Multimedia Tools and Applications
https://doi.org/10.1007/s11042-024-20175-1

1239: EMERGING TRENDS AND APPLICATIONS OF DEEP LEARNING FOR BIOMEDICAL DATA ANALYSIS

Residual encoder-decoder based architecture for medical image denoising

Abdesselam Ferdi[1,2] · **Said Benierbah**[1] · **Amir Nakib**[2]

Received: 14 April 2023 / Revised: 5 August 2024 / Accepted: 28 August 2024

Abstract
High-resolution computed tomography (CT) scans require high doses of X-rays, posing potential health risks to patients, including genetic damage and cancer. Conversely, low doses of X-rays result in noise and artifacts in the reconstructed CT scans. Consequently, the problem of denoising low-dose CT (LDCT) images has become a critical yet challenging issue in the field of CT imaging. However, existing deep learning-based LDCT image denoising methods frequently result in the loss of high-frequency features, such as edges and textures, due to the use of mean squared error loss. To address this issue, we propose a method based on high-frequency feature learning to enhance the denoising performance of existing models. Our method is designed to simultaneously learn the primary task of LDCT image denoising and the auxiliary task of LDCT edge detection, thereby improving the denoising performance without increasing the number of model parameters and the inference time. Our method significantly improves the denoising performance of the RED-CNN model, achieving competitive results compared to state-of-the-art denoising models on the AAPM and Qin-LUNG-CT datasets.

Keywords Low dose CT · Image denoising · Edge detection · Multi task learning

1 Introduction

High-resolution computed tomography (HRCT) represents a pivotal tool in modern medicine, offering detailed insights into the human body's interior through the use

✉ Amir Nakib
nakib@u-pec.fr

Abdesselam Ferdi
abdesselam.ferdi@gmail.com

Said Benierbah
said_benierbah@umc.edu.dz

1 SP-LAB, Constantine 1 - Frères Mentouri University, 25017 Constantine, Algeria

2 LISSI, Université Paris-Est Créteil, 94400 Paris, France

Published online: 30 September 2024

of collimated X-rays and advanced image processing techniques. HRCT is employed in both screening tests and disease diagnosis, allowing physicians to examine internal body structures in detail and assess features such as density, size, shape, and texture. Despite its widespread use, HRCT has prompted concerns regarding patient safety due to the potential adverse effects of augmented X-ray radiation exposure.

While CT scans provide high-resolution structural information, they also expose patients to cumulative X-ray radiation, which carries potential health risks, including an increased risk of cancer [1]. Consequently, there is a compelling rationale for reducing radiation doses. Nevertheless, this reduction in radiation dose may result in a degradation of image quality, leading to the introduction of noise and artifacts in the reconstructed images [2–6]. This necessitates the development of algorithms capable of denoising CT images obtained under low-dose radiation, ensuring accurate diagnoses and correct CT examination outcomes.

A plethora of algorithms have been developed with the aim of enhancing the quality of low-dose CT (LDCT) images. These can be broadly categorized into three groups: sinogram filtering [7–10], iterative reconstruction [11–17], and image post-processing [2, 3, 18–28].

Sinogram filtering entails the processing of raw data prior to the application of the filtered back-projection algorithm, which is used for image reconstruction. Commonly employed methods include structural adaptive filtering [7], bilateral filtering [8], and penalized weighted least-squares techniques [9]. Recently, M. Patwari et al. [10] proposed a complex reconstruction and denoising framework based on the bilateral filter for LDCT images. In contrast, iterative reconstruction algorithms reconstruct the CT image iteratively, using prior information on noise and image content [11–15, 17]. Despite the efficacy of these two categories, these algorithms are frequently constrained in clinical settings due to the challenge of acquiring vendor-specific projection data. Furthermore, these methods are subject to inherent limitations, including spatial resolution loss and substantial computational overhead costs. Image post-processing algorithms operate directly on the reconstructed LDCT images. This approach has gained popularity with the advent of deep learning technology in medical image denoising. Traditional image post-processing algorithms, such as block-matching 3D [18, 19], non-local means [20], dictionary learning-based algorithms [21, 22], and diffusion filters [23], have been largely superseded due to their inability to effectively remove non-uniform noise distributions in the reconstructed LDCT images. This results in over-smoothness and structural distortion. Nevertheless, deep learning has demonstrated considerable potential in addressing the limitations of traditional algorithms in various medical imaging tasks, including segmentation [29–31], classification [32–34], and denoising [2, 3, 24, 35].

A plethora of deep learning models have been proposed to denoise LDCT images, each employing a distinct loss function. The majority of these algorithms are trained using the mean squared error (MSE) loss between the predicted normal-dose CT (NDCT) and the ground-truth NDCT images [2, 4, 25, 36, 37]. Despite their impressive performance, methods based on MSE or weighted-MSE loss often result in the attenuation of high-frequency components, which can lead to overly smoothed images [5, 38]. To address this issue, alternative loss functions, including adversarial, perceptual, and similarity losses, have been employed. Adversarial loss can be employed to learn the mapping from LDCT to NDCT images. However, this approach may result in the loss of individual image content in certain instances. Perceptual loss employs a pre-trained convolutional neural network (CNN) model on ImageNet dataset to emulate the human visual system in recognizing images [39]. However, it is not as effective as other loss functions in removing noise. Similarity loss serves as a substitute

for perceptual loss, thereby preserving both structural and textual information. Q. Yang et al. [3] proposed the use of the generative adversarial network (WGAN) with the wasserstein distance and perceptual loss to denoise LDCT images. The model was trained and validated on the AAPM dataset, which consisted of 100K training and 2K validation image patches. The model demonstrated efficiency in improving image quality and statistical properties, achieving a peak signal-to-noise ratio (PSNR) of 23.39 and structural similarity index (SSIM) of 0.79. H. Shan et al. [27] developed a convolutional encoder-decoder model based on conveying paths in 2D and 3D configurations in the GAN framework for LDCT denoising. The model was trained and validated on the AAPM and MGH datasets, achieving a PSNR of 30.14 and SSIM of 0.9 on the AAPM test set. M. Li et al. [5] proposed a self-attention CNN (SACNN) with a perceptual loss function for LDCT denoising. The model achieved a PSNR of 27.74 and SSIM of 0.89. G. Wang et al. [40] proposed a progressive wasserstein GAN with the weighted structurally-sensitive hybrid loss function (PWGAN-WSHL) to address LDCT denoising. The model achieved a PSNR of 24.87, SSIM of 0.80, and RMSE of 0.06. Y. Tang et al. [28] developed a content-noise complementary learning model with contrastive learning (CCN-CL) to denoise LDCT images. The model achieved a PSNR of 44.98, SSIM of 0.96, and RMSE of 0.01. J. Liu et al. [4] proposed a deviant feature sensitive noise estimate network (DFSNE-Net) to address the LDCT denoising task. The model achieved a PSNR of 29.25, a SSIM of 0.89, and a RMSE of 0.03 on the AAPM test set.

The key question we address is: How can we design a model that effectively denoises LDCT images while preserving high-frequency features?

