

# Content-Noise Complementary Learning for Medical Image Denoising

Mufeng Geng<sup>✉</sup>, Xiangxi Meng<sup>✉</sup>, Jiangyuan Yu, Lei Zhu<sup>✉</sup>, Lujia Jin, Zhe Jiang<sup>✉</sup>, Bin Qiu, Hui Li, Hanjing Kong<sup>✉</sup>, Jianmin Yuan, Kun Yang, Hongming Shan<sup>✉</sup>, Hongbin Han, Zhi Yang, Qiushi Ren, and Yanye Lu<sup>✉</sup>

**Abstract**—Medical imaging denoising faces great challenges, yet is in great demand. With its distinctive characteristics, medical imaging denoising in the image domain requires innovative deep learning strategies. In this study,

Manuscript received June 16, 2021; revised July 22, 2021 and September 6, 2021; accepted September 13, 2021. Date of publication September 16, 2021; date of current version February 2, 2022. This work was supported in part by Beijing Natural Science Foundation under Grant Z210008, in part by Shenzhen Science and Technology Program under Grant 1210318663, in part by Hebei Natural Science Foundation under Grant H2019201378, in part by Shenzhen Nanshan Innovation and Business Development Grant, and in part by Shanghai Municipal of Science and Technology Project under Grant 20JC1419500. (*Corresponding author: Yanye Lu.*)

This work involved human subjects in its research. Approval of all ethical and experimental procedures and protocols was granted by Beijing Cancer Hospital Institutional Review Board under Approval Nos. 2020KT15 and 2017KT94, and performed in line with ethical standards of the institutional and/or national research committee, and the Declaration of Helsinki.

Mufeng Geng, Lei Zhu, Lujia Jin, Zhe Jiang, Bin Qiu, and Qiushi Ren are with the Institute of Medical Technology, Peking University Health Science Center, Peking University, Beijing 100191, China, also with the Department of Biomedical Engineering, College of Future Technology, Peking University, Beijing 100871, China, also with the Institute of Biomedical Engineering, Peking University Shenzhen Graduate School, Shenzhen 518055, China, also with the Institute of Biomedical Engineering, Shenzhen Bay Laboratory, Shenzhen 518071, China, and also with the National Biomedical Imaging Center, Beijing 100871, China (e-mail: 1701111665@pku.edu.cn; zhulei@stu.pku.edu.cn; jinlujia@pku.edu.cn; gjiang47@163.com; qiub@pku.edu.cn; qren@pku.edu.cn).

Xiangxi Meng, Jiangyuan Yu, Hui Li, and Zhi Yang are with the Key Laboratory of Carcinogenesis and Translational Research (Ministry of Education), Department of Nuclear Medicine, Peking University Cancer Hospital and Institute, Beijing 100142, China, and also with the Key Laboratory for Research and Evaluation of Radiopharmaceuticals (National Medical Products Administration), Beijing 100037, China (e-mail: mengxiangxi@pku.edu.cn; yujiangyuan@aliyun.com; huilee1990@163.com; pekyz@163.com).

Hanjing Kong and Jianmin Yuan are with the Central Research Institute, UIH Group, Beijing 100080, China (e-mail: hanjing.kong@united-imaging.com; jianmin.yuan@united-imaging.com).

Kun Yang is with the College of Quality and Technical Supervision, Hebei University, Baoding 071000, China (e-mail: hbyangkun@163.com).

Hongming Shan is with the Institute of Science and Technology for Brain-Inspired Intelligence, Fudan University, Shanghai 200433, China, and also with Shanghai Research Center for Brain Science and Brain-Inspired Technology, Shanghai 201210, China (e-mail: hmshan@fudan.edu.cn).

Hongbin Han and Yanye Lu are with the Institute of Medical Technology, Peking University Health Science Center, Peking University, Beijing 100191, China, and also with Beijing Key Laboratory of Magnetic Resonance Imaging Technology, Beijing 100191, China (e-mail: hanhongbin@bjmu.edu.cn; yanye.lu@pku.edu.cn).

This article has supplementary downloadable material available at <https://doi.org/10.1109/TMI.2021.3113365>, provided by the authors.

Digital Object Identifier 10.1109/TMI.2021.3113365

we propose a simple yet effective strategy, the content-noise complementary learning (CNCL) strategy, in which two deep learning predictors are used to learn the respective content and noise of the image dataset complementarily. A medical image denoising pipeline based on the CNCL strategy is presented, and is implemented as a generative adversarial network, where various representative networks (including U-Net, DnCNN, and SRDenseNet) are investigated as the predictors. The performance of these implemented models has been validated on medical imaging datasets including CT, MR, and PET. The results show that this strategy outperforms state-of-the-art denoising algorithms in terms of visual quality and quantitative metrics, and the strategy demonstrates a robust generalization capability. These findings validate that this simple yet effective strategy demonstrates promising potential for medical image denoising tasks, which could exert a clinical impact in the future. Code is available at: <https://github.com/gengmufeng/CNCL-denoising>.

**Index Terms**—Deep learning, low dose, CT, MR, PET, image restoration.

## I. INTRODUCTION

THE practice of computer vision frequently involves image denoising, which aims to remove noise from the corrupted image and restore the true image. Special significance is attached to denoising in medical imaging processing, for the noise may baffle the disease diagnosis and affect subsequent clinical decision-making [1]. Computed tomography (CT), magnetic resonance (MR) imaging, and positron emission tomography (PET) are three common medical imaging modalities in clinical diagnosis. CT and MR provide structural information at high resolution, whereas PET is a molecular imaging modality to provide metabolic and functional information [2]. The denoising algorithms for these modalities aim to break through the trade off among scanning time, radiation intensity, and image quality. As for CT, the past decade has seen a trend towards low-dose examinations to reduce the ionizing radiation exposure, usually by decreasing the tube current or shortening the exposure time of the X-ray tube [3]. Unlike CT and PET, there is no radiation risk in MR examinations, while the narrow space and the long acquisition time of MR may induce anxiety especially for claustrophobia patients [4]. To alleviate patients' discomfort, there are some attempts to under-sampling  $k$ -space to achieve fast MR acquisition [5]. In PET scan, there is also a desire to reconstruct perfect images with small number of coincidence events [6]. Notwithstanding these above-mentioned attempts greatly alleviate the physical

and psychological hazards of patients, the reconstructed image is often corrupted, introducing a variety of noise or artifacts.

A range of methods have been proposed to improve the image quality of CT, MR, and PET. These image restoration methods are implemented on various stages, including the raw data preprocessing before reconstruction, during reconstruction, or post-processing after reconstruction. The former two rely on raw data acquisition, which is often vendor-specific; while image post-processing directly operates on corrupted images and can be easily integrated into existing clinical procedures. Thus, many researchers solve medical image denoising problems in the image domain. In the field of medical image denoising, traditional image-processing algorithms include non-local means (NLM) [7], block-matching 3D (BM3D) [8], and diffusion filters [9], etc. Although these algorithms can reduce the noise to various extents, they result in over-smoothness and residual errors in the denoised images.

Recently, deep learning has shown its superior capability in the medical imaging tasks, such as segmentation [10], material decomposition [11], [12], as well as denoising. For low-dose CT denoising, Chen *et al.* proposed a residual encoder-decoder convolutional neural network (RED-CNN) to estimate routine-dose CT values [13]; Fan *et al.* designed a quadratic autoencoder (Q-AE) network for low-dose CT denoising [14]; furthermore, a series of generative adversarial networks (GAN) were applied to the low-dose CT denoising task [15], [16], and some of them introduced perceptual loss [17], [18]. As for MR, both [19] and [20] adopted U-Net to remove aliasing artifacts and streaking artifacts from distorted MR images, respectively; Hyun *et al.* jointly used the U-Net denoising and the  $k$ -space correction to restore under-sampled MR images [21]; Jiang *et al.* utilized a denoising convolutional neural network (DnCNN) to repair the MR images corrupted by Rician noise [22]; Kidoh *et al.* designed a shrinkage convolutional neural network (SCNN) and a deep learning-based reconstruction (dDLR) network to denoise brain MR images [23]. In the field of low-dose PET denoising, Xiang *et al.* estimated a full-dose PET image from a pair of 25% dose PET and MR images, with an auto-context CNN model [24]; Sano *et al.* modified U-Net to reconstruct full-dose PET images from low-dose PET images [25]; Wang *et al.* presented an approach based on a conditional generative adversarial (cGAN) to estimate the high-quality full-dose PET images from low-dose ones [26]; Kim *et al.* optimized a DnCNN to perform PET image denoising [27]; Lu *et al.* comprehensively investigated the denoising performance of a convolutional autoencoder (CAE) network, a U-Net, and a GAN in oncological PET data [28]; Zhou *et al.* proposed a unified motion correction and denoising adversarial network (DPET) to simultaneously perform denoising and motion estimation for low-dose gated PET [29]; both [30] and [31] focused on deep image prior for PET denoising; in addition, there were some low-dose PET denoising works based on cycleconsistent generative adversarial networks (Cycle GAN) [32]–[34] and Wasserstein generative adversarial network (WGAN) [35]. These above-mentioned methods have demonstrated their capabilities on medical

image denoising tasks, and most of them focused on optimizing the network structures or designing case-dependent loss functions. Furthermore, we observe that, they either learn content images (i.e., full-dose or full-sampled images) directly [15]–[18], [26]–[29], [33], [35], [36], or learn noise images to reversely obtain the content images by subtracting [13], [14], [19], [20], [22], [25], [28], [32], [34], [37]. Both content learning and noise learning have their inherent advantages: noise learning is able to prevent performance degradation and preserve more structural and contrast details [13], while content learning may show more stable noise cancellation performance [28].

