.1402	17.2595 ± 3.6674
DU-GAN (2021)	26.5931 ± 1.8440	0.8040 ± 0.1578	19.9064 ± 3.2466
CTformer (2023)	25.7333 ± 3.3623	0.7875 ± 0.1606	21.0923 ± 5.8735
UNAD (2024)	16.3359 ± 2.9448	0.6672 ± 0.1715	62.5012 ± 19.0477
AMIR (2024)	30.8731 ± 2.5537	0.8729 ± 1.220	11.6486 ± 2.7697
IDDPM (2021)	29.1107 ± 2.0075	0.8473 ± 0.1282	14.1550 ± 2.6367
CoreDiff (2023)	31.1011 ± 2.4976	0.8791 ± 0.1200	11.3584 ± 2.7907
Our method	**32.6722 ± 3.8469**	**0.8938 ± 0.1270**	**9.6788 ± 3.1957**

4.6. Experimental results and analysis

(1) Quantitative Evaluation

Table 3 details the performance comparison of different denoising methods on the AAPM-Mayo dataset. Table 4 and Table 5 present the denoising performance of each method under different noise levels introduced in the image domain and sinogram domain of the Qin_LUNG_CT dataset. Benefiting from the effective integration of the Cold diffusion and the multi-scale adaptive residual network, our proposed MarCoDiff achieves the best denoising results across all metrics, outperforming other models. Taking the AAPM-Mayo dataset as an example, compared with UNAD, the SOTA CNN-based model, our method improves PSNR by 0.2294, SSIM by 0.0065, while reducing RMSE by 0.3675. Compared with AMIR, the SOTA Transformer-based model, our method improves PSNR by 0.1417, SSIM by 0.0078, while reducing RMSE by 0.2247. Compared with CoreDiff, the SOTA diffusion-based model, our method improves PSNR by 0.0676 and SSIM by 0.0025, while reducing RMSE by 0.1057. The experimental results show the denoised images from our method have the best perceptual and texture similarity with the NDCT images. Notably, as shown in Fig. 8, as the noise level added to the image domain increases, the proposed method not only leads comprehensively in metrics but also exhibits the slowest rate of performance degradation, indicating that MarCoDiff continuously improves its denoising capability. This further highlights its superior generalization performance in handling high-intensity noisy images.

The AAPM-Mayo dataset has a slice thickness of 1 mm, while the QIN_LUNG_CT dataset has a slice thickness of 5 mm. We further validated the impact of CT slice thickness on model performance when

Table 6
Quantitative comparison of datasets under different slice thicknesses.

Datasets	PSNR	SSIM	RMSE
AAPM-Mayo (single)	29.4092 ± 4.7228	0.8720 ± 0.0944	13.8593 ± 4.7735
AAPM-Mayo (context)	29.5782 ± 4.7279	0.8754 ± 0.0963	13.5936 ± 4.7862
Qin_LUNG_CT σ = 15 (single)	39.1785 ± 3.0463	0.9697 ± 0.0331	4.5467 ± 0.9685
Qin_LUNG_CT σ = 15 (context)	39.2208 ± 3.0068	0.9701 ± 0.0332	4.5268 ± 0.9823
Qin_LUNG_CT σ = 30 (single)	36.1675 ± 3.0207	0.9378 ± 0.0642	6.3360 ± 1.3344
Qin_LUNG_CT σ = 30 (context)	36.2234 ± 3.0346	0.9388 ± 0.0665	6.3009 ± 1.3647
Qin_LUNG_CT σ = 45 (single)	34.7422 ± 2.8098	0.9212 ± 0.0829	7.4540 ± 1.5389
Qin_LUNG_CT σ = 45 (context)	34.8128 ± 2.8358	0.9220 ± 0.0805	7.3973 ± 1.5985
Qin_LUNG_CT σ = 60 (single)	33.8332 ± 2.5842	0.9098 ± 0.0921	8.2589 ± 1.5940
Qin_LUNG_CT σ = 60 (context)	33.9491 ± 2.6073	0.9108 ± 0.0919	8.1602 ± 1.6981

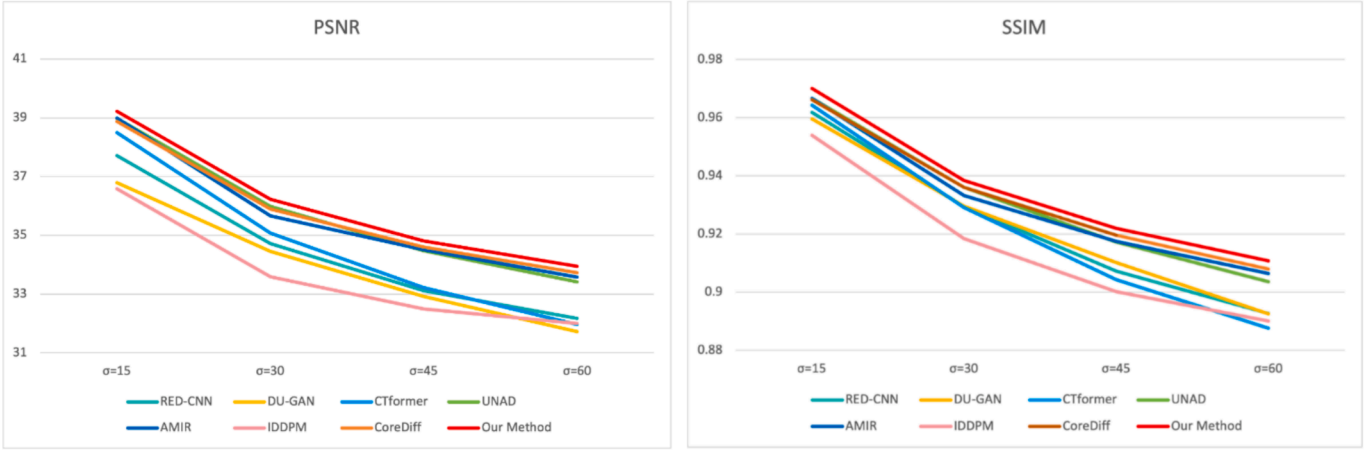


Fig. 8. PSNR and SSIM of denoising methods under varying Gaussian levels in the image domain of the Qin_LUNG_CT dataset.

