Computerized Medical Imaging and Graphics 120 (2025) 102491

Contents lists available at ScienceDirect

Computerized Medical Imaging and Graphics

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

Convergent–Diffusion Denoising Model for multi-scenario CT Image Reconstruction

Xinghua Ma [a,b], Mingye Zou [a], Xinyan Fang [a], Gongning Luo [a,b], Wei Wang [c,*], Suyu Dong [d,*], Xiangyu Li [a,*], Kuanquan Wang [a], Qing Dong [e], Ye Tian [f], Shuo Li [g,h]

[a] The Faculty of Computing, Harbin Institute of Technology, Harbin, Heilongjiang, China
[b] The Computational Bioscience Research Center, King Abdullah University of Science and Technology, Thuwal, Makkah, Saudi Arabia
[c] The Faculty of Computing, Harbin Institute of Technology, Shenzhen, Guangdong, China
[d] The College of Computer and Control Engineering, Northeast Forestry University, Harbin, Heilongjiang, China
[e] The Department of Thoracic Surgery at No. 4 Affiliated Hospital, Harbin Medical University, Harbin, Heilongjiang, China
[f] The Department of Cardiology at No. 1 Affiliated Hospital, Harbin Medical University, Harbin, Heilongjiang, China
[g] The Department of Computer and Data Science, Case Western Reserve University, Cleveland, OH, USA
[h] The Department of Biomedical Engineering, Case Western Reserve University, Cleveland, OH, USA

ARTICLE INFO

Keywords:
Image reconstruction
Diffusion-based model
Multi-scenario
Dual-domain
Sinogram
Low-dose CT
Sparse-view CT
Metal artifact

ABSTRACT

A generic and versatile CT Image Reconstruction (CTIR) scheme can efficiently mitigate imaging noise resulting from inherent physical limitations, substantially bolstering the dependability of CT imaging diagnostics across a wider spectrum of patient cases. Current CTIR techniques often concentrate on distinct areas such as Low-Dose CT denoising (LDCTD), Sparse-View CT reconstruction (SVCTR), and Metal Artifact Reduction (MAR). Nevertheless, due to the intricate nature of multi-scenario CTIR, these techniques frequently narrow their focus to specific tasks, resulting in limited generalization capabilities for diverse scenarios. We propose a novel Convergent–Diffusion Denoising Model (CDDM) for multi-scenario CTIR, which utilizes a stepwise denoising process to converge toward an imaging-noise-free image with high generalization. CDDM uses a diffusion-based process based on a priori decay distribution to steadily correct imaging noise, thus avoiding the overfitting of individual samples. Within CDDM, a domain-correlated sampling network (DS-Net) provides an innovative sinogram-guided noise prediction scheme to leverage both image and sinogram (*i.e.*, dual-domain) information. DS-Net analyzes the correlation of the dual-domain representations for sampling the noise distribution, introducing sinogram semantics to avoid secondary artifacts. Experimental results validate the practical applicability of our scheme across various CTIR scenarios, including LDCTD, MAR, and SVCTR, with the support of sinogram knowledge.

1. Introduction

Exploring multi-scenario CT image reconstruction (CTIR) has significant practical significance and considerable application prospects in medical systems. While CT is extensively employed as a medical examination, its imaging efficacy can be limited by the underlying physics, potentially reducing expected clinical outcomes (Benedict et al., 2010; Kazerouni et al., 2023). Imaging techniques utilizing lower radiation doses or sparser views have the potential to mitigate radiological risks of widespread public concern and broaden clinical applications (Zhang and Sejdić, 2019). However, adopting low-dose CT (LDCT) or sparse-view CT (SVCT) approaches can lead to significant degradation in the reconstructed image quality, thereby impacting diagnostic accuracy (Humphries et al., 2019). Moreover, the existence of metal implants in patients, such as artificial hip joints and spinal implants, can introduce undesirable streaking and shadowing artifacts during imaging process (Zhou et al., 2022b). The aforementioned scenarios will severely degrade the image quality, thus posing a major challenge for physicians to accurately diagnose lesions, as shown in Fig. 1. A robust and adaptable CTIR scheme facilitates dependable CT imaging diagnoses for a broader spectrum of patients, fostering the integration of a cohesive and intelligent healthcare ecosystem. Concurrently, harnessing distinct image information across various CTIR scenarios potentially enhances the performance of image reconstruction, thereby providing a more robust foundation for CT imaging diagnostics.

* Corresponding authors.
E-mail addresses: wangwei2019@hit.edu.cn (W. Wang), dongsuyu@nefu.edu.cn (S. Dong), lixiangyu@hit.edu.cn (X. Li).

https://doi.org/10.1016/j.compmedimag.2024.102491
Received 3 April 2024; Received in revised form 27 October 2024; Accepted 31 December 2024
Available online 4 January 2025

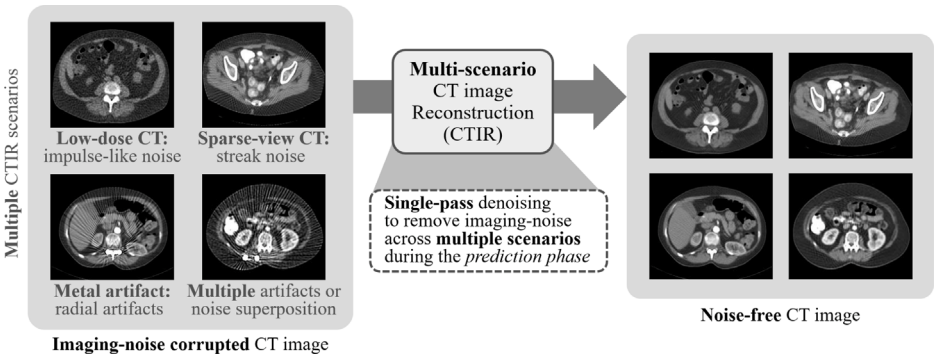

Fig. 1. Multi-scenario CT Image Reconstruction (CTIR) superimposes a variety of significantly different imaging noise and artifacts, including Low-Dose CT (LDCT), Sparse-View CT (SVCT), and metal artifacts.

Numerous specialized CTIR techniques have been developed for various scenarios (Ahishakiye et al., 2021). However, the lack of generalization in these methods potentially restricts their value and hinders their applicability in diverse clinical applications. Based on recent research progress in CTIR, relevant researchers have widely developed three meaningful scenarios: LDCT Denoising (LDCTD), SVCT Reconstruction (SVCTR), and Metal Artifact Reduction (MAR). (1) LDCTD is specifically designed to address the impulse-like noise generated by the reduction of radiation source energy while maintaining diagnostic information and reducing radiation risk. (2) SVCTR aims to restore the streak artifacts produced by the reduction in the number of projections caused by a sparse view acquisition protocol, which is adopted to minimize radiation exposure. (3) MAR, as a practical technique, effectively reduces radial artifacts caused by the high absorption coefficient of metal implants, thereby optimizing CT image quality in patients with metal implants. However, their scenario-specificity poses challenges in applying them to other scenarios and effectively handling situations involving multiple factors that impact CT imaging quality.

Handling multiple CTIR scenarios simultaneously poses significant challenges related to generalization, as shown in Fig. 2(a). The noise or artifacts in different CTIR scenarios show significant differences in both the sinogram domain and the image domain, necessitating multi-scenario CTIR to consider multiple completely different forms of noise distribution. In multi-scenario CTIR, diverse forms of imaging noise intertwine and overlay in CT images, leading to exacerbated image quality degradation, increased complexity in denoising efforts, and the potential introduction of unprecedented intricacies. On the other hand, the undesired secondary artifacts caused by the sinogram correction that fails to satisfy physical constraints are introduced into the CT image by the dual-domain conversion, despite previous research demonstrating the value of sinograms in improving imaging noise in the image domain Lin et al. (2019) and Lyu et al. (2020), as shown in Fig. 2(c). In the sinogram domain, imaging noise can be effectively corrected using inpainting techniques. However, when converting a sinogram that is not entirely noise-free into a CT image, errors and noise will propagate and become amplified in the image domain. Overall, the above challenges in CTIR impose exceptionally high demands on the scheme's ability to handle different scenarios, and require the scheme to judiciously utilize semantic information in both the image domain and the sinogram domain to avoid secondary artifacts.

Recently, the diffusion-based model has gained significant attention in the field of image generation and reconstruction (Croitoru et al., 2023). Its remarkable capacity for diversity has been substantiated through numerous studies, making it highly versatile across a wide range of vision tasks. However, in CTIR scenarios where the realism and accuracy of human tissue are crucial, the excessive diversity in images generated by diffusion-based models has emerged as a bottleneck (Müller et al., 2023). In this work, we present a novel diffusion-based model designed to simultaneously tackle multiple CTIR scenarios. By synergistically integrating semantics from both the image domain and the sinogram domain, our model achieves a high level of scenario generalization while ensuring the stability of the diffusion process convergence toward the denoising objective. Demonstrating a robust generalization, our model not only overcomes the challenges posed by the diversity of diffusion models but also effectively leverages the dual-domain imaging information, thereby further enhancing the accuracy and reliability of the reconstruction outcomes.

