CYBERNETICS AND PHYSICS, VOL. 13, NO. 2, 2024, 130–141

# COMPUTED TOMOGRAPHY IMAGE RECONSTRUCTION: INTEGRATING ITERATIVE METHODS WITH MAXIMUM LIKELIHOOD EXPECTATION MAXIMIZATION AND DEEP LEARNING MODELS

**Pham Cong Thang***
The University of Danang,
University of Science and Technology
Danang, Viet Nam
pcthang@dut.udn.vn

**Tran Thi Thu Thao**
The University of Danang,
University of Economics
Danang, Viet Nam
thaotran@due.udn.vn

**Huynh Duc Anh Bao, Nguyen Quoc Cuong, Nguyen Tien Hung**
The University of Danang,
University of Science and Technology
Danang, Viet Nam
102200245@sv1.dut.udn.vn; 102200249@sv1.dut.udn.vn; 102200258@sv1.dut.udn.vn

Article history:
Received 25.03.2024, Accepted 16.09.2024

**Abstract**

Computed Tomography (CT) imaging faces limitations, including low spatial resolution and noise, particularly in low-radiation-dose imaging. To address these challenges, researchers are exploring CT image reconstruction from sinogram data. Sinograms represent X-ray absorption throughout the body, and sophisticated image reconstruction methods, including machine learning algorithms and generative adversarial networks (GANs), can improve precision and resolution without increasing patient radiation exposure.

This study proposes an iterative reconstruction approach that combines filters from deep learning models (Convolutional Neural Networks and UNet) with the Maximum Likelihood Expectation Maximization (ML-EM) algorithm and the Enhanced Super-Resolution Generative Adversarial Networks (ESRGAN) model. Our method aims to enhance image quality and reconstruction speed. Experimental results show significant improvements in image quality and resolution, with the proposed method (DL-MLEM-IR-UNET-ESRGAN) achieving an average SSIM of 0.9980 and PSNR of 53.2119, outperforming other methods. Additionally, our method reduces reconstruction time, with an average runtime of 131 seconds.

**Key words**

CT Image Reconstruction, Iterative Reconstruction, ML-EM, U-Net, ESRGAN.

## 1 Introduction

Computed tomography (CT) has emerged as a cornerstone in clinical diagnostics, revolutionizing the field with its unparalleled ability to provide detailed internal images. Its impact on treatment strategies and decision-making across various medical disciplines has been profound. However, despite its widespread use and diagnostic benefits, concerns persist regarding the inherent ionizing radiation exposure associated with CT scans. Consequently, ongoing research and development efforts are dedicated to minimizing patient risk while preserving the integrity and diagnostic utility of the images produced.

The initial approach to CT image reconstruction, primarily utilizing filtered back projection, gained popularity due to its computational efficiency and ability to provide rapid results, which are crucial for real-time imaging during patient scans. However, despite these advantages, its limitations became increasingly evident in low-dose CT settings or when imaging patients with larger body habitus. In such scenarios, the method often led to heightened image noise and artifacts, thereby compromising overall image quality and potentially impacting diagnostic accuracy [Kalra, 2015; Granichin, 2018; Erofeeva, 2019; McLeavy, 2021; Tran, 2023].

*Corresponding author

Iterative reconstruction (IR) algorithms have emerged as a response to the limitations of traditional CT reconstruction methods. By employing a more sophisticated approach to image processing, IR algorithms have demonstrated a notable improvement in image quality, particularly in low-dose CT scans. These algorithms effectively reduce image noise and enhance image clarity [Pham, 2015]. Notably, Advanced Iterative Reconstruction (IR) methods have gained FDA (Food and Drug Administration) approval for significantly reducing X-ray dosage (ranging from 23 to 76%) without compromising image quality.

The evolution of machine learning and deep learning models has further propelled the development of model-based IR methods, showing potential for even greater reductions in X-ray dosage compared to hybrid IR and FBP (Filtered Back Projection) methods [Willemink, 2019; Shin, 2021]. However, these methods have faced criticism due to the altered texture and visual appearance of the processed images, which some users describe as having an unnatural, almost artificial look, raising concerns about their acceptability in clinical practice.

The integration of artificial intelligence (AI), particularly deep learning (DL), into CT imaging represents a significant advancement in the field of radiology. DL-based technologies, characterized by their ability to manage and analyze complex data models, have shown immense potential in enhancing CT image reconstruction. These methods can produce high-quality images at reduced radiation doses, overcoming the limitations of both traditional and IR methods. Moreover, they offer faster processing speeds, which are crucial in clinical settings where time is often of the essence [Akagi, 2019; Fallahpoor, 2024]. The ultimate goal of advancements in CT imaging, particularly those involving AI and DL, is to achieve a delicate balance between reducing radiation dose and maintaining, if not enhancing, the diagnostic performance of CT scans. As the medical community strides towards this goal, the promise of AI in image reconstruction emerges as a critical factor in achieving high-quality imaging at significantly lower radiation doses. This development signals the dawn of a new era in CT imaging, potentially leading to the realization of ultra-low-dose CT scans without compromising diagnostic accuracy.

Acknowledging the pivotal role of deep learning networks and algorithms in advancing image reconstruction techniques [Reader, 2023; Reader, 2024], we have undertaken comprehensive studies and experiments to explore the application of algorithms and deep learning models in the realm of image reconstruction. The primary contributions of our study encompass two main aspects: developing a CT image reconstruction method based on Sinogram images using IR with the Maximum Likelihood Expectation Maximization (MLEM) algorithm combined with the deep learning model UNet and the Enhanced Super-Resolution Generative Adversarial Networks (ESRGAN) model; and performing comparative analyses with other reconstruction methods.

The proposed method, which combines the MLEM algorithm with the UNet and ESRGAN deep learning models, has the potential to significantly improve CT image reconstruction by producing high-quality images at reduced radiation doses. This could lead to a reduction in radiation exposure for patients, improved diagnostic accuracy, and faster processing speeds, ultimately transforming the field of CT imaging and improving patient care.

This paper is organized as follows: In Section (2), we discuss related studies and terminology. The proposed method for solving the optimization problem is presented in Section (3). Section (4) provides numerical experiments to illustrate the outstanding performance of our approach. Finally, the conclusions of this work are drawn in Section (5).

## 2 Preliminaries

The projection process involves transforming CT images into their corresponding sinograms, as depicted in Figure (1). The sinogram is generated through the rotation of the CT scanner gantry, capturing multiple attenuation profiles at various angles (along the x-axis). Each profile represents a direct measurement of the X-ray absorption levels of tissues within the body, collected by an array of detectors (along the y-axis). As the gantry rotates, a point inside the patient is projected onto the detectors as a sinogram curve through the Radon transform. The collection of these curves forms the sinogram, which serves as the foundational dataset for reconstructing the anatomical structures within the body [Schofield, 2020].