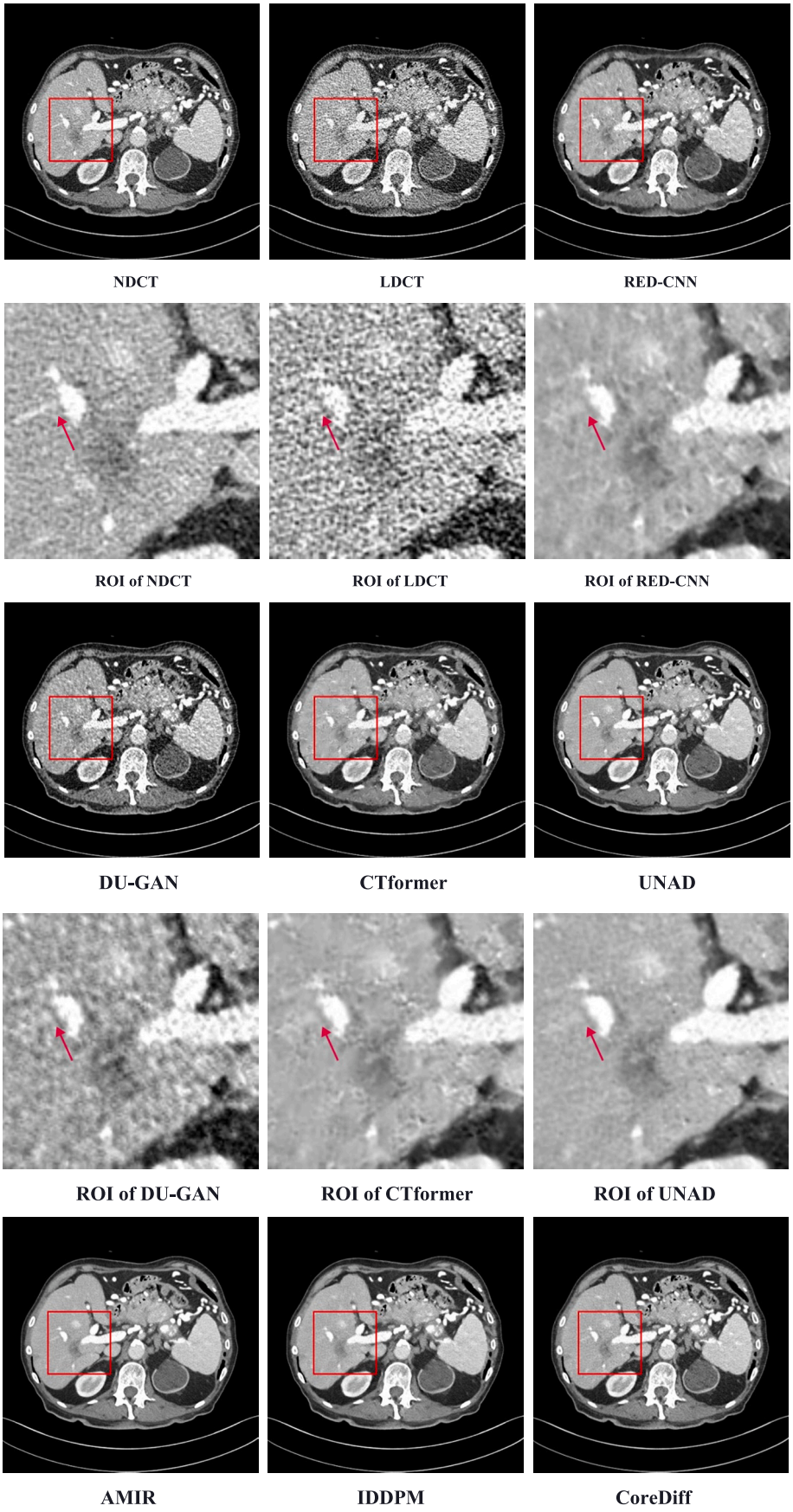


Fig. 9. Comparison of denoising methods on the AAPM-Mayo dataset.

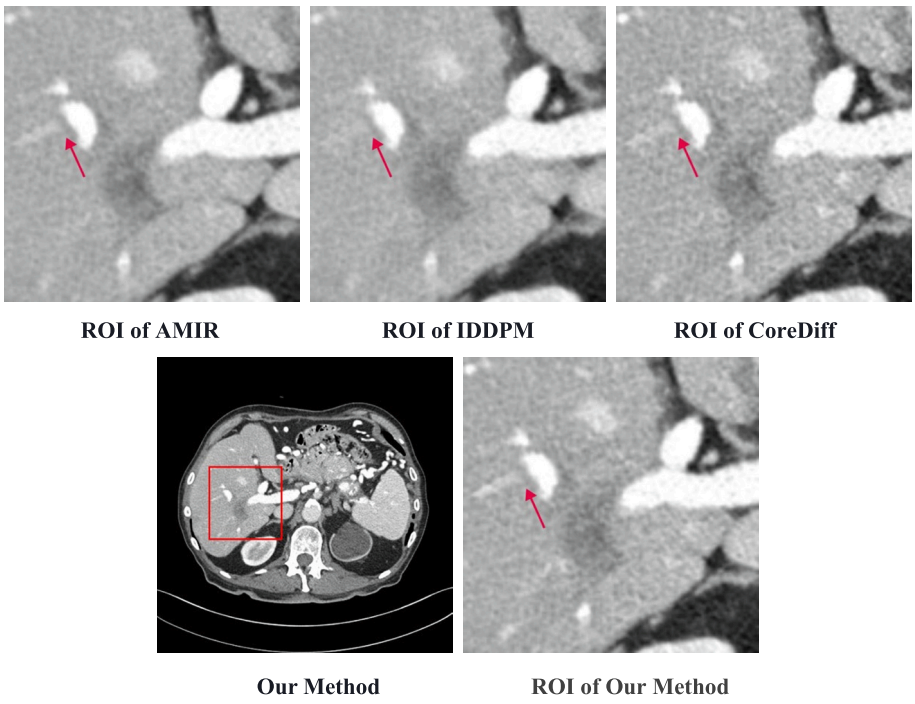


Fig. 9. (*continued*).

leveraging the context of three adjacent slices for modeling. As shown in Table 6, comparative experiments on the AAPM-Mayo dataset demonstrated improvements of 0.57 % in PSNR and 0.39 % in SSIM, along with a reduction of 1.92 % in RMSE, when using context modeling with adjacent slices compared to using a single input slice. For the QIN_LUNG_CT dataset, under optimal conditions ($\sigma = 60$), PSNR and SSIM increased by only 0.34 % and 0.11 % respectively, while RMSE decreased by 1.21 %. Our experiments confirm that thinner slices, compared to thicker ones, enable better enhancement of denoising effects when leveraging contextual information from adjacent slices. Thicker slices exhibit weaker spatial correlation between them, thus constraining the improvement in denoising performance.

