CT-Scan Denoising Using a Charbonnier Loss Generative Adversarial Network

BINIT GAJERA[1], **SIDDHANT RAJ KAPIL**[1], **DORSA ZIAEI**[1], **JAYALAKSHMI MANGALAGIRI**[1],
ELIOT SIEGEL[1,2], **AND DAVID CHAPMAN**[1]
[1]Department of Computer Science, University of Maryland, Baltimore County, Baltimore, MD 21250, USA
[2]Department of Diagnostic Radiology and Nuclear Medicine, University of Maryland School of Medicine, Baltimore, MD 21201, USA

Corresponding author: David Chapman (dchapm2@umbc.edu)

ABSTRACT We propose a Generative Adversarial Network (GAN) optimized for noise reduction in CT-scans. The objective of CT scan denoising is to obtain higher quality imagery using a lower radiation exposure to the patient. Recent work in computer vision has shown that the use of Charbonnier distance as a term in the perceptual loss of a GAN can improve the performance of image reconstruction and video super-resolution. However, the use of a Charbonnier structural loss term has not yet been applied or evaluated for the purpose of CT scan denoising. Our proposed GAN makes use of a Wasserstein adversarial loss, a pretrained VGG19 perceptual loss, as well as a Charbonnier distance structural loss. We evaluate our approach using both applied Poisson noise distribution in order to simulate low-dose CT imagery, as well as using an anthropomorphic thoracic phantom at different exposure levels. Our evaluation criteria are Peek Signal to Noise (PSNR) as well as Structured Similarity (SSIM) of the denoised images, and we compare the results of our method versus recent state of the art deep denoising GANs. In addition, we report global noise through uniform soft tissue mediums. Our findings show that the incorporation of the Charbonnier Loss with the VGG-19 network improves the performance of the denoising as measured with the PSNR and SSIM, and that the method greatly reduces soft tissue noise to levels comparable to the NDCT scan.

INDEX TERMS CT-scan denoising, machine learning, computed tomography, medical diagnostic imaging, generative adversarial network.

I. INTRODUCTION

Computed Tomography (CT) is an x-ray imaging procedure where a narrow beam of x-rays enters the patient's body from many angles and offsets; these are reconstructed to form cross-sectional images or ''slices'' of the body. Then these slices are in turn analyzed for diagnostic purposes. The scans capture detailed images of internal organs, bones, soft tissue, and blood vessels. CT has an advantage over the x-ray modality in that the 3D volume can allow the radiologist to look around bones and other anatomical structures that might disrupt the view of important regions. However, a disadvantage of CT scanning is that the requirement of taking slices from so many different angles increases the overall amount of radiation exposure. As such, techniques to improve the quality of the CT scan while simultaneously decreasing the radiation exposure of the patient are an important area of research.

The simplest way to reduce image noise and improve quality is to increase the radiation exposure to the patient,

The associate editor coordinating the review of this manuscript and approving it for publication was Hong-Mei Zhang.

either by increasing tube current and/or exposure time. However, increased radiation exposure carries health risks, including a long-term increased risk of developing cancer.

Alternatively, it is possible to reduce image noise without increasing exposure through the use of sophisticated Tomographic Reconstruction (TR) algorithms. The choice of TR algorithm can greatly affect the perceivable image noise even if the raw sinogram data is exactly the same.

In recent years, there has been much research interest in deep reconstruction algorithms. A major branch of deep reconstruction is image-to-image deep denoising. These algorithms take as input an already reconstructed CT-scan image and output an improved image with higher quality and lower noise. At the time of writing, Generative Adversarial Networks (GANs) are the prevailing deep learning technique for image-to-image CT-scan denoising. A challenge with deep denoising techniques, in general, is that the generated scans are often excessively blurry. The need to reduce this excessive blur has led to many innovations in the design of CT-scan denoising GANs, notably improvements to the *Adversarial loss* and introduction of *Perceptual loss* and
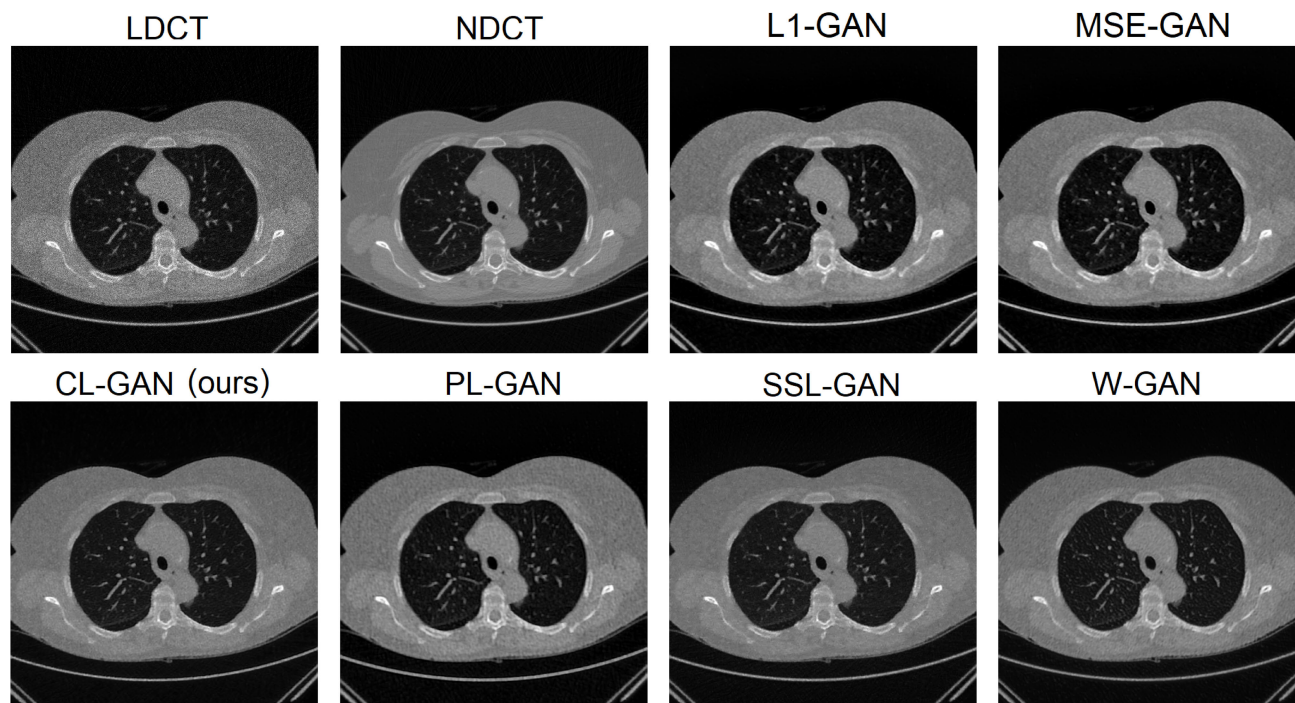
FIGURE 1. Result obtained from CL-GAN (ours), PL-GAN, SSL-GAN, L1-GAN and MSE-GAN on simulated dataset. Display window [−1024, 1024].

Structural loss terms. We present a novel GAN architecture for image-to-image deep denoising that incorporates Charbonnier distance as a *Structural loss* term. Figure 1 (right) shows an example generated slice from our method using a simulated Low Dose CT (LDCT) slice (left) as input in comparison with the ground truth Normal Dose CT (NDCT) slice (center). We demonstrate that our proposed CL-GAN architecture improves the overall quality of CT scan images and compares favorably against related deep denoising GANs that have been developed for this task.

A. CONTRIBUTIONS

- We introduce the Charbonnier distance as a structural loss term for a CT denoising GAN. Although the Charbonnier loss term has been recently demonstrated for related video-superresolution [1], to the best of our knowledge we are the first to evaluate this approach for CT-scan denoising.
- We analyze the L1 and L2 norms and describe a variance-bias tradeoff for denoising of bimodal distributions which is a theoretical justification for hybrid loss functions that such as Charbonnier that combine their advantages.
- We compare the results of our CT-scan denoising algorithm against two state-of-the-art CT-denoising GANs from literature and demonstrate that the proposed CL-GAN outperforms these techniques in terms of PSNR and SSIM.
- We evaluate the performance of the CL-GAN using published soft-tissue and smooth region noise metrics from Radiology literature in addition to PSNR and SSIM.

- We evaluate performance of CL-GAN and other methods using both simulated Poisson noise added, as well as a Phantom FDA dataset with scans obtained at higher and lower radiation exposure levels.

II. BACKGROUND
A. IMAGE QUALITY AND RADIATION EXPOSURE

The quality of CT-scans is a complex topic, although it is highly dependent on four basic factors: image contrast, spatial resolution, noise, and artifacts. Denoising algorithms attempt to reduce noise without introducing artifacts. The graininess in a LDCT is due to the limited number of photons that reach the photodetector [2]. Radiographic quantum mottle is fundamentally similar to the Poisson Distribution. As such, most studies of CT-scan denoising, including ours, attempt to simulate LDCT for evaluation purposes by applying Poisson noise to the image. The application of Poisson noise enables the investigator to simulate the appearance of an LDCT scan while retaining a ground truth image obtained from a higher dose. In addition to applied Poisson noise, we also perform analysis of the denoising capability of CL-GAN versus related techniques over a chest Phantom as scanned at different exposure levels. Image noise makes it particularly difficult to visualize small as well as low-contrast structures. As such, Radiologists prefer CT-scans with lower image noise when attempting to identify low-contrast structures for diagnostic purposes.

CT dose index (CTDI) is the most commonly used dose descriptor, which represents the dose to a location in a scanned volume. There are various versions of the descriptor such as $CTDI_{100}$ which takes a linear measure of dose

distribution over a pencil ionization chamber and hence does not take into account the topographical variation of a human body, $CTDI_w$ is a weighted dose index for periphery and center this makes it more relatable to the human body structure and is used in CT-scan instruments. $CTDI_{vol}$ is a type of dose index which performs similar to the weighted version but divides $CTDI_w$ by a pitch factor. Any form of CTDI is an estimate of radiation dose per body-mass expressed in the units of *Gray*, *(Gy)* which simplifies to *Joules/kilogram*, *(J/kg)*.

CTDI does not however take into account additional risks associated with distributing more radiation exposure over a larger region of the body. In order to capture these additional risks, another metric for radiation dose titled Dose-length product (DLP) is also measured as follows,

$$DLP = L \times CTDI_{vol} \qquad (1)$$

Here L is the total z-direction length of the examination which gives it the relationship with the level of depth as well. Some CT scanners also display DLP alongside the CTDI for operators.

B. TOMOGRAPHIC RECONSTRUCTION
The raw CT-scan is collected in the form of a *sinogram* which is a *Radon transform* of the actual volume radiodensity. This is because the device measures the X-ray attenuation from a series of beampaths from different angles and offsets around the body. The observed X-ray attenuation for a given beampath with angle ρ and offset θ, is given by the following equation:

$$p_{\rho,\theta} = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x, y)\delta(x\cos\theta + y\sin\theta - \rho)dxdy \qquad (2)$$

This represents a line integral of radiometric density over a pencil beam. In order to reconstruct a 2D slice of radiodensities, it is necessary to invert this equation. Exact evaluation of the inverse of this integral forms the branch of *Iterative Reconstruction* (IR) algorithms. The first CT scanners employed a variant of IR called *Algebraic Reconstruction Technique* (ART) in which the intersection of every grid cell and beampath form a system of linear equations of the canonical form, [3], [4].

$$Ax = b \qquad (3)$$

It is computationally expensive to solve this linear system directly for high-resolution scans. Furthermore, it is possible for this linear system to be ill-posed if the number of grid cells exceeds the number of beam paths for a slice. Finally, ART does not take into account the statistical distribution of radiographic quantum noise, because least-squares assume a Gaussian distribution but photon counts follow more closely the Poisson distribution.

Due to performance considerations, the industry-standard Tomographic Reconstruction (TR) algorithm is a variant of back-projection called *Filtered Back-Projection* (FBP) [5]. Backprojection approximates this inverse problem by smearing the line integrals over each of their contributing pixels, but

reconstructs a 2D slice that is excessively blurry. FBP reduces this excessive blur by performing a variation of highpass filtering in order to sharpen the image. This highpass filter magnifies high-frequency noise thereby leading to the excessive graininess of the image. As such, FBP images, although computationally cheap, are excessively grainy.

