A Generative Network with Dual-Domain Discriminators for Low-Dose Stationary Sources CT Imaging

Xiao Bai
School of Biomedical Engineering, ShanghaiTech University, Shanghai, China
baixiao@shanghaitech.edu.cn

Ying Cheng
School of Biomedical Engineering, ShanghaiTech University, Shanghai, China
chengying2022@shanghaitech.edu.cn

Linjie Chen
School of Biomedical Engineering, ShanghaiTech University, Shanghai, China
chenglj2022@shanghaitech.edu.cn

Shuo Yang
School of Biomedical Engineering, ShanghaiTech University, Shanghai, China
yangshuo1@shanghaitech.edu.cn

Huamin Wang
School of Biomedical Engineering, ShanghaiTech University, Shanghai, China
wanghm1@shanghaitech.edu.cn

Zhe Wang
School of Biomedical Engineering, ShanghaiTech University, Shanghai, China
wangzhe2@shanghaitech.edu.cn

Jiayi Wu
School of Biomedical Engineering, ShanghaiTech University, Shanghai, China
wujy12022@shanghaitech.edu.cn

Guohua Cao*
School of Biomedical Engineering, ShanghaiTech University, Shanghai, China
caogh@shanghaitech.edu.cn

ABSTRACT

Recent development of clinical Computed tomography (CT) technologies has led to research for novel CT systems that allow safer and faster imaging, such as low-dose cardiac CT imaging via stationary CT. However, the complex data acquisition schemes in stationary CT often cause severe artifacts and noise in the resulted images; this calls for the development of a new kind of image reconstruction algorithms. Recent advancements in deep learning have shown remarkable progress in medical image reconstruction, processing, and analysis. In this paper, we propose a generative network with dual-domain discriminators for low-dose CT reconstruction in a stationary CT system. The image-domain discriminator optimizes the generation network by comparing the generated CT images with the reference images, while the sinogram-domain discriminator preserves the structure of the sinograms and suppresses the noise. The network incorporates uncertainty to automatically adjust the weights of a multi-term loss function, eliminating the need for the manual tuning of hyperparameters in the loss function. The results from our numerical experiments demonstrate the effectiveness of our proposed reconstruction algorithm for low-dose imaging in stationary CT.

CCS CONCEPTS

• **Computing methodologies**; • **Artificial intelligence**; • **Reconstruction**;

KEYWORDS

Image reconstruction, Deep learning, Dual-domain, Stationary sources, Low-Dose CT

ACM Reference Format:
Xiao Bai, Ying Cheng, Linjie Chen, Shuo Yang, Huamin Wang, Zhe Wang, Jiayi Wu, and Guohua Cao. 2023. A Generative Network with Dual-Domain Discriminators for Low-Dose Stationary Sources CT Imaging. In *2023 6th International Conference on Digital Medicine and Image Processing (DMIP 2023), November 09–12, 2023, Kyoto, Japan.* ACM, New York, NY, USA, 5 pages. https://doi.org/10.1145/3637684.3637712

1 INTRODUCTION

Computed tomography (CT), utilizing X-ray radiation, provides cross-sectional images of the body and is one of the most important imaging modalities in clinical diagnosis [1]. One of the major developments in CT primarily focuses on improving the temporal resolution of cardiac CT, because higher temporal resolution in cardiac CT scans allows for more effective "freezing" of heart motion, resulting in clearer images with fewer artifacts. Currently, the speed of cardiac computed tomography (CT) scans is not fast enough for patients with very high or irregular heart rates, leading to noticeable motion artifacts in cardiac images. These artifacts may potentially impact the diagnostic accuracy of disease detection.

To achieve higher temporal resolution, researchers have previously investigated two strategies. The first strategy involves attaining a faster gantry rotation speed [2]. Over the past decades, significant improvements have been made in temporal resolution with increasingly faster rotation speeds. Typically, CT scanners with a single X-ray source can achieve scanning speeds of up to

*Corresponding author.

DMIP 2023, November 09–12, 2023, Kyoto, Japan

ACM ISBN 979-8-4007-0942-5/23/11
https://doi.org/10.1145/3637684.3637712

3 Hz. However, pushing for higher temporal resolution through faster gantry rotation speed is approaching the current mechanical limits. While rotating-gantry CT dominates in hospitals and clinics today, they may not deliver ideal imaging performance in challenging cases [3].

