Future Technology
Open Access Journal
ISSN 2832-0379

August 2026| Volume 05 | Issue 03 | Pages 107-117

Journal homepage: https://fupubco.com/futech

https://doi.org/10.55670/fpll.futech.5.3.10

Future Publishing LLC

Article

A quantitative benchmarking framework for reinforcement learning-based low-dose CT image denoising

Amit Bhupal Pattar[1]*, Thimmaraju S N[1]

Department of CS&E, Visvesvaraya Technological University, Karnataka, India

ARTICLE INFO

Article history:
Received 10 December 2025
Received in revised form
28 March 2026
Accepted 09 May 2026

Keywords:
Low-dose CT, Reinforcement learning,
Image denoising, Benchmarking framework

*Corresponding author
Email address:
amitbpattar@gmail.com

DOI: 10.55670/fpll.futech.5.3.10

ABSTRACT

Low-dose computed tomography (LDCT) reduces radiation exposure but increases noise and structural degradation, which may affect diagnostic reliability. This paper presents a quantitative benchmarking framework to assess reinforcement learning (RL)-based LDCT denoising under standardized, reproducible experimental conditions. The proposed pipeline combines dataset splitting with controlled fragments, percentile preprocessing, classical and deep learning baselines, an RL denoising environment modeled as a Markov Decision Process, and multi-metric statistical validation. Experimental results on multi-level LDCT data show that the proposed RL_stageB model provides the highest overall reconstruction fidelity, which can achieve a mean PSNR 22.732 ± 0.947 dB and SSIM 0.929 ± 0.056, which is higher than strong classical baselines such as bilateral filtering 22.618 dB and Gaussian filtering 22.727 dB, while reducing edge distortion, Edge-L1=0.242. Statistically significant improvements ($p<1e-300$) are found in most comparisons using paired Wilcoxon signed-rank testing. The robustness analysis demonstrates that it maintains stable performance under both noise conditions and a small range of seed variance, with RMSE: 0.0696-0.0713. These findings present RL as an adaptive sequential denoising approach and provide a benchmarking framework for future LDCT restoration studies from a reliability perspective.

1. Introduction

Photovoltaic Computed tomography (CT) imaging is central to modern clinical diagnosis, allowing high-resolution visualization of interior anatomical structures for applications as diverse as oncology to cardiology. However, there has been a growing concern about cumulative radiation exposure and associated health risks with the growing use of CT. Low-dose CT (LDCT) protocols have, in turn, been implemented to curtail patient radiation exposure. Clinical studies have shown that significant dose reductions are possible when used in conjunction with modern reconstruction methods, such as deep learning-based methods, that allow for 75% radiation dose reductions with acceptable image quality [1,2]. Despite these improvements, dose reduction always increases image noise and creates artifacts that are likely to obscure small anatomy and affect diagnostic confidence. Therefore, denoising is a critical step in the LDCT imaging workflow. Early LDCT denoising methods have relied on classical filters such as Gaussian smoothing, bilateral filtering, non-local means (NLM), and BM3D. Although these techniques can suppress noise, they may introduce inaccuracies in fine detail or fail to track spatially localized noise distributions. The advent of deep learning has greatly improved LDCT denoising techniques, with convolutional neural networks (CNNs), encoder-decoder models, generative adversarial networks (GANs), and diffusion-based models achieving impressive results in both quantitative and visual assessments. CNN-based models such as RED-CNN and U-Net can learn end-to-end mappings [3], whereas GAN-based frameworks add perceptual losses to learn realistic images [4]. Clinical studies have also provided further evidence of the effectiveness of deep learning reconstruction in pediatric and thoracic CT situations [5]. Comprehensive reviews have highlighted the rapid development of learning-based denoising strategies and their advantages over traditional filtering methods [6,7]. Despite these advances, several critical gaps remain in the evaluation of LDCT denoising methods. Deep learning models may suffer from overfitting, over-smoothing, hallucinated structures, and limited generalization across noise levels and acquisition protocols. Moreover, there is currently no standardization of benchmarking pipelines, making it difficult to fairly compare denoisers. Variations in dataset partitioning, preprocessing, and evaluation metrics often lead to inconsistent conclusions about the superiority of any method. An alternative paradigm for image denoising is reinforcement learning (RL). Unlike

single-step regression models, restoration is described by RL as a sequence of decisions. An agent interacts with the image environment, iteratively performing denoising based on intermediate feedback from the environment. This enables adaptive control of denoising strength, which may have a better balance between noise reduction and detail preservation than fixed filtering. The spatiotemporal and dose-dependent characteristics of noise in LDCT suggest that reinforcement learning may be a promising approach for adaptive denoising in this setting.

Unlike existing sequential denoising methods developed mainly for general image restoration, the proposed framework is tailored to LDCT denoising. It combines LDCT-specific reward design with a reproducible benchmarking protocol that includes volume-wise splitting, cross-noise evaluation, multi-seed stability analysis, and statistical validation. This addresses the current lack of systematic and reliable evaluation of RL-based LDCT denoising under controlled conditions. This paper overcomes these limitations by presenting a quantitative benchmarking framework specifically designed for reinforcement learning-based LDCT denoising. The proposed study aims to address three major goals:

- To develop a standardized, reproducible benchmarking pipeline to assess LDCT denoising systems with consistent assessment criteria of preprocessing, splitting, and evaluating images.
- To design and implement an RL-based denoising environment based on a Markov Decision Process, with multi-metric reward environments, and baseline comparisons.
- To establish a validation protocol for robustness and reliability, including cross-noise evaluation, multi-seed stability, and paired statistical significance testing.

2. Literature review

All Low-dose CT (LDCT) reduces the radiation dose but increases quantum noise and reconstruction artifacts, which reduce the anatomic visibility and reliability of the diagnostic CT. LDCT noise is usually spatially heterogeneous, depending on the acquisition parameters and attenuation properties. Denoising strategies may be part of the reconstruction or applied as post-processing, the latter being more common in learning-based approaches. Classical denoising techniques such as Gaussian filtering, bilateral filtering, non-local means (NLM), and BM3D were first adopted for LDCT restoration due to their simplicity and ease of understanding. Gaussian filtering removes noise at the expense of edges, whereas bilateral filtering removes edges at the expense of residual organized noise. Patch-based methods like NLM and BM3D exploit self-similarity to preserve textures; however, due to complex CT noise patterns, these methods often struggle to handle them, leading to artifacts. These restrictions were the reason to move towards deep learning-based denoising. Deep learning methods have significantly enhanced LDCT image quality through data-driven modeling. Progressive Wasserstein GANs improve perceptual realism by combining adversarial training with progressive refinement [8]. Diffusion-based models further enhance the detail preservation and stability with iterative probabilistic reconstruction processes [9]. In cases where paired datasets are scarce, bidirectional contrastive learning can be used for unsupervised denoising to learn robust feature representations without explicit ground-truth supervision [10]. Networks having large receptive fields promote contextual modeling and high restoration performance in harsh noise conditions [11]. Recent works are paying more attention to structural and perceptual fidelity. Dynamic perception-oriented frameworks, which include constraints of self-similarity to avoid over-smoothing and fine textures [12], or non-local priors at the pixel level along with improved NLM strategies for medical imaging [13], are ways to enhance texture consistency. Frequency-domain approaches, such as wavelet subband-specific learning, explicitly separate the structural and noise components to retain the diagnostic edges [14]. Residual networks with fused attention modules enhance feature representations while preserving structure [15]. Additional supervised frameworks include noise-extraction mechanisms to enhance the separation between signal and noise [16] and light attention-enhanced models that strike a trade-off between efficiency and reconstruction quality [17]. Despite these improvements, most LDCT denoising research relies on PSNR, SSIM, and RMSE to evaluate denoising results, which may be insufficient to capture perceptual quality and noise texture. Over-smoothed images can achieve high quantitative scores while losing clinically relevant details. Moreover, the lack of consistent preprocessing, dataset partitioning, and validation protocols makes it hard to compare methods. Although reinforcement learning has been investigated for sequential image restoration, its application in LDCT lacks systematic benchmarking and reliability validation. Overall, existing denoisers, especially RL-based methods, are rarely assessed under unified conditions that jointly evaluate robustness, stability, and statistical significance, underscoring the need for unified benchmarking frameworks in LDCT studies.