Some attempts had been made to apply the priors of both the content and the noise to denoising confocal laser endomicroscopy images with an encoder-decoder network, WhiteNNer [38]. However, WhiteNNer only used the noise prior as a loss regularization term for training under the assumption that the noise obeyed Gaussian distribution. Similarly, Liao *et al.* proposed an artifact disentanglement network (ADN) to solve CT metal artifact reduction problem [39], which learned not only the content but also the metal artifact in network training. The learned metal artifact coding in ADN was used to generate new metal artifact-corrupted images, enabling the subsequent calculation of cycle-consistent losses. That is, in the testing phase, both WhiteNNer and ADN discarded the noise features or the metal artifact coding, which meant that the learned noise of WhiteNNer and the learned metal artifact coding in ADN were not directly related to the final predicted content. Thus, for medical image denoising task, there is a demand to explore a new strategy which can make better use of the content prior and the noise prior for the final content prediction and integrate the strengths of the two learning paradigms.

In this work, we propose a content-noise complementary learning (CNCL) strategy for medical image denoising in image domain. In contrast to the existing methods that either learn the content or noise, the proposed denoising strategy learns the content and the noise by two predictors simultaneously in a complementary manner, and reconstructs the final content from the extracted features of both content predictor and noise predictor. Based on the proposed CNCL strategy, we present a universal medical image denoising pipeline, where the predictors are deep CNNs organized in a parallel way. We implement this pipeline based on a GAN framework, treating medical image denoising as an end-to-end task. We verify the proposed strategy over three medical image datasets (CT, PET, and MR), and one natural image dataset (the smartphone image denoising dataset, SIDD) [40]. The results show that, the proposed strategy outperforms state-of-the-art methods both qualitatively and quantitatively.

In summary, the contributions of this paper are as follows:

- 1) We propose a simple yet effective CNCL strategy for medical image denoising that complementarily learns content and noise to amply integrate the strengths of two paradigms.
- 2) A CNCL denoising pipeline is presented, and is implemented based on a GAN framework. We also investigate three representative classical CNN models

(i.e., U-Net [10], DnCNN [41], and SRDenseNet [42]) as the predictors in the generator part, and comprehensively investigate their denoising performance.

- 3) The proposed CNCL strategy is validated over multiple denoising datasets. The results show that our strategy is capable of dealing with a variety of noise types. The generalization capability and the regulation of the strategy are also comprehensively investigated.

## II. METHODS AND MATERIALS

### A. Content-Noise Complementary Learning

Assuming that  $\mathbf{I}_{\text{corrupted}}$  is the noise-corrupted image caused by low-dose or under-sampling, and  $\mathbf{I}_{\text{content}}$  is the corresponding full-dose or full-sampling image. In the field of deep learning-based medical image denoising,  $\mathbf{I}_{\text{corrupted}}$  can be regarded as the sum of its content  $\mathbf{I}_{\text{content}}$  and its noise  $\mathbf{I}_{\text{noise}}$  [13], [15]:

$$\mathbf{I}_{\text{corrupted}} = \mathbf{I}_{\text{content}} + \mathbf{I}_{\text{noise}}. \quad (1)$$

Existing deep learning-based image denoising methods can be divided into two categories. One employs a function to map the corrupted image to the content image directly:

$$\mathbf{I}'_{\text{content}} = p_c(\mathbf{I}_{\text{corrupted}}), \quad (2)$$

where  $\mathbf{I}'_{\text{content}}$  refers to the predicted content, and  $p_c$  denotes the content predictor.

The other predicts noise first and acquires final content image by subtracting the inputted corrupted image from the learned noise (i.e., residual mapping):

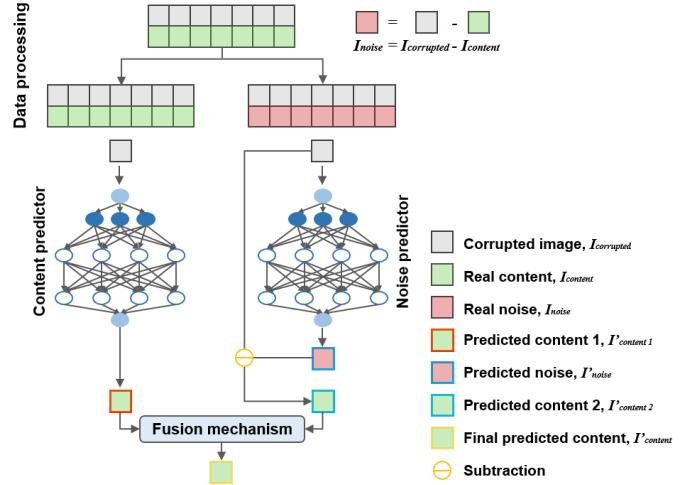
$$\begin{aligned} \mathbf{I}'_{\text{content}} &= \mathbf{I}_{\text{corrupted}} - \mathbf{I}'_{\text{noise}} \\ &= \mathbf{I}_{\text{corrupted}} - p_n(\mathbf{I}_{\text{corrupted}}), \end{aligned} \quad (3)$$

where  $\mathbf{I}'_{\text{noise}}$  indicates the predicted noise, and  $p_n$  represents the noise predictor. Both content learning and noise learning have their inherent advantages: the content learning shows more stable noise cancellation performance [28], and the noise learning is helpful for the performance degradation prevention and the structures preservation [13].

In order to combine the strengths of the two paradigms, we propose a CNCL strategy, where  $p_c$  and  $p_n$  are jointly utilized in a complementary manner. CNCL treats image denoising as an end-to-end task, whose input and output are  $\mathbf{I}_{\text{corrupted}}$  and  $\mathbf{I}'_{\text{content}}$ , respectively. The whole process of the CNCL denoising can be expressed with a deep learning function  $G$ :

$$\begin{aligned} \mathbf{I}'_{\text{content}} &= G(\mathbf{I}_{\text{corrupted}}) \\ &= f(\mathbf{I}'_{\text{content1}}, \mathbf{I}'_{\text{content2}}) \\ &= f(\mathbf{I}'_{\text{content1}}, \mathbf{I}_{\text{corrupted}} - \mathbf{I}'_{\text{noise}}) \\ &= f(p_c(\mathbf{I}_{\text{corrupted}}), \mathbf{I}_{\text{corrupted}} - p_n(\mathbf{I}_{\text{corrupted}})), \end{aligned} \quad (4)$$

where  $\mathbf{I}'_{\text{content1}}$  is the output of  $p_c$ ,  $\mathbf{I}'_{\text{noise}}$  denotes the output of  $p_n$ ,  $\mathbf{I}'_{\text{content2}}$  is the difference between  $\mathbf{I}_{\text{corrupted}}$  and  $\mathbf{I}'_{\text{noise}}$ , and  $f$  represents the fusion mechanism. First,  $\mathbf{I}_{\text{corrupted}}$  is used as the input of  $p_c$  and  $p_n$  to get the predicted content image  $\mathbf{I}'_{\text{content1}}$  and the predicted noise  $\mathbf{I}'_{\text{noise}}$ , respectively. Then,



**Fig. 1.** The proposed content-noise complementary learning (CNCL) pipeline for medical image denoising.

the other predicted content image  $\mathbf{I}'_{\text{content2}}$  is calculated by subtracting  $\mathbf{I}_{\text{corrupted}}$  from  $\mathbf{I}'_{\text{noise}}$ . After that, a fusion mechanism  $f$  is used to combine  $\mathbf{I}'_{\text{content1}}$  and  $\mathbf{I}'_{\text{content2}}$  to obtain the final predicted content  $\mathbf{I}'_{\text{content}}$ .

The learning process of CNCL can be formulated as an optimization problem:

$$\arg \min_G ||G(\mathbf{I}_{\text{corrupted}}) - \mathbf{I}_{\text{content}}||_2^2. \quad (5)$$

Based on the CNCL strategy, we propose a medical image denoising pipeline, as shown in Fig. 1. In the figure, given a pair of  $\{\mathbf{I}_{\text{corrupted}}, \mathbf{I}_{\text{content}}\}$ , the noise image  $\mathbf{I}_{\text{noise}}$  can be calculated, according to (1). As a result, we are able to expand one image pair,  $\{\mathbf{I}_{\text{corrupted}}, \mathbf{I}_{\text{content}}\}$ , to two image pairs,  $\{\mathbf{I}_{\text{corrupted}}, \mathbf{I}_{\text{content}}\}$  and  $\{\mathbf{I}_{\text{corrupted}}, \mathbf{I}_{\text{noise}}\}$ . Then,  $\{\mathbf{I}_{\text{corrupted}}, \mathbf{I}_{\text{content}}\}$  is applied for the training of  $p_c$ , and  $\{\mathbf{I}_{\text{corrupted}}, \mathbf{I}_{\text{noise}}\}$  is used to train  $p_n$ .

### B. Implementation Based on GAN

In order to optimize the deep learning function  $G$  in (5) better, we implemented the proposed CNCL strategy based on a GAN framework.  $G$  corresponds to the generator part in the GAN framework. It has been reported several times that using only some routine loss functions, such as L1 or L2, may result in oversmoothed and blurry images, losing some texture information [43]. With the help of discriminator, the details preserved in denoised images were visibly improved in previous works [15], [17], [18].

**1) Network Structure:** As shown in Fig. 2, like most GANs, the proposed GAN framework consists of two parts: a generator and a discriminator. The generator has two predictors that are organized in a parallel way. The predictors have the same network architecture, and can be implemented as CNNs. The fusion mechanism consists of a concatenation operation followed by a  $1 \times 1$  convolution operation. The fusion mechanism functions as a weighted averaging operation, and the weights are learnable due to the presence of convolution.