We propose a novel architecture designed for medical image denoising, specifically targeting LDCT images. Our objective is to denoise these images while preserving high-frequency features such as edges. To achieve this, we employed a dual-task training approach, wherein the main task focuses on LDCT image denoising, while the auxiliary task centers on LDCT image edge detection. We first integrated a low-frequency feature learning module in an encoder-decoder architecture to extract low-frequency features from the deepest layer of the model's encoder part. Then we subtracted the low-frequency features from the predicted NDCT image to obtain the high-frequency features (i.e., edges). In the main task, the optimizer will minimize the MSE loss between the predicted NDCT images and the ground-truth NDCT images. Concurrently, in the auxiliary task, the MSE loss between the output NDCT edge and the ground-truth NDCT edge will be minimized. In the inference phase, we removed the low-frequency feature learning module from the model. This step ensures that the model maintains the same number of parameters and inference time as the original model, thereby optimizing its denoising performance without compromising efficiency. We evaluated our method on the AAPM and Qin-LUNG-CT datasets with the RED-CNN model and obtained SOTA denoising results compared to existing methods.

Our contributions are as follows: (1) We propose a novel architecture designed for medical image denoising. (2) To the best of our knowledge, we are the first to introduce a method for LDCT image denoising that is based on the learning of high-frequency features. (3) Our method maintains the same inference time as the original model, thereby ensuring that efficiency is not compromised.

The rest of the paper is organized as follows: We present the problem of LDCT denoising and the proposed method in Section 2. In Section 3, we detail the experimental setup used to validate our method. In Section 4, we present the results obtained from our experiments and conduct an ablation study on the hyperparameters of our method. Finally, in Section 5, we summarize the work and propose a direction for future development of the proposed method.

2 Materials and methods

This section presents the formulation of the LDCT denoising problem and provides an overview of our proposed method. It includes an overview of multi-task learning, the proposed architecture, the low-frequency feature generation module, and the employed loss function.

2.1 LDCT denoising problem formulation

The problem of denoising LDCT images can be defined as a model designed to reduce noise within the spatial domain of the image. To illustrate, consider that I_{LD} denotes an LDCT image, and I_{ND} represents the corresponding NDCT image. A denoiser d is a model that maps an LDCT image to a corresponding NDCT image:

$$d(I_{LD}) = I_{ND} \tag{1}$$

2.2 Multi-task learning

Single-task learning (STL) is a training method that involves developing a model to perform a single task at a time. The performance of such a model is influenced by a number of factors, including the network architecture, loss function, optimizer, and dataset. However, STL models face significant challenges, particularly in the medical field, due to the necessity for a substantial amount of training data and the issue of data sparsity in different tasks where labeled data is limited. Furthermore, the training of multiple STL models to perform different tasks can be resource-intensive. In contrast, multi-task learning (MTL) can address these challenges. MTL enables a model to learn from the shared knowledge of other tasks using labeled data, thereby reducing the cost of manual data labeling and enabling the performance of multiple tasks with a single model. MTL is a learning paradigm in machine learning that aims to simultaneously learn different tasks in order to enhance the learning ability of a model for each task. This is achieved by leveraging the knowledge contained in all tasks. In comparison to their STL counterparts, MTL models frequently demonstrate superior performance with a reduced risk of overfitting for each task. This is due to the fact that, with a greater quantity of data from different tasks, the model is able to learn more robust and universal representations for multiple tasks. Consequently, MTL represents a more efficient and effective approach to machine learning [41].

In supervised learning tasks, a training dataset, denoted as $D^t = \{x_i^t, y_i^t\}_{i=1}^{N_t}$, is associated with a specific task, $t \in T$. Each training instance, x_i, in D^t has a corresponding label, y_i, and N_t represents the total number of training samples. The notation x_i^t signifies the training sample for task t. In the context of MTL, the concept of shared knowledge can be expressed in terms of features, instances, or parameters. Feature-based MTL is a method of learning common features across different tasks. Instance-based MTL facilitates the transfer of knowledge across tasks through instances, which are identified from data instances deemed useful in another task. Parameter-based MTL facilitates the learning of model parameters in different tasks by using the model parameters of another task. Among these, feature-based and parameter-based MTL are the most commonly used methods.

Feature-based MTL is based on feature learning, whereby the model learns to identify and use common feature representations for multiple tasks. These representations may be either a subset or a transformation of the original feature representation. This approach can

be divided into two categories based on the relationship between the original and learned representations: the feature transformation approach and the feature selection approach. In the feature transformation approach, the learned feature representation differs from the original one and is a linear or nonlinear transformation of the latter. In the feature selection approach, the learned representation is a subset of the original representation, selected by the model. Subsequently, irrelevant feature representations are eliminated based on various criteria.

Recently, a plethora of MT models have been proposed for a multitude of computer vision tasks [42–45]. This is due to the enhanced performance that can be achieved by learning shared feature representations from multiple supervisory tasks. In the context of image denoising, the denoising performance can be enhanced by training a MT model on an auxiliary task. In this work, we employed the feature-based MTL method in order to share low-frequency features extracted from the model's encoder with the decoder part, with the objective of predicting the NDCT image and its associated edges.

2.3 Proposed architecture

To enhance the performance of LDCT image denoising in deep learning models, we integrated a low-frequency feature learning module in an encoder-decoder architecture to extract low-frequency features from the deepest layer of the model's encoder part. Then we subtracted the extracted low-frequency features from the predicted NDCT image to obtain the high-frequency features (i.e., edges).

We demonstrated the effectiveness of our method using the RED-CNN model, a potent LDCT image denoising algorithm with an encoder-decoder architecture. A schematic representation of the proposed architecture is illustrated in Fig. 1.