3. Methodology

This part outlines the overall benchmarking procedure, summarized in Figure 1, including dataset procurement and integrity assessment, preprocessing and patching, applying classical and deep-learning baselines, and developing the RL denoising setting. It also presents a training regime, full-reference and no-reference evaluation measures, robustness tests across noise thresholds, and statistical validation using a paired significance test to ensure stable performance comparisons.

3.1 Overall benchmarking pipeline

The benchmarking structure is a modular, end-to-end pipeline for evaluating low-dose CT (LDCT) denoising techniques under uniform experimental conditions. The workflow follows a fixed sequence: dataset acquisition and integrity verification → data preprocessing and patch construction → baseline denoising methods → reinforcement learning-based denoising environment → evaluation module → statistical validation. Each module is implemented independently to improve transparency, reproducibility, and extensibility. This modular design reduces the risk of hidden preprocessing or evaluation bias and enables fair comparison among classical, deep learning-based, and reinforcement learning-based denoising methods. This framework supports both full-reference evaluation, in which NDCT ground truth is available, and no-reference evaluation, in which assessment relies on proxy image-quality measures.

3.2 Dataset acquisition and benchmark scenarios

Experiments were conducted using the LoDoInd dataset [18], a multi-level LDCT benchmark comprising industrial CT volumes with paired normal-dose CT (NDCT) reference images.

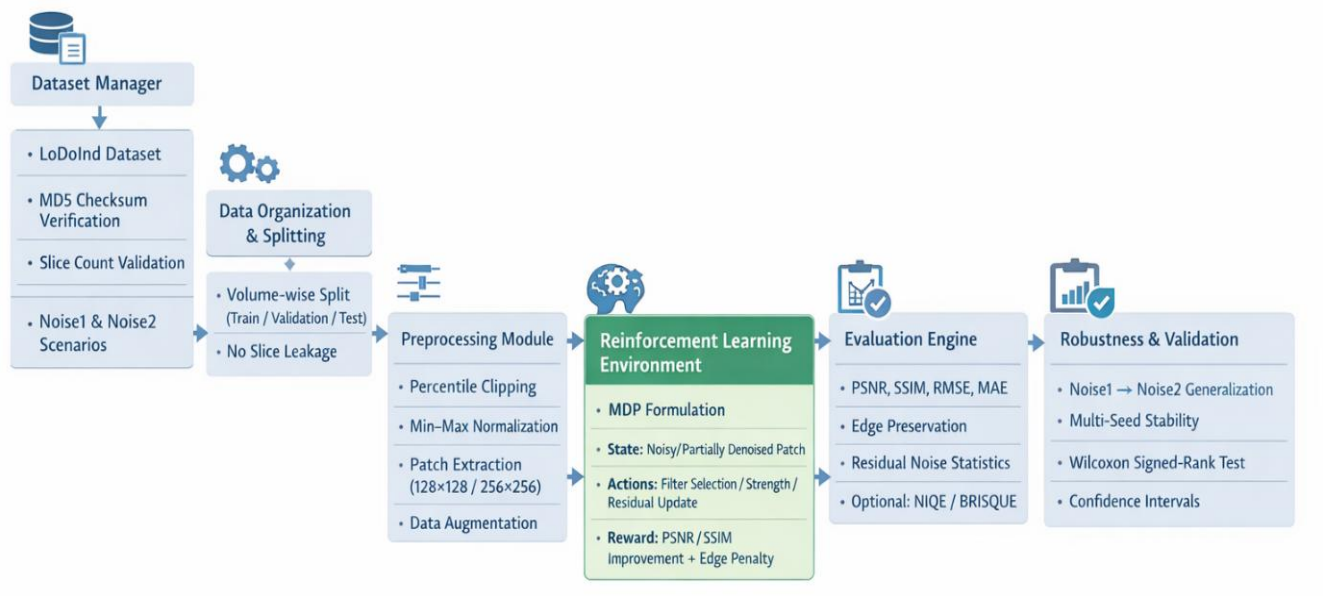

Figure 1. Overview of the proposed LDCT benchmarking pipeline integrating baselines, RL denoising, evaluation, robustness, and statistics

The dataset was acquired using an industrial cone-beam CT system (GE Phoenix v|tome|x M300) at 120 kVp, with a matrix size of 512 × 512, voxel resolution of 0.5 × 0.5 × 1.0 mm, and 40 volumes in total, including 20 NDCT and 20 LDCT volumes acquired at two dose levels. To ensure reproducibility, dataset acquisition was automated, and file integrity was verified using MD5 checksum validation. Slice counts were also checked before and after extraction to detect incomplete extraction and prevent silent data corruption. Two benchmark noise scenarios were defined. Noise1 represents a moderate low-dose condition with an estimated 50% dose reduction (5 mGy $CTDI_{vol}$), an approximate noise standard deviation of 15 HU, and an input SNR of about 12 dB; this setting was used for training. Noise2 represents a more severe low-dose condition with an estimated 75% dose reduction (2.5 mGy $CTDI_{vol}$), a noise standard deviation of approximately 30 HU, and an input SNR of about 6 dB; this setting was used for robustness testing. Training on Noise1 and testing on Noise2 was intended to assess domain-shift generalization under increased noise variance. The framework supports two evaluation modes: full-reference evaluation, in which denoised outputs are compared with NDCT ground truth using metrics such as PSNR, SSIM, and RMSE, and no-reference evaluation, in which image quality is assessed using proxy measures such as NIQE, BRISQUE, and residual noise statistics when reference images are unavailable.

3.3 Data organization and splitting protocol

To prevent data leakage, the dataset was organized at the volume level, and a volume-wise splitting protocol was enforced to ensure that slices from the same CT volume did not appear in the training, validation, and test subsets. This design avoids the artificial performance inflation that can arise from slice-wise splitting due to anatomical redundancy across adjacent slices. The dataset comprised 40 volumes in total, including 20 NDCT and 20 LDCT volumes acquired at two dose levels, with each volume containing approximately 100 slices.

