CrossMark

# Sharpness-Aware Low-Dose CT Denoising Using Conditional Generative Adversarial Network

Xin Yi[1] · Paul Babyn[1]

## Abstract

Low-dose computed tomography (LDCT) has offered tremendous benefits in radiation-restricted applications, but the quantum noise as resulted by the insufficient number of photons could potentially harm the diagnostic performance. Current image-based denoising methods tend to produce a blur effect on the final reconstructed results especially in high noise levels. In this paper, a deep learning-based approach was proposed to mitigate this problem. An adversarially trained network and a sharpness detection network were trained to guide the training process. Experiments on both simulated and real dataset show that the results of the proposed method have very small resolution loss and achieves better performance relative to state-of-the-art methods both quantitatively and visually.

**Keywords** Low-dose CT · Denoising · Conditional generative adversarial networks · Deep learning · Sharpness · Low contrast

## Introduction

The use of computed tomography (CT) has rapidly increased over the past decade, with an estimated 80 million CT scans performed in 2015 in the USA [8]. Although CT offers tremendous benefits, its use has lead to significant concern regarding radiation exposure. To address this issue, the as low as reasonably achievable (ALARA) principle has been adopted to avoid excessive radiation dose for the patient.

Diagnostic performance should not be compromised when lowering the radiation dose. One of the most effective ways to reduce radiation dose is to reduce tube current, which has been adopted in many imaging protocols. However, low-dose CT (LDCT) inevitably introduces more noise than conventional CT (convCT), which may potentially impede subsequent diagnosis or require more advanced algorithms for reconstruction. Many works have been devoted to CT denoising with promising results achieved by a variety of techniques, including those in the image, and sinogram domains and with iterative reconstruction techniques. One recent technique of increasing interest is deep learning (DL).

DL has been shown to exhibit superior performance on many image-related tasks, including low-level edge detection [6], image segmentation [69], and high-level vision problems including image recognition [26] and image captioning [61], with these advances now being brought into the medical domain [12, 13, 33, 67]. In this paper, we explore the possibility of applying generative adversarial neural net (GAN) [22] to the task of LDCT denoising.

In many image-related reconstruction tasks, e.g. super resolution and inpainting, it is known that minimizing the per-pixel loss between the output image and the ground truth alone generate either blurring or make the result visually not appealing [28, 37, 72]. We have observed the same effect in the traditional neural network-based CT denoising works [12, 13, 33, 67]. The adversarial loss introduced by GAN can be treated as a driving force that can push the generated image to reside in the manifold of convCTs, reducing the blurring effect. Furthermore, an additional sharpness detection network was also introduced to measure the sharpness of the denoised image, with focus on low-contrast regions. SAGAN (sharpness-aware generative adversarial network) will be used to denote this proposed denoising method in the remainder of the paper.

✉ Xin Yi
  xiy525@mail.usask.ca

[1] University of Saskatchewan, College of Medicine, Saskatoon, SK, Canada

## Related Works

**LDCT Denoising** algorithms can be broadly categorized into three groups, those conducted within the sinogram or image domains and iterative reconstruction methods (which iterate back and forth across the sinogram and image domains).

The CT sinogram represents the attenuation line integrals from the radial views and is the raw projection data in the CT scan. Since the sinogram is also a 2-D signal, traditional image processing techniques have been applied for noise reduction, such as bilateral filtering [41], structural adaptive filtering [4], etc. The filtered data can then be reconstructed to a CT image with methods like filtered back projection (FBP). Although the statistical property of the noise can be well characterized, these methods require the availability of the raw data which is not always accessible. In addition, by application of edge preservation smoothing operations (bilateral filtering), small edges would inevitably be filtered out and lead to loss of structure and spatial resolution in the reconstructed CT image.

Note that the above method only performs a single back projection to reconstruct the original image. Another stream of works performs an additional forward projection, mapping the reconstructed image to the sinogram domain by modelling the acquisition process. Corrections can be made by iterating the forward and backward process. This methodology is referred as model-based iterative reconstruction (MBIR). Usually, MBIR methods model scanner geometry and physical properties of the imaging processing, e.g. the photon counting statistics and the polychromatic nature of the source x-ray [5]. Some works add prior object information to the model to regulate the reconstructed image, such as total variation minimization [60, 74], Markov random fields based roughness or a sparsity penalty [7]. Due to its iterative nature, MBIR models tend to consume excessive computation time for the reconstruction. There are works that are trying to accelerate the convergence behaviour of the optimization process, for example, by variable splitting of the data fidelity term [49] or by combining Nesterov's momentum with ordered subsets method [34].

To employ the MBIR method, one also has to have access to the raw sinogram data. Image-based denoising methods do not have this limitation. The input and output are both images. Many of the denoising methods for LDCT are borrowed from natural image processing field, such as Non-Local means [9] and BM3D [18]. The former computes the weighted average of similar patches in the image domain while the latter groups similar 2D noisy image patches into 3D arrays and jointly filtered these patches by performing 3D transform and shrinkage of the coefficients in the transform domain. Both methods assume the redundancy of image information. Inspired by these two seminal works,

many applications have emerged applying them into LDCTs [14, 15, 23, 25, 40, 70, 71]. Another line of work focuses on compressive sensing, with the underlying assumption that every local path can be represented as a sparse combination of a set of bases. In the very beginning, the bases are from some generic analytic transforms, e.g. discrete gradient, contourlet [48], curvelet [10]. Chen et al. built a prior image constrained compressed sensing (PICCS) algorithm for dynamic CT reconstruction under reduced views based on discrete gradient transform [39]. It has been found that these transforms are very sensitive to both true structures and noise. Later on, bases learned directly from the source images were used with promising results. The K-SVD [1] algorithm generalizes the k-means clustering process, which alternates the learning of the bases from the supplied image data and sparse coding of the image based on the learnt bases. The base learning is efficient due to a Gauss-Seidel-like accelerated updating method and has inspired many applications in the medical domain [16, 17, 38, 39, 66].

Convolutional neural network (CNN)-based methods have recently achieved great success in image related tasks. Although its origins can be traced back to the 1980s, the resurgence of CNN can be greatly attributed to increased computational power and recently introduced techniques for efficient training of deep networks, such as BatchNorm [29], Rectifier linear units [20] and residual connection [26]. Chen et al. [12] first used CNN to denoise CT images by learning a patch-based neural net and later on refined it with a encoder and decoder structure for end-to-end training [13]. Kang et al. [33] devised a 24 convolution layer net with by-pass connection and contracting path for denoising but instead of mapping in the image domain, it performed end-to-end training in the wavelet domain. Yang et al. [67] adopted perceptual loss into the training, which measures the difference of the processed image and the ground truth in a high-level feature space projected by a pre-trained CNN. Suzuki et al. [58] proposed to use a massive-training artificial neural network (MTANN) for CT denoising. The network accepts local patches of the LDCT and regressed to the centre value of the corresponding patch of the convCT.

**Generative Adversarial Network** was first introduced in 2014 by Goodfellow et al. [22]. It is a generative model trying to generate real world images by employing a min-max optimization framework where two networks (Generator $G$ and Discriminator $D$) are trained against each other. $G$ tries to synthesize real appearing images from random noise whereas $D$ is trying to distinguish between the generated and real images. If the Generator $G$ get sufficiently well trained, the Discriminator $D$ will eventually be unable to tell if the generated image is fake or not.

The original setup of GAN does not contain any constraints to control what modes of data it can generate.

However, if auxiliary information were provided during the generation, GAN can be driven to output images with specific modes. GAN in this scenario is usually referred to as conditional GAN (cGAN) since the output is conditioned on additional information. Mirza et al. supplied class label encoded as one hot vector to generate MINIST digits [44]. Other works have exploited the same class label information but with different network architecture [45, 46]. Reed et al. fed GAN with text descriptions and object locations [50, 51]. Isola et al. proposed to do a image to image translation with GAN by directly supplying the GAN with images [30]. In this framework, training images must be aligned image pairs. Later on, Zhu et al. relaxed this restriction by introducing the cycle consistency loss so that images can be translated between two sets of unpaired samples [73]. But as also mentioned in their paper, the paired training remains the upper bound. Pathak et al. generated missing image patches conditioned on the surrounding image context [47]. Sangkloy et al. generated images constrained by the sketched boundaries and sparse colour strokes [53]. Shrivastava et al. refined synthetic images with GAN trying to narrowing the gap between the synthetic images and real image [56]. Walker et al. adopted a cGAN by conditioning on the predicted future pose information to synthesize future frames of a video [63].

Two works have also applied cGAN for CT denoising. In both their works, together with ours, the denoised image is generated by conditioning on the low-dose counterparts. Wolterink et al. employed a vanilla cGAN where the generator was a seven-layer all-convolutional network and the discriminator is a network to differentiate the real and denoised cardiac CT using cross entropy loss as the objective function [65]. Yang et al. adopted Wasserstein distance for the loss of the discriminator and incorporated perceptual loss to ensure visual similarity [68]. Using Wasserstein distance was claimed to be beneficial at stabilizing the training of GAN but not claimed to generate images of better quality [2, 24]. Our work differs in many ways and we would like to highlight some key points here. First, in our work, the generator used a U-Net style network with residual components and is deeper than the other two works. The superiority of the proposed architecture in retaining small details was shown in our simulated noise experiment. Second, our discriminator differentiates patches rather than full images which makes the resulted network have fewer parameters and applicable to arbitrary size image. Third, CT scans of a series of dose levels and anatomic regions were evaluated in this work for the generality assessment. Noise and artifacts differ throughout the body. Our work showed that a singe network could potentially denoise all anatomies. Finally, a sharpness loss was introduced to ensure the final sharpness of the image and the faithful reconstruction of low-contrast regions.