C. SINOGRAM FILTRATION AND ITERATIVE ALGORITMS
How can we achieve the best CT-scan image quality in a computationally efficient manner? This has been a hot topic for the past 50 years. Although our investigation has focused on image-to-image Deep denoising, it is important to recognize that alternative approaches exist and have been heavily studied. Two such broad categories of algorithms include Sinogram Filtration (SF) as well as IR.

IR spans a diverse set of methods that iterate between back-projection and forward projection thereby allowing the algorithm to take into account additional apriori information of the physical system [6]. The original TR algorithm was ART which is technically an IR algorithm [3], [4]. Today, IR comprises a wide variety of methods including major branches of statistical and model-based reconstruction. Statistical noise models [7]–[9] incorporate additional knowledge of quantum mottle distribution. Other explored techniques include Total Variation [10] and dictionary learning [11]. An underlying theme of these approaches, however, is their large computational overhead which has led to the investigation of alternate methods including deep denoising.

Another heavily studied approach involves filtering the sinogram in order to estimate and subtract the contribution of Poisson noise prior to TR [12], [13]. As such one might view SF as an attempt to improve the quality of the sinogram itself, although a limitation of these approaches is the possibility that filtration might reduce signal in addition to noise. One popular approach for sinogram filtration is the use of bilateral filtering [13] which has some ability to retain sharp edges. A recent approach by Karimi *et al.* [12] has shown that it is also possible to perform filtration using the sinogram data after the image reconstruction has been performed.

III. RELATED WORK
We now turn our attention to related works that have investigated the use of deep learning to improve CT-scan reconstruction through image-to-image denoising.

A. DENOISING NETWORKS
Deep reconstruction has a short history, with the first published method in 2017 (preprint in 2016) by Kang *et al.* [14], which developed a customized denoising autoencoder to produce wavelet coefficients. In the same year, additional Neural Network architectures were explored including pure convolutional approaches for artifact reduction [15], residual, and U-net based architectures [16], as well as persistent homology-based approaches [17]. The first methods were based loosely on denoising autoencoders with additional structural terms and did not make use of GANs.

Some researchers, such as Chen *et al.* [18], have followed an encoder/decoder approach of using residual networks. The network has two parts, Encoder, and Decoder. The Encoder consists of multiple layers of Convolutional kernels, while the Decoder consists of multiple layers of the De-convolutional kernels. These components have a residual connection between them. Their evaluation is shown using the Peak Signal-to-Noise ratio (PSNR) and Structural Similarity Index (SSIM) for the task of low-dose CT image denoising [18]. Shan *et al.* [19] improves on the convolutional encoder-decoder approach by incorporating 3D convolution and leveraging transfer learning from 2D CNNs for deep denoising. There is also a recent branch of denoising networks which attempt to incorporate Neural Networks as a regularization term for iterative reconstruction algorithms [20], [21].

B. DENOISING GANs

Recently, GAN based approaches are beginning to displace denoising autoencoders for image-to-image deep denoising. Much innovation in these areas has involved the development of structural and perceptual loss terms to improve the image quality and reduce excessive blur in the reconstruction [22]–[27]. The determination of the best perceptual and structural loss functions remains an open problem.

Wolterink *et al.* [22] were the first authors to build a GAN for Low-Dose CT-scan denoising. The GAN constructed by them incorporates voxelwise loss. The authors compared three novel loss functions: only voxelwise loss, voxelwise and adversarial loss, and only adversarial loss. It was reported that they were able to achieve the highest PSNR ratio with only voxelwise loss, but the addition of adversarial loss was able to capture the image statistics of normal dose scans.

Another innovation was the development of a sharpness detection network by Yi *et al.* [28]. Their GAN is a *Sharpness Aware GAN*, in the sense that the authors include an additional network to penalize the GAN for producing excessively blurry images. Reducing excessive blur is a focus of many new GAN architectures for this problem [23]–[27].

You *et al.* [24] proposed a deep denoising GAN based on a ResNet architecture with Wasserstein loss [23], [29]. The generator G proposed by the authors contains a Feature Extraction Network and a Reconstruction Network following an encoder/decoder model. The authors use the L1 loss function combined with Wasserstein adversarial loss and gradient penalty. This approach was later extended and modified to include a perceptual loss using the multi-scale Structural Similarity (SSIM) [25], [26]. At the time of writing the SL-GAN is considered a state-of-art method for image-to-image denoising.

Yang *et al.* [30] proposed a Wasserstein adversarial loss as combined with a pre-trained VGG-19 perceptual loss for deep denoising. The VGG19 network works as a feature extraction network with the intent of penalizing perceivable differences between the denoised and ground truth images [30]. The evaluation provided by the authors mainly consists of using Peak Signal-to-Noise Ratio (PSNR) and Structural Similarity Index (SSIM).

IV. METHODOLOGY

A. NETWORK ARCHITECTURE

The architecture for the proposed CL-GAN consists of three main parts: A Generator Network, a Discriminator Network, as well as a Perceptual and Structural Loss. Our Perceptual and Structural Loss is composed of a pretrained VGG network (VGG Loss) and additionally includes the Charbonnier distance as a structural loss term. A diagram of the overall network architecture is shown in Figure 2.

The generator G is a convolutional neural network with 8 layers each based on 3×3 kernels. Each of the first seven layers of the Generator have 32 filters and only the last layer generates the feature map with a single 3×3 filter. The Rectified Linear activation unit is applied after each of these layers [31].

The discriminator D as seen in figure 3. has 6 convolutional layers and is similar to the discriminator architecture of several recent works [25], [30], [32]. Here the first two layers have 64 filters, then the next have 128 filters, and the last two have 256 filters each. Among the convolutional layers, we also have added three batch normalization layers for stabilizing/optimizing the training for GAN. Similarly to the generator, the convolutional layers employ 3×3 kernels. The *head* exhibits two fully connected layers with 1024 and 512 outputs each. As we are using the Wasserstein GAN, and following the convention from the original authors [33], we do not employ a sigmoid cross entropy layer at the end of the discriminator.

The perceptual and structural loss block is shown in the top right of figure 2, consisting of the VGG perceptual loss as well as the Charbonnier structural loss. The incorporation of perceptual and structural loss terms greatly improves the performance of deep denoising GANs.

B. WASSERSTEIN GAN

The Wasserstein loss is an adversarial loss proposed by Arjovsky *et al.* [33] that greatly improves the stability of GANs relative to the original adversarial loss term of Goodfellow *et al.* [34]. The equation for Wasserstein distance is as follows,

$$W(\mathbb{P}_r, \mathbb{P}_g) = \inf_{\gamma \in \Pi(\mathbb{P}_r, \mathbb{P}_g)} \mathbb{E}_{(x,y) \sim \gamma}[\| x - y \|] \quad (4)$$

Wasserstein loss is based on the Earth-Mover (EM) distance between two probability density functions. Here, for the distance, we would be using a set of joint distributions whose marginals are $\mathbb{P}_r$ and $\mathbb{P}_g$. Here $\gamma(x, y)$ represents how much "mass" is required to transport from x to y so that the distribution $\mathbb{P}_r$ is transformed into the distribution $\mathbb{P}_g$ [33]. The EM distance then becomes the "cost" of the optimal transport plan. The authors [33] also describe that eq 4 is equivalent to the following due to the Kantorovich-Rubinstein duality [35],

$$W(\mathbb{P}_r, \mathbb{P}_\theta) = \sup_{\|f\|_L \leq 1} \mathbb{E}_{x \sim \mathbb{P}_r}[f(x)] - \mathbb{E}_{x \sim \mathbb{P}_\theta}[f(x)] \quad (5)$$
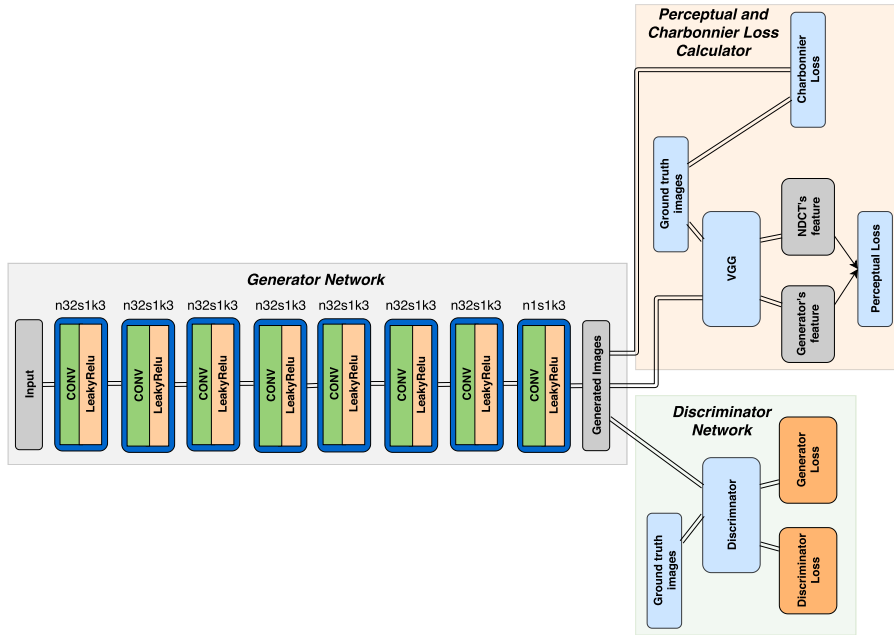
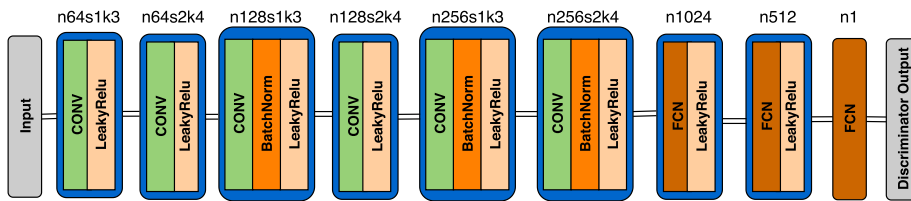**FIGURE 2.** Overall architecture of the proposed CL-GAN network network.

FIGURE 3. Discriminator has 6 blocks of convolution followed by 3 FC layers.

As such, the loss function is as follows [33], [36]:

$$\min_{G} \max_{D} L_{WGAN}(D, G) = -\mathbb{E}_x[D(x)] + \mathbb{E}_z[D(G(z))]$$
$$+\lambda \mathbb{E}_{\hat{x}}[(\|\nabla_{\hat{x}} D(\hat{x})\|_2 - 1)^2] \quad (6)$$

The first two terms perform the EM distance [35], and the last term is added for network regularization which is a gradient penalty term suggested by Gulrajani *et al.* [36]. $\hat{x}$ is uniformly sampled along straight lines connecting pairs of generated and real samples and λ is a constant weighting parameter [30]. From the equation, we can see that W-GAN removes the log function which can grow to infinity if the Discriminator learns to perfectly predict the samples, thereby leading to mode collapse.

C. VGG PERCEPTUAL LOSS
We employ VGG19 as a perceptual loss in order to maintain the low-contrast structures in the CT-scan [37]. VGG loss describes the square distance between VGG19 feature vectors of a pair of images as follows,

$$L_{PL} = \ell_{feat}^{\phi, j}(\hat{y}, y) = \frac{1}{C_j H_j W_j} \| \phi_j(\hat{y}) - \phi_j(y) \|_2^2 \quad (7)$$

VGG loss employs Mean Squared Error (MSE) over a pre-trained VGG19 feature vector rather than MSE directly

over the pixels. This approach was recently introduced to image-to-image denoising by Yang *et al.* [30].

In equation 7 C, H, and W represents depth, height and width respectively. $\hat{y}$ represents the result obtained from the generator and y represents the image patch from the real data distribution which for us would be the NDCT ground truth distribution over the training data. ϕ_j represents the feature extractor, here VGG-19 would be used as a feature extractor.