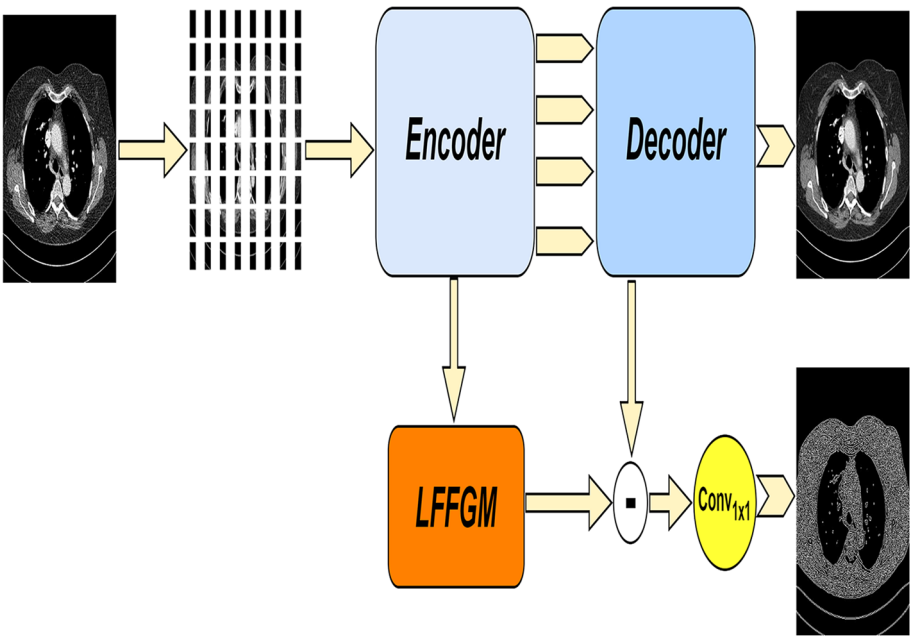

Fig. 1 Summary of our proposed architecture

The RED-CNN model, an encoder-decoder CNN comprising 10 layers, generates the low-frequency feature maps of the LDCT images as the output of the ReLU function following the final convolutional layer. The output of the ReLU function following the penultimate deconvolutional layer is responsible for generating the feature maps of the predicted NDCT image. Subsequently, a 1×1 deconvolutional layer with a ReLU activation function is applied to these feature maps, thereby generating the predicted NDCT image. [2]. The LDCT images contain both low- and high-frequency components. The edges, which are high-frequency features, can be computed by subtracting the low-frequency features from the whole features of the predicted NDCT image. To facilitate the extraction of low-frequency features of the input LDCT image from the model's encoder, we used a low-frequency feature generation module (LFFGM) [46]. Then we subtracted the output of the LFFGM from the feature maps of the predicted NDCT images to obtain the high-frequency features of the NDCT images. Subsequently, we applied a convolutional layer with a 1×1 kernel size and a sigmoid activation function to the high-frequency features in order to generate the predicted NDCT edges.

During the training phase, we trained the model on both the primary (LDCT image denoising) and auxiliary (LDCT edge detection) tasks using the corresponding ground truths. In the inference phase, we removed the LFFGM from the model, maintaining the same inference time as the original model.

2.3.1 Low-frequency feature generation module

The LFFGM, proposed by L. Liu et al. [46], is designed to enhance the extraction of low-frequency features in the brain stroke lesion segmentation task. As illustrated in Fig. 2, the LFFGM is comprised of a U-shaped structure with skip connections.

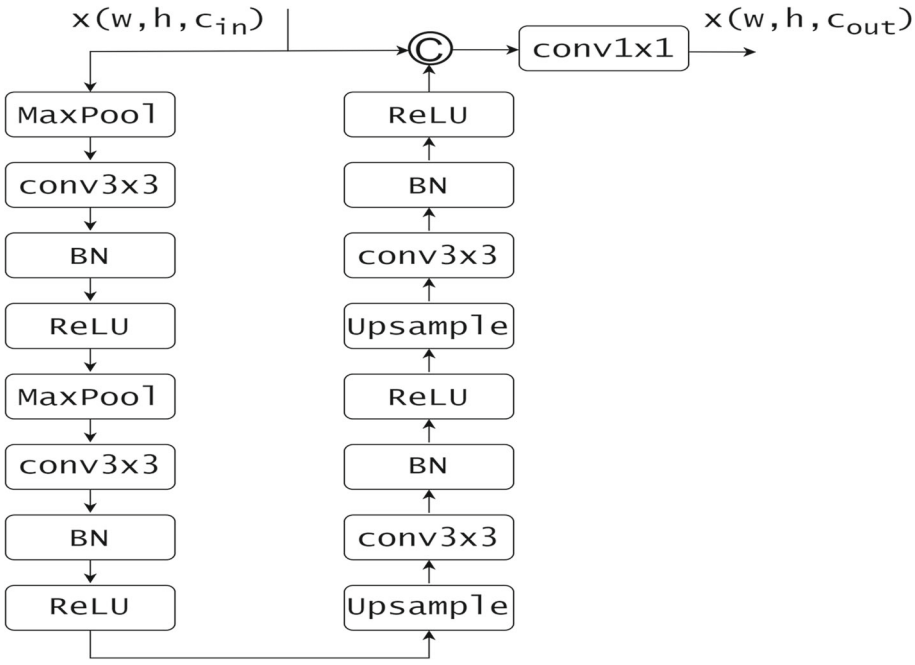

Fig. 2 Architecture of the LFFGM [46]

In our architecture, the input to the LFFGM is the low-frequency feature maps extracted from the encoder of the RED-CNN model. The initially is first processed by two blocks, each comprising a 2×2 max-pooling layer with a stride of 2, a 3×3 convolutional layer, a batch normalization layer, and a ReLU activation function. This sequence effectively reduces the size of the feature map by a factor of four. Subsequently, two blocks, each comprising a 2×2 up-sampling layer with a stride of two, a 3×3 convolutional layer, a batch normalization layer, and a ReLU activation function, are employed to restore the input size. Subsequently, the output of these blocks is then concatenated with the original LFFGM input using a concatenation layer. Finally, a 1×1 convolutional layer is applied to the concatenated output in order to reduce its size to match that of the feature maps of the predicted NDCT image. The final output, designated as F_{low}, represents the low-frequency features of the NDCT image.

2.3.2 Loss function

As detailed in Section 2.3, we designed the model architecture to learn two tasks: denoising and edge detection of LDCT images. To assess the influence of the proposed architecture on the selected RED-CNN model, we employed the loss function proposed in the original paper [2]. The loss function employed in the LDCT image denoising task is the MSE, denoted as $L_{denoising}$. For the LDCT edge detection task, the same loss function, denoted as L_{edge}, is employed to minimize the error between the predicted and ground-truth NDCT image edges. The total loss function of our method is the weighted sum of the two losses, $L_{denoising}$ and L_{edge}, with a weighting coefficient, α. Here, α represents the relative importance of the two tasks. L_{edge} is multiplied by α to control the influence of the auxiliary task on the main task. The proposed robust loss function, therefore, can be defined by the following equation:

$$\begin{aligned} L_{total} &= L_{denoising} + \alpha \cdot L_{edge} \\ &= \frac{1}{N}\sum_{i=1}^{N}(Y_i - \hat{Y}_i)^2 + \alpha \cdot (Z_i - \hat{Z}_i)^2 \end{aligned} \tag{2}$$

where $\hat{Y}_i$ and $\hat{Z}_i$ represent the NDCT ground truths of the image and edge, respectively. The total number of training samples is denoted by N. Y_i and Z_i correspond to the predicted NDCT image and edge, respectively. The weighting parameter, α, is adjusted through experimentation to minimize the value of the loss function.

3 Experimental setup

This section outlines the employed dataset, the techniques employed for image preprocessing, the training details, and the performance evaluation metrics.

3.1 Dataset

We used two datasets for experimental comparison: the AAPM dataset [47] and the Qin-LUNG-CT dataset [48].