As a famous discriminator, PatchGAN penalizes structure mismatch at the scale of patches, functioning as a form of

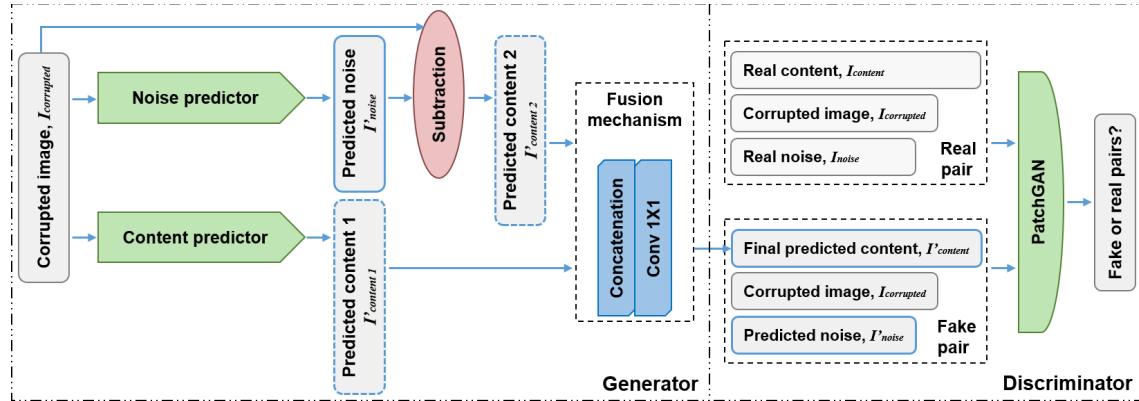


Fig. 2. A GAN-based implementation example of the content-noise complementary learning (CNCL) strategy for medical image denoising.

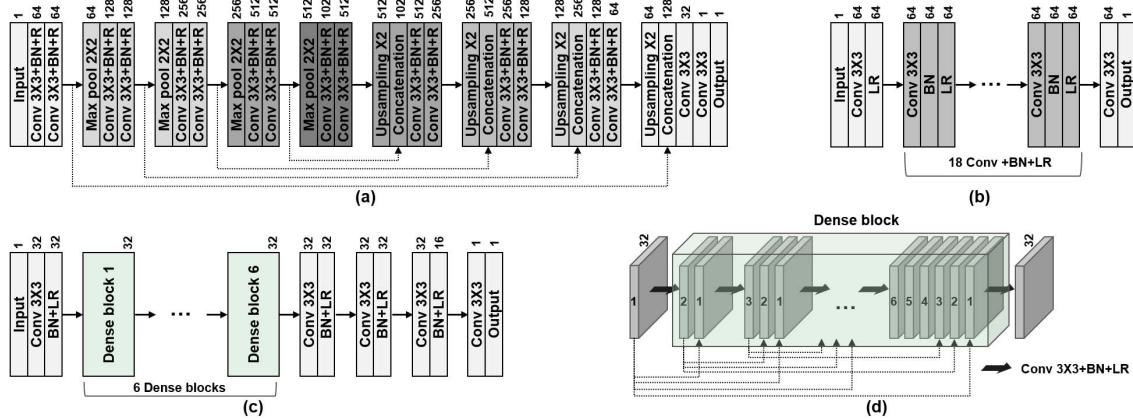


Fig. 3. Three classical CNNs adopted as the predictors in the GAN-based implementation of the CNCL strategy. (a) U-Net [10]; (b) DnCNN [41]; (c) SRDenseNet [42]; (d) the dense block in (c). BN refers to batch normalization, R denotes ReLU, and LR represents Leaky ReLU.

texture or style loss [43]. Hence, we incorporate PatchGAN as the discriminator in the proposed GAN framework. The inputs of PatchGAN are the real image pair  $\{I_{\text{content}}, I_{\text{corrupted}}, I_{\text{noise}}\}$  or the fake image pair  $\{I'_{\text{content}}, I_{\text{corrupted}}, I'_{\text{noise}}\}$ . The generator tries to generate some fake data to fool the discriminator, while PatchGAN is committed to distinguishing fake data from real data.

**2) Loss Function:** On the recommendation of [43], we jointly use PatchGAN loss and L1 loss as our entire loss. The loss  $L_{\text{CNCL-GAN}}$  can be expressed as:

$$L_{\text{CNCL-GAN}} = L_{\text{GAN}}(G, D) + \varphi L_{\text{L1}}(G), \quad (6)$$

where  $L_{\text{GAN}}(G, D)$  indicates the GAN loss contributed by PatchGAN,  $L_{\text{L1}}(G)$  represents L1 loss, and  $\varphi$  is the loss weight of  $L_{\text{L1}}(G)$ .

$L_{\text{GAN}}(G, D)$  can be written as:

$$L_{\text{GAN}}(G, D) = \mathbb{E}_{\mathbf{x}, \mathbf{y}}[\log(D(\mathbf{x}, \mathbf{y}))] + \mathbb{E}_{\mathbf{x}, \mathbf{y}}[\log(1 - D(\mathbf{x}, G(\mathbf{x})))], \quad (7)$$

where  $\mathbb{E}[\cdot]$  denotes the expectation operator,  $G$  indicates the generator,  $D$  represents the discriminator,  $\mathbf{x}$  is the corrupted image  $I_{\text{corrupted}}$ ,  $G(\mathbf{x})$  denotes the predicted noise and the predicted content (i.e.,  $I'_{\text{noise}}$  and  $I'_{\text{content}}$ ), and  $\mathbf{y}$  is the real

noise and the real content (i.e.,  $I_{\text{noise}}$  and  $I_{\text{content}}$ ).  $D$  tries to maximize this objective, while  $G$  tries to minimize it.

The L1 loss,  $L_{\text{L1}}(G)$ , is composed of two parts:

$$L_{\text{L1}}(G) = L_{\text{L1-content}} + \lambda L_{\text{L1-noise}}, \quad (8)$$

where  $L_{\text{L1-content}}$  represents the mean absolute error between the predicted content  $I'_{\text{content}}$  and the real content  $I_{\text{content}}$ ,  $L_{\text{L1-noise}}$  refers to the mean absolute error between the predicted noise  $I'_{\text{noise}}$  and real noise  $I_{\text{noise}}$ , and  $\lambda$  is the loss weight of  $L_{\text{L1-noise}}$ .

### C. Experimental Setup

**1) Predictors:** To demonstrate the effectiveness of the proposed CNCL strategy, we investigated three kinds of predictors in the GAN-based framework. They were three representative and widely-used CNN models, including U-Net [10], DnCNN [41], and SRDenseNet [42].

Fig. 3a shows the structure of U-Net used in this work. Compared with the original version [10], we modified U-Net in two aspects: padding was employed to ensure the same feature map size before and after convolution, and batch normalization [44] was added to offer regularization. The DnCNN architecture used in this work is illustrated in Fig. 3b. In contrast to the initial DnCNN in [41], the DnCNN used

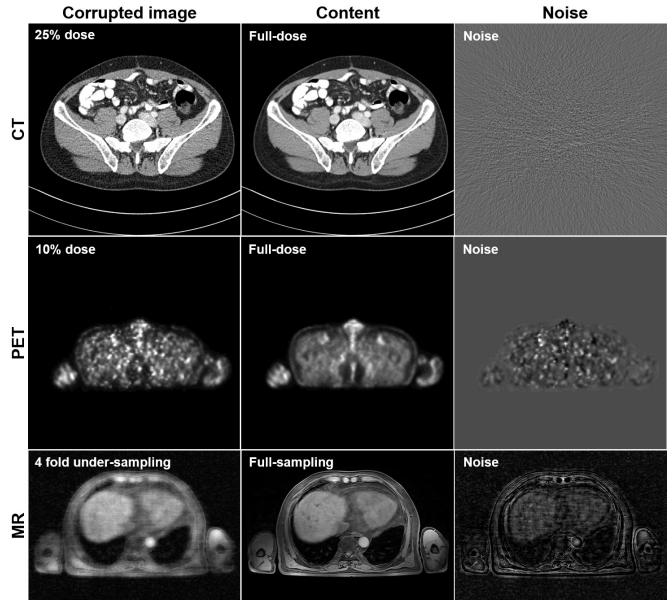
in this work discarded the residual learning, and replaced ReLU with Leaky ReLU. Fig. 3c shows the structure of SRDenseNet, and Fig. 3d presents the dense block in detail. Compared with the SRDenseNet in [42], we made the following changes: deconvolution, bottleneck, and reconstruction layers were replaced with four sets of convolution, batch normalization, and Leaky ReLU layers, to ensure the output image size was the same as the input size; the number of dense blocks and the number of the convolution layers in each dense block were both decreased from eight to six.

We named the GAN-based implementations of the proposed CNCL strategy as *CNCL-based networks*. When the predictors were two U-Nets, the CNCL-based network was abbreviated as CNCL-U-Net. Similarly, CNCL-DnCNN and CNCL-SRDenseNet represent the CNCL-based networks whose predictors were two DnCNNs and two SRDenseNets, respectively.

**2) Data Preparation:** To better evaluate the proposed denoising strategy, we investigated CNCL-U-Net, CNCL-DnCNN, and CNCL-SRDenseNet on three medical image datasets, across three different imaging modalities: CT, PET, and MR. These modalities are widely used in clinical diagnosis. The corrupted CT and the PET images were acquired under low dose, while the corrupted MR images were generated by under-sampling the  $k$ -space. Fig. 4 shows three slice examples, and it can be seen that the noise type of different imaging modalities varies greatly.

*a) CT dataset:* The proposed denoising strategy was first validated on a publicly available low-dose CT dataset, authorized for the 2016 NIH-AAPM-Mayo Clinic Low Dose CT Grand Challenge<sup>1</sup> by Mayo Clinic. In this dataset, there are 10 full-dose abdominal CT 2D slices from 10 anonymous patients, and the corresponding simulated 25% dose CT 2D slices. The full-dose data were acquired at 120 kV and 200 effective mAs, while the 25% dose data were generated by adding Poisson noise into the projection data. The dataset included 2,378 2D slice pairs ( $512 \times 512$  pixels). Referring to the data division in the previous works [14], [17], we selected 1,709 CT slice pairs from seven patients' scans for training, and 669 CT slice pairs from three patients' scans were set aside for test.