**Sharpness Detection** The sharpness detection network should be sensitive to low-contrast regions. Traditional methods based on local image energy have intrinsic limitations which is the sensitivity to both the blur kernel and the image content. Recent works have proposed more sophisticated measures by exploiting the statistic differences of specific properties of blur and sharp region, e.g. gradient [55], local binary pattern [69], power spectrum slope [55, 62], discrete cosine transform (DCT) coefficient [21]. Shi et al. used sparse coding to decompose local path and quantize the local sharpness with the number of reconstructed atoms [31]. There is research that tries to directly estimate the blur kernel but the estimated maps tend to be very coarse and the optimization process is very time consuming [11, 75]. There are other works that can produce a sharpness map, such as in depth map estimation [76], or blur segmentation [59], but the depth map is not necessarily corresponding to the amount of sharpness and they tend to highlight blurred edges or insensitive to the change of small amount of blur. In this work, we adopted the method of [69] given its sensitivity to sharp low contrast regions. Detailed description can be found in "Sharpness Detection Network" and "Training of the Sharpness Detection Network".

The rest of the paper is organized as follows. The proposed method is described in "Methods". Experiments and results are presented in "Experiment Setup" and "Results". Discussion of the potential of the proposed method is in "Discussion" with conclusion drawn in "Conclusion".

# Methods

## Objective

As shown in Fig. 1, SAGAN consists of three networks, the generator G, discriminator D and the sharpness detection network S. G learns a mapping $G : x \rightarrow \hat{y}$, where $x$ is the LDCT the generator is conditioned upon. $\hat{y}$ is the denoised CT that is expected to be as possible close to as to the convCT ($y$) and we call it virtual convCT here. D's objective is to differentiate the virtual image pair $(x, \hat{y})$ from the real one $(x, y)$. Note that the input to D is not just the virtual ($\hat{y}$) and real convCT ($y$), but also LDCT ($x$). $x$ is concatenated to both $y$ and $\hat{y}$ and is served as additional information for D to rely on so that D can penalize the mismatch. In simpler term, G tries to synthesize a virtual convCT that can fool D whereas D tries to not get fooled. The training of G against D forms the adversarial part of the objective, which can be expressed as

$$\mathcal{L}_{adv}(G, D) = \mathbb{E}_{x,y \sim p_{data}(x,y)}[(D(x, y) - 1)^2]$$
$$+ \mathbb{E}_{x \sim p_{data}(x)}[D(x, \hat{y})^2], \quad (1)$$
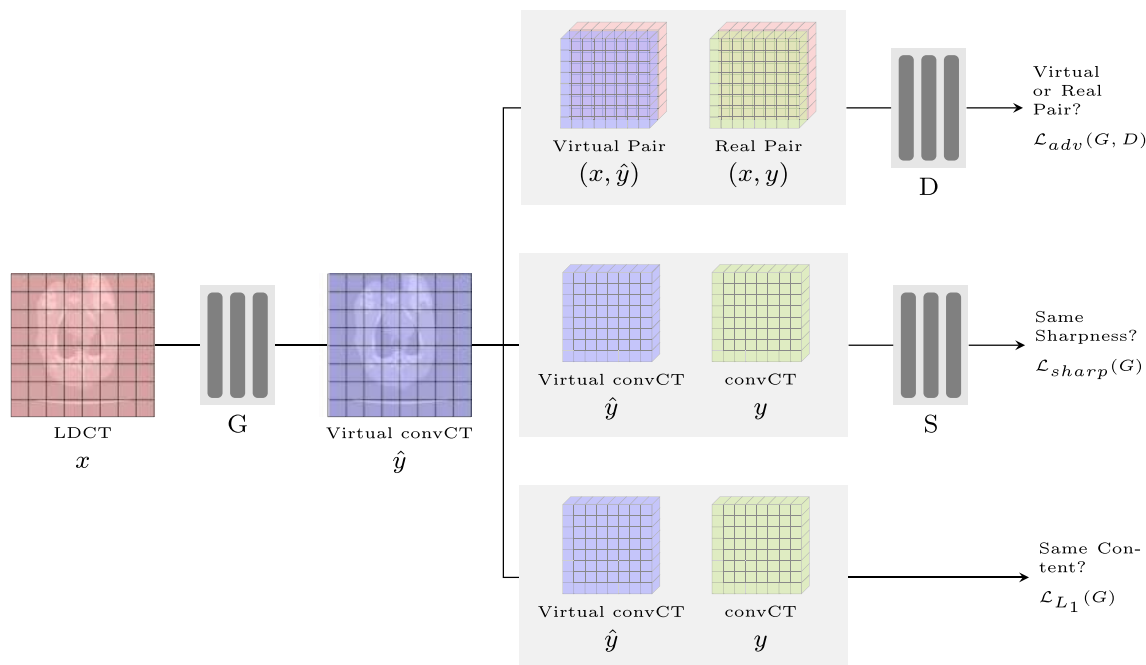
**Fig. 1** Overview of SAGAN. G is the generator that is responsible for the denoising. D is the discriminator employed to discriminate the virtual and real image pairs. S is a sharpness detection network and used to compare between the sharpness of the generated and real image. The system accepts the LDCT $x$ and convCT $y$ as the input and outputs virtual convCT (noise removed) $\hat{y}$

where $p_{data}$ is the real data distribution from which real images are sampled and the expectation $\mathbb{E}$ was calculated over these sampled data points. G is trying to minimize the above objective whereas D is trying to maximize it. We adopt the least square loss (with 0–1 binary coding scheme) instead of cross entropy loss in the above formulation because the least square loss tend to generate better images [42].

This loss is usually accompanied by traditional pixel-wise loss to encourage data fidelity for G, which can be expressed as

$$\mathcal{L}_{L_1}(G) = \mathbb{E}_{x,y \sim p_{data}(x,y)}[||y - \hat{y}||_1], \qquad (2)$$

where $|| \cdot ||_1$ is the $L_1$ distance and $y$ and $\hat{y}$ represents the convCT and virtual convCT.

Moreover, we proposed a sharpness detection network S to explicitly evaluate the denoised image's sharpness. The generator now not only has to fool the discriminator by generating a image with similar content to the real convCT in a $L_1$ sense, but also has to generate a similar sharpness map as close as to the real convCT. With $S$ denoting the mapping from the input to the sharpness map, the sharpness loss can be expressed as:

$$\mathcal{L}_{sharp}(G) = \mathbb{E}_{x,y \sim p_{data}(x,y)}[||S(\hat{y}) - S(y)||_2], \qquad (3)$$

where $|| \cdot ||_2$ is the $L_2$ distance.

Combining these three losses together, the final objective of SAGAN is

$$\mathcal{L}_{SAGAN} = \arg\min_G \max_D (\mathcal{L}_{adv}(G,D) + \lambda_1 \mathcal{L}_{L_1}(G) + \lambda_2 \mathcal{L}_{sharp}(G)), \qquad (4)$$

where $\lambda_1$ and $\lambda_2$ are the weighting terms to balance the losses. Note that in the traditional GAN formulation, the generator is also conditioned on random noise $z$ to produce stochastic outputs. But in practice, people have found that the adding of noise in the conditional setup like ours tends to be not effective [30, 43]. Therefore, in the implementation we have discarded $z$ so that the network only produces deterministic output.

## Network Architecture

### Generator

There are several different variants of generator architecture that have been adopted in the literature for image-to-image translation tasks: the Encoder-Decoder structure [27], the U-Net structure [30, 52], the Residual net structure [32] and one for removing of rain drops (denoted as Derain) [72]. The Encoder-Decoder structure has a bottleneck layer that requires all the information pass through it. The information consolidated by the encoder only encrypts the global structure of the input while discarding the textured

details. U-Net is similar to this architecture with a slight difference in that it adds long skip connections from encoder to decoder so that fine-grained details can be recovered [52]. The residual components, first introduced by He et al. [26] is claimed to be better for the training of very deep networks. The reason is that the short skip connection of the residual component can directly guide the gradient flow from deep layer to the shallow layer [19]. Later, we start to see many works incorporating the residual block into the network architecture [32, 53, 72, 73] when the network gets deep. The Detain architecture and its variants [68] share a common property which is that they maintained the spatial size of the feature maps during the processing. An adverse effect of this is that the number of feature maps need to remain small to avoid consuming too much memory.

Applications like style transfer do not require preservation of local textures and details of the content image (textures come from the style image) [32]. Therefore, it is rare to see long skip connections used in their network structure. However, for CT noise removal, the recovery of the underlying detail is of vital importance since the subtle structure could be a lesion that can develop into cancer. Therefore, in this work we adopted the unet256 structure [30] with long skip connections. The kernel stride is 1 for the first-stage feature extraction with no downsampling. We also incorporate several layers of the residual connection in the bottle neck layers for stabilizing the training of the network. Note that the feature of the bottleneck layers' spatial dimension is not reduced to $1 \times 1$ as in the Encoder-Decoder

structure to reduce the model size (similar to SegNet [3]) and we do not observe any significant performance drop by doing this. The architecture can be seen in Fig. 2. An experiment to compare the different generator architecture can be found in "Analysis of the Generator Architecture".

## Discriminator

The objective of the discriminator is to tell the difference between the virtual image pair $(x, \hat{y})$ and the real image pair $(x, y)$. Here, we adopt the PatchGAN structure from pix2pix framework [30], where instead of classifying the whole image as real or virtual, it will focus on overlapped image patches. By using $G$ alone with $L_1$ or $L_2$ loss, the architecture would degrade to a traditional CNN-based denoising methods.