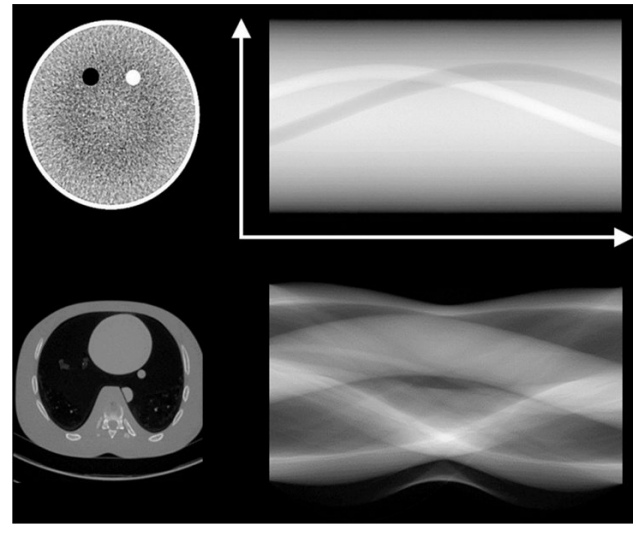

Figure 1. Examples CT slices and their simulated sinograms over a $180^{\circ}$ acquisition [Schofield, 2020]

The evolution of image reconstruction techniques, par-

ticularly in computed tomography (CT), has progressed through various phases of innovation, transitioning from the initial Inverse Radon Transform to the more advanced Iterative Reconstruction (IR) methods, paralleling the increasing computational power of computers [Stiller, 2018; Ustaoglu, 2023].

The fundamental process of back projection in CT imaging begins with the recording of X-ray absorption levels as they pass through the body from various angles. Each absorption profile, also known as a projection, is 'projected' back into the image space at the same angle it was acquired. The quality of the reconstructed image improves as the number of projections increases. When a sufficient number of projections from various angles are available, the absorbed data overlaps and begins to create a sharper image, resembling the final image in the back projection sequence. However, the simple back projection process has limitations, notably its inability to distinguish different X-ray absorption points along the same projection line. This limitation results in blurriness, leading to an image reconstruction that lacks sharpness and accuracy.

To address this issue, more advanced image reconstruction algorithms, such as filtered back projection and iterative reconstruction, have been developed, significantly improving image quality. Iterative Reconstruction (IR) methods are particularly noteworthy. IR methods combine a data term, which models the observed projections, with a regularization term that accounts for system nonuniformities and noise. Statistical IR introduces a weighting term to prioritize reliable data during reconstruction. Through iterative cycles of comparison and correction, the image is progressively refined, enhancing details and reducing artifacts until a satisfactory reconstruction is achieved. This technique capitalizes on powerful computing to handle the complexity of modern CT system geometry and data characteristics, which was not feasible in the early days of CT technology [Geyer, 2015; Stiller, 2018].

Filtered Back Projection (FBP) is a fundamental algorithm used to reconstruct 2D images from a series of CT scan projection profiles. This technique improves upon basic back projection by applying a spatial frequency filter, typically a ramp filter, to the attenuation profiles before back projecting them into the image space. This filtering step enhances high-frequency components associated with edges in the image, reducing blurriness and emphasizing anatomical boundaries. As a result, clearer images with improved structural details are produced [Schofield, 2020]. However, it's important to note that the FBP technique also amplifies noise present in the raw data, which can make the reconstructed images appear grainy or noisy. While FBP excels in reconstructing images quickly, which is advantageous for scanning large volumes, the trade-off is the increased noise that compromises spatial resolution improvements. More recently, Model-Based Iterative Reconstruction (MBIR) represents the cutting edge in this evolution [Willemink, 2019; Meyer, 2024].

Alongside these developments, modern deep learning technologies have also made significant strides. Models such as Convolutional Neural Networks (CNNs) [Roslin, 2023], UNets [Mizusawa, 2021], and Generative Adversarial Networks (GANs) [Fu, 2023; Liu, 2023] have been developed and applied in various problem-solving contexts. CNN models, in particular, have been employed as filters in traditional back projection methods, leading to the development of DL-enhanced versions like Deep Learned Filtered Backprojection (DL-FBP), Deep Learned Filtered Backprojection then Filtering (DL-FBP-F), and Deep Learned Backprojection then Filtering (DL-BPF) [Hashimoto, 2022; Reader, 2024; Tan, 2024]. An advanced framework for CT image reconstruction leverages deep learning (DL) techniques, integrating them into the raw data acquisition process to produce high-quality diagnostic images using artificial intelligence. The framework incorporates two neural networks in conjunction with traditional image reconstruction methods such as Back Projection (BP) and Forward Projection (FP), enhancing the refinement of image quality [Reader, 2024].

In DL-FBP, only the first DNN is employed to filter the raw data, and the traditional back projection is used afterward. The second DNN essentially performs an identity operation, meaning it does not alter the back-projected image further. This approach balances the use of deep learning with the simplicity of traditional reconstruction methods, offering a middle ground between complexity and image quality.

Meanwhile, the DL-FBP-F approach uses two deep neural networks (DNNs). The first DNN processes the raw measured data, enhancing it before the back projection. The back-projected image is then further refined by the second DNN, using both the original data and the forward-projected image to minimize reconstruction errors. This dual-network approach aims to filter out noise and improve image clarity more effectively than traditional methods. Finally, DL-BPF skips the initial deep learning filtering stage (first DNN) entirely, directly using back projection on the raw data. Then, the second DNN takes over to refine the back-projected image.

Another approach is the use of Maximum Likelihood Expectation Maximization (ML-EM). ML-EM is a method in statistics and machine learning often utilized in medical imaging applications such as tomography. The method comprises two steps: Maximum Likelihood (ML) and Expectation Maximization (EM) [Yang, 2022]. ML-EM is commonly employed in reconstructing high-quality images of the distribution of radioactive substances in the body. In the comprehensive investigation conducted in [Da Costa-Luis, 2017], a groundbreaking and inventive approach was developed for the reconstruction of positron emission tomography (PET) images. The study aimed to overcome the limitations inherent in the conventional Maximum Likelihood Ex-

pectation Maximization (ML-EM) method, with a specific focus on addressing the formidable challenge of reconstructing intricate structures influenced by resolution modeling.

The combination of IR and ML-EM is described in Figure (2) [Fan, 2017; Yang, 2022; Rodriguez, 2023]. To address these challenges, the researchers in [Mizusawa, 2021; Roslin, 2023; Fu, 2023; Xia, 2023; Reader, 2024] proposed an innovative post-processing step employing deep learning (DL) techniques, specifically utilizing a Convolutional Neural Network (CNN) (Figure 3). This strategic integration aimed to significantly enhance overall image quality while simultaneously paving the way for potential dose reduction in PET imaging—a crucial advancement in the field. The ML-EM update blocks will contain the Maximum Likelihood Expectation Maximization (MLEM) algorithm with the image being reconstructed over $k$ iterations. The initial image is the input image having the same dimensions as the reconstructed image (resulting from the back projection of the sinogram image). The outcome of the method is the reconstructed image after $k$ iterations. An overview of the differences between the mentioned methods and our proposed method is provided in Table (1).