*b) PET dataset:* PET scans were conducted on 12 patients with suspicious neural endocrine tumors, with an integrated 3-Tesla PET/MR scanner (uPMR 790, United Imaging Healthcare, China). The protocol was approved by Beijing Cancer Hospital Institutional Review Board (No. 2020KT15). The amount of  $^{18}\text{F}$ -labelled tracer injected was 3.7 MBq/kg, and the PET scans were conducted after a 30-minutes post-injection. For each acquired PET scan, two reconstructions were conducted after post-processing of the raw data (list mode): one involving all acquired events, the other utilizing 10% of the total events. The reconstruction matrices for both reconstructions were set as  $144 \times 144$ , and 4,271 2D slice pairs of torso PET slices were obtained in total. These pairs were assigned as either training or test data in the six-fold validation, and in each round, slice pairs from 10 patients were



**Fig. 4.** Corrupted medical images (the leftmost column), the corresponding clean images (the middle column), and the noise (the rightmost column) in three medical imaging modalities. From the top to bottom row: CT, PET, and MR. CT display window is  $[-160, 240]$  HU.

allocated to the training set and those from the rest two to the test set.

*c) MR dataset:* MR dataset contained abdomen 2D slices of 10 patients in the same clinical trial as the PET dataset. All MR data were performed by the same PET/MR scanner as the PET data acquisition, using a 3D water-fat imaging (WFI) technique, which is a two-point Dixon sequence. Four types of images were generated for each axial slice, including water, fat, in-phase, and opposed-phase images. To acquire the paired MR slices, we first collected the full-sampled MR slices from the scanner as the content images. Then the slices were transformed to  $k$ -space using a three dimensional Fourier transform. A four-fold pseudo-random under-sampling scheme was applied to get the under-sampled  $k$ -space. Finally, the three dimensional inverse fast Fourier transform was deployed to retrieve the under-sampled MR slices. There are 6,560 MR 2D slice pairs in total ( $544 \times 384$  pixels). The MR data were investigated by a five-fold cross validation strategy. That is, when eight patients were involved in the training phase, the other two patients were used for test.

**3) Training Details:** During the training phase, all the networks were implemented in the PyTorch framework [45] with an NVIDIA GeForce RTX 2080Ti GPU. Adam served as the optimizer for all networks' training. Momentum 1 and momentum 2 were set to 0.5 and 0.999, respectively. Learning rate was 0.0002, and batch size was 2. The weights  $\lambda$  and  $\varphi$  in (8) and (6), were set to 1 and 100, respectively.  $\varphi$  setting was derived from the PatchGAN paper [43], and  $\lambda$  was determined by a range of experiments to achieve the best denoising performance. Before training, the initial weights in convolutions and batch normalization were random numbers following the normal distributions of  $N(0, 0.02)$  and  $N(1.0, 0.02)$ , respectively, and their initial biases were both set

<sup>1</sup><https://www.aapm.org/GrandChallenge/LowDoseCT/>.

**TABLE I**  
TRAINING TIME FOR THE PROPOSED CNCL-BASED NETWORKS (HOURS)

	CNCL-U-Net	CNCL-DnCNN	CNCL-SRDenseNet
CT dataset	30	24	20
PET dataset	16	12	10
MR dataset	83	66	58

to 0. The training epoch number of all networks was 300, and the training time is listed in **Table I**. The detailed parameter settings for different algorithms including reference methods are shown in Table SI in the supplementary material.

**4) Quantitative Evaluation:** To quantitatively evaluate the performance of the CNCL-based networks, the following metrics were adopted: the structural similarity index (SSIM) [46], root mean square error (RMSE) [47], and peak signal to noise ratio (PSNR) [48]. They can be defined as follows:

$$\text{SSIM}_{\mathbf{x}, \mathbf{y}} = \frac{(2\mu_{\mathbf{x}}\mu_{\mathbf{y}} + C_1)(2\text{cov}(\mathbf{x}, \mathbf{y}) + C_2)}{(\mu_{\mathbf{x}}^2 + \mu_{\mathbf{y}}^2 + C_1)(\sigma_{\mathbf{x}}^2 + \sigma_{\mathbf{y}}^2 + C_2)}, \quad (9)$$

where  $\mu_{\mathbf{x}}$  and  $\mu_{\mathbf{y}}$  are the means of  $\mathbf{x}$  and  $\mathbf{y}$ , respectively;  $\text{cov}(\mathbf{x}, \mathbf{y})$  represents the covariance of  $\mathbf{x}$  and  $\mathbf{y}$ ;  $\sigma_{\mathbf{x}}^2$  and  $\sigma_{\mathbf{y}}^2$  indicate the variances of  $\mathbf{x}$  and  $\mathbf{y}$ , respectively;  $C_1$  and  $C_2$  are constants, which are set to 6.5 and 58.5, respectively.

$$\text{RMSE}_{\mathbf{x}, \mathbf{y}} = \sqrt{\frac{1}{m} \sum_{i=1}^m (\mathbf{x}_i - \mathbf{y}_i)^2}, \quad (10)$$

where  $m$  refers to the total number of pixels in  $\mathbf{x}$  and  $\mathbf{y}$ .

$$\text{PSNR}_{\mathbf{x}, \mathbf{y}} = 20 \log_{10} \left( \frac{\text{MAX}}{\text{RMSE}_{\mathbf{x}, \mathbf{y}}} \right), \quad (11)$$

where MAX is the peak intensity of the images.

### III. RESULTS

#### A. Ablation Studies

To verify the effectiveness of learning content and noise complementarily through the CNCL strategy, we conducted ablation studies over the NIH-AAPM-Mayo CT dataset. We compared the CNCL-based networks with their corresponding baseline networks which had only one predictor. In detail, we compared CNCL-U-Net with a single U-Net predictor that only learned content, and compared CNCL-U-Net with a single U-Net predictor that only learned noise. Analogically, CNCL-DnCNN was compared with a single DnCNN predictor that only learned content, and a single DnCNN predictor that only learned noise, separately; CNCL-SRDenseNet was compared with a single SRDenseNet predictor that only learned content, and a single SRDenseNet predictor that only learned noise, separately. For a fair comparison, all of the above-mentioned networks' training was under the same GAN framework in **Fig. 2**, except the generator part.

**Table II** enumerates the quantitative results of the ablation studies. The results demonstrate the effectiveness of CNCL. Generally speaking, U-Net experienced more significant improvement through the proposed CNCL strategy than DnCNN and SRDenseNet. **Fig. 5** presents a representative

**TABLE II**  
QUANTITATIVE RESULTS (MEAN  $\pm$  SDs) OF THE ABLATION STUDY OVER THE CT DATASET

	SSIM	RMSE	PSNR (dB)
<b>CNCL-U-Net</b>	<b>0.917 <math>\pm</math> 0.032</b>	<b>0.019 <math>\pm</math> 0.003</b>	<b>34.671 <math>\pm</math> 1.553</b>
one U-Net for content learning	0.882 $\pm$ 0.041	0.027 $\pm$ 0.004	31.269 $\pm$ 1.670
one wide-U-Net for content learning	0.893 $\pm$ 0.032	0.026 $\pm$ 0.004	31.727 $\pm$ 1.587
one U-Net for noise learning	0.888 $\pm$ 0.039	0.025 $\pm$ 0.004	31.864 $\pm$ 1.607
one wide-U-Net for noise learning	0.888 $\pm$ 0.041	0.025 $\pm$ 0.004	32.127 $\pm$ 1.596
<b>CNCL-DnCNN</b>	<b>0.916 <math>\pm</math> 0.032</b>	<b>0.018 <math>\pm</math> 0.003</b>	<b>35.122 <math>\pm</math> 1.544</b>
one DnCNN for content learning	0.911 $\pm$ 0.032	0.021 $\pm$ 0.003	33.508 $\pm$ 1.123
one DnCNN for noise learning	0.910 $\pm$ 0.032	0.020 $\pm$ 0.003	33.779 $\pm$ 1.360
<b>CNCL-SRDenseNet</b>	<b>0.917 <math>\pm</math> 0.031</b>	<b>0.017 <math>\pm</math> 0.003</b>	<b>35.256 <math>\pm</math> 1.519</b>
one SRDenseNet for content learning	0.911 $\pm$ 0.032	0.021 $\pm$ 0.003	33.550 $\pm$ 1.263
one SRDenseNet for noise learning	0.888 $\pm$ 0.043	0.027 $\pm$ 0.005	31.425 $\pm$ 1.345

CNCL-U-Net, CNCL-DnCNN, and CNCL-SRDenseNet refer to the proposed CNCL-based networks whose predictors are two U-Nets, two DnCNNs, and two SRDenseNets, respectively. wide-U-Net represents the U-Net whose convolutional layers are widened by adding more feature channels. All of the networks' training in this table was under the same GAN framework in **Fig. 2**, except the generator part.

