computers

MDPI

Article

Optimizing Natural Image Quality Evaluators for Quality Measurement in CT Scan Denoising

Rudy Gunawan [1,*], Yvonne Tran [2], Jinchuan Zheng [1], Hung Nguyen [1] and Rifai Chai [1,*]

[1] School of Science, Computing and Engineering Technologies, Swinburne University of Technology, Melbourne, VIC 3122, Australia; jzheng@swin.edu.au (J.Z.); hungnguyen@swin.edu.au (H.N.)
[2] Macquarie University Hearing (MU Hearing), Centre for Healthcare Resilience and Implementation Science, Macquarie University, Macquarie Park, Sydney, NSW 2109, Australia; yvonne.tran@mq.edu.au
* Correspondence: rgunawan@swin.edu.au (R.G.); rchai@swin.edu.au (R.C.)

Abstract: Evaluating the results of image denoising algorithms in Computed Tomography (CT) scans typically involves several key metrics to assess noise reduction while preserving essential details. Full Reference (FR) quality evaluators are popular for evaluating image quality in denoising CT scans. There is limited information about using Blind/No Reference (NR) quality evaluators in the medical image area. This paper shows the previously utilized Natural Image Quality Evaluator (NIQE) in CT scans; this NIQE is commonly used as a photolike image evaluator and provides an extensive assessment of the optimum NIQE setting. The result was obtained using the library of good images. Most are also part of the Convolutional Neural Network (CNN) training dataset against the testing dataset, and a new dataset shows an optimum patch size and contrast levels suitable for the task. This evidence indicates a possibility of using the NIQE as a new option in evaluating denoised quality to find improvement or compare the quality between CNN models.

Keywords: CT scan; neural network; denoising; Blind evaluator; reference less evaluator; NIQE; NIQE optimization

Academic Editor: Hersh Sagreiya Sagreiya

Received: 22 November 2024
Revised: 23 December 2024
Accepted: 3 January 2025
Published: 7 January 2025

Citation: Gunawan, R.; Tran, Y.; Zheng, J.; Nguyen, H.; Chai, R. Optimizing Natural Image Quality Evaluators for Quality Measurement in CT Scan Denoising. *Computers* **2025**, *14*, 18. https://doi.org/10.3390/computers14010018

1. Introduction

Cancer screening with low-dose CT scans (LDCT) aims to detect cancer or precancerous conditions before symptoms develop. The radiation used in these screenings is relatively low (approximately 10%) compared to the higher doses used in diagnostic imaging or cancer treatment. Early cancer detection, before it advances or becomes symptomatic, often leads to more effective treatment, improving outcomes and survival rates [1]. However, the reduced radiation dose can result in decreased detail and clarity of images. This is due to increased "noise", which appears as graininess or speckles because fewer photons are detected. This noise reduces the image's precision and signal strength, making accurate interpretation more difficult. One study showed that only 24.2% of cancers were correctly identified out of 53,454 patients, partly due to this noise issue [2].

Several strategies and technologies are used to manage and reduce noise artifacts. These include optimizing CT scanner parameters, using an advanced reconstruction algorithm, and applying image processing techniques. Within image processing, the denoising efforts range from traditional methods like noise reduction filters [3] and smoothing [4] to an advanced technique involving Convolutional Neural Networks (CNNs). The effectiveness of these denoising methods is assessed through quality evaluations that typically use FR image evaluators, such as the mean square error (MSE), which measures the average squared differences between the denoised and clean images; a lower score indicates

a better denoising result. The peak signal-to-noise ratio (PSNR) evaluates the ratio of the maximum pixel value to the power of corrupting noise (MSE), with a higher score indicating a better result. The Structural Similarity Index Measure (SSIM) assesses three image characteristics—brightness, contrast, and structure—between the noisy and original images, with improved scores reflecting better denoising results. These evaluations help determine whether new denoising techniques offer improvements over previous methods.

A Blind/No Reference evaluator was scarcely used within denoising CT scans, and within the few, there was a Natural Image Quality Evaluator (NIQE) [5,6]. The NIQE was introduced in 2013 for photolike images [7]; several NIQE variants followed, such as the Integrated Local NIQE [8] and the Multi-Orient NIQE [8]. NIQE application includes photolike images [7,9], stereoscopic images [8], remote sensing/radar/sonar images [10–12], and 3D point cloud images [13]. Unfortunately, there was no detailed discussion about the use of a NIQE that can draw a greater medical image community. It was briefly mentioned in tomosynthesis denoising [14], X-ray segmentation [15], herringbone artifacts removal in MRI [16], and ultrasound scanner [17]. Because the research community has limited information, this research presents a complete overview of NIQE usage. This paper's main contribution starts with directly comparing FR evaluators and finding the best patch size and contrast level using qualitative and quantitative analysis. It also covers finding the effect when the target images are not part of the model library. Ultimately, it can encourage more extensive use of a NIQE in CT scans.

This paper is structured as follows: Section 2 details the methodology, data usage, evaluator description, and comparison indicators. Section 3 presents the results of the relation finding about optimizing NIQE parameters based on the provided indicators. Section 4 observes the common practice approach, the relation of failed improvement perception against PSNR score, rectangular patch, and testing on a new dataset. Finally, Section 5 provides the conclusion.

2. Materials and Methods

2.1. The Dataset

The image data used for denoising assessment are sourced from the Cancer Imaging Archive (TCIA) [18], under the research of the LDCT-and-Projection-data. This dataset includes two types: the original CT image and the image with simulated noise. The original CT images were the scan results of the Somatom Definition AS+ (Siemens Healthcare, Erlangen, Germany) and Somatom Definition Flash CT scanner (Siemens AG, Muenchen, Germany) from a patient with and without solid, non-calcified nodules. A noise was added using a noise insertion tool to the projection data based on the noise model. The reconstruction utilized the scanner feature, and based on the assessment, the simulated noise produced images that closely resembled LDCT scans, with a deviation of about 6% [19].

There are two groups of datasets; the first one contains ten patients, amounting to 3300 image data, split into training, validation, and testing, each with 1650, 990, and 660 images. The NIQE library uses 3300 images from standard scans. The second one contains two patients for revalidation with 595 image data.

2.2. Denoising CNN

There are many approaches to address the image noise in CT scans; the earlier approach was a classical denoising such as Gaussian Smoothing, Non-Local Means (NLM), which uses the weighing principle on pixel similarity to the target [20], and 3D Block Matching (BM3D) on the weighing principle with an additional step on the Wiener filter [21]. CNN appears later as an advanced denoising option replicating human neuron connection

to recognize and minimize the noise from CT Scan images. The CNN evolves from a simple neural model into a deep layered model utilizing convolution processes [22].

Like all the neural processes, the CNN requires training before denoising by inputting a noisy image into the model and introducing the 'good' image as a target. Because of these training steps, an image pair with normal and low radiation doses is needed to represent both normal and screening CT scan images. This pair of images usually comes from a simulation in which the noise model is inserted into a normal CT scan image.

There are many variants of CNN based on the convolution types and the layer depth of the network. This research used four CNN models (see Figure 1) to train using the above dataset. The first model is derived from U-Net and was initially used for classification or segmentation [23]. The slightly modified U-Net retains convolutional, max pooling, and transpose convolution layers with Rectified Linear Unit (ReLU) activation and skips connection (concatenation layers); the only difference is the use of regression layer to facilitate the denoising/image regression process. The second CNN is a Double U-Net which stacked two modified U-Nets; the modification involves the use of a convolution layer on stride two as the contraction layer, using transpose convolution layers for the whole expansion layer, adding skips connection between two U-Nets and using the regression layer at the end for denoising process [5].