Experimental results demonstrate that compared to existing denoising methods, our approach achieves significant improvements across all metrics. It exhibits better performance and generalization capabilities when addressing denoising tasks on different datasets.

(2) Qualitative Evaluation

In this section, we visually analyze representative denoising results based on the characteristics of LDCT images. All CT images are displayed under a window of [-160, 240] HU.

Fig. 9 presents qualitative results of various denoising methods on the AAPM-Mayo dataset. To comprehensively demonstrate the local detail denoising capabilities and structural restoration performance of different approaches, the region of interest (ROI) marked by red rectangular boxes is enlarged. RED-CNN, employing an encoder-decoder architecture, effectively removes noise in the LDCT image but oversuppresses noise, resulting in overly smooth and blurred outputs. DU-GAN, based on GAN with adversarial training and a U-Net-based discriminator, avoids over-smoothing and improves visual fidelity but introduces detail distortion. CTformer leverages Transformer mechanisms for global structural modeling and detail recovery but retains residual speckle noise. UNAD enhances noise region prediction through a distribution regression layer and incorporates prior knowledge via predictive pre-training to exploit inter-slice correlations in CT images, achieving strong denoising effects, albeit with slight edge blurring. AMIR augments Restormer with routing modules to improve spatial and channel-wise denoising, but produces more over-smoothing in fine tissues compared to NDCT image. IDDPM introduces learnable variances to reduce sampling steps, recovering most structures but lacking detail richness.

CoreDiff also adopts a Cold diffusion and addresses error accumulation and misalignment between the input image and time step embedding features during reverse processes via a contextual error-modulated restoration network, achieving competitive denoising results. Its outputs demonstrate clear overall structures, confirming the superiority of Cold diffusion. However, compared to all methods, MarCoDiff delivers the highest denoised image quality.

As shown in Fig. 9, the results in the ROI area show that the outcomes of our method are more distinct at the blood vessels (indicated by the red arrow), with the highest naturalness of the vascular outline. Due to having structural features closer to those of the NDCT images, our method can assist doctors in understanding the morphology and course of the blood vessels, which is helpful in detecting whether there are any abnormal issues such as thrombosis. In the results of our method, the hypodense lesion tissues in the black area (indicated by the blue arrow) have the clearest boundaries and internal textures, which is conducive to the differentiation of hypodense abnormal tissues and reduces diagnostic uncertainty.

These improvements provide radiologists with a high-quality imaging basis that helps enhance medical efficiency. Their application value is not limited to lesion detection but can also extend to multiple clinical scenarios such as vascular assessment and surgical planning. It indicates that our denoising output has the feasibility of application in the clinical workflow. This is attributed to our multi-scale adaptive residual network, which dynamically refines global structures and local details from coarse to fine and adaptively adjusts denoising strategies.

When handling Gaussian or mixed noise of different levels, our model exhibits the strongest generalization. Visually, other methods retain noise, artifacts and structural blurring, and still exhibit certain shortcomings in the restoration of local details, while our model achieves the best visual outcomes.

Under the highest noise level ($\sigma = 60$), Fig. 10 presents the comparative results of methods applied to the image domain of the Qin_LUNG_CT dataset. Fig. 12 shows the denoising results of methods applied to the sinogram domain of the Qin_LUNG_CT dataset under a Poisson noise intensity of $\lambda = 1 \times 10^4$.

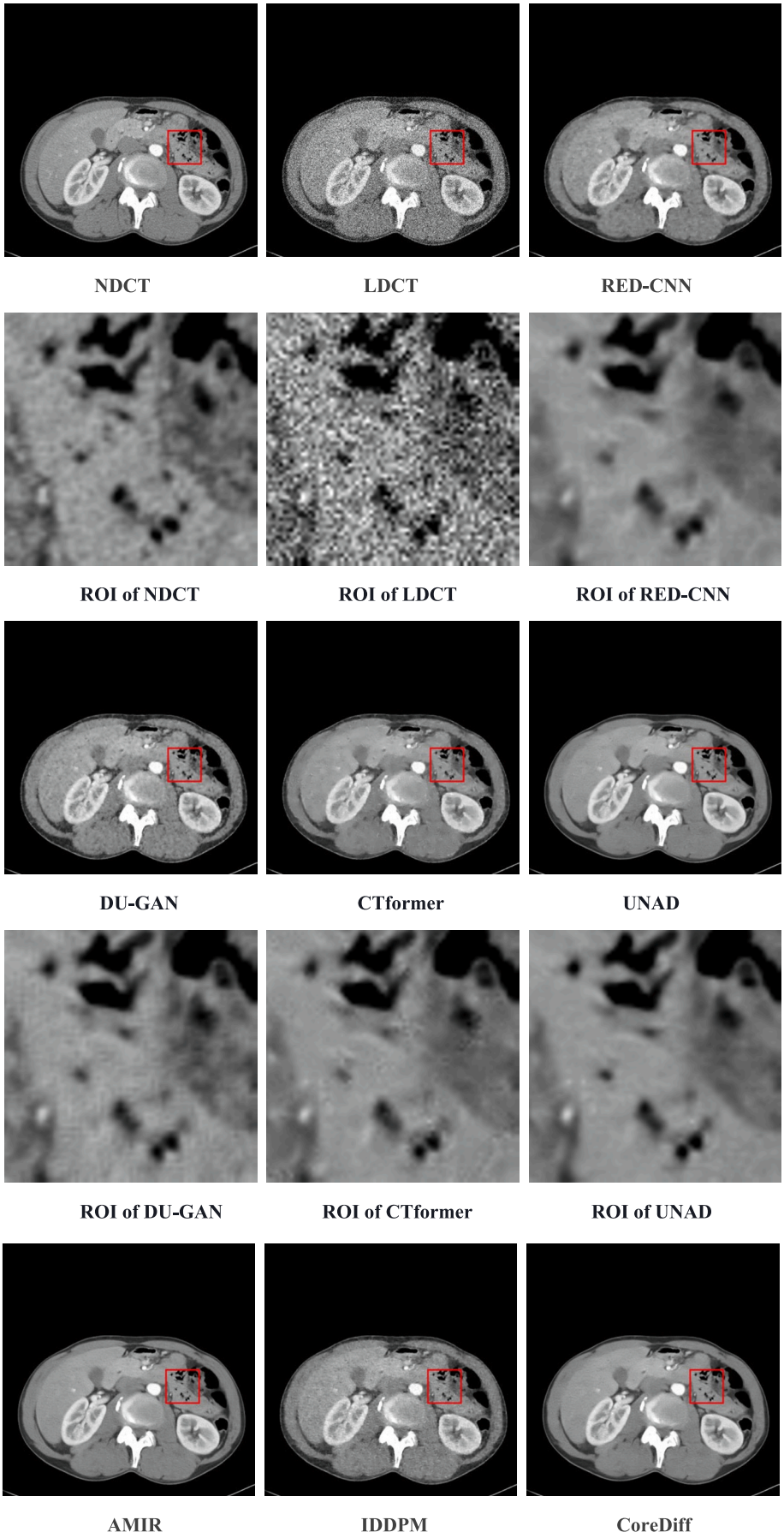


Fig. 10. Comparison of denoising methods for images with noise in the image domain of the Qin_LUNG_CT dataset.