The second strategy involves obtaining multiple projections simultaneously through multi-source-detector imaging chains. It is well-known that this strategy can enhance the temporal resolution of CT scanners. Liu et al. improved the image quality in a simulated five-source cone-beam micro-CT using the Feldkamp-type reconstruction algorithm [4]. Cao et al. proposed a CT system with a stationary-source-and-rotating-detector architecture, which includes three stationary X-ray source arrays and smaller rotating X-ray detectors [5]. Yueh et al. proposed a fixed-head CT prototype with an array of three carbon nanotube X-ray sources [6].

Another significant improvement in CT is the reduction of radiation dose. Computed tomography (CT) is an important medical imaging modality for diagnostic purposes. However, patient exposure to X-ray radiation during CT examinations increases the risk of cancer [7, 8]. Therefore, there is a pressing need for low-dose CT (LDCT) techniques. In CT scanning, the most common approach to reducing radiation dose is by decreasing X-ray tube current (or voltage). However, images reconstructed by conventional algorithms from LDCT always suffer from severe noise.

Stationary multi-source CT with multiple x-ray sources can improve temporal resolution and reduce radiation dose. However, the image reconstruction process is complex, and the reconstructed images often suffer from severe artifacts and noise. In recent years, deep learning methods have made significant advancements in medical image reconstruction, segmentation, post-processing, and analysis [9-11]. Using deep learning frameworks, low-dose CT denoising methods have achieved the state-of-the-art performance. Chen et al. proposed a residual encoder-decoder CNN (RED-CNN) that uses residual connections between the encoder and decoder to suppress noise in LDCT images [12]. Yang et al. added a perceptual loss to the Wasserstein generative adversarial network (WGAN) and proposed the WGAN-VGG network model to reduce image noise while preserving important image details [13]. Zhang et al. proposed a framework utilizing independent computing and searching units to achieve LDCT denoising [14]. Guo et al. proposed a dual-domain denoising network with an added texture-aware mechanism [15].

In this paper, we employ deep learning methods to reconstruct low-dose images generated by a stationary multi-source CT system. We propose a generative network with dual-domain discriminators (DUD-WGAN) for low-dose CT reconstruction of stationary-source CT. Our neural network consists of a "generator" network responsible for CT image reconstruction and two "discriminator" networks. The image-domain discriminator compares the generated CT images with reference CT images to optimize the generation network's ability to produce high-quality images. The sinogram-domain discriminator compares the forward-projected sinograms of the generated CT images with reference sinograms. The internal structure of the sinogram typically represents data characteristics in the sinogram domain, but this structure can be disrupted by randomly occurring noise during the imaging process. By designing a discriminator based on the internal structure of the sinogram, the structure in the sinogram domain can be effectively preserved. Furthermore, uncertainty-based weights are introduced to automatically adjust the hyperparameters of the loss function. Manual adjustment of the weights for each loss term is challenging and time-consuming, as the network's performance strongly depends on the relative weights between the loss terms. With automatic adjustment, the optimal parameters can be easily obtained. Experiments were conducted on a simulated dataset to demonstrate the effectiveness of the proposed method through qualitative and quantitative comparisons.

In the subsequent section, the overall architecture of the proposed network is described. Then, we present the comprehensive details regarding the datasets utilized in the experiments, the employed hyperparameters, and a comparative analysis of the results obtained with other reference methods. In the final section, the impact, limitations, future work, and conclusion of this study are discussed.

2 METHODS

In this section, each component of the proposed DUD-WGAN network is introduced. First, the generator part of the network is presented. Next, detailed information about the DUD-WGAN image-domain discriminator and sinogram discriminator are provided. Finally, the overall loss function used for training is presented.

The proposed DDU-WGAN aims to train an optimal generator G^* that can estimate NDCT (normal-dose CT) images from stationary source LDCT images, as shown in Equation (1):

$$G^*\left(I_{LD}\right) \approx I_{ND},, I_{LD} = FBP(P_{LD}) \tag{1}$$

Here, I_{LD} represents the stationary source low-dose CT image, I_{ND} represents the stationary source normal-dose CT image, and P_{LD} represents the measured low-dose projection. FBP refers to the filtered back projection algorithm.