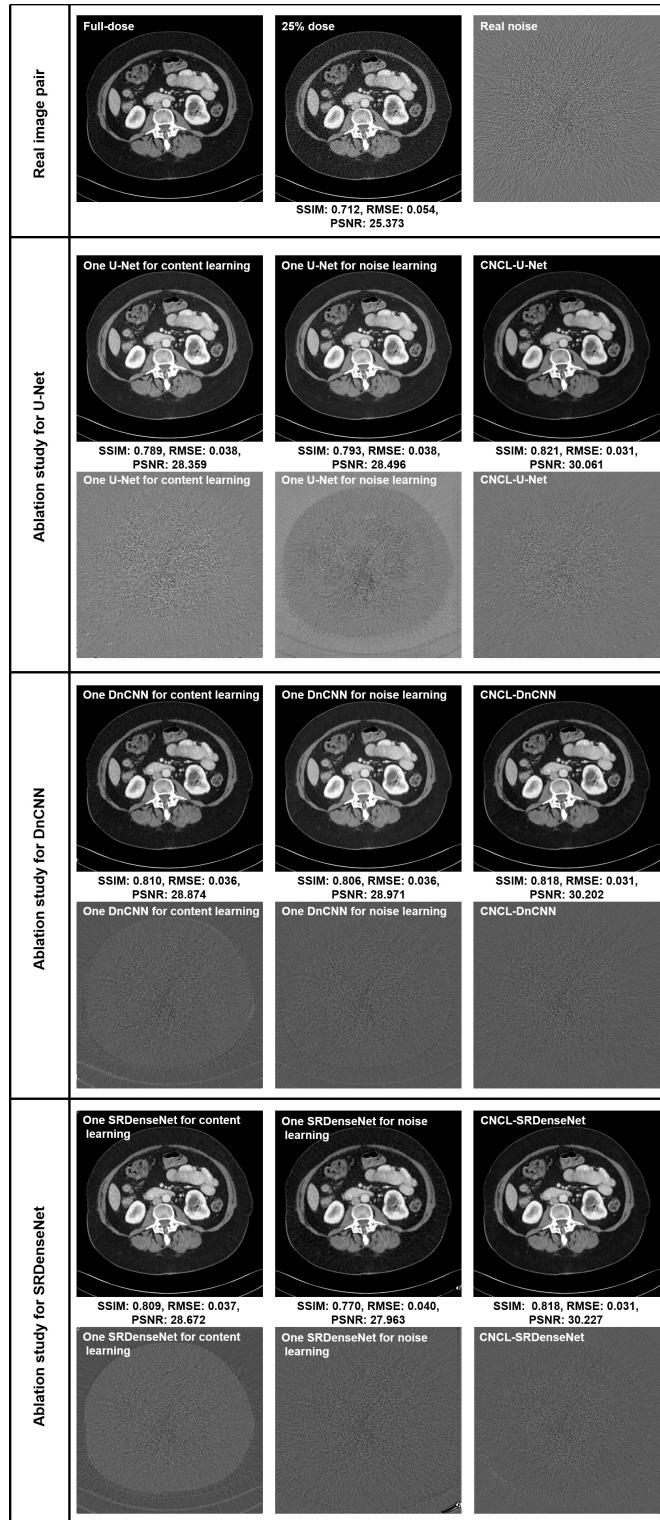
slice in the ablation studies. It can be seen that, when a single predictor was utilized to learn content or noise, the denoised outputs contained more noise residues than those of the CNCL-based networks.

CNCL caused the increase of network parameters inevitably. Two predictors were utilized through CNCL, which meant the parameters were doubled. To further verify the effectiveness of the CNCL strategy, we compared CNCL-U-Net with a single wide-U-Net predictor. The convolutional layers in wide-U-Net were widened by adding more feature channels. As a result, wide-U-Net and CNCL-U-Net had the same parameter number. For a fair comparison, the wide-U-Net's training was also under the GAN framework in **Fig. 2**. As presented in **Table II**, wide-U-Net was significantly inferior to CNCL-U-Net both in content learning and in noise learning.

In addition to the ablation studies on the network structure, we also carried out the ablation study on the loss function. Basing upon PatchGAN [43], we added a new L1 regularization term of noise into the whole loss function, as expressed in (8). In order to verify the effectiveness of the added L1 regularization term of noise, we compared two CNCL-U-Nets with different loss functions: one combining the GAN loss, the L1 loss of content, and the L1 loss of noise into the whole loss; the other adopted only the GAN loss and the L1 loss of content. The quantitative results of the two CNCL-U-Nets over the NIH-AAPM-Mayo CT dataset are presented in Table SII in the supplementary material. The results demonstrated the effectiveness of the added regularization term of noise. By combining the GAN loss, the L1 loss of content, and the L1 loss of noise into the whole loss, the content predictor and the noise predictor could be better supervised and trained, resulting in superior denoising performance.

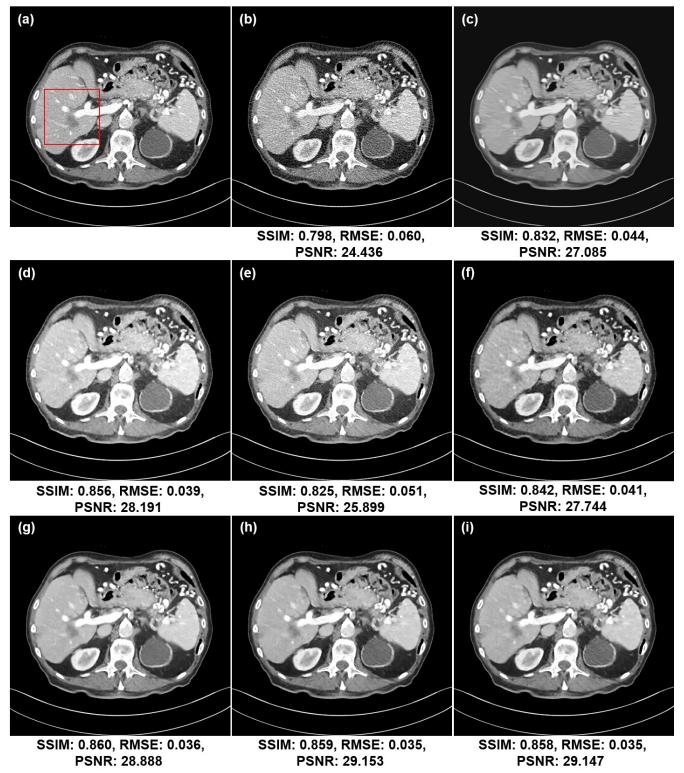
#### B. Comparison With Reference Methods

**1) Results on CT Dataset:** BM3D [8], RED-CNN [13], WGAN-VGG [18], and Q-AE [14] were selected as the reference methods in the CT denoising task. As shown in **Fig. 6**, **Fig. 7**, and **Fig. S1** in the supplementary material, two representative CT slices containing metastases, were chosen to visualize the denoising performance of different

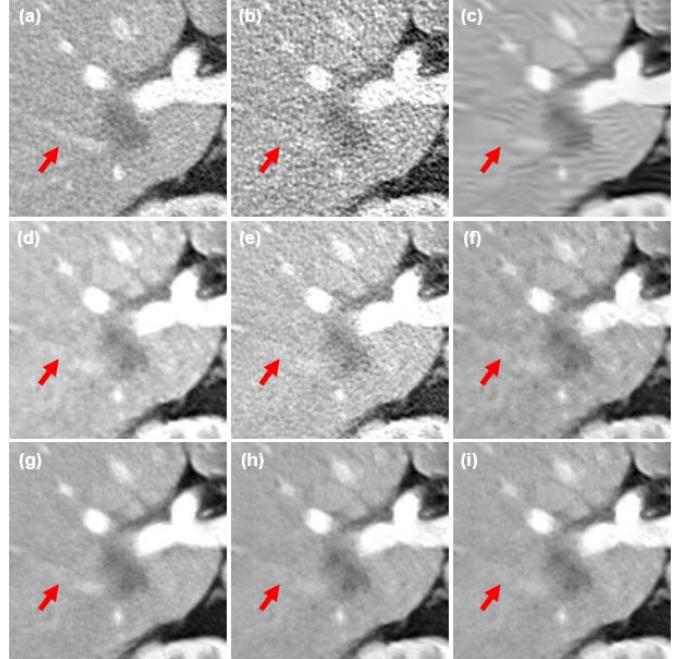


**Fig. 5.** Visual comparison in the ablation studies on case L310. The first row is the real image pair. The second and third rows are the ablation study for the situation where the predictor is U-Net. The fourth and fifth rows are the ablation study for DnCNN. The last two rows are the ablation study for SRDenseNet. Noise images are calculated by subtracting corresponding predicted content images and the real full-dose image. The CT display window is  $[-160, 240]$  HU.

methods. These two slices came from two patients, who were labeled L506 and L310, respectively. Generally speaking, all methods showed certain denoising capabilities. However,

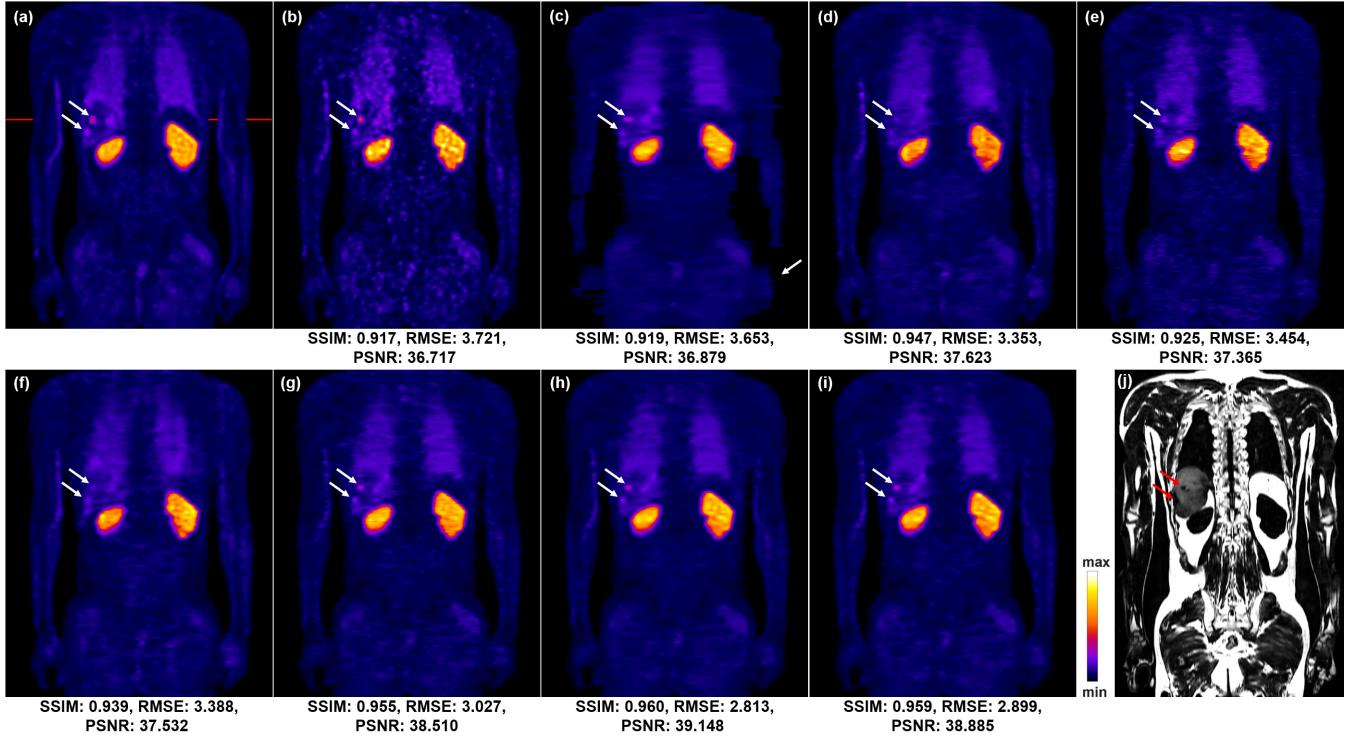


**Fig. 6.** Comparison of different methods for low-dose CT denoising on case L506. There is a metastatic lesion in the liver. (a) Full-dose CT. (b) 25% dose CT. (c) BM3D. (d) RED-CNN. (e) WGAN-VGG. (f) Q-AE. (g) CNCL-U-Net. (h) CNCL-DnCNN. (i) CNCL-SRDenseNet. The CT display window is  $[-160, 240]$  HU.