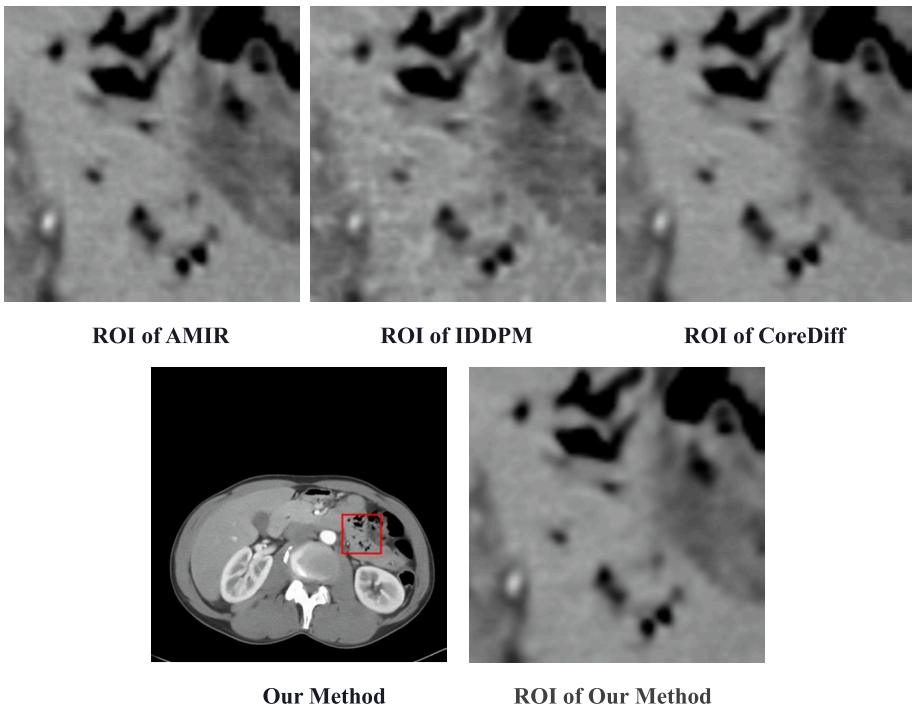


Fig. 10. (*continued*).

As shown in Fig. 11, we performed differential processing on the denoised image and the NDCT image. The denser and more obvious the white pixels in the differential image are, the greater the difference between the denoised image and the NDCT image is. For the edge area (indicated by the red arrow), the result of our method shows that the white pixels are invisible, while other methods exist to some extent. For the internal organizational area (indicated by the blue arrow), even though there are still some white pixels in the result of our method, it is the least obvious compared with other comparison methods. The differential images confirm that our method has the minimum prediction deviation, is closer to the NDCT images, and has the best denoising ability.

In summary, our proposed method MarCoDiff not only significantly enhances the denoising performance of LDCT images compared to existing approaches but also generates images with textural details that more closely resemble those of NDCT images, while demonstrating superior generalization capability when confronted with unknown noise patterns.

Remark 1 *Complexity analysis and computational overhead discussion of MarCoDiff.*

AFAM employs Fast Fourier Transform (FFT) operation, with the computational cost linearly related to the resolution of the feature map. AFAM contains fewer convolutional layers and has relatively fewer parameters. Moreover, we only apply it in the second stage, which enhances the computational efficiency. MRAB, due to the dynamic convolution kernel generation, has a complexity that grows quadratically with the resolution. HCB operates on low-resolution feature maps, and dilated convolution expands the receptive field while keeping the computational cost almost unchanged. Although the parameter count of a single network is not large, the two-stage denoising strategy does increase the parameter count. On the AAPM-Mayo dataset, compared to CoreDiff's parameter count (4.8 *M*), *although MarCoDiff's parameter count has increased* (7.5 *M*), *MarCoDiff's FLOPs and sampling speed are* 184.9*G and* 0.22 s, *respectively, which are superior to CoreDiff's* 327.3*G and* 0.24 s. *Our method has higher computational efficiency, requiring less actual computation for denoising inference. MarCoDiff also shows improved PSNR, SSIM, and RMSE performances* (Table 3). *In the visual comparison, the denoised images from MarCoDiff are clearer in the vascular and lesion areas* (Fig. 9). *This slight increase in parameter count due to the improved denoising effect is acceptable.*

4.7. Ablation Experiment

This section conducts an ablation study to evaluate the impact of each component in MarCoDiff on denoising performance, with experimental validation performed on the AAPM-Mayo dataset. Before incorporating the CMPD operator, we employed a mean-preserving degradation operator with linear scheduling strategy as the degradation method for cold diffusion. The comparative experimental results are shown in Table 7.

First, when using only the MRAB in the denoising network, the results demonstrate that MRAB achieves reconstruction of multi-scale features from coarse to fine during upsampling, where the MDC and ADC ensure denoising effectiveness, verifying the efficacy of this module. Subsequently, we independently validated the AFAM by introducing it exclusively in the second stage of denoising. Compared with MRAB, AFAM improved the PSNR and SSIM by 0.0774 and 0.0020, respectively, while reducing RMSE by 0.1155, thereby contributing more significantly to the final performance. Results indicate that this method can more effectively leverage time step to adaptively adjust the fusion of different frequency components of features. We further examined the combined effect of both MRAB and AFAM. As expected, their joint use improved the PSNR and SSIM by 0.1820 and 0.0041, respectively with respect to AFAM, and additionally brought about a reduction in the RMSE by 0.2826. In addition, we incorporated the HCB to enhance model performance (PSNR and SSIM increased by 0.1012 and 0.0018, respectively, while RMSE decreased by 0.1600) without further reducing resolution.

