

Learning Low-Dose CT Image Denoising From An Abstract Dataset

Paulo Alexandre Ribeiro de Carvalho

Data Science Master Student, École Polytechnique Fédérale de Lausanne, CH
paulo.ribeirodecarvalho@epfl.ch

Semester Research Project at IVRL
Code available [here](#)

Abstract. This paper introduces a novel approach for denoising low-dose Computed Tomography (CT) images using a synthetic dataset inspired by Dead Leaves (DL). The employed denoising model is a DnCNN. The synthetic DL dataset is leveraged for its unique properties, providing a versatile and realistic alternative for training denoising algorithms. Extensive experiments demonstrate the effectiveness of the proposed method in improving image quality and reducing noise artifacts in low-dose CT imaging. This approach has the potential to enhance diagnostic accuracy and clinical utility while addressing the challenges associated with the availability of real CT image pairs and the introduction of unpredictable artifacts by traditional methods. Our results indicate that models trained on DL images, especially those with added textures, can achieve performance nearly comparable to those trained on real CT images, highlighting the promise of synthetic datasets in medical image denoising.

Keywords: Computed Tomography · Synthetic Dataset · Machine Learning · Statistics · Image Quality · Noise Reduction

1 Introduction

The widespread use of Computed Tomography (CT) scans in medicine has sparked concerns about radiation exposure for patients. Lowering radiation often increases image noise, which can compromise diagnostic quality. Therefore, developing effective methods for denoising low-dose CT images is crucial for maintaining image quality while minimizing radiation exposure. Traditional denoising techniques, including filter-based and transform-based methods, often struggle to balance noise removal and detail preservation. Recently, deep learning-based methods have shown significant promise in denoising tasks due to their ability to learn complex mappings from noisy to clean images. Training deep learning models typically requires large datasets of noisy and clean image pairs, which are not always readily available. Synthetic datasets, generated using models such as

the Dead Leaves (DL), offer an alternative. These datasets, characterized by randomly placed overlapping objects, can simulate the variability and complexity found in real images.

This paper explores the feasibility of training the Denoising Convolutional Neural Network (DnCNN) model on synthetic datasets, specifically using DL images, to achieve nearly comparable performance to models trained on real low-dose high-dose CT images pair. By focusing on noise reduction rather than solely minimizing the mean squared error (MSE) of the CT image, this approach aims to limit the introduction of inexplicable artifacts that deep learning models may add in pursuit of their loss objective. We demonstrate that the abstract nature of the DL images provides sufficient diversity and complexity to train effective denoising models, potentially simplifying the dataset acquisition process for low-dose CT denoising tasks. This investigation aims to show that nearly similar performance can be achieved while limiting the uncontrolled hallucinations of models. Computational resources of this project: NVIDIA A100-SXM4-40GB

2 Related Works

Previous research has demonstrated various denoising methods for low-dose CT images [9], [7]. However, it has been shown that training using CT images can lead to hallucinations when denoising out-of-sample images [3]. These hallucinations refer to the introduction of artifacts that were not present in the original images, potentially compromising the diagnostic value of the denoised images.

Efforts have been made to control these hallucinations during training by incorporating feedback mechanisms from the model's confidence when reconstructing an image [8]. This approach aims to reduce the likelihood of hallucinations by ensuring that the model only generates denoised outputs that it deems reliable.

Additionally, there has been research into using synthetic datasets for image reconstruction [1]. This work investigates the use of synthetic images, such as those generated by the DL model, as prior datasets for training reconstruction models. The results indicate that models trained solely on synthetic images achieve performance close to those trained on natural images. Moreover, models trained with a combination of synthetic and natural images perform similarly to those trained exclusively on natural images. The study also highlights the simplicity of generating these synthetic images, which requires only a few parameters: α , r_{min} , and r_{max} . Our work builds on these foundations by exploring the feasibility of using synthetic datasets, specifically DL images, to train the DnCNN model for low-dose CT image denoising.

3 Computed Tomography

CT imaging is a widely utilized diagnostic tool in modern medicine, renowned for its ability to provide detailed cross-sectional images of the body's internal structures. Unlike traditional X-ray imaging, which compresses three-dimensional in-

formation into a two-dimensional plane, CT scans create a series of detailed slices that can be reconstructed into a comprehensive 3D representation [5]. This capability is particularly valuable for diagnosing and monitoring a variety of medical conditions, including cancers, cardiovascular diseases, and traumatic injuries. By offering superior contrast resolution and the ability to distinguish between different tissue types, CT imaging plays a crucial role in clinical decision-making and treatment planning. However, the high radiation doses associated with CT scans pose significant health risks [4].

3.1 Low Dose CT Grand Challenge

This work is based on the dataset provided by the *Low Dose CT Grand Challenge*, which is available for download from their website.