**Fig. 7.** Magnified parts from Fig. 6. (a) Full-dose CT. (b) 25% dose CT. (c) BM3D. (d) RED-CNN. (e) WGAN-VGG. (f) Q-AE. (g) CNCL-U-Net. (h) CNCL-DnCNN. (i) CNCL-SRDenseNet. The CT display window is  $[-160, 240]$  HU. The arrows point to a vessel.

BM3D blurred the low-contrast lesions, and introduced some streak artifacts in the denoised images. RED-CNN and Q-AE suffered from over-smoothness, and lost some texture



**Fig. 8.** Comparison of different methods for low-dose PET denoising in the coronal direction. (a) Full-dose PET. (b) 10% dose PET. (c) BM3D. (d) U-Net. (e) DnCNN. (f) cGAN. (g) CNCL-U-Net. (h) CNCL-DnCNN. (i) CNCL-SRDenseNet. (j) corresponding MR image. As indicated by the arrows, there are two tiny lesions.

TABLE III

QUANTITATIVE RESULTS (MEAN  $\pm$  SDs) ASSOCIATED WITH DIFFERENT ALGORITHMS FOR LOW-DOSE CT DENOISING ON COMPLETE TEST SET

	SSIM	RMSE	PSNR (dB)
25% dose CT	0.841 $\pm$ 0.063	0.033 $\pm$ 0.009	29.992 $\pm$ 2.339
BM3D [8]	0.887 $\pm$ 0.031	0.025 $\pm$ 0.003	31.879 $\pm$ 0.960
RED-CNN [13]	0.911 $\pm$ 0.032	0.024 $\pm$ 0.009	32.923 $\pm$ 2.698
WGAN-VGG [18]	0.885 $\pm$ 0.041	0.028 $\pm$ 0.011	31.512 $\pm$ 2.838
Q-AE [14]	0.913 $\pm$ 0.032	0.021 $\pm$ 0.003	33.473 $\pm$ 1.581
CNCL-U-Net	<b>0.917 <math>\pm</math> 0.032</b>	<b>0.019 <math>\pm</math> 0.003</b>	<b>34.671 <math>\pm</math> 1.553</b>
CNCL-DnCNN	<b>0.916 <math>\pm</math> 0.032</b>	<b>0.018 <math>\pm</math> 0.003</b>	<b>35.122 <math>\pm</math> 1.544</b>
CNCL-SRDenseNet	<b>0.917 <math>\pm</math> 0.031</b>	<b>0.017 <math>\pm</math> 0.003</b>	<b>35.256 <math>\pm</math> 1.519</b>

TABLE IV

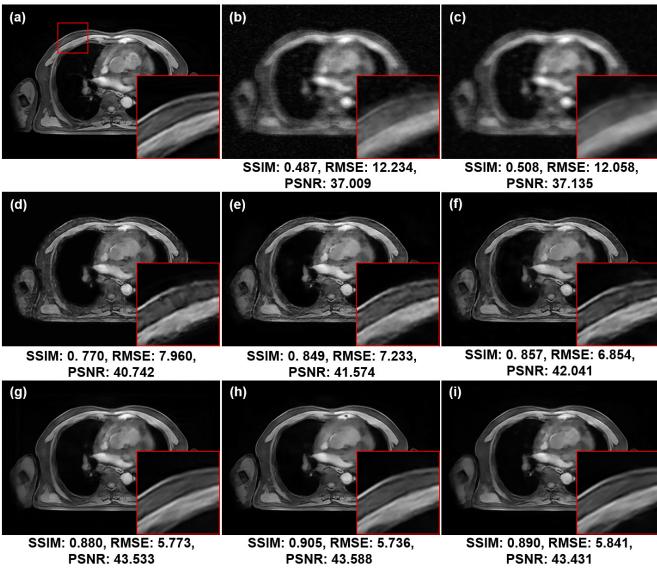
QUANTITATIVE RESULTS (MEAN  $\pm$  SDs) ASSOCIATED WITH DIFFERENT ALGORITHMS FOR LOW-DOSE PET DENOISING OF AXIAL SLICES ON THE FULL CROSS VALIDATION

	SSIM	RMSE	PSNR (dB)
10% dose PET	0.971 $\pm$ 0.018	2.308 $\pm$ 1.116	41.774 $\pm$ 3.967
BM3D [8]	0.976 $\pm$ 0.017	2.162 $\pm$ 0.814	41.929 $\pm$ 2.832
U-Net [28]	0.973 $\pm$ 0.057	1.963 $\pm$ 1.099	43.299 $\pm$ 4.096
DnCNN [27]	0.975 $\pm$ 0.015	2.155 $\pm$ 1.286	42.502 $\pm$ 4.010
cGAN [26]	0.980 $\pm$ 0.010	2.114 $\pm$ 1.184	42.751 $\pm$ 4.314
CNCL-U-Net	<b>0.984 <math>\pm</math> 0.010</b>	<b>1.786 <math>\pm</math> 1.081</b>	<b>44.292 <math>\pm</math> 4.446</b>
CNCL-DnCNN	<b>0.986 <math>\pm</math> 0.008</b>	<b>1.735 <math>\pm</math> 1.191</b>	<b>44.607 <math>\pm</math> 4.416</b>
CNCL-SRDenseNet	<b>0.986 <math>\pm</math> 0.009</b>	<b>1.759 <math>\pm</math> 1.195</b>	<b>44.500 <math>\pm</math> 4.466</b>

information, which could be attributed to the use of mean square error (MSE) as the loss function. Although the results of WGAN-VGG showed certain structural fidelity, there were a lot of noise residues. It is obvious that compared with comparison methods, the CNCL-based networks achieved better denoising performance in terms of both content preservation and noise suppression. Among the proposed CNCL-based networks, CNCL-U-Net outperformed CNCL-DnCNN and CNCL-SRDenseNet in the aspect of vascular structure preservation. Table III summarizes the quantitative results in the whole test set. This table shows that CNCL-U-Net, CNCL-DnCNN, and CNCL-SRDenseNet achieved better quantitative results than comparison methods. In general, the quantitative evaluation confirmed our visual observations.

**2) Results on PET Dataset:** BM3D [8], U-Net [28], DnCNN [27], and cGAN [26] were used as the reference methods in PET denoising task. Although U-Net, DnCNN,

and cGAN were not first proposed for PET denoising, they were modified and optimized in [26]–[28] to fit PET denoising tasks. Two representative slices containing lesions were selected for the visual comparison of PET denoising, one in the coronal direction (Fig. 8), and the other in the axial direction (Fig. S2 in the supplementary material). The full-dose images suggest the presence of two small lesions in the corresponding slices. Although U-Net, DnCNN, and cGAN reduced the noise to varying degrees, they erased the tiny lesions as noise. BM3D retained certain lesion information, but it suffered from severe blur, losing a lot of structural information. The proposed CNCL-based networks not only effectively removed noise, but also protected structural information well. After denoising, the tiny lesions could be clearly identified. Table IV presents the quantitative results on the full cross validation in terms of means  $\pm$  standard deviations. All of the proposed CNCL-based networks outperformed the comparison methods quantitatively.



**Fig. 9.** Comparison of different methods for under-sampling MR denoising in the water images of the WFI sequence. (a) Full-sampling MR. (b) Four-fold under-sampling MR. (c) NLM. (d) SCNN. (e) U-Net. (f) DnCNN. (g) CNCL-U-Net. (h) CNCL-DnCNN. (i) CNCL-SRDenseNet.

TABLE V

QUANTITATIVE RESULTS (MEAN  $\pm$  SDs) ASSOCIATED WITH DIFFERENT ALGORITHMS FOR UNDER-SAMPLING MR DENOISING OF AXIAL SLICES ON THE FULL CROSS VALIDATION

	SSIM	RMSE	PSNR (dB)
4 fold under-sampling MR	0.372 $\pm$ 0.086	23.150 $\pm$ 10.916	32.414 $\pm$ 4.182
NLM [7]	0.407 $\pm$ 0.080	21.069 $\pm$ 11.163	33.072 $\pm$ 4.512
SCNN [23]	0.706 $\pm$ 0.127	12.396 $\pm$ 8.302	37.959 $\pm$ 3.935
U-Net [20]	0.811 $\pm$ 0.049	12.614 $\pm$ 5.392	37.479 $\pm$ 3.563
DnCNN [22]	0.793 $\pm$ 0.104	10.112 $\pm$ 4.155	39.350 $\pm$ 3.429
CNCL-U-Net	<b>0.860 <math>\pm</math> 0.070</b>	<b>8.539 <math>\pm</math> 4.534</b>	<b>41.215 <math>\pm</math> 4.347</b>
CNCL-DnCNN	<b>0.863 <math>\pm</math> 0.070</b>	<b>8.452 <math>\pm</math> 4.009</b>	<b>41.008 <math>\pm</math> 3.624</b>
CNCL-SRDenseNet	<b>0.867 <math>\pm</math> 0.057</b>	<b>8.192 <math>\pm</math> 4.073</b>	<b>41.441 <math>\pm</math> 4.061</b>

**3) Results on MR Dataset:** For MR dataset, the reference methods included NLM [7], U-Net [20], DnCNN [22], and SCNN [23]. Compared with the original versions, the U-Net and DnCNN in [20], [22] were optimized to fit MR denoising tasks. The water images and the fat images of the WFI sequence were selected for visual comparison of MR denoising, as illustrated in Fig. 9 and Fig. S3 in the supplementary material. In the figures, all deep learning-based methods were able to effectively remove the noise, while NLM was not competent to the denoising task. However, SCNN introduced some artifacts. U-Net and DnCNN lost much textural information. The proposed CNCL-based networks were clearly superior to the comparison methods both in noise removal and in structural restoration. The quantitative results of the complete cross validation are provided in Table V. Compared with reference methods, the CNCL-based networks had better performance in terms of all metrics.

