Generative Adversarial Network with Robust Discriminator Through Multi-Task Learning for Low-Dose CT Denoising

Sunggu Kyung, Jongjun Won, Seongyong Pak, Sunwoo Kim, Sangyoon Lee, Kanggil Park, Gil-Sun Hong, and Namkug Kim, *Senior Member, IEEE*

***Abstract*—Reducing the dose of radiation in computed tomography (CT) is vital to decreasing secondary cancer risk. However, the use of low-dose CT (LDCT) images is accompanied by increased noise that can negatively impact diagnoses. Although numerous deep learning algorithms have been developed for LDCT denoising, several challenges persist, including the visual incongruence experienced by radiologists, unsatisfactory performances across various metrics, and insufficient exploration of the networks' robustness in other CT domains. To address such issues, this study proposes three novel accretions. First, we propose a generative adversarial network (GAN) with a robust discriminator through multi-task learning that simultaneously performs three vision tasks: restoration, image-level, and pixel-level decisions. The more multi-tasks that are performed, the better the denoising performance of the generator, which means multi-task learning enables the discriminator to provide more meaningful feedback to the generator. Second, two regulatory mechanisms, restoration consistency (RC) and non-difference suppression (NDS), are introduced to improve the discriminator's representation capabilities. These mechanisms eliminate irrelevant regions and compare the discriminator's results from the input and restoration, thus facilitating effective GAN training. Lastly, we incorporate residual fast Fourier transforms with convolution (Res-FFT-Conv) blocks into the generator to utilize both frequency and spatial representations. This approach provides mixed receptive fields by using spatial (or local), spectral (or global), and residual connections. Our model was evaluated using various pixel- and feature-space metrics in two denoising tasks. Additionally, we conducted visual scoring with radiologists. The results indicate superior performance in both quantitative and qualitative measures compared to state-of-the-art denoising techniques.**

***Index Terms*—Fourier transform, Generative adversarial network, Low-dose CT denoising, Multi-task learning, Robust discriminator**

I. INTRODUCTION

COMPUTED tomography (CT) provides volumetric images of the body's interior through X-ray radiation, serving as a vital diagnostic modality in modern medicine. However, in the course of receiving several CT scans, patients may experience cumulative radiation exposure, introducing potential health risks such as an increased possibility of cancer [1, 2]. To minimize such risks, radiologists have reduced CT radiation dose to as low as reasonably achievable (ALARA) [3]. Unfortunately, the decrease in radiation dose inevitably leads to heightened noise and artifacts within the CT images, which substantially undermine the diagnosis and patient care. Achieving satisfactory image quality in low-dose CT (LDCT) remains a challenging area of research due to the irregular distribution of CT noise [4]. With the introduction of the American Association of Physicists in Medicine (AAPM) clinic simulation dataset [5] and the recent development of deep learning (DL) technology, several studies [6-12] have demonstrated significant improvements in abdominal LDCT denoising. Despite their promising results, major challenges to real-world clinical application remain. First, as the images synthesized by DL-based networks are often over-smoothed and the clinically important signals are distorted, radiologists consider the synthesized images to be visually incongruent. Second, the evaluation of most previous denoising methods has often been limited to only a few metrics or a single metric type, which can produce biased results. Specifically, overprioritizing

S. Kyung, J. Won, S. Pak, and K. Park are with the Department of Biomedical Engineering, Asan Medical Institute of Convergence Science and Technology, Asan Medical Center, University of Ulsan College of Medicine, Seoul, South Korea (e-mail: babbu3682@gmail.com, wjj910@gmail.com, seongyong.pak@gmail.com, pkg777774@gmail.com).

S. Kim and S. Lee are with the Department of Convergence Medicine, University of Ulsan College of Medicine, Asan Medical Center, Seoul, South Korea (e-mail: ksunnywoo@gmail.com, lsangyoon7454@gmail.com).

G.S. Hong is with the Department of Radiology and Research Institute of Radiology, University of Ulsan College of Medicine, Asan Medical Center, Seoul, South Korea (e-mail: hgs2013@gmail.com).

N. Kim is with the Department of Convergence Medicine, University of Ulsan College of Medicine, Asan Medical Center, Seoul, South Korea (email: namkugkim@gmail.com) (G.S. Hong and N. Kim are the co-corresponding authors for this work. These authors contributed equally to this work.)

This research was supported by a grant from the Korea Health Technology R&D Project through the Korea Health Industry Development Institute (KHIDI), funded by the Ministry of Health & Welfare, Republic of Korea (HR20C0026, HI22C0471, HI21C1148).

pixel-based objectives and metrics improves the signal-to-noise ratio but risks over-smoothing and awkward outcomes, whereas overemphasizing feature-based objectives and metrics may preserve details but can potentially distort crucial signals and cause undesirable artifacts. Therefore, it takes considerable effort to find a clinically applicable denoising network that demonstrates robust performance across various pixel- and feature-wise metrics. Third, many LDCT denoising tasks have been primarily focused on abdominal CT images, and the exploration of other CT image domains remains relatively insufficient. Owing to the considerable variation in anatomy and characteristics across body locations, radiologists utilize a range of windowing Hounsfield unit (HU) settings tailored to each location to emphasize distinct targets and regions. For instance, when CT images exhibit a multitude of regions with varying intensities, such as the lung and chest, a wide windowing HU setting is applied. Conversely, when CT images contain numerous regions with similar intensities, as observed in the brain and soft tissues, a narrow windowing HU setting is utilized. For practical applications in medicine, robust networks that demonstrate excellent performance in various environments must be explored.

In this paper, we introduce the multi-task discriminator generative adversarial network (MTD-GAN) to address these issues with a few novel strategies:

- Develop a discriminator utilizing multi-task learning (MTL), which leverages three simultaneous tasks—restoration, image-level decision, and pixel-level decision—to transfer contextual, global, and local feedback between the real normal-dose and synthesized images to the generator.
- Propose two regulations to improve the representation capabilities of the discriminator: restoration consistency (RC), which compares the discriminator's outputs from the input data with the corresponding restoration data generated by our MTL discriminator for consistency, and non-difference suppression (NDS), which excludes areas that cause confusion in discriminator decisions.
- Design a novel generator that consists of residual fast Fourier transform with convolution (Res-FFT-Conv) blocks [13] that fuse frequency-spatial dual-domain representations. The proposed generator effectively captures rich information by simultaneously utilizing spatial (or local), spectral (or global), and residual connections. To the best of our knowledge, this represents an inaugural effort in employing the Res-FFT-Conv block within the generator for LDCT denoising, which demonstrates the versatility of the block.
- Evaluate our network with extensive experiments, including an ablation study and visual scoring using two distinct datasets of brain and abdominal CT images. Six metrics based on pixel- and feature-spaces were used, and the results indicated superior performances in both quantitative and qualitative measures compared to those of state-of-the-art denoising techniques.

This paper is an extension of our conference paper [14]. Relative to the conference paper, we have implemented the following major extensions: more comprehensive explanations of the methodology and additional experimental setups using different dataset configurations.

II. Related Research

A. Low-Dose CT Denoising Using Deep Learning

Convolution neural networks (CNNs), transformer networks (TRs), generative adversarial networks (GANs), and diffusion networks (DNs) have been the major approaches for denoiser.

CNNs have a powerful ability for high-frequency component extraction and mapping. Chen *et al.* proposed a residual encoder-decoder CNN (RED-CNN) consisting of convolution, deconvolution, and skip connections to suppress the noise of LDCT images [6]. To enhance sharpness, Liang *et al.* introduced an edge enhancement-based densely connected convolutional neural network (ED-CNN) comprising the trainable Sobel convolution to integrate edge information [9]. However, most of the CNN-based algorithms use pixel-based mean squared error (MSE) as the main objective function, which overlooks subtle image textures that are critical for human perception [15, 16].

TRs are based on attention mechanisms that form relationships between a local response and all other pixels, thus ensuring long-range dependencies in vision tasks. Zamir *et al.* proposed Restormer with design changes in multi-head attention and feed-forward network components to learn multi-scale local-global representations for high-resolution images [17]. Wang *et al.* designed a convolution-free token-to-token dilated vision transformer (CTformer) that leverages the spatial relations between adjacent tokens to circumvent the proclivity of conventional transformers [12]. However, the computational complexity of TR increases quadratically with the spatial resolution, and the use of overlapped inference, which often causes boundary artifacts, is inevitable due to patch tokenization.

GANs can preserve more texture information and generate realistic subtle patterns by dynamically comparing the similarity between the synthesized and real images in a normal-dose CT (NDCT) domain. Yang *et al.* employed the Wasserstein GAN with gradient penalty and perceptual loss for retaining important detailed image information [7]. Shan *et al.* used a modularized adaptive processing neural network (MAP-NN) that conducts end-to-process mapping for LDCT denoising [8]. Huang *et al.* designed two different U-Net based discriminators [18] to distinguish local and global differences between synthesized and real NDCT images in both the image and gradient domains [10]. However, due to the instability of GAN training, synthesized images with inconsistent and distorted anatomical signals are often produced [19].

DNs have been proposed recently to solve various image-to-image translations with the advent of the denoising diffusion probabilistic model (DDPM), which has incredible generative power [20]. Shi *et al.* adopted the DDPM as a means of data augmentation to ensure better circumstances for LDCT

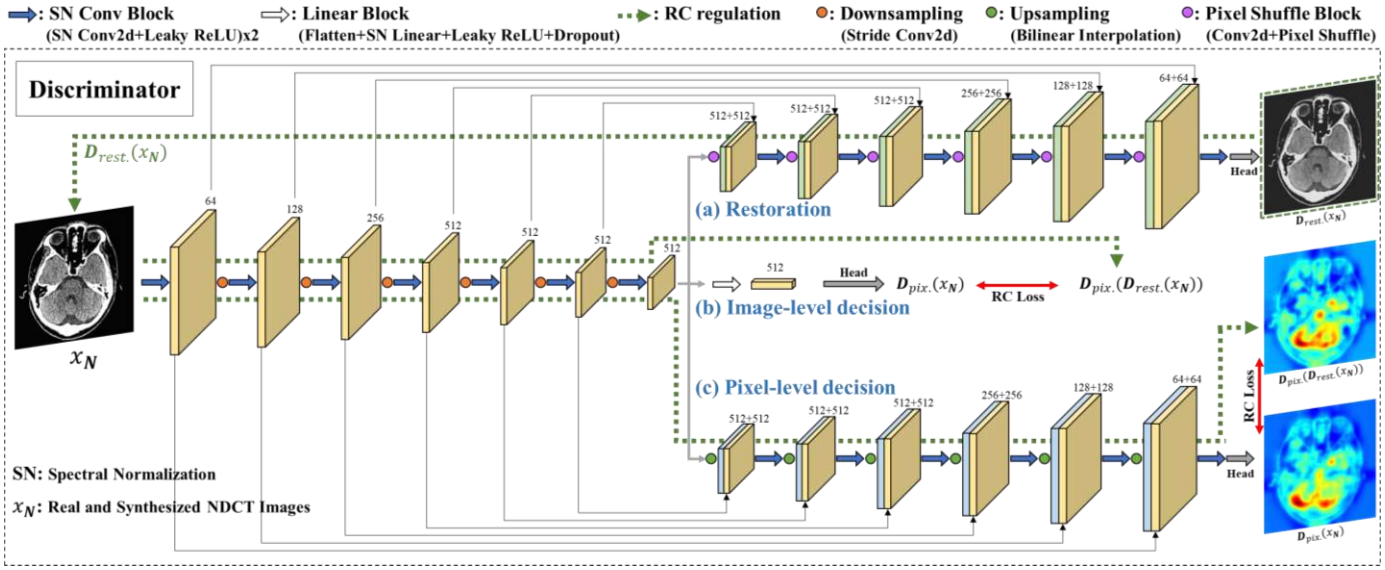

Fig. 1. Schematic overview of MTD-GAN. The discriminator aims at distinguishing the normal-dose and synthesized images and performs three multi-tasks: (a) restoration, (b) image-level decision, and (c) pixel-level decision. In the restoration consistency (RC) task, the restoration output is reused as the discriminator input, and the two outputs are compared with each other. Note: The value displayed above each feature map indicates the corresponding channel quantity.

denoising, which suggests that utilizing synthetic data may be a feasible approach to improving LDCT research [21]. Xia *et al.* also adopted the DDPM with conditional diffusion processing for solving LDCT denoising and reduced the high sampling cost with the DPM solver [11]. However, there is still an evident trade-off between sampling speed and performance, and it is sensitive to the specific solver and the number of function evaluations (NFE).

B. Evolution of Discriminator with Multi-Task Learning

Historically, building a powerful discriminator has been an important key to achieving high-quality results in the GAN framework since the generator is trained by the feedback from the discriminator. Representative methods include gradient penalties [22-24] and structural modifications to discriminators [18, 25, 26]. Recently, the training of GANs has been strengthened using MTL by utilizing comprehensive information simultaneously learned from multiple different tasks for the generalization of the performance of all tasks [27]. In MTL-based generators, Hang *et al.* designed a generator to undertake reconstruction and classification tasks [28]. Rad *et al.* presented a generator to incorporate semantic information by learning image super-resolution and semantic segmentation to super-resolve an input image [29]. In MTL-based discriminators, Wan *et al.* proposed a discriminator designed not only to determine whether an image is real or synthesized class but also to extract discriminative facial features [30]. Cha *et al.* employed a discriminator capable of separately classifying both character and font classes [31]. In the case of multiple discriminators, Liu *et al.* proposed two discriminators that are in charge of the content and style of the image [32]. Mordido *et al.* proposed an ensemble of discriminators that dynamically changes using adversarial dropout [33]. Unlike previous studies, we adopted an MTL-based discriminator to simultaneously conduct three different types of vision tasks—image-level decision, pixel-level decision, and restoration—to solve LDCT denoising.

C. Fourier Transform Applications

Lately, the introduction of Fourier transforms has demonstrated its potential to enhance the performance of DL-based methods in various vision tasks necessitating large receptive fields, as altering a single value in the spectral domain has a global impact on the image [34]. For image classification and detection, Chi *et al.* replaced the convolution with the spatial-Fourier interaction, which simultaneously uses spatial and spectral information for achieving cross-scale fusion [35]. For image segmentation, Yang *et al.* swapped the low-frequency spectrum between the source and target images to mitigate the influence caused by the domain-specific style [36]. For image inpainting, Suvorov *et al.* applied the structure of fast

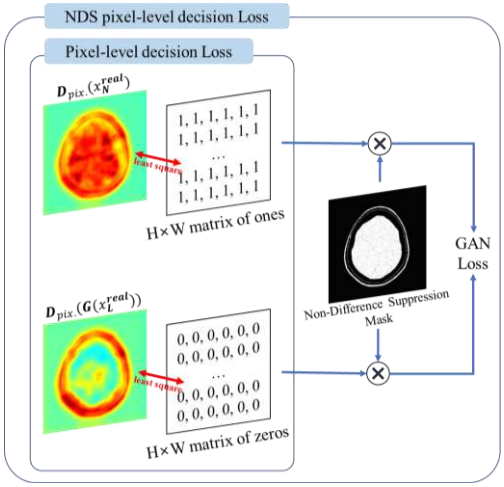

Fig. 2. Diagram illustrating the process of our non-difference suppression (NDS) regulation applied to the LSGAN loss in the pixel-level decision task of MTD-GAN. The NDS mask was created by subtracting the LDCT from NDCT images and then binarizing the result.