2.1 Overview of DUD-WGAN

The overall architecture is shown in Figure 1. The proposed DUD-WGAN network is optimized within the Wasserstein Generative Adversarial Network (WGAN) framework [16], which is one of the state-of-the-art frameworks capable of effectively reducing blurriness caused by the Mean Squared Error (MSE) loss. In this study, the proposed framework consists of three components: the generator network G, the image-domain discriminator network D_{IMG}, and the sinogram-domain discriminator D_{SIN}. The goal of G is to denoise and de-artifact CT images that contain artifacts and noise, resulting in high-quality CT images. D_{IMG} inputs images from G and real images from the ground truth dataset, while D_{SIN} inputs sinograms from G and real sinograms from the ground truth dataset, both aiming to discriminate between real and generated inputs. The introduction of the discriminators helps improve the texture of reconstructed images. All the three networks undergo optimization simultaneously during the training process.

Unlike conventional Generative Adversarial Networks (GAN) [17], the WGAN framework utilizes the Wasserstein distance instead of the logarithmic term in the loss function, which enhances training stability. The objective function of the WGAN framework

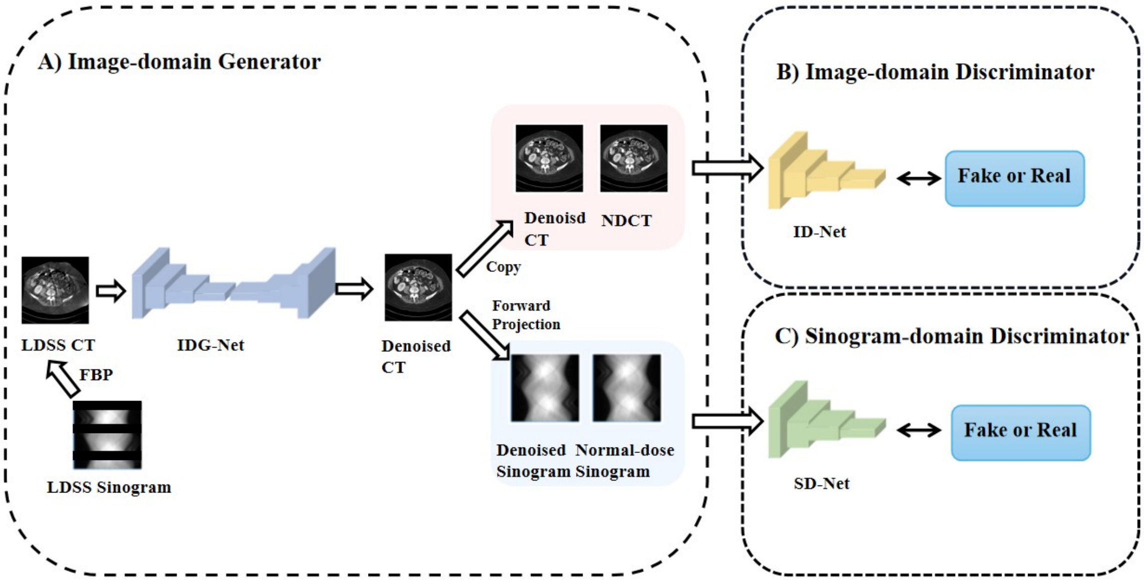

Figure 1: Architecture of the proposed DUD-WGAN. The generator generates denoised CT images, while two separate branches with discriminators operate in the image and sinogram domains. Where LDSS stands for low-dose stationary sources CT, IDG-NET refers to the image-domain generator network, ID-NET represents the image-domain discriminator network, and SD-net denotes the sinogram-domain discriminator network.