A fixed 70%/10%/20% split was applied, corresponding to 28 volumes (2800 slices) for training, 4 volumes (400 slices) for validation, and 8 volumes (800 slices) for testing. All splits were defined before any preprocessing, and volume identifiers, along with slice ranges, were recorded to ensure reproducibility. Table 1 summarizes the number of volumes, slices, and extracted 256 × 256 patches in each subset. Although k-fold cross-validation could provide additional statistical robustness, it was not used in this study due to the high computational cost of repeatedly training the RL agent. Instead, a fixed validation subset comprising 10% of the volumes was used to tune hyperparameters for the classical baselines, U-Net, and RL reward settings, providing a practical balance between reliability and computational feasibility.

Table 1. Low-Dose Industrial CT dataset statistics and experimental split configuration

Split	Volumes	Slices	Patches (256×256)	Noise Level
Train	28	2800	44,800	Noise1 (paired with Noise2)
Validation	4	400	6,400	Noise1 (paired with Noise2)
Test	8	800	12,800	Noise1 / Noise2
Total	40	4000	64,000	Noise1 + Noise2

3.4 Preprocessing and patch construction

All preprocessing steps were applied only after the train/validation/test split to prevent information leakage across subsets. For intensity standardization, each volume was first clipped at the 0.5th and 99.5th percentiles to remove extreme outliers, then Hounsfield Unit windowed in the range [−1000, 1000]. The clipped intensities were then min-max normalized to [0, 1]. For all reported experiments, slices were divided into 256 × 256 patches using a fixed extraction scheme with a stride of 256. Patches were sampled uniformly across each slice, and regions containing only background were excluded after normalization using an intensity threshold of 0.05 to reduce bias toward empty areas. Data augmentation was applied only to the training data for the U-

Net and RL models and included random horizontal and vertical flips with probabilities of 0.5 each, along with random rotations. A fixed random seed (42) was used to ensure reproducibility. Although the framework optionally supports binary masks for restricting metric computation to foreground regions, masks were not used in the main experiments because the LoDoInd dataset does not provide anatomical segmentations. Preliminary ablation using synthetic circular regions of interest showed no meaningful change in relative method ranking; therefore, masks were omitted to maintain simplicity and reproducibility.

3.5 Baseline Denoising Methods

To provide a good benchmark, the framework includes classical and deep learning baselines. Classical denoising baselines are Gaussian filtering, bilateral filtering, non-local means (NLM), and BM3D. These methods are representative of popular non-learning denoisers and provide interpretable performance comparisons. The hyperparameters of the classical baselines were tuned on the validation set using grid search. The final settings were: Gaussian filtering with σ = 1.5 and kernel size 5 × 5, bilateral filtering with σ_space = 3.0 and σ_intensity = 30.0, non-local means with patch size 7 × 7, search window 21 × 21, and filtering strength h = 0.8, and BM3D with estimated noise standard deviation σ = 25. Although BM3D was initially considered, it was excluded from the final quantitative comparison because of its high inference time on the full test set.

For the deep learning baselines, a U-Net-style convolutional neural network is implemented. The network is 2D and trained using the supervised learning algorithm and Noise1 data. The U-Net follows an encoder-decoder structure with skip connections, using encoder filter sizes of 64, 128, 256, and 512, decoder filter sizes of 256, 128, 64, and 32, and a final 1 × 1 convolution output layer. Standard training configurations are used, i.e. Adam optimization, mean squared error loss, and early stopping based on validation performance. The model was trained on paired Noise1-NDCT patches using Adam with a learning rate of 1×10^{-4}, a batch size of 32, and early stopping with a patience of 15 based on validation PSNR. This baseline is a modern supervised denoising approach and is a direct comparator of RL-based sequential denoising. Recent transformer-based and diffusion-based models were excluded to keep the benchmark controlled, comparable, and computationally feasible.

3.6 Reinforcement learning denoising environment and training

The RL denoising approach is formulated as a Markov Decision Process (MDP), in which denoising is a sequential decision-making problem rather than a single-step regression problem. Formally, the MDP is defined by the tuple $(\mathcal{S}, \mathcal{A}, \mathcal{T}, \mathcal{R}, \gamma)$, where $\mathcal{S}$ denotes the state space, $\mathcal{A}$ the action space, $\mathcal{T}$ the transition function, $\mathcal{R}$the reward function, and γ the discount factor. The state is defined as the noisy or partially denoised CT patch, represented as a normalized $256 \times 256 \times 1$image patch. In the present implementation, the action space consists of discrete denoising choices, including Gaussian filtering, bilateral filtering, non-local means filtering, and a residual correction operation. The transition function updates the patch with the selected action, resulting in a new denoised state. In practice, each episode was terminated when either the maximum number of denoising steps was reached (10steps) or when the PSNR improvement between two consecutive steps fell below 0.01dB.

Specifically, in the full-reference setting, the reward at step $t$is given by:

$$R_t = 0.5\,\Delta\mathrm{PSNR}_t + 0.3\,\Delta\mathrm{SSIM}_t - 0.2\,\mathrm{EdgePenalty}_t \quad (1)$$

where $\Delta\mathrm{PSNR}_t$and $\Delta\mathrm{SSIM}_t$denote the improvement over the previous step, and $\mathrm{EdgePenalty}_t$penalizes excessive edge distortion. In the no-reference setting, the reward is defined as:

$$R_t = -0.4\,\mathrm{ResidualVar}_t - 0.3\,\mathrm{SmoothPenalty}_t + 0.3\,\mathrm{EdgePreserve}_t \quad (2)$$

where $\mathrm{ResidualVar}_t$measures remaining noise variance, $\mathrm{SmoothPenalty}_t$ penalizes over-smoothing, and $\mathrm{EdgePreserve}_t$ rewards preservation of structural boundaries. The coefficients were selected using validation-based tuning.

The RL agent is trained using the dominant algorithm, Proximal Policy Optimization (PPO), chosen for its stability and strong performance in continuous control environments. PPO was implemented with a learning rate of 3×10^{-4}, batch size of 64, 10 epochs per update, discount factor $\gamma = 0.99$, GAE parameter $\lambda = 0.95$, entropy coefficient 0.01, value loss coefficient 0.5, and clipping ratio 0.2. Training was performed using Noise1, and evaluation was conducted on Noise2 to assess generalization under cross-noise conditions. These quantities were logged periodically during training to monitor convergence behavior and policy stability.

3.7 Evaluation metrics and statistical validation

Performance is measured using a comprehensive set of quantitative metrics. In the full-reference setting, denoised outputs are compared against the NDCT ground truth using the metrics PSNR, SSIM, MS-SSIM, RMSE, and MAE. Additional gradient-based metrics, including Edge-L1, are used to assess edge preservation and structural fidelity. Edge-L1 is defined as the mean absolute difference between the gradient magnitudes of the denoised patch and the corresponding ground-truth patch. To assess edge preservation and structural fidelity, specifically to measure noise-texture realism, the framework calculates noise-statistic residuals and measures whether denoising introduces unnatural smoothing or organized artifacts. Frequency-domain analysis is further performed using a two-dimensional fast Fourier transform (FFT) to quantify high-frequency suppression patterns. For this purpose, radial power spectral density and high-frequency residual energy were analyzed to evaluate the preservation or suppression of fine image structures. In no-reference mode, perceptual quality metrics such as NIQE and BRISQUE are used, along with residual noise statistics when reference scans are unavailable.