## Sharpness Detection Network

Bluring of the edges is a major problem faced by many image denoising methods. For traditional denoising methods using non-linear filtering, the edges will be inevitably blurred no matter by averaging out neighbouring pixels or self-similar patches. It is even worse in high-noise settings whereas noise can also produce some edge-like structures. Neural network based methods could also suffer the same problem if optimizing the pixel-wise differences between the generated image and the ground truth, because the result that averages out all possible solutions end up
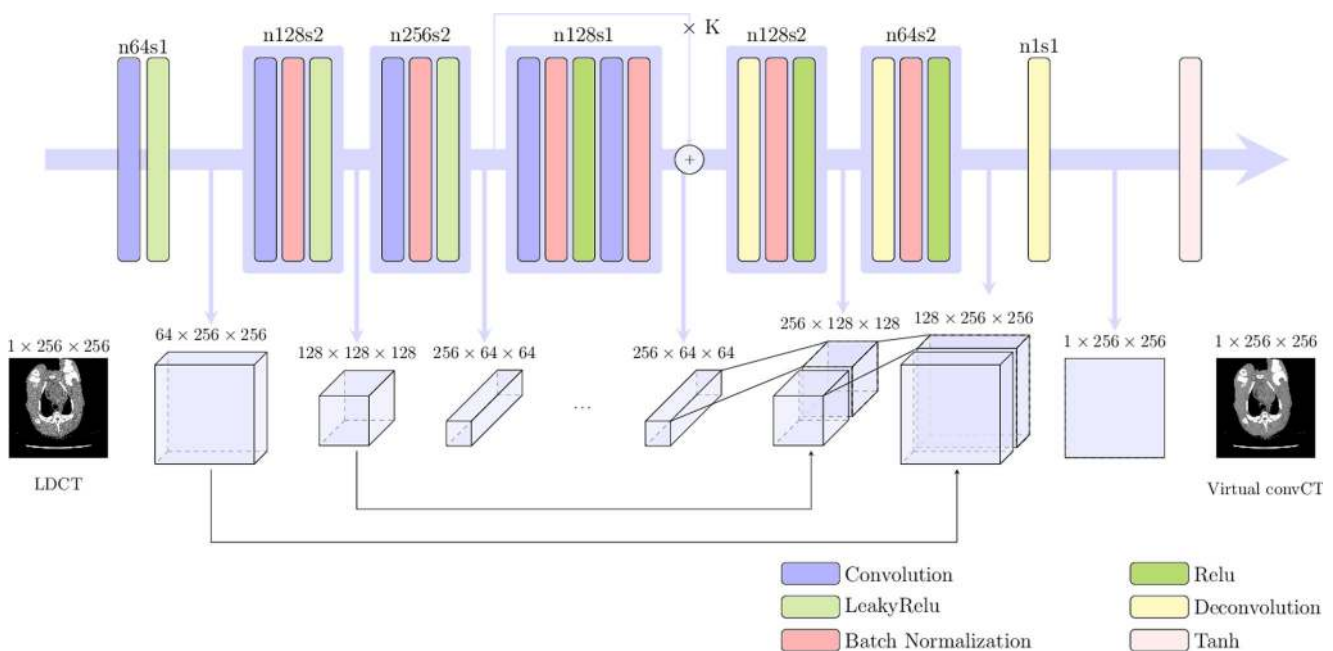


**Fig. 2** Proposed generator of the SAGAN. The residual block in the centre of the network is repeated K times and K was chosen as 9 for the experiment

giving the best quantitative measure. The adversarial loss used introduced by the discriminator of GAN is able to output a much sharper and recognizable image from the candidates. However, the adversarial loss does not guarantee the images to be sharply reconstructed, especially for low-contrast regions.

We believe that auxiliary guided information should be provided to the generator so that it can recover the underlying sharpness of the low-contrast regions from the contaminated noisy image, as similar to the frontal face predication [28] where the position of facial landmarks are supplied to the network. Since the direct markup of low-contrast sharp regions is not practical for medical images, an independent sharpness detection network S was trained in this work. During the training of SAGAN, the virtual convCT generated from G is sent to S and the output sharpness map is compared with the map of the ground truth image. The sharpness map is a grayscale image with the pixel intensity denotes the local sharpness of a selected local region and is the final feature map of the sharpness detection network S. We compute the mean square error between the two sharpness maps and this error was back-propagated through the generator to update its weights.

## Experiment Setup

The proposed SAGAN was applied to both simulated low-dose and real low-dose CT images to evaluate its effectiveness. In both settings, peak signal to noise ratio (PSNR) and structured similarity index (SSIM) [64] were adopted as the quantitive metrics for the evaluation (using abdomen window image). The former metric is commonly used to measure the pair-wise difference of two signals whereas the SSIM is claimed to better conform to the human visual perception. For the real dataset, the mean standard deviation of 42 smooth rectangular homogeneous regions (size of $21 \times 21$, $172.27$ mm$^2$) was computed as direct measures of the noise level.

To further evaluate the general applicability of the proposed method, we selected two patient's LDCTs from the Kaggle Data Science Bowl 2017[1] and applied our trained model to it. Visual results and noise levels are provided for evaluation in this case. Twenty rectangular homogeneous regions of size $21 \times 21$ were selected for the calculation.

### Simulated Noise Dataset

In this dataset, 239 normal-dose CT images were downloaded from the National Cancer Imaging Archive (NCIA).

---

[1] https://www.kaggle.com/c/data-science-bowl-2017/data

Each image has a size of $512 \times 512$ covering different parts of the human body. A fan-beam geometry was used to transform the image to the sinogram, utilizing 937 detectors and 1200 views.

The photons hitting the detector are usually treated as Poission distributed. Together with the electrical noise which starts to gain prominence in low-dose cases and is normally Gaussian distributed, the detected signal can be expressed as:

$$N \sim \text{Poisson}(N_0 \exp(-y)) + \text{Gaussian}(0, \sigma_e^2), \qquad (5)$$

where $N_0$ is the X-ray source intensity or sometimes called blank flux, $y$ is the sinogram data, and $\sigma_e$ is the standard deviation of the electrical noise [36, 70].

The blank scan flux $N_0$ was set to be $1 \times 10^5$, $5 \times 10^4$, $3 \times 10^4$, $1 \times 10^4$ to simulate effect of different dose levels and the electrical noise was discarded for simplicity. Since the network used is fully convolutional, the input can be of different size. Each image was further divided into four $256 \times 256$ sub-images to boost the size of the dataset. 700 out of the resultant 956 sub-images were randomly selected as the training set and the remaining 64 full images were used as the test set. Some sample images are shown in the first column of Fig. 5. Note that the simulated dose here is generally lower than that of [13].

### Real Datasets

CT scans of a deceased piglet were obtained with a range of different doses utilizing a GE scanner (Discovery CT750 HD) using 100 kVp source potential and 0.625 mm slice thickness. A series of tube currents were used in the scanning to produce images with different dose levels, ranging from 300 mAs down to 15 mAs. The 300 mAs served as the conventional full dose whereas the others served as low doses with tube current reductions of 50, 25%, 10 and 5% respectively. At each dose level we obtained 850 images of a size $512 \times 512$ in total. 708 of them were selected for training and 142 for testing. The training size of the real dataset was also boosted by dividing each image into four $256 \times 256$ sub-images, giving us 2832 images in total for training.

A CT phantom (Catphan 600) was scanned to evaluate the spatial resolution of the reconstructed image, using 120 kVp and 0.625 mm slice thickness. For this dataset, only two dose levels were used. The one with 300 mAs served as the convCT and the one with 10mAs served as the LDCT. The detailed doses is provided in Table 1.

Data Science Bowl 2017 is a challenge to detect lung cancer from LDCTs. It contains over a thousand high-resolution low-dose CT images of high risk patients. The corresponding convCTs and specific dosage level for each scan are not available. We selected two patients' scan to evaluate the generality of the proposed SAGAN method on unseen doses.

**Table 1** Detailed doses for the piglet and phantom datasets

(a) Doses used for the piglet dataset. In all 5 series, tube potential was 100 kV with 0.625 mm slice thickness. Tube currents were decreased to 50, 25, 10 and 5% of full-dose tube current (300 mAs) to obtain images with different doses. CTDI is the CT dose index and DLP is the dose-length product

| Dose level | Full | 50% | 25% | 10% | 5% |
|---|---|---|---|---|---|
| Tube current (mAs) | 300 | 150 | 75 | 30 | 15 |
| $CTDI_{vol}$ (mGy) | 30.83 | 15.41 | 7.71 | 3.08 | 1.54 |
| DLP (mGy-cm) | 943.24 | 471.62 | 235.81 | 94.32 | 47.16 |
| Effective dose (mSv) | 14.14 | 7.07 | 3.54 | 1.41 | 0.71 |

(b) Doses used for the Catphan 600 dataset. In both series, tube potential is 120 kV with 0.625 mm slice thickness

| Scan series | Full | 3.33% |
|---|---|---|
| Tube current (mAs) | 300 | 10 |
| $CTDI_{vol}$ (mGy) | 26.47 | 0.88 |

Four experiments were conducted. In the first experiment, we evaluated the effect of the generator and the sharpness loss by using the simulated noise dataset. In the second experiment, we evaluated the spatial resolution with the Catphan 600 dataset. In the third experiment the proposed SAGAN was applied on the piglet dataset to test its effectiveness on a wide range of real quantum noise. Finally in the last experiment, the SAGAN model trained on the piglet dataset was applied to the clinical patient data in the Data Science Bowl 2017. Two state-of-the-art methods: BM3D, K-SVD from two major line of traditional image denoising methods were selected for the comparison. For the real dataset, the available CT manufacture iterative reconstruction methods, ASIR (40%) and VEO were also compared. All experiments on the real datasets are trained on the full range DICOM image.

## Implementation Details

### Training of SAGAN

All the networks are trained on the Guillimin cluster of Calcul Quebec and the Cedar cluster of Compute Canada. Adam optimizer [35] with $\beta_1 = 0.5$ was used for the optimization with learning rate 0.0001. The generator and discriminator was trained alternatively across the training with $k = 1$ as similar in [22]. The implementation was based on the Torch framework. The training images have size of $256 \times 256$ whereas the testing is with full size $512 \times 512$ CT images. All the networks here are trained to 200 epochs. $\lambda_1$ was set to be 100 and $\lambda_2$ to 0.001. For the simulated dataset, one SAGAN was trained for each simulated dose.