is expressed as follows:

$$
\begin{aligned}
min_G max_{D_{IMG} D_{SIN}} \{[-E_{I_{ND}}\left(D_{IMG}\left(I_{ND}\right)\right) + E_{I_{LD}}\left(D_{IMG}\left(G\left(I_{LD}\right)\right)\right) \\
+\lambda E_{\hat{I}}\left(\left(\left\|\nabla_{\hat{I}} D_{IMG}\left(\hat{I}\right)\right\|_2 - 1\right)^2\right)] \\
+[-E_{S_{ND}}\left(D_{SIN}\left(S_{ND}\right)\right) + E_{S_{LD}}\left(D_{SIN}\left(G\left(I_{LD}\right)\right)\right) \\
+\lambda E_{\hat{S}}\left(\left(\left\|\nabla_{\hat{S}} D_{SIN}\left(\hat{S}\right)\right\|_2 - 1\right)^2\right)]\}
\end{aligned}
\tag{2}
$$

where I_{LD} represents the stationary-source low-dose CT image, I_{ND} represents the stationary-source normal-dose CT image, S_{ND} represents the sinogram obtained by performing the forward projection (FP) on I_{ND}, respectively. $E(a)$ denotes the expectation of a, and $\nabla(\cdot)$ represents the gradient. $\hat{I}$ represents uniform sampling along lines connecting corresponding points between the generated image and the original image. $\hat{S}$ represents uniform sampling along lines connecting corresponding points between the generated sinogram data and the original sinogram data. λ is a parameter used to balance the Wasserstein distance and the gradient penalty.

2.2 Generator

In this paper,the ResUNet [18] was utilized as the network generator. ResUNet consists of operations such as Conv, DeConv, Resblock, and Concat. The kernel size for all convolutions and deconvolutions is set to 3×3. The Resblock includes two consecutive convolutions and a residual connection. Leaky ReLU [19] is used as the activation function. The stride for both downsampling and upsampling is set to 2. During the downsampling process, the number of feature channels increases from 32 to 256. During the upsampling process, it decreases from 256 to 1.

2.3 Dual-domain Discriminators

The discriminator network D takes inputs from G or the ground truth dataset and attempts to classify the authenticity of the input data. DUD-WGAN consists of two discriminators, namely the image-domain discriminator and the sinogram discriminator, which have identical structures. The discriminator network comprises six convolutional layers with 64, 64, 128, 128, 256, and 256 filters, followed by two fully connected layers with 1024 and 1 neurons, respectively. Leaky ReLU activation function with a negative slope of 0.2 is applied after each layer. All convolutional layers employ 3×3 kernel size and zero-padding for two-dimensional convolutions. The stride for each layer is set to two.

2.4 Loss Functions for Generator

In this subsection, we introduce different objective functions used for reducing artifacts and noise. We employ a composite objective function L to optimize DUD-GAN. The functions used are shown in Equation 3:

$$
L = \frac{1}{2(\sigma_1)^2} L_{WGAN} + \frac{1}{2(\sigma_2)^2} L_{ML} + log\left(\sigma_1\sigma_2\sigma_3\sigma_4\sigma_5\right) \tag{3}
$$

Where $\sigma_k (k = 1, \dots, 5)$ represents trainable hyperparameters that are automatically adjusted through a neural network. We used uncertainty [20] as a mean of automatic weight learning, enabling us to learn the relative weights of each loss term in a principled and informed manner. This uncertainty-weighted loss is smooth, differentiable, and well-structured, preventing the task weights from converging to zero.

Mean Squared Error (MSE) [21, 22] is a popular choice for denoising and artifact removal. However, it can lead to over-smoothed images [23]. On the other hand, utilizing Multi-Scale Structural Similarity (MS-SSIM) [24] helps preserve the global structural similarity between images. In the image domain, we employ a mixed loss function of MSE and MS-SSIM, while in the sinogram domain, we only use the MSE loss to preserve the pixel accuracy in the sinogram domain. L_{ML} refers to the multi-loss that includes both the image domain and sinogram domain.

$$L_{ML} = \frac{1}{2(\sigma_3)^2}\|G(I_{LD}), I_{ND}\|_2 + \frac{1}{2(\sigma_4)^2}MS\_SSIM(G(I_{LD}), I_{ND}) + \frac{1}{2(\sigma_5)^2}\|S_{PRE}, S_{ND}\|_2 \quad (4)$$

where $S_{PRE}$is obtained from $G(I_{LD})$ using forward projection operation. S_{ND} is obtained from I_{ND} via forward projection operation $\|\cdot\|_2$ is the MSE term.