The VGG-19 network here works as a feature extractor. The pre-trained VGG model takes color images as input, but since CT-images are grayscale, we duplicate the CT slices to make RGB channels before feeding to VGG as input. VGG-19 has 16 convolutional layers and 3 fully connected layers [30]. The output obtained from the last convolutional layer is the feature that is extracted by the VGG network and is used in the perceptual loss function.

D. CHARBONNIER STRUCTURAL LOSS
We incorporate a Charbonnier structural loss term in order to pixelwise constrain the generated images to approximate the ground truth. Using only the adversarial loss of the GAN can construct artifacts in the images and show ringing patterns or unnecessary edges as investigated by Lucas *et al.* [1]. Such patterns can be reduced in GANs by incorporating a pixelwise structural loss function.
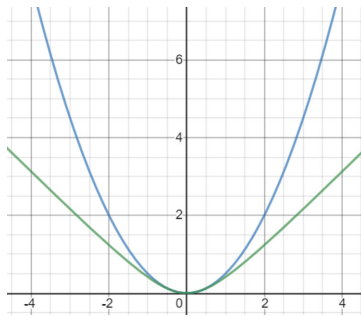
FIGURE 4. Comparison of MSE loss (blue) Charbonnier loss (green).

The proposed Charbonnier loss is a smooth approximation of the Huber loss. The Huber loss is an M-estimator which is quadratic for small deviations, but approximates L1 loss for larger deviations. As such, Huber loss (and furthermore Charbonnier loss) are able to exhibit the advantages of MSE over low variance regions (such as soft tissue), as well as the advantages of L1 over high variance regions (such as edges). A comparison of Charbonnier versus MSE is shown in figure 4. The Charbonnier loss can be written as follows [38],

$$L_{CL} = C(\hat{x}, x) = \sum_i \sum_j \sqrt{(\hat{x}_{i,j} - x_{i,j})^2 + \epsilon^2} \qquad (8)$$

where i, and j refer to pixel coordinates, $\hat{x}$ represents the estimated image obtained from the last layer of the generator and x represent the ground truth NDCT image. ϵ is the region for which the loss function changes from being approximately quadratic to approximately linear, and for our model this parameter was set to 0.001. The MSE as well as L1 and loss functions are common choices. However, neither is ideal for this task, and we show in Appendix A that there exists a variance/bias tradeoff between these estimators. MSE grows quadratically, which would overly penalize the model for large incorrect predictions. In practice MSE loss can lead to *blurry edges*, because the Maximum Likelihood Estimator (MLE) for MSE is the arithmetic mean but edges are bimodal: For example, consider an edge between air (−1000 HU) and soft tissue (−300 HU), the arithmetic mean would lie somewhere inbetween the two causing the model to favor overly blurring the boundary.

The use of the Charbonnier structural loss is justified by sampling theory, as it has the ability to improve the denoising capability over bimodal distirbutions (edges) without sacrificing accuracry over unimodal distributions (tissue interior). We show in Appendix A, that both the L1 and L2 norms suffer from a variance/bias tradeoff when sampling over bimodal distributions of the following form,
s

$$X_i \sim \alpha N(\mu_1, \sigma) + (1 - \alpha)N(\mu_2, \sigma) \qquad (9)$$

Variance and Bias are directly proporptional to MSE as follows,

$$MSE = E(\hat{\mu} - \mu_1)^2 = B^2 + V + \sigma^2 \qquad (10)$$

We show in Appendix A, that when sampling from this distribution, under reasonable assumptions such that the sample is sufficiently bimodal, more specifically N is large, $\alpha > \frac{4}{7}$ and $\mu_2 - \mu_1 \geq 2\sigma$, that a variance bias tradeoff between $L1$ and $L2$ estimators occurs analytically. Specifically, the $L1$ estimator has greater variance as follows,

$$\frac{\pi}{2} V_{L2} \leq V_{L1} \leq 2.21 V_{L2} \qquad (11)$$

whereas the $L2$ estimator has greater bias as follows,

$$B_{L2} \approx \sqrt{2} B_{L1} \qquad (12)$$

It is well known that when a variance/bias tradeoff is observed between estimators, that some combination of the estimators is justified to improve sampling performance. This formalizes our intuition that the MLE for L1 loss is the median which is typically a better estimator of population mode over bimodal distributions *(i.e. edges)*. However, the median may be less accurate than the mean if the probability distribution is sufficiently Gaussian, which is likely to occur over the interior of uniform regions such as air and smooth muscle. In uniform regions one might expect MSE loss to outperform L1 loss. As such, Charbonnier loss which is a smooth approximation of the Huber loss is likely to perform well under these circumstances by combining the best attributes of both the L1 and L2 norms.

E. COMBINED LOSS
Overall using the equation 6 and 7 and 8 we formulate a combined equation that represents the complete loss function for our W-GAN network as follows,

$$\min_G \{\lambda_1 [\max_D L_{WGAN}(D, G)] + \lambda_2 L_{PL}(G)$$
$$+ (1 - \lambda_1 - \lambda_2) L_{CL}(G)\} \qquad (13)$$

Here λ_1 and λ_2 are used as weighting parameters to control the trade-off between the three loss functions, that is, between W-GAN adversarial loss L_{WGAN} and the Perceptual loss from VGG L_{PL} and the Charbonnier Loss L_{CL}. The weights $\lambda_1 > 0$ and $\lambda_2 > 0$ with $\lambda_1 + \lambda_2 < 1$ are hyper-parameters.

V. DATA AND EXPERIMENTAL DESIGN
We evaluate the proposed deep denoising algorithm using two Thoracic CT datasets: the Kaggle Super Bowl 2017 [39] with simulated Poisson noise, as well as the Phantom FDA dataset from The Cancer Imaging Archive (TCIA) [40]. The Phantom FDA dataset contains scans of a anthropomorphic thoracic phantom at different exposure levels, thereby enabling us to evaluate the denoising performance over observational noise returns with minimal non-rigid motion.

A. SIMULATED POISSON NOISE DATASET
For evaluation purposes, we employ the Kaggle Super Bowl 2017 Data Science competition [39]. The dataset contains 1200 high-resolution DICOM chest CT-scans which were originally obtained from the National Cancer Institute

(NCI). All of the scans are obtained from high-risk individuals for lung cancer screening (60+, 30 pack years). Each image in the dataset contains a variable length series of 2D axial slices of the chest cavity, for which we uniformly select 20 axial slices per examination for inclusion in the dataset.

We randomly selected 75 patients out of 1200 as our train and test datasets. Of these 75 patients' examinations, 65 have been used for training, 6 for validation, and 4 patient's for testing. For each examination, we uniformly extract 20 or 21 slices. Separating the train and test datasets at the exam level, rather than the slice level, is necessary to ensure the train and test sets are fully decoupled.
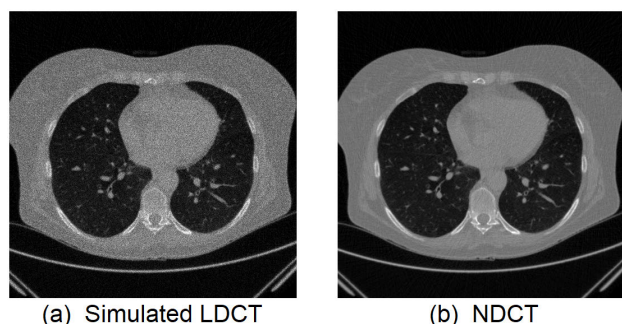

(a) Simulated LDCT   (b) NDCT

FIGURE 5. Simulated Low-Dose CT using Poisson Distribution.

We apply Poisson noise to the NDCT slices in order to simulate how they might appear if captured at a lower dose. We call this simulated LDCT imagery. The poisson distribution gives the probability that k events would occur given a mean of λ is described as follows,

$$f(k; \lambda) = \frac{\lambda^k e^{-k}}{k!} \tag{14}$$

For sufficently large values of λ, the Poisson distribution is approximately Gaussian as follows,

$$f(k; \lambda) \approx N(\mu = \lambda, \sigma = \sqrt{\lambda}) \tag{15}$$

For which σ is described as the noise power. We label the slices with Poisson noise added as the *simulated Low Dose CT (LDCT)* imagery, and the unmodified slices as *Normal Dose CT (NDCT)*. The simulated LDCT is created with the interval value of adding noise with the power of 50 *HU* leading to SNR of approximately 20 *db*. Application of Poisson noise is standard practice in order to simulate LDCT images while maintaining ground truth NDCT images for the evaluation of denoising algorithms.

B. PHANTOM DATASET
In order to evaluate the performance of the denoising GAN algorithm we employed the Phantom FDA dataset from TCIA [40]. This dataset uses an anthropomorphic thoracic phantom scanned at varying radiation exposure levels ranging from 25 mAs to 200 mAs. The Phantom FDA dataset consists of scans from two CT scanners a Philips 16-row scanner as well as a Siemens 64-row scanner.
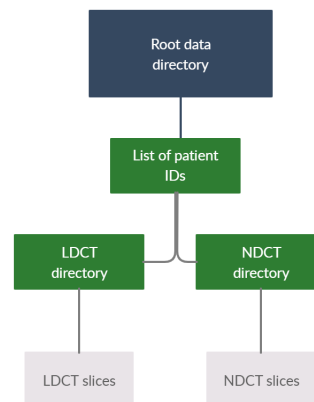

FIGURE 6. Directory tree for training data.

Utilizing the Phantom FDA dataset for deep denoising presents several challenges, most notably scans acquired at different noise levels may have slight rigid movement which makes subtraction of images to determine noise quality a nontrivial task. The use of an inanimate Phantom eliminates non-rigid movement between scans, but rigid-body motion may exist in the event that the positioning of the Phantom shifts slightly between successive scans. We correct for this rigid-body shift in two ways: (a) image registration of slices using rotation and translation, (b) the physical acquisition of scans on the same day.

We noticed that imagery collected on the same day using the same scanner had far fewer image registration artifacts than imagery collected on different days. A likely explanation is that the Phantom was scanned in place at varying settings within a single day, but the Phantom was likely replaced inbetween successive days of data acquisition.

To compare the performance of different denoising GANs, we used phantom scans of exposure 25 mAs and 100 mAs with pitch 1.2, and slice thickness 2 mm utilizing the Philips 16 row scanner. We utilize the 100 mAs scan as our ground truth NDCT scan, and the 25 mAs as the noisy LDCT scan in our comparison. These settings and exposure levels were selected after an investigation of the timestamps reveled these scans to be the largest discrepancy in exposure levels collected in the same day of observation without variation of other parameters. Utilization of scans acquired in the same day was necessary to enable high quality intercomparison of the denoising techniques because rigid-body motion between scans is nearly negligible and can be removed with basic image registration techniques.

C. PREPROCESSING AND TRAINING
We apply several preprocessing steps in order to facilitate the training of the GAN model. Firstly, we normalize the slices from Houndsfield units (−4096 to 4095) to a standard range of (−1.0 to 1.0). Secondly, we randomly select patches of size 64 × 64 in order to train the model. Selecting patches is technically a form of augmentation (random cropping), and increases the number of images available for training.

The validation and testing images were performed on whole images.

During training, the model is optimized using the Adam optimizer [41]. Using the validation data, we manually tuned the model over a variety of hyper-parameters, and consider the values in table 1 to be adequate. We also considered several patch sizes but suggest the size of 64×64. The gradient penalty weight is chosen as 10 suggested by Gulrajani *et al.* [36].

TABLE 1. Hyper parameters for training.

Parameters	Value	Parameters	Value
Epochs	800	Perceptual loss weight	0.3
Learning rate	0.0001	Adversarial loss weight	0.5
Batch size	128	Epsilon	0.001
Discriminator's iteration	3	Adam Beta1	0.5
Gradient penalty weight	10	Adam Beta2	0.9

D. EVALUATION METRICS

In order to evaluate the model in comparison with ground truth we make use of two standard evaluation metrics: Peek Signal to Noise (PSNR), and Structured SIMilarity (SSIM). We also report the level of soft tissue noise using the method of Samei *et al.* [42]. As suggested by National Instruments [43], the PSNR ratio can be considered as a good image quality metric. PSNR is the ratio between the maximum power of a signal and the power of distorting noise that affects the quality [43], and is typically reported in decibels as follows,

$$PSNR = 10 \, log \left(\frac{MAX_I^2}{MSE} \right) \qquad (16)$$

where MAX_I^2 is the maximum intensity of the image in HU, and *MSE* is the mean square error between denoised and ground truth images.