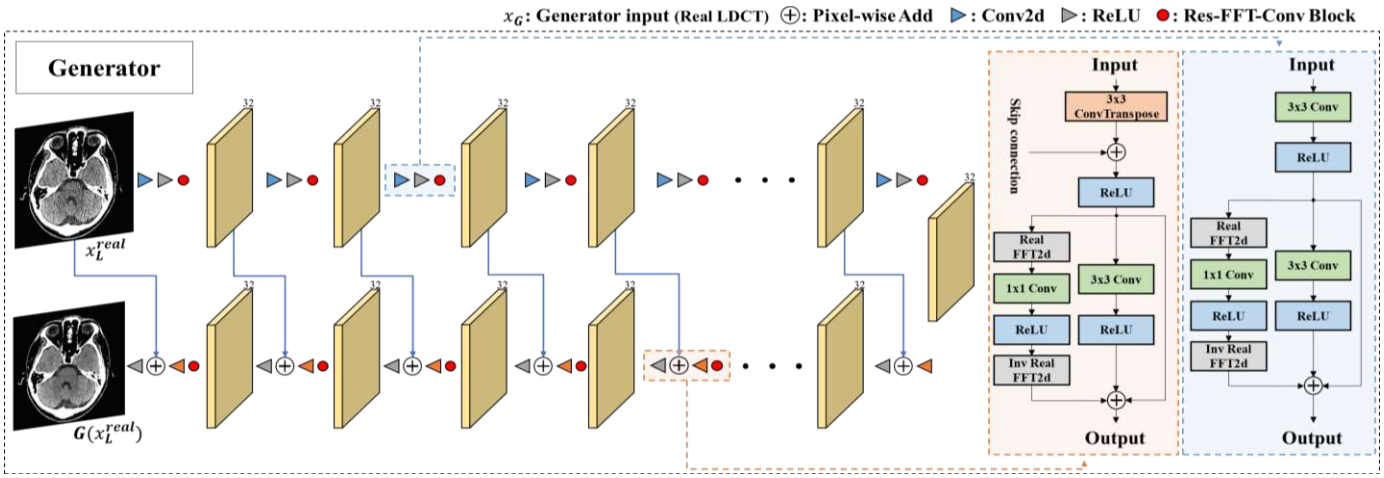

Fig. 3. Schematic overview of the MTD-GAN's generator—Res-FFT-Conv Generator. The RED-CNN was used as a base generator by adding a Res-FFT-Conv block at every layer. The Res-FFT-Conv block consists of three flows: identity, spectral, and spatial. Each path captures complementary information with different receptive fields. Note: The number presented atop each feature map signifies the associated channels.

Fourier convolution for acquiring an image-wide receptive field [37]. For image deblurring, Mao *et al.* introduced a residual fast Fourier transform with a convolution (Res-FFT-Conv) block for achieving mixed receptive fields [13]. Inspired by these applications of Fourier transforms, we attempted to apply compact Res-FFT-Conv blocks in our generator for LDCT denoising.

III. METHODS

The details of MTD-GAN are illustrated in **Fig. 1**. Our network consists of a single discriminator and a single generator, both of which are optimized through an adversarial strategy. First, our discriminator conducts three types of vision multi-tasks: (a) restoration, (b) image-level decision, and (c) pixel-level decision to transfer semantic and consistent feedback. Next, we introduced two regulations to enhance the discriminator's representation capabilities: one to eliminate the discriminator's confusion and the other to maintain consistency between the discriminator's outputs from the inputs and restored inputs. Lastly, we incorporated a compact Res-FFT-Conv block into our generator, effectively utilizing both frequency and spatial representations to provide mixed receptive fields.

A. Multi-Task Discriminator

As a classifier, a typical discriminator solely learns a representation to effectively penalize the generator by identifying the most discriminative disparities between authentic and synthetic images. This causes the discriminator to irregularly focus on only one global structure or one local detail, providing inconsistent feedback to the generator, which has negative effects on the denoising performance [18]. To solve this problem, we introduced an MTL-based discriminator that performed restoration, image-level decision, and pixel-level decision tasks. This approach prevents changeable bias by providing textural, global, and local coherent feedback to the generator. We utilized a hard-parameter sharing architecture [27], comprising a shared encoder and three distinct task-specific layers for the three vision tasks (see **Fig. 1**). Additionally, we added spectral normalization (SN) to all layers of the discriminator for stable training [24]. These architectural modifications result in a more robust discriminator that maintains a potent data representation, making it more challenging for the generator to fool the discriminator and ultimately improving the quality of the synthesized samples.

The restoration task is defined as an image restoration process for NDCT domain images, implemented through an encoding-decoding framework. This process is performed in the first target-specific layer, consisting of PixelShuffle upsample blocks [38], which are designed to super-resolve low-resolution images into high-resolution images (refer to **Fig. 1 (a)**). This task is conducted in an unsupervised manner, aiming to capture the contextual information of both synthesized and real NDCT images. This approach encourages the shared encoder to enhance its understanding of the semantic representation. The restoration loss was defined as the mean absolute error (MAE) loss.

$$L_{restore} = \mathbb{E}_{x_N}[|\boldsymbol{D}_{rest.}(x_N) - x_N|], \quad (1)$$

where $\boldsymbol{D}_{rest.}$ is the restoration output of the discriminator $\boldsymbol{D}$. x_N is the NDCT domain images including the real and synthesized NDCT images.

The image-level decision task is designed to determine whether images in the NDCT domain belong to a synthesized or real class, akin to conventional discriminators. This process is accomplished in the second task-specific layer, which includes a linear classifier (see **Fig. 1 (b)**). This task encourages the discriminator to concentrate on the most discernible differences between real and synthesized images in the NDCT domain, which then serve as global feedback to the generator. The classifier comprises flatten, linear with SN, leaky ReLU, and dropout layers. The image-level decision loss is defined at a logit level, using the least squares GAN (LSGAN) approach [39].

$$L_{image} = \mathbb{E}_{x_N^{real}}\left[\boldsymbol{D}_{pix.}(x_N^{real}) - 1\right]^2 + \mathbb{E}_{\boldsymbol{G}(x_L^{real})}[\boldsymbol{D}_{img.}(\boldsymbol{G}(x_L^{real}))]^2, \quad (2)$$

TABLE I
COMPARATIVE QUANTITATIVE ANALYSIS OF STATE-OF-THE-ART TECHNIQUES*

Type	Networks	FID↓	PL↓	TML↓	RMSE↓	PSNR↑	SSIM↑
				Brain			
CNNs	RED-CNN	37.4992	0.1505±0.0601	15.7372±6.9484	0.0332±0.0123	33.2597±7.4373	0.8998±0.0636
	ED-CNN	35.2580	0.1467±0.0588	14.4593±6.6161	0.0344±0.0127	32.9294±7.1878	0.8993±0.0630
TRs	Restormer	33.1093	0.1468±0.0583	15.3261±6.6683	**0.0322±0.0120**	**33.5436±7.5099**	0.9029±0.0631
	CTformer	33.4311	0.1437±0.0555	13.7465±5.7319	0.0337±0.0125	32.4808±6.7707	0.9006±0.0626
DNs	DDPM	21.3378	0.1479±0.0577	12.0669±5.7243	0.0444±0.0165	29.6737±6.7992	0.8558±0.0871
	DDIM	26.6446	0.1550±0.0548	12.6426±5.7009	0.0465±0.0162	27.8761±4.8821	0.8526±0.0736
	PNDM	25.9946	0.1541±0.0574	12.5484±5.8039	0.0459±0.0167	28.5441±5.8474	0.8570±0.0767
	DPM	23.6183	0.1632±0.0605	13.2447±5.8619	0.0437±0.0159	29.0033±5.8585	0.8686±0.0759
GANs	WGAN-VGG	24.6882	0.1499±0.0601	13.0983±5.9125	0.0447±0.0170	30.7714±7.0492	0.8646±0.0853
	MAP-NN	19.5778	0.1373±0.0544	11.4723±5.2817	0.0383±0.0144	32.0321±6.9669	0.8995±0.0623
	DU-GAN	18.8156	0.1285±0.0512	10.5245±5.0481	0.0364±0.0137	32.5533±7.6995	0.9035±0.0605
	MTD-GAN (Ours)	**17.3522**	**0.1235±0.0495**	**9.9207±4.7422**	0.0358±0.0136	32.7189±7.6286	**0.9054±0.0601**
GT (NDCT)		9.8355	0.0	0.0	0.0	100.0	1.0
Input (LDCT)		39.7583	0.1821	20.4448	0.0575	28.9550	0.8743
				Abdomen			
CNNs	RED-CNN	40.9710	0.1511±0.0232	14.8098±1.8305	0.0242±0.0048	32.4632±1.7106	0.9172±0.0265
	ED-CNN	35.0528	0.1484±0.0239	14.6234±1.7415	0.0244±0.0053	32.4189±1.6938	0.9204±0.0264
TRs	Restormer	39.6254	0.1414±0.0254	13.6422±2.0163	**0.0222±0.0049**	**33.2559±1.7395**	0.9242±0.0270
	CTformer	39.6326	0.1404±0.0167	13.0195±0.8481	0.0229±0.0057	33.0407±1.8822	0.9230±0.0254
DNs	DDPM	19.7057	0.1339±0.0240	10.6218±1.5560	0.0286±0.0067	31.0640±1.8146	0.8964±0.0356
	DDIM	24.1247	0.1401±0.0241	11.9858±1.5679	0.0259±0.0063	31.9465±1.8983	0.8963±0.0332
	PNDM	23.0774	0.1361±0.0240	11.6411±1.6190	0.0256±0.0062	32.0609±1.8891	0.8966±0.0301
	DPM	19.0566	0.1341±0.0230	10.9272±1.3019	0.0285±0.0070	31.1215±1.9019	0.8952±0.0383
GANs	WGAN-VGG	27.3975	0.1387±0.0276	12.0744±2.1688	0.0287±0.0068	31.0384±1.8187	0.9086±0.0312
	MAP-NN	23.7059	0.1327±0.0255	11.3263±1.9722	0.0264±0.0064	31.7779±1.8579	0.9169±0.0292
	DU-GAN	18.0828	0.1172±0.0219	9.3012±1.2724	0.0273±0.0090	31.6445±2.3974	0.9215±0.0282
	MTD-GAN (Ours)	**14.8930**	**0.1127±0.0215**	**8.8489±1.2725**	0.0239±0.0061	32.6641±1.9480	**0.9247±0.0275**
GT (NDCT)		0.8510	0.0	0.0	0.0	100.0	1.0
Input (LDCT)		42.4862	0.1663	16.5566	0.0356	29.2489	0.8883

Note: p-values were evaluated between MTD-GAN and other methods using a paired t-test. *MTD-GAN demonstrated significantly better performance (P < 0.0001) compared to all previous works in all t-test-applicable metrics, except with CTformer in terms of PSNR (p-value: 0.0255). The best performances are marked in bold.

where $\boldsymbol{D}_{img.}$ is the image-level decision output of the discriminator. x_N^{real} is the real NDCT image, and x_L^{real} is the real LDCT image. $\boldsymbol{G}(x_L^{real})$ is the NDCT image synthesized by the generator $\boldsymbol{G}$.

The objective of the pixel-level decision task is to discern whether each pixel in an image in the NDCT domain belongs to a synthesized or real class. This process utilizes an encoding-decoding framework and is executed in a third task-specific layer comprising upsampling blocks with bilinear interpolation (see **Fig. 1 (c)**). The discriminator can recognize variations in local details between real and synthesized images within the NDCT domain. A confidence map generated by the discriminator serves as local feedback to the generator. The pixel-level decision loss is also defined using the LSGAN approach at a pixel level.

$$L_{pixel} = \mathbb{E}_{x_N^{real}}\left[\boldsymbol{D}_{pix.}(x_N^{real}) - 1_{\boldsymbol{H}\times\boldsymbol{W}}\right]^2 + \mathbb{E}_{\boldsymbol{G}(x_L^{real})}[\boldsymbol{D}_{pix.}(\boldsymbol{G}(x_L^{real}))]^2, \quad (3)$$

where $\boldsymbol{D}_{pix.}$ is the pixel-level decision output of the discriminator. $1_{\boldsymbol{H}\times\boldsymbol{W}}$ is a matrix of ones with height and width.

B. Regulatory Mechanisms

While training our multi-task discriminator, negative phenomena, such as eccentric activations and discriminator inconsistency, were observed (see Section **IV-F**). To address these issues, we proposed two regulations.

First, we introduced the restoration consistency (RC) loss. Conventional consistency regularization [40] enforces the discriminator to remain unchanged by arbitrary semantic-preserving perturbations. However, our consistency regularization aims to improve the restoration ability and a broader understanding of the class by ensuring that the discriminator's representation remains unchanged by the generated restoration by the discriminator. Specifically, we reuse the restoration output of our MTL discriminator as an input to the discriminator, and then the corresponding image-level decision and pixel-level decision results are produced, respectively. The outputs of the input and corresponding restoration are then compared to make them closer to each other (see **Fig. 1 green path**). The RC loss is defined as follows:

$$L_{RC} = \mathbb{E}_{x_N}[\left\|\boldsymbol{D}_{img.}(x_N) - \boldsymbol{D}_{img.}(\boldsymbol{D}_{rest.}(x_N))\right\| + \left\|\boldsymbol{D}_{pix.}(x_N) - \boldsymbol{D}_{pix.}(\boldsymbol{D}_{rest.}(x_N))\right\|] \quad (4)$$

Second, we proposed non-difference suppression (NDS). Due to the nature of the medical image, the same regions in both LDCT and NDCT images, such as bone and background, occupy a large portion of the CT image, which causes confusion in the discriminator's decision-making. According to **Eq. (3)** for the pixel-level decision task, the confidence map is provided as a matrix of the same resolution, filled with a single value depending on whether the input is real or synthesized. To exclude the confusing region, we applied an NDS mask created