The purpose of the adversarial loss is to make the generator produce images that the discriminator network cannot distinguish from real ones. L_{WGAN} represents the adversarial loss of the DUD-WGAN network.

$$L_{WGAN} = -E_{I_{LD}}(D_{IMG}(G(I_{LD}))) - E_{S_{LD}}(D_{SIN}(G(I_{LD}))) \quad (5)$$

2.5 Stationary-Sources CT System

We employed the SSRD-CT system as described by Cao et al [5]. It consists of three fixed distributed X-ray sources and three rotating detectors. The three identical source-detector chains are symmetrically positioned around an object. It has parameters similar to most commercial CT scanners, with a source-to-isocenter distance (SID) of 540mm and a source-to-detector distance (SDD) of 950mm. The sinogram data generated by this system differs from sinograms produced by conventional CT, as it lacks data in the range of [0, 81.4○], [81.4°, 120°] and [201.4°, 240°] during a 360° scan. Consequently, CT images reconstructed using the FBP algorithm exhibit artifacts and noise.

3 EXPERIMENTAL DESIGN AND RESULTS

3.1 Datasets

To train and evaluate the proposed network, we utilized the publicly available "2016 NIH-AAPM-Mayo Clinic Low Dose CT Grand Challenge" dataset [9]. This dataset consists of real clinical data, including normal-dose abdominal CT images from 10 anonymous patients and the corresponding simulated quarter-dose CT images.

To simulate data acquired from a stationary x-ray sources CT system, we performed forward projection on the quarter-dose CT images. The distances from source to detector and from source to rotation center were set to 950.00 mm and 540.00 mm, respectively. There were 888 detector elements, and each had a dimension of 1 mm. The forward projections were collected from 720 views with a 0.5-degree angular interval in this study. The start angle is 0 degrees, and the end angle is 359.5 degrees. Only projections in [0, 81.4○], [120○, 201.4○], and [240○, 321.4○] are needed for the SSRD-CT architecture.

Table 1: Quantitative comparison of the proposed DUD-WGAN with other reference methods (Mean ± SD).

Method	PSNR	SSIM
FBP	18.7755±1.734	0.6622±0.0477
RED-CNN	27.8104±1.0351	0.8322±0.0249
WGAN-VGG	26.6742±1.5310	0.8367±0.0228
DUD-WGAN	**29.5319±1.0927**	**0.8511±0.0245**

3.2 Training and Implementation Details

The DUD-WGAN method was implemented using PyTorch and trained/validated on a workstation equipped with an NVIDIA A100 80GB PCIe GPU. During the end-to-end training process, the Adam algorithm was used to optimize all parameters. The learning rate for Adam was set to $\alpha = 1 \times 10^{-3}$, and the two exponential decay rates were set to $\beta_1 = 0.5$ and $\beta_2 = 0.9$. The hyperparameters $\sigma_k(k = 1, \ldots, 5)$ are automatically learned during the training process.

4 RESULTS

In order to evaluate the performance of the proposed method, it is compared with several state-of-the-art methods., namely RED-CNN and WGAN-VGG. RED-CNN is a CNN-based method, while WGAN-VGG is a GAN-based method. To visualize the reconstruction results of different algorithms, we present a representative slice of case L096 from the Mayo dataset in Figure 2. The display window was set to [-160, 240]. As shown in Figure 2, there are noticeable artifacts in the FBP image. The image generated by RED-CNN is overly smooth and lacks texture information. Compared to RED-CNN, WGAN-VGG improves the visual quality of the images. Our proposed DUD-WGAN achieves the best image quality. Figure 3 shows the absolute difference images, where brighter shades indicate larger errors. Table 1 provides a quantitative analysis of different methods. From Table 1, it can be observed that DUD-WGAN outperforms other methods, achieving the highest PSNR and SSIM.

5 CONCLUSION

In conclusion, we proposed a generative network with dual-domain discriminators for low-dose stationary sources CT reconstruction. The network consists of an image domain generator and two discriminators—one for the image domain and another for the sinogram domain. Through the incorporation of the sinogram discriminator, a better differentiation between the generated images and the ground truth images is achieved from the perspective of sinograms, thereby assisting in the denoising task by capturing sinogram-level details. Additionally, uncertainty is utilized to automatically adjust the weights of the composite loss function. Through this automatic adjustment, improved and optimized results can be readily obtained. Experimental results demonstrate that the proposed method effectively reconstructs low-dose stationary-source CT images with good artifact reduction and denoising capabilities.