The AAPM dataset, a publicly accessible LDCT dataset developed as part of the 2016 NIH-AAPM-Mayo Clinic Low Dose CT Grand Challenge. The dataset comprises full-dose

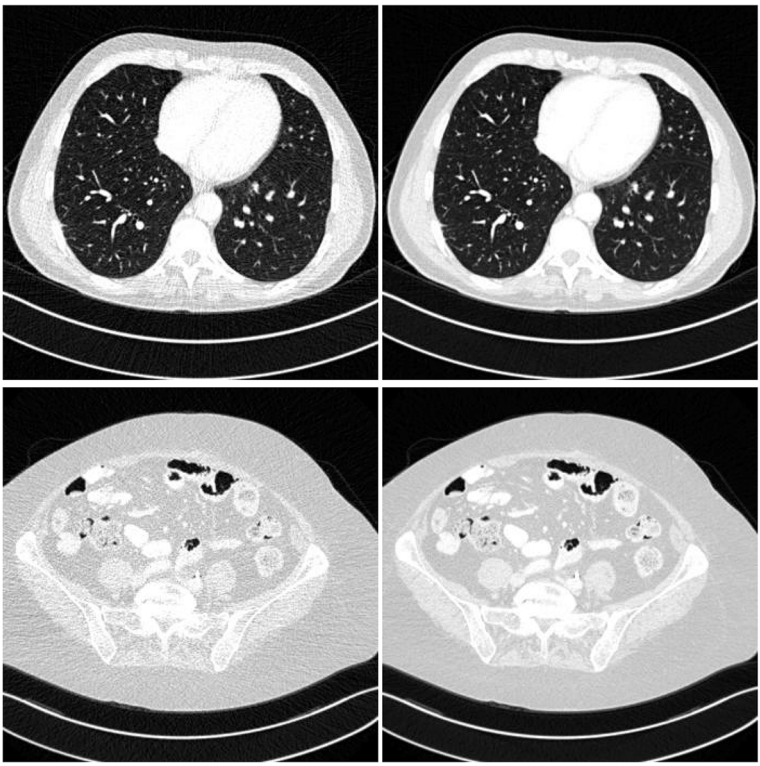

Fig. 3 Sample of images from the AAPM dataset [47]. On the left: LDCT images. On the right: the corresponding NDCT images

two-dimensional abdominal CT slices from 10 anonymized patients, along with their corresponding LDCT slices. The full-dose slices were captured at 120 kV and 200 effective mAs. The low-dose slices were generated by introducing Poisson noise to the projection data for each case, resulting in a noise level equivalent to 25% of the full dose. The CT images were reconstructed using both the medium smooth kernel (B30 kernel) and the medium sharp kernel (B45 kernel). Each reconstructed slice has a thickness of either 1 mm or 3 mm. In this study, we used the reconstructed scans (as illustrated in Fig. 3) of the B30 kernel with a thickness of 1mm. In accordance with the data division approach previously employed in related studies [4, 25], we selected a total of 2886 pairs of CT images from the scans of five patients in the dataset for model training. We designated the remaining 3050 pairs of CT images for model evaluation. The data division is detailed in Table 1.

The Qin-LUNG-CT dataset contains CT scans of 47 studies, which were obtained on patients diagnosed with non-small cell lung cancer with mixed stage and histology from the H. Lee Moffitt Cancer Center and Research Institute. The scans were obtained from patients who underwent surgical resection and had corresponding pre-surgery diagnostic CTs. The dataset comprises 3954 CT images of 512×512 pixels in the digital imaging and communications in medicine (DICOM) format. We used the CT images with subject IDs of R0273 and R0274 as the validation and test sets, respectively, while the remaining images constituted the training set.

Table 1 Division of the AAPM dataset [47]

Division	Case	# of images
Train	L067	560
	L096	823
	L109	318
	L143	585
	L192	600
	L286	525
	L291	856
Test	L310	533
	L333	310
	L506	526

3.2 Image preprocessing

The LDCT images in the AAPM dataset were generated by adding Poisson noise, which is confined to a window of Hounsfield units (HU) between -1231 and 1192, to the projection data. In accordance with the windowing technique described in [4], we employed a window of [-160, 240] HU to truncate the LDCT images. In order to speedup the model training process, we normalized the AAPM images to the range of [0, 1]. The dimensions of the AAPM images are 512×512 pixels. In order to increase the number of training images and thereby enhance the learning capability of the model, we applied data augmentation techniques, including rotation and flipping. Subsequently, we partitioned each image pair into smaller patches of size 64×64 pixels. This process facilitates the detection of perceptual differences in local regions, which are crucial for optimal denoising.

In order to generate the corresponding LDCT images of the NDCT images in the Qin-LUNG-CT dataset, we adhered to the image preprocessing methodology outlined in [49]. We first converted the images into the PNG format. Subsequently, we introduced artificial Gaussian noise with standard deviations of 15, 30, 45, and 60 to the converted images. This process was undertaken to simulate four distinct doses of LDCT images. Furthermore, we applied the same windowing and augmentation techniques we used in the AAPM dataset to the Qin-LUNG-CT dataset.

3.3 Training details

The AAPM dataset does not include corresponding segmentation masks or ground truths for the edges of the NDCT images. The process of annotating these images is a labor-intensive one that requires the expertise of radiologists. To address this challenge, we applied the Canny edge detector to the NDCT images in order to generate edges.

We trained the proposed model for 50 epochs with a mini-batch size of 16. We employed the Adam optimizer to update the model parameters, with the following values for the relevant parameters: beta_1, beta_2, and epsilon, which are set to 0.9, 0.999, and 1e-8, respectively. The initial learning rate is set to 1e-4, and it is decreased by a factor of 0.1 every two epochs. We used the TensorFlow framework for the implementation, and conducted the experiments on a computer equipped with the following specifications: an Intel® Core™ i7-7800X CPU @ 3.50 GHz processor, 64 GB of RAM, and an RTX 2080 TI GPU.

3.4 Performance evaluation metrics

We employed two commonly used metrics for image quality assessment to quantitatively evaluate the performance of models. The metrics employed include the Peak Signal-to-Noise Ratio (PSNR) [50] and the Structural Similarity Index Measure (SSIM) [51].

The PSNR is a metric that quantifies the distortion in a noise-free image that has been corrupted by noise. The PSNR is defined as follows:

$$PSNR = 20 \cdot \log_{10}\left(\frac{MAX_Y}{RMSE}\right) \tag{3}$$

where RMSE is defined by:

$$RMSE = \sqrt{\frac{1}{mn}\sum_{i=1}^{m}\sum_{j=1}^{n}(Y_{ij} - \hat{Y}_{ij})^2} \tag{4}$$

In these equations, $\hat{Y}$ and Y represent the noise-free image of size $m \times$n and its predicted output, respectively. MAX_Y represents the maximum possible pixel value of the image. A higher PSNR value (in decibels) indicates a greater degree of image denoising quality.

The SSIM is a metric that correlates well with perceived visual quality. The metric measures the degree of similarity between the noise-free image and the corresponding denoised image. The SSIM is defined by the following equation:

$$SSIM(\hat{Y}, Y) = \frac{(2\mu_{\hat{Y}}\mu_Y + C_1)(2\sigma_{\hat{Y}Y} + C_2)}{(\mu_{\hat{Y}}^2 + \mu_Y^2 + C_1)(\sigma_{\hat{Y}}^2 + \sigma_Y^2 + C_2)} \tag{5}$$

where $\mu_{\hat{Y}}$, μ_Y, $\sigma_{\hat{Y}}$, σ_Y, and $\sigma_{\hat{Y}Y}$ represent the means and variances of $\hat{Y}$ and Y, respectively, as well as the covariance between $\hat{Y}$ and Y. The symbols C_1 and C_2 are constants. The SSIM values lie between -1 and 1, with 1 indicating perfect similarity between $\hat{Y}$ and Y images.