For the real dataset, one SAGAN was trained for the piglet and phantom separately, with the training set of different doses of piglet combined.

### Training of the Sharpness Detection Network

The sharpness detection network follows the work of Yi et al. [69]. In that work, Yi et al. used a non-differentiable analytic sharpness metric to quantify the local sharpness of a single image. Here in this work, we trained a neural network to imitate its behaviour. To be more specific, the defocus segmentation dataset [54] that contains 704 defocused images was adopted for the training. Five sub-images of size $256 \times 256$ were sampled from the four corners and centre of each defocused image to boost the size of the training set. For the training of the sharpness detection network, the unet256 structure was adopted and the sharpness map created by the local sharpness metric of [69] was used for regression. Adam optimizer [35] with $\beta_1 = 0.5$ was also used for the optimization with learning rate also 0.0001. Some sample images and their sharpness map can be seen in Fig. 3.

## Results

### Analysis of the Generator Architecture

A variety of generator architectures were evaluated, including unet256 [52], res9 [32] and Derain [72]. In the analysis, only the architecture of the generator was modified. The discriminator was fixed to be the patchGAN with patch size of $70 \times 70$ [30]. The sharpness network was not incorporated in this experiment for simplicity.

The quantitative results are shown in Table 2. As can be seen, the performance generally improves when lowering the noise level (increase of $N_0$) no matter what architecture has been used. For every single noise level, the listed generators achieved comparative results since all of them optimized the PSNR as part of their loss function. However, the visual results shown in Fig. 4 have shed some light on the architectural differences. Comparing the unet256 with the proposed, we can see that the proposed recovered the low-contrast zoom region much sharper. It shows the benefits of maintaining the spatial size during the first stage of feature extraction. The major difference of res9 and the proposed one is the long skip connection. We can see by comparing results of the two that this connection can help recovering small details. As for the Derain architecture, the training was not stable and some grainy artifacts can be observed. We attribute this to the small size of the feature dimension of the bottleneck layer (only one in this case) which is not sufficient to encode the global features.
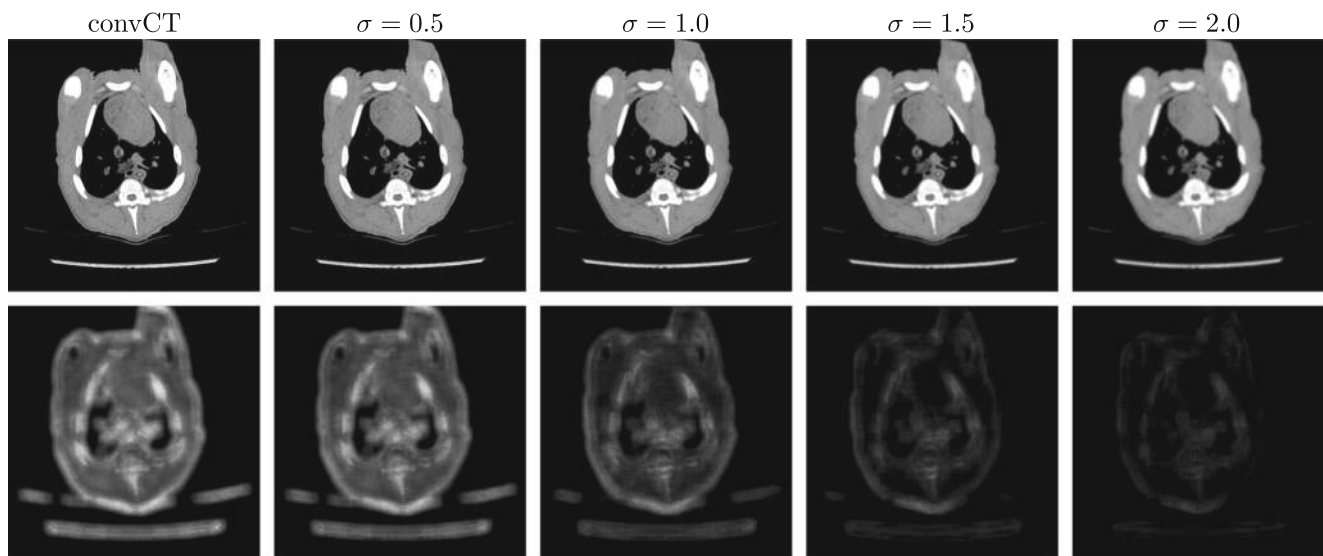
**Fig. 3** The output of the sharpness detection network. The upper row is the convCT of a lung region selected from the piglet dataset and its blurred versions with increasing amount of Gaussian blur ($\sigma$ shown on top). The lower row shows their corresponding sharpness map

**Table 2** Comparison of different generator architecture on the simulated dataset. The input noise level in terms of PSNR and SSIM is shown in the top row

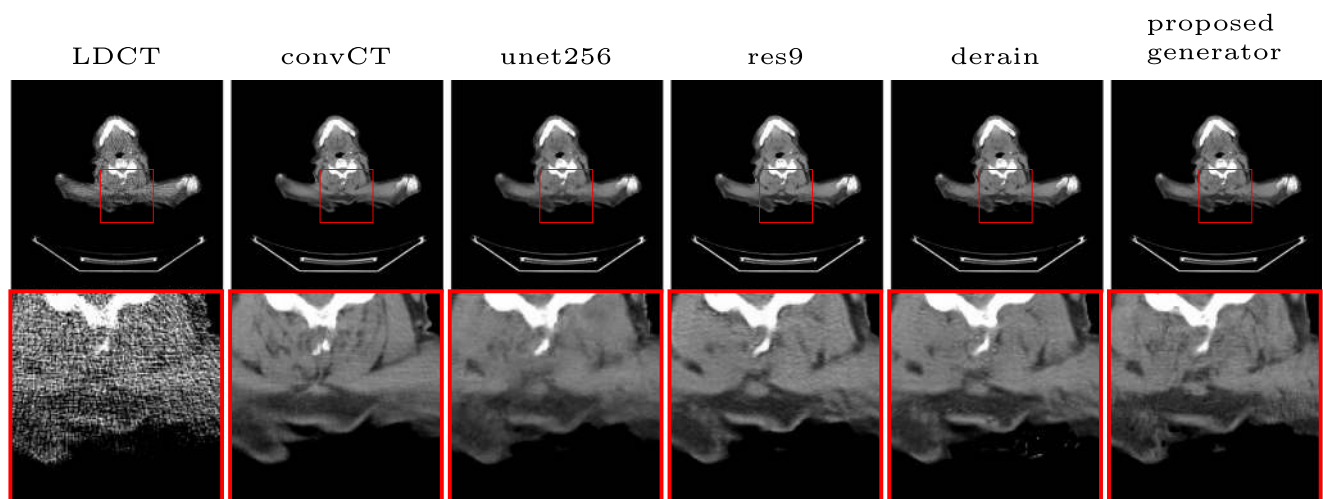| Generator Archetecture | $N_0 = 10000$ | | $N_0 = 30000$ | | $N_0 = 50000$ | | $N_0 = 100000$ | |
|---|---|---|---|---|---|---|---|---|
| | PSNR 18.3346 | SSIM 0.7557 | PSNR 21.6793 | SSIM 0.7926 | PSNR 23.1568 | SSIM 0.8119 | PSNR 24.8987 | SSIM 0.8387 |
| unet256 | 26.2549 | 0.8384 | 27.5819 | 0.8598 | 27.9269 | 0.8646 | 28.1243 | 0.8711 |
| res9 | 25.9032 | 0.8412 | 26.7443 | 0.8549 | 27.8765 | 0.8710 | 28.8179 | 0.8877 |
| Derain | 25.8094 | 0.8376 | 26.4167 | 0.8505 | 27.1724 | 0.8562 | 27.1307 | 0.8570 |
| Proposed | 26.6584 | 0.8438 | 27.3066 | 0.8533 | 27.8443 | 0.8622 | 28.1718 | 0.8701 |



**Fig. 4** Evaluation of the generator architecture. LDCT is with $N_0 = 10000$

**Table 3** Quantitive evaluation of the sharpness-aware loss

| Methods | $N_0 = 10000$ | | $N_0 = 30000$ | | $N_0 = 50000$ | | $N_0 = 100000$ | |
|---|---|---|---|---|---|---|---|---|
| | PSNR 18.3346 | SSIM 0.7557 | PSNR 21.6793 | SSIM 0.7926 | PSNR 23.1568 | SSIM 0.8119 | PSNR 24.8987 | SSIM 0.8387 |
| W/o sharpness loss | 26.6584 | 0.8438 | 27.3066 | 0.8533 | *27.8443* | 0.8622 | 28.1718 | 0.8701 |
| W sharpness loss (SAGAN) | *26.7766* | *0.8454* | 27.5257 | *0.8571* | 27.7828 | 0.8620 | *28.2503* | *0.8708* |
| BM3D | 24.0038 | 0.8202 | 25.6046 | 0.8485 | 26.0913 | 0.8589 | 26.7598 | 0.8726 |
| K-SVD | 21.9578 | 0.7665 | 24.0790 | 0.8167 | 25.0425 | 0.8379 | 26.0902 | 0.8620 |

## Analysis of the Sharpness Loss

In this experiment, we evaluated the effectiveness of the sharpness loss on the denoised result. Table 3 shows quantitive results before and after applying the sharpness detection network. The values in term of PSNR and SSIM are comparable to each other with slight differences that can be explained by the competition of the data fidelity loss and the sharpness loss. However, visual examples shown in Fig. 5 clearly demonstrates that the sharpness loss excels at suppressing noises on small structures without introducing too much blurring.