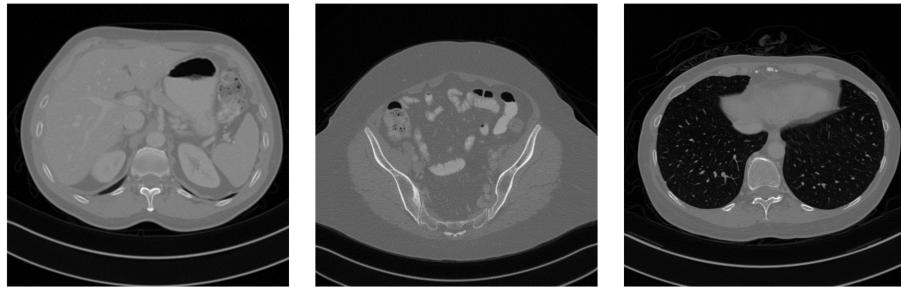


Fig. 1. Random example of images from Low Dose Grand Challenge dataset

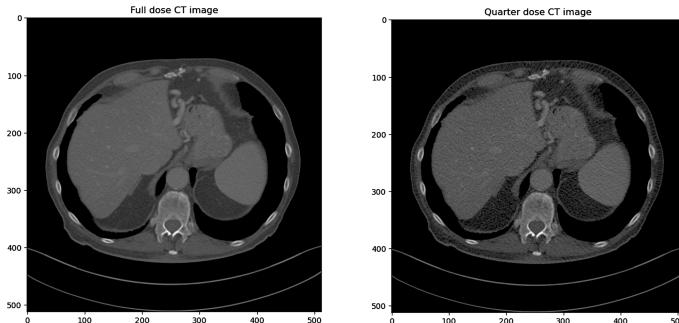


Fig. 2. Low-full dose pair CT¹

The dataset comprises over fifteen thousand CT slices pair (low-full dose, Fig. 2) from 10 patients, scanned under four different contexts: 1mm B30, 1mm

¹ Contrast of quarter dose CT have been augmented for better noise visualization.

D45, 3mm B30, and 3mm D45. These varying scan protocols provide a diverse range of imaging conditions, allowing for comprehensive evaluation of denoising algorithms.

3.2 Artifacts

Several challenges are induced while using real CT images in training a denoising model. Firstly, obtaining real CT datasets is inherently difficult due to privacy concerns and the complexity involved in acquiring paired full-dose and low-dose images. Secondly, models trained on real CT images are prone to hallucinations, where the denoising process introduces structures that do not exist in the original image. For instance, in Fig. 4, the true full-dose image shows no abnormalities in a specific region, while the denoised image, trained on real CT data, introduces artificial structures. These artifacts can result from the model replicating patterns observed during training on CT images, leading to potentially misleading or incorrect diagnostic information. These issues underscore the need for alternative approaches, such as using synthetic datasets, to train denoising models.

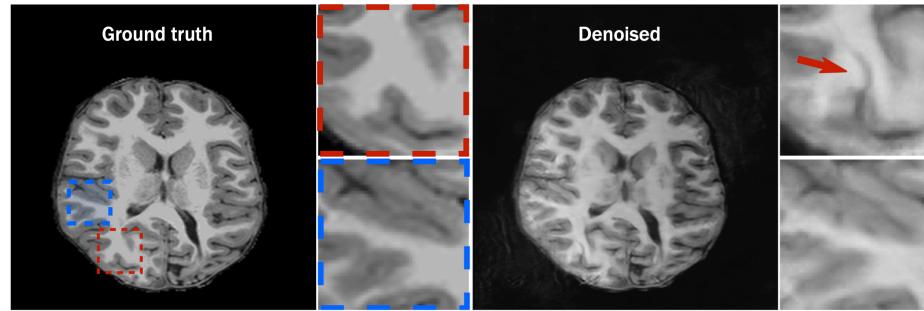


Fig. 3. Artifact introduction from CT denoising model. [3]

4 Synthetic Images

As discussed in Section 3.2, training denoising models with actual CT images has inherent limitations. These challenges motivate the exploration of using abstract and synthetic datasets for training denoising models. Synthetic images offer a controlled environment for model training. The key advantage of using synthetic images lies not in the elimination of artifacts but in the ability to better understand and explain the artifacts that do arise. Unlike models trained on real CT images, where the origin and nature of artifacts can be complex and unpredictable, synthetic datasets provide a structured framework where artifacts are more predictable and analyzable, leading to the development of more robust denoising techniques, ultimately improving the quality and reliability of low-dose CT denoised images.

4.1 Dead Leaves

DL images are a type of synthetic dataset characterized by randomly overlapping shapes, typically circles, ellipses, or square, that mimic the complex occlusion patterns found in natural scenes. This model was initially proposed by Matherton in 1968 [11] and later expanded upon by various researchers. One of the key advantages of DL images is their ability to simulate the variability and complexity found in real-world images while maintaining a controlled and predictable structure.