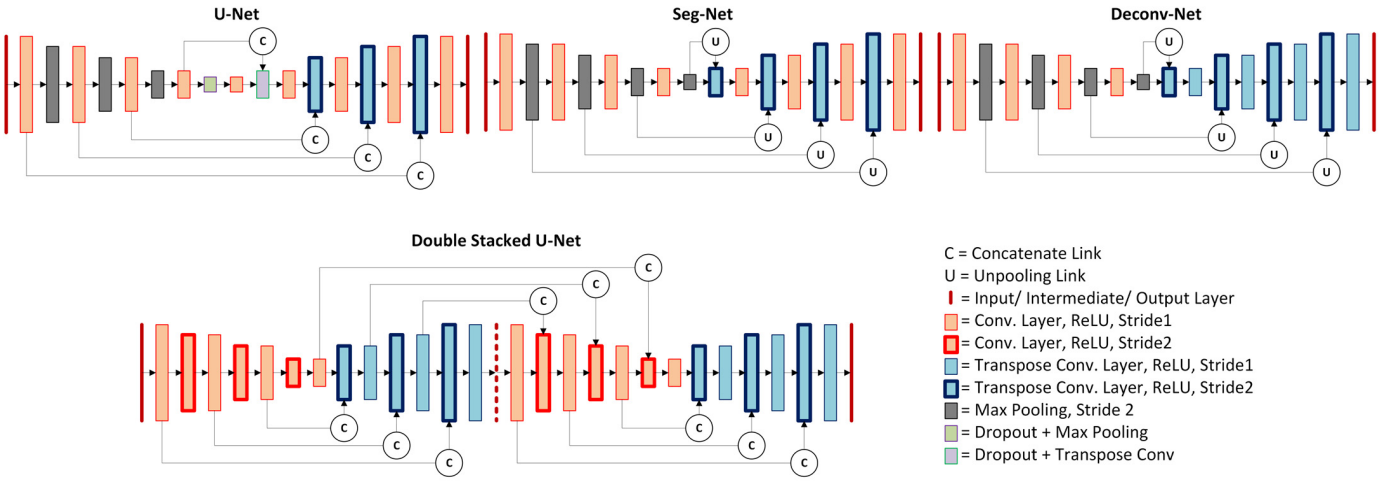

Figure 1. Four CNN models for denoising show layer type and connection variations.

The third and fourth CNNs are derived from the Segmentation Network (Seg-Net) [24] and Deconvolutional Network (DeConv-Net) [25]. These CNNs share similarities in the contraction layers, which use convolution and max pooling layers but differ in the expansion layers. The Seg-Net relies on transpose convolution stride two and convolution layers, while DeConv-Net relies on transpose convolution layers only. The last layer deviates from the original as it uses regression layers for denoising purposes. The assessment evaluates the difference in scoring before and after denoising using four quality evaluator techniques.

2.3. Full Reference Image Quality Evaluators

Several image quality evaluators utilize a Full Reference (FR) for measuring quality; the basic includes the mean square error (MSE), the peak signal-to-noise ratio (PSNR), and the Structural Similarity Index Measure (SSIM). There is an improved version of the basic evaluator, such as the Noise Quality Measure, which is a modified PSNR that adds an extra weighted frequency response of the Human Visual System (HVS) using the Low Pass Contrast Sensitivity Function (CSF) [26].

The Edge Strength Similarity-based Image quality Metric (ESSIM) focuses on structural component (edge) from SSIM calculated using the MSE formula [27], Saliency-Guided ESSIM (SG-ESSIM), which is a derivation of ESSIM and improved it by replacing component C in the equation with visual saliency pixel [28]. A Multiscale Similarity Index Measure (Multi SSIM) is derived from SSIM and uses a sliding window to calculate and set the SSIM score into an index; the final score comes from the index's average [29].

This paper focuses on the basic evaluators rather than the improved version since they were used in most denoising CT scans. The MSE is a standard metric that measures the average squared difference between values in two datasets. It is often used to evaluate the quality of reconstructed or compressed images compared to their original versions. In image processing, the MSE quantifies the average squared difference between the pixel values of the original image and those of the distorted or compressed image [30,31].

$$MSE = \frac{1}{N}\sum_{i=1}^{N}(r_i - t_i)^2 \ , \tag{1}$$

where N is the total number of pixels in the image, the pixel value at position i in the original/reference image r, and the noisy/target image t (1). The MSE measures how much the pixel values in the distorted image differ from those in the original image. Lower MSE values indicate better quality and less distortion, as the differences between the original and distorted images are smaller.

PSNR is a metric used to measure the quality of a reconstructed or compressed image compared to the original image. It is commonly used in image and video processing to evaluate the performance of compression algorithms and image restoration techniques. PSNR is the ratio between the maximum possible power of a signal (in this case, the pixel values of an image) and the power of the noise that affects the fidelity of its representation. It is calculated using the mean square error (MSE) between the original and distorted images [32–35].

$$PSNR = 20\log_{10}\left(f_{max}/\sqrt{MSE}\right) \ , \tag{2}$$

where f_{max} is the maximum possible pixel value of the image (e.g., 255 for an 8-bit image), and MSE is the mean square error between the original and distorted images (2). $PSNR$ is measured in decibels (dB), with higher values indicating better quality. Generally, a higher $PSNR$ means that the image has fewer distortions and is closer to the original.

SSIM is a metric used to measure the similarity between two images. It is commonly employed in image processing and computer vision to assess the quality of an image, often in the context of image compression or restoration. Unlike traditional metrics that might compare pixel-by-pixel differences, SSIM evaluates the perceived quality of an image by considering changes in structural information, luminance, and contrast [33,36].

$$SSIM_{(t,r)} = \left[\mathrm{s}_{(t,r)}\right]^{\alpha} + \left[\mathrm{l}_{(t,r)}\right]^{\beta} + \left[\mathrm{c}_{(t,r)}\right]^{\gamma} \ , \tag{3}$$

where (α, β, γ) is a weighted combination of three comparison measurements: $s_{(t,r)}$ the structural information, $l_{(t,r)}$ the luminance, and $c_{(t,r)}$ the contrast. Comparing t the target sample, and r the reference sample. The weight components are commonly set to 1 (3) for simplicity. It is designed to be more aligned with human visual perception, meaning it is better at identifying quality degradations that are noticeable to people. The SSIM score ranges from -1 to 1, where 1 indicates a perfect match between the two images, and lower values represent increasing levels of dissimilarity.

2.4. Blind/No Reference Evaluators

While the denoising result can use an FR evaluator for evaluation improvement due to the presence of an image pair, it becomes impossible to do so in the actual denoising process, which targets screening images. The evaluating process needs a Blind/No Reference (NR) evaluator. Several NR evaluators have various methods, such as Blind/Referenceless Image Spatial Quality Evaluator (BRISQUE), in which the model is trained on the database/library with a known distortion. The BRISQUE extracts three sets of features based on the statistics of natural images, distortion textures, and blur/noise; three regression models are trained for each feature set, and finally, a weighted combination of them is used to estimate the image quality [37].

The Natural Image Quality Evaluator (NIQE), in which the model is trained on the database/library of pristine images (in CT scan—a good image). The NIQE method extracts local features from an image and then fits the feature vectors to a multivariate Gaussian (MVG) model [7]. A Perception-based Image Quality Evaluator (PIQE) uses block-wise distortion estimation [38]. The discrete cosine transform (DCT) and self-organizing map (SOM) clustering use a neural network to train image patches (five vector values of the distorted patches per image) into a single score [39]. There are derivative methods, such as a Feature-enriched NIQE, which utilize the feature vector of each patch; thus, it has several local scores [9]. While there are many variants of NR evaluator, this paper chooses to use NIQE because of its simplicity and the 'good' images dependent on the model. Furthermore, a few other researchers on CT scan images have started using it without highlighting the setting.

2.5. Natural Image Quality Evaluator (NIQE)

The Natural Image Quality Evaluator (NIQE) measures image quality based on its perceived naturalness to human observers. Unlike conventional quality metrics that depend on reference images for comparison (such as PSNR or SSIM), NIQE assesses quality without needing an ideal reference. Instead, it uses statistical models derived from the characteristics of natural images (see Figure 2).