## 3 Proposed Method

Based on experimental results and comparisons of AI-based image reconstruction methods [Reader, 2023], the results indicate that "AI methods without training data" yield the poorest outcomes compared to the two methods utilizing supervised learning on labeled datasets, and approach the performance of traditional methods like Maximum Likelihood Expectation Maximization (ML-EM). From these results, we experimented and observed improvements in speed and quality when using supervised learning to fine-tune the ESRGAN model for enhancing image quality. Additionally, incorporating the UNet architecture into traditional methods like ML-EM enhanced convergence speed and ensured high accuracy in the reconstructed images.

In this paper, we utilize the Maximum Likelihood Expectation Maximization (ML-EM) image reconstruction method as the basis for our work. We have developed an integrated approach that incorporates a deep learning filter using UNet and enhances image resolution by fine-tuning an ESRGAN model specifically designed for CT image data.

Based on the structure of ML-EM integrated with CNN (Figure 3), we aimed to improve the basic CNN model (comprising convolutional layers and PReLU layers) by transforming it into a UNet model. Additionally, we enhanced the model by improving the output image quality using a fine-tuned ESRGAN deep learning model. We refer to our method as DL-MLEM-IR-UNET-ESRGAN.

UNet is a convolutional neural network architecture designed for biomedical image processing tasks. It consists of a contracting path to capture context and a symmetric expanding path that enables precise localization. UNet's architecture includes skip connections between corresponding layers, aiding in the preservation of spatial information and improving segmentation accuracy. UNet has been widely used in various medical image segmentation tasks, such as identifying organs, tumors, or abnormalities in medical scans like MRI or CT images [Ronneberger, 2015]. The improved UNet architecture proposed in the paper is described in Table (2).

The UNet architecture includes an encoder, middle, and decoder. The encoder part consists of two simple convolutional layers and a max-pooling operator. In both convolutional layers, we use a filter size of $3 \times 3$ with a stride of 1 and output with 64 channels. Each convolutional layer is followed by a ReLU (Rectified Linear Unit) activation function. A max-pooling operator $2 \times 2$ with a stride of 2 is used after the convolutional layers. The middle part consists of two convolutional layers, each of which is followed by a filter size of $3 \times 3$ with a stride of 1 and output with 128 channels. Finally, the decoder part includes a transposed convolutional layer and a convolutional layer. In the transposed convolutional layer, followed by a ReLU activation function, we use a filter size of $2 \times 2$ with a stride of 2 and output with 64 channels. The convolutional layer includes 64 neurons $3 \times 3$ with a stride of 1 and output with 1 channel. The structure of UNet is depicted in Figure (4).

The Enhanced Super-Resolution Generative Adversarial Network (ESRGAN) represents a significant leap in super-resolution image quality, building upon the foundation of the Super-Resolution Generative Adversarial Network (SRGAN). ESRGAN stands out due to its advanced architecture, which utilizes Residual in Residual Dense Blocks (RRDB) without batch normalization. This enables the modeling of intricate textures, surpassing the capabilities of SRGAN. It revolutionizes loss functions by using pre-activation features to enhance structure restoration and employing a perceptual GAN loss to assess relative realism, resulting in more lifelike textures [Wang, 2019].

The success of ESRGAN can be attributed to meticulous refinements in its architecture and loss functions. These modifications play a pivotal role in creating a robust and stable training process, thereby reducing the dependency on extensive hyperparameter tuning. This streamlined approach not only enhances the efficiency of the training phase but also contributes to the overall reliability of ESRGAN in delivering high-quality results consistently.

One of the noteworthy features of ESRGAN is its capacity to set a new standard in super-resolution technology. By pushing the boundaries of image enhancement, it has become a beacon of excellence in the field. The images produced by ESRGAN exhibit a level of realism that was previously unparalleled, making it the preferred choice for a wide array of applications (Figure 5).

The operational diagram illustrating the proposed DL-MLEM-IR-UNET-ESRGAN method is shown in Fig-

Table 1. An overview of the differences between the mentioned methods and our proposed method

| Method | Algorithm | Deep learning approach | Execution speed | Training data needs |
|---|---|---|---|---|
| FBP | One step | No | Fast | No |
| DL–FBP | One step | Yes | Fast | No |
| DL–BPF | One step | Yes | Fast | No |
| DL–FBP–F | One step | Yes | Fast | No |
| MLEM–IR | Iterative | No | Slow | No |
| DL–MLEM–IR | Iterative | Yes | Fast | No |
| DL–MLEM–IR–ESRGAN (ours) | Iterative | Yes | Fast | Yes |