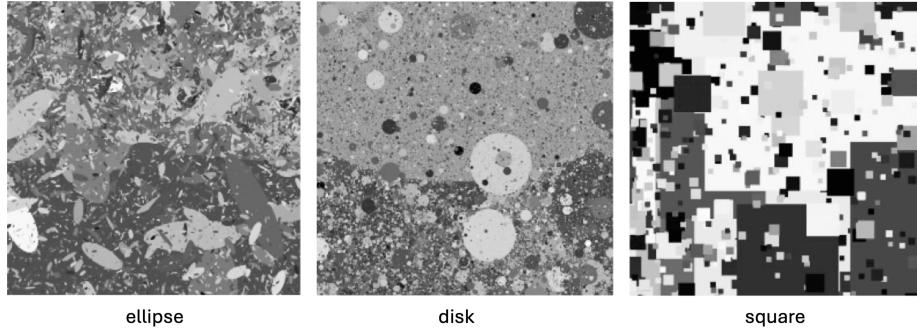


Fig. 4. Random example of DL using ellipse, disk and square object. [6]

4.2 Mimic Computed Tomography Distribution

Three statistical distributions were selected to quantify the statistical similarity between CT and its synthetic DL, i.e. color distribution, Fourier transformation distribution and directional gradient distribution. Then DL are computed to mimic as much as possible these distributions. During the generation process, DL images are initially created at three times their final size and then resized to the desired dimensions. This approach prevents over-definition in the images and results in a more natural appearance [1].

Morphological Analysis Synthetic image generation process involves selecting the appropriate object to construct the DL. This decision was informed by a morphological analysis, wherein various geometric shapes—specifically disk, ellipse, and square—were evaluated. To ensure comparability, each object type was adjusted to cover the same area. Morphological operations, namely opening (eq. 1) and closing (eq. 2), were performed on the shapes. An opening operation involves an erosion (ε) followed by a dilation (δ), while a closing operation reverses this sequence. The o in index stand for object used while applying the operator.

$$\gamma_o(X) = \delta_o(\varepsilon_o(X)) \quad (1)$$

$$\phi_o(X) = \varepsilon_o(\delta_o(X)) \quad (2)$$

To quantify which object produces the best image, the Structural Similarity Index Measure (SSIM) is used. SSIM values range from -1 to 1, where 1 indicates perfect similarity, 0 indicates no similarity, and -1 indicates perfect anti-correlation. SSIM between two images x and y is defined as:

$$\text{SSIM}(x, y) = \frac{(2\mu_x\mu_y + C_1)(2\sigma_{xy} + C_2)}{(\mu_x^2 + \mu_y^2 + C_1)(\sigma_x^2 + \sigma_y^2 + C_2)} \quad (3)$$

- μ_x : average of x
- μ_y : average of y
- σ_x^2 : variance of x
- σ_y^2 : variance of y
- σ_{xy} : covariance of x and y
- C_1 and C_2 : variables to stabilize the division with weak denominator

Subsequently, the object that yielded the highest similarity between the resultant image and real CT images was identified (Table 1). Following this analysis, the disk emerged as the most suitable object for generating synthetic CT images.

Table 1. Morphological Analysis

	SSIM Opening	SSIM Closing
Disk	0.83	0.89
Ellipse	0.81	0.88
Square	0.77	0.86

Color Distribution While generating DL images, each object needs to be associated with a color using a color distribution (Fig 5, CT line).

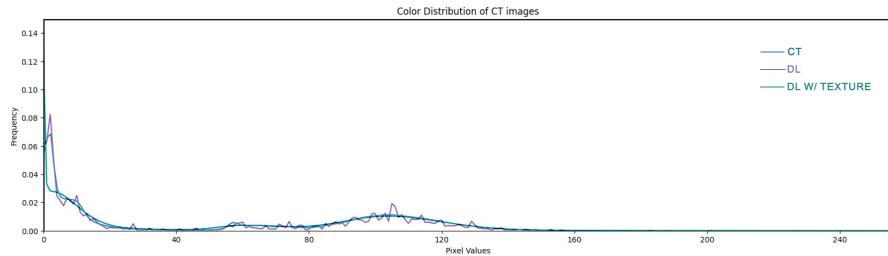


Fig. 5. Color distribution of full-dose CT, DL and DL w/ texture.

This process involves analyzing the grayscale intensity of all pixels and calculating the mean histogram across the entire full-dose CT dataset. Characterizing the color distribution in this manner ensures consistency and coherence in the synthetic image generation process. To address potential texture issues, a second synthetic dataset is generated by adding color noise as texture (Fig 6). The motivation for this texture choice is discussed in the next section.