Finally, we replaced the linear-scheduled mean-preserving degradation operator with CMPD. Although there is a slight improvement in the metrics，what matters more visually, as shown in Fig. 13, is that the denoised image integrating CMPD exhibits sharper outlines at the blood vessels' tips (red arrow). Furthermore, by changing the sampling step size T, the robustness of the proposed method to T was evaluated. The results are shown in Table 8. Essentially, this is to verify the sampling

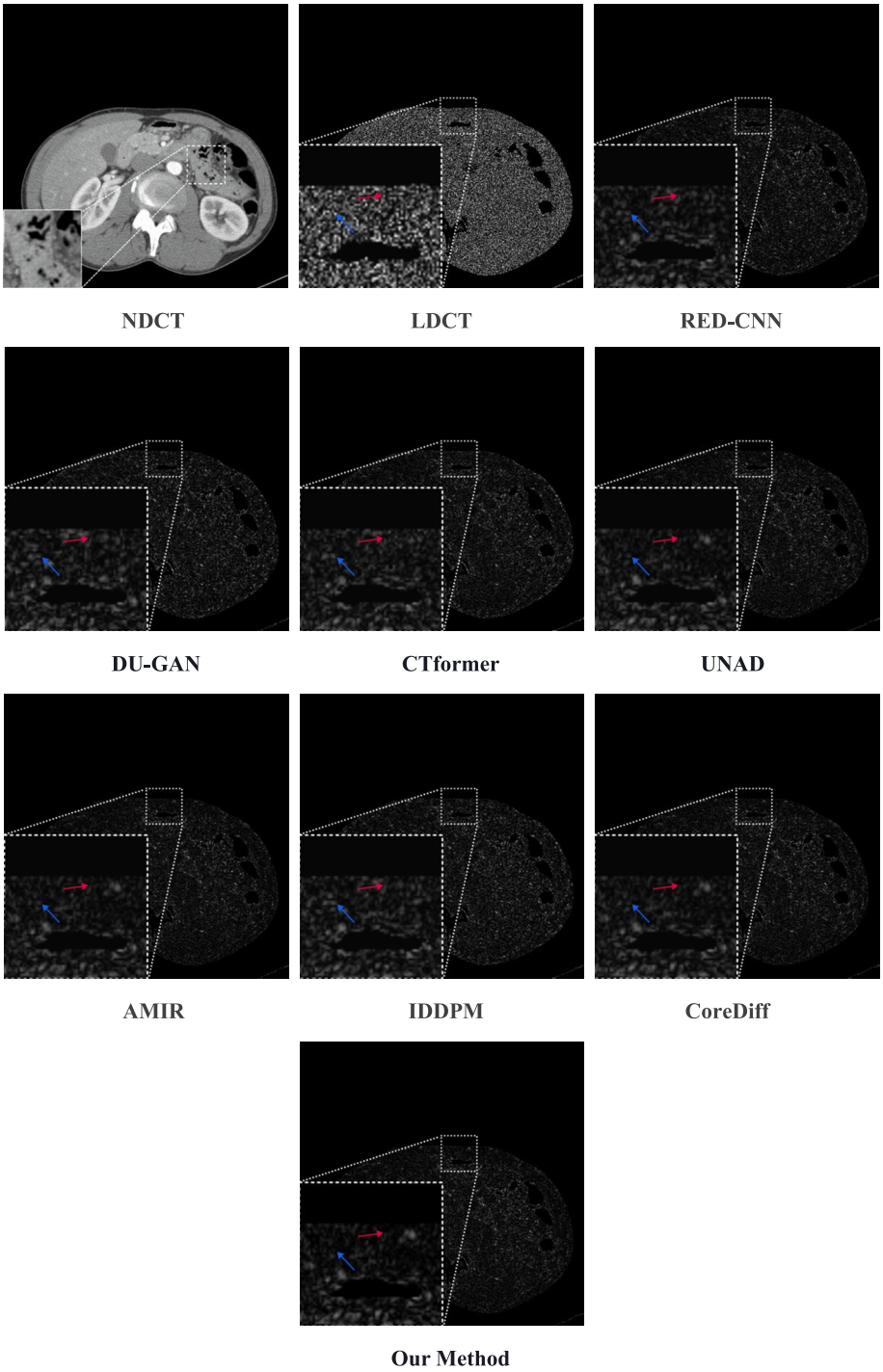


Fig. 11. Difference images between NDCT image and denoised images for denoising methods in Fig. 10.