## Denoising Results on Simulated Dataset

As can be also seen from Table 3 and Fig. 5, the performance of SAGAN in terms of PSNR and SSIM is better than BM3D and K-SVD in all noise levels. For the visual appearance, SAGAN is also sharper than BM3D and K-SVD and can recover more details. K-SVD could not remove all the noise
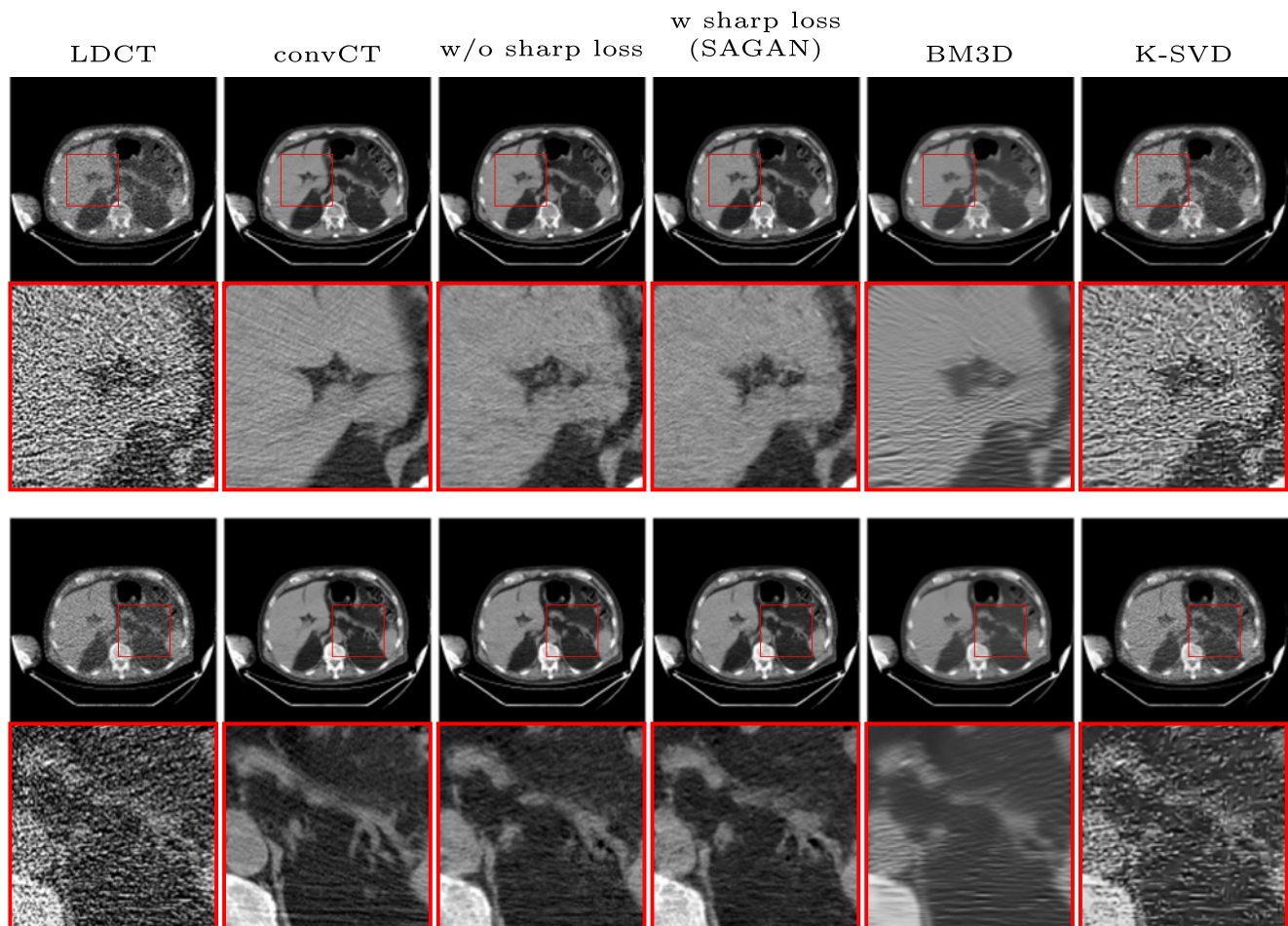


**Fig. 5** Visual examples to evaluate the effectiveness of the proposed sharpness loss with $N_0 = 10000$. Rows 1 and 3 are two examples with zoomed region shown below
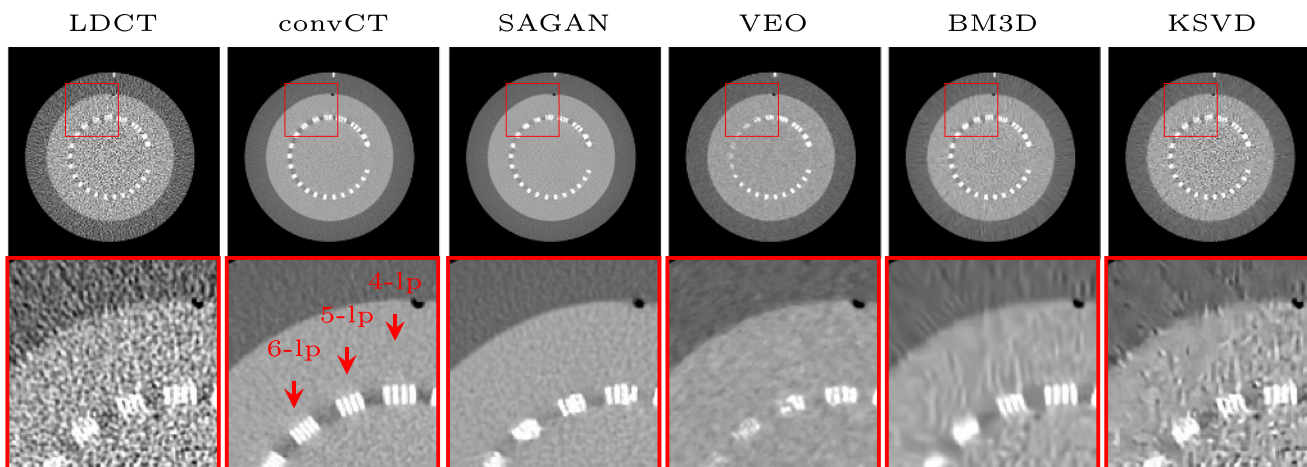
**Fig. 6** Visual comparison of the spatial resolution on the CTP 528 high-resolution module of the Catphan 600. Images are trained and tested on the full range DICOM image. Display window is [40, 400] HU

and sometimes make the resultant image very blocky. For example, in the zoomed region of row 1 of Fig. 5, the fat region of SAGAN is the sharpest among the comparators. In row 2, we have shown that the low-contrast vessel can be faithfully reconstructed by SAGAN whereas missed by the other methods. The streak artifact is another problem faced by BM3D in high quantum noise level as has already been pointed out by many works [13, 33]. We recommend the reader to zoom in for better appreciation of the results.

### Denoising Results on Catphan 600

Figure 6 gives the denoised visual result for the CTP 528 21 line pair high-resolution module of the Catphan 600. The 4 and 5-line pairs is clearly distinguishable for LDCT and we can observe these line pairs equally well on SAGAN reconstructed images which suggests that the amount of

spatial resolution loss is very small. Six-line pairs is distinguishable from the convCT but not the case for LDCT and all the reconstruction methods, which highlights the gap between the reconstruction methods and the convCT. Figure 7 shows the line profile along the line drawn across the 4- and 5-line pairs. Thirty points were sampled along the drawn line. SAGAN is the one among the comparative methods that achieves the highest spatial resolution. K-SVD behaves slightly better than BM3D and VEO demonstrates the lowest spatial resolution.

### Denoising Results on the Real Piglet Dataset

Here, we plotted a line graph of the PSNR and SSIM against the dosage in Fig. 9 for all the comparator methods. It can be seen that all methods except VEO have their performance improved with the increase of dose in terms of PSNR. SSIM
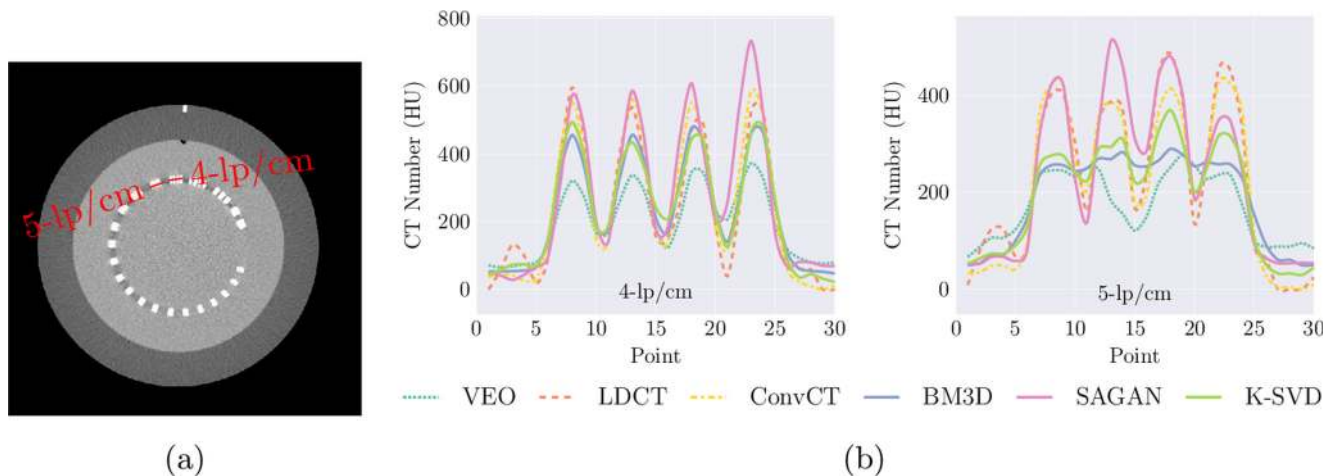


**Fig. 7** **a** shows the convCT of the CTP 528 high-resolution module of the Catphan 600. **b** shows the line profile for the 4-line pair per centimeter and 5-line pair per centimeter of different reconstruction methods. 30 data points were sampled along the red line as marked in (**a**)

is less affected because it penalizes structural differences rather than the pixel-wise difference. The average SSIM measure in the lowest dose level for SAGAN is 0.95 which is slightly higher than that of the second highest dose level for FBP. Figure 8 shows some visual examples from different anatomic region (from head to pelvis) at the lowest dose level and their reconstruction by all the comparator methods. We can see clearly that SAGAN produces results that are more visually appealing than the others.