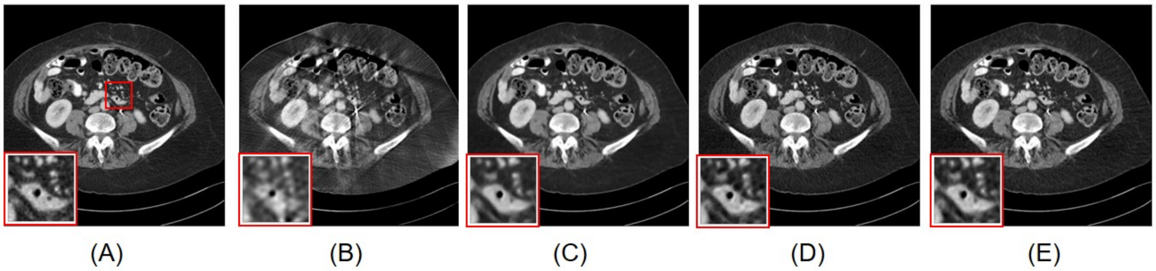

Figure 2: Results of different methods on the Mayo data. Zoomed ROI of the red rectangle is shown below the full-size one. (A) NDCT, (B) FBP, (C) RED-CNN, (D) DUD-WGAN. The display window is [−140, 260] HU.

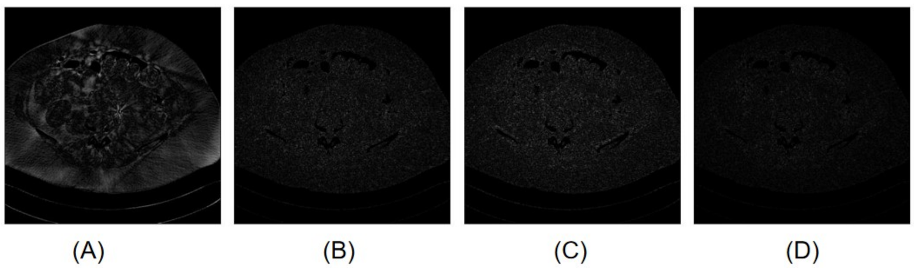

Figure 3: Absolute difference images relative to the NDCT image. (A) FBP, (B) RED-CNN, (C)WGAN-VGG, (D) DUD-WGAN. The display window is [0,200] HU.

ACKNOWLEDGMENTS

This work was partially supported by Dr. Cao's New Faculty Startup Fund at ShanghaiTech University, and his National Natural Science Foundation of China Grant no: 62273238.

REFERENCES

[1] Hounsfield, G. N. Computerized transverse axial scanning (tomography): Part 1. Description of system. *The British journal of radiology*, 46, 552 (1973), 1016-1022.

[2] Primak, A. N., McCollough, C. H., Bruesewitz, M. R., Zhang, J. and Fletcher, J. G. Relationship between noise, dose, and pitch in cardiac multi–detector row CT. *Radiographics*, 26, 6 (2006), 1785-1794.

[3] Samson, K. A Mobile Stroke CT Unit Cuts tPA Administration Times By One-Third: Would It Work in the US? *Neurology Today*, 12, 24 (2012), 1-16.

[4] Liu, Y., Liu, H., Wang, Y. and Wang, G. Half-scan cone-beam CT fluoroscopy with multiple x-ray sources. *Medical physics*, 28, 7 (2001), 1466-1471.

[5] Cao, G., Liu, B., Gong, H., Yu, H. and Wang, G. A stationary-sources and rotating-detectors computed tomography architecture for higher temporal resolution and lower radiation dose. *IEEE Access*, 2 (2014), 1263-1271.

[6] Spronk, D., Luo, Y., Billingsley, A., Inscoe, C. R., Zhou, O., Lu, J. and Lee, Y. Z. *A stationary head CT prototype with CNT x-ray source arrays*. SPIE, City, 2022.