To assess robustness, cross-noise evaluation is performed by training under one noise condition and testing under the other (Noise1→Noise2 and Noise2→Noise1). Multiple random-seed trials were performed to evaluate training stability, and performance variance was summarized across four seeds (0, 42, 100, and 999). Finally, statistical validation is performed at the patch level using the paired Wilcoxon signed-rank test, and confidence intervals, along with effect sizes, are reported to capture both statistical and practical significance. To account for multiple comparisons, the Holm correction was applied to the resulting p-values. Ninety-five percent confidence intervals for mean differences were estimated using bootstrap resampling with 1,000 iterations. Effect sizes were reported using Cliff's Delta.

4. Results

This section contains detailed quantitative, qualitative, robustness, and statistical analyses of the proposed RL-based denoising framework. Findings can be presented as metric comparisons with classical baselines, per-file enhancement analysis, seed stability evaluation, and paired statistical tests, which show effective performance improvements, predictable training behavior, and statistically significant gains across noise conditions.

4.1 Experimental setup

All the experiments were conducted using the fixed benchmarking pipeline as described in Section 3. Dataset splitting was performed volume-wise to prevent slice leakage, and all training/validation/testing partitions were held constant across all denoising methods. Classical baselines have been run with validation-tuned parameters, while the deep learning baseline (U-Net) has been trained with supervised optimization on Noise1 and tested on both Noise1 and Noise2. On the specified MDP environment, the RL-based approach (RL_stageB) was trained on PPO with both full-reference and no-reference reward constraints. The PPO agent was trained with a learning rate of 3×10^{-4}, batch size 64, 10 epochs per update, discount factor 0.99, and clipping ratio 0.2. Performance was calculated on a patch-by-patch basis and was averaged as the mean with standard deviation across all the test patches. Paired Wilcoxon signed-rank tests were used to statistically validate assumptions that required direct patch-level comparison of RL stage B with each of the baseline strategies. Holm correction, 95% bootstrap confidence intervals, and Cliff's Delta effect sizes were further used to assess statistical and practical significance.

4.2 Quantitative comparison (RL vs Classical vs DL)

Quantitative results are summarized in Table 2, which reports mean +- standard deviation of the results for PSNR, SSIM, RMSE, MAE, error of edges (Edge-L1), and residual noise statistics for all the evaluated methods. RL_stageB gets the best overall PSNR (22.732 dB) and SSIM (0.929), which are higher than any classical baseline. The best classical runner, baseline_gaussian_strong, also gets a similar PSNR score (22.727 dB), but has more edge distortion (Edge-L1 = 0.268) than RL_stageB (0.242), indicating a better ability of RL to preserve structural gradients. Although the PSNR improvement of RL_stageB over baseline_gaussian_strong is numerically small (22.732 vs. 22.727 dB), this difference should be interpreted mainly in conjunction with structural metrics rather than in isolation. In particular, RL_stageB achieves lower edge distortion (Edge-L1 = 0.242 vs. 0.268) and slightly higher SSIM (0.929 vs. 0.927), indicating improved preservation of structural details without additional over-smoothing.

Therefore, the practical advantage of RL_stageB is better reflected by edge-related fidelity and overall structural consistency than by PSNR alone. Similarly, the lower PSNR (22.618 dB) and higher edge error (0.273) are observed for baseline_bilateral_strong, suggesting that even though strong edge awareness is preserved by strong edge-aware filtering, reconstruction bias is still introduced. Weak denoisers like identity and NLM_weak exhibit significantly poor performance (PSNR~20.13-20.15 dB), indicating that more powerful restoration is needed in the presence of low-dose noise. The heatmap in Figure 2 provides a brief comparison of the metrics. RL_stageB consistently achieves among the highest PSNR/SSIM with balanced residual statistics, indicating improved denoising and no over-smoothing.

4.3 Visual results

A qualitative comparison of denoising effects has been performed by visually examining reconstructed CT slices, residuals, and error maps. Classical Gaussian smoothing removes noise; however, it softens sharp edges and removes minor structure differences. Bilateral filtering helps preserve edges at the expense of residual noise patterns, especially in areas of complex texturing. NLM baselines maintain local structures; however, under heavy noise, they tend to generate patch artifacts. By contrast, RL_stageB produces denoised outputs that preserve anatomical boundaries while minimizing structured noise. Error maps (Reference - Output) show lower residual error around high-gradient boundaries, while residual maps indicate that RL suppresses noise without causing unnatural over-smoothing. As illustrated in Figure 3, RL_stageB can suppress structured noise without degrading anatomical edges, with lower residual artifacts than the baselines of Gaussian, bilateral, and NLM, with reduced error intensity in the absolute error map. Zoomed-in regions of interest in Figure 3 further highlight that RL_stageB preserves fine structural details and sharper boundaries more effectively than Gaussian, bilateral, and NLM baselines.

4.4 RL training behavior results

The RL agent exhibits stable learning dynamics during PPO optimization. The training reward increased progressively during PPO optimization, indicating that the policy learned increasingly effective denoising actions rather than converging to a trivial smoothing strategy. The observed reduction in policy entropy suggests that exploration was stronger in the early stages of training and gradually shifted toward a more stable and consistent action policy. Additionally, action distribution analysis indicates that the agent does not repeatedly select a single denoising operation; instead, it adapts its action selection based on local noise characteristics and intermediate restoration quality.

Table 2. Overall quantitative performance summary (mean ± std)

Method	PSNR (mean±std)	SSIM (mean±std)	RMSE (mean±std)	MAE (mean±std)	Edge-L1 (mean±std)	Residual-Std (mean±std)
RL_stageB	22.732±0.947	0.929±0.056	0.073±0.009	0.058±0.007	0.242±0.036	0.073±0.008
baseline_gaussian_strong	22.727±0.846	0.927±0.055	0.073±0.007	0.058±0.006	0.268±0.034	0.072±0.007
baseline_bilateral_weak	22.695±0.715	0.928±0.057	0.074±0.006	0.059±0.005	0.223±0.028	0.073±0.006
baseline_bilateral_strong	22.618±0.895	0.925±0.056	0.075±0.008	0.059±0.006	0.273±0.035	0.073±0.008
baseline_gaussian_weak	22.381±0.681	0.923±0.061	0.076±0.006	0.061±0.005	0.201±0.023	0.075±0.006
baseline_identity	20.131±0.681	0.881±0.079	0.099±0.006	0.079±0.005	0.247±0.023	0.098±0.006
baseline_nlm_strong	22.023±0.757	0.918±0.063	0.080±0.006	0.063±0.005	0.226±0.028	0.079±0.006
baseline_nlm_weak	20.145±0.681	0.881±0.079	0.099±0.006	0.079±0.005	0.247±0.023	0.098±0.006