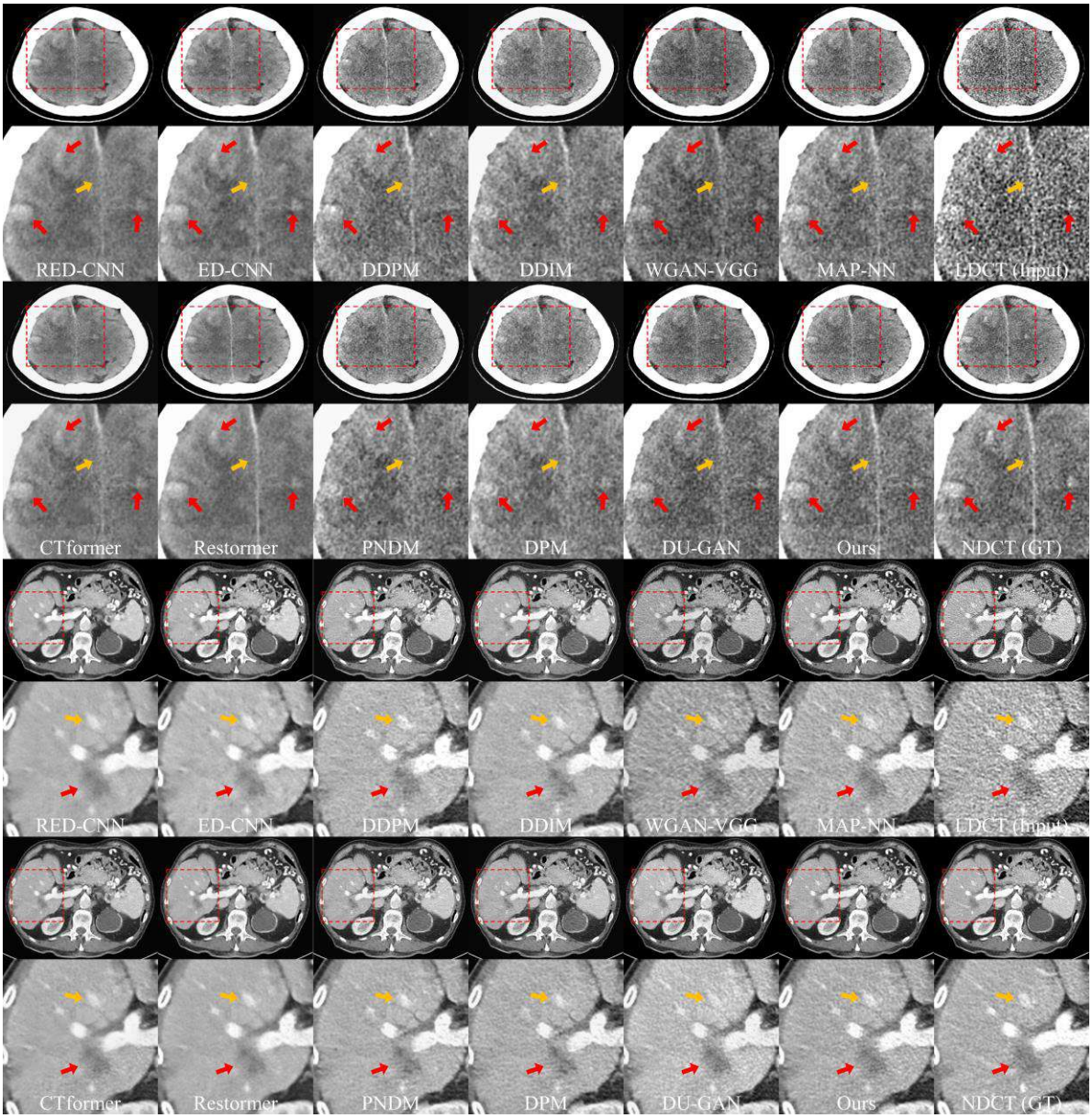

Fig. 4. Comparison of the denoising results of previous methods on the brain and abdomen test sets. The display window ranges from [0, 80] HU for the brain and [-160, 240] HU for the abdomen. The region enclosed by the red square was selected as the focal area for its perceived clinical significance. In the brain images, the intracerebral hemorrhage (ICH) and falx cerebri are denoted by red and yellow arrows, respectively. In the abdomen images, cancerous lesions and blood vessels are denoted by red and yellow arrows, respectively.

by subtracting the NDCT images from the corresponding LDCT images and then binarizing the result for the pixel-level decision objective function (see **Fig. 2**). This objective function is defined as follows:

$$L_{pixel}^{NDS} = L_{pixel} \times f(|x_L^{real} - x_N^{real}|),\ f(x) = \begin{cases} 0, if\ x = 0 \\ 1, if\ x \neq 0 \end{cases} \quad (5)$$

In the discriminator, the total loss function for the multi-tasks is defined by:

$$L_{\boldsymbol{D}} = \lambda_1 L_{image} + \lambda_2 L_{pixel}^{NDS} + \lambda_3 L_{restore} + \lambda_4 L_{RC}, \quad (6)$$

where $\lambda_{1\text{-}4}$ control the relative weighting of the different loss factors.

C. Generator Meets Res-FFT-Conv Block

A convolutional operator excels at capturing high-frequency details but might struggle to accurately represent low-frequency information [41]. Consequently, a CNN-based generator could be vulnerable when trying to synthesize both high and low-frequency features of NDCT. To address this problem, we proposed a generator consisting of Res-FFT-Conv blocks (the Res-FFT-Conv Generator) that take advantage of both the spatial and Fourier domains. Inspired by Huang *et al.* [10], we selected the lightweight version of RED-CNN [6], which features 10 stacked (de)convolutional layers in both the encoder and decoder, as the base denoiser. We then added a compact Res-FFT-Conv block [13] with a reduced number of 1x1 convolution layers to each layer in our generator. This allows

TABLE II
ABLATION STUDY OF THE MTD-GAN COMPONENTS

Type	Ablation Methods	FID↓	PL↓	TML↓	RMSE↓	PSNR↑	SSIM↑
Brain							
Single	(a): $D_{img.}$	19.8473	0.1347±0.0526	11.5021±5.3312	0.0375±0.0143	32.3521±7.8508	0.9025±0.0612
	(b): $D_{pix.}$	19.7945	0.1346±0.0529	11.5321±5.3486	0.0373±0.0141	32.3547±7.6271	0.9026±0.0609
Dual	(c): $D_{img.} + D_{pix.}$	19.1580	0.1340±0.0528	11.3019±5.1623	0.0372±0.0141	32.3948±7.7078	0.9026±0.0611
	(d): $D_{img.} + D_{rest.}$	18.8776	0.1338±0.0527	11.2503±5.1525	0.0370±0.0140	32.4054±7.5246	0.9027±0.0609
	(e): $D_{pix.} + D_{rest.}$	18.9772	0.1339±0.0528	11.3029±5.1514	0.0370±0.0139	32.4188±7.6954	0.9028±0.0608
Triple	(f): $D_{img.} + D_{pix.} + D_{rest.}$	18.5373	0.1332±0.0525	11.1399±5.1430	0.0371±0.0141	32.4062±7.7230	0.9027±0.0608
	(g): (f) + NDS reg.	18.4707	0.1331±0.0527	11.0717±5.0789	0.0370±0.0140	32.4279±7.6522	0.9028±0.0608
	(h): (f) + RC reg.	18.4853	0.1329±0.0525	11.0792±5.1206	0.0370±0.0140	32.4460±7.7966	0.9028±0.0609
	(i): (f) + NDS + RC reg.	18.2474	0.1327±0.0525	11.0335±5.0960	0.0369±0.0140	32.4613±7.7928	0.9031±0.0608
	(j): (i) + Res-FFT-Conv Generator	17.7768	0.1297±0.0512	10.5051±4.8779	0.0368±0.0140	32.4826±7.6717	0.9036±0.0605
	(k): (j) + PCGrad (Ours)	**17.3800**	**0.1277±0.0507**	**10.3253±4.8421**	**0.0366±0.0139**	**32.6347±8.0107**	**0.9043±0.0604**
	GT (NDCT)	9.8371	0.0	0.0	0.0	100.0	1.0
	Input (LDCT)	39.7624	0.1821	20.4434	0.0575	28.9550	0.8743
Abdomen							
Type	Ablation Methods	FID↓	PL↓	TML↓	RMSE↓	PSNR↑	SSIM↑
Single	(a): $D_{img.}$	24.2318	0.1231±0.0234	10.1723±1.4906	0.0274±0.0078	31.5471±2.1425	0.9216±0.0276
	(b): D_{pix}	23.9954	0.1227±0.0215	10.1604±1.1911	0.0272±0.0072	31.5652±2.0061	0.9218±0.0282
Dual	(c): $D_{img.} + D_{pix.}$	22.9336	0.1222±0.0227	10.0811±1.3084	0.0269±0.0070	31.6474±1.9691	0.9219±0.0277
	(d): $D_{img.} + D_{rest.}$	22.4710	0.1223±0.0220	10.0591±1.2697	0.0270±0.0062	31.5693±1.7908	0.9222±0.0270
	(e): $D_{pix.} + D_{rest.}$	22.3942	0.1213±0.0220	10.0575±1.2147	0.0268±0.0068	31.6575±1.9236	0.9220±0.0277
Triple	(f): $D_{img.} + D_{pix.} + D_{rest.}$	21.3416	0.1201±0.0237	9.7139±1.5484	0.0269±0.0064	31.6109±1.8351	0.9221±0.0276
	(g): (f) + NDS reg.	21.1750	0.1195±0.0234	9.6081±1.4841	0.0268±0.0067	31.6548±1.9064	0.9221±0.0277
	(h): (f) + RC reg.	20.1929	0.1193±0.0228	9.5559±1.3723	0.0265±0.0068	31.7825±1.9491	0.9222±0.0282
	(i): (f) + NDS + RC reg.	19.5482	0.1190±0.0226	9.5522±1.3616	0.0260±0.0063	31.9256±1.8782	0.9224±0.0274
	(j): (i) + Res-FFT-Conv Generator	17.8494	0.1179±0.0209	9.5047±1.1569	0.0241±0.0061	32.5787±1.9300	0.9232±0.0275
	(k): (j) + PCGrad (Ours)	**17.0614**	**0.1168±0.0215**	**9.3077±1.2832**	**0.0241±0.0060**	**32.5841±1.9164**	**0.9235±0.0275**
	GT (NDCT)	0.851	0.0	0.0	0.0	100.0	1.0
	Input (LDCT)	42.4862	0.1663	16.5566	0.0356	29.2489	0.8883

Note: The best performances are marked in bold. For a detailed comparison of statistical significance using the paired t-test for all combinations of MTD-GAN elements, refer to Appendix I-D.

the generator to benefit from modeling both high- and low-frequency features while simultaneously capturing long- and short-range interactions (see **Fig. 3**). The Res-FFT-Conv block consists of three flows: an identity flow that induces the generator to focus more on residual information, a spectral (or global) flow that captures the information from low- to high-frequency in the Fourier domain, and a spatial (or local) flow that obtains geometric information through convolutional operations. Each path can capture complementary information with different receptive fields. The final result is obtained by summing the three flows to leverage frequency-spatial dual-domain representations that are more concentrated through the residual connection. In the generator's objective function, the conventional adversarial LSGAN loss was selected, to which the NDS regulation is also applied.

$$L_{adv}^{NDS} = \mathbb{E}_{x_L^{real}}\left[\boldsymbol{D}_{pix.}\left(\boldsymbol{G}(x_L^{real})\right) - 1\right]^2 \times f(|x_L^{real} - x_N^{real}|) + \mathbb{E}_{x_L^{real}}\left[\boldsymbol{D}_{img.}\left(\boldsymbol{G}(x_L^{real})\right) - 1\right]^2, \quad (7)$$

where $f(x)$ is the same as **Eq. (5)**. Inspired by [42], we used the Charbonnier loss, which is more robust to outliers than MAE loss.

$$L_{pixel} = \mathbb{E}_{x_L^{real}}\left[\sqrt{\|x_N^{real} - \boldsymbol{G}(x_L^{real})\|^2 + \varepsilon^2}\right], \quad (8)$$

where ε is empirically set to 10^{-6}. To further enhance the fidelity of high-frequency details, we employed additional edge loss using the Laplacian.

$$L_{edge} = \mathbb{E}_{x_L^{real}}\left[\sqrt{\|\Delta(x_N^{real}) - \Delta(\boldsymbol{G}(x_L^{real}))\|^2 + \varepsilon^2}\right], \quad (9)$$

where Δ denotes the Laplacian operator. The total generator loss is defined as follows:

$$L_{\boldsymbol{G}} = \lambda_5 L_{adv}^{NDS} + \lambda_6 L_{pixel} + \lambda_7 L_{edge}, \quad (10)$$

where $\lambda_{5\text{-}7}$ determine the relative weighting of the individual loss elements.

IV. EXPERIMENT

Our model was trained and evaluated on two CT denoising datasets: brain and abdomen. First, we demonstrated both the quantitative and qualitative superiority of the MTD-GAN's denoising performance over the state-of-the-art methods (see **Section IV-C**). Next, to investigate the effects of the individual elements of the MTD-GAN (see **Section IV-D**), we conducted ablation studies. With the help of experienced radiologists, we conducted a visual scoring test to evaluate the synthesized images (see **Section IV-E**). Furthermore, we assessed the effects of the two regulations by comparing the confidence map of each type (see **Section IV-F**). To verify the effects of the Res-FFT-Conv Generator, we compared the differences relative to the NDCT images depending on the presence or absence of the Res-FFT-Conv block through a line profile in the Fourier domain (see **Section IV-G**). Next, we applied various MTL optimization algorithms and conducted comparative analyses to consider task-specific weights in our framework (see **Appendix**

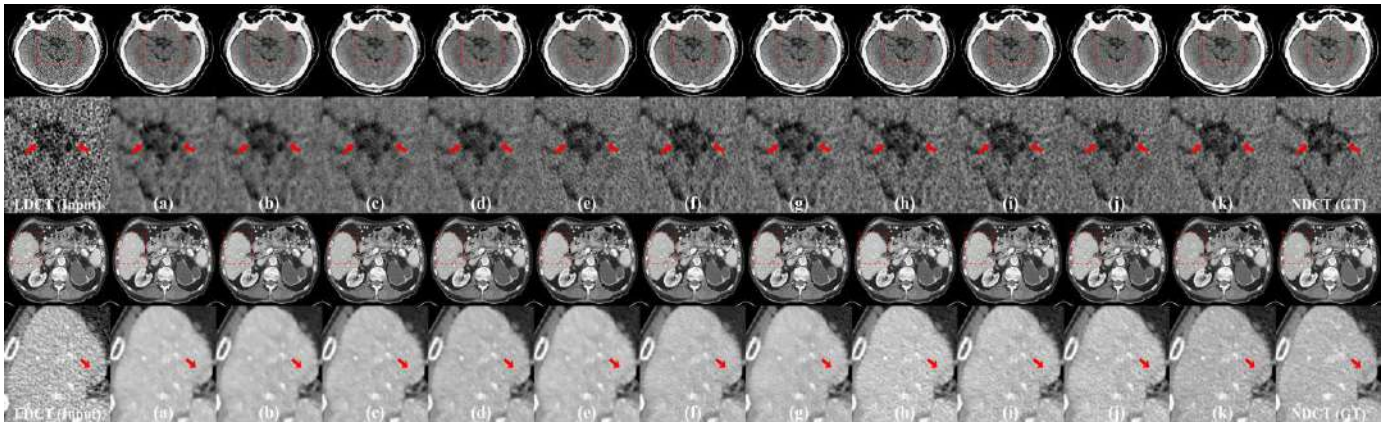

Fig. 5. Comparison of the denoising results of ablation studies on the brain and abdomen test sets. The ROI of the red box was enlarged for better view or the areas with clinical significance. In the zoomed images, the optic tract in the brain samples and the cancerous lesion (metastasis) in the abdomen samples are marked with red arrows. The display window ranges from [0, 80] HU for the brain and [-160, 240] HU for the abdomen.

I-A). The efficiency of the various methods was measured and compared in both training and inference settings (see **Appendix I-B**). In addition, we performed a comparative analysis of the different methods on supplemental clinical cases containing lesions (see **Appendix I-C**). A paired t-test was conducted for all combinations of MTD-GAN elements to assess statistical significance (see **Appendix I-D**). The factors influencing the multi-task discriminator were analyzed (see **Appendix I-E**). The details of evaluation criteria for a blind reader study were delineated (see **Appendix I-F**).

A. Datasets

In the brain dataset, the Asan Medical Center in Korea retrospectively acquired brain CT denoising data from patients who underwent consecutive CT scans from July to August 2020. This dataset, comprising scans from 130 patients, was randomly divided into training, validation, and test sets with 100, 15, and 15 patients, respectively. The scans were acquired using a Siemens Definition AS+ scanner at 120 kVp and 300 quality reference mAs (QRM).

For the abdominal dataset, the paired LDCT data was sourced from the 2016 NIH-AAPM-Mayo Clinic LDCT Grand Challenge [5]. It included abdominal NDCT and LDCT images of 10 anonymized patients. Data from Patient L506 were earmarked for evaluation, while the remaining nine patients' data were utilized for model training. These scans were acquired in the portal venous phase using a Siemens SOMATOM Flash scanner at 120 kVp and 200 QRM.

