Low-dose 2D CT Scan Image Denoising

Camilo A. Pérez Martínez
Ingeniería Biomédica
Universidad de Los Andes
ca.perez17@uniandes.edu.co

Daniel F. Baron Espitia
Ingeniería de Sistemas
Universidad de los Andes
df.baron10@uniandes.edu.co

Abstract

The X-ray radiation from computed tomography (CT) brought us the potential risk. Simply decreasing the dose makes the CT images noisy and diagnostic performance compromised. We propose to use a Generative Adversarial Network (GAN) [5] with Wassertein distance and a pretrained ResNet-152 network as a Perceptual Loss calculator. The idea of implementing ResNet-152 is that the inclusion of the outer layers presented in the ResNet-152 network architecture would be useful in the reconstruction of smaller structures in the input CT images. However, the results obtained in the experiments performed indicate that ResNet-152 as a feature extractor does not offer a great improvement in the reconstruction of smaller structures, but offers a slight improvement in execution time while preserving similar results in PSNR, SSIM and RMSE metrics. Additionally, We propose the implementation and characteristics of a hybrid loss that would improve the weaknesses identified in this article, as well as an evaluation metric to better understand the results of future experiments.

Keywords: Low-Dose, Full-Dose, CT-Scans, GAN, VGG-19, ResNet-152, Perceptual Loss, Wassertein Distance

1. Introduction

X-Ray computed tomography is being considered as an important medical imaging tool and has been widely used for diagnostic purpose. Moreover, CT scanning have revolutionized diagnosis and medical treatment, almost eliminating the need for once-common exploratory surgeries and many other invasive and potentially risky procedures [24]. However, the high radiation–dose exposure increase the probability to develop cancer in patients that requires recurrent CT Scans [27]. Currently, other techniques such as Low Dose CT Scans are being used in patients to avoid the risk of developing cancer. Those Low Dose CT Scan provides 1.4 mSv of radiation and a regular diagnostic CT scan provides 7 mSv. Low Dose CT scans are a common recommendation for adults who have a high risk of developing lung cancer, due to Chest and Abdomen CT scans have a high standard radiation dose compared to other body parts CT scans [19]. Nevertheless, there is a possibility that for the reconstructed images using Low Dose radiation CT Scans, some anatomical structures are not clear and even the diagnosis of lesions in these may be difficult. Besides, Kubo et al. [14] demonstrated that lesion characterization capability by low-dose CT images is not comparable to standard-dose CT images and therefore not sufficient for evaluation of localized lung lesions.

Image denoising aims to restore clean images from noisy input CT Scan. For instance, efforts have been made to find ways to reduce the dose, such as reducing the exposure levels of each projection angle [29]. Nonetheless, low exposures inevitably induce stronger quantum noise into the CT Scans, which introduces artifacts or blurs their image and thus diminishing the clinical value [29].

There has been a surge of research implementing deep learning techniques going towards this direction: projection domain-based denoising [1] [11], image domain-based denoising [12] [13], and regularized iterative reconstruction [6] [10] [28] [33]. Over the last decade, there has been a trend to develop new iterative algorithms for image denoising. Generally, these algorithms optimize an objective function that incorporates an accurate system model [16][18], a statistical noise model and prior information in the image domain respectively [2][7].

Previous efforts include total variation (TV) and its variants, includes dictionary learning [26]. [22] These iterative reconstruction algorithms increases image quality significantly, however, they still lose some details and suffer from remaining artifacts. Additionally, most of them require high computational cost, which truncates the applicability of these algorithms in a real context [23]. Furthermore, in terms of computational efficiency, sinogram pre-filtration and image post-processing are more convenient. Adequate noise characterization has been achieved in the sinogram domain. However, these methods can suffer from loss of resolution and edge blurring, which makes this method need to be thoroughly processed, otherwise artifacts can be incorporated in the reconstructed images [23]. This makes important the need for a new approach that takes into account these difficulties.

In the case of image post-processing, and with the Deep learning approach, these methods specially operate directly on the image. Recent efforts have been made in the image domain to reduce Low Dose CT scans(LDCT) noise and suppress artifacts, for example, applying methods such as the non-local means(NLM) that were adapted for CT image denoising [5]. Second, the adaptation

of K-SVD methods inspired by compressed methods, as proposed in [23] to reduce artifacts. Third, the block-matching 3D(BM3D) algorithm was used in image restoration for various CT imaging tasks. [23].

Finally, here we propose to use a Generative Adversarial Network (GAN) [5] with Wassertein distance and pretrained ResNet-152 network as Perceptual Loss calculator. We use the Wasserstein distance as the discrepancy measure between distributions and a perceptual loss that computes the difference between images in an established feature space [5] [2] [23]. Instead of VGG loss, we implement a ResNet as a feature extractor for Perceptual Loss calculator. We take into account, that ResNet model consist of blocks of "Bottleneck", therefore we chose only the outermost layers. Nevertheless, the resulting images exhibit unique characteristics that are not present in using VGG alone. Also, some recent efforts suggest that including this outer-layers used to improve the problem of artifacts occurrence but for denoising reconstructed facial images, the results of applying this in a grayscale format patches coming from 2D CT-Scan images suggests us that this could perform well but including ethical considerations. Our proposed GAN could estimate the distance of distribution between low-dose CT and normal-dose CT. In the process, the Perceptual Loss based on ResNet-152 could preserve as many image details as possible when suppressing the noise. The SSIM Loss preserves the structural and textural details after the denoising process, and L1 Loss keeps the sharpness of the denoised image, especially in the low contrast regions.

2. Related Work

In the last decades, great advances have been developed in the area of image denoising, including the use of different methods such as spatial domain, transform domain and CNN-based denoising methods. Spatial image denoising methods consist of removing noise by calculating the real pixel value from the correlation between different patches of the original image [8].