Table 2 Loss functions of LDCT denoising methods

Model	Loss
CNN [24]	MSE
RED-CNN [2]	MSE
Q-AE [25]	MSE
DFSNE-NET [4]	MSE, BCE
EDCNN [37]	MSE, Multi-scales perceptual
WGAN-VGG [3]	Adversarial, Wasserstein, Perceptual
CPCE-2D [27]	Adversarial, Perceptual
DUGAN [53]	Least-squares adversarial, Gradient
Uformer [54]	Charbonnier
CTformer [55]	MSE
MT-RED-CNN (**ours**)	MSE

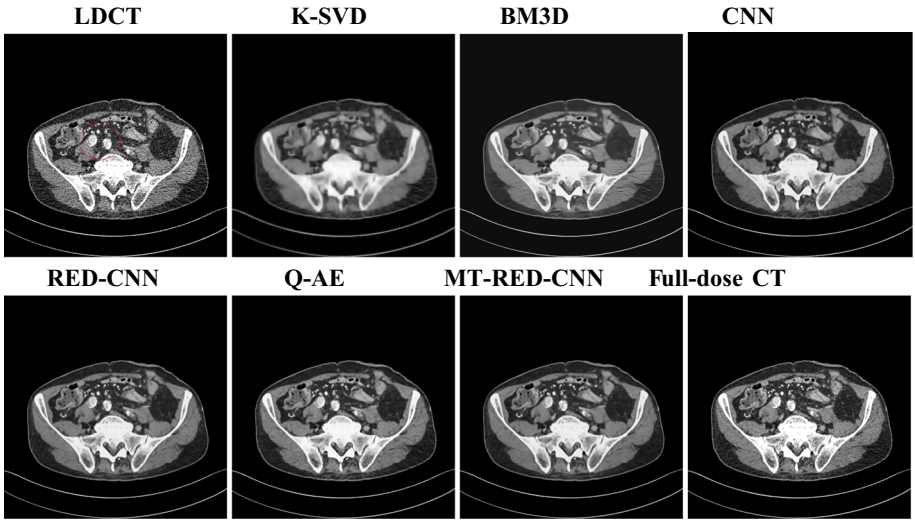

Fig. 4 L506 270^{th} image before and after denoising using different methods

4 Results and discussion

We evaluate our method both subjectively and objectively, comparing it to several existing methods: K-SVD [20], BM3D [19], CNN [24], RED-CNN [2], Q-AE [25], DFSNE-Net [4], DnCNN [52], EDCNN [37], WGAN-VGG [3], CPCE-2D [27], DUGAN [53], Uformer [54], and CTformer [55]. The loss functions for these methods are summarized in Table 2.

4.1 Qualitative comparison

We assess the denoising performance of the models subjectively through a qualitative comparison. In accordance with the methodology outlined in [25], we selected two representative

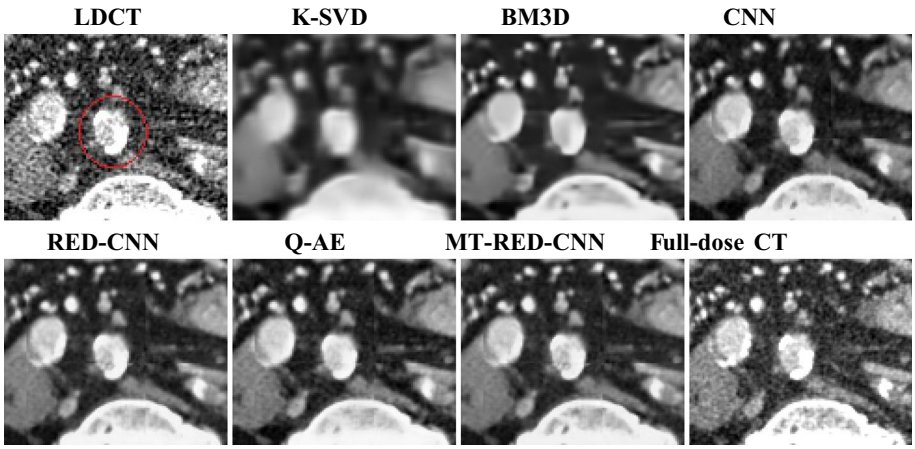

Fig. 5 Zoomed ROI of the L506 270^{th} image

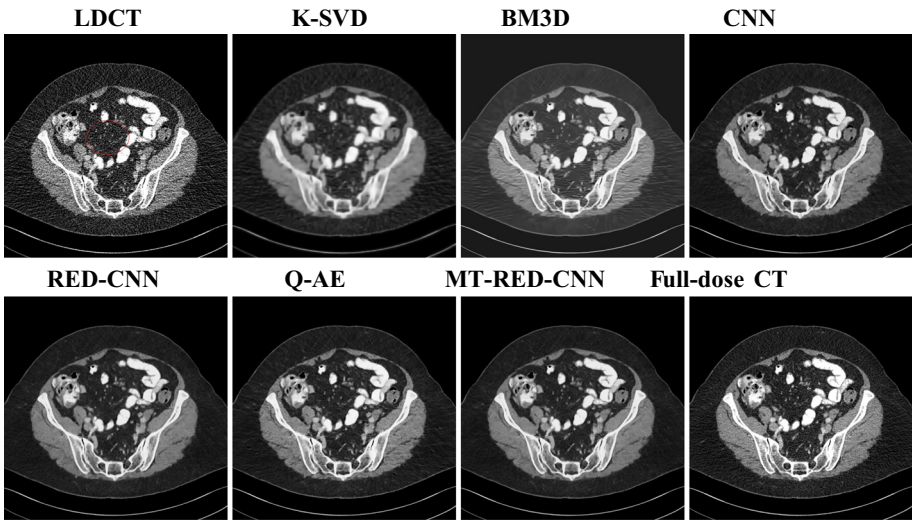

Fig. 6 L310 340^{th} image before and after denoising using different methods

LDCT images for the illustration of the model's denoising performance. The images, the 270th and 340th images from L506 and L310, respectively, feature low-attenuation lesions and blood vessels.

The denoising results are presented in Figs. 4 and 6. To facilitate a more detailed visualization and comparison of the denoising results, the regions of interest (ROIs), highlighted by red circles, were magnified and are presented in Figs. 5 and 7. The display window for these figures is [-160, 240].