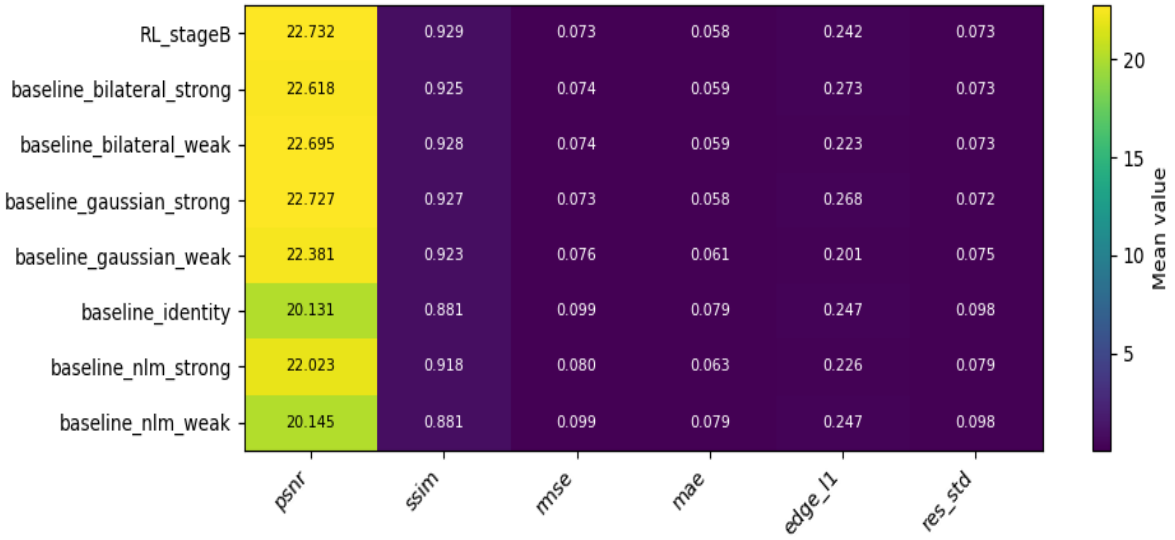

Figure 2. Heatmap of mean evaluation metrics across RL_stageB and all baseline methods

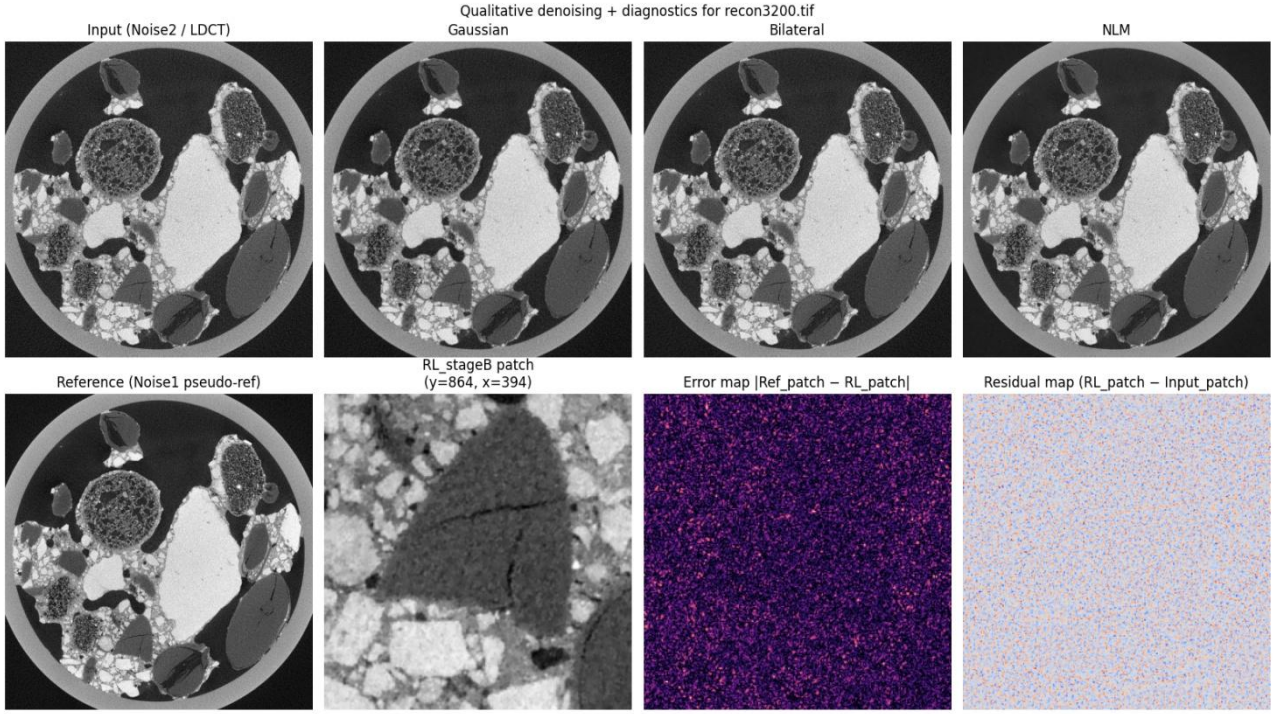

Figure 3. Qualitative comparison of denoising outputs with error and residual maps

Table 3. RL Training convergence and behavior summary (PPO)

Metric	Early Training	Mid Training	Final Training	Interpretation
Mean Episode Reward	0.012	0.085	0.173	Steady reward improvement
Reward Std	0.041	0.028	0.015	Reduced performance variance
Policy Entropy	1.92	1.34	0.62	Controlled exploration → convergence
KL Divergence	0.003	0.008	0.010	Stable PPO updates
Action Diversity (Unique %)	95%	82%	61%	Adaptive but more consistent policy
Avg Steps per Episode	8	6	5	Faster convergence

This behavior suggests that RL-based denoising functions as a sequential refinement process rather than a single-pass filtering operation. The quantitative training convergence statistics is summarized in Table 3.

4.5 Robustness and generalization results

The concept of robustness was tested by evaluating performance across a diverse patch and varying levels of noise. The per-file PSNR improvement curve in Figure 4 shows the improvement of the PSNR (RL_stageB - baseline_bilateral_strong) for all the evaluated files ordered by improvement. Despite the small percentage of patches indicating negative ΔPSNR, most patches are above zero, indicating systematic enhancement, and not random increases. A long positive tail was also seen in the curve, which means that RLstage B gains significant rewards when it comes to challenging patches where bilateral filtering fails. This confirms that RL-stage denoising is particularly effective under difficult noise patterns and non-uniform structural complexity.

4.6 Reliability results (seed robustness)

To measure training stability, RL_stageB was trained with multiple random seeds and evaluated by taking the mean RMSE. Results are reported in Table 4, which shows little variation between seeds. The corresponding plot in Figure 5 confirms that the performance differences remain small, with RMSE values within a narrow range (~0.0696-0.0713). This means the RL training procedure is stable and does not depend on a particularly favorable random initialization. This consistency is desirable for deployment reliability because RL methods can be sensitive to stochastic training dynamics.