In these methods, different linear or nonlinear filters are used to reconstruct the image and remove the noise. However, when applying these filters, it is not possible to preserve some image artifacts. On the other hand, transform domain filtering methods transform the noisy image to another domain, and then apply a denoising procedure on the transformed image according to its characteristics and noise [29]. [19] Transform domain methods were developed from the Fourier transform, nevertheless there are a variety of transform domain methods, such as cosine transform or wavelet domain methods [20]. Furthermore, CNN-based denoising methods have emerged thanks to the rise and development of Deep Learning models to solve different computer vision tasks[8]. CNN-based denoising methods attempt to learn a mapping function by optimizing a loss function over a dataset containing noisy images. The results obtained by CNN-based architectures are outstanding, thus becoming one of the most widely used methods for image denoising [8].

In the same way, with the advancement of medical imaging, the problem of image denoising has moved to the medical environment. Early models focus on denoising synthetic images due to the lack of real data. Because of this, any data set can be used in the image denoising problem because noise is artificially introduced using some distribution such as Gaussian distribution or Poisson distribution. For example, Michael and Yoon [20] used the ChestX-ray14 dataset to address the image denoising problem. This dataset consists of 112,000 X-ray images from 30,000 patients and noise was introduced using a Gaussian distribution with sigma between 0 and 50. Michael and Yoon implemented Kai Zhang's DnCNN and NVIDIA Research's CNN Noise2Noise. Both methods were evaluated using PSNR and SSIM metrics [20].

Besides, in 2016, the Low Dose CT Grand Challenge was published which consisted of reducing noise in head, chest and abdominal CT images. Also, noise using a Poisson distribution was introduced to these images in order to simulate the outcome of low radiation dose images. This dataset has been widely used to propose new methods to address the problem of image denoising. For example, Qingsong Yang et al. [32] proposed the use of GANs (Generative Adversarial Netwrok) with Wasserstein Distance and Perceptual Loss to reduce the image noise of this dataset. The experimental results presented in this work demonstrate that the use of GANs helps to significantly improve the image quality. Also, [29] shows that the use of Perceptual Loss helps to reduce the over smoothing and loss of artifacts generated during image reconstruction. The results obtained by the proposed architecture were evaluated using PSNR and SSIM.

Moreover, Ti Bai et al. [3] proposed the implementation of a High Resolution Deep Network (HRNet) to address the Low Dose CT Grand Challenge problem. This architecture has two stages, a feature extraction stage and a prediction stage. For the feature extraction stage, different branches (convolutional layers) are used to extract features at different scales. To evaluate the results obtained, Ti Bai et al. used RMSE, SSIM and CNR (Contrast-Noise-Ratio) metrics. They observe that this architecture is able to remove the added simulated noise, but also to remove the noise inherited from the target images.

Finally, Hongming Shan et al. [25] designed a new neural network architecture for low dose CT (LDCT) and compare it with commercial iterative reconstruction methods used for standard CT. They use the Low Dose CT Grand Challenge dataset and evaluate the denoised images in terms of two aspects: noise suppression and structural fidelity. They found that the best DL reconstruction outperforms the best IR reconstruction in terms of structural fidelity. The architecture used by Hongming Shan et al. was the Modularized Adaptive Processing Neural Network (MAP-NN) consisting of multiple identical denoising network modules. To replicate the results obtained by Hongming Shan et al. the code and documentation are available in a GitHub repository.

2.1. Wassertein GAN Framework

The GAN arquitecture has the difficulties of network training and the vanishing gradient [21] [9]. To deal with these limitations, the GAN with the Wasserstein distance (WGAN) has been widely used [23] [15] [29] [4], which made use of the Wasserstein distance as the measurement of the difference between the distribution loss and perceptual loss[25]. Besides, gradient penalty was employed as a regular accelerated method for training network (WGAN-GP) [23]. In other side, more recently the WGAN-VGG approach for low-dose CT, which achieved promising denoised CT images, applying the perceptual using a pretrained on natural images VGG-19 [23]. WGAN-VGG could overcome the problem

of image overblur. Also, SMGAN [4] combined the L1 loss and the multiscale structure loss so that it outperformed the WGAN-VGG in convergence accuracy. But sometimes, the reconstruction images were fuzzy. Besides, the gradient penalty term weakened the express ability of GAN [4]. Furthermore, researchers found the denoising model with the conveying path-based convolutional U-net denoising model, which is called as CPCE. Fan et al. [**?**] improved the method and proposed a denoising framework.

3. Approach

3.1. Denoising Framework

Generally, the noise distribution in CT images is treated as the combination of quantum Poisson and electronic Gaussian noise. But the noise in reconstruction images is complex, and its distribution is always nonuniform. Besides, the relationship between NDCT and LDCT cannot be described with an accuracy mathematical model. So, only with conventional methods, we could hardly obtain better results of denoising LDCT images. Fortunately, the uncertain noise model can be estimated by deep learning techniques, because of its strong ability of capturing features.

Let z ϵ R N×N denote a LDCT image and x ϵ R N×N denote the corresponding NDCT image. The goal of the 3 denoising process is to seek a function G that maps LDCT z to NDCT x: G : z $\rightarrow$ x. The generative and adversarial abilities of GAN can be applied to extract features from deep levels with the spatial information of reconstruction images, so that GAN can identify the noise and effective image details. GAN usually includes a pair of neural networks: a generator G and a discriminator D[18]. The generator G can learn the real distribution of NDCT, and the discriminator D can make the best effort to distinguish between real or fake samples generated by G [17]. This pair of networks is often trained alternatively, so the competition encourages the generated samples to be hardly distinguished from real ones. Finally, we could obtain CT images of better quality [4] [23].