In this work, we propose a novel diffusion-based CTIR framework that effectively mitigates the impact of CT's physical limitations on image quality and simultaneously handles various scenarios through highly generalized dual-domain learning (Fig. 3). Specifically, the **Convergent-Diffusion Denoising Model** (CDDM) converges on the imaging-noise-free CT with sampling inference, responding to multiple imaging-noise with high generalizability during the CTIR process, as shown in Fig. 2(b). CDDM systematically learns to invert the parametric Markov noising process (Ho et al., 2020) with the attenuation distribution as a priori, enabling the sampling inference to focus on imaging noise. The convergent sampling for noise correction, based on a noising schedule, avoids the overfitting of individual scenarios and adapts to the various data distributions. During the sampling inference, the **Domain-correlated Sampling Network** (DS-Net) samples the noise distribution, introducing valuable image and sinogram semantics while completely sidestepping the interference of secondary artifacts, as shown in Fig. 2(d). We innovatively treat images and sinograms as guidance knowledge to reinforce noise prediction, similar to textual information in the text-to-image task (Ramesh et al., 2022). DS-Net captures dual-domain semantics with correlation analysis, ensuring that sinogram errors are not amplified or carried over into the image domain.

Our main contributions are summarized as follows: (1) CDDM is the novel diffusion-based image reconstruction framework firstly tailored for the CTIR task, which is highly generalizable to various scenarios, employing stepwise inference focused on imaging noise. (2) DS-Net offers an innovative dual-domain integration scheme, incorporating sinogram semantics to guide artifact prediction to integrate dual-domain knowledge and avoid secondary artifacts. (3) The conducted experiments on public datasets demonstrate that our framework can reliably deal with various CTIR scenarios, with high generalization, practicality, and clinical potential.

2. Related work

2.1. CT image reconstruction

2.1.1. Low-dose CT denoising

Originally intended to enhance the image quality of high-resolution LDCT scans, model-based iterative reconstruction encounters limitations stemming from vendor-specific scan geometries and substantial

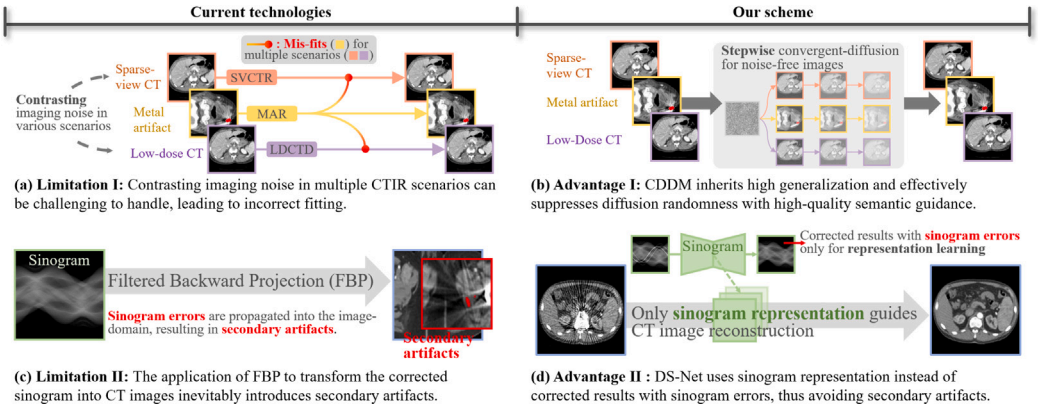

Fig. 2. The advanced generalization of CDDM and the efficient leverage of dual-domain semantics by DS-Net not only reliably handle diverse CTIR scenarios but also simultaneously alleviate secondary artifacts when utilizing sinogram representation.

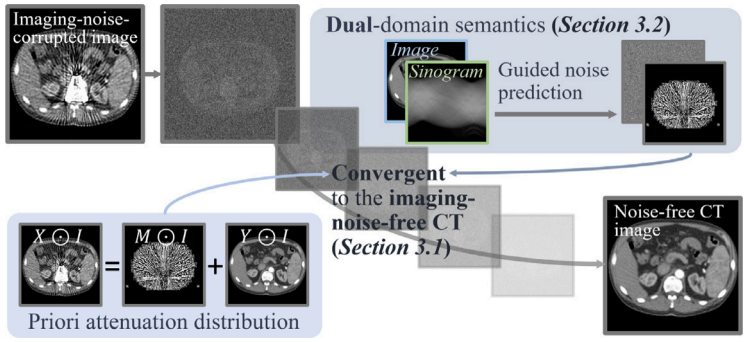

Fig. 3. CDDM utilizes convergent denoising diffusion, along with DS-Net's sinogram semantics, to concurrently handle multiple CTIR scenarios, achieving robust generalization through dual-domain learning.

computational demands (Fang et al., 2015; Green et al., 2016). These limitations, however, can be effectively alleviated through the application of deep learning methodologies. Chen et al. combined the autoencoder, deconvolution network, and shortcut connections to simultaneously balance the quality of noise reduction and the preservation of structural details (Chen et al., 2017). In the work by Yang et al. (2018), a Generative Adversarial Network (GAN) with Wasserstein distance and perceptual similarity was introduced to address the challenge of detail visibility in previous approaches. Feng et al. presented a novel strategy termed Content and Noise Concurrent Learning (CNCL) (Geng et al., 2021). This strategy enables simultaneous learning of noise characteristics and content information using two deep-learning predictors. Focusing on network architecture, Shen et al. applied Network Architecture Search (NAS) to LDCT reconstruction efforts, emphasizing the optimization of network design (Shen et al., 2022).

2.1.2. Sparse-view CT reconstruction

Numerous conventional algorithms have been proposed for reconstructing sparse-view CT images, which can be categorized into three main groups: sinogram completion (Li et al., 2014), iterative reconstruction (Kim et al., 2016), and image post processing (Han et al., 2016). As the number of available views significantly decreases, deep learning methods can effectively replace these techniques to achieve enhanced reconstruction outcomes. Jin et al. proposed employing direct inversion followed by a CNN to solve inverse problems involving normal-convolution scenarios (Jin et al., 2017). Zhang et al. introduced an innovative deep-learning algorithm that addresses the SVCTR problem by amalgamating the benefits of DenseNet and deconvolution (Zhang et al., 2018). Zhang et al. proposed a comprehensive domain network featuring spectral complementarity, resulting in improved noise reduction for Dual-Energy Computed Tomography (DECT) (Zhang et al., 2021). To synergize the strengths of deep learning and iterative reconstruction while adhering to the constraints of the projection domain, residual space, and image domain information, Zhang et al. presented a novel iterative reconstruction framework (Zhang et al., 2022).

2.1.3. Metal artifact reduction

The traditional MAR methods restore metal artifact signal (Meyer et al., 2010; Karimi et al., 2015) with iterative reconstruction (Jin et al., 2015; Chang et al., 2018), which have the disadvantages of relying heavily on artificial features and difficulty in ensuring data consistency. While deep learning-based methods demonstrate greater efficacy in restoring metal-damaged areas compared to traditional methods. These deep learning-based methods primarily fall into two categories: single-domain and dual-domain corrections. It is worth noting that sinogram-domain correction inevitably gives rise to secondary artifacts, while image-domain correction frequently misinterprets metal shadows as normal tissues (Zou et al., 2024). To address the limitations of single-domain correction, dual-domain correction methods have been developed. Lin et al. (2019), for the first time, processed sinogram and image domain information jointly based on a gradient-propagable Radon inversion layer. Wang et al. provided a deeper exploration of artifact prior structural features based on the convolutional dictionary model and a filter parameterization method based on Fourier series expansions (Wang et al., 2022b).

Lately, there has been a growing trend toward methods that collectively address the aforementioned scenarios. Zhou et al. processed sparse view reconstruction and metal artifact reduction tasks simultaneously for the first time, adding consistent information to the recurrent network to ensure reconstruction quality (Zhou et al., 2022b). In another effort by Zhou et al. (2022a), LDCTD and MAR were considered together for the first time, employing a combined approach involving

both image- and sinogram-domain processing to enhance image quality. For the first time, we contemplate the amalgamation of LDCTD, SVCTR, and MAR as a multi-scenario CTIR task, and then tailor a novel diffusion-based model to address these three CTIR scenarios simultaneously.

2.2. Diffusion-based model

Diffusion probabilistic models, inspired by non equilibrium statistical physics, systematically learn to reverse a gradual noising schedule for the structure destroying and restoring of a data distribution (Sohl-Dickstein et al., 2015). Similar to the early introduction of the diffusion concept, Ho et al. proposed denoising diffusion probabilistic models (DDPMs), which learn to invert the parametric Markov image noising process for high-quality image synthesis (Ho et al., 2020). Given a data distribution $x_0 \sim q(x_0)$ in DDPMs, the forward process produces a series of latents $x_{1:T}$ by adding Gaussian noise ($q(x_{1:T}|x_0) = \prod_{t=1}^{T} q(x_t|x_{t-1})$). And then, the reverse process is modeled by neural network architecture, predicting the parameters $\tilde{\mu}_t$ and $\tilde{\beta}_t$ of the Gaussian distribution to sample the x_{t-1} Gaussian distribution $q(x_{t-1}|x_t, x_0)$. Minimizing the log-likelihood (Benny and Wolf, 2022) enables diffusion-based probabilistic models (Dhariwal and Nichol, 2021; Nichol and Dhariwal, 2021) and their derivatives (Benny and Wolf, 2022; Song and Ermon, 2019) to learn the target distribution by sampling the reverse process (Sehwag et al., 2022).