In both cases, following the clinical protocol of the host institution, data acquisition involved automated exposure control (CAREDose 4D, Siemens Healthcare). Images were reconstructed using a 512 x 512 matrix, with the field of view tailored to patient size and a uniform slice thickness of 3mm. All CT scans employed the filtered back projection (FBP) algorithm with a B40F kernel for the brain and a B30F kernel for the abdomen, where "F" denotes a fast rotation time and "B" indicates "body." The numerical value in each kernel designates image sharpness, with higher numbers producing sharper images. The brain LDCT dataset of the same patients was generated using a low-dose simulation tool (ReconCT, Siemens Healthcare) [43], employing a method similar to that used for the abdomen LDCT dataset. This tool inserted extra noise, properly scaled to a quarter-dose using Poisson-distributed random numbers. The simulated LDCT datasets feature 120 kVp and 75 QRM for the brain dataset, and 120 kVp and 50 QRM for the abdomen dataset. Given the distinct features of brain and abdominal CT images, we trained two separate models for denoising these datasets and assessed them independently.

B. Implementation Details

Our model was compared with the previous state-of-the-art algorithms including CNN-based (RED-CNN [6] and ED-CNN [9]), TR-based (Restormer [17] and CTformer [12]), DN-based (DDPM [44], DDIM [45], PNDM [46], DPM [47]), and GAN-based (WGAN-VGG [7], MAP-NN [8], and DU-GAN [10]) methods. The models were trained using an official code, but the training settings, including the optimizer, patch size, number of epochs, and learning rate scheduler, were equally assigned. All the DN models used 50 NFE except DDPM, which used 1000 NFE. For a fair comparison within limited resources, the batch size for previous methods was set to the maximum for a single graphics processing unit (GPU) memory.

In preprocessing, we applied a brain window clipping of [0, 80] HU and an abdomen window clipping of [-160, 240] HU, and then scaled the values to [0, 1] for the network input. We cropped the foreground of the CT image and then randomly extracted 8 patches of size 64×64 in each epoch.

All experiments were implemented using Pytorch 1.13.1 and CUDA 11.6, accelerated by an NVIDIA TITAN RTX 24 GB GPU. The network was initialized with a uniform Xavier distribution, and an AdamW optimizer was employed with a learning rate of 1e-4, a warmup of 10 epochs, weight decay of 5e-4, and betas set to (0.9, 0.99). The learning rate was reduced during training using a polynomial learning rate schedule. The total number of epochs was up to 500.

The weights $\lambda_{5\text{-}7}$ in the generator objective function were manually adjusted to align with the methodology of Shan *et al.* [8] ($\lambda_5 = 1$ and $\lambda_{6,7} = 50$). For the multi-task weights $\lambda_{1\text{-}4}$, we selected the Projecting Conflicting Gradients (PCGrad) algorithm [48], based on the ablation study results (refer to **Appendix I-A**).

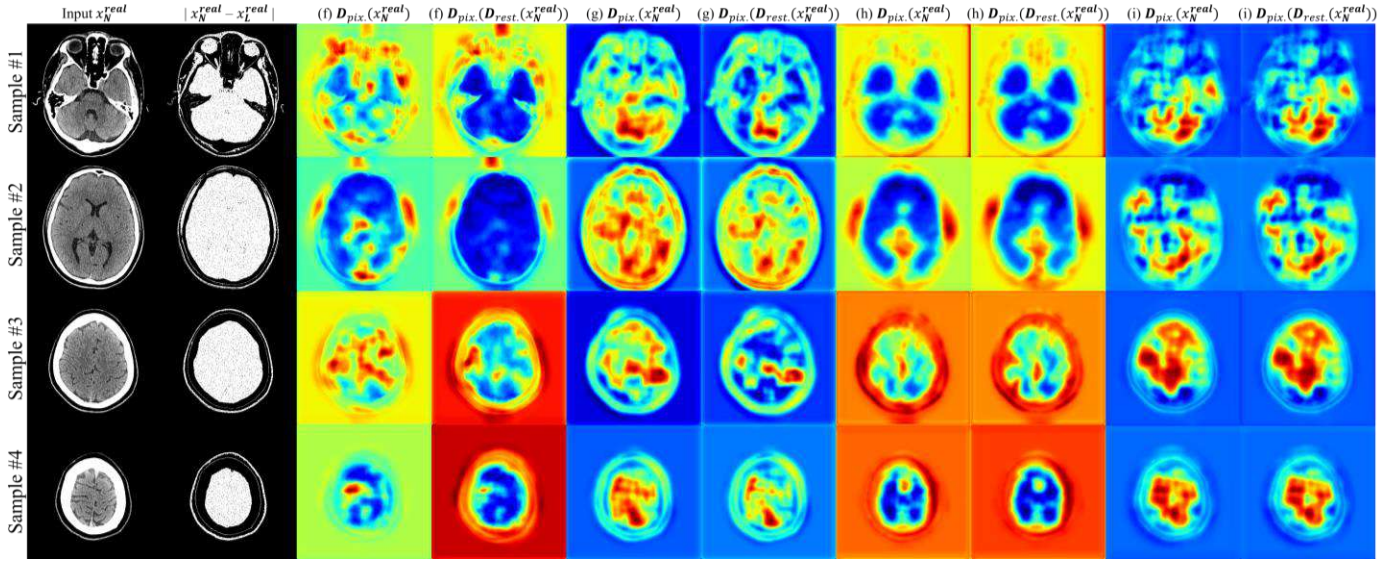

Fig. 6. A comparison of the consistency and concentration of the discriminator through the pixel-level decision confidence map according to the regulations. Four slices were selected according to the depth of the randomly sampled patient. (f)-(i) are the same labels as shown in Table II.

To ensure the suitability of our approach for clinical applications, we employed a variety of metrics to prevent biased results and evaluate the robustness of the networks. Specifically, we utilized a total of six metrics to quantitatively assess the synthesized images in all experiments. For pixel-based metrics, we used the peak signal-to-noise ratio (PSNR), structural similarity (SSIM), and root-mean-square error (RMSE). For feature-based metrics, we used perceptual loss (PL) [7], texture matching loss (TML) [49], and Fréchet inception distance (FID) [50]. To determine statistical significance, we performed paired t-tests on all metrics except FID.

C. Denoising Comparison

In **Table I**, both the CNN- and TR-based models demonstrate strong performance in pixel-based metrics (PSNR, RMSE); however, they exhibit over-smoothing, which leads to image blur and loss of structural details, as evident in **Fig. 4**. Specifically, ED-CNN sacrifices performance in pixel-based metrics compared to RED-CNN but shows improved performance in feature-based metrics. CTformer performs well in pixel-based metrics, but it requires significant overlapped inference to resolve the inevitable boundary artifacts. In contrast, the DN- and GAN-based models excel in feature-based metrics, albeit with some limitations in pixel-based metrics. The DDPM model occasionally distorts signals by erroneously creating blood vessels, and the WGAN-VGG's noise suppression performance, which includes undesirable artifacts, is still relatively poor. DU-GAN offers the most competitive performance for MTD-GAN. Our proposed MTD-GAN significantly outperforms other models in feature-based metrics for brain and abdomen test sets, and it also demonstrates competitive results in pixel-based metrics (SSIM). High TML and SSIM scores of MTD-GAN indicate its superior structural similarity in texture and content across both pixel-space and feature-space. As illustrated in **Fig. 4**, our MTD-GAN has the closest qualitative performance to NDCT in reducing artifacts and preserving significant anatomical structures, resulting in radiologist-friendly images.

D. Ablation Study of Proposed Methods

We conducted extensive ablation studies to investigate the effects of individual elements of the MTD-GAN on brain and abdomen denoising datasets, as detailed in Table II and Fig. 5. To ensure fair comparison, training settings were consistently maintained across all studies, with up to 200 epochs and a batch size of 20, considering the computational resource constraints. Our ablation study utilized a lightweight version of RED-CNN as the base generator, akin to Huang *et al.* [10], and a conventional classifier as the base discriminator. In **Table II** and **Fig. 5**, we define the following nomenclature: (a) a single-task discriminator for image-level decision, (b) a single-task discriminator for pixel-level decision, (c), (d), and (e) representing dual-task discriminators combining two of three tasks—an image-level decision, a pixel-level decision, and a restoration task—and (f) a triple-task discriminator integrating decision tasks with restoration. Further, we examined enhancements to the triple-task model, including NDS regulation (g), RC regulation (h), and their combination (i). We also explored the impact of replacing the generator with a Res-FFT-Conv Generator (j) and implementing PCGrad (k). Except for (i) and (j), all models used the base generator. **Table II** and **Fig. 5** illustrate significant step-by-step improvements in quantitative and qualitative performance with the gradual incorporation of MTD-GAN elements. Specifically, denoising performance in pixel- and feature-based metrics improved with the progression from single to dual and triple multi-task configurations. Additional improvements were observed through two specific regulations without increasing parameter count. The inclusion of the Res-FFT Conv block in the generator further enhanced performance, with the optimal outcome achieved by applying PCGrad to coordinate various objective functions. Detailed statistical analysis, including paired t-tests for each metric across all MTD-GAN elements, can be found in **Appendix I-D**.

TABLE III
SUBJECTIVE QUALITY ASSESSMENTS (MEAN ± SD) FOR VARIOUS ALGORITHMS USING A BLIND READER STUDY

Brain

Methods	Reader A			Reader B			Reader C			Average		
	Noise reduction	Contrast retention	Structure preservation	Noise reduction	Contrast retention	Structure preservation	Noise reduction	Contrast retention	Structure preservation	Noise reduction	Contrast retention	Structure preservation
RED-CNN	13.3333±0.4714	3.9333±0.2494	3.4000±0.4899	13.4000±0.4899	4.0000±0.0000	3.8000±0.4000	13.2667±0.4422	3.8667±0.3399	3.7333±0.4422	13.3333±0.0544	3.9333±0.0544	3.6444±0.1750
EDCNN	10.6667±0.6992	3.0000±0.0000	2.0000±0.0000	11.3333±0.4714	3.0667±0.4422	2.8667±0.3399	11.4667±0.4989	2.9333±0.2494	2.8667±0.3399	11.1556±0.3500	3.0000±0.0544	2.5778±0.4086
Restormer	13.6667±0.4714	4.0000±0.0000	2.0000±0.0000	13.6000±0.4899	3.9333±0.2494	4.0000±0.0000	13.7333±0.4422	4.0000±0.0000	4.0000±0.0000	13.6667±0.0544	**3.9778±0.0314**	**4.0000±0.0000**
CTformer	11.4000±0.7118	3.0000±0.0000	4.0000±0.0000	11.6000±0.6110	3.0000±0.0000	2.9333±0.2494	11.4667±0.6182	3.0000±0.0000	2.8667±0.3399	11.4889±0.0831	3.0000±0.0000	2.6000±0.4251
DDPM	6.0667±0.8537	2.9333±0.5735	2.9333±0.2494	6.2000±0.7483	2.8000±0.4000	2.8667±0.3399	5.8667±0.6182	3.0000±0.0000	3.0000±0.3651	6.0444±0.1370	2.9111±0.0831	2.9333±0.0544
DDIM	6.2667±0.7037	3.0000±0.0000	2.8667±0.3519	6.4000±0.6325	2.9333±0.2582	2.8000±0.4140	6.4000±0.6325	3.0000±0.0000	2.8667±0.3519	**6.3556±0.0629**	2.9778±0.0314	2.8444±0.0314
PNDM	6.3333±0.6992	3.0667±0.2494	2.9333±0.2494	6.2667±0.7717	3.0000±0.0000	2.9333±0.2494	6.2000±0.7483	3.0000±0.0000	3.0000±0.3651	6.2667±0.0544	3.0222±0.0314	2.9556±0.0314
DPM	6.0667±0.7717	2.8667±0.3399	2.9333±0.2494	6.2667±0.5735	2.8667±0.3399	2.9333±0.2494	6.2667±0.5735	2.9333±0.2494	2.9333±0.2494	6.2000±0.0943	2.8889±0.0314	2.9333±0.0000
WGAN-VGG	3.1333±0.4989	2.0000±0.0000	2.0000±0.0000	2.8667±0.3399	2.0667±0.2494	2.0667±0.2494	2.8667±0.3399	2.0000±0.0000	2.0000±0.0000	2.9556±0.1257	2.0222±0.0314	2.0222±0.0314
MAP-NN	5.8000±1.1662	3.0000±0.8944	2.7333±0.5735	5.4000±1.0832	2.7333±0.4422	2.8000±0.5416	5.5333±1.2579	2.9333±0.4422	3.0000±0.5164	5.5778±0.1663	2.8889±0.1133	2.8444±0.1133
DU-GAN	7.8667±1.0873	3.1333±0.4989	3.3333±0.5963	7.7333±1.2365	3.2667±0.5735	3.5333±0.6182	7.8667±1.3597	3.2000±0.5416	3.2000±0.5416	7.8222±0.0629	3.2000±0.0544	3.3556±0.1370
MTD-GAN	7.0000±0.0000	4.0000±0.0000	3.9333±0.2494	6.9333±0.2494	3.8000±0.4000	3.8667±0.3399	7.0000±0.0000	4.0000±0.0000	3.9333±0.2494	**6.9778±0.0314**	**3.9333±0.0943**	**3.9111±0.0314**
NDCT	7.0000±0.0000	5.0000±0.0000	5.0000±0.0000	7.0000±0.0000	5.0000±0.0000	5.0000±0.0000	7.0000±0.0000	5.0000±0.0000	5.0000±0.0000	7.0000±0.0000	5.0000±0.0000	5.0000±0.0000
LDCT	1.0000±0.0000	1.0000±0.0000	1.0000±0.0000	1.0000±0.0000	1.0000±0.0000	1.0000±0.0000	1.0000±0.0000	1.0000±0.0000	1.0000±0.0000	1.0000±0.0000	1.0000±0.0000	1.0000±0.0000

