LOW DOSE ABDOMINAL CT IMAGE RECONSTRUCTION: AN UNSUPERVISED LEARNING BASED APPROACH

Shiba Kuanar[1], Vassilis Athitsos[2], Dwarikanath Mahapatra[3], K.R. Rao[1], Zahid Akhtar[4], Dipankar Dasgupta[4]

Department of [1]Electrical and [2]Computer Science Engineering, University of Texas Arlington, USA
[3] Research Staff Member, IBM Research, Melbourne, Australia
[4] Department of Computer Science, University of Memphis, USA

ABSTRACT

In medical practice, the X-ray Computed tomography-based scans expose a high radiation dose and lead to the risk of prostate or abdomen cancers. On the other hand, the low-dose CT scan can reduce radiation exposure to the patient. But the reduced radiation dose degrades image quality for human perception, and adversely affects the radiologist's diagnosis and prognosis. In this paper, we introduce a GAN based auto-encoder network to de-noise the CT images. Our network first maps CT images to low dimensional manifolds and then restore the images from its corresponding manifold representations. Our reconstruction algorithm separately calculates perceptual similarity, learns the latent feature maps, and achieves more accurate and visually pleasing reconstructions. We also showed the effectiveness of our model on a number of patient abdomen CT images, and compare our results with existing deep learning and iterative reconstruction methods. Experimental results demonstrate that our model outperforms other state-of-the-art methods in terms of PSNR, SSIM, and statistical properties of the image regions. https://github.com/ShibaPrasad/CT-Image-Reconstruction

Index Terms— Auto-encoder, Low dose CT Image, De-noise, Manifold, and GAN.

1. INTRODUCTION

The X-ray Computed Tomography (CT) has become a valuable modality in medical imaging to obtain direct visualization of patient anatomy [15] and demonstrate tremendous clinical values. But the CT exam involves in high radiation exposure to patients and leads to the concerns of lifetime risk of cancer [16], and [19]. On the contrary, lowering radiation dose increases artifact and noise in the reconstructed image. Due to the ill-posed nature of low dose CT image reconstruction, more advanced methods are needed to improve the diagnostic performance. Therefore, it is the central topic to develop effective image processing methods to reduce X-ray exposure while maintaining the clinically acceptable image quality. A simple way to lower X-ray exposure is to reduce the milliampere-second (mAs) level in a CT scan. However, this approach reduces number of X-ray photons reaching the detector, which increases the quantum noise of X-ray projection data. If reconstruction algorithm remains unchanged, the quality of reconstructed CT images would be deteriorated by noise-contaminated projection data and yields the amplified noise and streak artifacts [18].

Over the years, numerous research efforts have been devoted reconstructing the CT images and broadly fall into three groups a) sinogram pre-filter for reconstruction, b) iterative optimization, and c) image post-processing. As included in [22], the noise characteristics in sinogram domain are well modeled for pre-filtering. However, the sinogram data's are not widely accessible to users and may suffer from edge blurring, artifacts, and resolution loss. Differently, post-processing methods [21] directly operates on images and often introduces over-smoothing in the processed images. Typically, the iterative algorithms optimize an objective function, include image priors like the total variation (TV) [15], and a statistical noise model [20], [17]. As a result, these iterative algorithms indeed improves the image quality but at cost of high computation. With the given non-uniform noise distribution in CT images, these problems are even more difficult to address. In spite of the above algorithmic implementations, reconstructed image still loses some visual details and needs improvement. Inspired by the recent deep learning success [3], [1], [8], [4], biomedical researchers also extended the deep learning techniques on low dose CT reconstruction problems. Subsequent, complex models were proposed to handle low dose CT problems such as RED-CNN [9], Wasserstein GAN [24] and wavelet network [10].