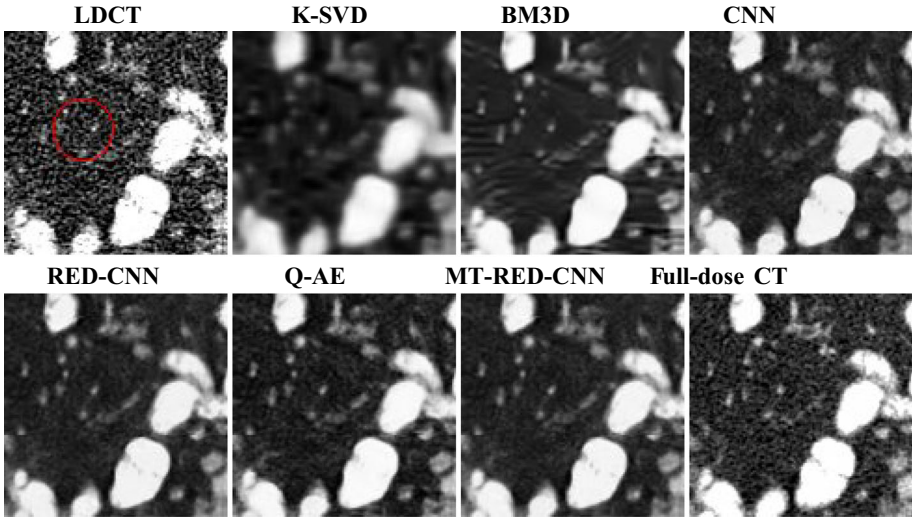

Fig. 7 Zoomed ROI of the L310 340^{th} image

Table 3 Quantitative comparison on the AAPM dataset. Best results are highlighted in bold

Model	PSNR					SSIM				
	L286	L291	L310	L333	L506	L286	L291	L310	L333	L506
LDCT	24.43	22.17	21.63	22.56	24.47	0.81	0.80	0.72	0.82	0.87
K-SVD [20]	24.46	23.14	24.04	24.02	24.16	0.81	0.78	0.76	0.81	0.83
BM3D [19]	27.97	25.98	26.30	26.69	27.97	0.87	0.85	0.82	0.87	0.90
CNN [24]	28.89	26.69	26.75	27.64	28.69	0.89	0.86	0.83	**0.89**	0.91
RED-CNN [2]	29.29	27.00	26.97	27.95	29.00	**0.90**	**0.87**	**0.84**	**0.89**	**0.92**
Q-AE [25]	28.85	26.64	26.66	27.68	28.80	**0.89**	0.86	0.83	0.89	**0.92**
DFSNE-Net [4]	29.25	**27.11**	**27.11**	27.72	**29.11**	0.89	**0.87**	0.83	0.88	**0.92**
MT-RED-CNN (ours)	**29.32**	27.02	27.00	**27.97**	29.03	**0.90**	**0.87**	**0.84**	**0.89**	**0.92**

In general, all methods denoise LDCT images to varying degrees. From the zoomed ROIs in Figures 5 and 7, we observe the following: (1) Deep learning methods outperform classical methods (K-SVD and BM3D) in denoising LDCT images. (2) LDCT images denoised by classical methods appear blurred and show stripe artifacts. (3) LDCT images denoised with deep learning methods are smoothed by using MSE loss. CNN-denoised LDCT images show some blurring, while RED-CNN-denoised LDCT images are less smooth than those processed by CNN due to the increased number of convolution filters. Q-AE-denoised LDCT images are similar to those of RED-CNN. MT-RED-CNN denoised LDCT images are visually similar to those of RED-CNN, but with reduced image noise and better fidelity for small structures.

Since radiologists rely on ROIs in scans to diagnose diseases, it is critical to preserve the edges of these ROIs after denoising to improve diagnostic accuracy. Furthermore, the quality of denoised LDCT images using our proposed MT-RED-CNN model can be further improved by learning the edges of ROIs rather than the entire NDCT edges.

4.2 Quantitative comparison

To objectively analyze the denoising performance of the methods, we quantitatively compare their performance using the PSNR and SSIM metrics. The computed metrics are shown in Tables 3 and 4.

Table 3 presents the results of methods evaluated on the AAPM dataset. In accordance with the qualitative analysis, deep learning-based denoising methods demonstrate superior performance compared to classical methods, with an improvement of at least 0.72% in PSNR and 0.01% in SSIM. Among the deep learning methods, RED-CNN demonstrates superior performance to CNN and Q-AE, with further improvements observed in the proposed MT-RED-CNN model in terms of PSNR. The proposed model demonstrates superior performance compared to existing methods, achieving results that are comparable to those of DFSNE-Net.

Table 4 presents the results of existing methods evaluated on the Qin-LUNG-CT dataset across four different noise levels. In general, our MT-RED-CNN model consistently demonstrates superior performance compared to existing methods methods. For example, our MT-RED-CNN outperforms RED-CNN by 2.29% in PSNR and 0.03% in SSIM at σ=15, and by 2.21% in PSNR and 0.07% at σ=45. Nevertheless, RED-CNN achieves superior PSNR at σ=60.

Table 4 Quantitative comparison on the Qin-LUNG-CT dataset across different noise levels

Model	$\sigma = 15$		$\sigma = 30$		$\sigma = 45$		$\sigma = 60$	
	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM
LDCT	33.76	0.93	27.76	0.84	24.28	0.79	21.86	0.76
DnCNN [52]	33.68	0.93	32.05	0.90	31.80	0.88	30.35	0.86
EDCNN [37]	36.38	0.95	33.43	0.91	31.23	0.88	30.28	0.86
RED-CNN [2]	36.38	0.95	34.74	0.93	31.74	0.89	**32.00**	0.90
WGAN-VGG [3]	33.73	0.93	32.44	0.90	30.89	0.88	29.96	0.87
CPCE-2D [27]	33.60	0.93	28.34	0.85	29.58	0.86	27.50	0.83
DUGAN [53]	38.10	0.96	34.65	0.93	33.13	0.91	31.80	0.89
Uformer [54]	35.96	0.96	32.90	0.92	31.67	0.90	30.54	0.88
CTformer [55]	36.02	0.95	32.12	0.91	30.26	0.88	29.05	0.86
MT-RED-CNN (**ours**)	**38.67**	**0.98**	**35.57**	**0.97**	**33.95**	**0.96**	30.25	**0.92**

This analysis provides empirical support for the hypothesis that leveraging the correlation between LDCT image denoising and LDCT edge detection enhances denoising performance.

4.3 Effect of hyperparameter α

The hyperparameter α controls the contribution of the edge loss to the total loss, as defined in (2). A larger α value indicates that the model will prioritize the learning of the task of NDCT edge detection over the task of NDCT image denoising. Conversely, a smaller α value directs the model's attention towards the learning of NDCT image denoising. To determine the optimal α value that balances the model's learning on NDCT image denoising and edge detection tasks, we evaluate the denoising performance of the proposed MT-RED-CNN model on the AAPM dataset using four distinct α values: 10^{-1}, 10^{-2}, 10^{-3}, and 10^{-4}. Table 5 indicates that the optimal α value is 10^{-3}.

4.4 Model efficiency

We assess the efficiency of the proposed model in comparison to existing methods in the inference phase, based on the number of parameters, floating point operations (FLOPs), and the inference time, as illustrated in Table 6. FLOPs represent the computational complexity of the model, quantifying the number of floating-point operations executed during the inference phase. Conversely, the inference time represents the duration of the forward pass in the model.