Abdomen

Methods	Reader A			Reader B			Reader C			Average		
	Noise reduction	Contrast retention	Structure preservation	Noise reduction	Contrast retention	Structure preservation	Noise reduction	Contrast retention	Structure preservation	Noise reduction	Contrast retention	Structure preservation
RED-CNN	13.6000±0.4899	3.8000±0.4000	2.4667±0.4989	13.8667±0.3399	3.8000±0.4000	3.0667±0.5735	13.5333±0.4989	3.6000±0.4899	2.9333±0.5735	13.6667±0.1440	**3.7333±0.0943**	2.8222±0.2572
EDCNN	10.2000±0.9092	3.1333±0.3399	2.3333±0.5963	10.6000±0.9522	3.2667±0.4422	2.7333±0.5735	10.4000±0.9522	3.1333±0.3399	2.8667±0.4989	10.4000±0.1633	3.1778±0.0629	2.6444±0.2266
Restormer	13.0000±1.0954	3.6000±0.4899	3.9333±0.2494	12.5333±1.0873	3.6667±0.4714	3.8667±0.3399	13.0667±1.1235	3.6667±0.4714	3.8667±0.3399	12.8667±0.2373	3.6444±0.0314	**3.8889±0.0314**
CTformer	11.8667±0.7180	3.4667±0.4989	2.4000±0.7118	11.8000±0.8327	3.4667±0.4989	2.9333±0.7717	11.8000±0.7483	3.4667±0.4989	2.9333±0.7717	11.8222±0.0314	3.4667±0.0000	2.7556±0.2514
DDPM	6.8667±0.4989	3.7333±0.5735	2.6667±0.5963	6.4000±1.2000	3.4000±0.6110	2.8667±0.4989	6.8000±0.6532	3.3333±0.5963	2.8000±0.6532	6.6889±0.2061	3.4889±0.1750	2.7778±0.0831
DDIM	7.7333±0.5936	3.6000±0.5071	3.0667±0.4577	7.5333±0.6399	3.4667±0.5164	3.1333±0.5164	7.6000±0.6325	3.4667±0.5164	3.2000±0.5606	7.6222±0.0831	3.5111±0.0629	3.1333±0.0544
PNDM	7.7333±0.5735	3.7333±0.4422	2.9333±0.5735	7.2000±0.7483	3.6667±0.4714	3.2000±0.6532	7.5333±0.8055	3.6000±0.4899	3.1333±0.6182	7.4889±0.2200	3.6667±0.0544	3.0889±0.1133
DPM	7.0667±0.2494	3.8000±0.4000	3.0000±0.3651	7.0667±0.2494	3.4667±0.4989	2.9333±0.2494	7.0667±0.2494	3.4000±0.4899	3.0000±0.3651	**7.0667±0.0000**	3.5556±0.1750	2.9778±0.0314
WGAN-VGG	3.0000±0.0000	2.0000±0.0000	2.1333±0.3399	2.8667±0.3399	2.0000±0.0000	2.1333±0.3399	2.8667±0.3399	2.0000±0.0000	2.0667±0.2494	2.9111±0.0629	2.0000±0.0000	2.1111±0.0314
MAP-NN	5.0000±0.6325	2.5333±0.6182	2.2000±0.4000	4.9333±0.4422	2.4000±0.4899	2.2000±0.4000	4.7333±0.6799	2.5333±0.4989	2.2667±0.4422	4.8889±0.1133	2.4889±0.0629	2.2222±0.0314
DU-GAN	6.2667±0.6799	3.2000±0.6532	2.3333±0.6992	6.6000±0.6110	3.2000±0.5416	2.8667±0.7180	6.4667±0.8055	3.0667±0.5735	2.8000±0.6532	6.4444±0.1370	3.1556±0.0629	2.6667±0.2373
MTD-GAN	6.7333±0.4422	3.9333±0.2494	3.7333±0.4422	6.8000±0.4000	3.7333±0.4422	3.7333±0.4422	6.8667±0.3399	3.8667±0.3399	3.8000±0.4000	**6.8000±0.0544**	**3.8444±0.0831**	**3.7556±0.0314**
NDCT	7.0000±0.0000	5.0000±0.0000	5.0000±0.0000	7.0000±0.0000	5.0000±0.0000	5.0000±0.0000	7.0000±0.0000	5.0000±0.0000	5.0000±0.0000	7.0000±0.0000	5.0000±0.0000	5.0000±0.0000
LDCT	1.0000±0.0000	1.0000±0.0000	1.0000±0.0000	1.0000±0.0000	1.0000±0.0000	1.0000±0.0000	1.0000±0.0000	1.0000±0.0000	1.0000±0.0000	1.0000±0.0000	1.0000±0.0000	1.0000±0.0000

Note: The mean and standard deviation values for the blind test samples per reader are presented. The 'Average' column shows the mean values across the three readers along with the inter-reader standard deviation (SD). Red indicates the best performance, while blue indicates the second-best performance. In all metrics, closer values to NDCT indicate better performance.

E. Blind Reader Study

For a reliable qualitative assessment, we conducted a blind reader study using image slices from 15 randomly selected groups, each comprising brain and abdomen test sets. Each group consisted of 14 pairs of images, including an LDCT image, an NDCT image, and denoised LDCT images. The study involved three board-certified radiologists, each with 15 years of experience in interpreting brain and abdominopelvic CT scans. They independently evaluated each denoised image in terms of noise reduction, contrast retention, and structure preservation using a scoring system that permitted identical scores against LDCT and NDCT benchmarks. A 1-14 point scoring system was employed to evaluate noise reduction, using NDCT as the baseline (7 points), and assessing the over-smooth state (8-14 points) and under-noise state (1-6 points) to determine the closest noise resemblance to NDCT. For evaluating contrast retention and structural preservation, a 1-5 point scoring system was used, with LDCT images rated at 1 and NDCT images at 5. For more detailed evaluation criteria and examples, please refer to **Appendix I-F**. This approach more accurately reflects the comparison of similarities between synthetic and NDCT images from the radiologist's point of view. **Table III** presents the average and standard deviation of blind test results per sample and per radiologist. As shown in **Table III**, MTD-GAN achieved performance closest to NDCT in terms of noise reduction in brain images and contrast retention in abdomen images, along with competitive scores in other evaluation metrics.

F. Effect of Regulatory Mechanisms

To demonstrate the effectiveness of the two regulations, we conducted confidence map comparisons with and without regulations (see **Fig. 6**). We randomly selected a patient from the test set and used four real NDCT slices of varying depths. The comparison is made for the four cases (f)-(i) in **Table II**. The pixel-level decision confidence maps for the input and its corresponding restoration are represented in each case. In these maps, the intensity of red indicates confidence in the real class, while the intensity of blue signifies confidence in the synthesized class. As illustrated in **Fig. 6**, with NDS regulation, the discriminator's focus was on the region of interest (ROI) within the brain. In contrast, without NDS regulation, the discriminator's activation occasionally concentrated on confusing areas, which had similar intensities in both LDCT and NDCT images, such as the background and bone regions. With RC regulation, the discriminator's output was unaffected by the restoration, leading to more consistent confidence maps. In the absence of RC regulation, however, the discriminator's activations differed between the input and the restoration. Employing both regulations resulted in the most ideal ROI confidence map and a notable improvement in performance.

G. Effect of Res-FFT-Conv Generator

To verify the potency of the Res-FFT-Conv Generator, we conducted experiments with and without its implementation (see **Fig. 7**). We randomly selected an NDCT slice from the brain test set and created an absolute difference map between

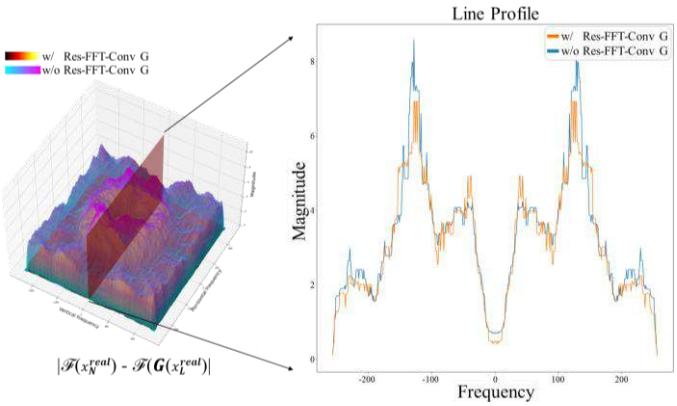

Fig. 7. Comparison of line profiles of the differences between the real and synthesized NDCT images in the Fourier domain according to the Res-FFT-Conv Generator.

the real NDCT image and the corresponding synthesized NDCT image, both with and without the Res-FFT-Conv Generator. Then, we transformed the difference map into the Fourier domain and compared their differences using a 3D surface plot for area comparison and a line profile plot for analyzing detailed signals in the frequency domain. When employing the Res-FFT-Conv Generator, the difference from the real NDCT image was relatively smaller in both the low- and high-frequency ranges. Across brain test data slices, the method using the Res-FFT-Conv Generator demonstrated an average of 6% closer alignment to the NDCT image in terms of area profile compared to the method that did not utilize it.

V. Discussion

Due to visual incongruence for radiologists, the unsatisfactory performance of various evaluation metrics, and insufficient exploration of denoising tasks in other CT domains, challenges have persisted in developing LDCT denoising methods. In this paper, we propose a novel GAN-based algorithm using a multi-task discriminator for LDCT denoising to overcome these challenges for clinical applications. The discriminator simultaneously conducts three vision tasks, transferring semantic feedback to the generator and synthesizing denoised high-quality images. Furthermore, two regulations are introduced to improve the representation capabilities of the discriminator. Lastly, Res-FFT-Conv blocks are applied in the generator for the integration of frequency and spatial information.

As shown in **Section IV-C**, each type of previous DL-based denoising method has its own tendencies and limitations. Previous CNN- and TR-based methods predominantly focus on pixel-based optimization, offering robustness in terms of PSNR and RMSE. However, despite their superiority in PSNR and RMSE, the methods often lead to over-smoothing and poorly correlate with human perception of image quality, making the synthesized images visually discomforting for radiologists [15, 16]. Specifically, ED-CNN tries to capture details by adding a trainable sharpness kernel, but this also emphasizes unnecessary noise. Additionally, CTformer exhibits boundary artifacts during the overlapped inference process, with the artifacts being sensitive to the degree of overlap. Previous DN- and GAN-based methods may also result in signal distortion, despite their considerable proficiency in utilizing detailed features. This is evident in WGAN-VGG and DN-based models, which risk erasing existing anatomical structures or generating wrong clinical signals during denoising. Moreover, DN-based models depend on specific solvers for rapid inference and remain vulnerable to NFE, indicating potential areas for further development. Furthermore, despite DU-GAN showing the most competitive performance, its gain is insufficient for the cost of adding an additional discriminator (see **Table V**). Our proposed MTD-GAN is the closest network to NDCT in terms of quantitative and qualitative evaluations, as well as in visual scoring by expert radiologists. Moreover, it demonstrated the most robust denoising performance, particularly showing superior structural similarity across pixel-space and feature-space in both brain and abdominal CT domains, potentially paving the way for clinical applications.

In **Section IV-D**, similar to Kyung *et al.* [51], a performance enhancement was observed with the increase in the diversity of MTL configurations (single, double, triple, etc.). This led to a more effective discriminator that provided meaningful feedback to the generator. Notably, performing both image-level and pixel-level decision tasks enables the discriminator to consistently provide global and local feedback to the generator. Additionally, the integration of the restoration task enhanced the discriminator's capabilities at both the image and pixel levels by extracting more contextual information, thereby facilitating the provision of more meaningful feedback to the generator.

Through **Section IV-F**, the active area of the discriminator is more focused on the ROI with NDS regulation. We interpret that NDS regulation prevents confusion in the discriminator and induces better focus on ROI regions between LDCT and NDCT images. The RC regulation leads the discriminator to more consistent confidence maps. By comparing the discriminator's outputs for the input with the restoration that retains the input's contextual information, the RC regulation offers a more comprehensive and profound understanding of the global and local aspects of real or synthesized classes. This contributes to mitigating overfitting to the input while improving the discriminator's performance in restoration, image-level, and pixel-level decision tasks. Lastly, the two regulations did not interfere with each other, and employing both regulations yields superior performance compared to using a single regulation, resulting in a more pertinent ROI confidence map.

As demonstrated in **Section IV-G**, the successful integration of the Res-FFT-Conv block into the generator for LDCT denoising not only proves its effectiveness but also showcases its adaptability. Furthermore, the performance of the Res-FFT-Conv Generator was particularly notable in the Fourier domain.

In **Appendix I-A**, it was observed that the performance of our multi-task framework was sensitive to task-specific weights. This issue was addressed through the implementation of an MTL optimization algorithm. Selecting the most robust algorithm within our framework resulted in a comprehensive improvement in performance.

TABLE IV
COMPARATIVE QUANTITATIVE ANALYSIS OF MULTI-OBJECTIVE OPTIMIZATION (MOO)

Brain						
Methods	FID↓	PL↓	TML↓	RMSE↓	PSNR↑	SSIM↑
All One	17.7768	0.1297±0.0512	10.5051±4.8779	0.0368±0.0140	32.4826±7.6717	0.9036±0.0605
MGDA	30.5830	0.1414±0.0572	13.4197±6.0028	**0.0338±0.0126**	**33.1172±7.4603**	0.9036±0.0616
PCGrad	**17.3800**	**0.1277±0.0507**	**10.3253±4.8421**	0.0366±0.0137	32.6347±8.0107	**0.9043±0.0604**
CAGrad	19.2791	0.1284±0.0506	10.4747±4.8640	0.0366±0.0138	32.5345±7.7455	0.9041±0.0605
GT	9.8355	0.0	0.0	0.0	100.0	1.0
Input	39.7583	0.1821	20.4448	0.0575	28.955	0.8743
Abdomen						
Methods	FID↓	PL↓	TML↓	RMSE↓	PSNR↑	SSIM↑
All One	17.8494	0.1179±0.0209	9.5047±1.1569	0.0241±0.0061	32.5787±1.9300	0.9232±0.0275
MGDA	29.0294	0.1401±0.0233	13.5391±1.6692	**0.0234±0.0050**	**32.8006±1.6795**	0.9238±0.0260
PCGrad	**17.0614**	**0.1168±0.0215**	**9.3077±1.2832**	0.0241±0.0060	32.5841±1.9164	0.9235±0.0275
CAGrad	17.3645	0.1185±0.0215	9.6379±1.2653	0.0237±0.0060	32.7489±1.9379	**0.9247±0.0274**
Gt	0.851	0.0	0.0	0.0	100.0	1.0
Input	42.4862	0.1663	16.5566	0.0356	29.2489	0.8883

Note: The best performances are marked in bold.

Although our method demonstrates enhanced performance, it still has certain limitations. First, various CT equipment, vendors, and kernels exist in clinical CT examinations, which may lead to domain gaps. The LDCT images used in this study were obtained from a single machine and vendor. Consequently, our network, trained in a controlled environment, may not show optimal performance when applied to real clinical patient images, resulting in degraded results. In future research, our study will focus on unsupervised learning incorporating actual clinical data, aiming to develop an approach robust against various factors that influence domain gaps. Second, our study should be extended beyond training on brain and abdomen images to include chest and whole-body images for our model to be effectively applied in the real-world medical community. Third, task balancing is a critical aspect of MTL. Although we implemented the PCGrad algorithm to automatically explore appropriate weights in various objective functions, we have localized the problem of multi-objective optimization to multi-task discriminators, which means that there is still room for further improvement. Furthermore, this approach might not be optimal in other domains, and ongoing studies on multi-task balancing are still needed.

VI. Conclusion

In this paper, we propose a multi-task discriminator generative adversarial network for LDCT image denoising, named MTD-GAN. MTD-GAN utilizes (i) multi-task learning in the discriminator to transfer contextual, global, and local feedback to the generator, (ii) RC and NDS regulations to improve the representation capabilities of the discriminator for consistency and ROI-focused activations, and (iii) Res-FFT-Conv blocks in the generator for fusing frequency-spatial dual-domain representations. Our comprehensive experimental results, including visual scoring, demonstrate MTD-GAN's effectiveness in both brain and abdominal CT domains. Notably, MTD-GAN achieved superior performance in both quantitative and qualitative measures compared to state-of-the-art denoising techniques across various metrics based on pixel- and feature-spaces.

Appendix I
Supplemental Materials

A. Exploration of MTL optimization algorithms

In this section, we address the optimization challenge inherent in MTL [27]. The primary objective in multi-task learning is to minimize the average loss across all tasks, which involves making predictions for multiple tasks simultaneously and sharing the knowledge among them. This learning improves generalization performance by reducing computational costs and improving data efficiency; however, as the gradients of these different tasks may conflict, it sometimes causes negative transfer, where it performs worse than the corresponding single task.

Recently, the main cause of this performance degradation has been found to be gradient collisions, and several model-agnostic methods [48, 54, 55] have been designed to manipulate gradients for identifying a more effective update vector.

MGDA. Sener *et al.* propose an upper bound for the multi-objective loss and employ the Multiple-Gradient Descent Algorithm (MGDA) for MTL, which is designed to directly optimize towards the Pareto set [54].

PCGrad. Yu *et al.* introduced Projecting Conflicting Gradients (PCGrad) to address gradient conflicts in MTL by projecting each task's gradient onto the normal plane of others, effectively balancing objective functions [48].