We also make use of SSIM as a image quality assessment metric. SSIM is a perceptual distance metric proposed by Wang *et al.* [26]. SSIM is defined between -1 and 1, in which 0 indicates no similarity between the denoised and ground truth images. The SSIM metric is defined as follows,

$$SSIM = \frac{(2\mu_x\mu_y + c_1)(2\sigma_{xy} + c_2)}{(\mu_x^2 + \mu_y^2 + c1)(\sigma_x^2 + \sigma_y^2 + c_2)} \qquad (17)$$

where μ_x and μ_y are the means of images x and y for comparison, σ_{xy} is the covariance of the images, and σ_x^2 and σ_y^2 are the individual pixel variances.

E. SOFT TISSUE NOISE

In addition to the PSNR and SSIM metrics, we also report the *Soft tissue noise*, using the method as proposed by Kamalanathan *et al.* which is largely similar to the method of Samei *et al.* [42], [44]. The purpose of this method is to calculate the standard deviation of the pixels over a soft tissue medium which is anatomically very smooth.

Soft tissue noise has been shown to be highly anticorrelated with CTDI and DLP, although it is an imperfect measure. In particular, the metric will not be able to detect if an

image blurs the tissue boundaries, as this metric discards edges in order to calculate the noise over uniform material only. As such, we do not consider *soft tissue* noise to be an evaluation metric, but rather a descriptive estimate of global image noise over uniform material.

Figure 7 shows how the uniform regions of *soft tissue* are selected by the algorithm of Kamalanathan *et al.* [42] which is largely similar to that of Samei *et al.* [44]. We see that non-overlapping boxes of 30×30 pixels are selected if these boxes have a low standard deviation, and lie within a range of radiodensities associated with smooth muscle. Smooth muscle is associated with roughly 300 *HU*.
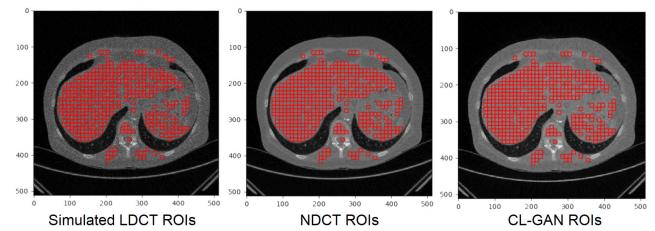

Simulated LDCT ROIs   NDCT ROIs   CL-GAN ROIs

FIGURE 7. ROIs obtained for calculation of soft tissue noise.

We determine the regions of *uniform air* and *smooth muscle* over the NDCT scan, but then apply these same regions to the denoised scan for each of the algorithms in our comparison. This is seen in figure 7 in which the denoised image has the same regions of interest as calculated over the original NDCT scan. The reported *Air noise* and *Soft tissue noise* are standard deviations in HU over the smooth Regions of Interest (ROI), and serve as an estimate of global image noise.

VI. RESULTS

In this section, we present the results obtained using the proposed CL-GAN model in comparison with other models, including two state-of-the-art published methods. We present results both in comparison with simulated Poisson noise as well as the Phantom FDA returns.

For both the Qualitative and Quantitative results, we compare our proposed CL-GAN model versus 5 alternative GAN methods. Two of these methods are considered state-of-the-art CT denoising GANs from recent literature [25], [30]. We also compare against a L1 structural loss GAN, the MSE structural loss GAN, and the Wasserstein adversarial loss GAN. Finally, as a naive baseline, we compare against the original noisy LDCT scan. We find that CL-GAN achieves the highest PSNR on the test dataset of any of the methods in this comparison as seen in Fig. 13 and Fig. 15.

We describe each of these methods as follows, SSL-GAN is the structurally sensitive loss SSIM equation as proposed by You *et al.* [25]. PL-GAN is the Wasserstein GAN architecture as proposed by Yang *et al.* [30]. W-GAN is a baseline GAN using Wasserstein adversarial loss but without the use of any perceptual loss term. L1-GAN and MSE-GAN represent the Wasserstein GAN with L1 loss as well as the MSE perceptual loss functions respectively. LDCT represents the original test

image scan with Poisson noise and is expected to be lower than the rest of the models.

A. QUALITATIVE RESULTS

Figures 8, 9, and 10 show example denoising results of different slices from the test dataset with added Poisson noise. Figure 8 shows an example slice with highlighted regions, whereas 9 and 10 show zoomin details of the denoising output over pulmonary vessels and spinal cord respectively. Figures 11 and 12 show example denoising results using the Phantom FDA dataset, where 11 shows whole CT clides, and 12 show zoom-in results over a specific region containing bone, soft tissue and lung anatomy.

All images are shown using a display window of [−1024, 1024] Houndsfield Units, and compare the LDCT, NDCT, L1-GAN, MSE-GAN, CL-GAN (ours), PL-GAN,
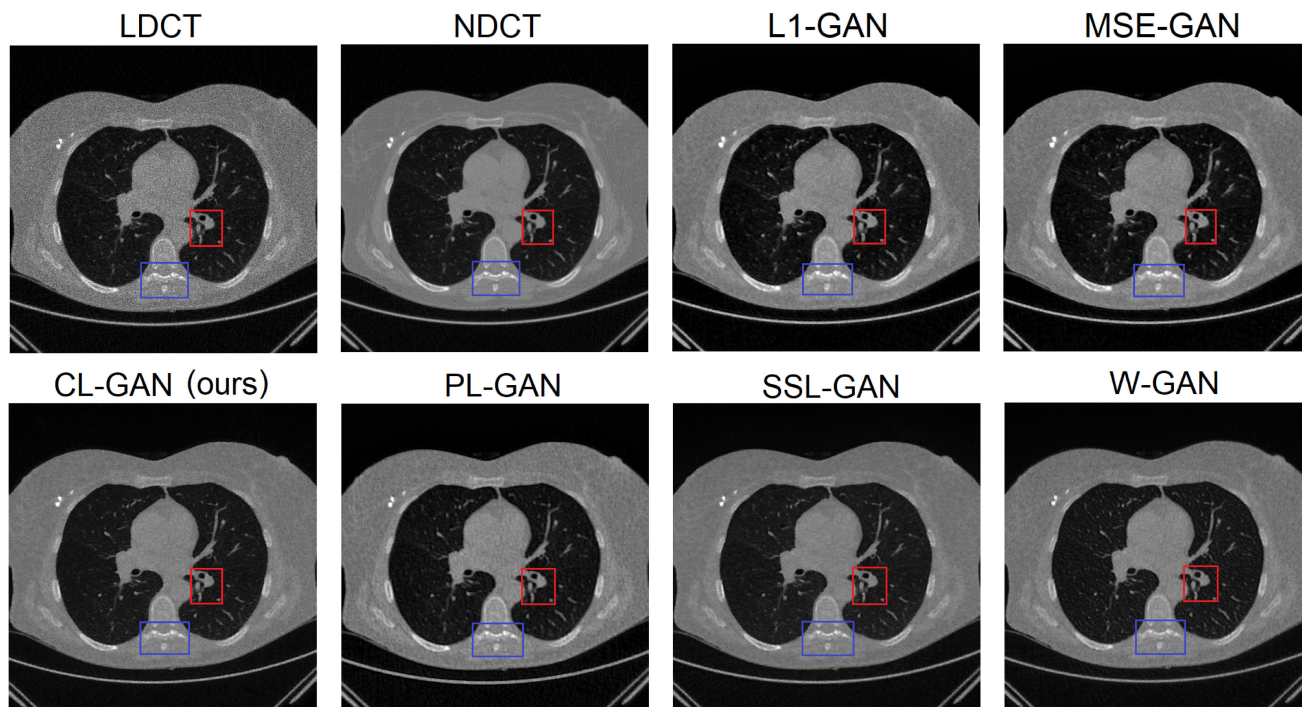

FIGURE 8. Result obtained from PL-GAN, SSL-GAN, CL-GAN (ours) on simulated dataset. Display window [−1024, 1024].
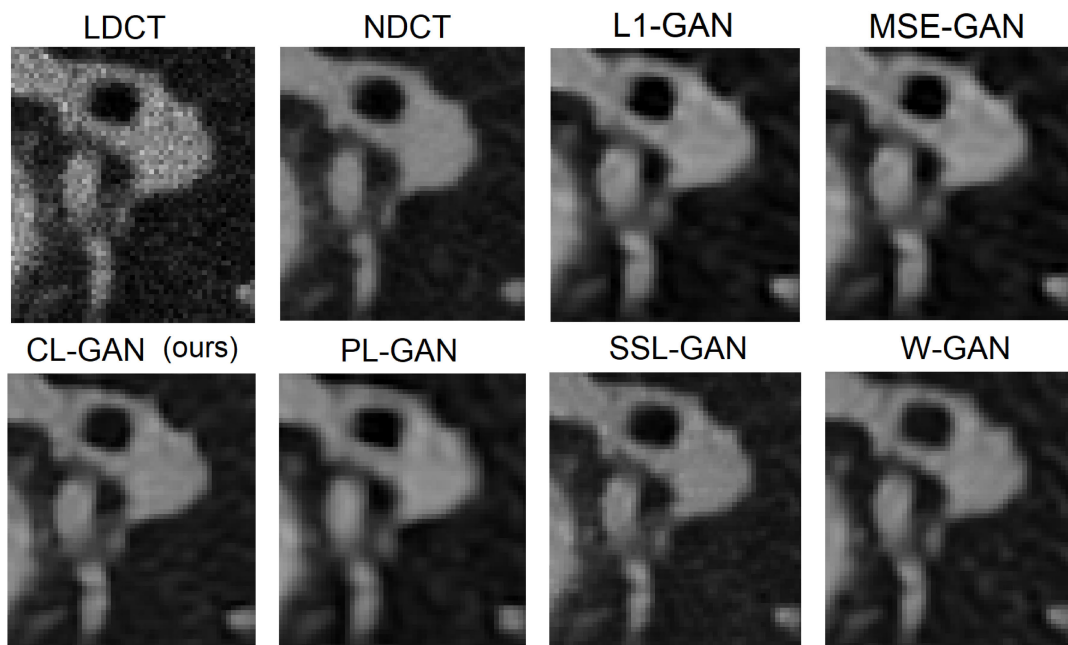