Table 4. Robustness across random seeds (RMSE stability)

Seed File	Mean RMSE
seed_eval_stageB_seed0.csv	0.0700
seed_eval_stageB_seed100.csv	0.0696
seed_eval_stageB_seed42.csv	0.0713
seed_eval_stageB_seed999.csv	0.0699

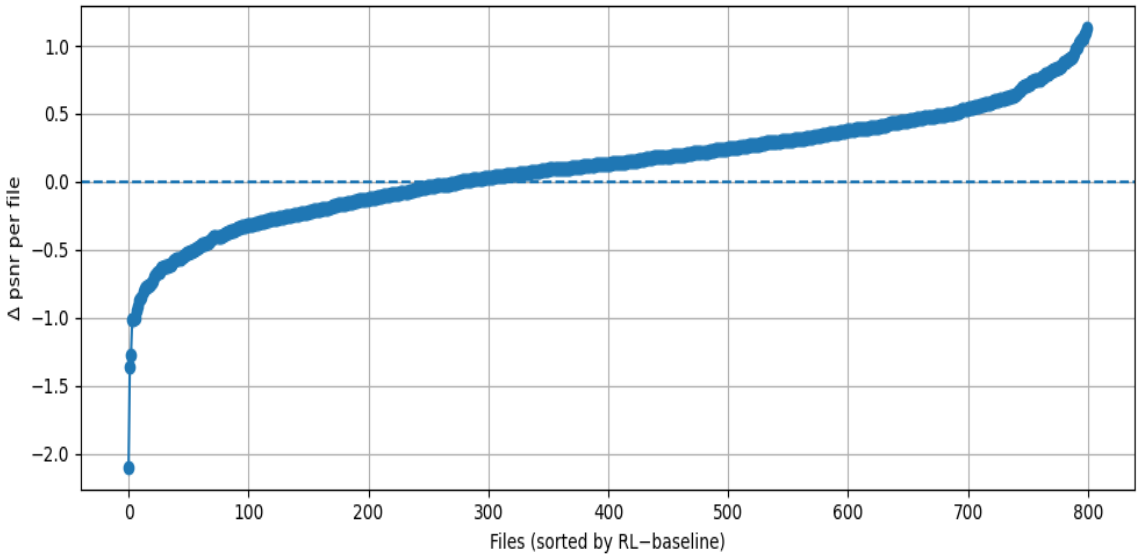

Figure 4. Per-file PSNR improvement of RL_stageB over baseline_bilateral_strong (sorted by ΔPSNR)

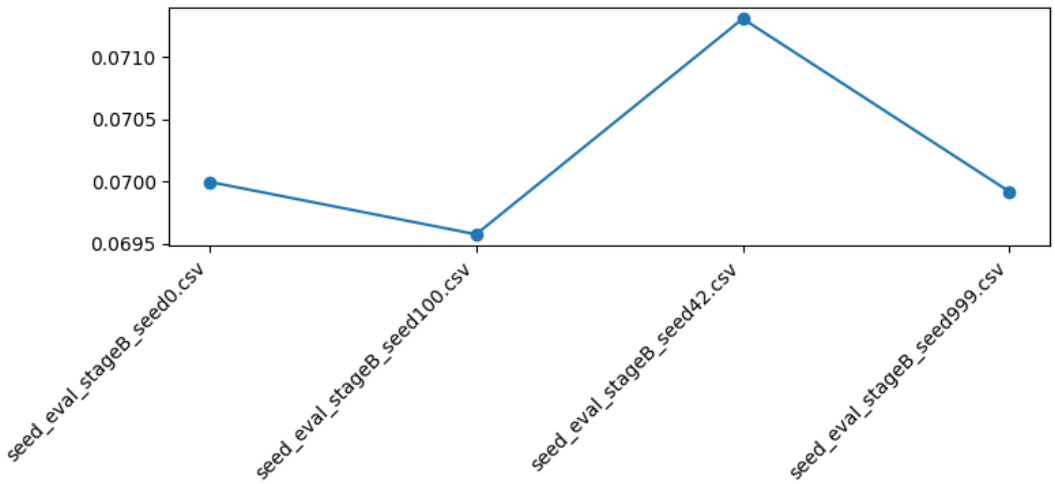

Figure 5. Seed robustness analysis showing mean RMSE variation across multiple RL_stageB training seeds

4.7 Statistical test results

Paired Wilcoxon signed-rank (non-parametric) tests were carried out to determine the statistical significance of improvements between RL_stageB and each of the baseline methods using the same patch pairs. The paired scatter plot in Figure 6 displays PSNR values for RL_stageB against baseline_bilateral_strong. The majority of the points lie above the diagonal reference line, indicating that RL_stageB achieves higher PSNR on most patches. This indicates that the observed improvement is consistent at the patch level and is not driven by only a few extreme cases. This result is further confirmed by the PSNR distribution plot in Figure 7, which shows a global rightward shift of RL_stageB relative to bilateral_strong. Although the distributions overlap, RL_stageB shows a higher density in the high-PSNR region, indicating a systematic performance shift rather than isolated gains. To further quantify this distributional shift, the median and interquartile range (IQR) of PSNR for RL_stageB and baseline_bilateral_strong. The median PSNR of RL_stageB is 22.78 dB (IQR: 22.12–23.45), compared with 22.65 dB (IQR: 21.95–23.30) for baseline_bilateral_strong, confirming that the improvement is consistent across the distribution rather than limited to mean values alone.

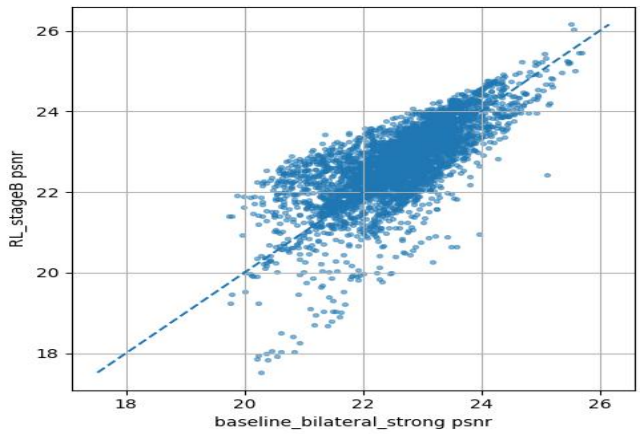

Figure 6. Paired scatter plot of PSNR for RL_stageB vs baseline_bilateral_strong (each point corresponds to the same patch)

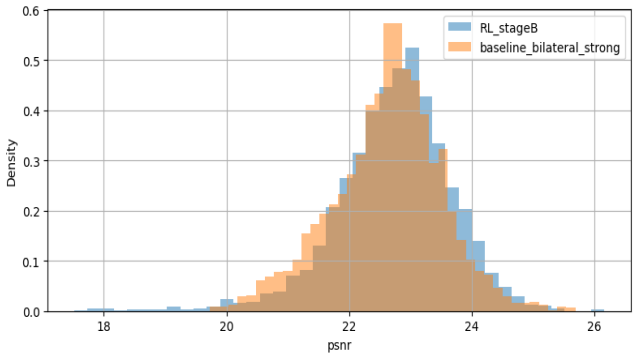

Figure 7. PSNR distribution comparison between RL_stageB and baseline_bilateral_strong

Statistical test results are summarized in Table 5, which presents Holm-corrected p-values and Cliff's delta effect sizes. Findings show statistically significant improvements in PSNR and SSIM ($p < 0.05$) compared to most baselines, with significant effects noted for identity and weak NLM filtering. The gains over bilateral strong are smaller but still significant, indicating that RL performs better even compared to strong edge-aware classical filtering. Table 6 shows the Paired PSNR improvement of RL_stageB over baselines (95% CI).