CAGrad. Liu *et al.* used Conflict-Averse Gradient descent (CAGrad), which minimizes the average loss across tasks by searching the update vector that maximizes the worst improvement among all tasks in a neighborhood of the average gradient [55].

In our study, we paid considerable attention to these aspects and conducted a series of ablation studies to identify the most suitable MTL optimization algorithms for our framework. Our primary focus was on the multi-task discriminator, aiming to understand its performance within the MTL context. Regarding the generator's objective function, we set the weights λ_{5-7} in accordance with Shan *et al.* [8], specifically setting λ_5 to 1 and $\lambda_{6,7}$ to 50. For the multi-task discriminator, we explored several scenarios for the discriminator's weights λ_{1-4}, including setting them to "1," updating the discriminator using the MGDA algorithm, applying the PCGrad algorithm, and utilizing the

TABLE V
COMPARISON OF RESOURCES AND PARAMETERS ACROSS DIFFERENT NETWORKS

Type	Methods	Generator (Inference)		Generator (Train)		Discriminator (Train)		Total (Train)	
		MACs (G)↓	Param. (M)↓	MACs (G)↓	Param. (M)↓	MACs (G)↓	Param. (M)↓	MACs (G)↓	Param. (M)↓
Previous works									
CNNs	RED-CNN	462.0927	1.8489	5.0467	1.8489	-	-	5.0467	1.8489
	ED-CNN	21.0806	0.0809	0.3294	0.0809	-	-	0.3294	0.0809
TRs	Restormer	50.8362T	26.0940	8.8013	26.0940	-	-	8.8013	26.0940
	CTformer	4.9727T	1.3824	0.8609	1.3824	-	-	0.8609	1.3824
DNs	DDPM	1042.4371T	18.0257	16.2896T	18.0257	-	-	16.2896T	18.0257
	DDIM	52.1219T	18.0257	814.4814	18.0257	-	-	814.4814	18.0257
	PNDM	52.1219T	18.0257	814.4814	18.0257	-	-	814.4814	18.0257
	DPM	52.1219T	18.0257	814.4814	18.0257			814.4814	18.0257
GANs	WGAN-VGG	24.2431	0.0925	0.3788	0.0925	0.2834	17.9235	0.6622	18.0160
	MAP-NN	79.7498	0.0620	1.0855	0.0620	0.2834	17.9235	1.3689	17.9855
	DU-GAN	48.4694	0.1856	0.7573	0.1856	3.9654	114.4276	4.7227	114.6132
	MTD-GAN	110.5220	0.4671	1.7317	0.4671	2.7204	68.4320	4.4521	68.8991
Ablation study									
Single	(a): $D_{img.}$	48.4694	0.1856	0.7573	0.1856	1.2174	28.8731	1.9747	29.0587
	(b): $D_{pix.}$	48.4694	0.1856	0.7573	0.1856	1.9352	46.6405	2.6925	46.8261
Dual	(c): $D_{img.} + D_{pix.}$	48.4694	0.1856	0.7573	0.1856	1.9355	46.9037	2.6928	47.0893
	(d): $D_{img.} + D_{rest.}$	48.4694	0.1856	0.7573	0.1856	2.0023	50.4014	2.7596	50.5870
	(e): $D_{pix.} + D_{rest.}$	48.4694	0.1856	0.7573	0.1856	2.7201	68.1689	3.4774	68.3545
Triple	(f): $D_{img.} + D_{pix.} + D_{rest.}$	48.4694	0.1856	0.7573	0.1856	2.7204	68.4320	3.4777	68.6176
	(g): (f) + NDS reg.	48.4694	0.1856	0.7573	0.1856	2.7204	68.4320	3.4777	68.6176
	(h): (f) + RC reg.	48.4694	0.1856	0.7573	0.1856	2.7204	68.4320	3.4777	68.6176
	(i): (f) + NDS + RC reg.	48.4694	0.1856	0.7573	0.1856	2.7204	68.4320	3.4777	68.6176
	(j): (i) + Res-FFT-Conv Generator	110.5220	0.4671	1.7317	0.4671	2.7204	68.4320	4.4521	68.8991
	(k): (j) + PCGrad (Ours)	110.5220	0.4671	1.7317	0.4671	2.7204	68.4320	4.4521	68.8991

Note: Param. refers to trainable parameters, and MACs denotes Multiply-Accumulate Operations. A model with fewer parameters and lower MACs is generally more efficient, meaning it requires less computational power and memory.

CAGrad algorithm. The results detailed in **Table IV** show that the PCGrad algorithm was particularly robust and effective in feature-based metrics (FID, PL, TML) across brain and abdomen datasets, and excelled in pixel-based metrics (SSIM) in the brain dataset. Consequently, we integrated PCGrad into our framework due to its superior performance across these metrics.

B. Computational Efficiency of Networks

To compare computational resources across different models, we utilized the pytorch-OpCounter library [56] to count the trainable parameters (Param.) and Multiply-Accumulate Operations (MACs), serving as metrics for memory and computational efficiency, respectively. Generally, models with fewer trainable parameters and lower MACs are more efficient, implying reduced memory needs and computational load. We measured the metrics in both the training and inference settings. The image size for inference was 512 × 512, matching the original CT scans, while for training, the image size was 64 × 64, based on the patch-based training method. As detailed in **Table V**, during inference, DNs have shown improved sampling speeds with specific solvers, but the inference speed remains relatively slow and more memory-intensive compared to other models. TRs also exhibit prolonged processing times due to the necessity of overlapped inference. For GANs, the introduction of a discriminator for adversarial training requires additional parameters compared to other models. Considering DU-GAN's performance as the most competitive to MTD-GAN, as shown in **Table II**, MTD-GAN's superiority in both efficiency and effectiveness is noteworthy, despite its increased parameters owing to its multi-tasks.

C. Supplemental clinical structure comparison

For the clinical task-based assessment, additional figures featuring cancerous lesions and important tissues were included from the brain and abdomen test sets. In **Fig. 8**, the top part showcases a brain CT image focusing on intracerebral hemorrhage (ICH) and subarachnoid hemorrhage (SAH), while the bottom part displays an abdominal CT image focusing on the kidney and spleen. **Fig. 9** presents a brain CT image at the top, concentrating on the sulcus and falx cerebri, and an abdominal CT image at the bottom, centering on the liver. In both **Figs. 8** and **9**, the Regions of Interest (ROIs) marked by red rectangles are magnified below each corresponding CT image for enhanced visibility. The arrows in the images indicate clinically important areas (marked in red) and areas of greatest variation (marked in yellow). As demonstrated in **Figs. 8** and **9**, MTD-GAN effectively suppresses noise and artifacts while preserving more anatomical details crucial for diagnosis, achieving a quality closest to NDCT images and outperforming other models.

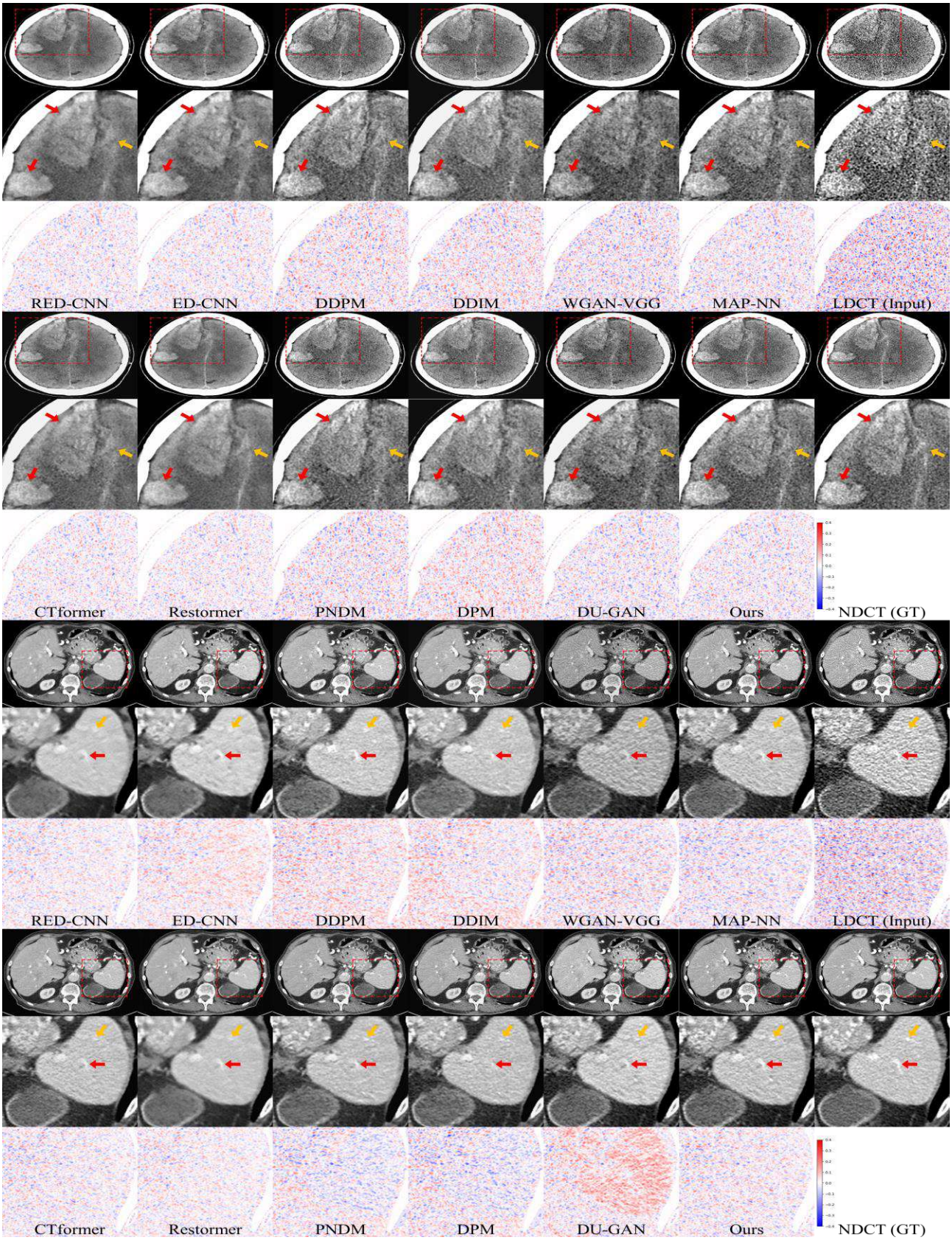

Fig. 8. Results of denoising for comparison on the test sets of brain with intracerebral hemorrhage (ICH) and subarachnoid hemorrhage (SAH), and abdomen of spleen. The ROI marked by the red rectangles are zoomed in the below row, respectively. The red arrow indicates clinically prominent points, and the yellow one highlights areas with significant differences. In brain case, the display window is [0, 80] HU. The red arrow points to ICH and the yellow one points to SAH. In abdomen case, the display window is [-160, 240] HU. The red and yellow arrows indicate vessels in spleen. The last row of each model shows a difference map, created by calculating the difference between the synthesized image and the NDCT image, with positive values depicted in red and negative values in blue.

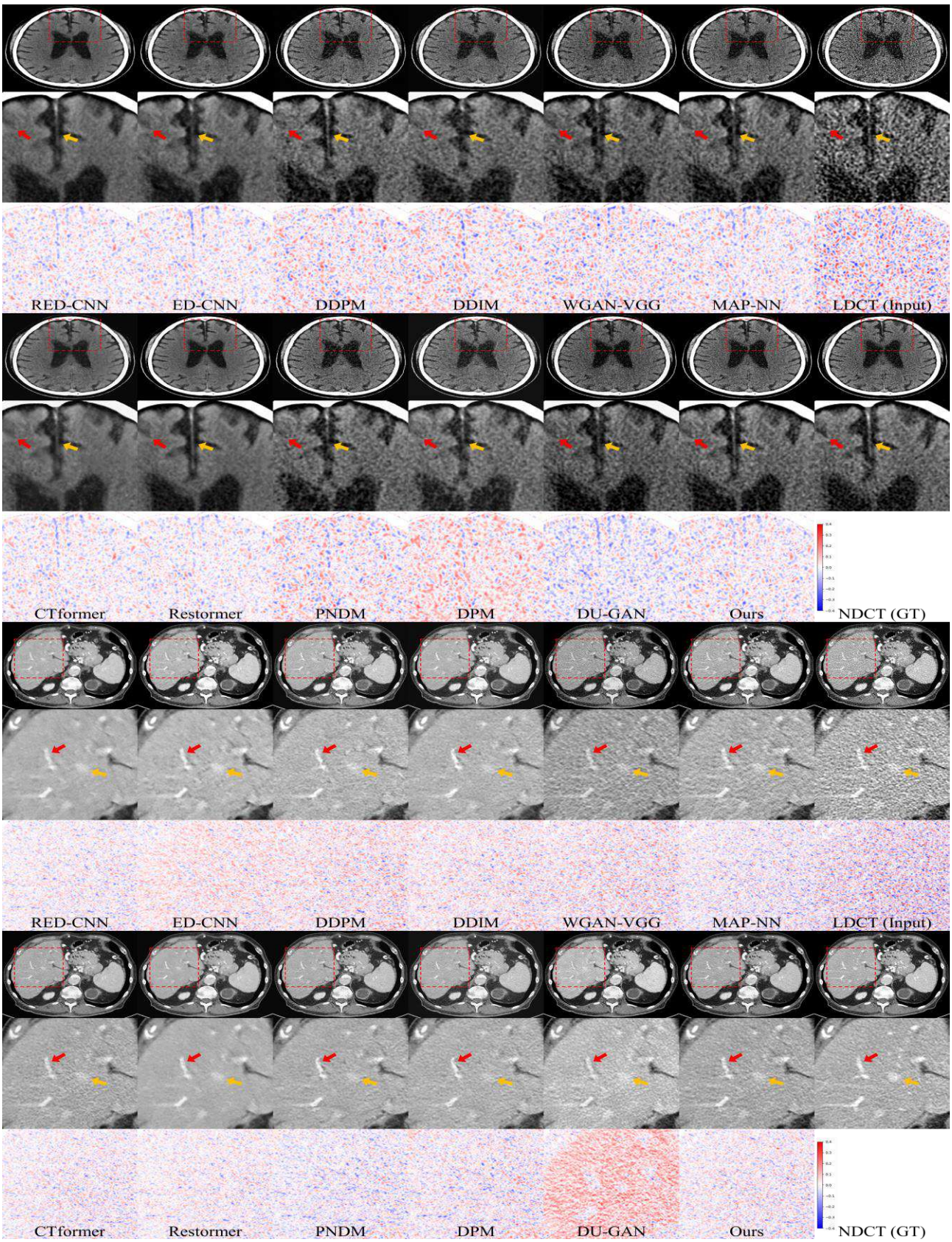

Fig. 9. Results of denoising for comparison on the test sets of brain and abdomen of liver. The ROI marked by the red rectangles are zoomed below, respectively. The red arrow indicates clinically prominent points, and the yellow one highlights areas with significant differences. In brain case, the display window is [0, 80] HU. The red arrow points to sulcus and the yellow one points to falx cerebri. In abdomen case, the display window is [-160, 240] HU. The red and yellow arrows indicate vessels in liver. The red and yellow arrows indicate vessels in spleen. The last row of each model shows a difference map, created by calculating the difference between the synthesized image and the NDCT image, with positive values depicted in red and negative values in blue.