Fourier Distribution Determining the size distribution of the disks is the next consideration. By leveraging the distinct distribution properties of DL, particularly the correlation between the size distribution and Fourier Transformation, a strategy was devised [6]. It was observed that using a power law (Eq. 4) for the disk size distribution directly influenced the slope of the log-log Fourier transform values at various distances from the image center.

$$P(\text{disk radius} = r) = r^\alpha \quad (4)$$

Aligning the alpha coefficient of the size distribution with the desired Fourier transform characteristics allows DL images with similar Fourier transforms to be generated. The coefficient α is determined to be approximately -1.66 for full-dose CT images (short visualization²). As mentioned in the previous section, uniform color can lead to texture issues. It has been qualitatively identified that a range between pink and brownian noise (Fig 6) can produce textures similar to those of full-dose CT images and exhibit interesting properties in the Fourier domain. Each type of color noise is associated with a specific Fourier distribution slope, and both pink and brownian noise have slopes similar to those of CT images.

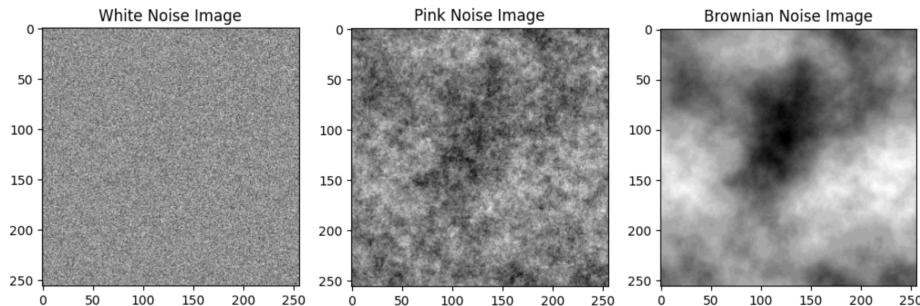


Fig. 6. Color noise example as texture.

Using the power law for the disk size distribution and setting the α value to match the slope values of the log-log Fourier distribution from full-dose CT

² <https://youtu.be/ke5Db2ElD-w>

images results in DL and DL textured images with log-log Fourier distribution slopes of -1.65 and -1.68 , respectively (Fig. 7).

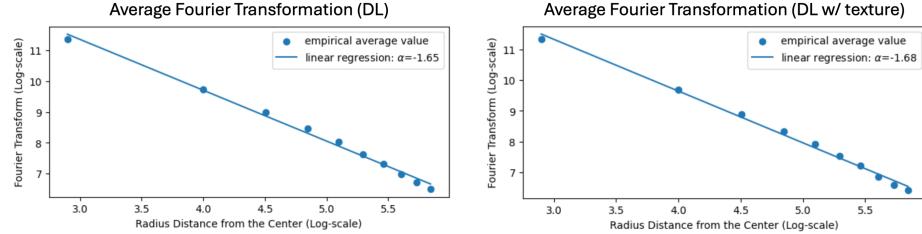


Fig. 7. Average Fourier distribution amplitude over distance from center.

Directional Gradient Distribution Finally, it is crucial to ensure that the directional gradient distribution of the DL images closely resembles that of the full-dose CT images. Objects of uniform color inadequately reproduce this distribution, as demonstrated in previous studies [10]. The DL images with color-noised texture are expected to better match the directional gradient distribution of full-dose CT images compared to uniformly colored DL images. In fact, the improvement is minor, nevertheless the use of textured DL images is retained to address texture issues during the denoising process.

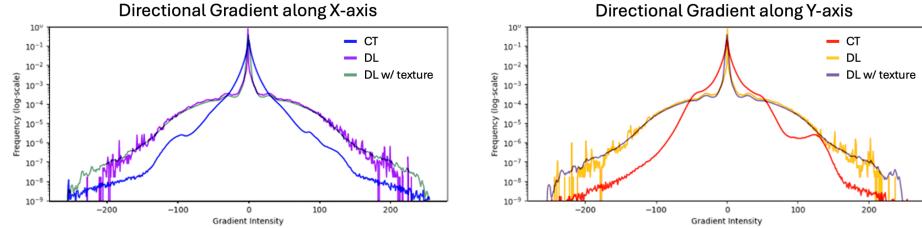


Fig. 8. Directional gradient of each dataset.

Dead Leaves As Synthetic Computed Tomography The generation process results in two distinct datasets: DL without texture and DL with texture, as illustrated in Figure 9.