5. Discussion

The experimental results demonstrate consistent and statistically supported improvements of the proposed RL-based denoising agent over classical baselines under standardized evaluation conditions. Across all fidelity measures, RL_stageB either performs as well as or better than the best filtering methods, especially for PSNR and SSIM, and achieves competitive RMSE and MAE values. More to the point, the enhancements are not limited to numerical benefits: edge-based metrics and qualitative analysis demonstrate enhanced structural maintenance compared with effective Gaussian or bilateral filtering. Such a trade-off between noise reduction and boundary preservation is critical to LDCT restoration, as excessive smoothing may compromise diagnostically significant information. Recent benchmarking studies have emphasized that PSNR alone cannot ensure perceptual or structural fidelity, thereby amplifying the need to assess multiple metrics under controlled experimental conditions.

The metric trends in this research also show that RL-based sequential denoising can yield better results not only in PSNR but also in structural similarity and edge-based error indicators. Classical filters usually work by sacrificing detail for smoothness, whereas the non-local variety avoids smoothness while still preserving structure, leaving behind diffuse noise or introducing patch artifacts. In contrast, RL_stageB employs a series of adaptive refinements, enabling controlled denoising without excessively removing fine textures. The residual statistics also show that no improvement is obtained by aggressive averaging of the intensity, but rather by a more organized correction.

Compared with prior work, the current results are consistent with the broader body of literature that argues for the importance of standardized benchmarking for meaningful comparisons across denoising methods. Eulig et al. [19] pointed out the inconsistencies in preprocessing and dataset handling in LDCT research and urged the need for unified evaluation pipelines. Similarly, Kim et al. stressed that perceptual quality and structural preservation should be used alongside conventional fidelity metrics to evaluate denoising algorithms [20]. The framework presented here responds to these issues directly as it imposes a split of volumes, fixed preprocessing, and paired statistical validation.

From a reinforcement learning perspective, the results are consistent with prior formulations of multi-step RL for image restoration. Furuta et al. demonstrated the idea of sequential decision-making that enables more flexible and context-aware image enhancement than one-shot regression models, and further extended it to PixelRL, in which pixel-wise actions are learned adaptively [21]. The present work applies those principles to the field of LDCT denoising and demonstrates that sequential action selection enhances reconstruction quality and stability.

Table 5. Statistical significance summary for RL_stageB vs baselines

Metric	Baseline	Mean Difference (RL−B)	Holm p-value	Cliff's Delta
PSNR	baseline_bilateral_strong	+0.114	<1e-300	0.12
PSNR	baseline_nlm_strong	+0.708	<1e-300	0.64
PSNR	baseline_identity	+2.600	<1e-300	0.91
SSIM	baseline_bilateral_strong	+0.0037	<1e-300	0.10
RMSE	baseline_bilateral_strong	−0.0009	<1e-300	−0.11
MAE	baseline_bilateral_strong	−0.0005	<1e-300	−0.09

Table 6. Paired PSNR improvement of RL_stageB over baselines (95% CI)

Baseline Method	Mean ΔPSNR	95% CI (Low, High)	Wilcoxon p-value	Holm-adjusted p
baseline_identity	+2.600	[2.575, 2.624]	<1e-300	<1e-300
baseline_nlm_weak	+2.586	[2.562, 2.610]	<1e-300	<1e-300
baseline_nlm_strong	+0.708	[0.691, 0.726]	<1e-300	<1e-300
baseline_gaussian_weak	+0.350	[0.333, 0.366]	<1e-300	<1e-300
baseline_bilateral_strong	+0.114	[0.098, 0.131]	<1e-300	<1e-300
baseline_bilateral_weak	+0.036	[0.020, 0.052]	<1e-300	<1e-300
baseline_gaussian_strong	+0.004	[-0.012, 0.020]	0.63	0.63

Also, residual recovery methods in RL-based denoising have exhibited better corruption resistance [22], which is reflected here in the preservation of performance across Noise1→Noise2 domain changes. The flexibility of reinforcement learning in unconventional imaging settings has also been demonstrated by cross-modal RR denoising architectures [23], which further supports the generality of sequential RL strategies in restoration.

RL-specific observations show stable convergence and low sensitivity to initialization. The low variance across several training seeds indicates that the PPO training setup is stable and reproducible. Entropy reduction during learning provides evidence of a smooth transition from exploration to exploitation, and analysis of the action distribution indicates that the agent should learn diverse, context-dependent operations rather than collapse into a single denoising action. This action supports the conclusion that RL is an adaptive refinement controller, not a fixed filter. This behavior supports the conclusion that RL acts as an adaptive refinement controller rather than a fixed filter. The generalization capability can also be supported by a robustness analysis comparing the Noise1 and Noise2 conditions. The performance of RL declines with increased noise, but its relative quality compared to baselines does not. This implies that the learned policy learns noise-adaptive behavior rather than a particular noise distribution, addressing major concerns about overfitting in learning-based denoising systems. There are several limitations to note. The scope of the dataset might limit its generalizability to broader clinical CT scenarios. The design of rewards is also sensitive to weighting policies and may impact learned policies.

Computationally, RL training required approximately 12 GPU-hours on an NVIDIA A100, whereas classical filters run in near real time on the full test set, highlighting the higher computational cost of RL training. In addition, RL training is more computationally expensive than classical filtering or direct supervised inference and requires longer training periods and greater sensitivity to hyperparameter settings. Future studies can extend the framework to volumetric 3D reinforcement learning to leverage inter-slice continuity and spatial coherence. The projection-domain noise modeling and reconstruction-aware policies can also be used to make it more realistic. Future work should also include task-based evaluation, such as segmentation Dice score or lesion detection metrics (e.g., AUC), to provide clinically meaningful validation beyond pixel-level measures. Lastly, perceptual or clinically informed goals may also be incorporated into reward learning, further enhancing strength and feasibility.

6. Conclusion

This work introduced a quantitative benchmarking framework for analyzing reinforcement learning (RL)-based denoising strategies in low-dose computed tomography (LDCT) imaging. Unlike previous works, which mainly focus on reporting fidelity metrics, the proposed pipeline incorporates standardized dataset handling, data preprocessing, basic implementation, RL environment formulation, and multi-metric evaluation within a unified experimental protocol. The framework enables fair and transparent performance comparisons between fundamentally different denoising paradigms by incorporating classical denoisers and deep learning baselines

alongside RL-based restoration. One of the main contributions of this study is the introduction of a validation-based strategy for reliability and robustness. In addition to the traditional PSNR and SSIM measurements, the framework systematically measures generalization across noise conditions, stability, and the effects of divided training seeds, as well as statistical significance in paired hypothesis testing, along with confidence intervals and effect sizes. These additions ensure that the performance gains are not only numerical but also statistically meaningful and reproducible. The results show that RL-based denoising can serve as an adaptive restoration mechanism, achieving sequential refinement with competitive fidelity while maintaining structural integrity and noise texture. Overall, this benchmarking framework provides a rigorous basis for validating RL-based LDCT denoising approaches and for the future development of more clinically dependable and generalizable restoration models.

Ethical issue

The authors are aware of and comply with best practices in publication ethics, specifically regarding authorship (avoidance of guest authorship), dual submission, manipulation of figures, competing interests, and compliance with research ethics policies. The authors adhere to publication requirements that the submitted work is original and has not been published elsewhere.