Usually, the CT images are not uniformly distributed in a high dimensional Euclidean space. Mathematically, true CT images reside on a low-dimensional manifold structure formed by the image pixel values. The success of the above mentioned image processing approaches essentially makes an assumption about the structure of low-dimensional subspace, describes it in a geodesic distance term, and uses that to constrain the sharper and noise details in the image. Hence, the effectiveness of these methods lies in the accurate approximation of true feature map or manifold

learning on latent space. Deep neural networks are extremely effective to accurately capture the underlying manifold structures by studying data points in the high-dimensional space. With the numerous available patient CT images, it is potentially possible to derive the manifolds of the abdomen CT images using advanced neural networks. The derived manifold can be then employed in image reconstruction problem as a constraint to regulate the solution to image quality and remove contaminations. With this idea in mind, we propose a RegNet based auto-encoder network that learns manifold as an image prior and performs the image reconstruction. In auto-encoder model, the generative (G) and discriminator network (D) compete for one against the other and train in an end to end framework. While D is trained to discriminate between generated and real images, G is trained to produce realistic images from a random vector. The G and D are then trained by following a nonlinear mini-max optimization as described in [6], [23].

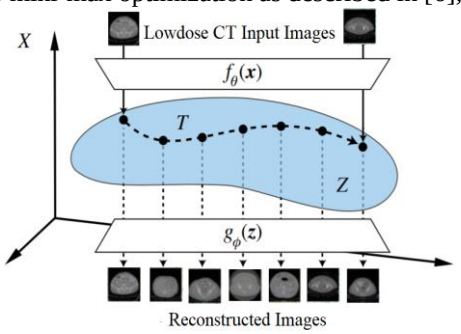

Fig. 1: Proposed reconstruction by manifold in latent space.

2. METHODS

To build an effective model for image reconstruction, we implemented a paired encoder-decoder interface of a simple GAN architecture [12] (Fig. 2). Let $x \in R^{N\times N}$ denotes a low dose CT image and $x' \in R^{N\times N}$ correspond to the normal dose CT image. The goal of our reconstruction process is to search for a function F that maps x to x', i.e. F: $x \rightarrow x'$. On the other hand, sample x is taken from low dose CT image with distribution P_L and x' from normal dose CT images with real distribution P_r. But our function F map samples from P_L into a certain distribution Pg. By varying function F, we treat the reconstruction operator as moving from one data distribution (P_g) close to other distribution (P_r). As shown in Fig. 1, our encoder maps input x into a latent space z: $f_\theta(x)$, and decoder learns function $g_\varphi(z)$ which brings back image x' close to the input image. The generator f_θ () transforms noisy samples to mimic a real sample, whereas $g_\varphi()$ is trained to do the classification task on the decoder side. During our training, the discriminator is trained to become an optimal classifier for a fixed G and generator minimize search by minimizing the Wasserstein distance between P_r and P_g distributions. Typically, noises in x-ray measurement are statistically modeled as a combination of quantum Poisson noise and additive white Gaussian noise. But the noise model in image reconstruct is usually complicated and distributed non-uniformly across the whole image. These uncertainties of noise models can be ignored in deep learning, as neural networks extract feature representations from input image patches through series of coarser-to-finer convolution and learn data distribution efficiently.

2.1. Perceptual Loss

In many real-time applications, a mathematical equation is used to compare the image similarity. Two images may look same to the normal human eyes but vary mathematically and may not produce the satisfactory results. During the reconstruction process, the color, texture, and object shapes are not properly preserved from layers of learning [3], [24]. To overcome the dissimilarity, a perceptual loss function is included to our network for feature learning. A separate perceptual loss function is included in the feature space to keep the image details and represented as below:

$$L_{RegNet}(D, G) = E_{x,z}\left[\frac{1}{WHd}\|RegNet(f_\theta(x)) - RegNet(NDCT)\|_F^2\right] \quad (1)$$

Where RegNet(f_θ(x)) is a feature vector extractor, and d, W, and H stand for depth, width, and height of the feature space respectively. In our implementation, we adopted a RegNet-50 pre-trained Keras model as our feature extraction. The input CT images are duplicated to make color channels and are fed into the RegNet network. The RegNet network contains fifty convolution layers followed by one fully connected layer (FC). The output of the final convolution layer is a feature vector extracted by our network and used in the perceptual loss function. For our convenience, we designate the perceptual loss as RegNet loss. So our final loss function is expressed as:

$$\min_G \max_D L_{WGAN}(D, G) + \lambda_1 L_{RegNet}(G) \quad (2)$$

where λ1 is a regularization term which controls the trade-off between RegNet perceptual loss and WGAN adversarial loss [5].