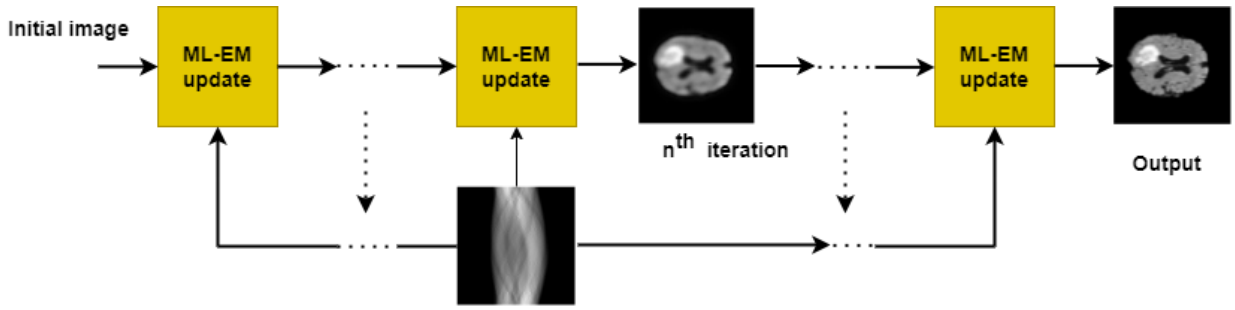


Figure 2. The iterative structure of ML-EM

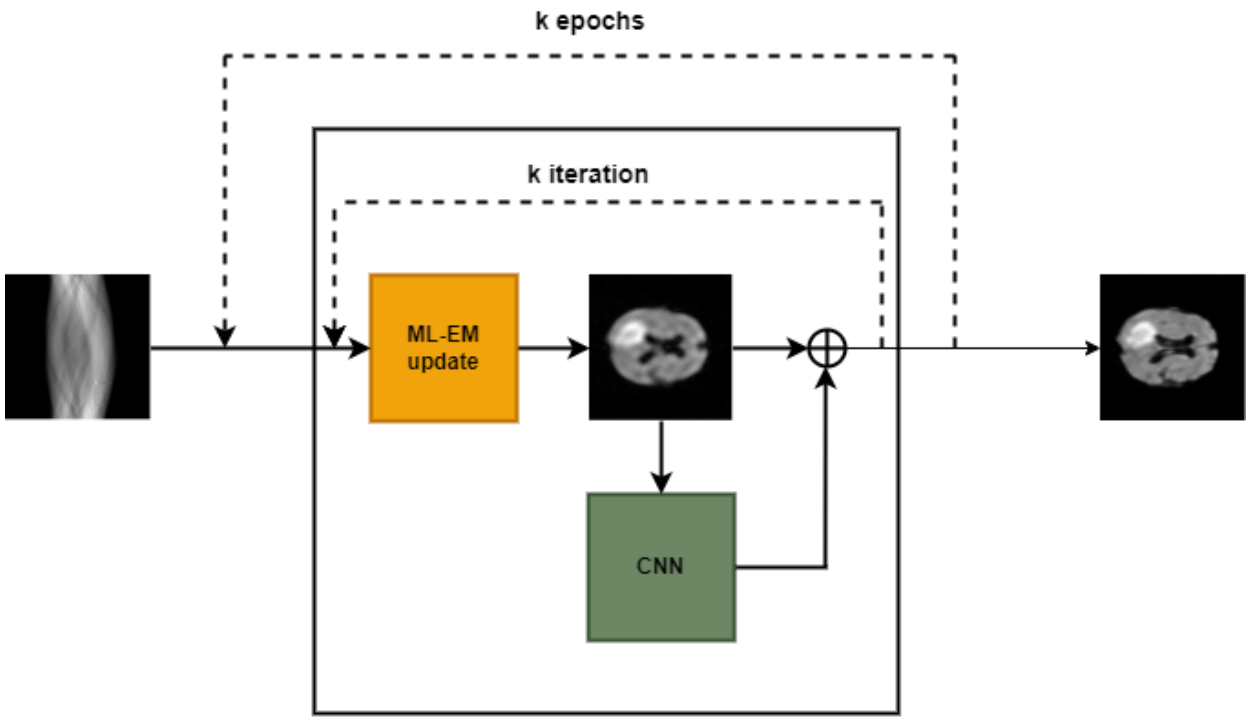


Figure 3. The iterative structure of ML-EM integrated with CNN

ure (7), which enhances image reconstruction from sinogram data. The method begins with a sinogram, which undergoes a series of ML-EM updates through $k$ iterations, aiming to produce an initial reconstruction. Following ML-EM processing, the UNet, a deep learning model (as shown in Table (2)), further refines the image by reducing artifacts and improving structural details. This refined image then passes through the fine-

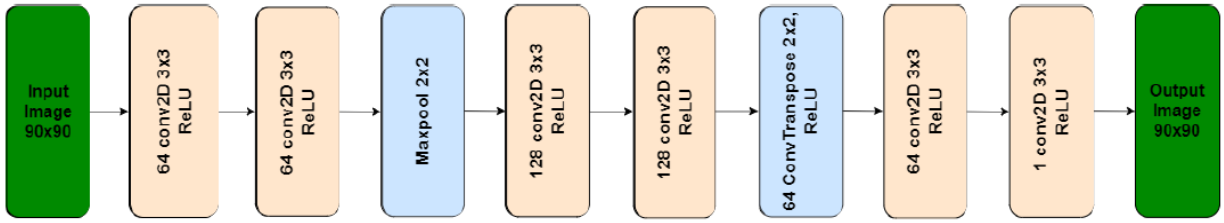


Figure 4. Structure of the proposed UNET

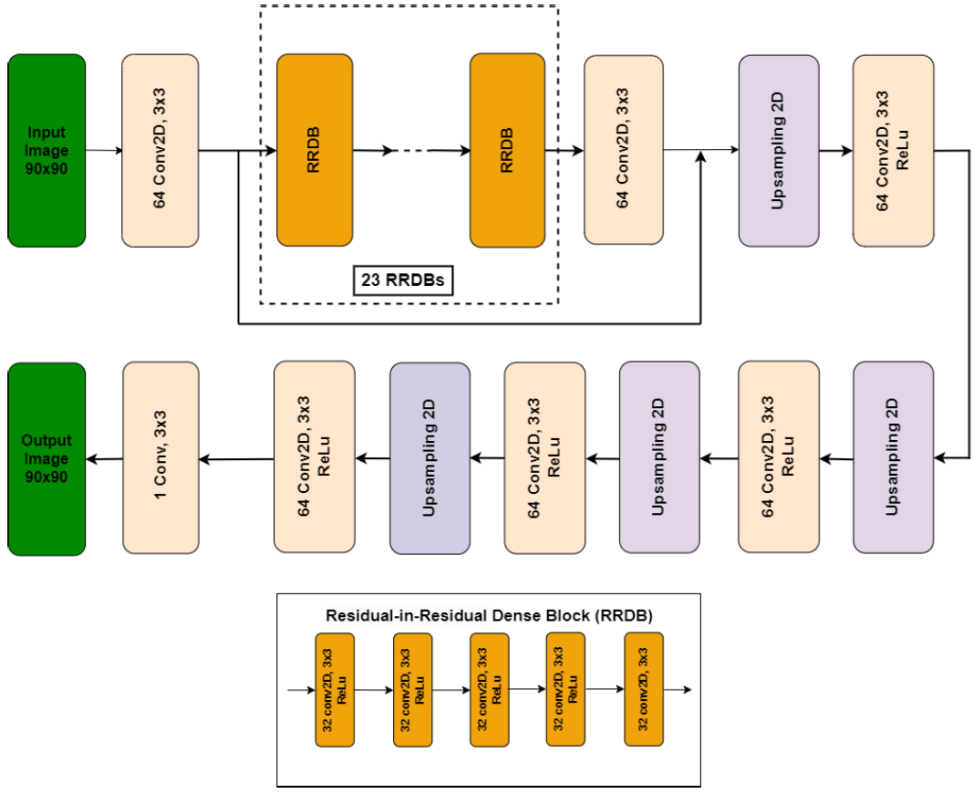


Figure 5. Structure of the ESRGAN

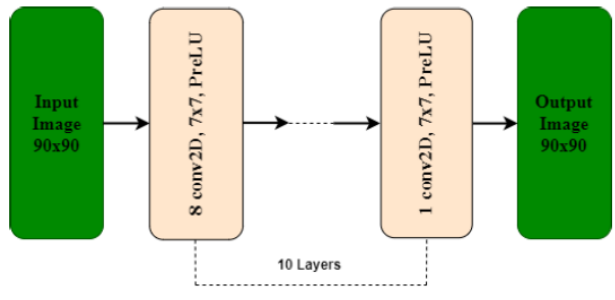


Figure 6. Structure of the CNN

tuned Enhanced Super-Resolution Generative Adversarial Network (ESRGAN), which elevates the image resolution and sharpness. This process is reiterated for $k$ epochs, progressively enhancing the reconstruction. The final output is a high-resolution, detailed reconstructed image derived from the original sinogram input. The ESRGAN model used in this study has been fine-tuned from the pretrained RRDB-ESRGAN-x4 model [Wang, 2019] on the CT image BraTs20 dataset [1].