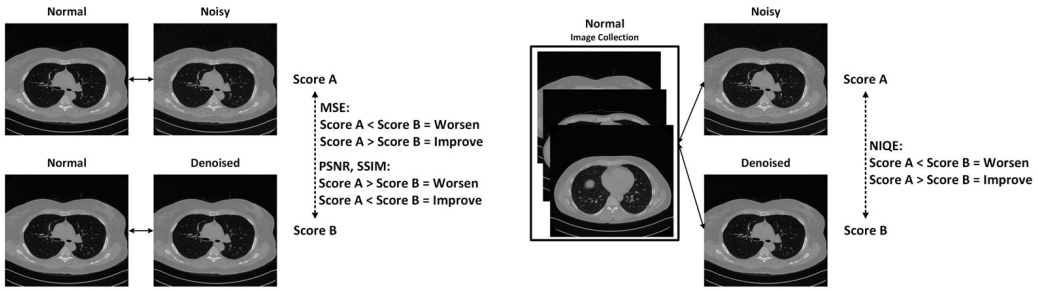

Figure 2. The scoring difference between improved and worsened (deteriorated) quality was between the MSE, PSNR, SSIM, and NIQE.

It uses Natural Scene Statistic (NSS) obtainable from Generalized Gaussian Distribution (GGD) and Asynchronous GGD (AGGD) fitting of the Mean-Subtracted Contrast Normalized (MSCN) image collection [40]. The statistical reference is built from a large dataset of natural images, and the properties of the target image are compared against this reference to determine how unnatural or distorted the image is. An image that closely matches the statistical properties of natural images has a lower NIQE score, indicating higher perceived quality and naturalness.

In the context of CT scans, images taken at standard radiation doses with minimal noise can be used to create a dataset for extracting these statistical properties. Both noisy and denoised CT images can then be evaluated using NIQE to assess the effectiveness of

noise reduction techniques. Image patches were extracted and selected based on the NSS coefficient and the variance data related to the sharpness information (4).

$$\delta_b = \sum\sum_{(i,j)\in patch_b} \sigma_{(i,j)} \ , \tag{4}$$

with b as the patch index, $\sigma_{(i,j)}$ is the variance data of spatial indices i and j of the image, and δ is the local patch sharpness. The patches with sharpness $\delta > \tau$ are selected in the patch pool for the scoring calculation.

$$NIQE = \sqrt{(v_1 - v_2)^\tau (\Sigma_1 + \Sigma_2/2)^{-1} (v_1 - v_2)} \ , \tag{5}$$

where v_1 and Σ_1 are the vector and covariance of the NIQE model, while v_2 and Σ_2 are the vector and covariance of the target image, with the sharpness threshold of the patch τ (5). The image's properties, such as size and intensity level, can affect the scoring due to patch selection, NSS scoring, and the image collections' sharpness threshold.

2.6. Performance Indicators

This paper uses three indicators to determine the optimum NIQE setting for CT scan denoising: the average scoring assessment, the quantity analysis of perceived improvement, and the correct scoring rank quantity. These three indicators are being compared to the Full Reference image evaluator: MSE, PSNR, and SSIM. The average scoring assessment is a common method to find the average quality score across all testing images and make the comparison to find the best denoising method. The quantity of perceived improvement means finding several images that show an improved score after denoising [41]; this approach differs from the average scoring of testing samples used by most denoising research [22,42–44]. The correct scoring rank quantity relies on the individual score comparison between the denoising method; the number of images that correctly find the best denoising method is counted.

The scoring ranks depend on the quality of the denoised image, which comes from CNN models' denoising ability. The CNN denoising ability comes mainly from the type of layers, the layer arrangement, and the number of layers [40]. The layer type was defined earlier in Section 2.3, and the number of layers of four CNN models is as follows: Seg-Net and DeConv-Net (43 layers), U-Net (57 layers), and Double U-Net (95 layers). The trend of going deeper into CNN is due to its improved ability for denoising; with the correct arrangement, the risk of vanishing gradient can be minimized [45]. Based on the depth alone, DU-Net with 95 layers should excel in denoising quality. Other research has proven the quality lead by 0.54 points on PSNR to U-Net [5].

The quantity of perceived improvement has two categories of denoising results: improvement (*Im*) and deterioration (*De*); each of the image quality evaluators has a different scoring on *Im* (6) and *De* (7). The following formula can indicate the denoising result.

$$Im = S_A - S_B \begin{cases} Im > 0 : PSNR,\ SSIM \\ Im < 0 : MSE,\ NIQE \end{cases} \ , \tag{6}$$

$$De = S_A - S_B \begin{cases} De < 0 : PSNR,\ SSIM \\ De > 0 : MSE,\ NIQE \end{cases} \ , \tag{7}$$

where S_A is the noisy score, S_B is the denoised score, *Im* is the improvement score indicator, and *De* is the deterioration score indicator. The condition of improvement or deterioration for each evaluator is shown in both formulas.

3. Results

3.1. CNN Trainings

The patient designator IDs used in this training are C002, C004, C012, C016, C027, C030, C050, C52, C67, and C77. The training was limited to 200 epochs with an initial learning rate of 0.0001, and the learning rate dropped by 0.8 every ten epochs. The training batch was set to 5, and the validation interval was set to 5 epochs to preserve memory usage. The training utilized an automatic weight collection based on the best validation loss from MATLAB 2023a. This paper does not intend to propose a new CNN model; it only used the existing model with minor adjustments (see Figure 1) and trained on the dataset.

The validation RMSE shown in Figure 3 indicates almost the exact figure between Seg-Net and DeConv-Net at 29,743.56 and 29,771.56. The U-Net performance is a bit better than the previous two at 29,222.38, which occurred at 140 epochs, and Double U-Net reached the peak of performance at 27,638.64. Besides the U-Net, the other three have a minimum RMSE at the epoch ends.

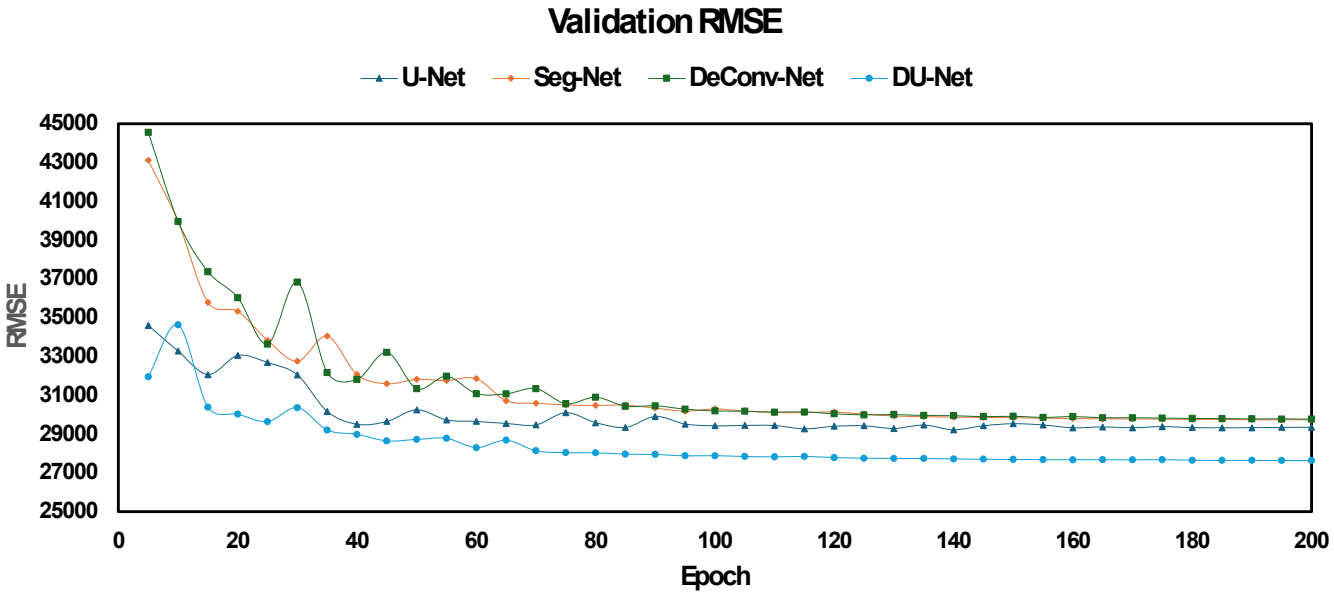

Figure 3. The validation RMSE between four CNN Models shows that DU-Net takes the lead with a large margin.