3. AUTO-ENCODER NETWORK

As shown in the Fig. 2, our model includes a scalable architecture with three components. The first component is the CNN generator which includes six convolution layers. The input CT image to the generator passes through a stack of convolution layers with various receptive fields with a kernel size of 3×3 and followed by a 2×2 max-pooling in each layer. All hidden layers are equipped with Rectified Linear Unit non-linearity (ReLU, max (0, x)) and applied thresholding on filter responses. The second part of the network contains a pre-trained RegNet network [4] (upper half in Fig. 2) and calculates perceptual loss for better image enhancement. The output manifold (z) from the generator and ground truth are fed into the RegNet pre-trained network for respective feature extraction. As explained in Eq. (1), the objective loss is computed using the extracted features from the above two specified layers. The reconstruction errors are

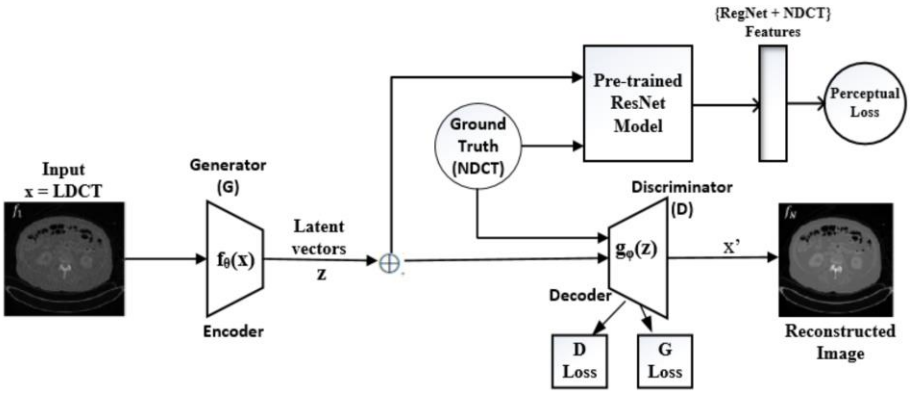

Fig. 2: Auto-encoder architecture: Generator, Discriminator, and RegNet for perceptual loss.

then back propagated to update the generator weight while keeping RegNet parameters intact. The third part of our model is a discriminator D, which includes eight convolution layers with a structure inspired by model [2], [4], and [7]. The first two convolution layers have 64 filters, followed by four convolution layers of 96 filters, and last two layers with 128 filters. All convolutional layers in discriminator have a kernel size of 3×3. After eight convolution layers, two fully-connected (FC) layers are included, of which first has 1024 outputs and the last layer has a single output. As explained in [5], the sigmoid cross entropy layer is not included at the end of the discriminator. The network is trained using image patches derived from entire images. After each epoch, we calculated the loss over all the image patches for validation. In Fig. 3 it is observed that increasing the number of epochs reduces the Wasserstein distance, as the decay rate becomes smaller. This indicates the effectiveness of our RegNet loss introduced in WGAN-RegNet model.

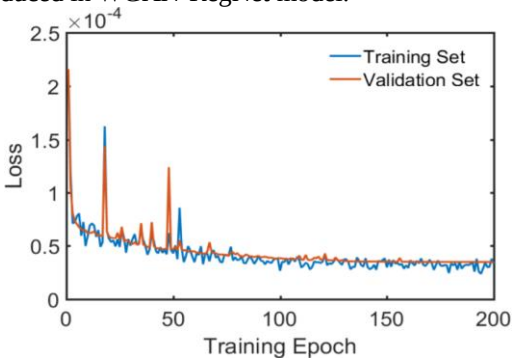

Fig. 3: Loss graphs of our model on training and validation dataset

4. EXPERIMENT

In our experiments, 4250 patient abdomen quarter dose CT images have been collected from publicly available cancer imaging archive (TCIA) [13] and AAPM-Mayo Clinic Low Dose 2018 datasets [11] for model training and validation. Among them, 3800 randomly selected images are used to train encoder-decoder network to generate image manifolds in latent space and rest are considered for testing purpose. During the training step, 10% of training data are employed as validation to monitor the network performance and tuned hyper parameters. All the CT images have a resolution of 512×512 pixels and a pixel size of 0.875×0.875 mm^2. In our data set, images with a different resolution are resampled to this resolution via bilinear interpolation.