### C. Comparison Among CNCL-Based Networks

In the three denoising tasks above, all the proposed CNCL-based networks verified their superior capabilities of

denoising. Besides, it is notable that the images generated by the CNCL-based networks were slightly different in detail.

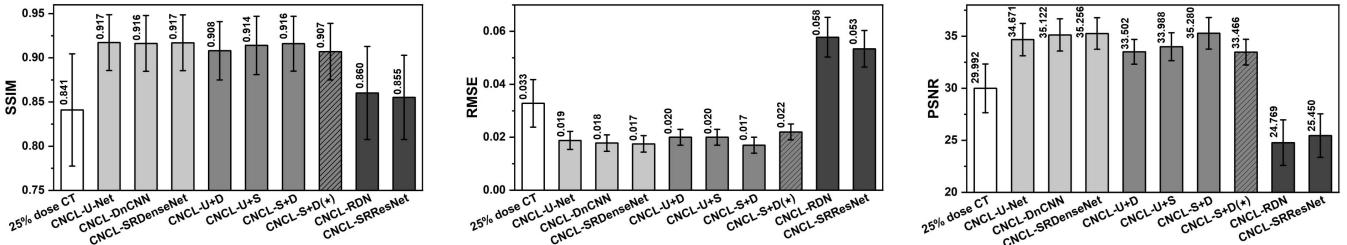
For low-dose CT denoising, although CNCL-DnCNN and CNCL-SRDenseNet obtained better RMSE and PSNR (Table III), they blurred some structural objects, like vessels pointed out by the red arrows in Fig. 6. CNCL-U-Net could restore more structural information from the corrupted image, resulting in clearer boundaries of blood vessel and lesion. In low-dose PET denoising task, the detectability of tiny lesions is one of the critical challenges. All the CNCL-based networks were able to conserve the tiny lesions to various extents, while CNCL-DnCNN and CNCL-SRDenseNet maintained higher intensity in the lesion areas (Fig. S2 in the supplementary material). As for under-sampling MR denoising, according to Table V, CNCL-DnCNN achieved the best quantitative results. In terms of image vision, all of them produced satisfactory visual quality of MR denoising. In general, predictor selection is task-based and is different depending on the desired application.

## IV. DISCUSSIONS

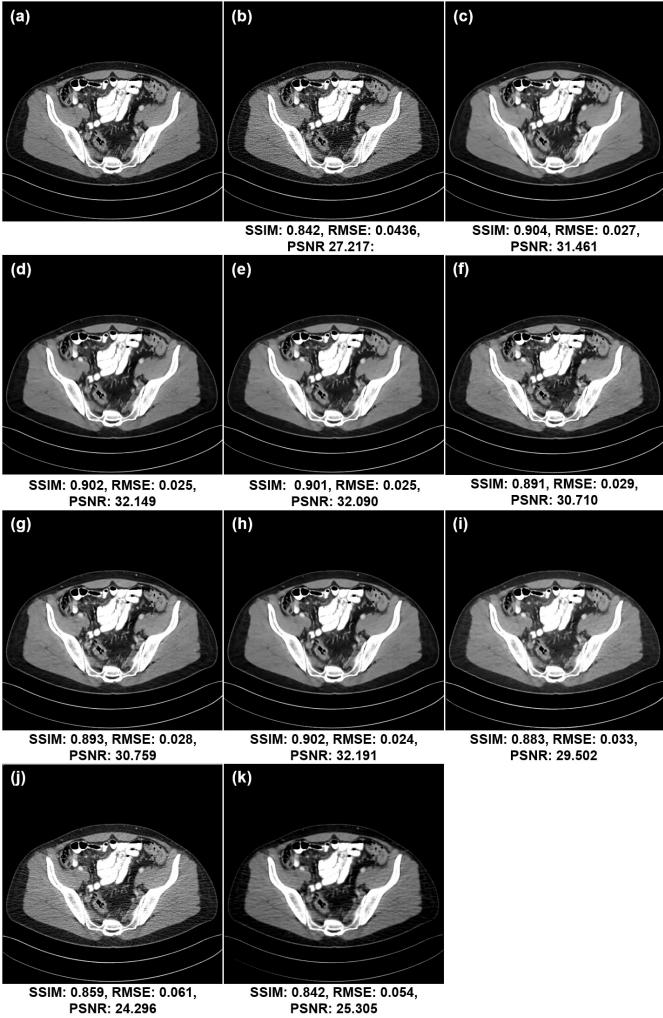
### A. Predictor Selection

The adopted content predictor and noise predictor can be homogeneous (with the same architecture) or heterogeneous (with different architectures). Here, we investigated the denoising performance of three CNCL-based networks with heterogeneous predictors under the GAN framework of Fig. 2, over the NIH-AAPM-Mayo CT dataset. Specifically, the first one adopted U-Net and DnCNN as the content predictor and the noise predictor respectively (namely CNCL-U+D); the second one utilized U-Net and SRDenseNet as the content predictor and the noise predictor respectively (namely CNCL-U+S); the last one applied SRDenseNet and DnCNN as the content predictor and the noise predictor respectively (namely CNCL-S+D). The quantitative results are shown in Fig. 10, and a representative slice was chosen for visual comparison, as illustrated in Fig. 11. In general, the CNCL-based networks with heterogeneous predictors achieved comparable denoising performance with the ones with homogeneous predictors. CNCL-U+D and CNCL-U+S were slightly inferior to CNCL-U-Net, CNCL-DnCNN, and CNCL-SRDenseNet, while CNCL-S+D outperformed them. The results revealed that the predictor selection should be case-dependent.

When the predictors are homogeneous, the selected predictor is expected to own both the capability of learning noise, and the capability of learning content, which is not met by all CNN models. When a CNN model had owned the capability of learning content (such as U-Net, DnCNN, and SRDenseNet, etc.), it could own the capability of learning noise by replacing the label from content to noise. However, some CNN models (such as RDN [49] and SRResNet [50], etc.) might not be suitable for learning content, due to the presence of internal residual layers. The internal residual layers were designed to help the network extract noise feature better. When the label was content, these residual layers might interfere the feature extraction of content. Thus, we also tried RDN and SRResNet as the predictors under



**Fig. 10.** Quantitative comparison in denoising performance of different predictor selections under the proposed strategy over the CT dataset. CNCL-U-Net, CNCL-DnCNN, CNCL-SRDenseNet, CNCL-RDN, and CNCL-SRResNet adopted two homogeneous CNNs to predict content and noise respectively. CNCL-U+D, CNCL-U+S, and CNCL-S+D utilized two heterogeneous CNNs to predict content and noise respectively. CNCL-S+D(\*) employed two heterogeneous CNNs which both predicted noise. All of them were trained under the GAN framework in Fig. 2.



**Fig. 11.** Visual comparison in the predictor selection studies over the CT dataset. (a) Full-dose CT. (b) 25% dose CT. (c) CNCL-U-Net. (d) CNCL-DnCNN. (e) CNCL-SRDenseNet. (f) CNCL-U+D. (g) CNCL-U+S. (h) CNCL-S+D. (i) CNCL-S+D(\*). (j) CNCL-RDN. (k) CNCL-SRResNet. The CT display window is  $[-160, 240]$  HU.

the GAN implementation (Fig. 2) of the CNCL strategy over the NIH-AAPM-Mayo CT dataset (namely CNCL-RDN and CNCL-SRResNet, respectively). The quantitative results are presented in Fig. 10, and the visual comparison is illustrated in Fig. 11. It can be seen that, compared with baseline (i.e., 25% dose CT), CNCL-RDN and CNCL-SRResNet had a

slight improvement in SSIM, while CNCL-U-Net, CNCL-DnCNN, and CNCL-SRDenseNet improved SSIM significantly. CNCL-RDN and CNCL-SRResNet even achieved worse RMSE and PSNR results than baseline, whereas CNCL-U-Net, CNCL-DnCNN, and CNCL-SRDenseNet optimized RMSE and PSNR greatly.

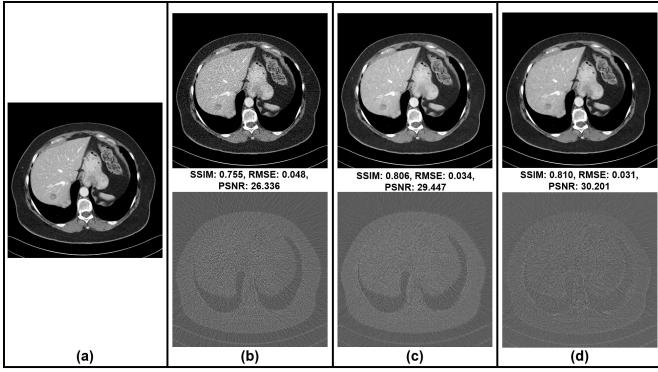
To further verify the effectiveness of learning content and noise complementarily for the CNCL strategy, we compared CNCL-S+D with another heterogeneous network which utilized DnCNN as a noise predictor and used SRDenseNet as the other noise predictor (namely CNCL-S+D(\*)). The quantitative and qualitative results are shown in Fig. 10 and Fig. 11, respectively. It can be seen that CNCL-S+D outperformed CNCL-S+D(\*) both quantitatively and qualitatively. The results verified the effectiveness of the proposed CNCL strategy in terms of learning content and noise complementarily.