Figure 9 also shows the mean standard deviation of CT numbers on 42 hand selected rectangular homogeneous regions as a direct measure of noise level. The red horizontal dashed line is the performance of the convCT and serves as

reference and it can be seen that the available commerical methods do not surpass the reference line. In general, the mean standard deviation of SAGAN results are pretty constant across all dose levels and very close to the convCT. It implicitly shows that SAGAN can simulate the statistical properties of convCT. BM3D and K-SVD on the other hand obtained smaller numbers by over-smoothing the result images. At the highest noise level, the measure was 25.35 for FBP and 8.80 for SAGAN, corresponding to a noise reduction factor of 2.88. Considering both the quantitative measures and the visual appearance, SAGAN is no doubt the best method among the comparators in the highest noise level.
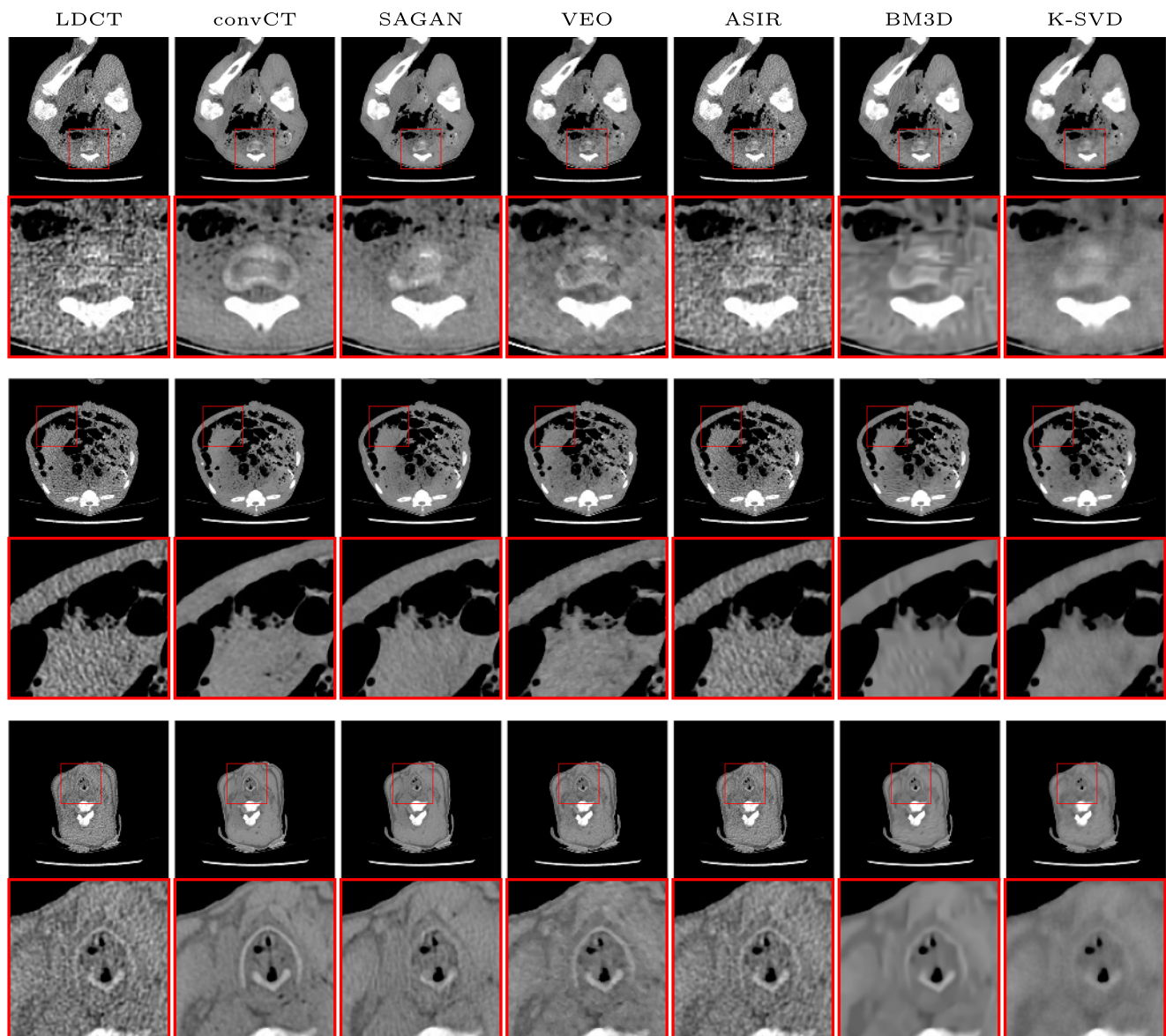


**Fig. 8** Visual examples for the denoised images on the piglet dataset. Rows are 3 selected samples from pelvis to head each with a zoomed up local region. The first column is the LDCT (5% of full dose reconstructed by FBP, 0.71 mSv). The second column is the convCT (100% dose reconstructed by FBP, 14.14 mSv). The last 5 columns are results from different denoising methods. Display window is [40, 400]
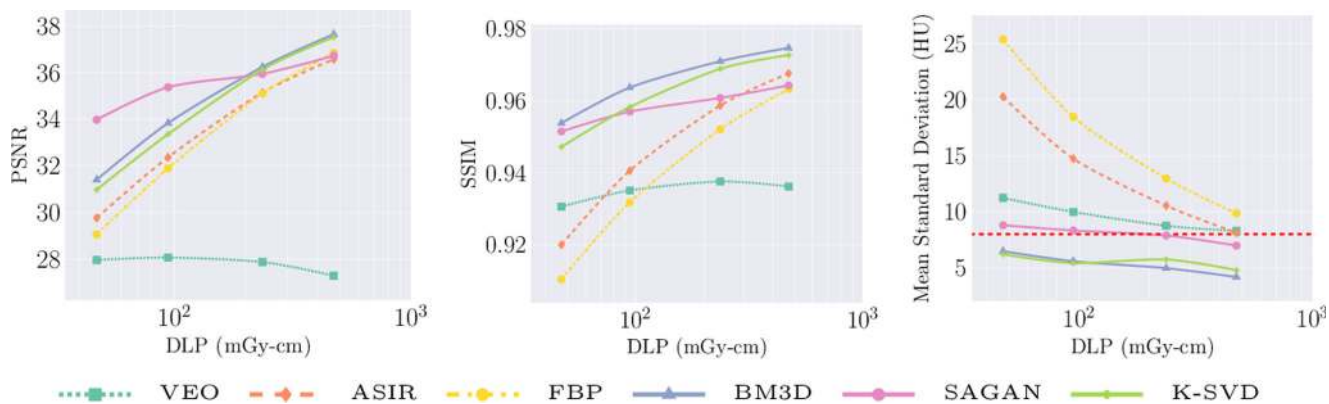
**Fig. 9** The left two figures show the PSNR and SSIM plotted against dose-length product (DLP) for different reconstruction methods for the piglet dataset. The rightmost figure gives the mean standard deviation of image noise against dose-length product (DLP) for different reconstruction methods. Red dashed line refers to the standard deviation of the FBP-reconstructed convCT

## Denoising Results on the Clinical Patient Data with Unknown Dose Level

Figure 10 shows the results on the clinical patient data with unknown dose levels. The dose level of these data are unlikely to coincide with the dose level of our training set but we can see that it performs reasonable well on these images with decreased noise level.

## Discussion

These quantitative results demonstrate that SAGAN excels in recovering underlying structures in great uncertainty. The adversarially trained discriminator guarantees the denoised texture to be close to convCT. This is an advantage over VEO, which produces a different texture. The sharpness detection network guarantees that the generated CT is with
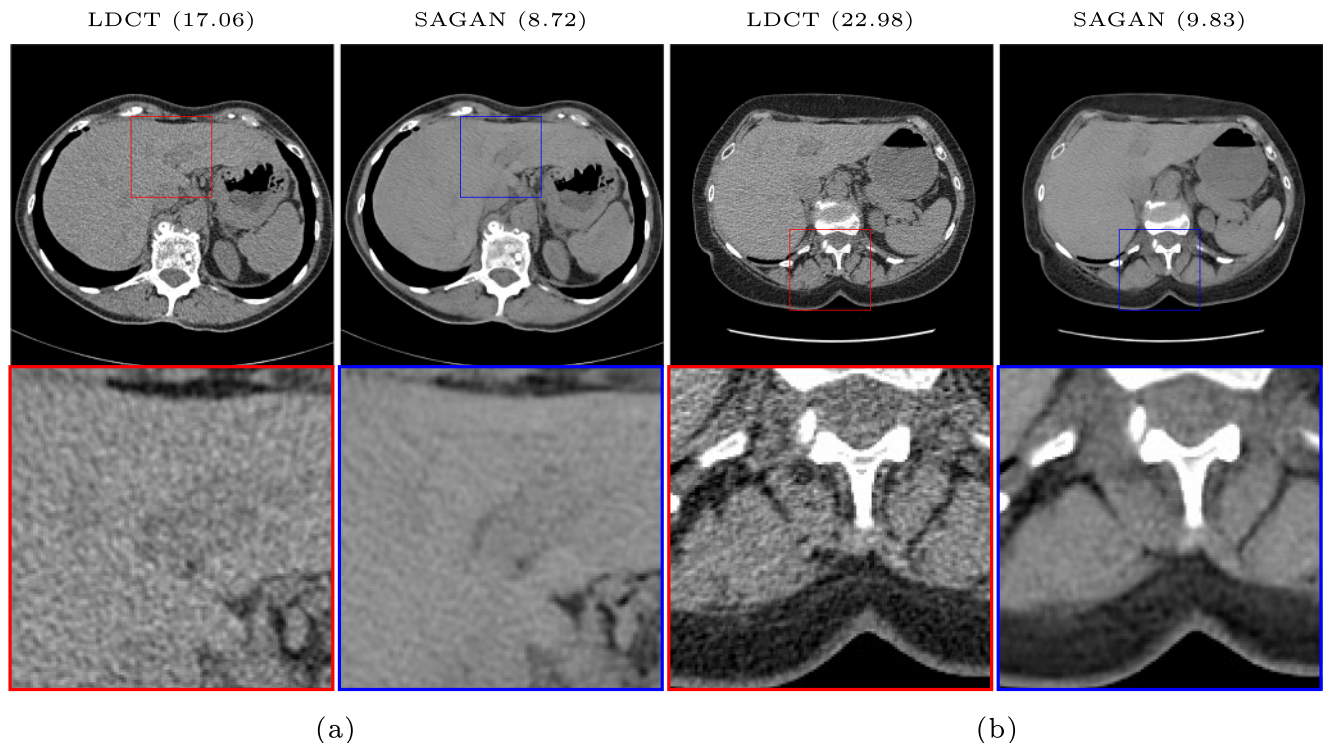