4.1. Training

To use dataset efficiently we adopted two data augmentation techniques, i.e. horizontal flipping, and rotation by degrees of 90, 180 and 270. The objective loss functions L on Eq. (2) was minimized using the stochastic gradient descent optimization with standard back propagation. The layer weight parameters are initialized from a zero-mean Gaussian with 0.05 standard deviation. We followed the training steps explained in [12], [14] and performed our end-to-end optimization. The momentum parameter was set to 0.99 and training was regulated by a weight decay of 0.0001. The hyper-parameter λ_1 value was set to 7, learning rate to 1×10^{-4}, number of epochs to 200, and the number of iterations the discriminator followed were set to 10. For each epoch, the discriminator sampled a batch of 64 ground truth NDCT images. The generator output the latent space LDCT manifolds from respective input patches and updated the discriminator loss function recursively. Similarly, the generator processed a batch of input LDCT images, NDCT images (perceptual loss update), and updated generator loss. The whole process continued for a total number of 120 epochs. So both generator and discriminator counterfeited each other and optimized their respective loss functions with a global minimum. Our implementation was derived on python based Tensorflow/Keras framework with ten hours of training on an NVIDIA K40 GPU computer. The network parameters are converged after 600 iterations.

5. RESULTS AND ANALYSIS

In this section, we reported the effectiveness of our model in low dose CT reconstruction. The performance of proposed model was evaluated by using the PSNR and SSIM metrics on test images (Fig. 4). The experimental results of various models are summarized in Table II. It is observed that our WGAN-RegNet achieves the best in terms of SSIM, PSNR and RMSE metrics, and TV [15] method performs the worst.

Table I: Statistical Analysis (Mean ± SD) of Image Quality associated with different Models

Evaluation Metrics	Reference		TV [15]	FBPConvNet[26]	Wavelet-CNN [10]	Our Model
Artifact Reduction	R1	3.48 ± 0.33	2.16 ± 018	3.22 ± 0.39	3.78 ± 0.01	4.18 ± 0.29
	R2	3.55 ± 0.21	2.21 ± 0.13	3.27 ± 0.44	3.61 ± 0.24	4.06 ± 0.15
Contrast Retention	R1	3.32 ± 0.88	2.44 ± 0.48	3.41 ± 0.11	4.17 ± 0.55	4.22 ± 0.02
	R2	3.74 ± 0.91	2.26 ± 0.11	3.38±0.15	3.82 ± 0.39	4.19 ± 0.16
Noise Suppression	R1	3.49 ± 0.66	2.29 ± 0.72	3.18 ± 0.10	3.86 ± 0.63	3.95 ± 0.72
	R2	3.71 ± 0.14	2.31 ± 0.35	3.26 ± 0.64	3.65 ± 0.13	4.08 ± 0.17
Over all Image Quality	R1	3.83 ± 0.53	2.47 ± 0.12	3.28 ± 0.55	3.92 ± 0.41	4.23 ± 0.59
	R2	3.97 ± 0.47	2.58 ± 0.14	3.05 ± 0.33	3.56 ± 0.27	3.92 ± 0.66