FIGURE 9. Enlarged view of blood vessels (blue bounding box in figure 8) and comparison of result obtained from PL-GAN, SSL-GAN, CL-GAN (ours) on simulated dataset. Display window [−1024, 1024].
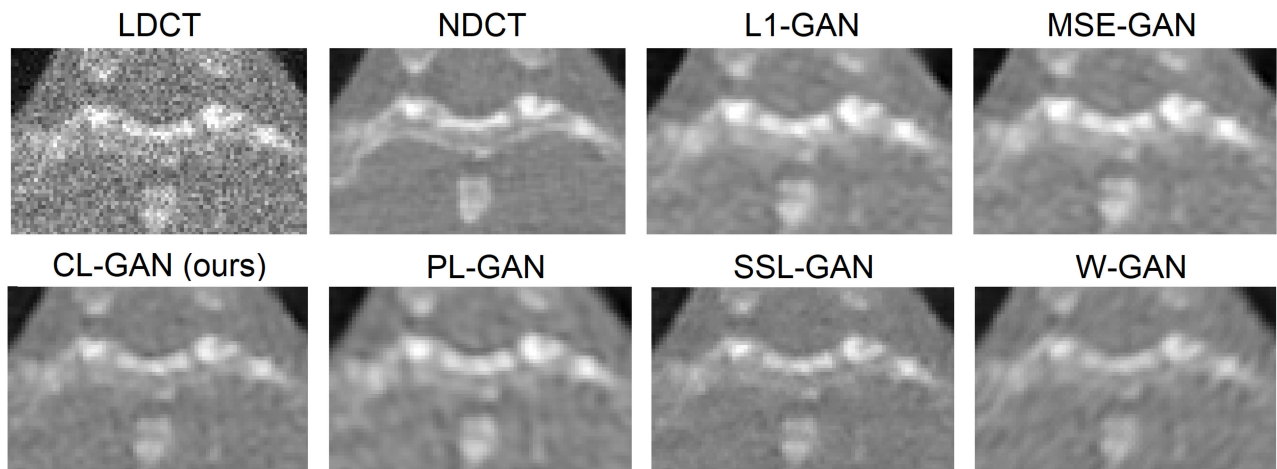
FIGURE 10. Enlarged view of spinal cord (red bounding box in figure 8) and comparison of result obtained from PL-GAN, SSL-GAN, CL-GAN (ours) on simulated dataset. Display window [−1024, 1024].
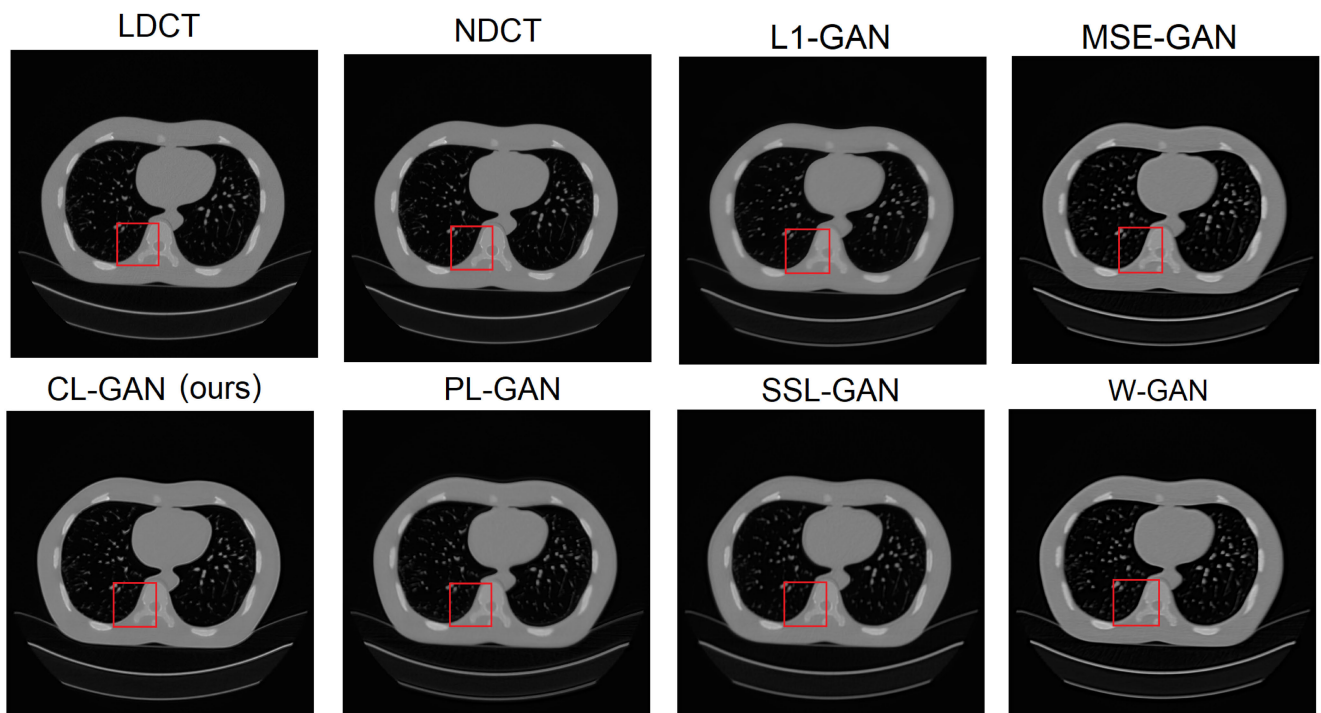

FIGURE 11. Denoising result of axial slice over phantom dataset. Display window [−1024, 1024].

SSL-GAN, and W-GAN. For the Kaggle dataset with simulated Poisson noise, the LDCT represents the NDCT image with added noise. Over the Phantom FDA dataset, the LDCT represents the image as acquired at 25 mAs, whereas the NDCT represents the most similar slice at 100 mAs after sub-pixel image registration with translation and rotation was performed

We can see in all figures that the GAN algorithms have substantially lower noise and better represent the true NDCT performance relative to the LDCT for both the simulated noise, as well as the true noise Phantom scans. However, we see that there is subtle variation in image quality amongst the different GAN algorithms, qualitatively we believe CL-GAN to produce the highest quality resulting image relative to the other algorithms in comparison.

We see in figures 9 and 10 that CL-GAN has greater textural detail than MSE-GAN, while achieving smoother returns on uniform structures than L1-GAN, although variation amongst GAN denoising results is relatively subtle over the simulated noise imagery.

In Figure 12 we see that over the phantom vessels, that the CL-GAN is highly accurate at determining both the shape and size of the blood vessels, and reproduces much of the bone structure. We see that MSE-GAN and W-GAN appear to
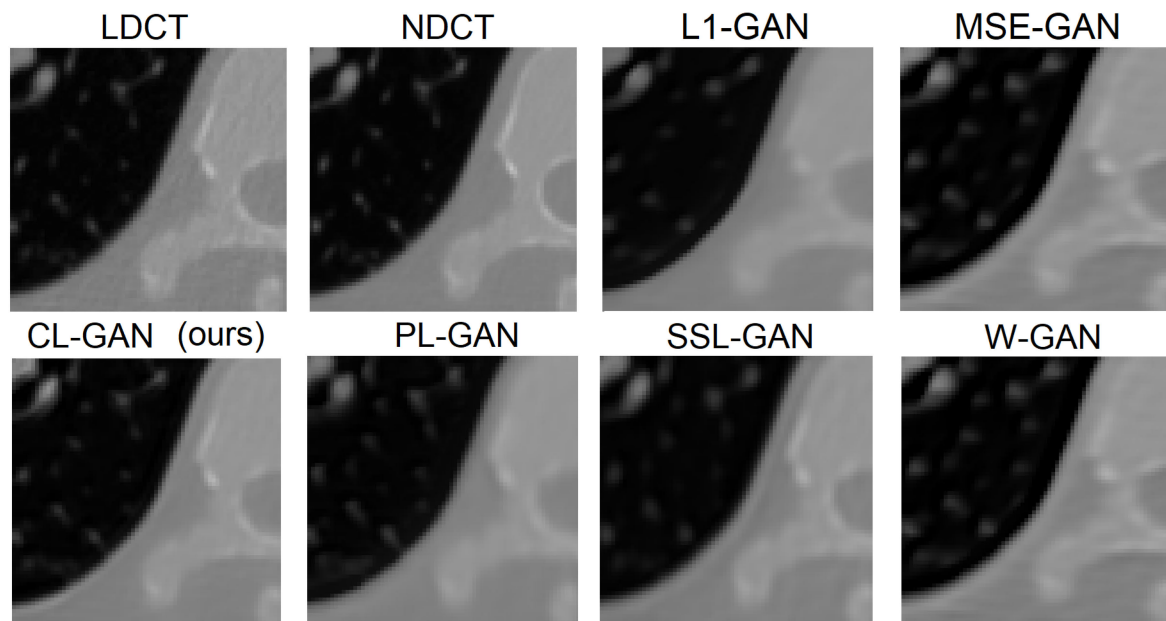
FIGURE 12. Enlarged view of denoising result over axial slice in phantom dataset. Display window [−1024, 1024].

qualitatively over-blur the resulting image, and that PL-GAN and SSL-GAN appear to have less ability to reconstruct high-contrast details of the spinal cord relative to CL-GAN. Although all GAN algorithms in this comparison achieve greater quality than LDCT without increasing radiation exposure, we see that the quality improvement is easier to discern over the Phantom slices potentially due to the higher quality of the LDCT imagery providing more apriori information to the GANs of small scale structures.

Qualitatively, we see that over the simulated imagery, all GAN algorithms appear to improve the image quality, whereas over the Phantom imagery the LDCT scan is relatively higher quality and as such we see that overblurring of edges may cause image qualty to decrease slightly with several GAN algorithms, although the CL-GAN appears able to suppress noise without overblurring the edges or high-contrast features.

B. QUANTITATIVE ANALYSIS AND COMPARISON

Figures 13 and 14 show the PSNR and SSIM metrics respectively of the generated images for 82 test slices from 4 different patients for each model in the comparison. Figures 15 and 16 show the PSNR and SSIM metrics respectively for 300 slices taken from Phantom dataset.

It can be seen that the proposed CL-GAN outperforms all other GAN algorithms in terms of PSNR on both the simulated noise dataset as well as the Phantom FDA scans. We also see that the CL-GAN outperforms all other GAN algorithms in SSIM on the simulated dataset, and outperforms all other GAN algorithms except possibly L1-GAN in terms of SSIM on the Phantom dataset.

We see that on the Simulated noise dataset in Figures 13 and 14 that the that L1-GAN and MSE-GAN are not competitive with the more sophisticated algorithms CL-GAN, PL-GAN and SSL-GAN. In figure 13 we can see that the proposed CL-GAN was able to produce the highest PSNR when compared to these other methods, followed by PL-GAN [30], SSL-GAN [25], and W-GAN. The MSE-GAN and L1-GAN achieved a lower PNSR. Figure 14 shows the comparison of the Structural Similarity Index (SSIM) which is a perceptual metric for image quality over the simulated noise dataset. We see that the proposed CL-GAN outperforms the other methods in accordance with this metric as well. Overall, the ordering of the results with SSIM is largely similar to the PSNR results, although we see that W-GAN performs more competitively in comparison to the PL-GAN [30] and SL-GAN [25] models for this evaluation metric.