TABLE VI
STATISTICAL SIGNIFICANCE COMPARISON OF MTD-GAN ELEMENTS

Brain

PL	(a)	(b)	(c)	(d)	(e)	(f)	(g)	(h)	(i)	(j)	(k)
(a)			***	***	***	***	***	***	***	***	***
(b)			***	***	***	***	***	***	***	***	***
(c)				***		***	***	***	***	***	***
(d)						***	***	***	***	***	***
(e)						***	***	***	***	***	***
(f)								***	***	***	***
(g)								***	***	***	***
(h)									**	***	***
(i)										***	***
(j)											***
(k)											

TML	(a)	(b)	(c)	(d)	(e)	(f)	(g)	(h)	(i)	(j)	(k)
(a)		***	***	***	***	***	***	***	***	***	***
(b)			***	***	***	***	***	***	***	***	***
(c)				***		***	***	***	***	***	***
(d)					***	***	***	***	***	***	***
(e)						***	***	***	***	***	***
(f)							***	***	***	***	***
(g)								***	***	***	***
(h)									***	***	***
(i)										***	***
(j)											***
(k)											

RMSE	(a)	(b)	(c)	(d)	(e)	(f)	(g)	(h)	(i)	(j)	(k)
(a)		***	***	***	***	***	***	***	***	***	***
(b)			***	***	***	***	***	***	***	***	***
(c)				***	***	***	***	***	***	***	***
(d)						***	***	***	***	***	***
(e)						***	***	***	***	***	***
(f)							***	***	***	***	***
(g)								***	***	***	***
(h)								***	***	***	***
(i)									***	***	***
(j)										***	***
(k)											

PSNR	(a)	(b)	(c)	(d)	(e)	(f)	(g)	(h)	(i)	(j)	(k)
(a)				*	***	**	***	***	***	***	***
(b)			***	**	***	***	***	***	***	***	***
(c)					*		**	*	**	***	***
(d)									*	***	***
(e)									**	***	***
(f)							*	*	**	***	***
(g)										***	***
(h)									**		***
(i)											***
(j)											***
(k)											

SSIM	(a)	(b)	(c)	(d)	(e)	(f)	(g)	(h)	(i)	(j)	(k)
(a)		***	***	***	***	***	***	***	***	***	***
(b)				***	***	***	***	***	***	***	***
(c)				***	***	***	***	***	***	***	***
(d)					***	***	***	**	***	***	***
(e)						***		**	***	***	***
(f)							***	***	***	***	***
(g)								***	***	***	***
(h)									***	***	***
(i)										***	***
(j)											***
(k)											

Abdomen

PL	(a)	(b)	(c)	(d)	(e)	(f)	(g)	(h)	(i)	(j)	(k)
(a)			***	***	***	***	***	***	***	***	***
(b)			*		***	***	***	***	***	***	***
(c)					***	***	***	***	***	***	***
(d)					***	***	***	***	***	***	***
(e)						***	***	***	***	***	***
(f)							***	***	***	***	***
(g)									***	***	***
(h)									***	***	***
(i)										***	***
(j)											***
(k)											

TML	(a)	(b)	(c)	(d)	(e)	(f)	(g)	(h)	(i)	(j)	(k)
(a)			***	***	***	***	***	***	***	***	***
(b)			***	***	***	***	***	***	***	***	***
(c)						***	***	***	***	***	***
(d)						***	***	***	***	***	***
(e)						***	***	***	***	***	***
(f)							***	***	***	***	***
(g)								***	***	**	***
(h)											***
(i)											***
(j)											***
(k)											

RMSE	(a)	(b)	(c)	(d)	(e)	(f)	(g)	(h)	(i)	(j)	(k)
(a)		**	***	*	***	**	***	***	***	***	***
(b)			***		***	**	***	***	***	***	***
(c)								***	***	***	***
(d)					**	***	***	***	***	***	***
(e)								***	***	***	***
(f)							**	***	***	***	***
(g)								***	***	***	***
(h)									***	***	***
(i)										***	***
(j)											**
(k)											

PSNR	(a)	(b)	(c)	(d)	(e)	(f)	(g)	(h)	(i)	(j)	(k)
(a)			***		**		***	***	***	***	***
(b)			***		***		***	***	***	***	***
(c)				***		*		***	***	***	***
(d)					***	***	***	***	***	***	***
(e)						***		***	***	***	***
(f)							***	***	***	***	***
(g)								***	***	***	***
(h)									***	***	***
(i)										***	***
(j)											
(k)											

SSIM	(a)	(b)	(c)	(d)	(e)	(f)	(g)	(h)	(i)	(j)	(k)
(a)		*	***	***	***	***	***	***	***	***	***
(b)				**	***	***	***	***	***	***	***
(c)				***	***	***	***	***	***	***	***
(d)									***	***	***
(e)							**	***	***	***	***
(f)							*	**	***	***	***
(g)									***	***	***
(h)									*	***	***
(i)										***	***
(j)											***
(k)											

Note: Statistical analysis in the ablation study on the test set. P-values were calculated as follows: *, P < 0.01; **, P < 0.001; ***, P < 0.0001. Single-task discriminators for decisions at the image level (a) and pixel level (b), dual-task discriminators merging any two of the following three tasks: image-level decisions, pixel-level decisions, and a restoration task (c, d, e), as well as a triple-task discriminator that combines decision tasks with restoration (f). Furthermore, the enhancements applied to the triple-task model encompassed NDS regulation (g), RC regulation (h), the integration of both NDS and RC regulations (i), the substitution of the generator with a Res-FFT-Conv Generator (j), and the implementation of PCGrad (k).

D. Statistical Significance Comparison

The statistical significance of all combinations was assessed using the MedCalc program [57]. Paired t-tests were performed for each metric, excluding FID, and the results are presented in **Table VI**, utilizing nomenclature consistent with **Table II**. The analysis included single-task discriminators for image-level (a) and pixel-level (b) decisions; dual-task discriminators that combine two of three tasks—an image-level decision, a pixel-level decision, and a restoration task (c, d, e); and a triple-task discriminator integrating decision tasks with restoration (f). Enhancements to the triple-task model included NDS regulation (g), RC regulation (h), a combination of both NDS and RC

TABLE VII
ANALYSIS OF FACTORS THAT AFFECT THE DISCRIMINATOR'S DECISIONS

Brain				
Ablation Study	$\boldsymbol{D}_{img.}(x_L^{real})\downarrow$	$\boldsymbol{D}_{pix.}(x_L^{real})\downarrow$	$\boldsymbol{D}_{img.}(\boldsymbol{D}_{rest.}(x_L^{real}))\downarrow$	$\boldsymbol{D}_{pix.}(\boldsymbol{D}_{rest.}(x_L^{real}))\downarrow$
$D_{img.}$	0.6905±0.0032	-	-	-
$D_{img.} + D_{rest.}$	0.6294±0.1194	-	0.6019±0.1219	-
$D_{pix.}$	-	0.6840±0.0898	-	-
$D_{pix.} + D_{rest.}$	-	0.6397±0.1059	-	0.6159±0.1061
$D_{img.} + D_{pix.}$	0.6383±0.0909	0.6412±0.0886	-	-
$D_{img.} + D_{pix.} + D_{rest.}$	0.5925±0.0014	0.6164±0.0064	0.5826±0.0013	0.6057±0.0063
$D_{img.} + D_{pix.} + D_{rest.}$ + NDS reg.	0.5744±0.0492	0.6004±0.0471	0.5720±0.0514	0.5972±0.0497
$D_{img.} + D_{pix.} + D_{rest.}$ + RC reg.	0.5779±0.1085	0.5954±0.1074	0.5575±0.1097	0.5612±0.1087
$D_{img.} + D_{pix.} + D_{rest.}$ + NDS + RC reg.	0.5593±0.1140	0.5905±0.1141	0.5394±0.1153	0.5510±0.1155
Abdomen				
Ablation Study	$\boldsymbol{D}_{img.}(x_L^{real})\downarrow$	$\boldsymbol{D}_{pix.}(x_L^{real})\downarrow$	$\boldsymbol{D}_{img.}(\boldsymbol{D}_{rest.}(x_L^{real}))\downarrow$	$\boldsymbol{D}_{pix.}(\boldsymbol{D}_{rest.}(x_L^{real}))\downarrow$
$D_{img.}$	0.7936±0.0016	-	-	-
$D_{img.} + D_{rest.}$	0.6962±0.0384	-	0.6551±0.0451	-
$D_{pix.}$	-	0.7609±0.0515	-	-
$D_{pix.} + D_{rest.}$	-	0.6989±0.0531	-	0.6893±0.0554
$D_{img.} + D_{pix.}$	0.6997±0.0631	0.7005±0.0597	-	-
$D_{img.} + D_{pix.} + D_{rest.}$	0.6391±0.0442	0.6319±0.0451	0.5141±0.0400	0.5092±0.0366
$D_{img.} + D_{pix.} + D_{rest.}$ + NDS reg.	0.6102±0.0543	0.5907±0.0507	0.4971±0.0410	0.4994±0.0351
$D_{img.} + D_{pix.} + D_{rest.}$ + RC reg.	0.6018±0.0498	0.5538±0.0474	0.4852±0.0510	0.4868±0.0484
$D_{img.} + D_{pix.} + D_{rest.}$ + NDS + RC reg.	0.5334±0.0478	0.5299±0.0456	0.4704±0.0438	0.4794±0.0409

Note: x_L^{real} represents the real LDCT images fed to the discriminator. $\boldsymbol{D}_{img.}(\cdot)$ and $\boldsymbol{D}_{pix.}(\cdot)$ denote the image-level and pixel-level decision outputs of the discriminator $\boldsymbol{D}$, respectively, with lower values indicating superior recognition performance. $\boldsymbol{D}_{rest.}(x_L^{real})$ is the restored image of x_L^{real} by the discriminator.

regulations (i), replacement of the generator with a Res-FFT-Conv Generator (j), and the addition of PCGrad (k). **Table VI** shows significant differences when the multi-task level was varied from single to dual, and then to triple. The implementation of two regulations also demonstrated significant improvement. Moreover, replacing the Res-FFT-Conv Generator and applying the PCGrad algorithm resulted in a notable difference.

E. Factors that affect the discriminator's decisions

In this section, we analyze factors influencing the image-level and pixel-level decisions of the multi-task discriminator, particularly regarding the impact of the restoration task and regulations. We also examine the effect of reusing restored images as input on the discriminator's decisions when the restoration task exists.

Accurately measuring the discriminator's performance in image-level and pixel-level decision tasks within a GAN framework, which is based on the zero-sum game principle, is challenging [19, 23]. Specifically, in the image-translation GAN framework, the generator aims to convert LDCT images to NDCT images to a degree that sufficiently confuses the discriminator in differentiating between real and synthesized images in the NDCT domain. Owing to this challenge, it is necessary to focus on evaluating the discriminator's performance in the LDCT domain, where LDCT images are recognized as a synthesized class, to ensure more accurate measurement. In our experiments, we fed real LDCT images (x_L^{real}) and their restored counterparts into the discriminator. Lower outputs of $\boldsymbol{D}_{image}(\cdot)$ and $\boldsymbol{D}_{pixel}(\cdot)$ indicate the discriminator's effectiveness in recognizing the inputs as a synthesized class.

As illustrated in **Table VII**, the discriminator's ability to identify LDCT images as a synthesized class is influenced by the integration of the restoration task and regulations. Interestingly, reintroducing restored images to the discriminator improved its ability to discern images in the LDCT domain. This improvement is interpreted as a result of the "refinement" process inherent in the encoding-decoding pathway, where the discriminator filters out marginal information while retaining critical elements, thereby enhancing its ability to differentiate between real and synthesized images.

F. Detailed evaluation criteria for a Blind Reader Study

Our criteria aim to determine which converted images most closely resemble NDCT images, focusing on noise reduction, contrast retention, and structure preservation. Noise reduction is evaluated primarily by changes in particle size, while image homogeneity and contrast resolution are also considered. Contrast retention is evaluated by measuring the density differences between adjacent areas and assessing the consistency of these differences compared to NDCT. Structural preservation focuses on maintaining normal anatomy without introducing imprecise elements or compromising contrast resolution. To validate these subjective visual quality assessments, radiologists performed a preliminary review using non-test images not included in the blind reader study, ensuring that the evaluation criteria were standardized through ample discussion and practice. Furthermore, it should be noted that these evaluation criteria contain subjective elements that may vary depending on the radiologist's expertise and experience.

Given the wide degree of noise distribution, a 1-14 point scoring system was used for the noise reduction metric in the blind reader study. NDCT images were given a standard score

of 7 points. Scores above 7 indicated overly smoothed images, while scores below 7 suggested insufficient noise reduction. As detailed in **Table VIII**, contrast retention and structure preservation metrics were assessed using a 1-5 point scoring system, where LDCT images scored 1 point and NDCT images scored 5 points.

Fig. 10 illustrates representative sample images to aid in understanding our evaluation standards for the blind reader study. Different colored arrows indicate key areas that influence heterogeneity. Notably, yellow arrows indicate newly formed structures not previously present, mimicking lesions. Blue arrows highlight areas where image quality is degraded, introducing blotchy and noisy heterogeneity. Red arrows mark areas where existing structures are damaged or removed. Each evaluation criterion is arranged and expressed in ascending order, with the numbers in (*) representing the value of the corresponding metric.

In the case of A. Noise Reduction in **Fig. 10**, under-denoised images (scores 1-6) exhibit larger particle sizes compared to NDCT images, leading to poor contrast resolution between neighboring structures. Conversely, over-denoised images (scores 8-14) feature smaller particle sizes. Despite this reduction, significant image heterogeneity persists in the ED-CNN and CTformer models, which can mimic pathological lesions and create artifacts. Additionally, abdomen CT images generated by DU-GAN also exhibit heterogeneity.

In the case of B. Contrast Retention in **Fig. 10**, this parameter assesses the density differences between adjacent structures, such as gray and white matter in brain CTs, and vascular structures in abdomen CT images. The DU-GAN model shows diminished density contrast in the deep gray matter of brain CTs compared to NDCT. Similarly, image heterogeneity, marked by yellow and blue arrows in the ED-CNN and CTformer models, also reduces density contrast in certain regions.

In the case of C. Structure Preservation in **Fig. 10**, this assessment focuses on the clarity with which normal anatomical structures are visualized, including the cerebral cortex and deep gray matter in brain CTs, and vascular structures in abdomen CT images. WGAN models generally perform poorly in visualizing anatomical structures due to inadequate denoising. Conversely, ED-CNN, CTformer, and other diffusion-based models, except for MTD-GAN, often yield unrealistic or missing normal structures.