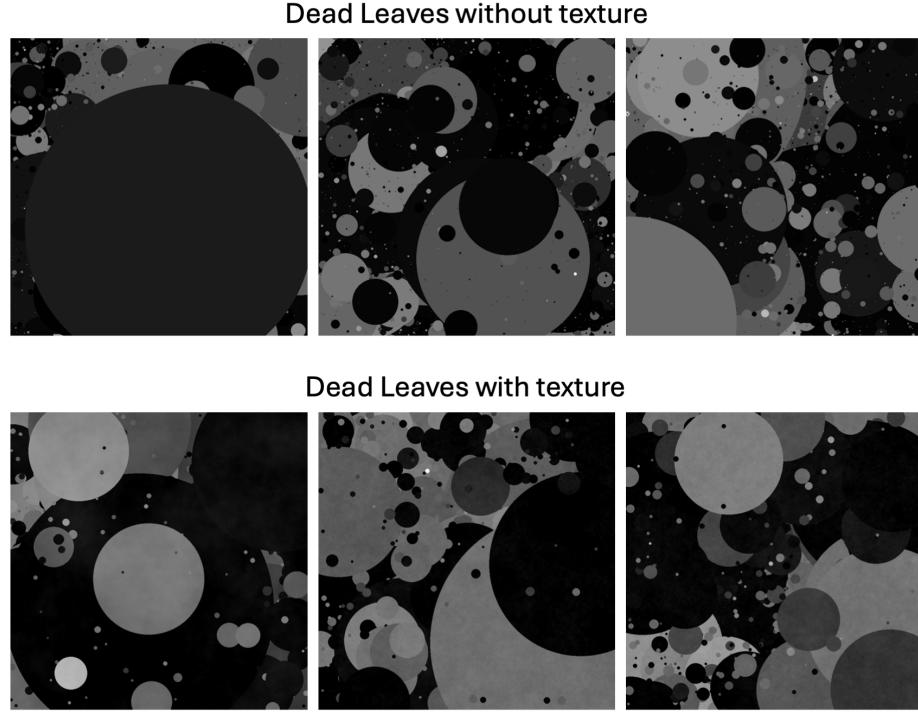


Fig. 9. DL generated from mimicking CT distributions.

5 Noise Analysis

With our synthetic dataset designed to mimic full-dose CT images, the next step involves adding the appropriate noise to train our denoising model effectively. Initially, based on several studies, we considered incorporating Poisson-Gaussian noise, which combines noise from Poisson and Gaussian distributions (eq. 5). To determine the parameters for these distributions, we examined the variance of pixel intensities. The intercept of the variance-intensity plot indicates the variance of the Gaussian distribution, while the slope represents the multiplicative factor of the lambda parameter in the Poisson distribution.

$$\text{Expected Noise} = \mathcal{P}(\lambda) + \mathcal{N}(\mu, \sigma^2) = \frac{\lambda^k e^{-\lambda}}{k!} + \frac{1}{\sigma \sqrt{2\pi}} e^{-\frac{1}{2} \left(\frac{x-\mu}{\sigma} \right)^2} \quad (5)$$

However, a preliminary analysis of the noise model in our CT dataset revealed that the noise characteristics were more complex than initially assumed (Fig. 10). It appeared that some correlations existed within the noise, necessitating a more in-depth analysis. Future work will involve a more detailed investigation of the noise characteristics in CT images. Other studies suggest that Poisson-Gaussian noise is present at the sinogram level of CT images. To address this, one could retrieve the sinogram using the Radon transformation, add the noise at this level, and then convert back to the CT image level.

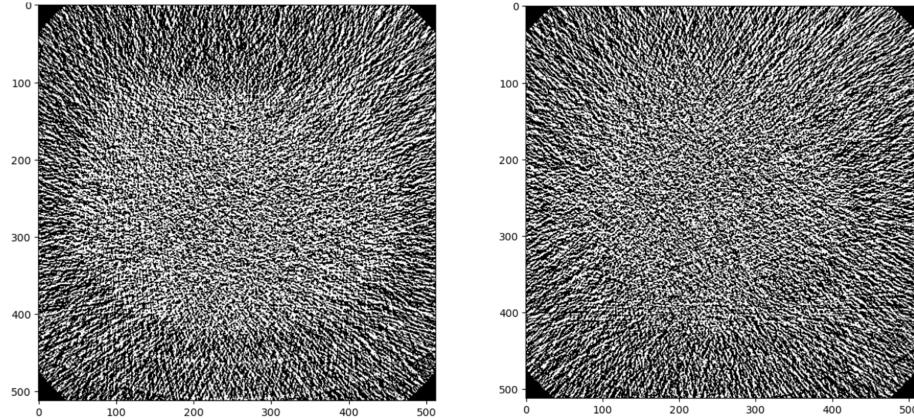


Fig. 10. CT observed noise example.

For the current study, given that our primary goal was to assess the suitability of Dead Leaves images for denoising tasks in CT, we opted to use simple Gaussian noise. This approach allows us to compare the performance of models trained on both CT and Dead Leaves images, providing a baseline for evaluating the effectiveness of synthetic datasets in denoising applications.

6 Denoising Model

The denoising model employed in this study is the basic DnCNN, with its architecture, counting 557'057 parameters, depicted in Figure 11. The DnCNN was selected for its well-documented effectiveness and its straightforward, yet powerful architecture, which strikes a balance between complexity and performance, making it an ideal choice for this research. Several hyperparameters were set during the training phase. All training sessions were performed under the following conditions: `batch size` set to 15, `validation set` set to 20%, `number of epochs` set to 50, `initial learning rate` set to 2×10^{-3} , `weight decay` set to 1×10^{-3} , `loss` set to `MSELoss()`, `optimizer` set to `Adam()`, and `learning rate scheduler` set to `CosineAnnealingLR()`.