**Fig. 10** SAGAN denoising results on the clinical patient data (from the Kaggle Data Science Bowl 2017) with unknown dose levels. Image pairs of (**a**) and (**b**) come from two different patients. The noise level were shown above the images and were computed from 20 homogeneous regions selected from the patient scan. Display window of [40, 400] HU

similar sharpness as the convCT. Another advantage is time efficiency. Neural network based methods, including SAGAN only need one forward pass in the testing and the task could be accomplished in less than a second. BM3D showed better denoising at the highest dose, however SAGAN was better at lower doses. BM3D and K-SVD also had evident streak artifacts across the image surface at low doses as seen in Fig. 3.

Another phenomenon we have observed is that the SAGAN can also help to mitigate the streak artifacts. As can be seen from the first row of Fig. 5, the lower half of the convCT had mild streak artifacts but was less evident in the SAGAN result (4th column). The reason we think is that the discriminator discriminates patches and the number of patches containing artifacts is significantly smaller than the number of normal ones. Therefore these patches were considered as outliers in the discrimination process. A straightforward extension of this work would be for limited view CT reconstruction.

The proposed generator here adopts the Unet [30, 52] architecture and incorporates the residual connection for the ease of training. We have empirically demonstrate its effectiveness on the denoising task and have observed much more stable training statistics than the comparators in the adversarial training scheme.

The sharpness-aware loss proposed here is similar to the methodology of the content loss as used in [32, 37] but differs in the final purpose. The similarity lies in that we both measure the high-level features of the generated and input image. In their work, the high-level features are from the middle layer of the pre-trained VGG network [57] and used to ensure the perceptual similarity. On the contrary, the high-level features used here are extracted from a specifically trained network and directly correspond to the visual sharpness.

The work of Wolterink et al. [65] and Yang et al. [68] also employed GAN for CT denoising and some technical differences from ours have been highlighted in "Related Works". Here we also want to emphasize two of their weaknesses. These two works either centred on cardiac CT or abdominal CT. It is not clear whether their trained model can be applied to CTs of different anatomies. Our work considered a wider range of anatomic regions ranging from head to pelvis and has demonstrated that a single network in cGAN setting would be sufficient to denoise CT of the whole body. Moreover, their work only employed a single dose level in the training whereas ours covered a wider range of dose levels. We have also empirically shown that the trained model not only suitable for denoising images with the training dose level but also applicable to unseen dose levels as long as the noise level is within our training range.

The dose reduction achieved by SAGAN is very high. According to the measurement of PSNR and SSIM, SAGAN

reconstructed result in the lowest dose level has a measurement almost equivalent to the CT in the second highest dose level (7.07 mSv), corresponding to a dose reduction of 90%. Meanwhile, if we measure the dose reduction factor with respect to the mean standard deviation of attenuation, SAGAN's result in the lowest dose level (0.71 mSv) has a noise level similar to that of the convCT (14.14 mSv), corresponding to a dose reduction of 95%. Each measurement has their own strengths and weaknesses in measuring the CT image quality. PSNR and SSIM take into consideration the spatial information but would penalize a lot of the texture difference even with similar underlying image content. For example, although VEO has been shown to have superior performance in many clinical studies than ASIR and FBP and has received clearance released by the Food and Drug Administration of the USA, it obtained the worst PSNR and SSIM measurements. On the other hand, mean standard deviation of attenuation direct measures the noise level, but completely discarded the spatial information. Therefore, we reported both results for the sake of fair comparison. The work of Suzuki et al. [58] reported a dose reduction of 90% from 1.1 mSv to 0.11 mSv with MTANN. Their network is patched based and need to train multiple networks corresponding to different anatomic regions. It is unclear how much dose reduction was achieved for the other deep learning based approaches [13, 33, 65, 67].

We evaluated the spatial resolution of the reconstructed image using the Catphan 600 high-resolution module. This analysis was generally missed for the other deep learning based approaches. These methods could achieve high PSNR and SSIM by directly treat them as the optimization objective but at the cost of losing spatial resolution. GAN-based methods, including ours, mitigate this problem by incorporating the adversarial objective. We think it is crucial to bring the spatial resolution into assessment for the deep learning based approaches when PSNR and SSIM become less effective. An alternative way of quantifying the denoising performance would be to measure the performance of subsequent higher level tasks, e.g. lung nodule detection, anatomical region segmentation. We would like to leave this to the future work.

## Conclusion

In this paper, we have proposed sharpness-aware network for low-dose CT denoising. It utilizes both the adversarial loss and the sharpness loss to leverage the blur effect faced by image-based denoising method, especially under high noise levels. The proposed SAGAN achieves improved performance in the quantitative assessment and the visual results are more appealing than the tested competitors.

However, we acknowledge that there are some limitations of this work that are waiting to be solved in the future. First of all, the sharpness detection network is trained to compute the sharpness metric of [69] which is not very sensitive to just noticeable blur. This could limit the final sharpness of the denoised image, especially some small low-contrast regions.

Second, for all the deep learning-based methods, the network need to be trained against a specific dosage level. Even though we trained our method on a wild range of doses and have applied it to clinical patient data, the analysis is mostly centred on the visual quality assessment of the denoised image. The image diagnosis performance in clinical practice remains to be evaluated.

# References

1. Aharon M, Elad M, Bruckstein A: K-SVD: An algorithm for designing overcomplete dictionaries for sparse representation. IEEE Transactions on signal processing 54(11):4311–4322, 2006
2. Arjovsky M, Chintala S, Bottou L (2017) Wasserstein gan. arXiv:170107875
3. Badrinarayanan V, Kendall A, Cipolla R (2015) Segnet: A deep convolutional encoder-decoder architecture for image segmentation. arXiv:151100561
4. Balda M, Hornegger J, Heismann B: Ray contribution masks for structure adaptive sinogram filtering. IEEE transactions on medical imaging 31(6):1228–1239, 2012
5. Beister M, Kolditz D, Kalender WA: Iterative reconstruction methods in x-ray ct. Physica medica 28(2):94–108, 2012
6. Bertasius G, Shi J, Torresani L: Deepedge: A multi-scale bifurcated deep network for top-down contour detection. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp 4380–4389, 2015
7. Bouman C, Sauer K: A generalized gaussian image model for edge-preserving map estimation. IEEE Transactions on Image Processing 2(3):296–310, 1993
8. Brenner DJ (2016) What do we really know about cancer risks at dose pertinent to ct scans. http://amos3.aapm.org/abstracts/pdf/115-31657-387514-119239.pdf
9. Buades A, Coll B, Morel JM: A non-local algorithm for image denoising. In: IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2005. CVPR 2005., IEEE, vol 2, 2005, pp 60–65
10. Candes EJ, Donoho DL: Recovering edges in ill-posed inverse problems: optimality of curvelet frames . Annals of statistics 30:784–842, 2002
11. Chakrabarti A, Zickler T, Freeman WT: Analyzing spatially-varying blur. In: 2010 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). IEEE, 2010, pp 2512–2519
12. Chen H, Zhang Y, Zhang W, Liao P, Li K, Zhou J, Wang G (2016) Low-dose ct via deep neural network. arXiv:160908508
13. Chen H, Zhang Y, Kalra MK, Lin F, Liao P, Zhou J, Wang G (2017) Low-dose ct with a residual encoder-decoder convolutional neural network (red-cnn). arXiv:170200288
14. Chen Y, Gao D, Nie C, Luo L, Chen W, Yin X, Lin Y: Bayesian statistical reconstruction for low-dose x-ray computed tomography using an adaptive-weighting nonlocal prior. Computerized Medical Imaging and Graphics 33(7):495–500, 2009
15. Chen Y, Yang Z, Hu Y, Yang G, Zhu Y, Li Y, Chen W, Toumoulin C, et al: Thoracic low-dose ct image processing using an artifact suppressed large-scale nonlocal means. Physics in Medicine and Biology 57(9):2667, 2012
16. Chen Y, Yin X, Shi L, Shu H, Luo L, Coatrieux JL, Toumoulin C: Improving abdomen tumor low-dose ct images using a fast dictionary learning based processing. Physics in medicine and biology 58(16):5803, 2013
17. Chen Y, Shi L, Feng Q, Yang J, Shu H, Luo L, Coatrieux JL, Chen W: Artifact suppressed dictionary learning for low-dose ct image processing. IEEE Transactions on Medical Imaging 33(12):2271–2292, 2014
18. Dabov K, Foi A, Katkovnik V, Egiazarian K: Image denoising by sparse 3-d transform-domain collaborative filtering. IEEE Transactions on image processing 16(8):2080–2095, 2007
19. Drozdzal M, Vorontsov E, Chartrand G, Kadoury S, Pal C: The importance of skip connections in biomedical image segmentation. In: International Workshop on Large-Scale Annotation of Biomedical Data and Expert Label Synthesis, Springer, pp 179–187, 2016
20. Glorot X, Bordes A, Bengio Y: Deep sparse rectifier neural networks. In: Aistats, vol 15, 2011, p 275
21. Golestaneh SA, Karam LJ (2017) Spatially-varying blur detection based on multiscale fused and sorted transform coefficients of gradient magnitudes. arXiv:170307478
22. Goodfellow I, Pouget-Abadie J, Mirza M, Xu B, Warde-Farley D, Ozair S, Courville A, Bengio Y: Generative adversarial nets. In: Advances in neural information processing systems, pp 2672–2680, 2014
23. Green M, Marom EM, Kiryati N, Konen E, Mayer A: Efficient low-dose ct denoising by locally-consistent non-local means (lc-nlm). In: International Conference on Medical Image Computing and Computer-Assisted Intervention, Springer, pp 423–431, 2016
24. Gulrajani I, Ahmed F, Arjovsky M, Dumoulin V, Courville A (2017) Improved training of wasserstein gans. arXiv:170400028
25. Ha S, Mueller K: Low dose ct image restoration using a database of image patches. Physics in medicine and biology 60(2):869, 2015
26. He K, Zhang X, Ren S, Sun J: Deep residual learning for image recognition. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp 770–778, 2016
27. Hinton G E, Salakhutdinov RR: Reducing the dimensionality of data with neural networks. Science 313(5786):504–507, 2006
28. Huang R, Zhang S, Li T, He R (2017) Beyond face rotation: Global and local perception gan for photorealistic and identity preserving frontal view synthesis. arXiv:170404086
29. Ioffe S, Szegedy C (2015) Batch normalization: Accelerating deep network training by reducing internal covariate shift. arXiv:150203167
30. Isola P, Zhu JY, Zhou T, Efros AA (2016) Image-to-image translation with conditional adversarial networks. arXiv:161107004
31. Jianping Shi JJ, Li XU: Just noticeable defocus blur detection and estimation. In: 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). IEEE, 2015
32. Johnson J, Alahi A, Fei-Fei L: Perceptual losses for real-time style transfer and super-resolution. In: European Conference on Computer Vision, Springer, pp 694–711, 2016
33. Kang E, Min J, Ye JC (2016) A deep convolutional neural network using directional wavelets for low-dose x-ray ct reconstruction. arXiv:161009736