We utilized 3,880 high-quality CT images with a size

[1] https://www.med.upenn.edu/cbica/brats2020/data.html

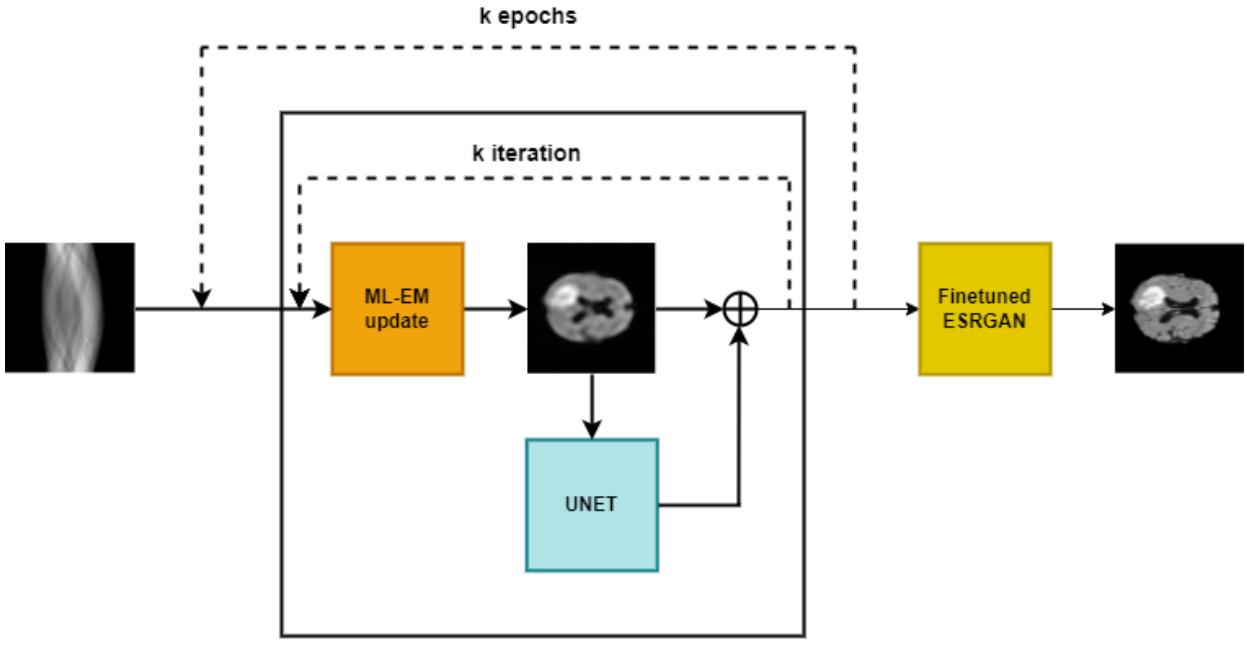


Figure 7. Structure of the DL-MLEM-IR-UNET-ESRGAN Method

Table 2. Proposed U-Net Model

| Stage | Layer | Kernel size |
|---|---|---|
| Encoder | Conv2d(1,64),s1, ReLU | 3 |
| Encoder | Conv2d (64,64)s1, ReLU | 3 |
| Encoder | MaxPool2d, s2 | 2 |
| Middle | Conv2d (64,128), s1, ReLU | 3 |
| Middle | Conv2d (128,128),s1, ReLU | 3 |
| Decoder | ConvTranspose2d (128,64), s2, ReLU | 2 |
| Decoder | Conv2d (64, 1), s1,ReLU | 3 |
| Total params | 292417 | |

Table 3. Comparison of Fine-Tuned Models' Results

| Model | PSNR | SSIM |
|---|---|---|
| RRDB_ESRGAN_x4 (pretrained) | 45.3524 | 0.9951 |
| 400 frozen layers + 10 epoch (finetuned) | 41.8152 | 0.9981 |
| 500 frozen layers + 10 epoch (finetuned) | 42.6025 | 0.9983 |
| 600 frozen layers + 10 epoch (finetuned) | 45.4058 | 0.9983 |

of 369 × 369 pixels[2] and corresponding low-quality images, automatically resized to a resolution of 90×90 pixels, for fine-tuning. With a total of 702 layers in the pretrained model, we experimented with three cases of fine-tuning with frozen layers, including 400, 500, and 600 frozen layers. We selected the fine-tuned model with 600 frozen layers as it produced the best results compared to the model without fine-tuning (as shown in Table 3).

[2] https://github.com/pacotha/Dataset_CT-images_.git

## 4 Experiments

The results were obtained by calculating the average PSNR and SSIM [Bovik, 2006], after testing on an evaluation image dataset[2]. Some images from the dataset will be displayed in Figure (8). The SSIM formula used for evaluation is as follows:

$$\text{SSIM}(x,y) = \frac{(2\mu_x\mu_y + c_1)(2\sigma_{xy} + c_2)}{(\mu_x^2 + \mu_y^2 + c_1)(\sigma_x^2 + \sigma_y^2 + c_2)} \quad (1)$$

where $\mu_x, \mu_y$ are the mean values of images $x$ and $y$, respectively. $\sigma_x, \sigma_y$ are the standard deviations of images $x$ and $y$, respectively, $\sigma_{xy}$ is the cross-covariance between $x$ and $y$, and $c_1, c_2$ are two constants. These factors help SSIM provide an overall assessment of the similarity between the structures of two images, considering not only brightness and color but also examining the structure and transformations within the images. PSNR (Peak Signal-to-Noise Ratio) is defined through Mean Squared Error (MSE). MSE is a statistical concept, representing the average of the squared differences between estimates and the ground truth. PSNR is calculated as follows:

$$PSNR = 10\log_{10}\left(\frac{\text{MAX}_I^2}{MSE}\right) = 20\log_{10}\left(\frac{\text{MAX}_I}{\sqrt{MSE}}\right)$$

where $\text{MAX}_I$ is the maximum value of a pixel in the image, typically 255 for an 8-bit image. MSE is the Mean

Squared Error, the average of the squared differences between corresponding pixels of the two images:

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

where $m$ and $n$ are the dimensions of the image (number of rows and columns). $I(i,j)$ is the value of the pixel at row $i$ and column $j$ in the original image. $K(i,j)$ is the value of the corresponding pixel in the comparison or result image.

We trained for 2600 epochs using Mean Squared Error (MSE) loss. The experimental dataset comprises CT images obtained from the evaluation dataset. These images were preprocessed into 2D images with dimensions of 90×90 pixels and then forward-projected to create input images for the methods. The detailed content of the experimental methods is described in Table (4). A visual depiction of the experimental process is encapsulated in Figure (6), which features 10 representative images from the experimental dataset alongside their corresponding reconstructions. The architectures of the CNN and UNET models used are illustrated in Figures (4) and (6), as described in Table (2).