To date, diffusion-based models have been widely applied to various tasks in computer vision, encompassing both generative modeling (Dhariwal and Nichol, 2021; Li et al., 2022) and discriminative tasks (Amit et al., 2021; Zimmermann et al., 2021). Recent research has also highlighted the effectiveness of diffusion models in medical image denoising and reconstruction applications (Hu et al., 2022; Gong et al., 2023; Liu et al., 2023). To address the limitations of diffusion models in generalizing to unfamiliar measurement processes, Song et al. proposed a completely unsupervised technique for solving inverse problems (Song et al., 2021). They implemented a fraction-based generative model, demonstrating favorable results in the MAR task. Building on these advances, diffusion-based models tailored to specific CTIR applications have been developed to enhance CT image quality. For low-dose CT, related methods aimed at minimizing contextual errors and optimizing diffusion rates have been proposed to reduce impulse-like noise (Xia et al., 2022b; Gao et al., 2023). In sparse-view CT, probability- and frequency-based diffusion models effectively managed radial artifacts (Xia et al., 2022a; Liu et al., 2023; Xu et al., 2024). For metal artifact, task-specific priors, such as metal and tissue anatomy information, have been integrated to mitigate streak artifacts (Liu et al., 2024; Cai et al., 2024).

Diverging from previous diffusion-based models that address only a single CTIR scenario, we introduce an innovative convergent-diffusion denoising model to accomplish multi-scenario CTIR. Our model converges toward the denoising objective with guidance from dual-domain semantics, offering significant application value and promising clinical potential.

3. Methodology

3.1. Convergent-diffusion denoising model

CDDM, a latent variable model, systematically learns to invert a gradual noising schedule, enabling effective disruption and restoration with high generalizability in CTIR scenarios. As shown in Fig. 4, it corrects the imaging noise m from an imaging-noise-corrupted image x_0 using *forward image destruction* and *reverse convergent inference* to obtain a noise-free image y_0. According to Wang et al. (2022a), the imaging-noise-corrupted images are treated as superposition states of the attenuation distributions of human tissue and imaging noise. Therefore, we introduce the priori attenuation distribution $X \odot I = Y \odot I + M \odot I$ (*i.e.*, $x_0 = y_0 + m$ in CDDM) into the convergent inverse inference, where X and Y are an imaging-noise-corrupted image and a noise-free image respectively, M is the imaging noise, and I is a binary non-metal mask.

Algorithm 1: Training of Convergent-diffusion

```
repeat
    ϵ ~ N(ϵ; 0, I)
    t ← Uniform ({1, ..., T})
    x_t ← sqrt(ᾱ_t) x_0 + sqrt(1 − ᾱ_t) ϵ_x
    m_t, ϵ_t ← f_θ(x_t, t)
    Take gradient descent step on
        ∇_θ(‖m_t − x_0 + y_0‖² + ‖ϵ_t − ϵ‖²)
until converged
```

For the *forward image destruction*, x_0 and y_0 are fixed in two independent Markov chains in parallel, and are destroyed into Gaussian noises x_T and y_T with imaging information bias via T-step Gaussian noise ϵ superposition (Song et al., 2020). For the *reverse convergent inference*, the reverse process converges x_t (t is random) to y_{t-1} using the attenuation-based inference sampling in the *training phase* (Algo. 1). Based on the characteristics of diffusion-based model, sampling of x_t or y_t in closed form at an arbitrary step t during forward destruction is defined as Eq. (1)(Top), while sampling the $t-1$-step distribution based on x_t at an arbitrary step t is defined as Eq. (1)(Bottom).

$$\begin{aligned} \left\{q(x_t|x_0), q(y_t|y_0)\right\} &= \left\{\mathcal{N}(x_t; \sqrt{\bar{\alpha}_t}x_0, (1-\bar{\alpha}_t)\mathbf{I}), \mathcal{N}(y_t; \sqrt{\bar{\alpha}_t}y_0, (1-\bar{\alpha}_t)\mathbf{I})\right\} \\ q(y_{t-1}|x_t, y_0, m) &= \mathcal{N}(y_{t-1}; \tilde{\mu}_t(x_t, y_0, m), \tilde{\sigma}_t^2\mathbf{I}) \end{aligned} \tag{1}$$

where α_t and $\bar{\alpha}_t$ are $1-\beta_t$ and $\prod_{i=1}^{t}\alpha_i$, respectively, and β_t follows the variance schedule $\{\beta_t \in (0,1)\}_{t=1}^{T}$.

According to Bayes' rule, the Gaussian distribution $q(y_{t-1}|x_t, y_0, m)$ can be transformed into a joint distribution, *i.e.*, the product of $q(x_t|y_{t-1}, m)$ and the ratio $q(y_{t-1}|y_0)$ to $q(x_t|x_0)$. In this context, $q(y_{t-1}|y_0)$ and $q(x_t|x_0)$ are computed as $\mathcal{N}(y_{t-1}; \sqrt{\bar{\alpha}_{t-1}}y_0, (1-\bar{\alpha}_{t-1})\mathbf{I})$ and $\mathcal{N}(x_t; \sqrt{\bar{\alpha}_t}x_0, (1-\bar{\alpha}_t)\mathbf{I})$, respectively, following Eq. (1). With the introduction of priori attenuation distribution, $q(x_t|y_{t-1}, m)$ is computed as $\mathcal{N}(x_t; \sqrt{\alpha_t}y_{t-1} + \sqrt{\bar{\alpha}_t}m, \chi^2\mathbf{I})$, where χ is defined as $\sqrt{1-\bar{\alpha}_t - \sqrt{\alpha_t - \bar{\alpha}_t}}$. After integrating the y_{t-1}-related terms of this joint distribution, $\tilde{\mu}_t$ and $\tilde{\sigma}_t$ in Eq. (1) are expressed as:

$$\begin{aligned} \tilde{\mu}_t &= ((-\sqrt{\alpha_t}x_t + \alpha_t\sqrt{\bar{\alpha}_{t-1}}m)/\chi^2 - \sqrt{\bar{\alpha}_{t-1}}y_0/(1-\bar{\alpha}_{t-1})) \cdot \tilde{\sigma}_t^2 \\ \tilde{\sigma}_t &= \left(\alpha_t/\chi^2 + 1/(1-\bar{\alpha}_{t-1})\right)^{-1/2} \end{aligned} \tag{2}$$

where y_0 is $1/\sqrt{\bar{\alpha}_t}(x_t - \sqrt{\bar{\alpha}_t}m - \sqrt{1-\bar{\alpha}_t}\epsilon)$. For the necessary noise distribution in Eq. (2), the DS-Net (the next section for details) predicts the imaging noise m and Gaussian noise ϵ. In the *sampling phase* (Algo. 2), a total of T consecutive steps of sampling inference are performed, where the $t-1$-step Gaussian distribution y'_{t-1} is inferred from the t-step sampled distribution y'_t (y'_T is equivalent to x_T).

3.2. Domain-correlated sampling network

DS-Net correlates the sinogram semantics obtained by representation learning with the image-domain noise prediction during noise prediction of CDDM, enabling the utilization of sinograms to avoid secondary artifacts. As shown in Fig. 5, a time-step t and a t-step distribution x_t (in *training phase*) or y'_t (in *sampling phase*) are the network input. First, t and the t-step distribution are computed into image representation f_{img} and time-step embedding e_t using an encoder $\mathcal{E}_{img}(\cdot)$ and a Multi-Layer Perceptron (MLP) projection, respectively. Next, the image representation f_{img} and a sinogram representation f_{sin} (*Section - Sinogram Representation Learning.* for details) are transformed

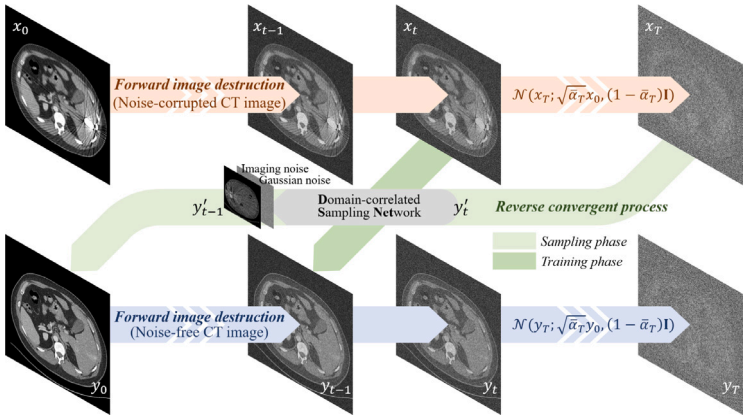

Fig. 4. CDDM. Sampling inference converging on a noise-free CT image y_0 responds to various CTIR scenarios with high generalizability during the correction of an imaging-noise-corrupted CT image x_0.