Figure 15 shows the resulting PSNR over the Phantom FDA imagery, for which the 25 mAs scan was used as LDCT, and 100 mAs scan was used as NDCT. We see that CL-GAN is the highest performing algorithm followed by SSL-GAN ad L1-GAN, then PL-GAN, MSE-GAN, and W-GAN achieve lower performance. This ordering is similar to that of the Simulated datset, except that the L1-GAN achieves comparably higher relative performance than on the Simulated dataset, but still below that of CL-GAN. We see similar results when we apply the SSIM metric to the Phantom FDA imagery as seen in Figure 16 with two notable exceptions, firstly the LDCT scan itself achieves higher SSIM than the lower performing MSE-GAN and W-GAN, and this can be attributable because the LDCT scan at 25 mAs has much lower noise than the simulated imagery. A second difference is that, although
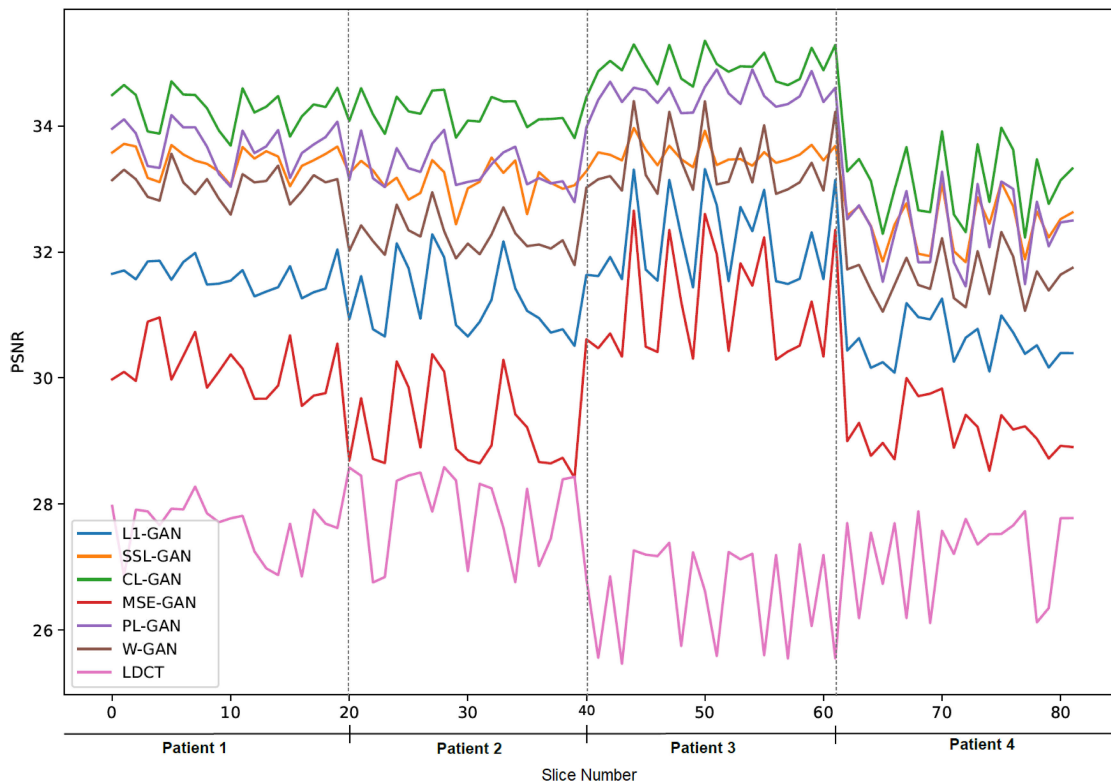
FIGURE 13. PSNR obtained during testing on simulated dataset for four patients.
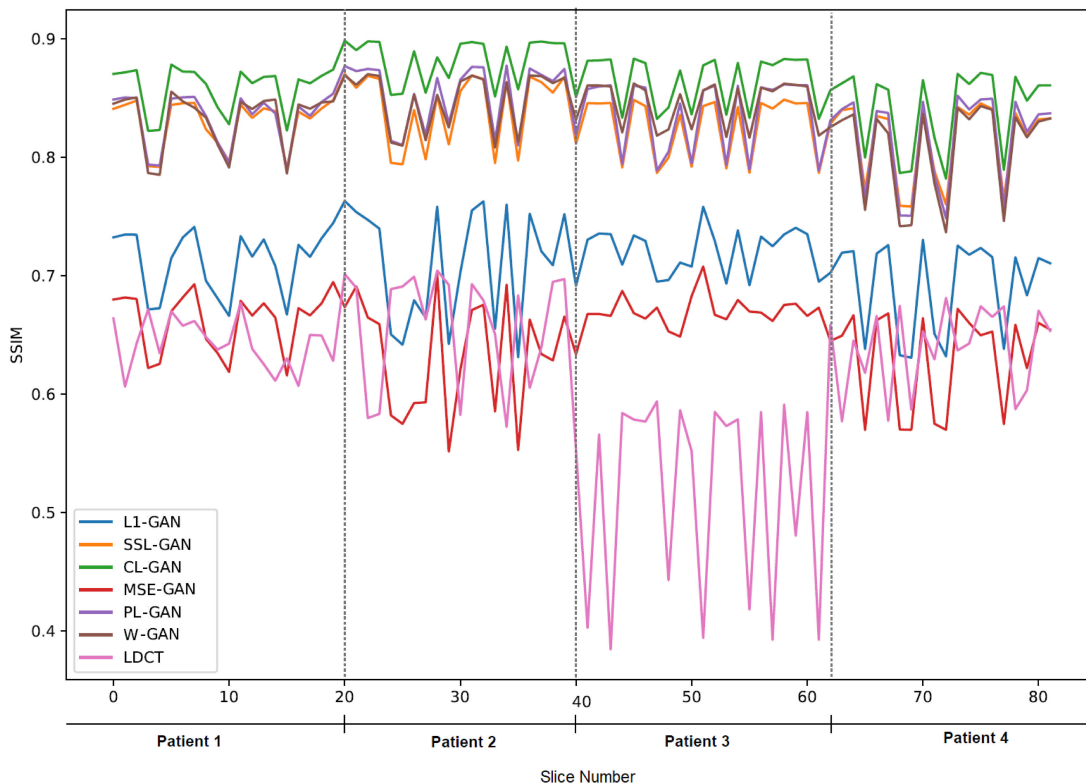

FIGURE 14. SSIM obtained during testing on simulated dataset for four patients.

CL-GAN achieves higher performance than the other methods in terms of SSIM, the L1-GAN and SSL-GAN are competitive. As SSIM is a perceptual metric, it is possible that the noise within the LDCT scan is relatively less perceptible than other artifacts such as overblurring of edges which occurs in the MSE-GAN and W-GAN techniques.
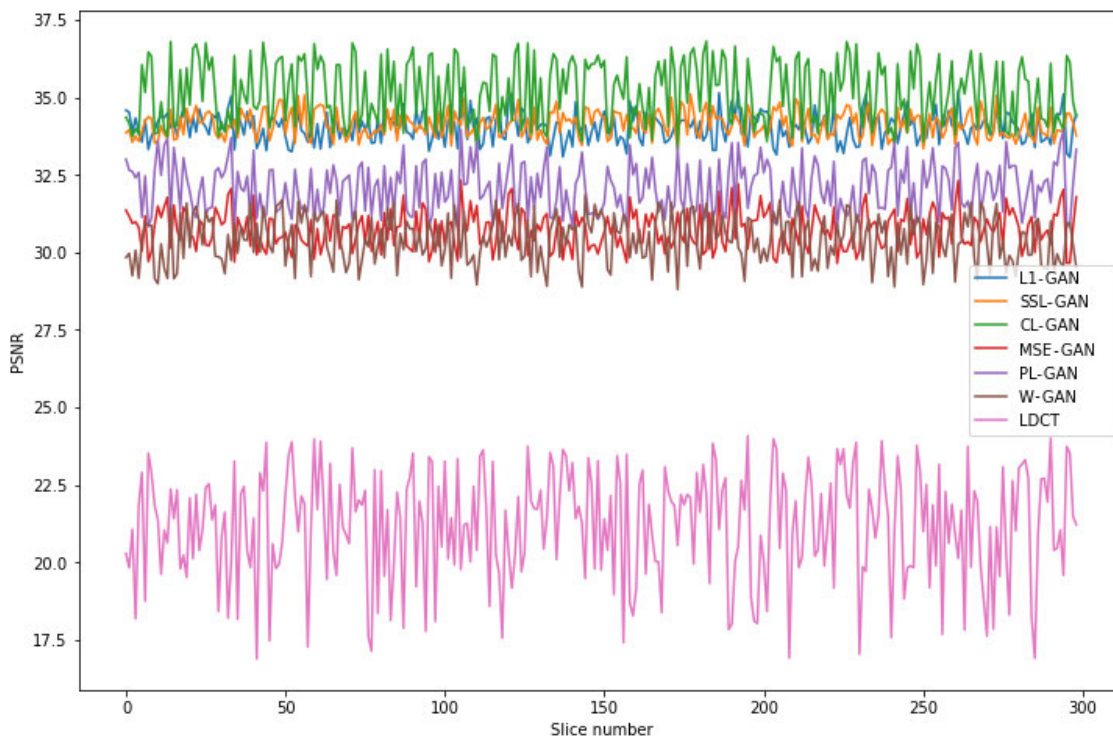
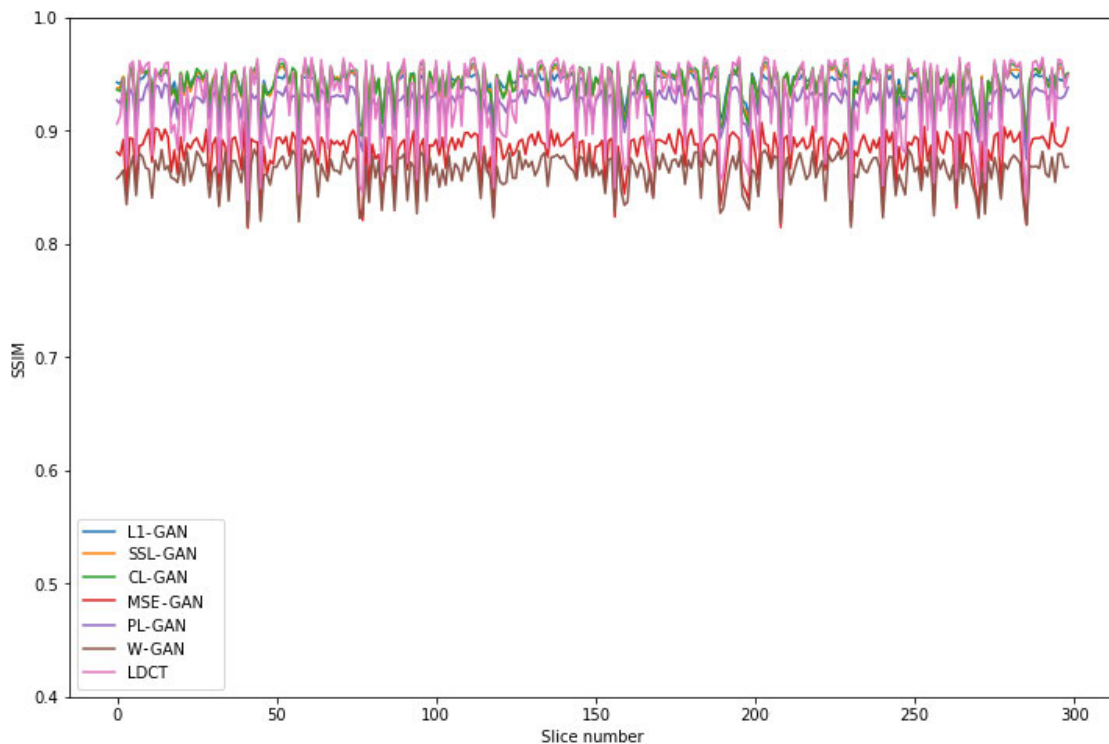**FIGURE 15.** PSNR obtained during testing on Phantom Dataset.

FIGURE 16. SSIM obtained during testing on Phantom Dataset.

MSE loss in particular has been shown by several researchers to produce overly blurry images, particularly over edges [22], [30], thereby reducing PSNR and SSIM. CL-GAN improves this performance because Charbonnier loss is an M-estimator that approximates MSE over small deviations.

C. SOFT TISSUE NOISE METRICS

We now report the soft tissue noise using the method of Kamalanathan *et al.* based on the work of Samei *et al.* [42], [44]. These results are consistent with the quantitative results obtained using PSNR and SSIM as shown in the previous sub-sections.

TABLE 2. Noise Algorithm results on figure 8.

	SD Noise
LDCT	169.41
NDCT	42.93
CL-GAN	46.13
SSL-GAN	47.01
PL-GAN	52.42
MSE-GAN	51.90
L1-GAN	51.43
W-GAN	41.96

In table 2 we see the soft tissue noise over the slice presented in Fig. 8. In an ideal scenario, a denoising algorithm should achieve the soft tissue noise which is as close as possible to that of the NDCT ground truth slices. As it can be seen in table 2, CL-GAN provides the lowest soft tissue noise estimate in this comparison without going lower than the original NDCT image. For completeness, the original LDCT slice with simulated noise is also shown exhibiting a much higher value of soft-tissue noise. The next lowest is the SSL-GAN does not provide close results to us relatively for PSNR and SSIM L1-GAN produces very close results to MSE-GAN. We see that PL-GAN produces higher soft tissue noise estimates than the other models tested. Finally, W-GAN appears producing values lower than that of the ground truth image which means that the generated images from only using the adversarial loss are over-smoothed which also supports our visual observation.

VII. CONCLUSION

We present an image-to-image deep denoising GAN that improves performance over related methods by incorporating a Charbonnier structural loss term. Image-to-image denoising is the most heavily studied approach to deep reconstruction, and denoising GANs are at present the most accurate technique. The proposed CL-GAN makes use of Charbonnier structural loss in combination with Wasserstein adversarial loss and VGG perceptual loss. The Charbonnier loss is a variant of Huber loss known as pseudo-Huber loss. It acts as a pixelwise regularizer to constrain the generated images to closely approximate the ground truth over the training dataset. Charbonnier structural loss is less sensitive to bimodal distributions that can occur over tissue boundaries as compared to MSE pixelwise loss. This property thereby results in a greater ability to suppress noise without compromising structural detail. In conclusion, we demonstrate that the CL-GAN architecture which incorporates Charbonnier structural loss is a state of the art approach for the purposes of improving the signal to noise ratio of CT-scans without increasing radiation exposure to the patient.