[7] Brenner, D. J. and Hall, E. J. Computed tomography—an increasing source of radiation exposure. *New England journal of medicine*, 357, 22 (2007), 2277-2284.

[8] Pearce, M. S., Salotti, J. A., Little, M. P., McHugh, K., Lee, C., Kim, K. P., Howe, N. L., Ronckers, C. M., Rajaraman, P. and Craft, A. W. Radiation exposure from CT scans in childhood and subsequent risk of leukaemia and brain tumours: a retrospective cohort study. *The Lancet*, 380, 9840 (2012), 499-505.

[9] Chen, H., Zhang, Y., Zhang, W., Liao, P., Li, K., Zhou, J. and Wang, G. Low-dose CT via convolutional neural network. *Biomedical optics express*, 8, 2 (2017), 679-694.

[10] Ghani, M. U. and Karl, W. C. *Deep learning-based sinogram completion for low-dose CT*. IEEE, City, 2018.

[11] Shen, L., Zhao, W. and Xing, L. Patient-specific reconstruction of volumetric computed tomography images from a single projection view via deep learning. *Nature biomedical engineering*, 3, 11 (2019), 880-888.

[12] Chen, H., Zhang, Y., Kalra, M. K., Lin, F., Chen, Y., Liao, P., Zhou, J. and Wang, G. Low-dose CT with a residual encoder-decoder convolutional neural network. *IEEE transactions on medical imaging*, 36, 12 (2017), 2524-2535.

[13] Yang, Q., Yan, P., Zhang, Y., Yu, H., Shi, Y., Mou, X., Kalra, M. K., Zhang, Y., Sun, L. and Wang, G. Low-dose CT image denoising using a generative adversarial network with Wasserstein distance and perceptual loss. *IEEE transactions on medical imaging*, 37, 6 (2018), 1348-1357.

[14] Shen, J., Luo, M., Liu, H., Liao, P., Chen, H. and Zhang, Y. MLF-IOSC: Multi-Level Fusion Network With Independent Operation Search Cell for Low-Dose CT Denoising. *IEEE Transactions on Medical Imaging*, 42, 4 (2022), 1145-1158.

[15] Wang, H., Zhao, X., Liu, W., Li, L. C., Ma, J. and Guo, L. Texture-aware dual domain mapping model for low-dose CT reconstruction. *Medical Physics*, 49, 6 (2022), 3860-3873.

[16] Gulrajani, I., Ahmed, F., Arjovsky, M., Dumoulin, V. and Courville, A. C. Improved training of wasserstein gans. *Advances in neural information processing systems*, 30 (2017).

[17] Creswell, A., White, T., Dumoulin, V., Arulkumaran, K., Sengupta, B. and Bharath, A. A. Generative adversarial networks: An overview. *IEEE signal processing magazine*, 35, 1 (2018), 53-65.

[18] Diakogiannis, F. I., Waldner, F., Caccetta, P. and Wu, C. ResUNet-a: A deep learning framework for semantic segmentation of remotely sensed data. *ISPRS Journal of Photogrammetry and Remote Sensing*, 162 (2020), 94-114.

[19] Xu, J., Li, Z., Du, B., Zhang, M. and Liu, J. *Reluplex made more practical: Leaky ReLU*. IEEE, City, 2020.

[20] Kendall, A., Gal, Y. and Cipolla, R. *Multi-task learning using uncertainty to weigh losses for scene geometry and semantics*. City, 2018.

[21] Wang, Z. and Bovik, A. C. Mean squared error: Love it or leave it? A new look at signal fidelity measures. *IEEE signal processing magazine*, 26, 1 (2009), 98-117.

[22] Wolterink, J. M., Leiner, T., Viergever, M. A. and Išgum, I. Generative adversarial networks for noise reduction in low-dose CT. *IEEE transactions on medical imaging*, 36, 12 (2017), 2536-2545.

[23] Zhao, H., Gallo, O., Frosio, I. and Kautz, J. Loss functions for image restoration with neural networks. *IEEE Transactions on computational imaging*, 3, 1 (2016), 47-57.

[24] Wang, Z., Simoncelli, E. P. and Bovik, A. C. *Multiscale structural similarity for image quality assessment*. Ieee, City, 2003.