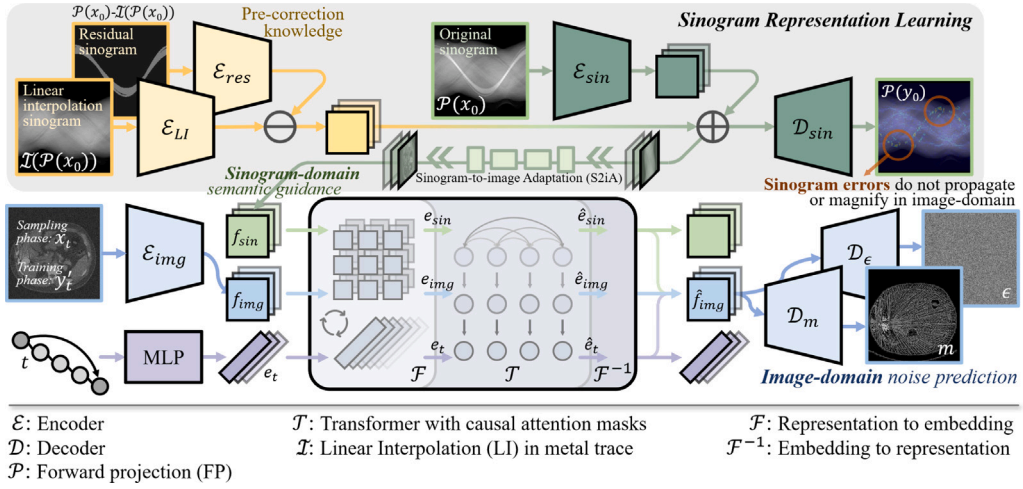

Fig. 5. DS-Net. Dual-domain correlation analysis introduces sinogram semantics as guidance knowledge for noise prediction, while the sinogram error is not amplified or carried over into the image domain.

Algorithm 2: Sampling of Convergent-diffusion

$\epsilon \sim \mathcal{N}(\epsilon; 0, \mathbf{I})$
$x_T \leftarrow \sqrt{\bar{\alpha}_T} x_0 + \sqrt{1-\bar{\alpha}_T}\epsilon$
$t = T$
while $t > 0$ **do**
  $y'_t \leftarrow x_T$ **if** $t = T$
  $m_t, \epsilon_t \leftarrow f_\theta(y'_t, t)$
  $y'_{t-1} \sim \mathcal{N}(y'_{t-1}; \tilde{\mu}_t(m_t, \epsilon_t), \tilde{\sigma}_t^2 \mathbf{I})$
  $t \leftarrow t - 1$
end
return y'_0

into two embeddings by a series of operations, that consists of dividing the representation into patches and performing an MLP projection on each patch. Then, a Transformer architecture $\mathcal{T}(\cdot)$ with causal attention masks analyzes sequence correlation to incorporate sinogram semantics and time-step constraints into image embedding. Causal attention masks (Aghajanyan et al., 2022) are utilized to maintain causal relationships in sequence modeling, which find applications in various contexts (*e.g.*, DALL·E2 Ramesh et al., 2022). The analyzed sequence consists of, in order: the sinogram embedding, the image embedding, and the time-step embedding. The series of operations described above can be expressed as:

$$\begin{aligned} e_{img}, e_{sin} &= \mathcal{F}(f_{img}), \mathcal{F}(f_{sin}) \\ \hat{e}_{sin}, \hat{e}_{img}, \hat{e}_t &= \mathcal{T}(e_{sin}, e_{img}, e_t) \\ \hat{f}_{img} &= \mathcal{F}^{-1}(\hat{e}_{img}) \end{aligned} \tag{3}$$

Finally, m and ϵ are predicted by decoding the final representation using two decoders $\mathcal{D}_\epsilon(\cdot)$ and $\mathcal{D}_m(\cdot)$, where the representation is transformed from image embedding.

Sinogram Representation Learning. Given the incremental effectiveness of Linear Interpolation (LI) pre-correction in sinogram-domain correction (Wang et al., 2021; Yu et al., 2020), we introduce pre-correction knowledge into the learning of the *sinogram representation*. Reconstruction of artifact-free sinogram $\mathcal{P}(y_0)$ is treated as the proxy task for representation learning, where $\mathcal{P}(\cdot)$ denotes the forward projection. During the reconstruction, we introduce a LI sinogram (Kalender et al., 1987) and a residual sinogram (Yu et al., 2020) as pre-corrected sinograms. The two are determined as $\mathcal{I}(\mathcal{P}(x_0))$ and the signal difference between the $\mathcal{P}(x_0)$ and $\mathcal{I}(\mathcal{P}(x_0))$, where $\mathcal{I}(\cdot)$ denotes the LI operation in the metal trace. Multiple encoders $\mathcal{E}_{sin}(\cdot)$, $\mathcal{E}_{LI}(\cdot)$, and $\mathcal{E}_{res}(\cdot)$ process the above sinograms into latent representations, including $\mathcal{P}(x_0)$, $\mathcal{I}(\mathcal{P}(x_0))$, and $\mathcal{P}(x_0) - \mathcal{I}(\mathcal{P}(x_0))$. For representation learning, the LI sinogram provides pre-corrected knowledge and the residual sinogram focuses on correcting the LI error in the metal trace. The encoding difference between the representations of the LI sinogram and the residual sinogram is element-wise added with the representation of

$\mathcal{P}(x_0)$ to obtain the reconstruction representation:

$$\begin{aligned} f'_{sin} &= \mathcal{E}_{LI}(\mathcal{I}(\mathcal{P}(x_0))) - \mathcal{E}_{res}(\mathcal{P}(x_0) - \mathcal{I}(\mathcal{P}(x_0))) \\ f_{sin} &= \mathcal{E}_{sin}(\mathcal{P}(x_0)) + f'_{sin} \end{aligned} \tag{4}$$

The reconstructed representation translates the image domain into the *sinogram representation* through a Sinogram-to-Image Adaptation (S2iA) process. This adaptation encompasses a sequence comprising two up-convolutional layers and two max-pooling convolutional layers. S2iA enhances alignment for semantic guidance while ensuring consistent parameters across all time steps, as the sinogram is not involved in diffusion. A decoder $D_{sin}(\cdot)$ outputs a metal-free sinogram based on the reconstruction representation to complete the proxy task.

The DS-Net objective function $\mathcal{L}_{DS}$ consists of the sinogram reconstruction loss of s and the noise predicted loss of m/ϵ, which is defined as:

$$\begin{aligned} \mathcal{L}_{DS} = \mathbb{E}_{\epsilon\sim\mathcal{N}(\epsilon;0,\mathbf{I}),t\sim[1,T]}[&\|m_\theta(x_t,t) - x_0 + y_0\|^2 \\ &+ \|e_\theta(x_t,t) - \epsilon\|^2 + \|\mathcal{D}(f_{sin}) - \mathcal{P}(y_0)\|^2] \end{aligned} \tag{5}$$

4. Experiments

4.1. Dataset

The dataset utilized in this work includes both the DeepLesion dataset and the CLINIC-metal dataset. The DeepLesion dataset[1] is utilized for comparative experiments and ablation analyses conducted on synthetic data. Meanwhile, the CLINIC-metal dataset,[2] featuring authentic metal artifacts, is employed to validate the clinical reliability. The DeepLesion dataset comprises 32,120 axial CT slices extracted from 10,594 CT scans, representing 4427 distinct patients. Within this dataset, a diverse array of lesion types is present, facilitating applications in tasks such as lesion detection, classification, segmentation, retrieval, and related applications. The CLINIC-metal dataset is a sub-dataset of the CTPelvic1K and is collected for pelvic fracture segmentation, which consists of 14 metal-corrupted volumes with pixel-wise annotations of multiple bone structures. CTPelvic1K aggregates large pelvic CT datasets from multiple sources and different manufacturers, which include 1184 CT volumes and over 320,000 slices with different resolutions and a variety of the above-mentioned appearance variations.

4.2. Pre-processing

We employed realistically simulated low-dose CT images and sparse-view CT images with metallic implants, adhering to the consensus simulation protocol outlined in Zhang and Yu (2018), Lin et al. (2019) and Zhou et al. (2022b,a). For SVCT and LDCT, we adopted an equiangular fan-beam projection geometry with a 120 kVp poly-energetic X-ray source. Two low-dose CT scenarios with Poisson noise in the sinogram were simulated, using 2×10^5 photons for the 1/2 dose level and 1×10^5 photons for the 1/4 dose level. Regarding sparse-view sinograms, we uniformly sampled 90 and 180 projection views from the total 360 views. The sparse-view image was reconstructed from the sinogram using filtered backward projection (FBP). For MAR, we randomly selected 1200 full-dose CT 2D images from the DeepLesion (Yan et al., 2018) dataset. From these, we collected 100 manually segmented metal implants with varying locations, shapes, and sizes, categorized into five levels (small to large) based on size. Out of these, 90 were used as training masks, and the remaining 10 as testing masks. Subsequently, we randomly selected 1000 CT images, and for each size-level, each slice was synthesized using a randomly selected training mask to generate five metal artifact images as the training set. The synthesis strategy for this training data ensures both size-level balance and sample diversity. Following this, we randomly chose 200 CT slices and synthesized them with the testing mask to create testing samples. To validate the robust generalization of our approach, we combined the two low-dose scenarios and the two sparse-view settings, generating multiple combinatorial approaches within the context of various types of metallic implants for experimentation purposes.