TABLE VIII
VISUAL QUALITY EVALUATION CRITERIA OF DENOISED IMAGES IN TERMS OF CONTRAST RETENTION AND STRUCTURE PRESERVATION

Score	Contrast Retention	Structure Preservation
1	Very poor contrast resolution	Very poor visualization of anatomical structures
2	Poor contrast resolution	Poor visualization of anatomical structures
3	Fair contrast resolution, but some areas exhibit heterogeneity	Fair visualization of structures with some missing anatomical features or pseudo-lesions
4	Good contrast resolution, consistent across the image (homogeneity)	Good visualization of structures, consistent across the image (homogeneity)
5	Excellent contrast resolution	Excellent visualization of anatomical structures

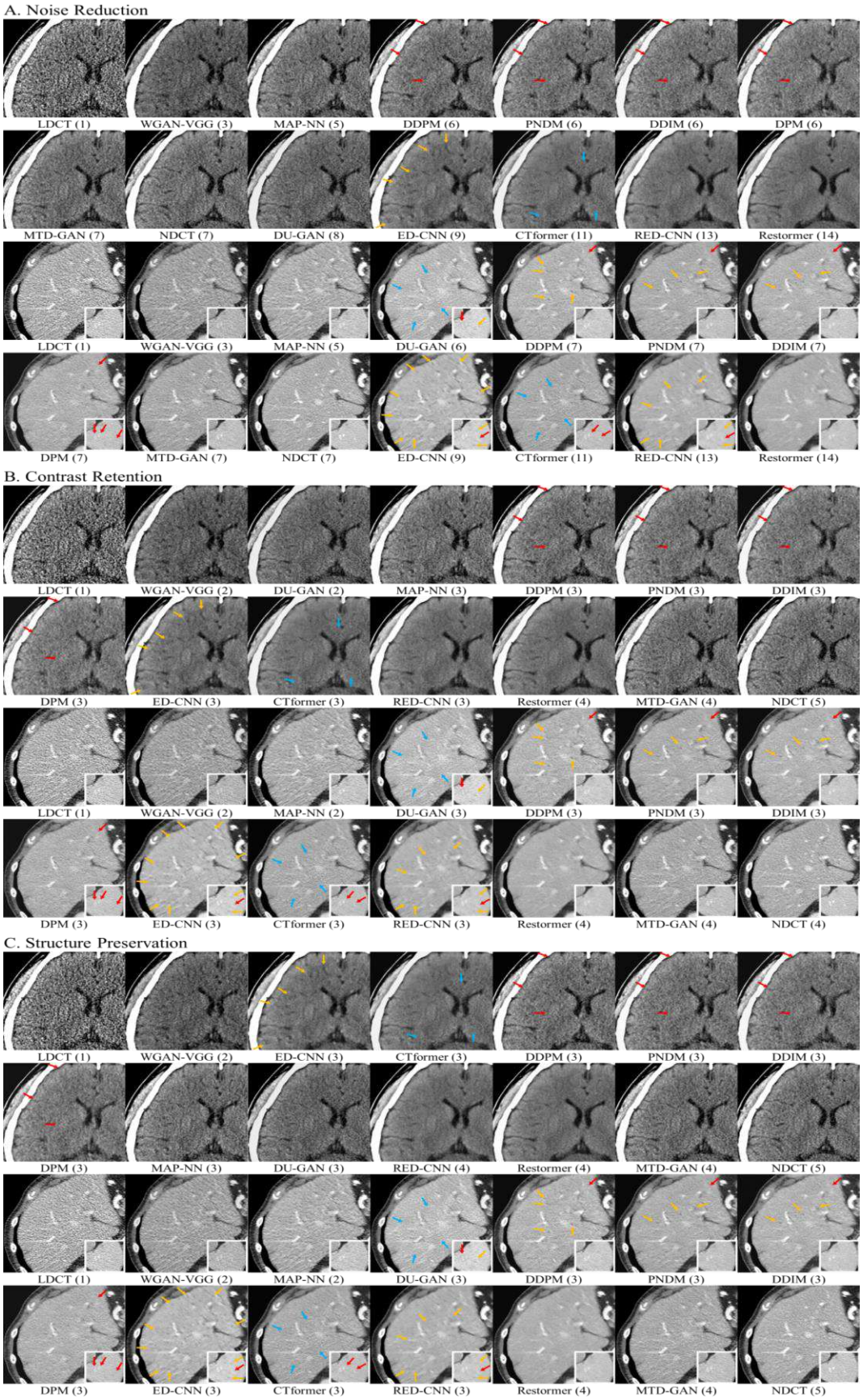

Fig. 10. Examples of the blind test evaluation standards in brain and abdomen CT images in terms of noise reduction, contrast retention, and structure preservation. Different colored arrows indicate key areas affecting heterogeneity: yellow arrows for newly formed structures mimicking lesions, blue arrows for degraded image quality, and red arrows for damaged or removed existing structures. Each evaluation criterion is expressed in ascending order, with numbers (*) representing the corresponding metric values.

ACKNOWLEDGMENT

This work was supported by grants from the Korea Health Technology R&D Project through the Korea Health Industry Development Institute (KHIDI), funded by the Ministry of Health & Welfare, Republic of Korea (HR20C0026, HI22C0471, HI21C1148). We are grateful to Ji Eun Shin, MD for participating in the blinded reader study. We acknowledge the open-source libraries, including the Diffuser [52] and MONAI Generative Models [53], which enabled valuable comparisons in this study, and we extend our thanks to the pioneering authors. In the spirit of contributing to the research community, we will share our code and research materials at https://github.com/babbu3682/MTD-GAN, hoping that our work will serve as a valuable resource for the community.

REFERENCES

[1] A. B. de Gonzalez, and S. Darby, "Risk of cancer from diagnostic X-rays: estimates for the UK and 14 other countries," *The Lancet,* vol. 363, no. 9406, pp. 345-351, 2004.

[2] D. J. Brenner, and E. J. Hall, "Computed tomography—an increasing source of radiation exposure," *New England J Med,* vol. 357, no. 22, pp. 2277-2284, 2007.

[3] K. J. Strauss, and S. C. Kaste, "The ALARA (as low as reasonably achievable) concept in pediatric interventional and fluoroscopic imaging: striving to keep radiation doses as low as possible during fluoroscopy of pediatric patients—a white paper executive summary," *Radiology,* vol. 240, no. 3, pp. 621–622, 2006.

[4] J. Wang, Y. Tang, Z. Wu *et al.*, "Domain-adaptive denoising network for low-dose CT via noise estimation and transfer learning," *Med Phys,* vol. 50, no. 1, pp. 74–88, 2023.

[5] C. H. McCollough, A. C. Bartley, R. E. Carter *et al.*, "Low-dose CT for the detection and classification of metastatic liver lesions: results of the 2016 low dose CT grand challenge," *Medical physics,* vol. 44, no. 10, pp. e339-e352, 2017.

[6] H. Chen, Y. Zhang, M. K. Kalra *et al.*, "Low-dose CT with a residual encoder-decoder convolutional neural network," *IEEE Trans Med. Imaging,* vol. 36, no. 12, pp. 2524–2535, 2017.

[7] Q. Yang, P. Yan, Y. Zhang *et al.*, "Low-dose CT image denoising using a generative adversarial network with Wasserstein distance and perceptual loss," *IEEE Trans Med. Imaging,* vol. 37, no. 6, pp. 1348–1357, 2018.

[8] H. Shan, A. Padole, F. Homayounieh *et al.*, "Competitive performance of a modularized deep neural network compared to commercial algorithms for low-dose CT image reconstruction," *Nat. Mach. Intell.,* vol. 1, no. 6, pp. 269–276, 2019.

[9] T. Liang, Y. Jin, Y. Li *et al.*, "Edcnn: Edge enhancement-based densely connected network with compound loss for low-dose ct denoising." pp. 193–198.

[10] Z. Huang, J. Zhang, Y. Zhang *et al.*, "DU-GAN: Generative adversarial networks with dual-domain U-Net-based discriminators for low-dose CT denoising," *IEEE Trans. Instrum. Meas.* vol. 71, pp. 1–12, 2021.

[11] W. Xia, Q. Lyu, and G. Wang, "Low-Dose CT Using Denoising Diffusion Probabilistic Model for 20× Speedup," *arXiv preprint arXiv:2209.15136*, 2022.

[12] D. Wang, F. Fan, Z. Wu *et al.*, "CTformer: convolution-free Token2Token dilated vision transformer for low-dose CT denoising," *Phys. Med. Biol.,* vol. 68, no. 6, pp. 065012, 2023.

[13] X. Mao, Y. Liu, F. Liu *et al.*, "Intriguing findings of frequency selection for image deblurring." pp. 1905–1913.

[14] S. Kyung, J. Won, S. Pak *et al.*, "MTD-GAN: Multi-task Discriminator Based Generative Adversarial Networks for Low-Dose CT denoising." pp. 133–144.

[15] Z. Wang, A. C. Bovik, H. R. Sheikh *et al.*, "Image quality assessment: from error visibility to structural similarity," *IEEE Trans. Image Process.,* vol. 13, no. 4, pp. 600–612, 2004.

[16] I. Goodfellow, J. Pouget-Abadie, M. Mirza *et al.*, "Generative adversarial networks," *Commun. ACM,* vol. 63, no. 11, pp. 139–144, 2020.

[17] S. W. Zamir, A. Arora, S. Khan *et al.*, "Restormer: Efficient transformer for high-resolution image restoration." pp. 5728–5739.

[18] E. Schonfeld, B. Schiele, and A. Khoreva, "A u-net based discriminator for generative adversarial networks." pp. 8207–8216.

[19] Z. Wang, Q. She, and T. E. Ward, "Generative adversarial networks in computer vision: A survey and taxonomy," *ACM Comput. Surveys (CSUR),* vol. 54, no. 2, pp. 1–38, 2021.

[20] F.-A. Croitoru, V. Hondru, R. T. Ionescu *et al.*, "Diffusion models in vision: A survey," *IEEE Trans. Pattern Anal. and Mach. Intell.*, 2023.

[21] Y. Shi, and G. Wang, "Conversion of the Mayo LDCT Data to Synthetic Equivalent through the Diffusion Model for Training Denoising Networks with a Theoretically Perfect Privacy," *arXiv preprint arXiv:2301.06604*, 2023.

[22] K. Roth, A. Lucchi, S. Nowozin *et al.*, "Stabilizing training of generative adversarial networks through regularization," *Adv. Neural inf. Process. Syst.,* vol. 30, 2017.

[23] L. Mescheder, A. Geiger, and S. Nowozin, "Which training methods for GANs do actually converge?." pp. 3481–3490.

[24] T. Miyato, T. Kataoka, M. Koyama *et al.*, "Spectral normalization for generative adversarial networks," *arXiv preprint arXiv:1802.05957*, 2018.

[25] T. Karras, T. Aila, S. Laine *et al.*, "Progressive growing of gans for improved quality, stability, and variation," *arXiv preprint arXiv:1710.10196*, 2017.

[26] C. Yang, Y. Shen, Y. Xu *et al.*, "Improving gans with a dynamic discriminator," *Adv. Neural Inf. Process. Syst.,* vol. 35, pp. 15093–15104, 2022.

[27] S. Vandenhende, S. Georgoulis, W. Van Gansbeke *et al.*, "Multi-task learning for dense prediction tasks: A survey," *IEEE Trans. Pattern Anal. Mach. Intell.,* vol. 44, no. 7, pp. 3614–3633, 2021.

[28] R. Hang, F. Zhou, Q. Liu *et al.*, "Classification of hyperspectral images via multitask generative adversarial networks," *IEEE Trans. Geosci. Remote Sens.,* vol. 59, no. 2, pp. 1424–1436, 2020.

[29] M. S. Rad, B. Bozorgtabar, C. Musat *et al.*, "Benefiting from multitask learning to improve single image super-resolution," *Neurocomput.,* vol. 398, pp. 304–313, 2020.

[30] W. Wan, and H. J. Lee, "Generative adversarial multi-task learning for face sketch synthesis and recognition." pp. 4065–4069.

[31] J. Cha, S. Chun, G. Lee *et al.*, "Few-shot compositional font generation with dual memory." pp. 735–751.

[32] Y. Liu, Z. Wang, H. Jin *et al.*, "Multi-task adversarial network for disentangled feature learning." pp. 3743–3751.

[33] G. Mordido, H. Yang, and C. Meinel, "Dropout-gan: Learning from a dynamic ensemble of discriminators," *arXiv preprint arXiv:1807.11346*, 2018.

[34] Y. Katznelson, *An introduction to harmonic analysis*: Cambridge University Press, 2004.

[35] L. Chi, B. Jiang, and Y. Mu, "Fast fourier convolution," *Adv. Neural Inf. Process. Syst.,* vol. 33, pp. 4479–4488, 2020.

[36] Y. Yang, and S. Soatto, "Fda: Fourier domain adaptation for semantic segmentation." pp. 4085–4095.

[37] R. Suvorov, E. Logacheva, A. Mashikhin *et al.*, "Resolution-robust large mask inpainting with fourier convolutions." pp. 2149–2159.

[38] W. Shi, J. Caballero, F. Huszár *et al.*, "Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network." pp. 1874–1883.

[39] X. Mao, Q. Li, H. Xie *et al.*, "Least squares generative adversarial networks." pp. 2794–2802.

[40] H. Zhang, Z. Zhang, A. Odena *et al.*, "Consistency regularization for generative adversarial networks," *arXiv preprint arXiv:1910.12027*, 2019.

[41] H. Wang, X. Wu, Z. Huang *et al.*, "High-frequency component helps explain the generalization of convolutional neural networks." pp. 8684–8694.

[42] S. W. Zamir, A. Arora, S. Khan *et al.*, "Multi-stage progressive image restoration." pp. 14821–14831.

[43] S. Ellmann, F. Kammerer, M. Brand *et al.*, "A novel pairwise comparison-based method to determine radiation dose reduction potentials of iterative reconstruction algorithms, exemplified through circle of Willis computed tomography angiography," *Invest. Radiol.,* vol. 51, no. 5, pp. 331–339, 2016.

[44] J. Ho, A. Jain, and P. Abbeel, "Denoising diffusion probabilistic models," *Adv. Neural Inf. Process. Syst.,* vol. 33, pp. 6840–6851, 2020.

[45] J. Song, C. Meng, and S. Ermon, "Denoising diffusion implicit models," *arXiv preprint arXiv:2010.02502*, 2020.

[46] L. Liu, Y. Ren, Z. Lin *et al.*, "Pseudo numerical methods for diffusion models on manifolds," *arXiv preprint arXiv:2202.09778*, 2022.

[47] C. Lu, Y. Zhou, F. Bao *et al.*, "Dpm-solver: A fast ode solver for diffusion probabilistic model sampling in around 10 steps," *Adv. Neural Inf Process. Syst.,* vol. 35, pp. 5775–5787, 2022.

[48] T. Yu, S. Kumar, A. Gupta *et al.*, "Gradient surgery for multi-task learning," *Adv. Neural Inf. Process. Syst.,* vol. 33, pp. 5824–5836, 2020.

[49] M. S. Sajjadi, B. Scholkopf, and M. Hirsch, "Enhancenet: Single image super-resolution through automated texture synthesis." pp. 4491–4500.

[50] M. Heusel, H. Ramsauer, T. Unterthiner *et al.*, "Gans trained by a two time-scale update rule converge to a local nash equilibrium," *Adv. Neural Inf Process. Syst.,* vol. 30, 2017.

[51] S. Kyung, K. Shin, H. Jeong *et al.*, "Improved performance and robustness of multi-task representation learning with consistency loss between pretexts for intracranial hemorrhage identification in head CT," *Med. Image Anal.,* vol. 81, pp. 102489, 2022.

[52] P. von Platen, S. Patil, A. Lozhkov *et al.*, "Diffusers: State-of-the-art diffusion models," 2022.

[53] W. H. Pinaya, M. S. Graham, E. Kerfoot *et al.*, "Generative ai for medical imaging: extending the monai framework," *arXiv preprint arXiv:2307.15208*, 2023.

[54] O. Sener, and V. Koltun, "Multi-task learning as multi-objective optimization," *Adv. Neural Inf. Process. Syst.,* vol. 31, 2018.

[55] B. Liu, X. Liu, X. Jin *et al.*, "Conflict-averse gradient descent for multi-task learning," *Adv. Neural Inf. Process. Syst.,* vol. 34, pp. 18878–18890, 2021.

[56] L. Zhu, "Thop: pytorch-OpCounter," Available online: https://github.com/Lyken17/pytorch-OpCounter.

[57] F. Schoonjans, A. Zalata, C. Depuydt *et al.*, "MedCalc: a new computer program for medical statistics," *Compu. Methods Programs Biomed.,* vol. 48, no. 3, pp. 257–262, 1995.