APPENDIX A
VARIANCE/BIAS TRADEOFF OF CT DENOISING FILTERS

We now describe a theoretical variance/bias tradeoff between the use of the popular L1 and L2 norms for CT denoising, and which justifies the use of a hybrid Charbonnier loss in order to achieve a balance between these factors to reduce reconstruction error. Our intent for this analysis not to model the complexities of neural architectures or GANs, but rather to formalize our argument that under basic assumptions a simple non-linear Charbonnier filtering kernel would outperform a L1 (median) kernel, or L2 (mean) filtering kernel for the purposes of image denoising. In particular we wish to demonstrate that a Charbonnier loss filter would produce lower *MSE* over edges than L2 loss, and lower *MSE* over smooth regions than L1 loss, and furthermore would produce good values in both situations. Finally, we demonstrate that a properly tuned Charbonnier loss kernel has the ability to outperform even the L1 loss kernel over edge boundaries.

We assume a piecewise smooth model of a CT image with white gaussian noise. Gaussian noise is a reasonable first order approximation of Poisson noise due to the normal approximation of the Poisson distribution for photon counts greater than 20. The piecewise smooth model is reasonable to first order because signal variation within tissue regions is often of comparable magnitude to noise power. We consider cases of pixels within a smooth region, as well as pixels that cross the edge boundaries between regions.

We define the unobserved true radiodensity of the target pixel as μ_1, and define the radiodensity estimate of the denoising filter as $\hat{\mu}$. We assume that the denoising filter must estimate μ_1 as closely as possible given a sample of N surrounding radiodensities $X_1 \ldots X_N$. The N pixels follow bimodal Normal distribution between two regions of nominal radiodensities μ_1 and μ_2 with noise power σ as follows.

$$X_i \sim \alpha N(\mu_1, \sigma) + (1 - \alpha)N(\mu_2, \sigma) \qquad (18)$$

Equation 18 simplifies to a unimodal Normal distribution in the special case when $\alpha = 1$. The unimodal case is anticipated when sampling over the interior of smooth tissue, whereas the bimodal case $0.5 < \alpha < 1$ is anticipated when sampling over edge boundaries. We constrain $\alpha > 0.5$, such that the true radiodensity μ_1 is the population mode.

When sampling over edge boundaries, we define K as the number of sample inliers as drawn from $N(\mu_1, \sigma)$ and $N - K$ as the number of outliers drawn from $N(\mu_2, \sigma)$ due to contamination of the sample by the adjacent smooth region. The value K varies from region to region, but does not vary from denoising kernel to denoising kernel. i.e. the L1 filter, L2 filter, and Charbonnier filters all encounter the same number of outliers when sampling the same region of tissue. As such, we can reorder this sample such that the first K samples are the inliers drawn from the target smooth region and the remaining $N - K$ are drawn from the adjacent smooth region as follows,

$$X_1 \ldots X_K \sim N(\mu_1, \sigma)$$
$$X_{K+1} \ldots X_N \sim N(\mu_2, \sigma) \qquad (19)$$

A denoising kernel must attempt to estimate population mode μ_1 given K observed samples $X_1..X_K$. This kernel will produce a sample estimate $\hat{\mu}$, which minimizes the loss L to K candidate values as follows,

$$\underset{\hat{\mu}}{\text{argmin}} \sum_{i=1}^{K} L(\mu - X_i) \quad (20)$$

The loss L could be either the L1, L2, or Charbonnier functions. It is well known that for the L1 loss, $\hat{\mu}$ simplifies to the sample median, whereas for the L2 loss $\hat{\mu}$ becomes the sample mean. However, for the Charbonnier loss, estimation of $\hat{\mu}$ to minimize equation 20 does not have explicit form. Rather, this minimum must be calculated with convex numerical methods.

We must quantify and compare the bias and variance of each of the proposed kernel loss functions for estimating the true tissue radiodensity μ_1. Bias B is defined as the expected difference between the sample estimate and the true radiodensity as follows,

$$B = E(\hat{\mu}) - \mu_1 \quad (21)$$

Variance V is defined as the square error of the sample estimate relative to the sample mean as follows,

$$V = E(\hat{\mu} - E(\hat{\mu}))^2 \quad (22)$$

A suitable estimator should strive to reduce bias and variance simultaneously. The expected MSE of an estimator is equally affected by the sample square bias as well as the sample variance as follows.

$$MSE = E(\hat{mu} - \mu_1)^2 = B^2 + V + \sigma^2 \quad (23)$$

For any given value of $K > 0$, the L1 norm has lower bias whereas the L2 norm has lower variance. For $K = 0$ the L1 and L2 norms have zero bias but the L2 norm has lower variance. Given K, the samples $X_1 \dots X_K$ and $X_{K+1} \dots X_N$ can be seen as drawn from two unimodal distributions with means μ_1 and μ_2.

A. BIAS OF L1 AND L2 NORMS
The L2 norm has a relatively high bias as it is influenced by the $N - K$ outliers as drawn from the adjacent tissue as follows.

$$B_{L2} = \frac{N-K}{N}(\mu_2 - \mu_1) \quad (24)$$

The L1 norm is a robust estimator and is therefore not heavily influenced by outliers so long as $K > N/2$. The expected value of $\hat{\mu}_{L1}$ is the sample median, but in the presence of outliers even the median is biased relative to the population mode μ_1. The expected value of the median is the 50th percentile of a bimodal PDF as follows,

$$\int_{-\infty}^{E(\hat{\mu})} b(x)dx = \frac{1}{2} \quad (25)$$

where,

$$b(x) = \frac{K}{N}f(x, \mu_1, \sigma) + \frac{N-K}{N}f(x, \mu_2, \sigma) \quad (26)$$

Assuming that $\mu_2 > \mu_1 + 2\sigma$, the bias of the L1 estimator is well approximated by adjusting the percentile of the dominant distribution as follows,

$$\int_{-\infty}^{E(\hat{\mu})} f(x, \mu_1, \sigma)dx \approx \frac{N}{2K} \quad (27)$$

Under this approximation the bias of the L1 estimator simplifies to the following which depends only on σ and is robust to large differences in the value of μ_2.

$$B_{L1} \approx \sigma \sqrt{2} \, \text{erf}^{-1}\left(\frac{N-K}{K}\right) \quad (28)$$

In order to compare bias, note that $\text{erf}^{-1}(\lambda) \approx \lambda$ for $-\frac{3}{4} \le a \le \frac{3}{4}$. Therefore, for $K > \frac{4}{7}N$ and for $\mu_2 - \mu_1 > 2\sigma$, we arrive at the following ratio of bias between L2 and L1 norms,

$$\frac{B_{L2}}{B_{L1}} \approx \frac{\mu_2 - \mu_1}{\sigma\sqrt{2}}$$

In the case $\mu_2 - \mu_1 \approx 2\sigma$, then the bias between the L2 and L1 norms differ by approximately a factor of $\sqrt{2}$ as follows,

$$B_{L2} \approx \sqrt{2} \, B_{L1} \quad (29)$$

B. VARIANCE OF L1 AND L2 NORMS
Variance of the L1 and L2 norms is straightforward to calculate due to our decomposition of the sample into two unimodal normal distributions of equal variance. As their are $X_1 \dots X_K \, N(\mu_1, \sigma)$ and $X_{K+1}..X_N \, N(\mu_2, \sigma)$, the sample variance follows a t distribution that is approximately normal for large N as follows,

$$V_{L2} = \frac{\sigma_2}{N} \quad (30)$$

We now proceed to show that under basic assumptions V_{L2} is proportional yet larger than V_{L1}. We assume as before that $K > \frac{4}{7}N$, and $\mu_2 - \mu_1 > 2\sigma$ such that the sample median of the bimodal distribution is well approximated by a sample $\frac{N}{2K}$ percentile of the dominant unimodal distribution. For large K we may apply the central limit theorem of medians and percentiles, in that in which the variance of the $q = \frac{N}{2K}^{th}$ percentile converges to

$$V_{L1} = \frac{N}{2K}\frac{2K-N}{2K}\frac{1}{Kf^2(x_q, \mu_1, \sigma)}$$
$$where \quad x_q = \mu_1 + \sigma\sqrt{2} \, erf^{-1}(\frac{N-K}{K}) \quad (31)$$

Note again that due to Taylor expansion $erf^{-1}(\lambda) \approx \lambda$ for $-\frac{3}{4} \le \lambda \le \frac{3}{4}$ which holds when $K > \frac{4}{7}N$ as assumed. As such, the above equation simplifies to the following,

$$V_{L1} \approx \frac{(2KN - N^2)\sigma^2 2\pi}{4K^3 e^{-2(\frac{N-K}{K})^2}} \quad (32)$$

We may substitute $K = \beta N$ where $\frac{4}{7} < \beta < 1$. In doing so, the asymptotic variance of the $L1$ estimator simplifies to a constant function of β times the asymptotic variance of the $L2$ estimator as follows,

$$V_{L1} \approx \frac{(2\beta - 1)\pi}{2\beta^3 e^{-2(1-\beta)^2}} V_{L2} \qquad (33)$$

Numerical evaluation of this coefficient over the valid interval $\frac{4}{7} < \beta < N$ demonstrates that the $L1$ denoising kernal exhibits proportional yet greater variance than the $L2$ kernel as follows,

$$\frac{\pi}{2} V_{L2} \le V_{L1} \le 2.21 V_{L2} \qquad (34)$$

REFERENCES

[1] A. Lucas, S. Lopez-Tapia, R. Molina, and A. K. Katsaggelos, "Generative adversarial networks and perceptual losses for video super-resolution," *IEEE Trans. Image Process.*, vol. 28, no. 7, pp. 3312–3327, Jul. 2019.

[2] L. W. Goldman, "Principles of CT: Radiation dose and image quality," *J. Nucl. Med. Technol.*, vol. 35, no. 4, pp. 213–225, Dec. 2007.

[3] R. Gordon, R. Bender, and G. T. Herman, "Algebraic reconstruction techniques (ART) for three-dimensional electron microscopy and X-ray photography," *J. Theor. Biol.*, vol. 29, no. 3, pp. 471–481, Dec. 1970.

[4] G. N. Hounsfield, "Computerized transverse axial scanning (tomography): Part 1. Description of system," *Brit. J. Radiol.*, vol. 46, no. 552, pp. 1016–1022, Dec. 1973.

[5] C. J. Dasch, "One-dimensional tomography: A comparison of Abel, onion-peeling, and filtered backprojection methods," *Appl. Opt.*, vol. 31, no. 8, pp. 1146–1152, 1992.

[6] M. Beister, D. Kolditz, and W. A. Kalender, "Iterative reconstruction methods in X-ray CT," *Phys. Medica*, vol. 28, no. 2, pp. 94–108, Apr. 2012.

[7] B. R. Whiting, P. Massoumzadeh, O. A. Earl, J. A. O'Sullivan, D. L. Snyder, and J. F. Williamson, "Properties of preprocessed sinogram data in X-ray computed tomography," *Med. Phys.*, vol. 33, no. 9, pp. 3290–3303, Aug. 2006.

[8] I. A. Elbakri and J. A. Fessler, "Statistical image reconstruction for polyenergetic X-ray computed tomography," *IEEE Trans. Med. Imag.*, vol. 21, no. 2, pp. 89–99, Aug. 2002.

[9] S. Ramani and J. A. Fessler, "A splitting-based iterative algorithm for accelerated statistical X-ray CT reconstruction," *IEEE Trans. Med. Imag.*, vol. 31, no. 3, pp. 677–688, Mar. 2012.

[10] E. Y. Sidky and X. Pan, "Image reconstruction in circular cone-beam computed tomography by constrained, total-variation minimization," *Phys. Med. Biol.*, vol. 53, no. 17, p. 4777, Sep. 2008.

[11] Q. Xu, H. Yu, X. Mou, L. Zhang, J. Hsieh, and G. Wang, "Low-dose X-ray CT reconstruction via dictionary learning," *IEEE Trans. Med. Imag.*, vol. 31, no. 9, pp. 1682–1697, Sep. 2012.