Table 4. Details of the Experimental Methods

| Abbreviation | Method |
|---|---|
| DL-FBP | FBP + CNN |
| DL-FBP-F | FBP-F + CNN |
| DL-BPF | BPF + CNN |
| ML–EM-IR | IR |
| DL-ML–EM-IR | IR + CNN |
| DL–ML–EM-IR-UNET | IR+ U-Net |
| DL-ML–EM-IR-UNET-ESRGAN | IR + U-Net+ ESRGAN |

The experimental evaluation of the proposed methods is comprehensively presented in Tables (5) and (6), showcasing PSNR and SSIM values across the distinct methods as follows: DL-FBP, DL-FBP-F, DL-BPF, ML-EM-IR, DL-ML-EM-IR, DL-ML-EM-IR-UNET (Ours without ESRGAN), and DL-ML-EM-IR-UNET-ESRGAN (Ours with ESRGAN). Corresponding to each method, the labels A, B, C, D, E, F, G, H, I, J are assigned to represent the individual images within the experimental dataset. Table (7) consolidates the average PSNR, SSIM, and runtime values derived from the experimental outcomes.

The methods that do not use ESRGAN (DL-FBP-F, DL-BPF, ML-EM-IR, DL-ML-EM-IR, DL-ML-EM-IR-UNET) were built based on the source code and methods mentioned in the Preliminaries section. The architectures were constructed and tested in the same setup environment to ensure fairness among the methods. Deep learning models such as CNN and UNet were optimally selected during the experimentation process, and the methods all use the same CNN or UNet architecture. The experiments were executed on Google Colab, leveraging a GPU T4, with a standardized configuration of 2600 iterations. Notably, both input and reconstructed images adhered to dimensions of $90 \times 90$ pixels.

Methods incorporating deep learning techniques, particularly those utilizing U-Net and ESRGAN, demonstrate superior results in both structural similarity index (SSIM) and peak signal-to-noise ratio (PSNR). Notably, the DL-ML-EM-IR-UNET-ESRGAN method achieves the highest SSIM (0.9980) and PSNR (53.2119), indicative of superior reconstruction quality.

Compared to the baseline DL-FBP method, all techniques exhibit improvements in SSIM and PSNR, with the exception of DL-FBP-F and DL-BPF. However, computational efficiency varies among methods, with iterative reconstruction (IR) methods like ML-EM-IR requiring the longest runtimes. Conversely, approaches combining IR with U-Net, such as DL-MLEM-IR-UNET and DL-ML-EM-IR-UNET-ESRGAN, strike a balance between performance and runtime efficiency. These findings underscore the effectiveness of hybrid methodologies, emphasizing the potential of integrating iterative reconstruction with advanced deep learning architectures to achieve high-quality image reconstructions in a computationally efficient manner.

The proposed method DL-ML-EM-IR-UNET-ESRGAN emerges as a compelling approach, offering the highest levels of image reconstruction quality as evidenced by its superior SSIM and PSNR scores. Moreover, its relatively modest runtime signifies a favorable balance between computational efficiency and reconstruction performance. This highlights the method's potential as a practical solution for high-fidelity image reconstruction tasks, underscoring the significance of integrating iterative reconstruction techniques with advanced deep learning architectures like U-Net and ESRGAN.

## 5 Conclusion

In this paper, we have proposed a method that combines the potential of Iterative Reconstruction techniques with the ML-EM algorithm, alongside the U-Net deep learning model and the ESRGAN model. The research results demonstrate the superior capabilities of this method, as the U-Net model significantly improves reconstruction time while maintaining high accuracy between the reconstructed image and the original image. Furthermore, the involvement of image generation models like ESRGAN enhances the quality of the reconstructed images. The comparison results also show that the IR image reconstruction technique outperforms older techniques like FBP, FBP-F, and BPF, even though all of them apply deep learning models to reconstruction. The research also highlights the potential application of

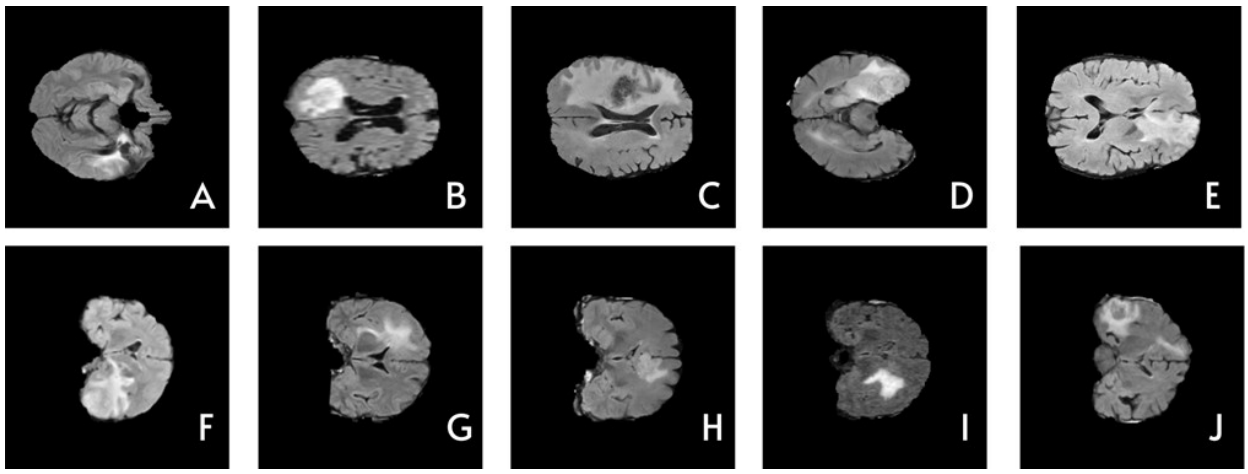


Figure 8. CT images extracted from the experimental dataset

Table 5. SSIM results during experimental process

| Image | DL–FBP | DL–FBP–F | DL–BPF | ML–EM–IR | DL–ML–EM–IR | Ours (Unet) | Ours (Unet+ESRGAN) |
|---|---|---|---|---|---|---|---|
| A | 0.8869 | 0.7102 | 0.5884 | 0.9942 | 0.9932 | **0.9983** | 0.9956 |
| B | 0.8757 | 0.7056 | 0.5952 | 0.9992 | 0.9963 | 0.9985 | **0.9993** |
| C | 0.8823 | 0.711 | 0.5874 | 0.9965 | 0.9977 | 0.9950 | **0.9991** |
| D | 0.8735 | 0.7035 | 0.5802 | **0.9983** | 0.9946 | 0.9962 | 0.9976 |
| E | 0.8703 | 0.7068 | 0.5729 | 0.9977 | 0.9952 | 0.9971 | **0.9988** |
| F | 0.8726 | 0.7341 | 0.6015 | 0.9952 | 0.9948 | 0.9974 | **0.9976** |
| G | 0.8815 | 0.7118 | 0.6041 | 0.9956 | 0.9974 | 0.9968 | **0.9978** |
| H | 0.8747 | 0.7076 | 0.5905 | **0.9994** | 0.9982 | 0.9986 | 0.9969 |
| I | 0.8763 | 0.7102 | 0.5809 | **0.9981** | 0.9954 | 0.9969 | **0.9981** |
| J | 0.8725 | 0.7037 | 0.5787 | 0.9993 | 0.9966 | 0.9972 | **0.9988** |
| Average | 0.8766 | 0.7105 | 0.5301 | 0.9974 | 0.9959 | 0.9972 | **0.9980** |