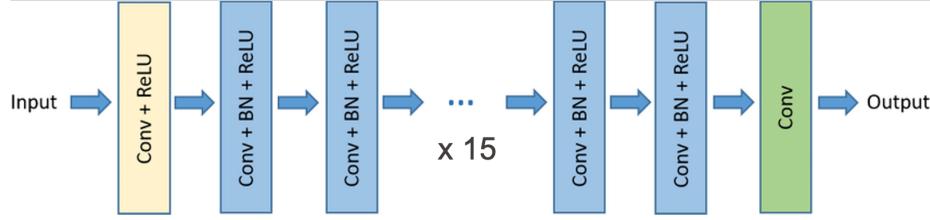


Fig. 11. DnCNN architecture used for denoising [2].

7 Experimental Results

During training, the models exhibited a decreasing training loss, as shown in the upper left graph of each subplot in Figure 12, indicating proper convergence. To gain more insight into the training evolution, additional metrics such as PSNR (bottom left) and SSIM (bottom right) were tracked over the epochs. All training processes demonstrated good behavior, with both supplementary metrics showing an upward trend. The transparent blue and orange zones represent the variance of train and validation set. Additionally, the learning rate evolution is depicted in the top right of each subplot, illustrating the adjustments made throughout the training.

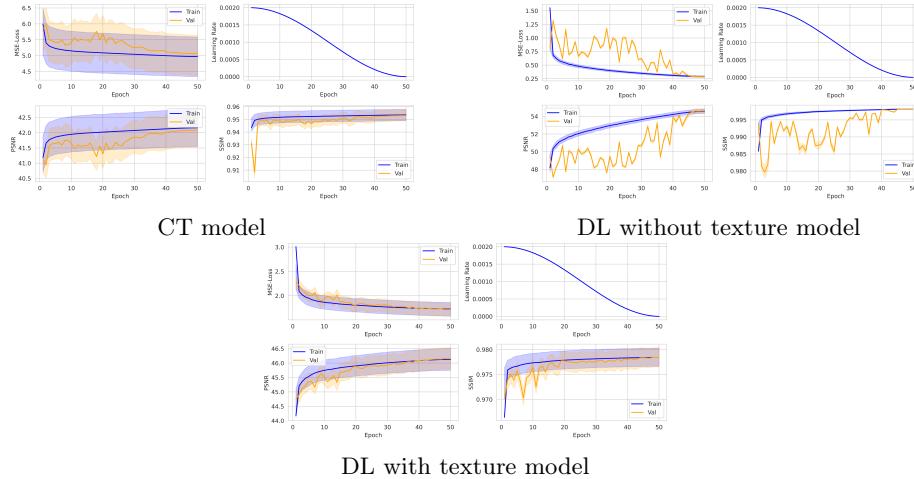


Fig. 12. Training evolution for each model.

The performance metrics, displayed in Figure 13, were computed by comparing the denoised images to the ground truth images. These metrics were obtained using a test set composed of full-dose CT images that had not been seen by the models during training. Gaussian noise was added to the test images, and the

models attempted to denoise them. The figures show the results of each metric for every image in the test set as blurry circles, with the mean represented by a vertical line and the 95% confidence interval of the mean shown as horizontal lines.

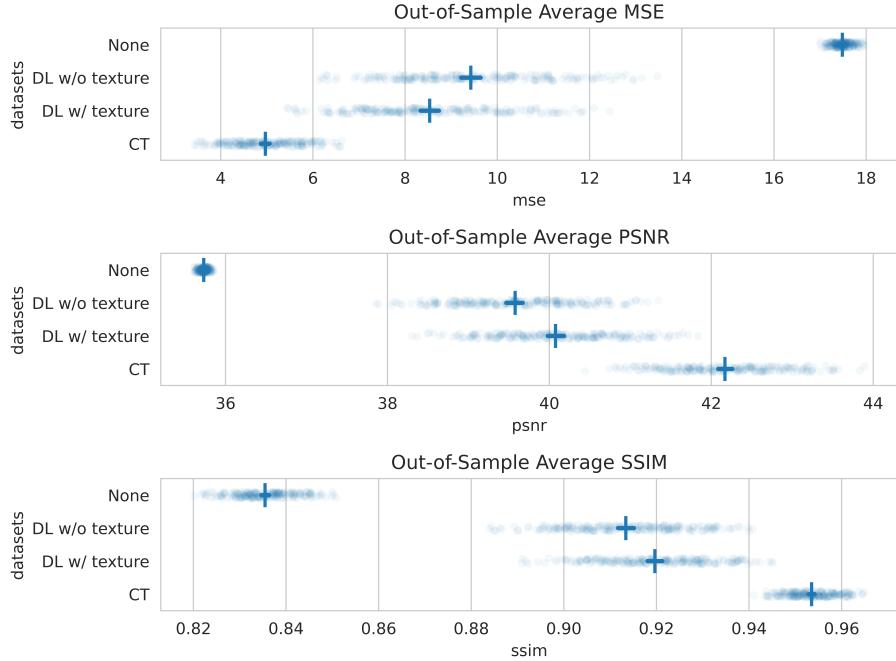


Fig. 13. Models performance comparison.