Table 6 indicates that CNN has the fewest parameters due to its use of only three convolutional layers. Q-AE has fewer parameters than RED-CNN and MT-RED-CNN. RED-CNN shares the same number of parameters as our MT-RED-CNN. With regard to the number of

Table 5 Effect of the hyperparameter α on denoising performance with the AAPM Dataset

Model	$\alpha = 10^{-1}$		$\alpha = 10^{-2}$		$\alpha = 10^{-3}$		$\alpha = 10^{-4}$	
	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM
MT-RED-CNN	27.70	**0.88**	27.64	0.87	**28.07**	**0.88**	27.64	0.87

Table 6 Efficiency comparison of existing methods

Model	# of parameters	FLOPS (B)	Inference time (s)
CNN [24]	24513	10	0.319
RED-CNN [2]	1848865	505	0.562
Q-AE [25]	49818	72.87	0.58
MT-RED-CNN (**ours**)	1848865	505	0.562

FLOPs, CNN exhibits the lowest number of operations, followed by Q-AE. RED-CNN and MT-RED-CNN exhibit comparable FLOPs. We measured the inference time for processing the same 512×512 LDCT image. CNN produces the predicted NDCT image more rapidly than RED-CNN, Q-AE, and MT-RED-CNN. The inference times for RED-CNN and MT-RED-CNN are identical. For traditional methods, K-SVD and BM3D have inference times of 60 seconds and 4.139 seconds, respectively.

5 Conclusion

We hypothesized that the correlation between the tasks of LDCT image denoising and LDCT edge detection could be exploited to enhance the LDCT denoising results. Consequently, we proposed a novel architecture for LDCT image denoising. Our architecture incorporates a high-frequency learning module within the model, enabling concurrent training on both the primary and secondary tasks of LDCT image denoising and LDCT edge detection, respectively. Our experiments yielded evidence supporting the hypothesis, indicating an improvement in the performance of the RED-CNN model for LDCT denoising, while maintaining the same number of parameters and inference time. In the future, we intend to extend the applicability of our method to address the challenges faced by the medical image denoising community.

Funding Open access funding provided by Université Paris-Est Créteil.

Declarations

Conflict of interest The authors declare no conflicts of interest.

References

1. Brenner DJ, Hall EJ (2007) Computed tomography—an increasing source of radiation exposure. N Engl J Med 357(22):2277–2284

2. Chen H, Zhang Y, Kalra MK, Lin F, Chen Y, Liao P, Zhou J, Wang G (2017) Low-dose ct with a residual encoder-decoder convolutional neural network. IEEE Trans Med Imag 36(12):2524–2535
3. Yang Q, Yan P, Zhang Y, Yu H, Shi Y, Mou X, Kalra MK, 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 Imag 37(6):1348–1357
4. Liu J, Jiang H, Ning F, Li M, Pang W (2022) Dfsne-net: Deviant feature sensitive noise estimate network for low-dose ct denoising. Comput Biol Med 149:106061
5. Li M, Hsu W, Xie X, Cong J, Gao W (2020) Sacnn: self-attention convolutional neural network for low-dose ct denoising with self-supervised perceptual loss network. IEEE Trans Med Imag 39(7):2289–2301
6. Trung NT, Trinh D-H, Trung NL, Luong M (2022) Low-dose ct image denoising using deep convolutional neural networks with extended receptive fields. SIViP 16(7):1963–1971
7. Balda M, Hornegger J, Heismann B (2012) Ray contribution masks for structure adaptive sinogram filtering. IEEE Trans Med Imag 31(6):1228–1239
8. Manduca A, Yu L, Trzasko JD, Khaylova N, Kofler JM, McCollough CM, Fletcher JG (2009) Projection space denoising with bilateral filtering and ct noise modeling for dose reduction in ct. Med Phys 36(11):4911–4919
9. Wang J, Li T, Lu H, Liang Z (2006) Penalized weighted least-squares approach to sinogram noise reduction and image reconstruction for low-dose x-ray computed tomography. IEEE Trans Med Imag 25(10):1272–1283
10. Patwari M, Gutjahr R, Raupach R, Maier A (2022) Limited parameter denoising for low-dose x-ray computed tomography using deep reinforcement learning. Med Phys 49(7):4540–4553
11. Sidky EY, Pan X (2008) Image reconstruction in circular cone-beam computed tomography by constrained, total-variation minimization. Phys Med Biology 53(17):4777
12. Tian Z, Jia X, Yuan K, Pan T, Jiang SB (2011) Low-dose ct reconstruction via edge-preserving total variation regularization. Phys Med Biology 56(18):5949
13. Liu Y, Ma J, Fan Y, Liang Z (2012) Adaptive-weighted total variation minimization for sparse data toward low-dose x-ray computed tomography image reconstruction. Phys Med Biology 57(23):7923
14. Cai J-F, Jia X, Gao H, Jiang SB, Shen Z, Zhao H (2014) Cine cone beam ct reconstruction using low-rank matrix factorization: algorithm and a proof-of-principle study. IEEE Trans Med Imag 33(8):1581–1591
15. Xu Q, Yu H, Mou X, Zhang L, Hsieh J, Wang G (2012) Low-dose x-ray ct reconstruction via dictionary learning. IEEE Trans Med Imag 31(9):1682–1697
16. Cong W, Wang G, Yang Q, Li J, Hsieh J, Lai R (2019) Ct image reconstruction on a low dimensional manifold. Inverse Prob imag (Springfield, Mo.) 13(3):449
17. Zhang Y, Zhang W, Lei Y, Zhou J (2014) Few-view image reconstruction with fractional-order total variation. JOSA A 31(5):981–995
18. Feruglio PF, Vinegoni C, Gros J, Sbarbati A, Weissleder R (2010) Block matching 3d random noise filtering for absorption optical projection tomography. Phys Med Biology 55(18):5401
19. Kang D, Slomka P, Nakazato R, Woo J, Berman DS, Kuo C-CJ, Dey D (2013) Image denoising of low-radiation dose coronary ct angiography by an adaptive block-matching 3d algorithm. In: Medical imaging 2013: image processing, SPIE, vol 8669, pp 671–676
20. Ma J, Huang J, Feng Q, Zhang H, Lu H, Liang Z, Chen W (2011) Low-dose computed tomography image restoration using previous normal-dose scan. Med Phys 38(10):5713–5731
21. Aharon M, Elad M, Bruckstein A (2006) K-svd: an algorithm for designing overcomplete dictionaries for sparse representation. IEEE Trans Signal Process 54(11):4311–4322
22. Chen Y, Yin X, Shi L, Shu H, Luo L, Coatrieux J-L, Toumoulin C (2013) Improving abdomen tumor low-dose ct images using a fast dictionary learning based processing. Phys Med Biology 58(16):5803
23. Mendrik AM, Vonken E-J, Rutten A, Viergever MA, Ginneken B (2009) Noise reduction in computed tomography scans using 3-d anisotropic hybrid diffusion with continuous switch. IEEE Trans Med Imag 28(10):1585–1594
24. Chen H, Zhang Y, Zhang W, Liao P, Li K, Zhou J, Wang G (2017) Low-dose ct via convolutional neural network. Biomed Opt Express 8(2):679–694
25. Fan F, Shan H, Kalra MK, Singh R, Qian G, Getzin M, Teng Y, Hahn J, Wang G (2019) Quadratic autoencoder (q-ae) for low-dose ct denoising. IEEE Trans Med Imag 39(6):2035–2050
26. Ma L, Xue H, Yang G, Zhang Z, Li C, Yao Y, Teng Y (2023) Scrdn: residual dense network with self-calibrated convolutions for low dose ct image denoising. Nucl Instrum Methods Phys Res, Sect A 1045:167625
27. Shan H, Zhang Y, Yang Q, Kruger U, Kalra MK, Sun L, Cong W, Wang G (2018) 3-d convolutional encoder-decoder network for low-dose ct via transfer learning from a 2-d trained network. IEEE Trans Med Imag 37(6):1522–1534