3.2. Improvement Scores of the Full Reference Evaluators

The first assessment uses three FR evaluators scoring to indicate the denoising improvement between four CNN models, as shown in Table 1. A greater value on PSNR and SSIM indicates a better denoising result; the Double U-Net holds higher scores than the other three at 10.59 and 0.0111. It has a 0.55 score difference in PSNR to the next contender U-Net (10.04), Seg-Net (9.62), DeConv-Net (9.55), and a 0.0003 difference in SSIM with the other three. On the other hand, the MSE provides better denoising when the score difference is negative; the greater negative offers a better quality, which also seems to occur on the Double U-Net with a −37,217 score. U-Net follows next with −36,869, then SegNet (−36,536) and DeConv-Net (−36,477). This result confirms that the Double U-Net reaches Rank1 compared to the other three.

Table 1. The average improvement scores on denoising CNNs using three evaluators. This approach is commonly used in much research as evaluation metrics. Green and orange fonts indicate the first and second-best ranks between CNN models.

CNN	Average Improvement Scores		
	MSE	PSNR	SSIM
U-Net	−36,869.51	10.04	0.0108
Seg-Net	−36,536.65	9.62	0.0108
DeConv-Net	−36,477.61	9.55	0.0108
DU-Net	−37,217.25	10.59	0.0111

These data contain another detail: 660 images under the test indicate improvement, albeit at different levels. Therefore, these FR evaluators can provide high confidence (quantity-wise) in evaluating image improvement.

3.3. Average Improvement Scores on NIQE

In the NIQE evaluator, the assessment was performed on seven patch sizes and in combination with four contrast levels using the same average scoring as the above FR evaluators (there are a total of twenty-eight combinations). As discussed earlier, Double U-Net assumed the highest denoising CNN model; it is imperative to find the same scoring that puts the Double U-Net at the top (in the NIQE case, the average scoring of Double U-Net should be the lowest values). Table 2 shows the green highlight on the patch–contrast combination that follows the trend, excluding any positive scores.

Table 2. The average NIQE scoring on denoising CNNs using different patch sizes and contrast. The green highlights indicate the combination in which DU-Net is the best scorer.

[Patch]−Contrast	Average NIQE Scoring			
	U-Net	Seg-Net	DeConv-Net	DU-Net
[8 × 8]−0.2	−1.44	−1.51	−1.53	−1.57
[8 × 8]−0.4	−2.15	−2.25	−2.26	−2.22
[8 × 8]−0.6	−2.67	−2.78	−2.76	−2.69
[8 × 8]−0.8	−6.03	−6.39	−6.14	−6.06
[16 × 16]−0.2	−0.90	−0.68	−0.74	−1.12
[16 × 16]−0.4	−1.89	−1.77	−1.81	−2.03
[16 × 16]−0.6	−3.68	−3.72	−3.74	−3.79
[16 × 16]−0.8	−8.17	−8.40	−8.32	−8.18
[32 × 32]−0.2	0.37	0.58	0.49	0.17
[32 × 32]−0.4	−0.61	−0.29	−0.39	−0.73
[32 × 32]−0.6	−2.97	−2.82	−2.90	−3.03
[32 × 32]−0.8	−11.99	−11.77	−11.98	−11.69
[64 × 64]−0.2	−0.69	−0.74	−0.77	−0.90
[64 × 64]−0.4	−2.34	−2.32	−2.34	−2.45
[64 × 64]−0.6	−9.30	−9.32	−9.42	−9.40
[64 × 64]−0.8	−41.42	−41.84	−42.04	−41.66
[128 × 128]−0.2	−20.91	−21.32	−21.32	−21.28
[128 × 128]−0.4	−30.26	−30.80	−30.82	−30.43
[128 × 128]−0.6	−57.32	−57.86	−57.90	−57.49
[128 × 128]−0.8	−115.38	−116.93	−117.10	−116.57

Table 2. *Cont.*

[Patch]−Contrast	Average NIQE Scoring			
	U-Net	**Seg-Net**	**DeConv-Net**	**DU-Net**
[16 × 32]−0.2	−0.23	0.02	−0.08	−0.54
[16 × 32]−0.4	−1.41	−1.22	−1.30	−1.66
[16 × 32]−0.6	−3.19	−3.10	−3.17	−3.32
[16 × 32]−0.8	−10.57	−11.18	−11.25	−10.94
[32 × 16]−0.2	0.37	0.58	0.49	0.17
[32 × 16]−0.4	−1.12	−0.83	−0.90	−1.18
[32 × 16]−0.6	−2.70	−2.38	−2.43	−2.66
[32 × 16]−0.8	−6.36	−6.41	−6.11	−6.34

For example, the patch [8 × 8] at contrast 0.2 has the lowest Double U-Net score (−1.57) of the other three, which means this combination meets the requirement. Meanwhile, [8 × 8] at contrast 0.4 has the lowest score on DeConv-Net (−2.26), which did not meet the requirement. Then there is [32 × 32] at contrast 0.2, while the Double U-Net has the smallest score, but the score is in the positive domain, causing it to fail to meet the requirement.

Twelve patch–contrast combinations from twenty-eight combinations meet the lowest average scores on Double U-Net. Patches [8 × 8] and [32 × 16] have one contrast level that meets the demand, patches [32 × 32] and [64 × 64] have two contrast levels, and patches [16 × 16] and [16 × 32] have three contrast levels.

3.4. Quantitative Analysis of Improvement Scores on NIQE

While the average scoring can narrow the patch–contrast combination to a selected few, further quantitative analysis is required to pinpoint the optimum combination. The quantitative approach uses the number of testing images that fall under the category improved because, unlike the FR evaluator, which has a one-by-one comparator, the NIQE uses a statistical figure from a collection of good images. The FR evaluators provide 660 improved images out of 660 testing images; Table 3 indicates the number of improved images from NIQE viewpoints that can lead to the optimal NIQE setting.

A lower image count can lower the confidence level of the evaluator assessment. Thus, the higher, the better. While there is no formula for the proper minimum count, this research took 80% of the total 660 images as the minimum (528 images). It gives a new lead on selecting a better patch–contrast combination, and for a stringent implementation, all CNN models need to have a minimum of 528 images.

From Table 3, four patch–contrast combinations meet the requirements (highlighted in green); patches [8 × 8] and [16 × 32] contribute one contrast level, and patch [16 × 16] contributes two contrast levels. Most of the remaining contenders have several improved images below 70%, and only one has a close call with only DU-Net above 80%.

An assessment of the scoring improvement based on the image quantity can narrow the selection. This method utilizes the scoring comparison between CNN models on every image and quantifies the number of images with the highest score on Double U-Net. Table 4 shows that the patch [16 × 16] with a 0.4 contrast level came up to the top with 534 images (80.9%)—highlighted in green. The other three contenders have the number of images with DU-Net on top below 60%.

Table 3. The number of improved images from NIQE viewpoints between the four CNNs and the twelve combinations. The green highlights indicate the combination in which the number of improved images (perceived by NIQE) is above 80% of the total testing images.