Currently, the model trained on real CT images demonstrates slightly better performance according to the numerical metrics. However, a closer examination of the denoised images reveals important insights into this performance disparity. The images denoised by the model trained on synthetic DL data exhibit more predictable artifacts, suggesting that while the numerical metrics favor the real CT-trained model, the visual quality of the denoised images from the synthetic dataset-trained model is comparable in certain aspects (Figure 14). The most significant finding is the superior performance of the model using DL with texture. The results indicate that the PSNR of our enhanced synthetic model is statistically better at a 95% confidence interval. This clearly shows that the current approach is promising and can be improved by continually adding better textures to our DL images.

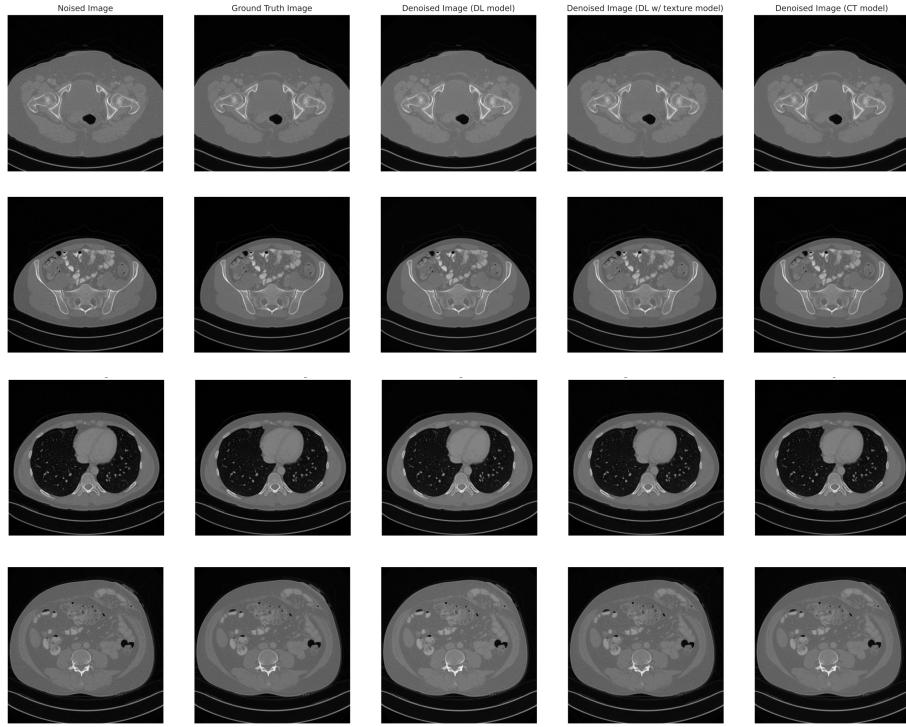


Fig. 14. Models denoising results³.

Finally, we examined the behavior of our metrics when reducing the CT dataset size and when mixing both CT and DL with texture datasets. Figure 15 illustrates these two aspects concurrently. The blue line represents the performance evolution of each metric as the CT dataset size increases, while the orange line indicates the performance when increasing the proportion of CT images in a mixed CT-DL with texture dataset. For example, a CT ratio of 1% means that only 133 CT images populate the training dataset for the blue line, and 133 CT images are augmented by 13,167 DL with texture images for the orange line. Once again, the current synthetic dataset, DL with texture, does not seem to improve the results significantly. This suggests that the primary advantage of the synthetic dataset lies in training a model with a fully synthetic dataset. While this approach may result in slightly lower denoising performance, it allows the model to produce more predictable artifacts, enhancing the overall reliability and interpretability of the denoised images.

³ more denoised images can be found in my GitHub repository.