4.3. Implementation details

All inputs are resized to 256 × 256, enabling the encoders to naturally align the sinogram and the CT image. In DS-Net, both the encoders ($\mathcal{E}_{img}$, $\mathcal{E}_{sin}$, $\mathcal{E}_{LI}$, and $\mathcal{E}_{res}$) and decoders ($\mathcal{D}_m$, $\mathcal{D}_\epsilon$, and $\mathcal{D}_{sin}$) adopt the U-Net (Ronneberger et al., 2015) architecture with 4 layers, where the encoding features are combined with the decoding features through concatenation during the decoding process. Feature maps $f \in \mathbb{R}^{16\times16\times512}$ are divided into 2×2 patches, with each patch transformed into embedding $e \in \mathbb{R}^{512}$; The number of heads and head dimensions for Transformer encoders is consistently set at 8 and 64, respectively. The maximum value of T is set to 250, and the variance schedule β_t increases linearly from 0 to 0.9 during the *forward image destruction* process. To optimize DS-Net, we employ the Adam optimizer with hyperparameters $\beta_1 = 0.9$ and $\beta_2 = 0.6$, a learning rate of $1e$–4, and $5.0e$–2 weight decay, following a warm-up period of 5000 iterations. The peak signal-to-noise ratio (PSNR) and the structured similarity index (SSIM) were adopted for the CTIR evaluation.

4.4. Comparison with SOTAs

To evaluate the significance of CDDM in reconstructing multi-scenario CT images, we conducted experiments encompassing various tasks, including LDCTD, SVCTR, and MAR. Additionally, we investigated two multi-scenario tasks: LDCTD combined with MAR and SVCTR combined with MAR. State-of-the-art (SOTA) methods were compared for each task, including RED-Net (Chen et al., 2017), WGAN (Yang et al., 2018), CNCL (Geng et al., 2021), and MLF-IOSC (Shen et al., 2022) for LDCTD, and DD-Net (Zhang et al., 2018), CD-Net (Zhang et al., 2021), SGM (Song et al., 2021), and DREAM-Net (Zhang et al., 2022) for SVCTR. MAR comparisons involved SOTA methods such as Lin (Kalender et al., 1987), DuDoNet (Lin et al., 2019), SGM (Song et al., 2021), and OSCNet (Wang et al., 2022b). LDCTD + MAR and SVCTR + MAR tasks involved assessments of methods such as Lin (Kalender et al., 1987), DuDoDRNet (Zhou et al., 2022b), DuDoUFNet (Zhou et al., 2022a), and SGM (Song et al., 2021). In the comparisons mentioned above, SGM served as the diffusion model-based baseline. This systematic comparison facilitated a comprehensive assessment of CDDM's efficacy across diverse scenarios, providing a thorough understanding of its CTIR performance.

As demonstrated in Tables 1, 2, and 3, CDDM outperforms SOTA methods by achieving the highest PSNR and SSIM scores across various CTIR tasks — MAR, SVCTR, and LDCTD. In MAR tasks with varying metal sizes, CDDM showed improvements (PSNR/SSIM) over SOTA methods: 1.77/0.38% for the smallest metal size, 1.85/1.04% for the second smallest size, 1.49/0.58% for the middle size, 0.74/0.51% for the second largest size, and 0.49/0.93% for the largest size. For SVCTR tasks with 90 and 180 projection views, CDDM exhibited performance improvements (PSNR/SSIM) of 2.43/0.54% and 1.41/0.95%, respectively, compared to SOTA methods. In LDCTD tasks with dose levels of 1/2 and 1/4, CDDM achieved improvements (PSNR/SSIM) of 1.45/1.36% and 0.09/1.16%, respectively. Furthermore, compared to SOTA methods concurrently addressing MAR+LDCT and MAR+SVCT scenarios, CDDM consistently outperformed in various multi-scenario CTIR tasks. As illustrated in Table 4, experimental results demonstrated that CDDM seamlessly handled multiple CTIR scenarios without requiring task decomposition, leading to optimal performance in both PSNR and SSIM metrics.

[1] https://nihcc.app.box.com/v/DeepLesion.
[2] https://github.com/MIRACLE-Center/CTPelvic1K.

Table 1
Quantitative results (PSNR/SSIM) on different size-levels demonstrate the superior MAR performance of CDDM.

Methods	Small metal ⟶ Large metal					Average
Input	31.30/0.9375	29.54/0.9262	27.92/0.9085	25.80/0.8901	21.14/0.8465	27.14/0.9018
Lin (Kalender et al., 1987)	32.46/0.9606	29.94/0.9458	29.17/0.9390	24.93/0.9039	21.49/0.8674	27.60/0.9233
DuDoNet (Lin et al., 2019)	37.65/0.9791	37.19/0.9778	36.46/0.9742	35.01/0.9704	32.39/0.9616	35.74/0.9726
SGM-MAR (Song et al., 2021)	40.17/0.9894	39.62/0.9821	39.23/0.9853	38.24/0.9829	35.96/0.9748	38.64/0.9829
OSCNet (Wang et al., 2022b)	39.41/0.9825	38.76/0.9787	37.74/0.9754	36.28/0.9730	34.03/0.9721	37.24/0.9763
CDDM (Ours)	**41.94/0.9932**	**41.47/0.9925**	**40.72/0.9911**	**38.98/0.9880**	**36.45/0.9841**	**39.91/0.9898**

Table 2
Quantitative results (PSNR/SSIM) on different view-number (90°&180°) demonstrate the superior SVCTR performance of CDDM.

Methods	90°	180°	Average
Input	23.77/0.7647	34.00/0.9306	28.89/0.8477
DD-Net (Zhang et al., 2018)	25.13/0.7532	34.38/0.9062	29.76/0.8297
CD-Net (Zhang et al., 2021)	31.40/0.9340	35.33/0.9415	33.37/0.9378
SGM-SVCT (Song et al., 2021)	35.18/0.9677	38.85/0.9712	37.02/0.9695
DREAM-Net (Zhang et al., 2022)	34.50/0.9506	36.30/0.9739	35.40/0.9623
CDDM (Ours)	**37.61/0.9731**	**40.26/0.9807**	**38.94/0.9769**

Table 3
Quantitative results (PSNR/SSIM) on different CT doses (1/4&1/2) demonstrate the superior LDCTD performance of CDDM.

Methods	1/4	1/2	Average
Input	30.47/0.8783	33.54/0.9186	32.01/0.8985
RED-Net (Chen et al., 2017)	32.49/0.9130	35.28/0.9366	33.89/0.9248
WGAN (Yang et al., 2018)	33.91/0.9298	35.06/0.9304	34.49/0.9301
CNCL (Geng et al., 2021)	35.55/0.9554	37.61/0.9692	36.58/0.9623
MLF-IOSC (Shen et al., 2022)	37.77/0.9632	40.94/0.9675	39.36/0.9654
CDDM (Ours)	**39.22/0.9768**	**41.03/0.9791**	**40.13/0.9780**

Table 4
Quantitative results (PSNR/SSIM) on multiple scenarios (Top: 1/2-LDCTD+MAR Bottom: 180°-SVCTR+MAR) demonstrate the superior performance of CDDM for multi-scenario CTIR.

Methods	Small metal ⟶ Large metal					Average
Input	28.91/0.8824	27.89/0.8745	26.66/0.8642	25.07/0.8475	21.11/0.8131	25.93/0.5863
Lin (Kalender et al., 1987)	29.96/0.8972	28.45/0.8863	27.98/0.8840	24.58/0.8561	21.43/0.8317	26.48/0.8711
DuDoUFNet (Zhou et al., 2022b)	38.20/0.9620	36.06/0.9553	35.99/0.9493	35.35/0.9363	31.60/0.9194	35.44/0.9445
CDDM (Ours)	**40.56/0.9878**	**39.30/0.9798**	**38.56/0.9724**	**37.88/0.9714**	**35.37/0.9674**	**38.33/0.9758**
Input	27.52/0.8587	25.90/0.8329	24.92/0.8264	23.24/0.7977	19.82/0.7534	25.93/0.8563
Lin (Kalender et al., 1987)	30.39/0.9066	28.78/0.8952	28.29/0.8917	24.75/0.8646	21.52/0.8414	26.75/0.8799
DuDoDRNet (Zhou et al., 2022a)	36.31/0.9748	35.62/0.9692	35.75/0.9506	33.99/0.9426	31.74/0.9333	34.68/0.9541
SGM (MAR→SVCT) (Song et al., 2021)	34.85/0.9673	34.29/0.9740	33.67/0.9460	33.79/0.9243	30.69/0.9030	33.46/0.9429
CDDM (Ours)	**39.70/0.9843**	**39.66/0.9861**	**37.72/0.9739**	**37.46/0.9711**	**33.05/0.9595**	**37.52/0.9750**

It is noteworthy that CDDM's advantage becomes more pronounced when addressing multi-scenario CTIR compared to single-scenario scenarios. This observation suggests that CDDM demonstrates a high level of generalization across various CTIR scenarios, underscoring its remarkable versatility in a spectrum of CTIR tasks. This superiority can be attributed to the high-quality dual-domain semantic-guided convergent diffusion process, which reliably preserves the original image information.

For qualitative comparison, the visual sample in Figs. 6 and 7 illustrates that CDDM not only effectively removes imaging noise in various scenarios but also excels in restoring the authentic appearance of human tissues. Notably, the dual-domain semantic coupling, closely aligned with CTIR, effectively mitigates various imaging noise in multiple CTIR scenarios while simultaneously minimizing the introduction of secondary artifacts arising from sinogram correction.