Table 6. PSNR results during experimental process

| Image | DL–FBP | DL–FBP–F | DL–BPF | ML–EM–IR | DL–ML–EM–IR | Ours (Unet) | Ours (Unet+ESRGAN) |
|---|---|---|---|---|---|---|---|
| A | 29.0838 | 29.0547 | 24.7382 | 52.1824 | 46.1259 | 51.8272 | **53.1852** |
| B | 29.8876 | 28.7805 | 24.1236 | **52.6621** | 46.1351 | 52.5388 | 53.2863 |
| C | 30.1794 | 28.7764 | 24.6512 | 52.4982 | 45.8926 | 51.8783 | **52.8956** |
| D | 29.8932 | 28.9321 | 24.4345 | 52.6749 | 45.7639 | 52.0095 | **53.1269** |
| E | 29.9892 | 28.8763 | 24.3442 | 52.1299 | 46.1006 | 51.8342 | **53.1008** |
| F | 29.5928 | 28.8632 | 24.7327 | 52.8861 | 45.9821 | 52.4981 | **53.4561** |
| G | 29.6511 | 29.0051 | 24.4852 | 52.3786 | 45.8782 | 52.1194 | **53.4562** |
| H | 29.932 | 28.7563 | 24.8214 | 52.6712 | 46.1006 | 52.3242 | **52.8968** |
| I | 29.7452 | 29.0023 | 24.6731 | 52.4251 | 46.1118 | 52.1191 | **53.1248** |
| J | 30.1265 | 28.8761 | 24.6651 | 52.7382 | 45.7425 | 52.3224 | **53.5901** |
| Average | 29.8081 | 28.8923 | 24.5669 | 52.5247 | 45.9833 | 52.1471 | **53.2119** |

deep learning models in CT image reconstruction methods to improve image quality and reduce radiation exposure during the imaging process.

Despite the promising results, the proposed method has several limitations. Firstly, the computational cost associated with the implementation of the U-Net and ESRGAN models is relatively high, which may hinder real-time application and widespread clinical adoption. Secondly, the training process requires a large dataset of high-quality images, which may not always be available,

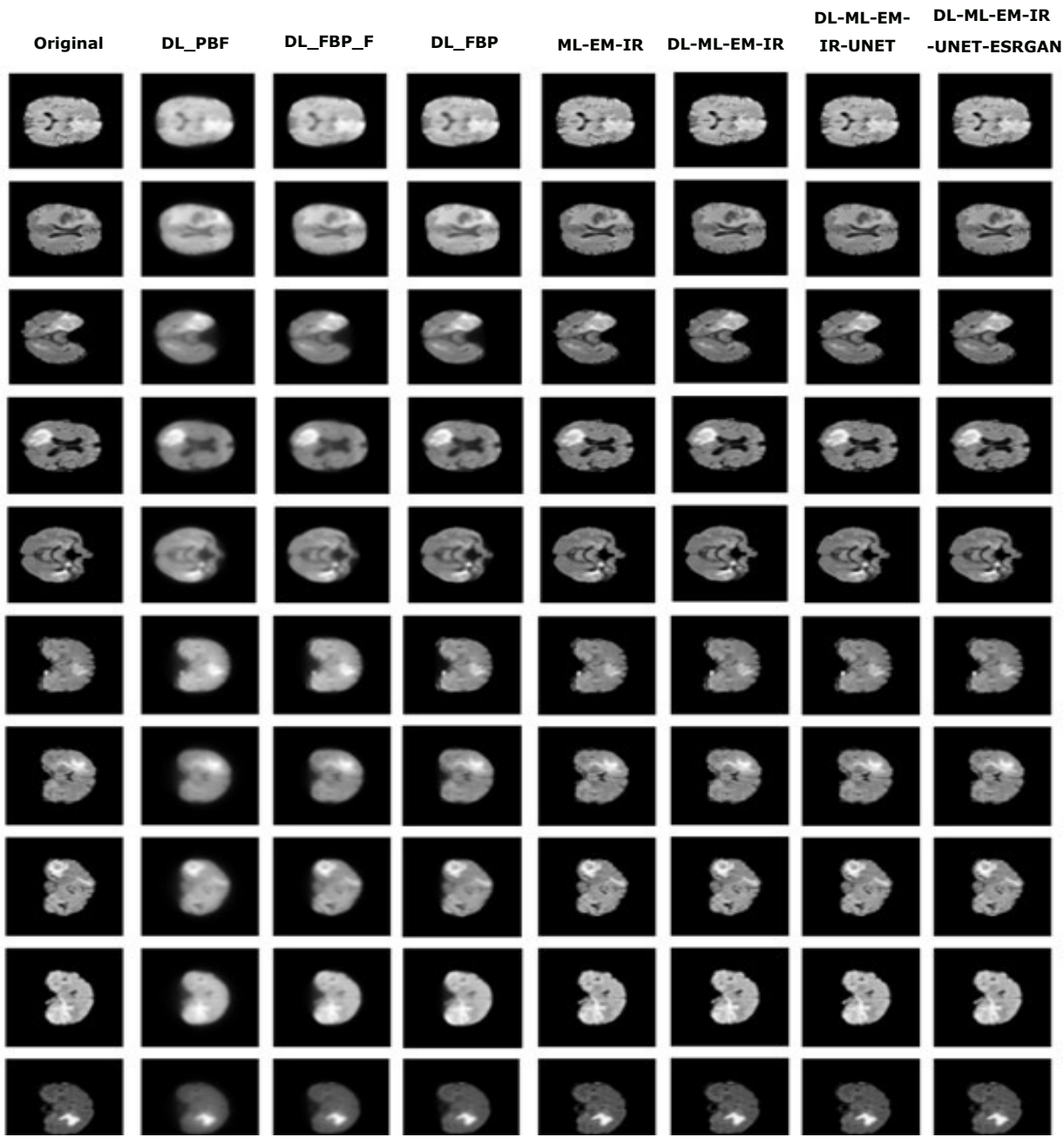


Figure 9. Comparison of the reconstruction results (2600 iterations) obtained by compared methods

especially in specific medical imaging contexts. Additionally, the method's performance has been evaluated on a limited dataset, and further validation on diverse datasets is needed to ensure robustness and generalizability.

Future research should focus on optimizing the computational efficiency of the proposed method, potentially through model compression techniques or the use of more efficient architectures. Expanding the dataset and including more diverse imaging conditions in the training process will be essential to enhance the method's generalizability. Moreover, integrating this approach with other advanced techniques, such as attention mechanisms or hybrid models, could further improve image reconstruction quality. Lastly, clinical trials and real-world testing should be conducted to validate the practical benefits of the method in reducing radiation exposure and improving diagnostic accuracy in medical imaging.

**Acknowledgements**

Pham Cong Thang (corresponding author) would like to thank his colleagues at IT Faculty, DUT, Danang, Vietnam, for their helpful comments. This work was supported by The University of Danang–University of Science and Technology, code number of Project: **T2023–02–05MSF**.