3.2. Baseline

We follow the strategy implemented in Generative Adversarial Network with Wasserstein Distance and Perceptual Loss introduced by Qingsong Yang et al. [23]. This network consists of three parts: Generator, Discriminator and Perceptual Loss Calculator.

3.2.1 Generator Architecture

The Generator is a convolutional neural network (CNN) composed of 8 convolutional layers. The first 7 hidden layers have 32 filters, and the last layer has a 1x1x1 convolution filter. The kernel used to compute the feature maps in the CNN is 3x3 in size. In each layer, the activation function is the Rectified Linear Unit (ReLU).

3.2.2 Discriminator Architecture

The Discriminator is a convolutional neural network with 6 layers. The first two layers have 64 filters, the next two layers have 128 filters, and the last two layers have 256 filters. The kernel size of these convolutional layers is 3x3. After these convolutional layers, there are two fully connected layers with 1024 and 1 outputs, respectively.

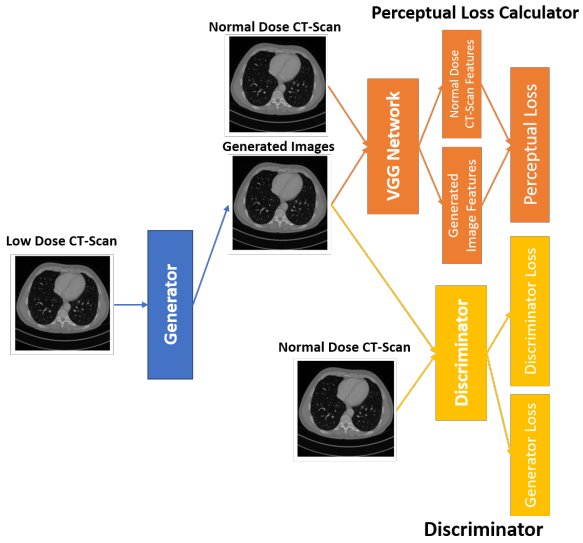

Figure 1: Architecture of Baseline Network.

Figure 2: Detailed architecture of Generator.

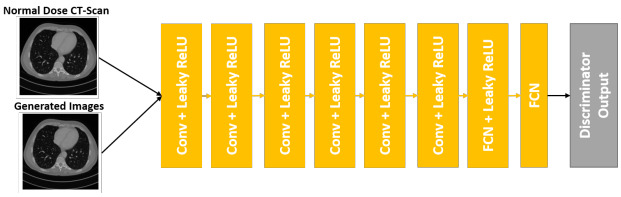

Figure 3: Detailed architecture of Discriminator.

3.2.3 Perceptual Loss Calculator

The Perceptual Loss Calculator consists of a VGG-19 network pretrained in ImageNet, and this is used to extract the features from the Normal Dose images and those generated by the Generator. The perceptual Loss is then calculated using the extracted features and L1 Loss formula. In addition, Since the pretrained VGG network takes color images as input while CT images are in grayscale, we duplicated the CT images to make RGB channels before they are fed into the VGG network. Finally, The VGG-19 network contains 16 convolutional layers followed by 3 fully-connected layers. The output of the 16th convolutional layer is the feature extracted by the VGG network and used in the perceptual loss function, then, the error is backpropagated to update the Generator weights.

$$L_{VGG}(G) = \frac{1}{whd} \|VGG(G(z)) - VGG(x)\|^2 \qquad (1)$$

where w, h and d are the width, height and depth of the feature space, respectively. For convenience, we call the perceptual loss computed by VGG network VGG loss.

3.2.4 Wasserstein Loss

Wasserstein distance solves the following min G and max D problem to obtain both Discriminator and Generator:

$$L_{WGAN} = -E_x\left[D(x)\right] + E_z\left[D(G(z))\right] + \lambda E_{\hat{x}}\left[(\| \bigtriangledown_{\hat{x}} D(\hat{x})\|^2 - 1)^2\right] \quad (2)$$

where the first two terms perform a Wasserstein distance estimation; the last term is the gradient penalty term for network regularization; xˆ is uniformly sampled along straight lines connecting pairs of generated and real samples; and λ is a constant weighting parameter. Also, Wassertein drops the last sigmoid in the implementation of the Discriminator and also removes the log function in the losses. Specifically, the networks D and G are trained alternatively by fixing one and updating the other [23]

Combining Perceptual Loss and Wasserstein Loss, we get the following equation expressed as:

$$Loss = \min_G \max_D L_{WGAN}(D,G) + \lambda_1 L_{VGG}(G) \quad (3)$$

3.3. Proposed Method

We propose a small modification in Perceptual Loss. The baseline method uses VGG-19 pretrained network, and we propose to use ResNet-152 pretrained network. The overall view of the proposed network structure is shown in Fig. 4. For convenience, we name this network WGAN-RESNET152, and it consists of three parts. The first part is the generator G, which is a convolutional neural network (CNN) of 8 convolutional layers, as is shown in Fig. 2. Following the common practice in the deep learning community [23] [31] [29], we used small 3 × 3 kernels in each convolutional layer. Due to the stacking structure, we think that such a network can cover a large enough receptive field efficiently. Each of the first 7 hidden layers of G have 42 filters, we also added more filters in this version, in order to increase the receptive field even more. The last layer generates only one feature map with a single 3×3 filter, which is also the output of G. We use Rectified Linear Unit (ReLU) as the activation function.

The second part of the Network is the Perceptual loss calculator using a pre-trained ResNet-152 in COCO dataset [29]. A denoised output image G(z) from the generator G and the ground truth image x are fed into the pre-trained ResNet-152 network for feature extraction. Then, the objective loss is computed using the extracted features from a specified layer according to Eq. 3. Following the strategy applied by [23], the reconstruction error is then backpropagated to update the weights of G only, while keeping the ResNet-152 parameters intact.