4.5. Ablation experiment

To validate the effectiveness of the network architecture design, we introduced several ablation variants for comparative analysis with CDDM:

- **ABLT-SRL:** This variant excludes Sinogram Representation Learning and the semantic guidance of sinograms.
- **ABLT-LI:** This variant omits LI pre-correction, essential for reinforcement representation learning in sinogram representation.
- **ABLT-S2iA:** This variant eliminates S2iA, a component aimed at enhancing the adaptability of sinogram representations to image representations.

These ablation variants allow us to evaluate the specific contributions of each component to the overall performance of the network. It is crucial to note that the remaining network structural components, not included in the ablation analysis, play a role in maintaining the end-to-end integrity of CDDM and DS-Net.

In order to assess the efficacy of dual-domain correlation for denoising guidance, a *reverse convergent inference* in DS-Net was conducted by removing sinogram representation learning (ABLT-SRL, Table 5). The noticeable decline in performance without sinogram conditional guidance highlights the positive impact of the interaction and coupling of two-domain semantics on CT image reconstruction. This observation underscores the necessity of robust semantic support for diffusion-based models with generative diversity to be effectively employed in authenticity-focused image denoising.

Additionally, the effectiveness of LI pre-correction (ABLT-LI, Table 5) for obtaining sinogram embeddings was examined. Despite its inherent imprecision, the prior pre-correction proves advantageous for sinogram-domain correction and description. This suggests that LI pre-correction, focusing on trace areas, can effectively incorporate potential

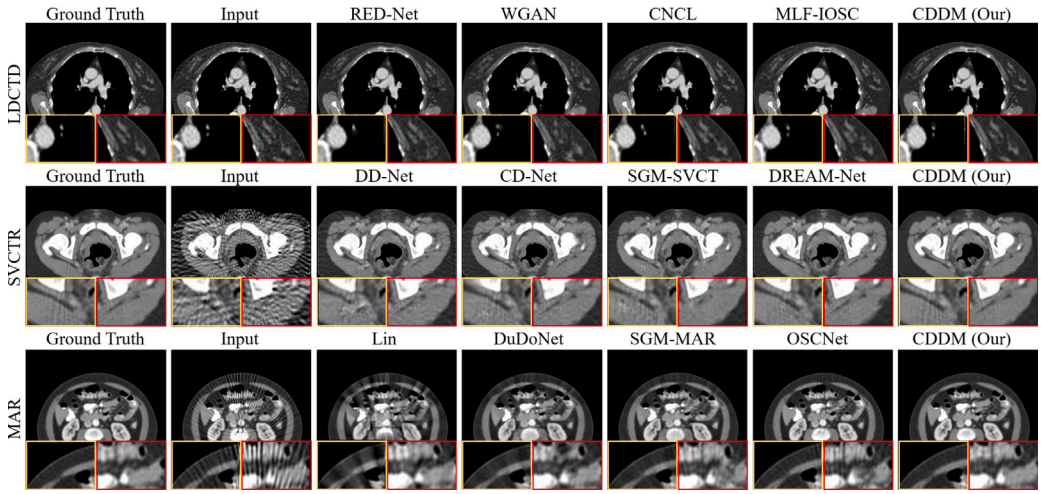

Fig. 6. Qualitative results demonstrate the capability of CDDM to correct imaging noise in scenarios involving low-dose CT, sparse-view CT, and metal artifacts.

Table 5
Ablation analysis demonstrates that the effectiveness of DS-Net's architecture design and the reliability of dual-domain information in guiding the denoising process, where MAR *avg.* represents the average performance across different size-level MARs.

Methods	SVCTR 90°	SVCTR 180°	LDCTD 1/4	LDCTD 1/2	MAR *avg.*
ABLT-SRL	26.37/0.7862	33.31/0.9366	33.53/0.9254	36.92/0.9419	35.95/0.9793
ABLT-LI	36.40/0.9682	39.73/0.9708	37.16/0.9545	39.07/0.9628	38.56/0.9835
ABLT-S2iA	34.36/0.9427	37.27/0.9526	36.74/0.9477	38.62/0.9554	36.73/0.9768
CDDM	37.61/0.9731	40.26/0.9807	39.22/0.9768	41.03/0.9791	39.91/0.9897

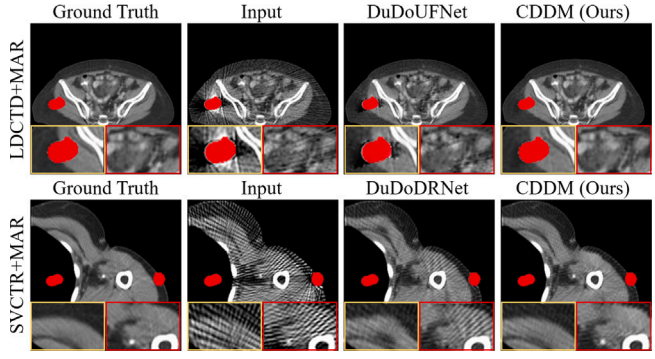

Fig. 7. Qualitative results demonstrate CDDM's effectiveness in addressing complex imaging noise in multi-scenario CTIR.

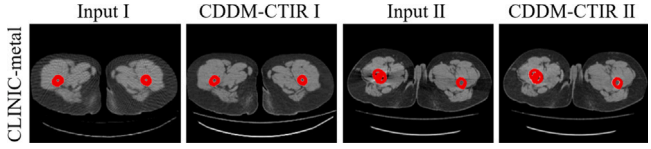

Fig. 8. Qualitative results on CLINIC-metal demonstrate the clinical feasibility and application potential of CDDM.

prior semantics to alleviate secondary artifacts in sinogram-domain correction, providing better continuity for the targeted guidance of image-domain estimation.

Furthermore, the effectiveness of S2iA was evaluated in comparison between ABLT-S2iA and CDDM in Table 5. Although the shapes of the sinogram representation and image representation are naturally aligned due to the uniformity of input shape, a significant difference in the distribution of image signals between the sinogram and CT images still exists. S2iA effectively addresses this issue, enabling the sinogram representation to more efficiently guide noise prediction in the image domain and contribute to improved image reconstruction.

4.6. Clinical study

To enhance the validation of CDDM's generalization and applicability in clinical scenarios, we conducted a qualitative comparison using the CLINIC-metal dataset (Liu et al., 2021). This dataset comprises 14 volumes intentionally distorted by metal artifacts, presenting a realistic and challenging clinical scenario. In this study, we initiated the analysis by segmenting the clinical metal masks, utilizing a threshold of 2500 Hounsfield Units (HU) to precisely isolate regions affected by metal artifacts.

The results, presented in Fig. 8, unequivocally demonstrate CDDM's effectiveness in correcting the majority of imaging noise associated with metal artifacts. Notably, CDDM achieves this correction without introducing secondary artifacts or interfering with unaffected human tissues, underscoring its capability to address the intricacies of complex clinical scenarios.

Furthermore, our clinical study indicates that CDDM performs comparably to methods trained on synthesized data when handling clinical datasets. This finding not only emphasizes the potential of CDDM but also highlights its adaptability and reliability in practical applications within clinical settings. The robust performance of CDDM in a clinical context enhances its credibility and positions it as a promising solution for real-world medical imaging applications.

5. Discussion

The addressed challenges in multi-scenario CT image reconstruction have been tackled with the introduction of a novel diffusion-based framework. The core of the proposed solution lies in CDDM and DS-Net. CDDM, as a diffusion-based model, is introduced to simultaneously handle multiple CTIR scenarios. DS-Net plays a crucial role in integrating

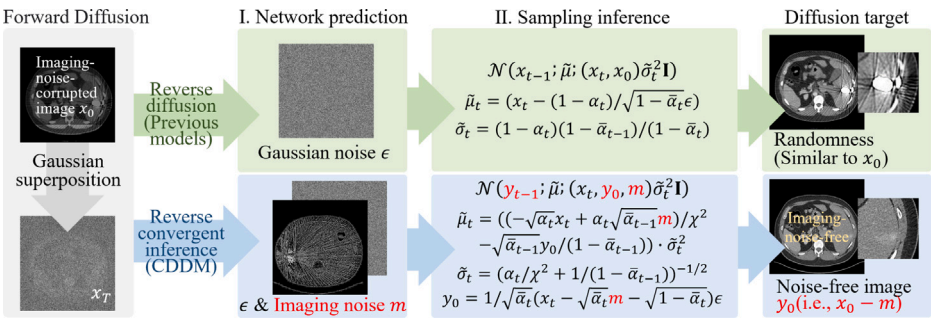

Fig. 9. CDDM effectively corrects CT imaging noise by constraining previous diffusion randomness, assisted by an imaging-noise-focused network prediction (I) and a novel sampling inference design (II).

semantics from both image and sinogram domains, ensuring multi-scenario generalization and stable convergence toward the denoising objective.