Table 7. Average Results of comparing the experimented methods during experimental process

| Method | SSIM | PSNR | Runtime (s) |
|---|---|---|---|
| DL-FBP | 0.8766 | 29.8081 | 139 |
| DL-FBP-F | 0.7105 | 28.8923 | 144 |
| DL-BPF | 0.5301 | 24.5669 | 193 |
| MLEM-IR | 0.9974 | 52.5247 | 1705 |
| DL-MLEM-IR | 0.9959 | 45.9833 | 239 |
| DL-MLEM-IR-UNET | 0.9972 | 52.1471 | **124** |
| DL-MLEM-IR-UNET-ESRGAN | **0.9980** | **53.2119** | 131 |

## References


Akagi M. et al. (2019). Deep learning reconstruction improves image quality of abdominal ultra-highresolution CT. *European Radiology*, **29** 11, pp. 6163–6171.

Bovik A.C., Wang Z. (2006). *Modern Image Quality Assessment, Synthesis Lectures on Image, Video, and Multimedia Processing*. Morgan and Claypool Publishers, 146 pages.

Da Costa-Luis. C.O, Reader A. J. (2017). Deep Learning for Suppression of Resolution-Recovery Artefacts in MLEM PET Image Reconstruction. *IEEE Nuclear Science Symposium and Medical Imaging Conference (NSS/MIC)*, pp. 1–3.

Erofeeva V., Galyamina V., Oleg Granichin O. and et al. (2019). Detection of specific areas and densities for ultrasound tomography. *Cybernetics and Physics*, **8** 3, pp. 121–127.

Fallahpoor M. et al. (2024). Deep learning techniques in PET/CT imaging: A comprehensive review from sinogram to image space. *Computer Methods and Programs in Biomedicine*, **243**, pp. 1–13.

Fan H., Zhu H., Zhao X., Zhang J., Wu D., Han Q. (2017). Ultrasonic image reconstruction based on maximum likelihood expectation maximization for concrete structural information. *Computers & Electrical Engineering*, **62**, pp. 293–301.

Fu Y. et al. (2023) AIGAN: Attention–encoding Integrated Generative Adversarial Network for the reconstruction of low-dose CT and low-dose PET images. *Medical Image Analysis*, **86**, pp. 1–14.

Geyer L.L. et al. (2015). State of the Art: Iterative CT Reconstruction Techniques. *Radiology*, **276** 2, pp. 339–357.

Granichin O., Erofeeva V., Senin I. (2018) Modifying the physical process of ultrasound tomography scanning through compressive sensing. *Cybernetics and Physics*, **7** 2, pp. 66–71.

Hashimoto F., Ote K., Onishi Y. (2022) PET Image Reconstruction Incorporating Deep Image Prior and a Forward Projection Model. *IEEE Transactions on Radiation and Plasma Medical Sciences*, **6** 8, pp. 841–846.

Kalra M.K., Sodickson A.D., Mayo-Smith W.W. (2015). CT radiation: key concepts for gentle and wise use. *Radiographics*, **35** 6, pp. 1706–1721,

Liu Y., Wu G., Lv Z. SDGAN: A novel spatial deformable generative adversarial network for low-dose CT image reconstruction. *Displays*, **78**, pp. 1–9.

McLeavy C.M. et al. The future of CT: deep learning reconstruction. *Clinical Radiology*, **76**, pp. 407–415.

Meyer N. K. et al. (2024). Model-based iterative reconstruction for direct imaging with point spread function encoded echo planar MRI. *Magnetic Resonance Imaging*, **109**, pp. 189–202

Mizusawa S., Sei Y., Orihara R., Ohsuga A. (2021). Computed tomography image reconstruction using stacked U-Net. *Computerized Medical Imaging and Graphics*, **90**, pp. 1–10.

Pham, C.T., Kopylov, A.V. (2015) Multi-quadratic dynamic programming procedure of edge-preserving denoising for medical images. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, **XL-5/W6**, pp. 101–106.

Reader A. J. (2024). Self-Supervised and Supervised Deep Learning for PET Image Reconstruction. *AIP Conference Proceedings*, 3061, 030003, pp. 1–13.

Reader A.J., Pan B. (2023). AI for PET image reconstruction. *The British Journal of Radiology*, **96** 1150, pp. 1–17.

Rodriguez-Alvarez M.J. (2023). Expectation maximization (EM) algorithms using polar symmetries for computed tomography (CT) image reconstruction. *Computers in Biology and Medicine*, **43** 8, pp. 1053-1061

Ronneberger O., Fischer P., Brox T. (2015). UNet: Convolutional Networks for Biomedical Image Segmentation. *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015. MICCAI 2015, Lecture Notes in Computer Science*, **9351**, Springer, Cham, pp. 234–241.

Roslin A., Lebedev M., Mitchell T.R., Onederra I.A.,Leonardi C.R. (2023). Processing of micro-CT images of granodiorite rock samples using convolutional neural networks (CNN). Part III: Enhancement

of Scanco micro-CT images of granodiorite rocks using a 3D convolutional neural network super-resolution algorithm. *Minerals Engineering*, **195**, pp. 1–10.

Schofield R. et al. (2020). Image reconstruction: Part 1–understanding filtered back projection, noise and image acquisition. *Journal of Cardiovascular Computed Tomography*, **14** 3, pp. 219–225.

Shin H.B. et al. (2021). Application of sigmoidal optimization to reconstruct nuclear medicine image: Comparison with filtered back projection and iterative reconstruction method. *Nuclear Engineering and Technology*, **53** 1, pp. 258–265.

Stiller W. (2018). Basics of iterative reconstruction methods in computed tomography: A vendor–independent overview. *European Journal of Radiology*, **109**, pp. 147–154.

Tan X., Liu X., Xiang K.,Wang J., Tan S. (2024). Deep Filtered Back Projection for CT Reconstruction. *IEEE Access*, **12**, pp. 20962–20972.

Tran T.T.T., Pham C.T. et al. (2023) Non-Convex Hybrid Total Variation for Restoring Medical Image Corrupted by Poisson Noise *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, **XLVIII-2/W3-2023**, pp. 255–260.

Ustaoglu Z. (2023). Iterative inversion of Radon transform via discretization by fuzzy basic functions. *Journal of Computational and Applied Mathematics*, **430**, pp. 1–15.

Wang X. et al. (2019). ESRGAN: Enhanced Super Resolution Generative Adversarial Networks. *Proceedings of the European Conference on Computer Vision (ECCV) Workshops*, pp. 63–79.

Willemink M.J., Noel P.B. (2019). The evolution of image reconstruction for CT—from filtered back projection to artificial intelligence. *European Radiology*, **29** 5, pp. 2185–2195.

Xia Z., Liu J., Kang Y., Wang Y.,Hu D., Zhang Y. (2023) Dynamic controllable residual generative adversarial network for low-dose computed tomography imaging. *Quantitative Imaging in Medicine and Surgery*, **13**, pp. 5271–5286.

Yang H. et al. (2022). Transmission reconstruction algorithm by combining maximum-likelihood expectation maximization and a convolutional neural network for radioactive drum characterization. *Applied Radiation and Isotopes*, **184**, pp. 1–12.