The third part of the network is the discriminator D. As shown in Fig. 3. D has 6 convolutional layers, with the structure inspired by others' work [23] [21] [9]. The first two convolutional layers have 64 filters, then followed by two convolutional layers of 128 filters, and the last two convolutional layers have 256 filters. Following the same logic as in G, all the convolutional layers in D have a small 3 × 3 kernel size. After the six convolutional layers, there are two fully-connected layers, of which the first has 1024 outputs and the other has a single output. Following the practice proposed by [4], there is no sigmoid cross entropy layer at the end of D. The network is trained using image patches and applied on entire images, we also proposed experiments changing the patch size and their number, the details are provided in Section of experiments.

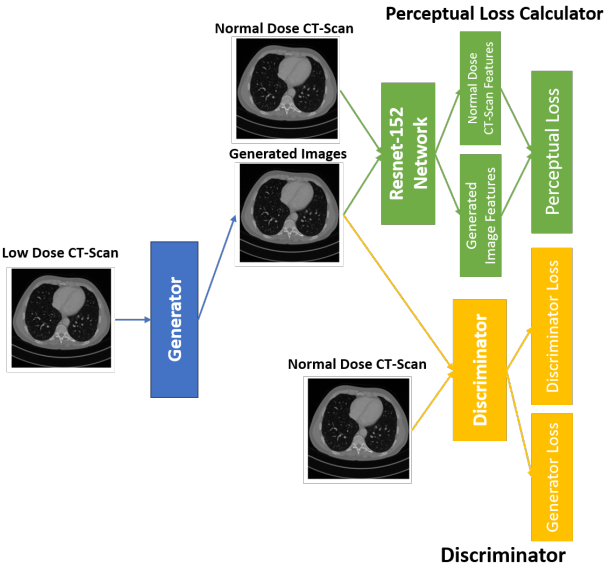

Figure 4: Architecture of proposed Method Network.

4. Experiments

4.1. Dataset

We use the Clinic Low Dose CT Grand Challenge dataset, a real clinical CT Scans published by Mayo Clinic in 2017. This dataset consists of 2378 normal and quarter doses 3D CT scans coming from 10 anonymous patients, the images have 512x512 slices with 1 mm thickness. The images related to quarter dose CT-Scans were simulated using a Poisson distribution Noise to reach a noise level that corresponded to 25% of the full dose CT-Scan images.

Specifically, in our experiments, we randomly extracted 10 pairs of 120 x 120 patches from every training image slices. It is important to mention that in the random selection of image patches, we randomly selected 80x80 patches where 70% corresponds to anatomical structures, that is, excluding those patches that have a high probability of having an area with a majority corresponding to air. In addition, figure 5 shows one example of Full Dose and Quarter Dose CT Scan, we have comparisons of low and normal resolution annotations and are available for the 2D original slices.

Additionally, it is important to note that we take all images of a patient to perform the tests of the proposed model. Specifically, the total number of images selected as testing dataset is 211 images, about 8% of the complete dataset.

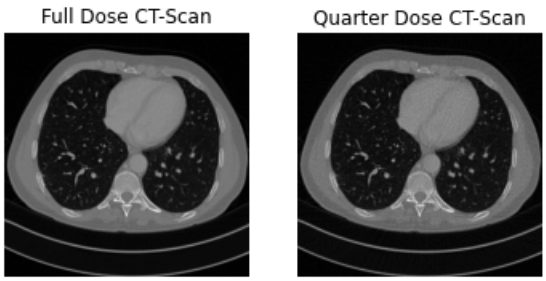

Figure 5: Comparison of Full Dose CT Scan and Quarter Dose CT Scan images from Low Dose Grand Challenge Dataset.

4.1.1 Annotations

Due to the nature of the image denoising problem, where we need to reproduce a high-resolution image using a low-resolution image, the available annotations for this problem are the high-resolution images or, in our case, the Full Dose CT-Scan images. Based on the random choice of patches for each image, we perform the same random transformation to the Full Dose and Low Dose CT-Scan images to avoid introducing bias in the proposed method.

In order to give a measure to make reliable comparisons, we use the high-resolution images as ground-truth. Based on some recent works that propose to use the image pairs as labels themselves. Therefore, we follow the strategy proposed in [23], and we will use the nominal dose CT (NDCT) images as the gold standard because they have the best image quality in this same data set [31] [14].

4.1.2 Evaluation Methodology

We use PSRN, SSIM, Mean CT numbers and standard deviations, with the objective to perform a quantitative evaluation. The aim is, define the ratio between the maximum possible energy of the low-dose images, thus evaluating the performance of the model by applying the same form of evaluation on the processed images [34]. This would allow us to predict the perceived quality of the resulting images, and their difference concerning low-dose images, to evaluate the performance of the denoising process [30]. In addition, to gain more insight into the output images from different approaches, we inspect the statistical properties calculating the mean CT numbers (Hounsfield Units) and standard deviations (SDs) of two flat regions [34].

Regarding that, the most important for medical images is to keep the necessary features used in pathologic diagnosis. [17], some state-of-the-art methods implements Perceptual Loss as a strategy, adding to the loss function a way to maintain image details or information content [23]. Perceptual loss functions have several properties that make them appealing in the context of denoising. (i) they do not suffer from regression-to-the-mean problem like point-estimate loss functions, (ii) the CNN-based architecture of the pre-trained network makes them more stable to local deformations in the LR image, and (iii) they demonstrate a lower variance for stationary textures in the input, which are abundant in natural images [22]. In other words, perceptual loss functions are low variance estimators that can produce stable high-frequency content and, consequently, sharp output images [23]. Our main aim is to efficiently filter out the artifacts introduced by a pre-trained perceptual loss function

Table 1: Summary of experiments performed.