[12] D. Karimi, P. Deman, R. Ward, and N. Ford, "A sinogram denoising algorithm for low-dose computed tomography," *BMC Med. Imag.*, vol. 16, no. 1, p. 11, Dec. 2016.

[13] A. Manduca, L. Yu, J. D. Trzasko, N. Khaylova, J. M. Kofler, C. M. McCollough, and J. G. Fletcher, "Projection space denoising with bilateral filtering and CT noise modeling for dose reduction in CT," *Med. Phys.*, vol. 36, no. 11, pp. 4911–4919, Oct. 2009.

[14] E. Kang, J. Min, and J. C. Ye, "A deep convolutional neural network using directional wavelets for low-dose X-ray CT reconstruction," *Med. Phys.*, vol. 44, no. 10, pp. e360–e375, Oct. 2017.

[15] H. Zhang, L. Li, K. Qiao, L. Wang, B. Yan, L. Li, and G. Hu, "Image prediction for limited-angle tomography via deep learning with convolutional neural network," 2016, *arXiv:1607.08707*. [Online]. Available: http://arxiv.org/abs/1607.08707

[16] K. H. Jin, M. T. McCann, E. Froustey, and M. Unser, "Deep convolutional neural network for inverse problems in imaging," *IEEE Trans. Image Process.*, vol. 26, no. 9, pp. 4509–4522, Sep. 2017.

[17] Y. S. Han, J. Yoo, and J. C. Ye, "Deep residual learning for compressed sensing CT reconstruction via persistent homology analysis," 2016, *arXiv:1611.06391*. [Online]. Available: http://arxiv.org/abs/1611.06391

[18] H. Chen, Y. Zhang, M. K. Kalra, F. Lin, Y. Chen, P. Liao, J. Zhou, and G. Wang, "Low-dose CT with a residual encoder-decoder convolutional neural network," *IEEE Trans. Med. Imag.*, vol. 36, no. 12, pp. 2524–2535, Dec. 2017.

[19] H. Shan, Y. Zhang, Q. Yang, U. Kruger, M. K. Kalra, L. Sun, W. Cong, and G. Wang, "3-D convolutional encoder-decoder network for low-dose CT via transfer learning from a 2-D trained network," *IEEE Trans. Med. Imag.*, vol. 37, no. 6, pp. 1522–1534, Jun. 2018.

[20] L. Cheng, S. Ahn, S. G. Ross, H. Qian, and B. D. Man, "Accelerated iterative image reconstruction using a deep learning based leapfrogging strategy," in *Proc. Int. Conf. Fully Three-Dimensional Image Reconstruct. Radiol. Nucl. Med.*, 2017, pp. 715–720.

[21] H. Chen, Y. Zhang, Y. Chen, J. Zhang, W. Zhang, H. Sun, Y. Lv, P. Liao, J. Zhou, and G. Wang, "LEARN: Learned experts' assessment-based reconstruction network for sparse-data CT," *IEEE Trans. Med. Imag.*, vol. 37, no. 6, pp. 1333–1347, Jun. 2018.

[22] J. M. Wolterink, T. Leiner, M. A. Viergever, and I. Isgum, "Generative adversarial networks for noise reduction in low-dose CT," *IEEE Trans. Med. Imag.*, vol. 36, no. 12, pp. 2536–2545, Dec. 2017.

[23] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," 2015, *arXiv:1512.03385*. [Online]. Available: https://arxiv.org/abs/1512.03385

[24] C. You, L. Yang, Y. Zhang, and G. Wang, "Low-dose CT via deep CNN with skip connection and network in network," 2018, *arXiv:1811.10564*. [Online]. Available: http://arxiv.org/abs/1811.10564

[25] C. You, W. Cong, G. Wang, Q. Yang, H. Shan, L. Gjesteby, G. Li, S. Ju, Z. Zhang, Z. Zhao, and Y. Zhang, "Structurally-sensitive multi-scale deep neural network for low-dose CT denoising," *IEEE Access*, vol. 6, pp. 41839–41855, 2018.

[26] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: From error visibility to structural similarity," *IEEE Trans. Image Process.*, vol. 13, no. 4, pp. 600–612, Apr. 2004.

[27] E. Kang, H. J. Koo, D. H. Yang, J. B. Seo, and J. C. Ye, "Cycle-consistent adversarial denoising network for multiphase coronary CT angiography," *Med. Phys.*, vol. 46, no. 2, pp. 550–562, Feb. 2019.

[28] X. Yi and P. Babyn, "Sharpness-aware low-dose CT denoising using conditional generative adversarial network," *J. Digit. Imag.*, vol. 31, no. 5, pp. 655–669, Oct. 2018.

[29] C. Frogner, C. Zhang, H. Mobahi, M. Araya, and T. A. Poggio, "Learning with a Wasserstein loss," in *Proc. Adv. Neural Inf. Process. Syst.*, 2015, pp. 2053–2061.

[30] Q. Yang, P. Yan, Y. Zhang, H. Yu, Y. Shi, X. Mou, M. K. Kalra, Y. Zhang, L. Sun, and G. Wang, "Low-dose CT image denoising using a generative adversarial network with Wasserstein distance and perceptual loss," *IEEE Trans. Med. Imag.*, vol. 37, no. 6, pp. 1348–1357, Jun. 2018.

[31] V. Nair and G. E. Hinton, "Rectified linear units improve restricted Boltzmann machines," in *Proc. 27th Int. Conf. Mach. Learn. (ICML)*, 2010, pp. 807–814.

[32] J. Johnson, A. Alahi, and L. Fei-Fei, "Perceptual losses for real-time style transfer and super-resolution," in *Proc. Eur. Conf. Comput. Vis.* Amsterdam, The Netherlands: Springer, Oct. 2016, pp. 694–711.

[33] M. Arjovsky, S. Chintala, and L. Bottou, "Wasserstein generative adversarial networks," in *Proc. Int. Conf. Mach. Learn.*, Jul. 2017, pp. 214–223.

[34] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative adversarial nets," in *Proc. Adv. Neural Inf. Process. Syst.*, Montreal, QC, Canada, Dec. 2014, pp. 2672–2680.

[35] C. Villani, *Optimal Transport: Old and New*, vol. 338. Berlin, Germany: Springer, 2009, p. 23.

[36] I. Gulrajani, F. Ahmed, M. Arjovsky, V. Dumoulin, and A. C. Courville, "Improved training of Wasserstein GANs," in *Proc. Adv. Neural Inf. Process. Syst.*, 2017, pp. 5767–5777.

[37] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," 2014, *arXiv:1409.1556*. [Online]. Available: http://arxiv.org/abs/1409.1556

[38] P. Charbonnier, L. Blanc-Feraud, G. Aubert, and M. Barlaud, "Two deterministic half-quadratic regularization algorithms for computed imaging," in *Proc. 1st Int. Conf. Image Process.*, 1994, pp. 168–172.

[39] (2017). *Data Science Bowl*. [Online]. Available: https://www.kaggle.com/c/data-science-bowl-2017/data

[40] M. A. Gavrielides, L. M. Kinnard, K. J. Myers, J. Peregoy, W. F. Pritchard, R. Zeng, J. Esparza, J. Karanian, and N. Petrick, "Data from phantom FDA. The cancer imaging archive," Nat. Cancer Inst., Bethesda, MD, USA, Tech. Rep., 2015, doi: 10.7937/K9/TCIA.2015.ORBJKMUX.

[41] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," 2014, *arXiv:1412.6980*. [Online]. Available: http://arxiv.org/abs/1412.6980

[42] X. Tian and E. Samei, "Accurate assessment and prediction of noise in clinical CT images," *Med. Phys.*, vol. 43 no. 1, pp. 475–482, 2016.

[43] *Peak Signal-to-Noise Ratio as an Image Quality Metric*. Accessed: Jun. 2020. [Online]. Available: https://www.ni.com/en-us/innovations/white-papers/11/peak-signal-to-noi%se-ratio-as-an-image-quality-metric.html

[44] O. Christianson, J. Winslow, D. P. Frush, and E. Samei, "Automated technique to measure noise in clinical CT examinations," *Amer. J. Roentgenology*, vol. 205, no. 1, pp. W93–W99, Jul. 2015.

BINIT GAJERA received the B.Tech. degree in computer engineering from Dharmsinh Desai University, Nadiad, India, in 2018, and the M.S. degree in computer science from the University of Maryland, Baltimore County, Baltimore, MD, USA, in 2020.

From 2017 to 2018, he was a Software Engineer Intern with Trueline Solution, India. He is currently a Software Engineer with Softheon Inc., Stony Brook, NY, USA. His research interests include machine learning, medical imaging, and computer vision.
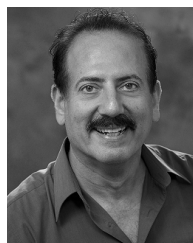
SIDDHANT RAJ KAPIL received the B.Tech. degree in computer engineering from Guru Gobind Singh Indraprastha University, in 2016. He is currently pursuing the M.S. degree in computer science with the University of Maryland, Baltimore County, Baltimore, MD, USA.

From 2016 to 2019, he worked with Nagarro as a Data Scientist and solved several AI use-cases for Nagarro's clients. He also won the Nagarro Excellence Award. His research interests include machine learning, object tracking, and computer vision.

DORSA ZIAEI received the B.S. degree in computer science from Azad University, Tehran, in 2005, the M.S. degree in computer science from the Islamic Azad University, Science and Research Branch, Tehran, in 2009, and the M.S. degree in computer science from the University of Maryland, Baltimore County, in 2017, where she is currently pursuing the Ph.D. degree in computer science.

She has won several awards, including an ORISE Fellow at Food and Drug Organization for project: Characterization of color normalization methods in digital pathology whole slide imaging. She is the winner of the SPIE Medical Imaging 2020, "Image Processing Conference," Best Paper Poster Award for "Segmentation of stem cell colonies in fluorescence microscopy images with transfer learning." She is also the winner of the F1000 Award: Outstanding Presentation Prize, "NYC Symposium: Deep Learning for Drug Discovery" for the project "Assessment of Deep Convolutional Neural Networks: Segmentation of Large High-Resolution Stem Cell Images."

JAYALAKSHMI MANGALAGIRI received the B.Tech. degree in information technology from GITAM University, India, in 2014, the master's degree in computing and information systems from Youngstown State University, USA, in 2016, and the Ph.D. degree in information technology from the University of Cumberlands, in 2019. She is currently pursuing the Ph.D. degree in computer science with the University of Maryland, Baltimore County.

Her research interests include machine learning and its applications to medicine and health care. She is also studying artificial intelligence, image processing, computer vision, data analysis, and visualization.

ELIOT SIEGEL received the B.Sc. degree in computer science from the University of Maryland, College Park, MD, USA, in 1978, and the M.D. degree from the University of Maryland School of Medicine, in 1982. He completed his residency in radiology and nuclear medicine from the University of Maryland Medical Center, in 1987.

He is currently the Vice Chair of the Department of Diagnostic Radiology, University of Maryland School of Medicine, and the Chief of Radiology and Nuclear Medicine for the Veterans Affairs Maryland Healthcare System. He is also the Director of the Maryland Imaging Research Technologies Laboratory, and has adjunct appointments as a Professor of Bioengineering with the University of Maryland, and as a Professor of Computer Science with the University of Maryland, Baltimore County.

DAVID CHAPMAN received the B.S., M.S., and Ph.D. degrees in computer science from the University of Maryland, Baltimore County (UMBC), in 2006, 2008, and 2012, respectively.

From 2012 to 2014, he was a Postdoctoral Fellow with the Lamont Doherty Earth Observatory (LDEO), Columbia University. From 2014 to 2018, he was a Design Engineer with Oceaneering International, Inc. He is currently an Assistant Professor with the Department of Computer Science, University of Maryland, Baltimore County, and the Head of the Vision and Image Processing Algorithms Research Group (VIPAR). VIPAR focuses on computer visions, machine learning, and medical imaging informatics. VIPAR is working in collaboration with the University of Maryland School of Medicine, the University of California San Francisco, and the Mercy Medical Center in Baltimore on problems related to AI in radiologicial imaging.

• • •