[Patch]−Contrast	Number of Improved Testing Images			
	U-Net	Seg-Net	DeConv-Net	DU-Net
[8 × 8]−0.2	651	651	651	652
[16 × 16]−0.2	450	426	446	489
[16 × 16]−0.4	626	625	632	635
[16 × 16]−0.6	648	653	645	645
[32 × 32]−0.4	362	324	348	391
[32 × 32]−0.6	447	445	452	467
[64 × 64]−0.2	381	396	401	411
[64 × 64]−0.4	422	436	436	439
[16 × 32]−0.2	359	314	330	387
[16 × 32]−0.4	511	504	523	567
[16 × 32]−0.6	571	568	579	591
[32 × 16]−0.4	461	439	434	471

Table 4. The number of images in which the DU-Net scores best between the last four combinations. The green highlight indicates the highest number of images in which DU-Net is the best scorer.

[Patch]−Contrast	Number of Testing Images−Best Score on DU-Net
[8 × 8]−0.2	384
[16 × 16]−0.4	526
[16 × 16]−0.6	339
[16 × 32]−0.6	376

3.5. Detailed Quantitative Observation on [16 × 16] Patch

The selected combination falls on the [16 × 16] patch with a contrast level of a factor of 0.2. While the contrast of 0.4 has been chosen, there is a gap in the contrast level implementation. Further quantitative observation is required to understand the relationship between the number of improved images and the number of images with the best score on DU-Net (see Tables 3 and 4) across all contrast levels with a smaller factor of 0.02. The assessment covered the contrast level between 0.2 and 0.8, even though the contrast of 0.8 did not meet the requirement (see Table 2).

Rather than using the number of images, the percentage of improved images and the number of images with the best scores on DU-Net represent a percentage of the total images. Figure 4 shows the percentage of enhanced images on four CNN models, and they show almost similar trends; the data indicate a rapid ascend in percentage from a contrast level of 0.2 to 0.34, then a steady increase until 0.74 before changing to a gradual descent—a slight percentage variation between CNN models in the contrast level above 0.36.

Figure 5 shows the number of images in which the NIQE score on DU-Net is better than the other three in percentage. The figure in each contrast level excludes any NIQE positive score (NIQE positive score indicates the denoised image comes out worse than the noisy image). The chart has an ascending trend from contrast 0.2 to 0.32 (71.2% to 82%), then starting to descend to 30.5% (≈201 images) at contrast 0.8. The chart reaches its peak at a contrast level of 0.32 with 82% (≈541 images), which indicates this contrast is much better than the previous 0.4 with 79.7% (≈526 images). The contrast levels 0.28, 0.3, 0.34, and 0.36 follow closely.

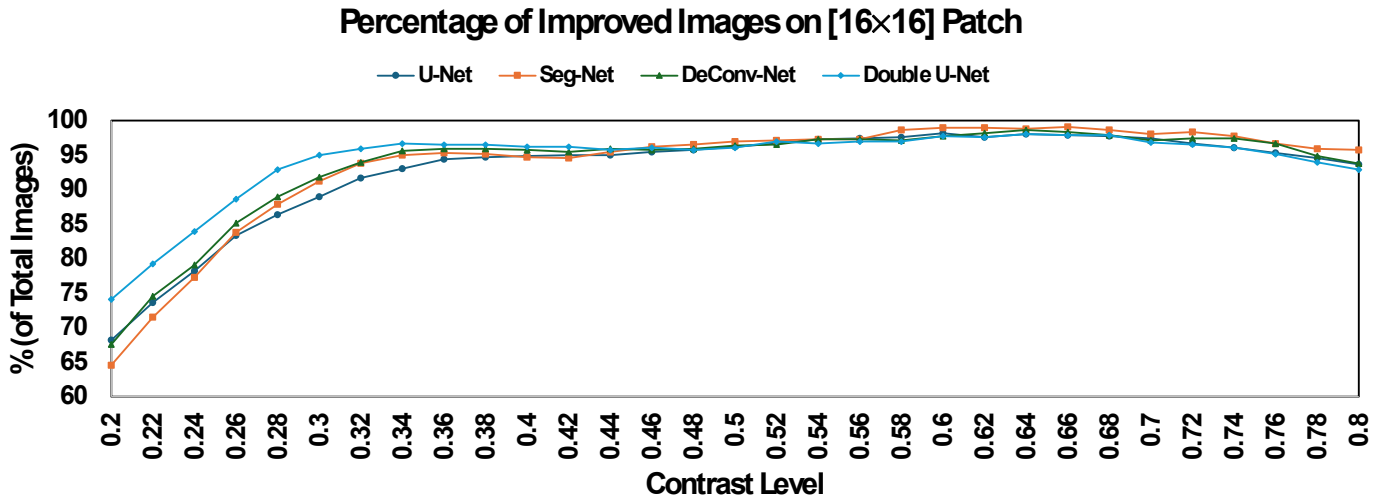

Figure 4. The percentage of improved images on [16 × 16] patch between four CNNs and different contrast levels.

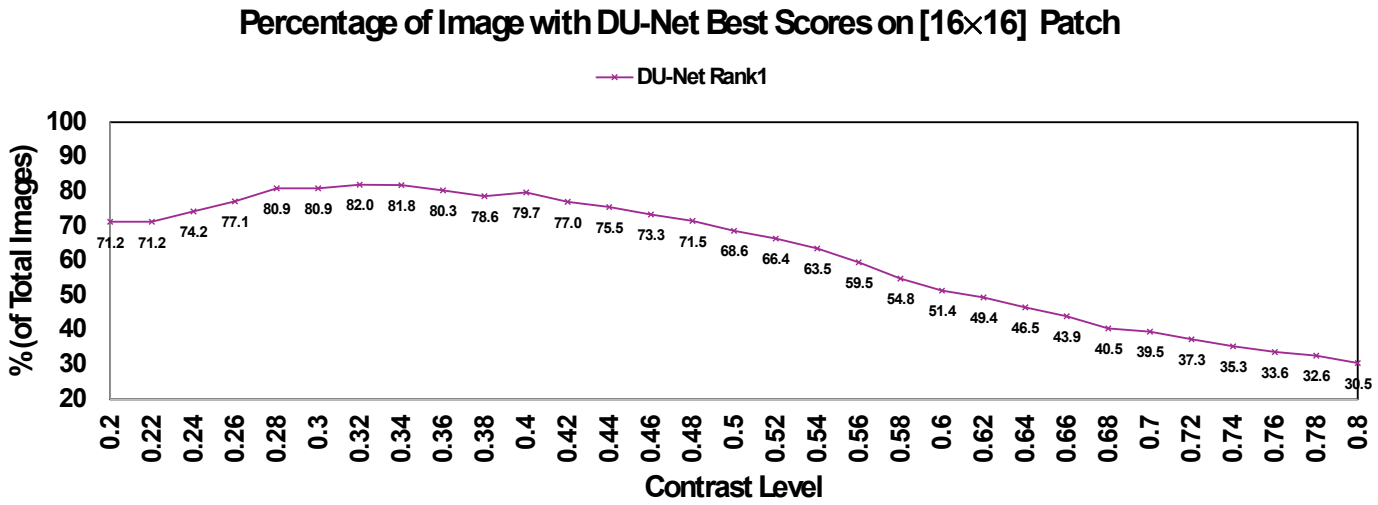

Figure 5. The percentage of images where the DU-Net scores best on [16 × 16] patch between different contrast levels.

4. Discussion

While the direct testing on the denoising result managed to find the most optimum patch size and contrast level, the library of images still used the 'good' images from the CNN training output. As NIQE should be able to detect the improvement unassisted in the photolike image, there is a need for an additional test using a different CT scan dataset to confirm that the ability is maintained. The two dataset patient designators were C081 and C095, with 595 images. The library of images still uses the previous ten patients' datasets from CNN training with 3300 images.