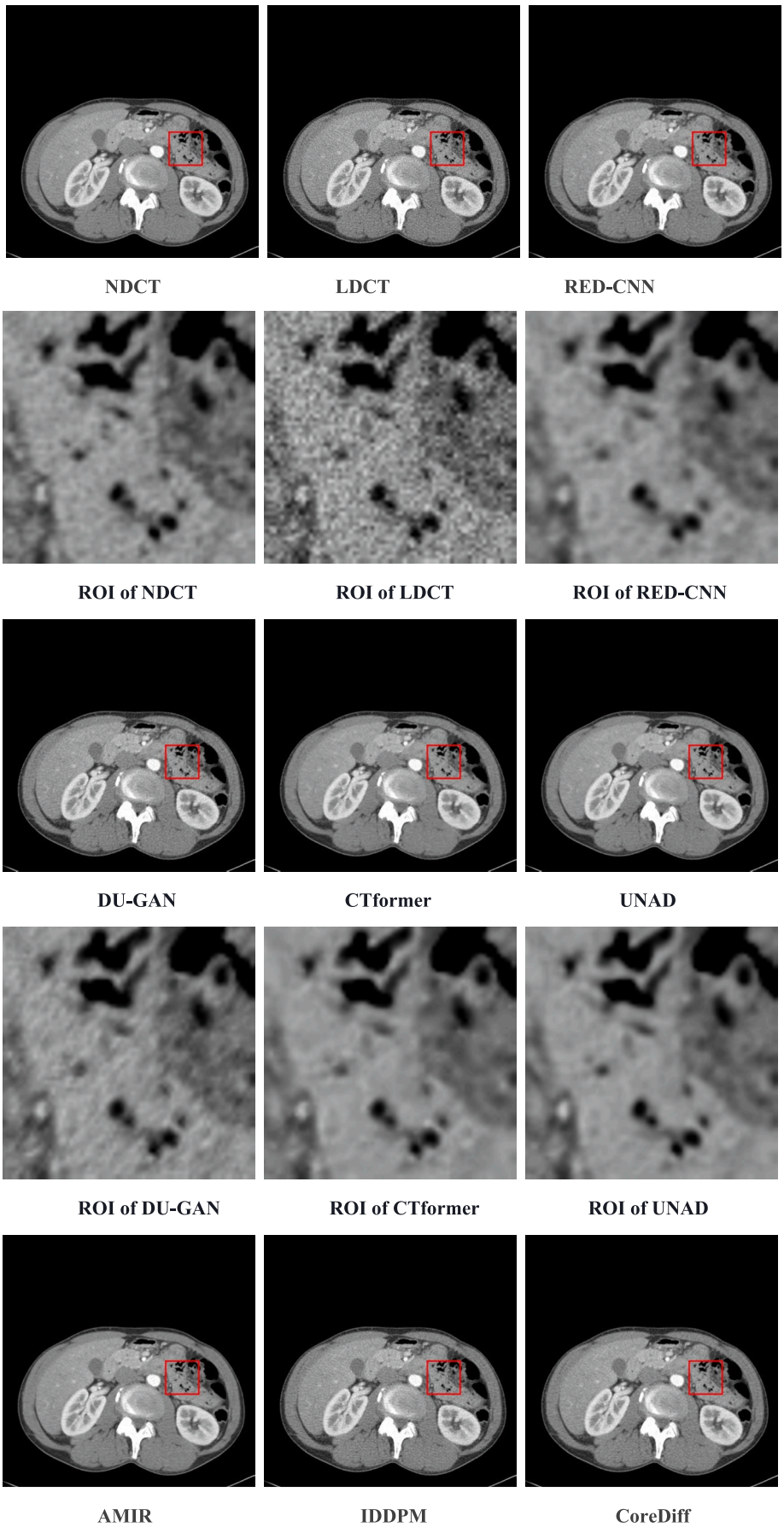


Fig. 12. Comparison of denoising methods for images with noise in the sinogram domain of the Qin_LUNG_CT dataset.

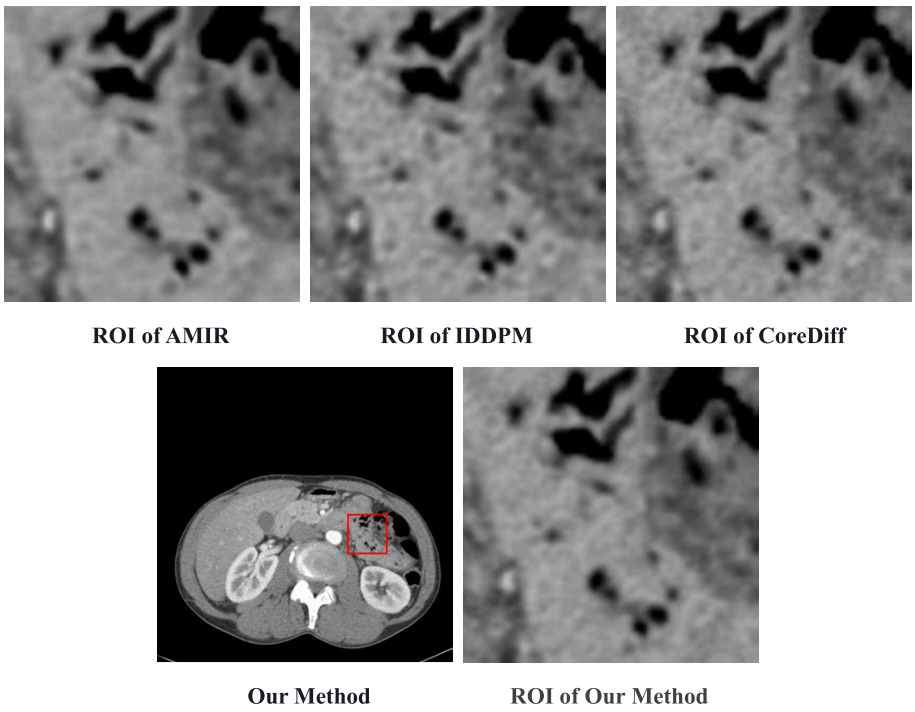


Fig. 12. (*continued*).

Table 7
Ablation experiments.

Methods	PSNR	SSIM	RMSE
MRAB	29.1780 ± 4.6488	0.8672 ± 0.0933	14.2135 ± 4.8192
AFAM	29.2554 ± 4.7107	0.8692 ± 0.0945	14.0980 ± 4.8154
MRAB + AFAM	29.4374 ± 4.7479	0.8733 ± 0.0963	13.8154 ± 4.8396
MRAB + AFAM + HCB	29.5386 ± 4.7192	0.8751 ± 0.0965	13.6554 ± 4.7912
MRAB + AFAM + HCB + CMPD	29.5782 ± 4.7279	0.8754 ± 0.0963	13.5936 ± 4.7862

stability of the intermediate steps of the cosine scheduling strategy. The selection of T needs to take into account both the quality of image denoising and the efficiency of computation. As shown in Table 8, compared with the linear scheduling strategy, our cosine scheduling strategy has smaller output differences under different sampling step sizes T. The model is insensitive to the change in step size T, which to some extent verifies the robustness and stability of CMPD. The results demonstrate that our CMPD exhibits advantages in modeling noise distributions and achieves superior denoising performance under identical network architectures.

The experimental results confirm that adding any component to the baseline improves the model's denoising performance, with combined components outperforming individual implementations. The model achieves optimal denoising performance when all four components are employed. These findings validate the effectiveness of the proposed components in enhancing the model's denoising capabilities.

5. Conclusions

This paper proposes a novel Multi-scale Adaptive Residual Cold Diffusion Model (MarCoDiff) for LDCT image denoising. The cosine mean-preserving degradation (CMPD) operator effectively simulates the physical degradation process of CT images while enhancing sampling efficiency. The cosine degradation strategy emphasizes the learning of intermediate time steps, preserving image details and improving the stability of reverse process sampling. We introduce three key components—the adaptive frequency adjustment module (AFAM), multi-scale residual adaptive block (MRAB) and hybrid-dilated-convolution context block (HCB)—into the designed multi-scale adaptive residual network. This architecture effectively suppresses noise while maintaining sharp image details. Experimental results on the AAPM-Mayo and Qin_LUNG_CT datasets demonstrate that MarCoDiff outperforms existing methods in both qualitative and quantitative evaluations, exhibiting strong generalization capabilities under varying noise levels. The model provides superior LDCT denoising performance, enhancing the reliability of clinical diagnostic assistance.

Although MarCoDiff achieves remarkable results in LDCT image denoising, several limitations persist. First, the relatively limited size of medical image datasets may restrict comprehensive coverage of diverse scenarios involving different anatomical regions and pathological types, thereby constraining generalization potential. Second, the inherent large parameter size of diffusion models affects computational efficiency, and the two-stage architecture may limit practical applications in resource-constrained environments. Future work will focus on deeper optimization of MarCoDiff. We plan to explore denoising methods without paired data to reduce dependency on aligned image pairs and improve real-world generalization. Concurrently, we aim to refine the model algorithm by investigating efficient, unified network architectures to reduce parameter complexity. These advancements are expected to further enhance CT image quality, providing more accurate and reliable foundations for medical diagnosis.