Execution	Backbone Percep. Loss	Patch Size	Num. Patches	Discriminator
Baseline	VGG-19	(80, 80)	10	Enabled
Experiment 1	VGG-19	(80, 80)	16	Enabled
Experiment 2	ResNet-152	(120, 120)	8	Enabled
Experiment 3	VGG-19	(120, 120)	8	Enabled
Experiment 4	ResNet-152	(80, 80)	16	Disable
Experiment 5	VGG-19	(80, 80)	16	Disable
Experiment 6	ResNet-152	(80, 80)	16	Enabled
Experiment 7	ResNet-152	(80, 80)	10	Enabled

Table 2: Summary of hyperparameters used to execute experiments.

Hyperparameter	Value
Number of epochs	60
Lambda (λ)	10
Learning Rate	10^{-6}

4.2. Validation Experiments

4.2.1 Proposed Method

We performed the modification of the pretrained network used to extract features and calculate the perceptual loss. In this experiment, we use the parameters recommended by Qingsong Yang et al. [23]. In addition, we realize that if we increase the number of the randomly extracted patches, the proposed method will focus on small structures that may not appear if we use less number of patches in the training dataset. Due to the above, we increase the number of patches extracted from the low-dose CT-Scans to 16.

In addition, we want to evaluate the runtime of different configuration models to find out which configuration outperforms the best possible results in a considerable runtime. The summary of the experiments performed is presented in the table 1.

In the proposed experiments, we want to analyze the performance implications of varying the size and number of patches randomly extracted from the input images, in addition to the implications on runtime and quantitative metrics of disabling the discriminative model in the GAN model.

4.3. Evaluation Experiments

4.3.1 Baseline

In accordance with table 5 the results of the proposed baseline are comparable with the results obtained at [23]. The difference between applying 60 and 200 epochs explain the differences in quantitative results between the two approaches, which congruently, would be approximately half as good quantitatively as only have less than the half as many training epochs. However, the qualitative results, which can be seen in the 12,which can be seen in the table below, show a result where the PSNR and SSIM are approximately half of those obtained in [23], but the RMSE indi-

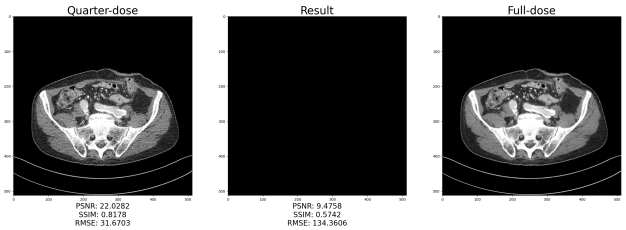

Figure 6: Comparison of results obtained by the Baseline Method.

Table 3: Summary of results obtained in the different experiments performed.

Execution	Runtime (h)	PSNR	SSIM	RMSE
Baseline	10,41	10,2259	0,575	124,2412
Experiment 1	18,69	19,559	0,6696	42,1995
Experiment 2	23,11	19,684	0,6716	41,6102
Experiment 3	26,66	20,3784	0,7113	38,4763
Experiment 4	5,468	10,2259	0,575	124,2412
Experiment 5	5,871	10,9268	0,5761	114,9027
Experiment 6	17,91	19,3576	0,6605	43,1869
Experiment 7	11,07	19,4038	0,6559	43,0236

cates a clear difference between the pixel intensities predicted by the network and the ground-truth.

Additionally, in the figures 11 and 9 qualitatively similar results can be appreciated, in spite of being variations with a deeper network and even a greater number of patches in each epoch. Also, quantitatively, similarities can be seen in the RMSE value, and in the execution time. This allowed us to identify a strategy that increased the execution time but apparently allows the generator network to create better candidates, as can again be seen in the figures with the discriminative network disabled. This strategy consists of increasing the amount of patches and decreasing the size of those. This allows the G network to increase the receptive field and get more information in each iteration, inevitably, this increases the execution of time, so we decided to also implement smaller patches and therefore equalize the overall run time.

Thus, this strategy allowed us to obtain results as the execution time per batch increased, as can be seen in the figures 11, 9 and even 13, this last one, indicates us, the threshold to start getting acceptable results from G but now, with Discminator network enabled.

In figures 6, 7, 8, 9 it is possible to observe the results obtained using VGG19 as feature extractor for the Perceptual Loss calculator. If we compare the results obtained by the baseline experiment, we can see that the quantitative and qualitative results have a great improvement.

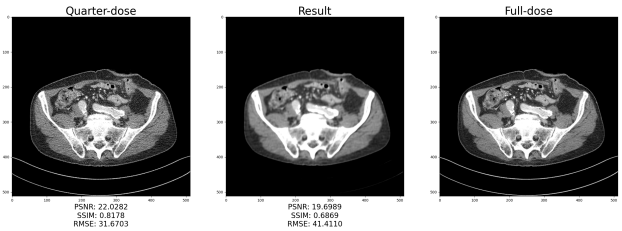

Figure 7: Comparison of results obtained by the Experiment 1.

For instance, in the results of experiment 1 it is possible to see a similar image compared with ground truth image. However, the predicted image is blurry, and it is not possible to distinguish the internal structures of the image.

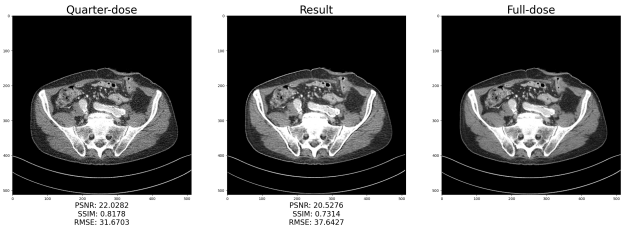

Figure 8: Comparison of results obtained by the Experiment 3.