The new testing followed the same procedure: finding the NIQE average scores on patch [16 × 16] and 31 contrast levels (from 0.2 to 0.8 with a factor of 0.02) and determining which contrast has the highest improvement score on DU-Net. It was followed by a qualitative analysis involving the number of improved images and the number of images in which DU-Net came out at the top.

4.1. Average NIQE Scoring on the New Dataset

Rather than using a table, the score comparison is presented as a chart, as shown in Figure 6. The chart shows the best score on each contrast level, represented by a bigger

marker. The DU-Net came up to the top between the contrast level of 0.2 and 0.44, the DeConv-Net covers the contrast level between 0.46 and 0.56 with a single occurrence of 0.72, the Seg-Net has the rest of the contrast level, and the U-Net has no share. Thus, the following qualitative analysis focused on 0.2 to 0.44.

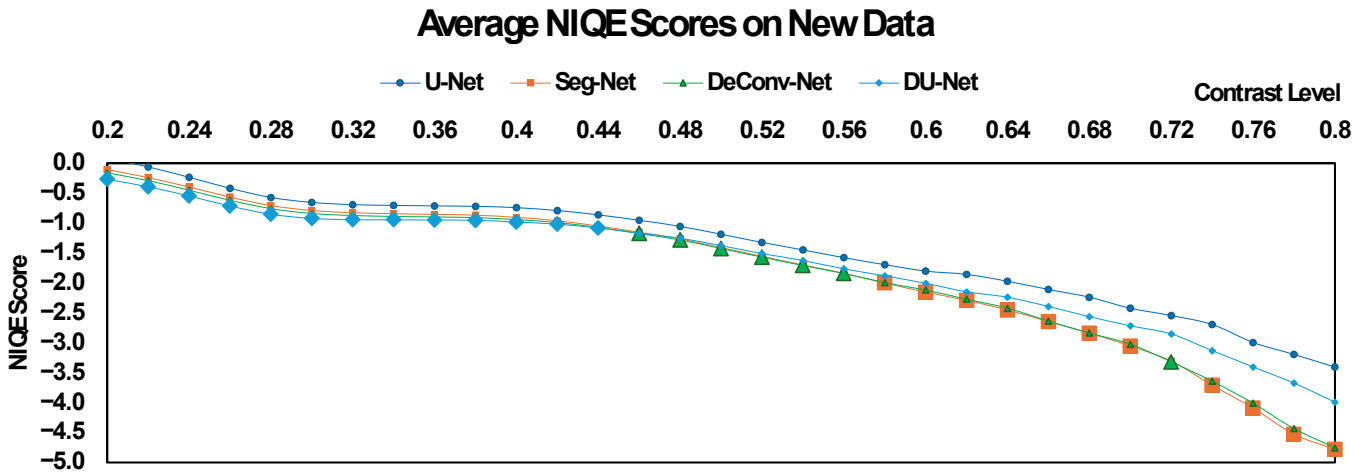

Figure 6. The new data average NIQE scores on different CNN models using [16 × 16] patch and a range of contrast levels.

4.2. Qualitative Analysis of the New Dataset

Following the highlight from NIQE average scoring, the first analysis of the number of improved images is shown in Figure 7. It shows 43% to 92% across the contrast of 0.2 and 0.44; those percentages are slightly lower than the previous assessment (see Figure 4) (64% to 95%) on the same contrast level. Therefore, the minimum rate was adjusted to 70% against 595 images. In this new percentage, the contrast level that meets the criteria is between 0.3 and 0.44.

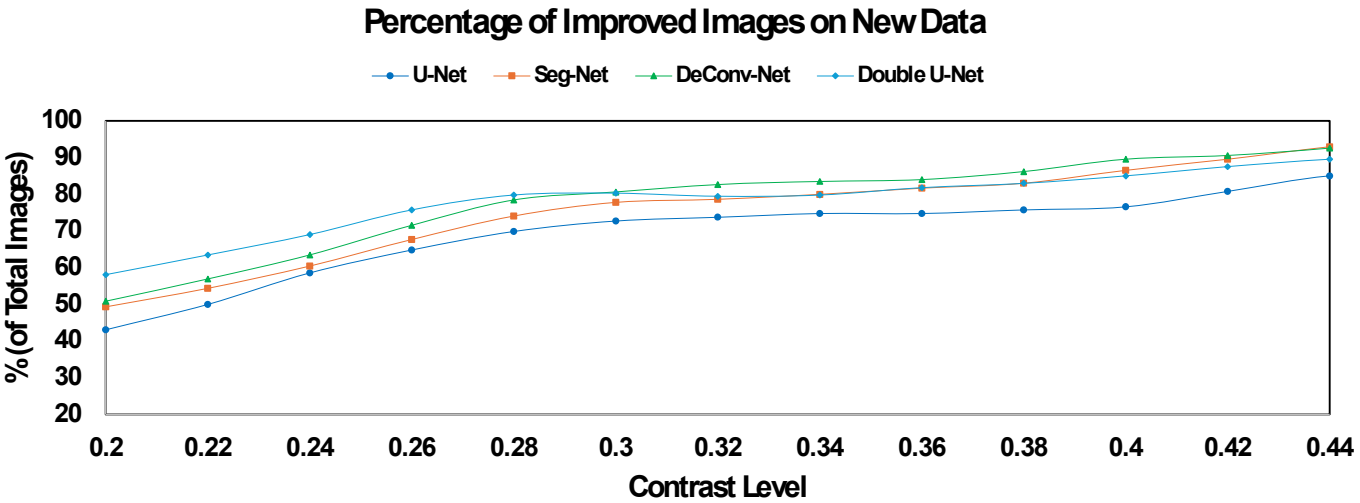

Figure 7. The NIQE perceived improvement in the new data between CNN models using [16 × 16] patch on selective contrast levels.

A follow-up analysis was performed on the new contrast level, as shown in Figure 8. It indicates the peak percentage at contrast level 0.3, and the contrast of 0.32, 0.34, and 0.36 follows closely.

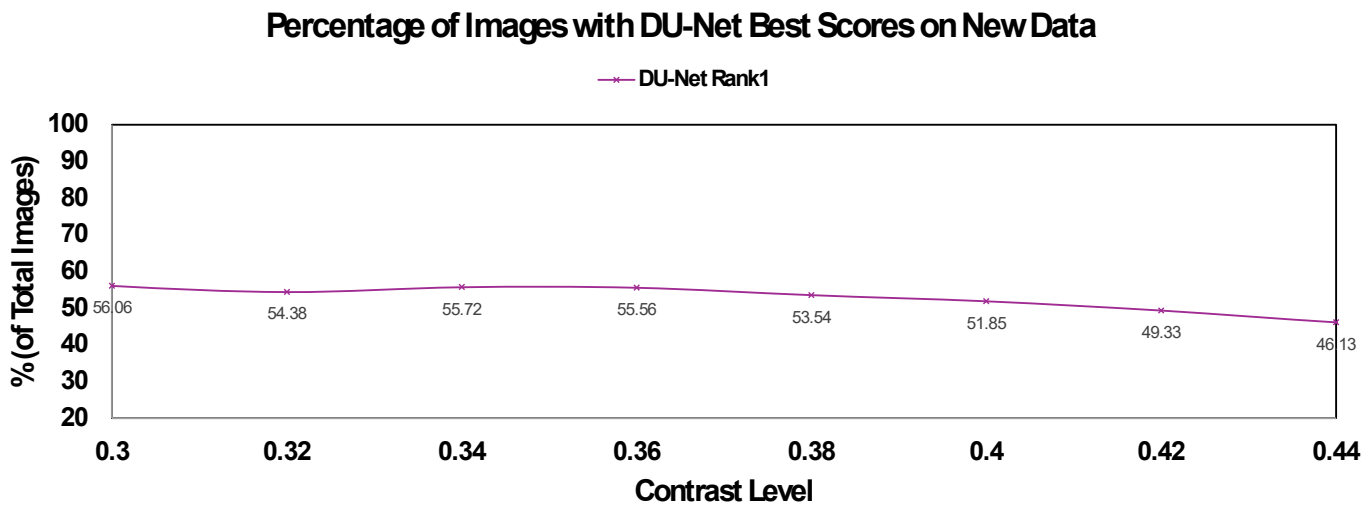

Figure 8. The percentage of images where the DU-Net scores best in the new data using [16 × 16] patch on the targeted contrast levels.