34. Kim D, Ramani S, Fessler JA: Combining ordered subsets and momentum for accelerated x-ray ct image reconstruction. IEEE transactions on medical imaging 34(1):167–178, 2015

35. Kingma D, Ba J (2014) Adam: A method for stochastic optimization. arXiv:14126980

36. La Riviere PJ: Penalized-likelihood sinogram smoothing for low-dose ct. Medical physics 32(6):1676–1683, 2005

37. Ledig C, Theis L, Huszár F, Caballero J, Cunningham A, Acosta A, Aitken A, Tejani A, Totz J, Wang Z, et al (2016) Photo-realistic single image super-resolution using a generative adversarial network. arXiv:160904802

38. Li S, Fang L, Yin H: An efficient dictionary learning algorithm and its application to 3-D medical image denoising. IEEE Transactions on Biomedical Engineering 59(2):417–427, 2012

39. Lubner MG, Pickhardt PJ, Tang J, Chen GH: Reduced image noise at low-dose multidetector ct of the abdomen with prior image constrained compressed sensing algorithm. Radiology 260(1):248–256, 2011

40. Ma J, Huang J, Feng Q, Zhang H, Lu H, Liang Z, Chen W: Low-dose computed tomography image restoration using previous normal-dose scan. Medical physics 38(10):5713–5731, 2011

41. Manduca A, Yu L, Trzasko JD, Khaylova N, Kofler JM, McCollough CM, Fletcher JG: Projection space denoising with bilateral filtering and CT noise modeling for dose reduction in CT. Medical physics 36(11):4911–4919, 2009

42. Mao X, Li Q, Xie H, Lau RY, Wang Z (2016) Least squares generative adversarial networks. arXiv:161104076

43. Mathieu M, Couprie C, LeCun Y (2015) Deep multi-scale video prediction beyond mean square error. arXiv:151105440

44. Mirza M, Osindero S (2014) Conditional generative adversarial nets. arXiv:14111784

45. Odena A (2016) Semi-supervised learning with generative adversarial networks. arXiv:160601583

46. Odena A, Olah C, Shlens J (2016) Conditional image synthesis with auxiliary classifier gans. arXiv:161009585

47. Pathak D, Krahenbuhl P, Donahue J, Darrell T, Efros AA: Context encoders: feature learning by inpainting. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp 2536–2544, 2016

48. Po DY, Do MN: Directional multiscale modeling of images using the contourlet transform. IEEE Transactions on image processing 15(6):1610–1620, 2006

49. Ramani S, Fessler JA: A splitting-based iterative algorithm for accelerated statistical X-ray CT reconstruction. IEEE transactions on medical imaging 31(3):677–688, 2012

50. Reed S, Akata Z, Yan X, Logeswaran L, Schiele B, Lee H: Generative adversarial text to image synthesis. In: Proceedings of The 33rd International Conference on Machine Learning, vol 3, 2016

51. Reed SE, Akata Z, Mohan S, Tenka S, Schiele B, Lee H: Learning what and where to draw. In: Advances in Neural Information Processing Systems, pp 217–225, 2016b

52. Ronneberger O, Fischer P, Brox T: U-net: Convolutional networks for biomedical image segmentation. In: International Conference on Medical Image Computing and Computer-Assisted Intervention, Springer, pp 234–241, 2015

53. Sangkloy P, Lu J, Fang C, Yu F, Hays J (2016) Scribbler: Controlling deep image synthesis with sketch and color. arXiv:161200835

54. Shi J, Xu L, Jia J (2014) Blur detection dataset. http://www.cse.cuhk.edu.hk/leojia/projects/dblurdetect/dataset.html

55. Shi J, Xu L, Jia J (2014) Discriminative blur detection features

56. Shrivastava A, Pfister T, Tuzel O, Susskind J, Wang W, Webb R (2016) Learning from simulated and unsupervised images through adversarial training. arXiv:161207828

57. Simonyan K, Zisserman A (2014) Very deep convolutional networks for large-scale image recognition. arXiv:14091556

58. Suzuki K, Liu J, Zarshenas A, Higaki T, Fukumoto W, Awai K: Neural network convolution (nnc) for converting ultra-low-dose to "virtual" high-dose ct images. In: International Workshop on Machine Learning in Medical Imaging, Springer, pp 334–343, 2017

59. Tang C, Wu J, Hou Y, Wang P, Li W: A spectral and spatial approach of coarse-to-fine blurred image region detection. IEEE Signal Processing Letters 23(11):1652–1656, 2016

60. Tian Z, Jia X, Yuan K, Pan T, Jiang SB: Low-dose ct reconstruction via edge-preserving total variation regularization. Physics in medicine and biology 56(18):5949, 2011

61. Vinyals O, Toshev A, Bengio S, Erhan D: Show and tell: A neural image caption generator. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp 3156–3164, 2015

62. Vu CT, Phan TD, Chandler DM: A spectral and spatial measure of local perceived sharpness in natural images. IEEE Transactions on Image Processing 21(3):934–945, 2012

63. Walker J, Marino K, Gupta A, Hebert M (2017) The pose knows: Video forecasting by generating pose futures. arXiv:170500053

64. Wang Z, Bovik AC, Sheikh HR, Simoncelli EP: Image quality assessment: from error visibility to structural similarity. IEEE transactions on image processing 13(4):600–612, 2004

65. Wolterink JM, Leiner T, Viergever MA, Isgum I (2017) Generative adversarial networks for noise reduction in low-dose ct. IEEE Transactions on Medical Imaging

66. Xu Q, Yu H, Mou X, Zhang L, Hsieh J, Wang G: Low-dose x-ray ct reconstruction via dictionary learning. IEEE Transactions on Medical Imaging 31(9):1682–1697, 2012

67. Yang Q, Yan P, Kalra MK, Wang G (2017) Ct image denoising with perceptive deep neural networks. arXiv:170207019

68. Yang Q, Yan P, Zhang Y, Yu H, Shi Y, Mou X, Kalra MK, Wang G (2017) Low dose ct image denoising using a generative adversarial network with wasserstein distance and perceptual loss. arXiv:170800961

69. Yi X, Eramian M: Lbp-based segmentation of defocus blur. IEEE Transactions on Image Processing 25(4):1626–1638, 2016

70. Zhang H, Ma J, Wang J, Liu Y, Lu H, Liang Z: Statistical image reconstruction for low-dose ct using nonlocal means-based regularization. Computerized Medical Imaging and Graphics 38(6):423–435, 2014

71. Zhang H, Ma J, Wang J, Liu Y, Han H, Lu H, Moore W, Liang Z: Statistical image reconstruction for low-dose ct using nonlocal means-based regularization. part ii: An adaptive approach. Computerized Medical Imaging and Graphics 43:26–35, 2015

72. Zhang H, Sindagi V, Patel VM (2017) Image de-raining using a conditional generative adversarial network. arXiv:170105957

73. Zhu JY, Park T, Isola P, Efros AA (2017) Unpaired image-to-image translation using cycle-consistent adversarial networks. arXiv:170310593

74. Zhu M, Wright SJ, Chan TF: Duality-based algorithms for total-variation-regularized image restoration. Computational Optimization and Applications 47(3):377–400, 2010

75. Zhu X, Cohen S, Schiller S, Milanfar P: Estimating spatially varying defocus blur from a single image. IEEE Transactions on Image Processing 22(12):4879–4891, 2013

76. Zhuo S, Sim T: Defocus map estimation from a single image. Pattern Recognition 44(9):1852–1858, 2011