This study acknowledges the existence of several specialized CTIR techniques designed for specific scenarios such as LDCTD, SVCTR, and MAR. However, it highlights their significant limitations in universality, restricting their broader applicability across diverse clinical conditions. The diversity introduced by diffusion-based models brings excellent generalization performance, but the inherent randomness poses a bottleneck in faithfully restoring the true attenuation distribution of human tissues. CDDM addresses this challenge by employing powerful semantic guidance, ensuring both generalization performance and the authenticity of generated results. The term "convergent" refers to constraining diffusion diversity through task-specific architecture and inference, guiding reconstruction toward noise-free CT images. In contrast to previous diffusion models, CDDM incorporates the design of network prediction and sampling inference, correcting Gaussian ϵ and imaging noise m in a step-by-step *reverse convergent inference* process, as shown in Fig. 9. This process converges toward the noise-free CT image y_0 with minimal diffusion randomness. The constraints on diffusion randomness are attributed to *reverse convergent inference*, specifically aimed at rectifying imaging noise m, with contributions from both ϵ and m predictions. Experimental results demonstrate that, for the same x_0, CDDM achieves diffusion errors less than 10^{-4}, highlighting its robust performance. The interference of various forms of imaging noise in multi-scenario CTIR adds complexity, making it crucial to effectively leverage dual-domain information and mitigate the impact of secondary artifacts. DS-Net emerges as an effective solution to this challenge, offering strong generalization and utilizing dual-domain information to enhance the accuracy and reliability of reconstruction results. DS-Net's capability to avoid "secondary artifacts" is primarily due to CDDM's use of sinogram semantics to guide image domain predictions, minimizing errors in sinogram and achieving more accurate reconstructions.

The presented investigation provides a comprehensive assessment of the proposed CDDM in the domain of multi-scenario CTIR. The systematic comparison of CDDM with SOTA methods encompasses various CTIR tasks. Significantly, CDDM consistently outperforms SOTA methods across all tasks, underscoring its superior efficacy in single-scenario CTIR endeavors. CDDM's prowess is further highlighted in multi-scenario tasks, and ablation experiments offer insights into the specific contributions of its components. The study extends its scrutiny to a clinical setting using the CLINIC-metal dataset, incorporating volumes distorted by metal artifacts. Under realistic conditions, CDDM is evaluated, demonstrating its effectiveness in rectifying imaging noise associated with metal artifacts without introducing secondary artifacts or impacting unaffected human tissues. This substantiates CDDM's potential as a robust solution for real-world challenges in medical imaging, particularly in scenarios dominated by metal artifacts. Rigorous quantitative evaluations, ablation experiments, and real-world clinical validations collectively establish the credibility and potential of CDDM as a SOTA solution in the field of CTIR.

Table 6
Correspondence between the reduction of the maximum sampling steps T and the performance degradation (Mean Squared Error, MSE) of CDDM.

T	10	15	25	50	100	150	250
MSE	0.0042	0.0019	0.0008	0.0006	0.0004	0.0001	0.0001

However, CDDM also has certain limitations. When T exceeds 25, the model's performance degradation is below 10^{-3}, as shown in Table 6. At $T = 25$, the average inference time for 2000 samples is 1736 ms. This affirms the reliable real-time denoising capability of CDDM, ensuring practical applicability without compromising quality. Nevertheless, when compared to input-to-output network models for one-time prediction, the diffusion-based model necessitates multiple network prediction iterations, resulting in a time cost disadvantage. In future work, our primary focus will be on improving prediction efficiency based on the diffusion model while maintaining the reducibility and authenticity of the reconstruction results for various human tissues. Additionally, we plan to explore the potential of diffusion-based models for reconstructing images from a greater variety of medical imaging modalities.

6. Conclusion

In this work, we introduce a novel diffusion-based denoising framework, denoted as CDDM, designed to address diverse image reconstruction scenarios in CT images, employing an innovative dual-domain approach facilitated by DS-Net. CDDM showcases robust generalization capabilities by incorporating convergent sampling inference throughout the denoising procedure. DS-Net plays a crucial role in seamlessly integrating sinogram semantics into CDDM denoising, ensuring the absence of secondary artifacts through a comprehensive analysis of dual-domain correlations. The experimental findings robustly validate the efficacy of our proposed method in tackling multi-scenario CT image reconstruction tasks, encompassing LDCTD, SVCTR, and MAR. The results underscore the versatility and reliability of CDDM, emphasizing its potential as a promising solution for various challenges in the realm of CT image reconstruction.

CRediT authorship contribution statement

Xinghua Ma: Writing – review & editing, Writing – original draft, Methodology, Formal analysis, Data curation, Conceptualization. **Mingye Zou:** Conceptualization. **Xinyan Fang:** Conceptualization. **Gongning Luo:** Conceptualization. **Wei Wang:** Conceptualization. **Suyu Dong:** Methodology, Validation. **Xiangyu Li:** Methodology, Validation. **Kuanquan Wang:** Conceptualization. **Qing Dong:** Conceptualization. **Ye Tian:** Conceptualization. **Shuo Li:** Writing – review & editing.

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.

Acknowledgments

This work was supported by the National Natural Science Foundation of China under Grants 62272135, 62372135, 62202092, and the Key Research and Development Program of Heilongjiang Province under Grant 2023X01A08.

Data availability

The authors do not have permission to share data.

References

Aghajanyan, A., Huang, B., Ross, C., Karpukhin, V., Xu, H., Goyal, N., Okhonko, D., Joshi, M., Ghosh, G., Lewis, M., et al., 2022. Cm3: A causal masked multimodal model of the internet. arXiv preprint arXiv:2201.07520.

Ahishakiye, E., Bastiaan Van Gijzen, M., Tumwiine, J., Wario, R., Obungoloch, J., 2021. A survey on deep learning in medical image reconstruction. Intell. Med. 1 (03), 118–127.

Amit, T., Shaharbany, T., Nachmani, E., Wolf, L., 2021. Segdiff: Image segmentation with diffusion probabilistic models. arXiv preprint arXiv:2112.00390.

Benedict, S.H., Yenice, K.M., Followill, D., Galvin, J.M., Hinson, W., Kavanagh, B., Keall, P., Lovelock, M., Meeks, S., Papiez, L., et al., 2010. Stereotactic body radiation therapy: the report of AAPM task group 101. Med. Phys. 37 (8), 4078–4101.

Benny, Y., Wolf, L., 2022. Dynamic dual-output diffusion models. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 11482–11491.

Cai, T., Li, X., Zhong, C., Tang, W., Guo, J., 2024. DiffMAR: A generalized diffusion model for metal artifact reduction in CT images. IEEE J. Biomed. Health Inf..

Chang, Z., Ye, D.H., Srivastava, S., Thibault, J.-B., Sauer, K., Bouman, C., 2018. Prior-guided metal artifact reduction for iterative x-ray computed tomography. IEEE Trans. Med. Imaging 38 (6), 1532–1542.

Chen, H., Zhang, Y., Zhang, W., Liao, P., Li, K., Zhou, J., Wang, G., 2017. Low-dose CT denoising with convolutional neural network. In: 2017 IEEE 14th International Symposium on Biomedical Imaging. ISBI 2017, IEEE, pp. 143–146.

Croitoru, F.-A., Hondru, V., Ionescu, R.T., Shah, M., 2023. Diffusion models in vision: A survey. IEEE Trans. Pattern Anal. Mach. Intell..

Dhariwal, P., Nichol, A., 2021. Diffusion models beat gans on image synthesis. Adv. Neural Inf. Process. Syst. 34, 8780–8794.

Fang, R., Zhang, S., Chen, T., Sanelli, P.C., 2015. Robust low-dose CT perfusion deconvolution via tensor total-variation regularization. IEEE Trans. Med. Imaging 34 (7), 1533–1548.

Gao, Q., Li, Z., Zhang, J., Zhang, Y., Shan, H., 2023. CoreDiff: Contextual error-modulated generalized diffusion model for low-dose CT denoising and generalization. IEEE Trans. Med. Imaging.

Geng, M., Meng, X., Yu, J., Zhu, L., Jin, L., Jiang, Z., Qiu, B., Li, H., Kong, H., Yuan, J., et al., 2021. Content-noise complementary learning for medical image denoising. IEEE Trans. Med. Imaging 41 (2), 407–419.

Gong, S., Chen, C., Gong, Y., Chan, N.Y., Ma, W., Mak, C.H.-K., Abrigo, J., Dou, Q., 2023. Diffusion model based semi-supervised learning on brain hemorrhage images for efficient midline shift quantification. In: International Conference on Information Processing in Medical Imaging. Springer, pp. 69–81.

Green, M., Marom, E.M., Kiryati, N., Konen, E., Mayer, A., 2016. Efficient low-dose CT denoising by locally-consistent non-local means (LC-NLM). In: Medical Image Computing and Computer-Assisted Intervention-MICCAI 2016: 19th International Conference, Athens, Greece, October 17-21, 2016, Proceedings, Part III 19. Springer, pp. 423–431.

Han, Y.S., Jin, K.H., Kim, K., Ye, J.C., 2016. Sparse-view X-ray spectral CT reconstruction using annihilating filter-based low rank hankel matrix approach. In: 2016 IEEE 13th International Symposium on Biomedical Imaging. ISBI, IEEE, pp. 573–576.

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

Hu, D., Tao, Y.K., Oguz, I., 2022. Unsupervised denoising of retinal OCT with diffusion probabilistic model. In: Medical Imaging 2022: Image Processing. Vol. 12032, SPIE, pp. 25–34.

Humphries, T., Si, D., Coulter, S., Simms, M., Xing, R., 2019. Comparison of deep learning approaches to low dose CT using low intensity and sparse view data. In: Medical Imaging 2019: Physics of Medical Imaging. Vol. 10948, SPIE, pp. 1048–1054.

Jin, P., Bouman, C.A., Sauer, K.D., 2015. A model-based image reconstruction algorithm with simultaneous beam hardening correction for X-ray CT. IEEE Trans. Comput. Imaging 1 (3), 200–216.