4.3. Overall NIQE Performance Assessment

Patch size is a critical component of the NIQE setting. While a smaller patch can deliver an outstanding result in the number of improved images, it cannot correctly distinguish the quality between different CNN models. Meanwhile, a bigger patch is difficult to analyze due to the number of pixels involved in the statistical assessment. It can make the statistical number too generic for checking any improvement.

On the contrast level, a smaller contrast makes it difficult to see the improved image, which indicates that the image still has more information after the thresholding process (it does not filter out most of the lung tissue). On the other hand, the higher contrast level has the opposite impact since most of the tissue is gone, as well as most of the noise; the assessment focuses only on the tissue with high contrast, such as bone. It is impossible to see the denoised difference from the bone tissue only. The high contrast can provide a good detection of the number of improved images but fails to see the difference between CNN quality.

The NIQE result indicates a fluid response when evaluating denoising in a CT scan. At the same time, there is a gap when using the same 'good' image for training and statistical library to the independent image dataset. A trend puts contrast levels between 0.3 and 0.36 at the top.

5. Conclusions

A complete assessment of NIQE shows the possibility of usage within denoising CT scans with the optimum setting. It does not have the same level of surety as in the Full Reference (FR) evaluators due to the generic statistical value, but it is proven to work well using different target images, confirming the No Reference (NR) status around the application. The optimum contrast level is related to the noise characteristics and subsequent tissue, and it works well when the thresholding covers the entire image to include various tissues and their noise artifacts.

As this research uses 3300 images for statistical reference on the model, the question remains whether the presented modeling uses the correct quantity and quality. The quantity represents the number of images for reference, while the quality represents the type and method of CT scanner producing those images. The possibility of mixing image output from different CT scanner brands or versions could become an interesting topic underlying the quality of 'good' images. The current dataset uses a 10% noise model, yet an ultra-

low-dose CT scan typically uses a 2% noise model. While it does not affect the statistical value from the reference model, it could be another direction to explore the impact of noise variation on the NIQE score and its setting. The last direction is utilizing NR evaluators other than NIQE to find their compatibility with denoising CT scans; BRISQUE or PIQE could be the starting point.

Author Contributions: Conceptualization, R.G.; investigation, R.G.; methodology, Y.T.; resources, J.Z. and R.C.; supervision, R.C.; validation, J.Z. and H.N.; writing—original draft, R.G.; writing—review and editing, Y.T., H.N. and R.C. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: Publicly available datasets were analyzed in this study. Data were obtained from The Cancer Imaging Archive (TCIA) hosted by The National Cancer Institute (NCI), Washington University in St. Louis, and the University of Arkansas for Medical Sciences (UAMS), Low Dose CT Image and Projection Data (LDCT-and-Projection-data) (Version 4) [Dataset] at https://doi.org/10.7937/9NPB-2637.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. de Koning, H.J.; van der Aalst, C.M.; de Jong, P.A.; Scholten, E.T.; Nackaerts, K.; Heuvelmans, M.A.; Lammers, J.J.; Weenink, C.; Yousaf-Khan, U.; Horeweg, N.; et al. Reduced Lung-Cancer Mortality with Volume CT Screening in a Randomized Trial. *N. Engl. J. Med.* **2020**, *382*, 503–513. [CrossRef]
2. Aberle, D.R.; Abtin, F.; Brown, K. Computed Tomography Screening for Lung Cancer: Has It Finally Arrived? Implications of the National Lung Screening Trial. *J. Clin. Oncol.* **2013**, *31*, 1002–1008. [CrossRef]
3. Zeng, D.; Huang, J.; Huang, H.; Bian, Z.; Niu, S.; Zhang, Z.; Feng, Q.; Chen, W.; Ma, J. Spectral CT Image Restoration via an Average Image-Induced Nonlocal Means Filter. *IEEE Trans. Biomed. Eng.* **2016**, *63*, 1044–1057. [CrossRef] [PubMed]
4. Schaap, M.; Schilham, A.M.R.; Zuidervelt, K.J.; Prokop, M.; Vonken, E.; Niessen, W.J. Fast Noise Reduction in Computed Tomography for Improved 3-D Visualization. *IEEE Trans. Med. Imaging* **2008**, *27*, 1120–1129. [CrossRef] [PubMed]
5. Gunawan, R.; Tran, Y.; Zheng, J.; Nguyen, H.; Chai, R. Image Recovery from Synthetic Noise Artifacts in CT Scans Using Modified U-Net. *Sensors* **2022**, *22*, 7031. [CrossRef] [PubMed]
6. Gunawan, R.; Tran, Y.; Nguyen, H.; Zheng, J.; Chai, R. Implementing Natural Image Quality Evaluator for Performance Indicator on Noise Artefacts Recovery in CT Scan. In Proceedings of the 45th Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC), Sydney, Australia, 24–27 July 2023; p. 4.
7. Mittal, A.; Soundararajan, R.; Bovik, A.C. Making a "Completely Blind" Image Quality Analyzer. *IEEE Signal Process. Lett.* **2013**, *20*, 209–212. [CrossRef]
8. Appina, B. A 'Complete Blind' No-Reference Stereoscopic Image Quality Assessment Algorithm. In Proceedings of the 2020 International Conference on Signal Processing and Communications (SPCOM), Bangalore, India, 19–24 July 2020.
9. Zhang, L.; Zhang, L.; Bovik, A.C. A Feature-Enriched Completely Blind Image Quality Evaluator. *IEEE Trans. Image Process* **2015**, *24*, 2579–2591. [CrossRef] [PubMed]
10. Rubel, A.; Ieremeiev, O.; Lukin, V.; Fastowicz, J.; Okarma, K. Combined No-Reference Image Quality Metrics for Visual Quality Assessment Optimized for Remote Sensing Images. *Appl. Sci.* **2022**, *12*, 1986. [CrossRef]
11. Wang, Z.; Li, Z.; Teng, X.; Chen, D. LPMsDE: Multi-Scale Denoising and Enhancement Method Based on Laplacian Pyramid Framework for Forward-Looking Sonar Image. *IEEE Access* **2023**, *11*, 132942–132954. [CrossRef]
12. Li, Z.; Li, J.; Zhang, Y.; Guo, J.; Wu, Y. A Noise-Robust Blind Deblurring Algorithm with Wavelet-Enhanced Diffusion Model for Optical Remote Sensing Images. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2024**, *17*, 16236–16254. [CrossRef]
13. Yang, Q.; Chen, H.; Ma, Z.; Xu, Y.; Tang, R.; Sun, J. Predicting the Perceptual Quality of Point Cloud: A 3D-to-2D Projection-Based Exploration. *IEEE Trans. Multimed.* **2021**, *23*, 3877–3891. [CrossRef]
14. Harron, N.A.; Osman, N.F.; Sulaiman, S.N.; Karim, N.K.; Ismail, A.P.; Soh, Z.H.C. An Image Denoising Model using Deep Learning for Digital Breast Tomosynthesis Images. In Proceedings of the 13th Control and System Graduate Research Colloquium (ICSGRC), Shah Alam, Malaysia, 23 July 2022.
15. Dhore, S.; Abin, D. Chest X-ray Segmentation Using Watershed and Super Pixel Segmentation Technique. In Proceedings of the 2021 International Conference on Communication information and Computing Technology (ICCICT), Mumbai, India, 25–27 June 2021.