On the other hand, in the image obtained using the model of the experiment 3 it is possible to see a huge improvement in the qualitative and quantitative results. The predicted image is very similar to the ground truth and in this case, it is possible to distinguish the smaller and internal structures of the image.

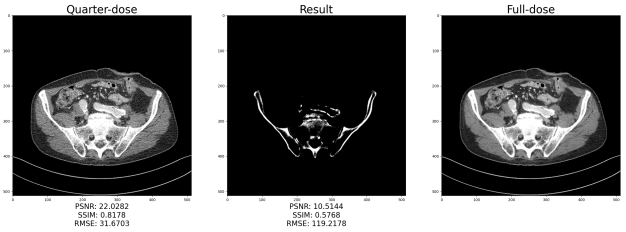

Figure 9: Comparison of results obtained by the Experiment 5.

Finally, if we analyze the results obtained with experiment 5, it is not possible to observe a huge improvement in the quantitative and qualitative results. In this image, it is only possible to distinguish the borders of some internal structure of the image.

4.3.2 Proposed Method

In figures 10, 11, 12, 13 it is possible to observe the results of the different configurations for the proposed method. Specifically, it is observed the results of change ResNet-152 as the pretrained network used as feature extractor for the Perceptual Loss calculator.

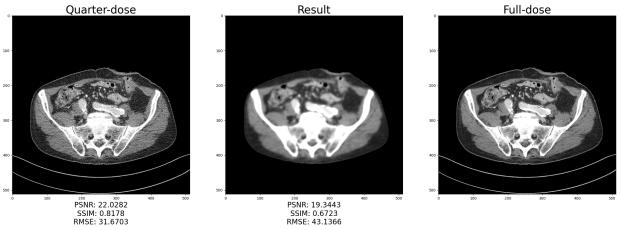

Figure 12: Comparison of results obtained by the Experiment 6.

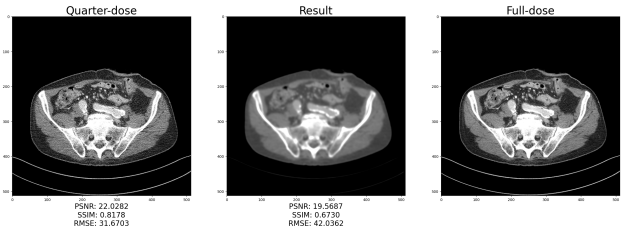

Figure 13: Comparison of results obtained by the Experiment 7.

If we compare the results obtained for the different experiments of baseline method and proposed method, we conclude that the results are quantitative and qualitative similar, but the runtime is slightly lower.

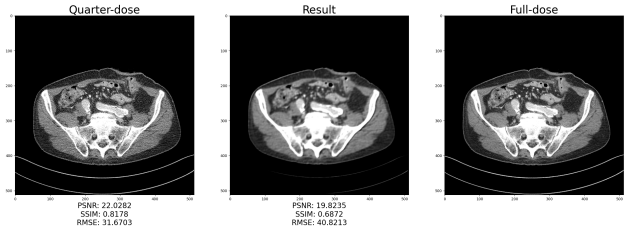

Figure 10: Comparison of results obtained by the Experiment 2.

Comparing the results of experiment 2 and experiment 3, we can see similar values for PSNR, SSIM and RMSE metrics, but a difference of almost 3 hours in the training run time. However, the qualitative results are very different because in experiment 2 the internal structures have a low resolution, and are not well distinguished, but in experiment 3 the internal structures have a similar resolution compared to ground truth and the smaller structures are very well appreciated.

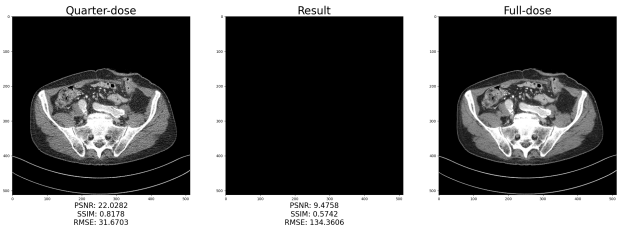

Figure 11: Comparison of results obtained by the Experiment 4.

5. Discussion

firstly, It is important to emphasize again, that the low dose images are simulated images, which again can be considered as a limitation of the model, as we are exposed to any bias that has been introduced since the creation of the images.

The method proposes to use ResNet-152 as a backbone and evaluate both its role and that of implementing a greater number of patches. It was identified that, in experiment 6 and 7, quantitatively the difference does not seem to be significant, even though the difference in execution is approximately 6 hours as 1 demonstrates. On the other hand, where a significant difference can be seen is in the qualitative results of both experiments, as shown in the following figure 12 and 13.

This can be considered as a limitation of the model, because it could be stated that the G-generator is creating artificial candidates with very good quantitative results but that to human perception, they do not take into account the definition of structures that, anatomically, can make the difference between a highly relevant medical decision, for example. Therefore, a model with a deeper and more updated backbone, such as ResNet-152, does not necessarily provide better and more useful results in this case, according to the objectives set. Furthermore, the qualitative results of 12 and 13 allow identifying the role of the number of patches, demonstrating that it helps the network to create realistic candidates with clear anatomical structures and with fewer included artifacts caused by excessive image blurring, again highlighting the importance of balancing the influence of the discriminator and the generator.

In order to take into account this balance and not have these inconsistencies between qualitative and quantitative results, some recent works have proposed to implement a hybrid loss that includes a special metric and loss function. [17] proposed a hybrid loss function with impact on the denoising results, trying to identify the optimal weighting parameter for each loss item in their hybrid loss function. We also identify that as 12 and 8 shows, the perceptual loss makes the edge more obvious, and it easily results in the artifacts and the adversarial loss makes the edge sharper (shown in 9. This suggests us that to implement a hybrid loss, with the aim of to preserve as much texture detail as possible when reducing noise. The hybrid loss function can keep the training process of the proposed method within bounds.