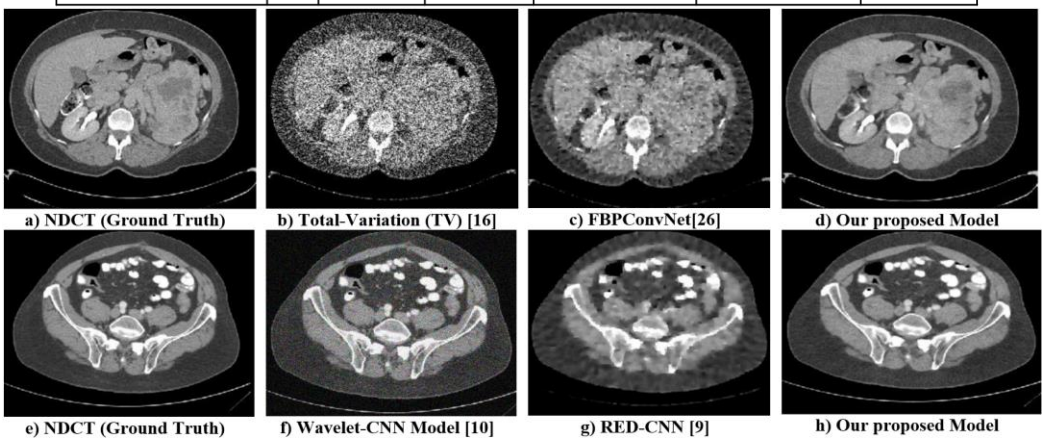

Fig. 4: Reconstruction comparison results on test samples. (a), (e) Ground truth; (b) Iterative total-variation; (c) FBPConvNet; (d), (h) Our proposed method; f) Wavelet-CNN method; g) RED-CNN method respectively. The CT images are displayed in a window [-210, 300] HU.

The comparisons are made between input CT images and its corresponding outputs from heuristics TV method, and CNN methods. We find that the reconstructed output images from our model (in Fig. 4), preserves most of the original input CT image structures, which demonstrates the capability of our model learning from the latent space. Quantitative evaluations of the low-dose CT reconstruction results using testing images are summarized in Table I, and II. It is observed that our proposed model outperformed other methods in all the metrics significantly.

Table II: Quantitative results from different model outputs

Models	PSNR	SSIM	RMSE
TV [15]	24.21	0.562	0.0392
FBPConvNet[26]	29.94	0.865	0.0574
RED-CNN [9]	30.82	0.892	0.0418
Wavelet-CNN [10]	33.51	0.913	0.045
Our Model	37.76	0.944	0.0092

For our quantitative evaluation, 25 references low dose CT images are used for the testing. Using different models the corresponding reconstructed images are randomly selected for expert evaluation (Table I). Artifact reduction, contrast retention, image quality, and noise suppression are included as qualitative indicators with five assessment grades: from 5 (best) to 1 (worst). Two radiologists R1 and R2 with years of clinical experience scored these images. The reference images are used as the gold standard. For each set of images, the scores are reported as (mean ± standard deviations). As demonstrated in Table I, the images reconstructed by TV are much poorer than that on the reference images in terms of the scores. The scores of our proposed model are closer to the ones of reference images which indicates the robustness of our model. The student's t-test with $p < 0.05$ is performed to assess discrepancy and results are summarized in Table I. The t-test results also show a similar trend that differences between reference images and the results from our model are statistically significant in all qualitative indices.

6. CONCLUSIONS

In this paper, we describe an autoencoder-RegNet based framework for the low-dose CT image reconstruction and compare it to baseline algorithms. The experimental results demonstrate that our model helps in improving image quality and avoids noise effect commonly suffered by MSE based image generators. The quantitative analysis shows that our model provides higher PSNR, SSIM, and better statistical properties of denoised CT images relative to those of normal CT images. However, using the GAN model alone reduces noise, but at the expense of losing some critical features. Therefore, an additive loss function i.e. RegNet perceptual loss is added to final cost functions that guide the denoising process. As a result, the reconstructed images become closer to the gold standard. The experimental results on the clinical real images show that our proposed auto-encoder model not only removes sharp features effectively but also generates an image with increased contrast. On a future direction, we would like to extend our model to find the image similarity search on latent space over huge clinical image dataset. This can be helpful in finding better treatment plans and spatial accuracy in the dose delivery for diagnosis and prognosis.

7. REFERENCES

[1] J. Johnson, A. Alahi, and F. Li, "Perceptual losses for real-time style transfer and super- resolution," ECCV, Springer, Amsterdam, Netherlands, pp. 694–711, Oct, 2016.

[2] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," CVPR, Las Vegas, June, 2016.