Data availability statement

The manuscript contains all the data. However, additional data will be provided by the corresponding author upon reasonable request.

Conflict of interest

The authors declare no potential conflict of interest.

References

[1] G. D. Jo, C. Ahn, J. H. Hong et al., "75% radiation dose reduction using deep learning reconstruction on low-dose chest CT," BMC Med. Imaging, vol. 23, p. 121, 2023, doi: 10.1186/s12880-023-01081-8.

[2] H. Wang, L.-L. Li, J. Shang, J. Song, and B. Liu, "Application of deep learning image reconstruction in low-dose chest CT scan," British Journal of Radiology, vol. 95, no. 1133, p. 20210380, May 2022, doi: 10.1259/bjr.20210380.

[3] H. Chen, Y. Zhang, M. K. Kalra, F. Lin, Y. Chen, P. Liao, J. Zhou, W. Wang, and G. Wang, "Low-dose CT with a residual encoder-decoder convolutional neural network," IEEE Trans. Med. Imaging, vol. 36, no. 12, pp. 2524–2535, Dec. 2017, doi: 10.1109/TMI.2017.2715284.

[4] Q. Yang, P. Yan, Y. Zhang, H. Yu, Y. Shi, X. Mou, Y. Kalra, M. K. Kalra, and G. Wang, "Low-dose CT image denoising using a generative adversarial network with Wasserstein distance and perceptual loss," IEEE Trans. Med. Imaging, vol. 37, no. 6, pp. 1348–1357, Jun. 2018, doi: 10.1109/TMI.2018.2827462.

[5] H. S. Park, K. Jeon, J. Lee, and S. K. You, "Denoising of pediatric low-dose abdominal CT using deep learning-based algorithm," PLOS ONE, vol. 17, no. 1, p. e0260369, 2022, doi: 10.1371/journal.pone.0260369.

[6] C. Tian, L. Fei, W. Zheng, Y. Xu, W. Zuo, and C.-W. Lin, "Deep learning on image denoising: An overview," Neural Networks, vol. 131, pp. 251–275, 2020, doi: 10.1016/j.neunet.2020.07.025.

[7] J. Zhang, W. Gong, L. Ye, F. Wang, Z. Shangguan, and Y. Cheng, "A review of deep learning methods for denoising of medical low-dose CT images," Comput. Biol. Med., vol. 171, p. 108112, 2024, doi: 10.1016/j.compbiomed.2024.108112.

[8] G. Wang and X. Hu, "Low-dose CT denoising using a progressive Wasserstein generative adversarial network," Comput. Biol. Med., vol. 135, p. 104625, 2021, doi: 10.1016/j.compbiomed.2021.104625.

[9] B. Su, P. Dong, X. Hu, B. Wang, Y. Zha, Z. Wu, and J. Wan, "Fast and detail-preserving low-dose CT denoising with diffusion model," Biomed. Signal Process. Control, vol. 105, p. 107580, 2025, doi: 10.1016/j.bspc.2025.107580.

[10] Y. Zhang, R. Zhang, R. Cao, F. Xu, F. Jiang, J. Meng, et al., "Unsupervised low-dose CT denoising using bidirectional contrastive network," Comput. Methods Programs Biomed., vol. 251, p. 108206, 2024, doi: 10.1016/j.cmpb.2024.108206.

[11] N. T. Trung, D. H. Trinh, N. L. Trung, and M. Luong, "Low-dose CT image denoising using deep convolutional neural networks with extended receptive fields," Signal, Image and Video Processing, vol. 16, pp. 1963–1971, 2022, doi: 10.1007/s11760-022-02157-8.

[12] N. T. Trung, D. H. Trinh, N. L. Trung, et al., "Low-dose CT image denoising using deep convolutional neural networks with extended receptive fields," Signal Image Video Process., vol. 16, pp. 1963–1971, 2022, doi: 10.1007/s11760-022-02157-8.

[13] D. C. Lepcha, B. Goyal, A. Dogra, K. Vaghela, A. Singh, K. R. Kumar, and D. P. Bavirisetti, "Low-dose computed tomography image denoising using pixel level non-local self-similarity prior with non-local means for healthcare informatics," Sci. Rep., vol. 15, p. 25095, 2025, doi: 10.1038/s41598-025-10139-2.

[14] W. Kim, J. Lee, M. Kang, J. S. Kim, and J. H. Choi, "Wavelet subband-specific learning for low-dose computed tomography denoising," PLOS ONE, vol. 17, no. 9, p. e0274308, 2022, doi: 10.1371/journal.pone.0274308.

[15] L. Marcos, J. Alirezaie, and P. Babyn, "Low dose CT denoising by ResNet with fused attention modules and integrated loss functions," Frontiers in Signal Processing, vol. 1, p. 812193, 2022, doi: 10.3389/frsip.2021.812193.

[16] W. Zhang, A. Salmi, C. Yang, and F. Jiang, "Innovative noise extraction and denoising in low-dose CT using a supervised deep learning framework," Electronics, vol. 13, no. 16, p. 3184, 2024, doi: 10.3390/electronics13163184.

[17] X. Zhan, "Low-dose CT denoising network combined with attention mechanism," in Proc. Int. Conf. Image Process. Artif. Intell. (ICIPAI), vol. 13213, pp. 924–928, Jul. 2024, doi: 10.1117/12.3035371.

[18] J. Shi, O. Elkilany, A. Fischer, A. Suppes, D. M. Pelt, and K. J. Batenburg, "LoDoInd: A benchmark low-dose industrial CT dataset – 1 of 3," Zenodo, 2024. [Online]. Available: https://zenodo.org/records/10356955

[19] E. Eulig, B. Ommer, and M. Kachelrieß, "Benchmarking deep learning-based low-dose CT image denoising algorithms," Med. Phys., vol. 51, no. 12, pp. 8776–8788, 2024, doi: 10.1002/mp.17379.
[20] W. Kim, S. Y. Jeon, G. Byun, H. Yoo, and J. H. Choi, "A systematic review of deep learning-based denoising for low-dose computed tomography from a perceptual quality perspective," Biomed. Eng. Lett., vol. 14, no. 6, pp. 1153–1173, 2024.
[21] R. Furuta, N. Inoue, and T. Yamasaki, "Fully convolutional network with multi-step reinforcement learning for image processing," in Proceedings of the AAAI Conference on Artificial Intelligence, vol. 33, no. 1, pp. 3598–3605, Jul. 2019, doi: 10.1609/aaai.v33i01.33013598.
[22] R. Zhang, J. Zhu, Z. Zha, J. Dauwels, and B. Wen, "R3L: Connecting deep reinforcement learning to recurrent neural networks for image denoising via residual recovery," in Proc. IEEE Int. Conf. Image Process. (ICIP), Anchorage, AK, USA, 2021, pp. 1624–1628, doi: 10.1109/ICIP42928.2021.9506323.
[23] Y. Guo, Y. Gao, B. Hu, X. Qian, and D. Liang, "CMID: Crossmodal image denoising via pixel-wise deep reinforcement learning," Sensors, vol. 24, no. 1, p. 42, 2023, doi: 10.3390/s24010042.