Jin, K.H., McCann, M.T., Froustey, E., Unser, M., 2017. Deep convolutional neural network for inverse problems in imaging. IEEE Trans. Image Process. 26 (9), 4509–4522.

Kalender, W.A., Hebel, R., Ebersberger, J., 1987. Reduction of CT artifacts caused by metallic implants. Radiology 164 (2), 576–577.

Karimi, S., Martz, H., Cosman, P., 2015. Metal artifact reduction for CT-based luggage screening. J. X-Ray Sci. Technol. 23 (4), 435–451.

Kazerouni, A., Aghdam, E.K., Heidari, M., Azad, R., Fayyaz, M., Hacihaliloglu, I., Merhof, D., 2023. Diffusion models in medical imaging: A comprehensive survey. Med. Image Anal. 102846.

Kim, H., Chen, J., Wang, A., Chuang, C., Held, M., Pouliot, J., 2016. Non-local total-variation (NLTV) minimization combined with reweighted L1-norm for compressed sensing CT reconstruction. Phys. Med. Biol. 61 (18), 6878.

Li, S., Cao, Q., Chen, Y., Hu, Y., Luo, L., Toumoulin, C., 2014. Dictionary learning based sinogram inpainting for CT sparse reconstruction. Optik 125 (12), 2862–2867.

Li, H., Yang, Y., Chang, M., Chen, S., Feng, H., Xu, Z., Li, Q., Chen, Y., 2022. Srdiff: Single image super-resolution with diffusion probabilistic models. Neurocomputing 479, 47–59.

Lin, W.-A., Liao, H., Peng, C., Sun, X., Zhang, J., Luo, J., Chellappa, R., Zhou, S.K., 2019. Dudonet: Dual domain network for CT metal artifact reduction. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 10512–10521.

Liu, J., Anirudh, R., Thiagarajan, J.J., He, S., Mohan, K.A., Kamilov, U.S., Kim, H., 2023. DOLCE: A model-based probabilistic diffusion framework for limited-angle CT reconstruction. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 10498–10508.

Liu, P., Han, H., Du, Y., Zhu, H., Li, Y., Gu, F., Xiao, H., Li, J., Zhao, C., Xiao, L., et al., 2021. Deep learning to segment pelvic bones: large-scale CT datasets and baseline models. Int. J. Comput. Assist. Radiol. Surg. 16, 749–756.

Liu, X., Xie, Y., Diao, S., Tan, S., Liang, X., 2024. Unsupervised CT metal artifact reduction by plugging diffusion priors in dual domains. IEEE Trans. Med. Imaging.

Lyu, Y., Lin, W.-A., Liao, H., Lu, J., Zhou, S.K., 2020. Encoding metal mask projection for metal artifact reduction in computed tomography. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer, pp. 147–157.

Meyer, E., Raupach, R., Lell, M., Schmidt, B., Kachelrieß, M., 2010. Normalized metal artifact reduction (NMAR) in computed tomography. Med. Phys. 37 (10), 5482–5493.

Müller, N., Siddiqui, Y., Porzi, L., Bulo, S.R., Kontschieder, P., Nießner, M., 2023. Diffrf: Rendering-guided 3d radiance field diffusion. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 4328–4338.

Nichol, A.Q., Dhariwal, P., 2021. Improved denoising diffusion probabilistic models. In: International Conference on Machine Learning. PMLR, pp. 8162–8171.

Ramesh, A., Dhariwal, P., Nichol, A., Chu, C., Chen, M., 2022. Hierarchical text-conditional image generation with clip latents. arXiv preprint arXiv:2204.06125.

Ronneberger, O., Fischer, P., Brox, T., 2015. U-net: Convolutional networks for biomedical image segmentation. In: Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: 18th International Conference, Munich, Germany, October 5-9, 2015, Proceedings, Part III 18. Springer, pp. 234–241.

Sehwag, V., Hazirbas, C., Gordo, A., Ozgenel, F., Canton, C., 2022. Generating high fidelity data from low-density regions using diffusion models. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 11492–11501.

Shen, J., Luo, M., Liu, H., Liao, P., Chen, H., Zhang, Y., 2022. MLF-IOSC: Multi-level fusion network with independent operation search cell for low-dose CT denoising. IEEE Trans. Med. Imaging 42 (4), 1145–1158.

Sohl-Dickstein, J., Weiss, E., Maheswaranathan, N., Ganguli, S., 2015. Deep unsupervised learning using nonequilibrium thermodynamics. In: International Conference on Machine Learning. PMLR, pp. 2256–2265.

Song, Y., Ermon, S., 2019. Generative modeling by estimating gradients of the data distribution. Adv. Neural Inf. Process. Syst. 32.

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

Song, Y., Shen, L., Xing, L., Ermon, S., 2021. Solving inverse problems in medical imaging with score-based generative models. arXiv preprint arXiv:2111.08005.

Wang, H., Li, Y., Meng, D., Zheng, Y., 2022a. Adaptive convolutional dictionary network for CT metal artifact reduction. arXiv preprint arXiv:2205.07471.

Wang, T., Xia, W., Huang, Y., Sun, H., Liu, Y., Chen, H., Zhou, J., Zhang, Y., 2021. Dual-domain adaptive-scaling non-local network for CT metal artifact reduction. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer, pp. 243–253.

Wang, H., Xie, Q., Li, Y., Huang, Y., Meng, D., Zheng, Y., 2022b. Orientation-shared convolution representation for CT metal artifact learning. In: Medical Image Computing and Computer Assisted Intervention–MICCAI 2022: 25th International Conference, Singapore, September 18–22, 2022, Proceedings, Part VI. Springer, pp. 665–675.

Xia, W., Cong, W., Wang, G., 2022a. Patch-based denoising diffusion probabilistic model for sparse-view CT reconstruction. arXiv preprint arXiv:2211.10388.

Xia, W., Lyu, Q., Wang, G., 2022b. Low-dose CT using denoising diffusion probabilistic model for 20x speedup. arXiv preprint arXiv:2209.15136.

Xu, K., Lu, S., Huang, B., Wu, W., Liu, Q., 2024. Stage-by-stage wavelet optimization refinement diffusion model for sparse-view CT reconstruction. IEEE Trans. Med. Imaging.

Yan, K., Wang, X., Lu, L., Summers, R.M., 2018. DeepLesion: automated mining of large-scale lesion annotations and universal lesion detection with deep learning. J. Med. Imaging 5 (3), 036501.

Yang, Q., Yan, P., Zhang, Y., Yu, H., Shi, Y., Mou, X., Kalra, M.K., Zhang, Y., Sun, L., Wang, G., 2018. Low-dose CT image denoising using a generative adversarial network with wasserstein distance and perceptual loss. IEEE Trans. Med. Imaging 37 (6), 1348–1357.

Yu, L., Zhang, Z., Li, X., Xing, L., 2020. Deep sinogram completion with image prior for metal artifact reduction in CT images. IEEE Trans. Med. Imaging 40 (1), 228–238.

Zhang, Y., Hu, D., Hao, S., Liu, J., Quan, G., Zhang, Y., Ji, X., Chen, Y., 2022. DREAM-Net: Deep residual error iterative minimization network for sparse-view CT reconstruction. IEEE J. Biomed. Health Inf. 27 (1), 480–491.

Zhang, Z., Liang, X., Dong, X., Xie, Y., Cao, G., 2018. A sparse-view CT reconstruction method based on combination of DenseNet and deconvolution. IEEE Trans. Med. Imaging 37 (6), 1407–1417.

Zhang, Y., Lv, T., Ge, R., Zhao, Q., Hu, D., Zhang, L., Liu, J., Zhang, Y., Liu, Q., Zhao, W., et al., 2021. CD-Net: Comprehensive domain network with spectral complementary for DECT sparse-view reconstruction. IEEE Trans. Comput. Imaging 7, 436–447.

Zhang, Z., Sejdić, E., 2019. Radiological images and machine learning: trends, perspectives, and prospects. Comput. Biol. Med. 108, 354–370.

Zhang, Y., Yu, H., 2018. Convolutional neural network based metal artifact reduction in x-ray computed tomography. IEEE Trans. Med. Imaging 37 (6), 1370–1381.

Zhou, B., Chen, X., Xie, H., Zhou, S.K., Duncan, J.S., Liu, C., 2022a. DuDoUFNet: Dual-domain under-to-fully-complete progressive restoration network for simultaneous metal artifact reduction and low-dose CT reconstruction. IEEE Trans. Med. Imaging 41 (12), 3587–3599.

Zhou, B., Chen, X., Zhou, S.K., Duncan, J.S., Liu, C., 2022b. DuDoDR-Net: Dual-domain data consistent recurrent network for simultaneous sparse view and metal artifact reduction in computed tomography. Med. Image Anal. 75, 102289.

Zimmermann, R.S., Schott, L., Song, Y., Dunn, B.A., Klindt, D.A., 2021. Score-based generative classifiers. arXiv preprint arXiv:2110.00473.

Zou, M., Ma, X., Wang, W., Wang, K., 2024. Conversion-based reconstruction: a discretized clinical convergence generative network for CT metal artifact reduction. In: International Conference on Future of Medicine and Biological Information Engineering. MBIE 2024, Vol. 13270, SPIE, pp. 25–30.