In other side, with the hybrid loss function, the differences between the generated images and real images can be measured and the weights of generator can be updated by back propagation (BP). In order to improve the denoising network, some recent jobs [4]

and [17] have implemented two more losses, those are, sharpness loss, and structural similarity loss, and we propose that our hybrid loss function should have four parts, which is adversarial loss, perceptual loss, sharpness loss, and structural similarity loss, respectively. Furthermore, these two losses would help the network to first, with the sharpness loss, usually used in sharpness detection network to evaluate the sharpness of images, our generator would be asked to not only generate the image as similar to the real one as possible but also generate the clear image as close to the real image as possible [9]. The structural similarity loss, which is usually used in CT images [32], would help us to control the existing strong feature correlation in the CT images with lose dose, in addition, according to [17] Structural similarity index (SSIM) includes three parts, which are luminance, contrast, and structure and these makes SSIM a better evaluating indicator than RMSE and peak signal-to-noise ratio (PSNR) in visual tasks. It is worth noting that the SSIM loss can be back-propagated to update the parameters of our network, when giving its property of differentiability.

Moreover, recent jobs [21], [9] and [4] implements the feature similarity indexing method (FSIM), which gives a normalized mean value of feature similarity between the two images. We believe that using this evaluation method would significantly help us to better understand the qualitative and quantitative results obtained from our future experiments.

6. Ethical Considerations

First of all, it is important to mention that the low quality images used are simulated images, i.e., these data have been modified with the aim of inducing a noise similar to that which real conditions force having images of a quarter of the conventional quality. Which implies that, the results of this method could be subject to any bias induced directly from the simulation data, where artifacts, anatomical structures and lesions can be considered. Therefore, our method should be robust enough to be able to avoid such biases coming from pre-processing, area-specific slices without human tissue, and data with altered lesions. Thus, we can achieve performance that has a significant impact in fields of real-world applicability.

Secondly, being images with a growing tendency to be used in medical diagnosis, especially for certain medical specialties such as radiology, oncology, orthopedics and others [15], it is important that the CT images resulting from the denoising process can support the decision-making process of health professionals. This implies the design and architecture of a model that guarantees high confidence in the images generated in the denoising process.

Thirdly, the model must take into account, during its training, specifically for GANs-related proposals, a careful control of the balance between G and D, either by means of the loss function or by implementing optimized training methods. In this way, the model should preserve a high fidelity to the real data, without producing new artifacts, nor generating bias towards anatomical structures that may be confusing for medical applications, reducing the probability of presenting false positives and false negatives of lesions or tumors, for example.

Fourthly, the actual case of low quality images may be due to the fact that different hospitals may differ in their CT scan image acquisition devices. Therefore, this would not only imply that the model must deal with different data sources having different kinds of noise, artifacts, intensity ranges and saturations, but also that the model must take into account that the difference between the actual quality difference and the simulated one. Additionally, the model should generate denoised images with the same quality, regardless of the different input variants and the differences between them. Finally, it is recognized the influence that the model can have directly on the health of a patient, because physicians will rely on the results. Therefore, the use of this model should be regulated, taking into account the precautions, possible biases and considerations made explicit here.

References

[1] J. D. T. N. K. J. M. K. C. M. M. A. Manduca, L. Yu and J. G. Fletcher. Projection space denoising with bilateral filtering and ct noise modeling for dose reduction in ct. *Medical physics*, 36(11):4911–4919, 2009.

[2] O. A. E. J. A. O. D. L. S. B. R. Whiting, P. Massoumzadeh and J. F. Williamson. "properties of preprocessed sinogram data in x-ray computed tomography". *Med. Phys.*, 33(9):3290–3300, 2006.

[3] T. Bai, D. Nguyen, B. Wang, and S. Jiang. Deep high-resolution network for low-dose x-ray ct denoising. *Journal of Artificial Intelligence for Medical Sciences*, 2(1-2):33, 2021.

[4] F. H. J. C. A. C. A. A. A. A. A. T. J. T. Z. W. C. Ledig, L. Theis and W. Shi. Photo-realistic single image super-resolution using a generative adversarial network. *IEEE Transactions on Medical Imaging*, page 1–19, 2017.

[5] R. N. J. W. D. S. B. C.-C. J. K. D. Kang, P. Slomka and D. Dey. "image denoising of low-radiation dose coronary ct angiography by an adaptive block-matching 3d algorithm. *SPIE Med. Imag. International Society for Optics and Photonics*, 21(2):86–92, 2013.

[6] I. A. Elbakri and J. A. Fessler. "statistical image reconstruction for polyenergetic x-ray computed tomography. *IEEE transactions on medical imaging*, 21(2):1272–1283, 2002.

[7] I. A. Elbakri and J. A. Fessler. "statistical image reconstruction for polyenergetic x-ray computed tomography". *IEEE Transactions on Medical Imaging*, 21(2):88–99, 2002.

[8] L. Fan, F. Zhang, H. Fan, and C. Zhang. Brief review of image denoising techniques. *Visual Computing for Industry, Biomedicine, and Art*, 2(1), 2019.

[9] A. A. J. Johnson and L. Fei-Fei. "perceptual losses for real-time style transfer and super-resolution. *IEEE Transactions on Medical Imaging*, 29(4):1–12, 2017.

[10] H. L. J. Wang, T. Li and Z. Liang. "penalized weighted least-squares approach to sinogram noise reduction and image reconstruction for low-dose x-ray computed tomography. *IEEE Transactions on Medical Imaging*, 25(10):88–98, 2006.

[11] T. L. J. Wang, H. Lu and Z. Liang. Sinogram noise reduction for low-dose ct by statisticsbased nonlinear filter. *Medical Imaging: Image Processing*, 5747(1):2058–2066., 2005.