[3] Jun-Yan Zhu et al., "Generative Visual Manipulation on the Natural Image Manifold," ECCV, Amsterdam, Netherlands, Oct, 2016.

[4] K. Simonyan, and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," April, 2015. [Online]. Available: https://arxiv.org/abs/1409.1556.

[5] M. Arjovsky, S. Chintala, and L. Bottou, "Wasserstein GAN," Dec, 2017. [Online]. Available: https://arxiv.org/abs/1701.07875.

[6] I. Goodfellow et. al., "Generative adversarial nets," in Advances Neural Information Process. Systems (NIPS), pp. 2672–2680, Dec, 2014.

[7] C. Ledig, et al., "Photo-realistic single image super-resolution using a generative adversarial network," May, 2016. [Online]. Available: https://arxiv.org/pdf/1708.00961.pdf

[8] L. Gatys, A. Ecker, and M. Bethge, "A neural algorithm of artistic style", Sep, 2015. [Online] https://arxiv.org/abs/1508.0657

[9] H. Chen et al., "Low-dose CT with a residual encoder-decoder convolutional neural network," IEEE Transactions on Medical Imaging (TMI), vol. 36, pp. 2524–2535, Dec, 2017.

[10] E. Kang, J. Min, and J. Ye, "A deep convolutional neural network using directional wavelets for low-dose x-ray CT reconstruction," 2016. Available: arXiv:1610.09736

[11] AAPM, "Low dose CT grand challenge," 2017. [Online]. Available: http://www.aapm.org/GrandChallenge/LowDoseCT/#

[12] Diederik P Kingma, Max Welling, "Auto-Encoding Variational Bayes," Proceedings of the 2nd International Conference on Learning Representations (ICLR), Banff, Canada, April, 2014.

[13] Clark, K. et al., "The Cancer Imaging Archive (TCIA): maintaining and operating a public information repository. J. Digit. Image," pp. 1045–1057, July, 2013.

[14] K. H. Jin et al., "Deep convolutional neural network for inverse problems in imaging," IEEE TIP, pp. 4509–4522, vol. 26, Sep, 2017.

[15] Y. Liu et. al., "Adaptive-weighted total variation minimization for sparse data toward low-dose x-ray computed tomography image reconstruction," Physics in Medicine and Biology, vol. 57, pp. 7923-7956, Nov. 2012.

[16] R. Smith-Bindman et al., "Radiation dose associated with common computed tomography examinations and the associated lifetime attributable risk of cancer," Arch Intern Med., vol. 169, issue. 22, pp. 2078-86, Dec, 2009.

[17] H. Chen et al., "LEARN: Learned Experts' Assessment based Reconstruction Network for Sparse-data CT," IEEE Transactions on Medical Imaging, vol. 37, pp: 1333-1347, June, 2018.

[18] A. Kak, and M. Slaney, "Principles of computerized tomographic imaging," IEEE Engineering in Medicine and Biology Society, 1998.

[19] D. Brenner et al., "Estimated risks of radiation-induced fatal cancer from pediatric CT," American Journal of Roentgenology, vol. 176, pp. 289-296, Feb, 2001.

[20] S. Ramani, and J. Fessler, "A splitting-based iterative algorithm for accelerated statistical x-ray CT reconstruction," IEEE TMI, vol. 31, pp. 677–688, Mar. 2012.

[21] Y. Chen et al., "Improving abdomen tumor low-dose CT images using a fast dictionary learning based processing," Physics in Medicine and Biology, IOP Publishing, , pp. 5803-5820, 2013.

[22] J. Wang et al., "Sinogram noise reduction for low-dose CT by statistics based nonlinear filters," Medical Imaging, SPIE, vol. 5747, pp. 2058-2066, San Diego, United States, April, 2005.

[23] J. Zhu, T. Park, P. Isola, A. Efros, 'Unpaired Image-To-Image Translation Using Cycle-Consistent Adversarial Networks," IEEE ICCV, pp. 2223-2232, Venice, Italy, Oct, 2017.

[24] Q. Yang et al., "Low dose CT image denoising using a generative adversarial network with Wasserstein distance and perceptual loss", IEEE TMI, vol. 37, pp. 1348-1357, June 2018.