16. Pankaj, D.; Govind, D.; Narayanankutty, K.A. Edge Preserved Herringbone Artifact Removal from MRI Using Two-Stage Variational Mode Decomposition. In Proceedings of the 2019 National Conference on Communications (NCC), Bangalore, India, 20–23 February 2019.
17. Outtas, M.; Zhang, L.; Deforges, O.; Hammidouche, W.; Sriri, A.; Cavaro-Menard, C. A study on the usability of opinion-unaware no-reference natural image quality metrics in the context of medical images. In Proceedings of the International Symposium on Signal, Image, Video and Communications (ISIVC), Tunis, Tunisia, 21–23 November 2016.
18. Clark, K.; Vendt, B.; Smith, K.; Freymann, J.; Kirby, J.; Koppel, P.; Moore, S.; Phillips, S.; Maffitt, D.; Pringle, M.; et al. The Cancer Imaging Archive (TCIA): Maintaining and operating a public information repository. *J. Digit. Imaging* **2013**, *26*, 1045–1057. [CrossRef] [PubMed]
19. McCollough, C.H.; Chen, B.; Holmes, D.R.I.; Duan, X.; Yu, Z.; Yu, L.; Leng, S.; Fletcher, J.G. Low Dose CT Image and Projection Data (LDCT-and-Projection-data) (Version 4) [Data set]. *Cancer Imaging Arch.* **2020**. [CrossRef]
20. Buades, A.; Coll, B.; Morel, J.M. A non-local algorithm for image denoising. In Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05), San Diego, CA, USA, 20–25 June 2005.
21. Dabov, K.; Foi, A.; Katkovnik, V.; Egiazarian, K. Image Denoising by Sparse 3-D Transform-Domain Collaborative Filtering. *IEEE Trans. Image Process.* **2007**, *16*, 2080–2095. [CrossRef] [PubMed]
22. Fan, L.; Zhang, F.; Fan, H.; Zhang, C. Brief review of image denoising techniques. *Vis. Comput. Ind. Biomed. Art* **2019**, *2*, 7. [CrossRef] [PubMed]
23. Ronneberger, O.; Fischer, P.; Brox, T. U-Net: Convolutional Networks for Biomedical Image Segmentation. In Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI), Munich, Germany, 5–9 October 2015; pp. 234–241.
24. Badrinarayanan, V.; Kendall, A.; Cipolla, R. SegNet: A Deep Convolutional Encoder-Decoder Architecture for Image Segmentation. *IEEE Trans. Pattern Anal. Mach. Intell.* **2017**, *39*, 2481–2495. [CrossRef] [PubMed]
25. Noh, H.; Hong, S.; Han, B. Learning Deconvolution Network for Semantic Segmentation. In Proceedings of the IEEE International Conference on Computer Vision (ICCV), Santiago, Chile, 13–16 December 2015.
26. Damera-Venkata, N.; Kite, T.D.; Geisler, W.S.; Evans, B.L.; Bovik, A.C. Image Quality Assessment Based on a Degradation Model. *IEEE Trans. Image Process* **2000**, *9*, 636–650. [CrossRef] [PubMed]
27. Zhang, X.; Feng, X.; Wang, W.; Xue, W. Edge Strength Similarity for Image Quality Assessment. *IEEE Signal Process. Lett.* **2013**, *20*, 319–322. [CrossRef]
28. Varga, D. Saliency-Guided Local Full-Reference Image Quality Assessment. *Signals* **2022**, *3*, 483–496. [CrossRef]
29. Wang, Z.; Simoncelli, E.P.; Bovik, A.C. Multiscale structural similarity for image quality assessment. In Proceedings of the 37th Asilomar Conference on Signals, Systems & Computers, Pacific Grove, CA, USA, 9–12 November 2003.
30. Pu, Y.; Wang, W.; Xu, Q. Image Change Detection Based on the Minimum Mean Square Error. In Proceedings of the Fifth International Joint Conference on Computational Sciences and Optimization, Harbin, China, 23–26 June 2012.
31. Abdusalomov, A.B.; Nasimov, R.; Nasimova, N.; Muminov, B.; Whangbo, T.K. Evaluating Synthetic Medical Images Using Artificial Intelligence with the GAN Algorithm. *Sensors* **2023**, *23*, 3440. [CrossRef] [PubMed]
32. Korhonen, J.; You, J. Peak signal-to-noise ratio revisited: Is simple beautiful? In Proceedings of the Fourth International Workshop on Quality of Multimedia Experience, Melbourne, Australia, 5–7 July 2012.
33. Kushwaha, S.; Amuthachenthiru, K.; Geetha, K.; Narasimharao, J.; Kumar, D.; Gadde, S.S. Development of Advanced Noise Filtering Techniques for Medical Image Enhancement. In Proceedings of the 5th International Conference on Intelligent Communication Technologies and Virtual Mobile Networks (ICICV), Tirunelveli, India, 11–12 March 2024.
34. Deepa, B.; Sumithra, M.G. Comparative analysis of noise removal techniques in MRI brain images. In Proceedings of the IEEE International Conference on Computational Intelligence and Computing Research (ICCIC), Madurai, India, 10–12 December 2015.
35. Taassori, M.; Vizvari, B. Enhancing Medical Image Denoising: A Hybrid Approach Incorporating Adaptive Kalman Filter and Non-Local Means with Latin Square Optimization. *Electronics* **2024**, *13*, 2640. [CrossRef]
36. Wang, Z.; Bovik, A.C.; Sheikh, H.R.; Simoncelli, E.P. Image quality assessment: From error visibility to structural similarity. *IEEE Trans. Image Process.* **2004**, *13*, 600–612. [CrossRef] [PubMed]
37. Mittal, A.; Moorthy, A.K.; Bovik, A.C. No-Reference Image Quality Assessment in the Spatial Domain. *IEEE Trans. Image Process* **2012**, *21*, 4695–4708. [CrossRef] [PubMed]
38. Venkatanath, N.; Praneeth, D.; Chandrasekhar, M.; Channappayya, S.; Medasani, S.S. Blind image quality evaluation using perception based features. In Proceedings of the 21st National Conference on Communications (NCC), Mumbai, India, 27 February–1 March 2015.
39. Zamani, M.; Azar, F.T. No Reference Image Quality Assessment Based on DCT and SOM Clustering. *IEEE Access* **2024**, *12*, 47258–47270. [CrossRef]
40. Athar, S.; Wang, Z. A Comprehensive Performance Evaluation of Image Quality Assessment Algorithms. *IEEE Access* **2019**, *7*, 140030. [CrossRef]

41. Zhang, K.; Zuo, W.; Zhang, L. FFDNet: Toward a Fast and Flexible Solution for CNN based Image Denoising. *IEEE Trans. Image Process.* **2018**, *27*, 4608–4622. [CrossRef] [PubMed]
42. Liu, G.; Dang, M.; Liu, J.; Xiang, R.; Tian, Y.; Luo, N. True wide convolutional neural network for image denoising. *Inf. Sci.* **2022**, *610*, 171–184. [CrossRef]
43. Chen, H.; Zhang, Y.; Kalra, M.K.; Lin, F.; Chen, Y.; Liao, P.; Zhou, J.; Wang, G. Low-Dose CT With a Residual Encoder-Decoder Convolutional Neural Network. *IEEE Trans. Med. Imaging* **2017**, *36*, 2524–2535. [CrossRef] [PubMed]
44. Gurrola-Ramos, J.; Dalmau, O.; Alarcon, T.E. A Residual Dense U-Net Neural Network for Image Denoising. *IEEE Access* **2021**, *9*, 31742–31754. [CrossRef]
45. Kumar, S.; Kurmi, Y. CNN-based denoising system for the image quality enhancement. *Multimed. Tools Appl.* **2022**, *81*, 20147–20174. [CrossRef]