[12] V. K. K. Dabov, A. Foi and K. Egiazarian. "image denoising with block-matching and 3d filtering. *n Image Process-*

ing: Algorithms and Systems, Neural Networks, and Machine Learning, 6064(1):188–203, 2012.

[13] V. K. K. E. Kostadin Dabov, Alessandro Foi. Image denoising by sparse 3-d transform-domain collaborative filtering. *IEEE Transactions on image processing*, 16(8):2080–2090, 2007.

[14] T. Kubo, Y. Ohno, D. Takenaka, M. Nishino, S. Gautam, K. Sugimura, H. U. Kauczor, and H. Hatabu. Standard-dose vs. low-dose ct protocols in the evaluation of localized lung lesions: Capability for lesion characterization—ilead study. *European Journal of Radiology Open*, 3:67–73, Mar 2016.

[15] T. Kubo, Y. Ohno, D. Takenaka, M. Nishino, S. Gautam, K. Sugimura, H. U. Kauczor, and H. Hatabu. Standard-dose vs. low-dose ct protocols in the evaluation of localized lung lesions: Capability for lesion characterization—ilead study. *European Journal of Radiology Open*, 3:67–73, Mar 2016.

[16] R. M. Lewitt. "multidimensional digital image representations using generalized kaiser–bessel window function. *J. Opt. Soc. Amer. A,*, 7(10):1834–1862, 1990.

[17] Z. Li, W. Shi, Q. Xing, Y. Miao, W. He, H. Yang, and Z. Jiang. Low-dose ct image denoising with improving wgan and hybrid loss function. *Computational and Mathematical Methods in Medicine*, 2021:1–14, 2021.

[18] B. D. Man and S. Basu. ""distance-driven projection and backprojection in three dimensions. *Phys. Med. Biol*, 49(11):234–259, 2004.

[19] F. A. Mettler, W. Huda, T. T. Yoshizumi, and M. Mahesh. Effective doses in radiology and diagnostic nuclear medicine: A catalog. *Radiology*, 248(1):254–263, Jul 2008.

[20] P. Michael and H.-J. Yoon. Survey of image denoising methods for medical image classification. *Medical Imaging 2020: Computer-Aided Diagnosis*, 2020.

[21] G. A. Nao Takano. "generator from edges: Reconstruction of facial images". *CVPR*, 33(9), 2020.

[22] X. M. L. Z. J. H. Q. Xu, H. Yu and G. Wang. "low-dose xray ct reconstruction via dictionary learning". *IEEE Transactions on Medical Imaging*, 31(9):1682–1690, 2012.

[23] Y. Z. H. Y. Y. S. X. M.-M. K. K. Y. Z. L. S. Qingsong Yang, Pingkun Yan and G. Wang. Low dose ct image denoising using a generative adversarial network with wasserstein distance and perceptual loss. *IEEE Transactions on Medical Imaging*, 21(2):1–12, 2021.

[24] H. M. School. Radiation risk from medical imaging, Sep 2021.

[25] H. Shan, A. Padole, F. Homayounieh, U. Kruger, R. D. Khera, C. Nitiwarangkul, M. K. Kalra, and G. Wang. Competitive performance of a modularized deep neural network compared to commercial algorithms for low-dose ct image reconstruction. *Nature Machine Intelligence*, 1(6):269–276, 2019.

[26] E. Y. Sidky and X. Pan. "image reconstruction in circular cone-beam computed tomography by constrained, total-variation minimization". *Phys. Med. Biol*, 53(17), 2008.

[27] A. Sodickson, P. F. Baeyens, K. P. Andriole, L. M. Prevedello, R. D. Nawfel, R. Hanson, and R. Khorasani. Recurrent ct, cumulative radiation exposure, and associated radiation-induced cancer risks from ct of adults. *Radiology*, 251(1):175–184, 2009.

[28] X. J. S. J. G. W. T. Bai, H. Yan and X. Mou. "z-index parameterization for volumetric ct image reconstruction via 3-d dictionary learning. *IEEE Transactions on Medical Imaging*, 36(12):2466–2478, 2017.

[29] B. W. Ti Bai, Dan Nguyen and S. Jiang. Deep high-resolution network for low dose x-ray ct denoising, 2021.

[30] X. Wang, L. Xie, C. Dong, and Y. Shan. Real-esrgan: Training real-world blind super-resolution with pure synthetic data, 2021.

[31] G. Wu, M. Kim, Q. Wang, B. C. Munsell, and D. Shen. Scalable high-performance image registration framework by unsupervised deep feature representations learning. *IEEE Transactions on Biomedical Engineering*, 63(7):1505–1516, 2016.

[32] Q. Yang, P. Yan, Y. Zhang, H. Yu, Y. Shi, X. Mou, M. K. Kalra, Y. Zhang, L. Sun, G. Wang, and et al. Low-dose ct image denoising using a generative adversarial network with wasserstein distance and perceptual loss. *IEEE Transactions on Medical Imaging*, 37(6):1348–1357, 2018.

[33] C. You, W. Cong, M. W. Vannier, P. K. Saha, E. A. Hoffman, G. Wang, G. Li, Y. Zhang, X. Zhang, H. Shan, and et al. Towards the clinical implementation of iterative low-dose cone-beam ct reconstruction in image-guided radiation therapy: cone/ring artifact correction and multiple gpu implementation. *Med Phys*, 41(11), 2014.

[34] C. You, W. Cong, M. W. Vannier, P. K. Saha, E. A. Hoffman, G. Wang, G. Li, Y. Zhang, X. Zhang, H. Shan, and et al. Ct super-resolution gan constrained by the identical, residual, and cycle learning ensemble (gan-circle). *IEEE Transactions on Medical Imaging*, 39(1):188–203, 2020.