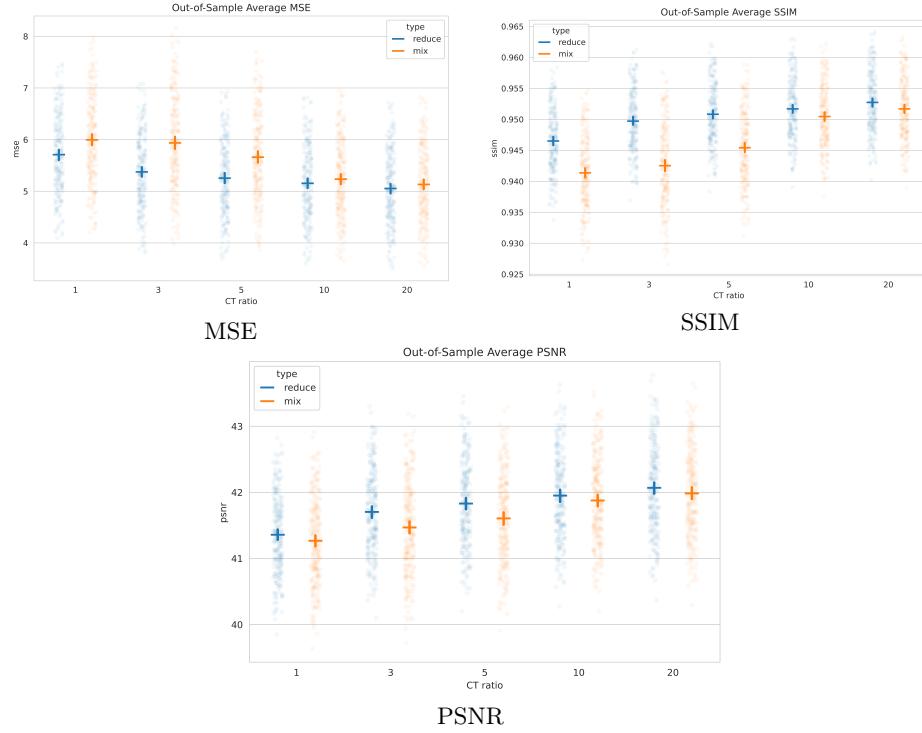


Fig. 15. Models performance evolution.

8 Conclusion And Future Works

This study shows the feasibility and effectiveness of using synthetic DL datasets for training denoising models for low-dose CT images. Our DnCNN model trained with synthetic datasets with statistical metrics mimicking the distribution characteristics of real CT images, shows promising results in reducing noise and preserving image quality. The experimental results indicate that models trained on DL datasets, particularly those with texture, can achieve performance metrics nearly comparable to models trained on real CT datasets. This suggests that synthetic datasets can serve as a viable alternative, addressing challenges related to data availability and unpredictable artifact introduction. Future work will focus on refining the texture and noise characteristics of the synthetic datasets to further improve model performance and generalizability. This research paves the way for more accessible and efficient methods for low-dose CT image denoising, potentially improving diagnostic accuracy and patient safety.

References

1. Achddou, R., Gousseau, Y., Ladjal, S.: Synthetic images as a regularity prior for image restoration neural networks. In: Eighth International Conference on Scale Space and Variational Methods in Computer Vision (SSVM). Cabourg (virtuel), France (May 2021), <https://hal.science/hal-03186499>
2. Anwar, S.: Data-Driven Image Restoration. Ph.D. thesis (01 2018)
3. Bhadra, S., Kelkar, V.A., Brooks, F.J., Anastasio, M.A.: On hallucinations in tomographic image reconstruction (2021)
4. Cao, C.F., Ma, K.L., Shan, H., et al.: Ct scans and cancer risks: A systematic review and dose-response meta-analysis. *BMC Cancer* **22**(1), 1238 (Nov 2022). <https://doi.org/10.1186/s12885-022-10310-2>
5. Flohr, T.G., Stierstorfer, K., Ulzheimer, S., Bruder, H., Primak, A.N., McCollough, C.H.: Image reconstruction and image quality evaluation for a 64-slice ct scanner with z-flying focal spot. *Medical Physics* **32**(8), 2536–2547 (Aug 2005). <https://doi.org/10.1118/1.1949787>
6. Lee, A.B., Mumford, D.B., Huang, J.: Occlusion models for natural images: A statistical study of a scale-invariant dead leaves model. *International Journal of Computer Vision* **41**(1-2), 35–59 (2001)
7. Li, Z., Zhou, S., Huang, J., Yu, L., Jin, M.: Investigation of low-dose ct image denoising using unpaired deep learning methods. *IEEE Transactions on Radiation and Plasma Medical Sciences* **5**(2), 224–234 (2021). <https://doi.org/10.1109/TRPMS.2020.3007583>
8. Liang, Q., Cassayre, F., Owsianko, H., Helou, M.E., Süsstrunk, S.: Image denoising with control over deep network hallucination (2022)
9. Ma, Y., Wei, B., Feng, P., He, P., Guo, X., Wang, G.: Low-dose ct image denoising using a generative adversarial network with a hybrid loss function for noise learning. *IEEE Access* **8**, 67519–67529 (2020). <https://doi.org/10.1109/ACCESS.2020.2986388>
10. Madhusudana, P., Lee, S.J., Sheikh, H.: Revisiting dead leaves model: Training with synthetic data. *IEEE Signal Processing Letters* **PP**, 1–1 (12 2021). <https://doi.org/10.1109/LSP.2021.3132289>
11. Matheron, G.: Modele séquentiel de partition aléatoire. Tech. rep., Technical report, CMM (1968)