28. Tang Y, Du Q, Wang J, Wu Z, Li Y, Li M, Yang X, Zheng J (2022) Ccn-cl: a content-noise complementary network with contrastive learning for low-dose computed tomography denoising. Comput Biol Med 147:105759
29. Ferdi A, Benierbah S, Ferdi Y (2022) U-net-based covid-19 ct image semantic segmentation: a transfer learning approach. In: 2022 7th International Conference on Image and Signal Processing and Their Applications (ISPA), IEEE, pp 1–5
30. Aslan MF (2022) A robust semantic lung segmentation study for cnn-based covid-19 diagnosis. Chemom Intell Lab Syst 231:104695
31. Albishri A, Shah SJH, Lee Y (2022) Tlu-net: transfer learning framework using u-net convolutional neural networks for ct-based lungs and covid-19 segmentation. In: 2022 IEEE International Conference on Bioinformatics and Biomedicine (BIBM), IEEE, pp 2238–2246
32. Bhosale YH, Patnaik KS (2023) Bio-medical imaging (x-ray, ct, ultrasound, ecg), genome sequences applications of deep neural network and machine learning in diagnosis, detection, classification, and segmentation of covid-19: a meta-analysis & systematic review. Multimed Tool Appl 82(25):39157–39210
33. Narin A, Kaya C, Pamuk Z (2021) Automatic detection of coronavirus disease (covid-19) using x-ray images and deep convolutional neural networks. Pattern Anal Appl 24:1207–1220
34. Astaraki M, Zakko Y, Dasu IT, Smedby Ö, Wang C (2021) Benign-malignant pulmonary nodule classification in low-dose ct with convolutional features. Physica Med 83:146–153
35. Xia K, Zhou Q, Jiang Y, Chen B, Gu X (2022) Deep residual neural network based image enhancement algorithm for low dose ct images. Multimed Tool Appl 81(25):36007–36030
36. Wu D, Kim K, Fakhri GE, Li Q (2017) A cascaded convolutional neural network for x-ray low-dose ct image denoising. arXiv:1705.04267
37. Liang T, Jin Y, Li Y, Wang T (2020) Edcnn: edge enhancement-based densely connected network with compound loss for low-dose ct denoising. In: 2020 15th IEEE International Conference on Signal Processing (ICSP), IEEE, vol 1, pp 193–198
38. Ledig C, Theis L, Huszár F, Caballero J, Cunningham A, Acosta A, Aitken A, Tejani A, Totz J, Wang Z, et al (2017) Photo-realistic single image super-resolution using a generative adversarial network. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 4681–4690
39. Johnson J, Alahi A, Fei-Fei L (2016) Perceptual losses for real-time style transfer and super-resolution. In: Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11-14, 2016, Proceedings, Part II 14, Springer, pp 694–711
40. Wang G, Hu X (2021) Low-dose ct denoising using a progressive wasserstein generative adversarial network. Comput Biol Med 135:104625
41. Zhang Y, Yang Q (2021) A survey on multi-task learning. IEEE Trans Knowl Data Eng 34(12):5586–5609
42. Chen F, Wu F, Gao G, Ji Y, Xu J, Jiang G-P, Jing X-Y (2022) Jspnet: Learning joint semantic & instance segmentation of point clouds via feature self-similarity and cross-task probability. Pattern Recogn 122:108250
43. Han C, Yao H, Zhao B, Li Z, Shi Z, Wu L, Chen X, Qu J, Zhao K, Lan R et al (2022) Meta multi-task nuclei segmentation with fewer training samples. Med Image Anal 80:102481
44. Wang K, Zhan B, Zu C, Wu X, Zhou J, Zhou L, Wang Y (2022) Semi-supervised medical image segmentation via a tripled-uncertainty guided mean teacher model with contrastive learning. Med Image Anal 79:102447
45. Wei J, Yang P, Pi Y, Cai H, Jiang L, Xiang Y, Zhao Z, Yi Z (2022) Cross-granularity multi-task network for ischemia diagnosis and defect detection in the myocardial perfusion imaging. Knowl-Based Syst 251:108877
46. Liu L, Huang C, Cai C, Zhang X, Hu Q (2022) Multi-task learning improves the brain stoke lesion segmentation. In: ICASSP 2022-2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), IEEE, pp 2385–2389
47. Low Dose CT Grand Challenge. https://www.aapm.org/GrandChallenge/LowDoseCT/
48. Goldgof D, Hall L, Hawkins S, Schabath M, Stringfield O, Garcia A, Balagurunathan Y, Kim J, Eschrich S, Berglund A, Gatenby R, Gillies R (2015) Data From QIN LUNG CT (version 2) [Data set]. The Cancer Imaging Archive
49. Zhang J, Gong W, Ye L, Wang F, Shangguan Z, Cheng Y (2024) A review of deep learning methods for denoising of medical low-dose ct images. Comput Biol Med 171:108112
50. Huynh-Thu Q, Ghanbari M (2008) Scope of validity of psnr in image/video quality assessment. Electron Lett 44(13):800–801
51. Wang Z, Bovik AC, Sheikh HR, Simoncelli EP (2004) Image quality assessment: from error visibility to structural similarity. IEEE Trans Image Process 13(4):600–612
52. Zhang K, Zuo W, Chen Y, Meng D, Zhang L (2017) Beyond a gaussian denoiser: residual learning of deep cnn for image denoising. IEEE Trans Image Process 26(7):3142–3155

53. Huang Z, Zhang J, Zhang Y, Shan H (2021) Du-gan: generative adversarial networks with dual-domain u-net-based discriminators for low-dose ct denoising. IEEE Trans Instrum Meas 71:1–12
54. Wang Z, Cun X, Bao J, Zhou W, Liu J, Li H (2022) Uformer: a general u-shaped transformer for image restoration. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, pp 17683–17693
55. Wang D, Fan F, Wu Z, Liu R, Wang F, Yu H (2023) Ctformer: convolution-free token2token dilated vision transformer for low-dose ct denoising. Phys Med Biology 68(6):065012