CRediT authorship contribution statement

Ju Zhang: Conceptualization, Methodology, Project administration, Funding acquisition. **Guangyu Liu:** Investigation, Methodology, Software, Writing – original draft. **Jikang Chen:** Software, Visualization, Validation. **Yun Cheng:** Writing – review & editing, Validation, Funding acquisition.

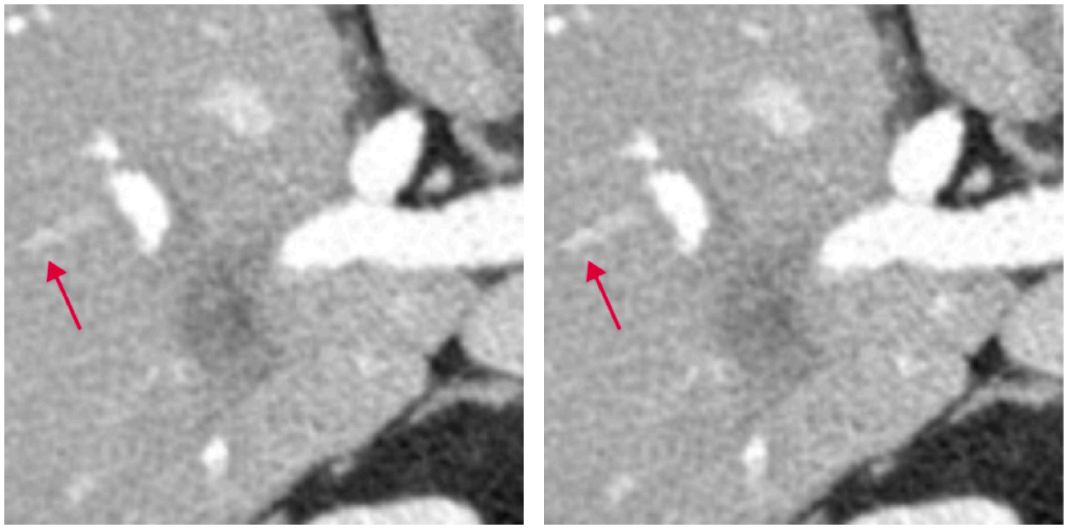


Fig. 13. Comparison between the linear scheduling and CMPD on the AAPM-Mayo dataset.

Table 8
Scheduling strategies with different T on the AAPM-Mayo.

Scheduling Strategies	PSNR	RMSE
T = 8		
linear	29.4364	13.8212
cosine	29.5413	13.6534
T = 9		
linear	29.5024	13.7142
cosine	29.5713	13.6048
T = 10		
linear	29.5386	13.6554
cosine	29.5782	13.5936

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This work was supported by the National Natural Science Foundation of China grants 60974042 and Interdisciplinary Research Project of Hangzhou Normal University grant 2024JCXK03.

Data availability

Data will be made available on request.

References

Goldman, L. W. (2007). Principles of CT: Radiation dose and image quality. *Journal of Nuclear Medicine Technology, 35*(4), 213–225.
Pan, X., Sidky, E. Y., & Vannier, M. (2009). Why do commercial CT scanners still employ traditional, filtered back-projection for image reconstruction ? . *Inverse Problems, 25*, Article 123009.
Wang, J., Li, T., Lu, H., et al. (2006). A penalized weighted least-squares approach to sinogram noise reduction and image reconstruction for low-dose X-ray computed tomography. *IEEE Transactions on Medical Imaging, 25*, 1272.
Manduca, A., Yu, L., Trzasko, J. D., et al. (2009). Projection space denoising with bilateral filtering and CT noise modeling for dose reduction in CT. *Medical Physics, 36*, 4911.
Kaur, R., Juneja, M., & Mandal, A. K. (2018). A comprehensive review of denoising techniques for abdominal CT images. *Multimedia Tools and Applications, 77*(17), 22735–22770.
Manduca, A., Yu, L., Trzasko, J. D., et al. (2009). Projection space denoising with bilateral filtering and CT noise modeling for dose reduction in CT. *Medical Physics, 36* (11), 4911–4919.
Xu, Q., Yu, H., Mou, X., et al. (2012). Low-dose X-ray CT reconstruction via dictionary learning. *IEEE Transactions on Medical Imaging, 31*, 1682.
Zheng, X., Ravishankar, S., Long, Y., et al. (2018). PWLS-ULTRA: An efficient clustering and learning-based approach for low-dose 3D CT image reconstruction. *IEEE Transactions on Medical Imaging, 37*, 1498.
Liu, Y., Ma, J., Fan, Y., et al. (2012). Adaptive-weighted total variation minimization for sparse data toward low-dose x-ray computed tomography image reconstruction. *Physics in Medicine and Biology, 57*(23), 7923.
Ma, J., Zhang, H., Gao, Y., et al. (2012). Iterative image reconstruction for cerebral perfusion CT using a pre-contrast scan induced edge-preserving prior. *Physics in Medicine and Biology, 57*(22), 7519.
Li, Z., Yu, L., Trzasko, J. D., et al. (2014). Adaptive nonlocal means filtering based on local noise level for CT denoising. *Medical Physics, 41*(1), Article 011908.
Sheng, K., Gou, S., Wu, J., et al. (2014). Denoised and texture-enhanced MVCT to improve soft tissue conspicuity. *Medical Physics, 41*(10), Article 101916.
Chen, H., Zhang, Y., Kalra, M. K., et al. (2017). Low-dose CT with a residual encoder-decoder convolutional neural network. *IEEE Transactions on Medical Imaging, 36*, 2524.
Liang T, Jin Y, Li Y, et al. Edcnn: Edge enhancement-based densely connected network with compound loss for low-dose ct denoising[C]//2020 15th IEEE International conference on signal processing (ICSP). IEEE, 2020, 1: 193-198.
Mazandarani F N, Babyn P, Alirezaie J. UNeXt: a Low-Dose CT denoising UNet model with the modified ConvNeXt block[C]//ICASSP 2023-2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). IEEE, 2023: 1-5.
Zhou J , Jampani V , Pi Z , et al. Decoupled dynamic filter networks [C] // Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Nashville: IEEE , 2021: 6647.
Qiao, Z., & Du, C. (2022). Rad-unet: A residual, attention-based, dense unet for CT sparse reconstruction. *Journal of Digital Imaging, 35*(6), 1748–1758.
Wu, T., Li, P., Sun, J., et al. (2024). Adaptive edge prior-based deep attention residual network for low-dose CT image denoising. *Biomedical Signal Processing and Control, 98*, Article 106773.
Lu, Z., Xia, W., Huang, Y., Hou, M., Chen, H., Zhou, J., Shan, H., & Zhang, Y. (2023). M3NAS: Multi-scale and multi-level memory-efficient neural architecture search for low-dose CT denoising. *IEEE Transactions on Medical Imaging, 42*(3), 850–863.
Rombach, R., Blattmann, A., Lorenz, D., et al. (2022:). High-resolution image synthesis with latent diffusion models. *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition.*, 10684–10695.
Xia, B., Zhang, Y., Wang, S., et al. (2023:). Diffir: Efficient diffusion model for image restoration. *Proceedings of the IEEE/CVF International Conference on Computer Vision.*, 13095–13105.
Saharia, C., Ho, J., Chan, W., et al. (2022). Image super-resolution via iterative refinement. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 45*(4), 4713–4726.
Shuai Z, Chen Y, Mao S, et al. DiffSeg: A Segmentation Model for Skin Lesions Based on Diffusion Difference. arXiv preprint arXiv:2404.16474, 2024.
Güngör, A., Dar, S. U. H., Öztürk, Ş., et al. (2023). Adaptive diffusion priors for accelerated MRI reconstruction. *Medical Image Analysis, 88*, Article 102872.

Xia W, Lyu Q, Wang G. Low-Dose CT Using Denoising Diffusion Probabilistic Model for 20×Speedup. arXiv:2209.15136, 2022.

J. Ho, A. Jain, and P. Abbeel, Denoising diffusion probabilistic models, in Proc. Adv. Neural Inf. Process. Syst., vol. 33, 2020, pp. 6840–6851.

A. Nichol and P. Dhariwal, Improved denoising diffusion probabilistic models, in Proc. 38th Int. Conf. Mach. Learn., 2021, pp. 8162–8171.

Bansal, A., Borgnia, E., Chu, H. M., et al. (2024). Cold diffusion: Inverting arbitrary image transforms without noise. *Advances in Neural Information Processing Systems, 36*.

Zhang, K., Zuo, W., Chen, Y., et al. (2017). Beyond a Gaussian denoiser: Residual learning of deep CNN for image denoising. *IEEE Transactions on Image Processing, 26*(7), 3142–3155.

Guo, S., Yan, Z., Zhang, K., et al. (2019:). Toward convolutional blind denoising of real photographs. *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition.*, 1712–1722.

Anwar S, Barnes N. Real image denoising with feature attention[C]//Proceedings of the IEEE/CVF international conference on computer vision. 2019: 3155-3164.

Chen, L., Chu, X., Zhang, X., et al. (2022). *Simple baselines for image restoration[C]// European conference on computer vision* (pp. 17–33). Cham: Springer Nature Switzerland.

Gu L, Deng W, Wang G. UNAD: Universal Anatomy-Initialized Noise Distribution Learning Framework Towards Low-Dose CT Denoising[C]//ICASSP 2024-2024 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). IEEE, 2024: 1671-1675.

Wang, D., Fan, F., Wu, Z., et al. (2023). CTformer: Convolution-free Token2Token dilated vision transformer for low-dose CT denoising. *Physics in Medicine and Biology, 68*(6), Article 065012.

Yang, Z., Chen, H., Qian, Z., et al. (2024). *All-in-one medical image restoration via task-adaptive routing[C]//International Conference on Medical image Computing and Computer-Assisted intervention* (pp. 67–77). Cham: Springer Nature Switzerland.

Goodfellow, I., Pouget-Abadie, J., Mirza, M., et al. (2020). Generative adversarial networks. *Communications of the ACM, 63*(11), 139–144.

Yang, Q., Yan, P., Zhang, Y., et al. (2018). Low-dose CT image denoising using a generative adversarial network with Wasserstein distance and perceptual loss. *IEEE Transactions on Medical Imaging, 37*(6), 1348–1357.

Huang, Z., Zhang, J., Zhang, Y., et al. (2021). DU-GAN: Generative adversarial networks with dual-domain U-Net-based discriminators for low-dose CT denoising. *IEEE Transactions on Instrumentation and Measurement, 71*, 1–12.

Song J, Meng C, Ermon S. Denoising diffusion implicit models[J]. arXiv:2010.02502, 2020.

Saharia, C., Ho, J., Chan, W., et al. (2022). Image Super-Resolution Via Iterative Refinement. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 3204461*.

Zhu, Y., & Chen, Y. (2024). A new diffusion method for blind image denoising. *Pattern Analysis and Applications, 27*(4), 124.

Chen, J., & Shen, Y. (2024). Memorizing Swin-Transformer Denoising Network for Diffusion Model. *Electronics, 13*(20), 4050.

Gao, Q., & Shan, H. (2022). CoCoDiff: A contextual conditional diffusion model for low-dose CT image denoising[C]//Developments in X-Ray Tomography XIV. *SPIE, 12242*, 92–98.

Gao, Q., Li, Z., Zhang, J., et al. (2023). CoreDiff: Contextual error-modulated generalized diffusion model for low-dose CT denoising and generalization. *IEEE Transactions on Medical Imaging*.

Du, Y., Liu, Y., Wu, H., et al. (2025). Combination of edge enhancement and cold diffusion model for low dose CT image denoising. *Biomedical Engineering/ Biomedizinische Technik, 70*(2), 157–169.

Shan, H., et al. (2018). 3-D convolutional encoder–decoder network for low-dose CT via transfer learning from a 2-D trained network. *IEEE Trans. Med. Imag., 37*(6), 1522–1534.

He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. In *In: Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 770–778).

Maas, A. L., Hannun, A. Y., & Ng, A. Y. (2013). Rectifier nonlinearities improve neural net-work acoustic models. *In: Proc. icml., 30*, 3.

Wang X, Chai L, Chen J. Frequency-Domain Refinement with Multiscale Diffusion for Super Resolution. arXiv:2405.10014, 2024.

Perez E, Strub F, De Vries H, et al. Film: Visual reasoning with a general conditioning layer[C]//Proceedings of the AAAI conference on artificial intelligence. 2018, 32(1).

Zhang, X., Han, Z., Shangguan, H., et al. (2021). Artifact and detail attention generative adversarial networks for low-dose CT denoising. *IEEE Transactions on Medical Imaging, 40*, 3901.

Li Y, Chen Y, Dai X, et al. Revisiting dynamic convolution via matrix decomposition. arXiv:2103.08756, 2021.