## B. Universality Studies

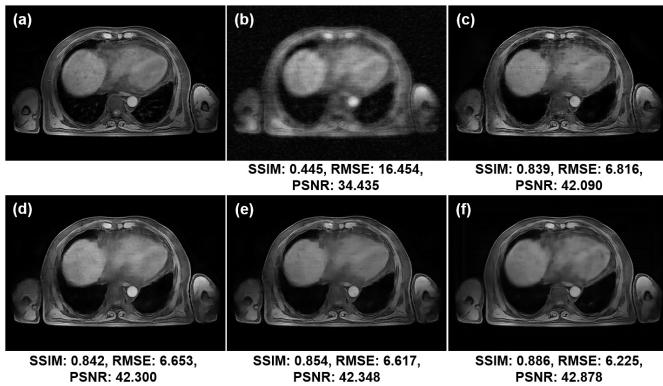
**1) Extension to WGAN-VGG:** To further verify the applicability of our proposed strategy, we extended it to a WGAN-VGG framework [18], and the NIH-AAPM-Mayo CT dataset was used for training and validation. As shown in Fig. S4 in the supplementary material, we modified the generator part of WGAN-VGG through CNCL (namely CNCL-WGAN-VGG): an extra CNN with the same structure was added. The original CNN learns content, and the added CNN learns noise.

To compare the denoising performance of WGAN-VGG and CNCL-WGAN-VGG, one representative slice is visualized in Fig. 12. It can be observed that the result of CNCL-WGAN-VGG had fewer noise residues. In terms of quantitative metrics, compared with WGAN-VGG, CNCL-WGAN-VGG increased SSIM from 0.885 to 0.891, decreased RMSE from 0.028 to 0.025, and improved PSNR from 31.512 dB to 32.103 dB, on the complete test set. This result demonstrated that our CNCL strategy is not only adapted to the framework in Fig. 2, but also has the potential to be applied to other frameworks.

**2) Extension to Non-GAN Framework:** We also verified the effectiveness of our strategy based on the non-GAN framework. We modified U-Net according to the CNCL strategy, and used only MSE as the loss function (namely CNCL-U-Net(\*)). The MR dataset was used for training and validation.



**Fig. 12.** Comparison before and after using the CNCL strategy in WGAN-VGG. (a) Full-dose CT. (b) 25% dose CT. (c) Original WGAN-VGG [18]. (d) CNCL-WGAN-VGG. The bottom row is the noise images. The CT display window is  $[-160, 240]$  HU.



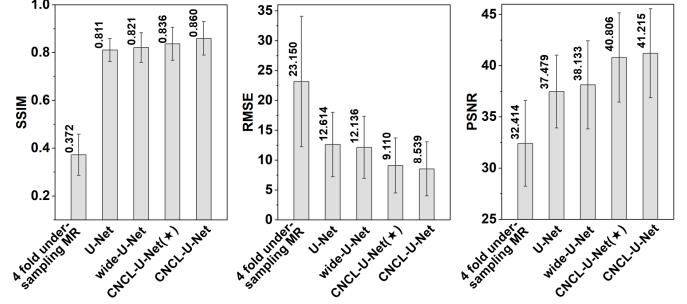
**Fig. 13.** Denoising performance comparison among (a) Full-sampling MR, (b) 4 fold under-sampling MR, (c) U-Net, (d) wide-U-Net, (e) CNCL-U-Net(\*), and (f) CNCL-U-Net. U-Net, wide-U-Net, and CNCL-U-Net(\*) were trained in non-GAN manners, while CNCL-U-Net was trained under the GAN framework in Fig. 2.

As demonstrated in Fig. 13, the denoising performance of CNCL-U-Net(\*) was better than those of U-Net and wide-U-Net, but worse than that of CNCL-U-Net. Note that, U-Net, wide-U-Net, and CNCL-U-Net(\*) were trained with MSE loss, while CNCL-U-Net was trained under the GAN framework.

Fig. 14 presents the quantitative results, which can be ranked in the following order: CNCL-U-Net > CNCL-U-Net(\*) > wide-U-Net > U-Net. From the results, it is revealed that the CNCL strategy is also applicable to non-GAN framework, and introducing a discriminator during training could improve the denoising performance.

**3) Extension to Natural Image Dataset:** We also explored the denoising capability of CNCL on a natural image dataset, SIDD [40]. SIDD was created from ten scenes under different lighting conditions, using five representative smartphone cameras. In this work, we used a small SIDD Dataset, named SIDD-Small.<sup>2</sup> The used SIDD-Small dataset contained 1,024 image pairs ( $256 \times 256$  pixels). We selected 704 image pairs for training, and 320 image pairs were used for the test. BM3D, NLM, U-Net, DnCNN, SRDenseNet, and Tyan-Net [51] were chosen as the reference methods in the SIDD

<sup>2</sup><https://www.eecs.yorku.ca/~kamel/sidd/dataset.php/>.



**Fig. 14.** Quantitative results of U-Net, wide-U-Net, CNCL-U-Net(\*), and CNCL-U-Net over the MR dataset. U-Net, wide-U-Net, and CNCL-U-Net(\*) were trained in non-GAN manners, while CNCL-U-Net was trained under the GAN framework.

**TABLE VI**  
QUANTITATIVE RESULTS (MEAN  $\pm$  SDs) ASSOCIATED WITH DIFFERENT ALGORITHMS FOR SIDD DENOISING ON THE COMPLETE TEST SET

	SSIM	RMSE	PSNR (dB)
Before denoising	$0.436 \pm 0.168$	$12.862 \pm 6.011$	$26.756 \pm 3.675$
BM3D	$0.863 \pm 0.097$	$7.288 \pm 3.613$	$31.987 \pm 4.520$
NLM	$0.681 \pm 0.181$	$7.888 \pm 4.049$	$31.246 \pm 4.289$
U-Net	$0.911 \pm 0.048$	$4.404 \pm 2.008$	$35.937 \pm 3.287$
DnCNN	$0.636 \pm 0.183$	$9.716 \pm 5.056$	$29.491 \pm 4.473$
SRDenseNet	$0.898 \pm 0.064$	$4.270 \pm 2.006$	$36.342 \pm 3.716$
Tyan-Net	$0.924 \pm 0.053$	$3.943 \pm 1.905$	$37.036 \pm 3.670$
CNCL-U-Net	$0.959 \pm 0.041$	$3.116 \pm 1.917$	$39.560 \pm 4.665$
CNCL-DnCNN	$0.928 \pm 0.050$	$3.809 \pm 1.912$	$37.444 \pm 3.968$
CNCL-SRDenseNet	$0.937 \pm 0.043$	$3.703 \pm 1.914$	$37.729 \pm 4.035$

denoising task. Fig. S5 in the supplementary material shows images denoised by different methods in the SIDD-Small dataset. We can observe that BM3D and Tyan-Net blurred the image to various extents; NLM, DnCNN, and SRDenseNet suffered from some degree of noise residues; U-Net caused a color cast problem; whereas the CNCL-based networks eliminated the noise, and retained the texture information. Furthermore, CNCL-U-Net outperformed CNCL-DnCNN and CNCL-SRDenseNet in terms of detail retention, which was consistent with our observations in low-dose CT denoising. Table VI summarizes the quantitative metrics on the complete test set. The quantitative results supported our visual observations.

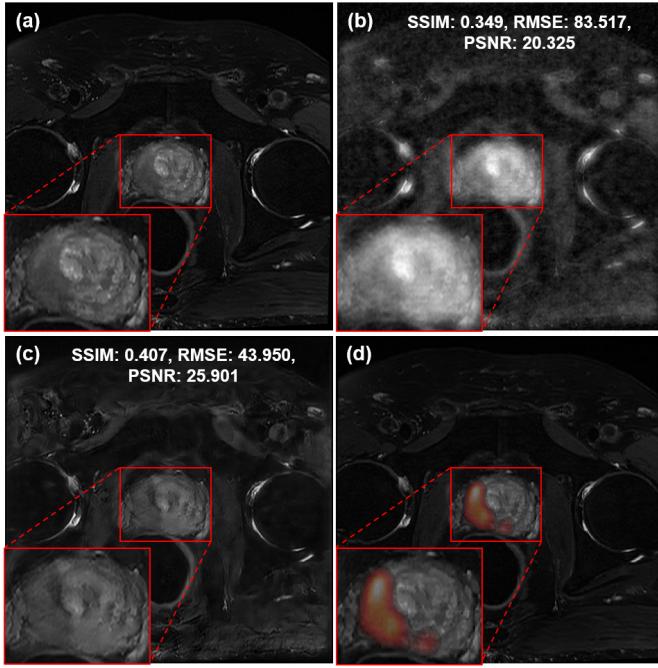
### C. Generalization Studies

In practice, the acquired medical images tend to suffer from the domain shift problem (i.e., distribution difference) [52], as a result of different scanning protocols, devices, and patient populations, etc. Therefore, it is necessary to evaluate the generalization capability of a trained deep learning model. In this section, we investigated the generalization capability of the proposed CNCL-based networks by two preliminary cross-domain tests.

First, we applied the CNCL-based networks trained on the NIH-AAPM-Mayo CT dataset to a piglet CT dataset [16]. The full-dose and 25% dose CT images in the piglet data were obtained from actual scans. Besides, the two datasets vary in terms of the scanner model, tube voltage, tube current, and scanning object. Given that Q-AE achieved the best performance among reference methods over the NIH-AAPM-Mayo

**TABLE VII**  
QUANTITATIVE RESULTS (MEAN  $\pm$  SDs) OF THE PRELIMINARY CROSS-DOMAIN TESTS ON THE PIGLET DATASET WITH DIFFERENT ALGORITHMS

	SSIM	RMSE	PSNR (dB)
Before denoising	$0.839 \pm 0.025$	$0.052 \pm 0.010$	$25.754 \pm 2.968$
Q-AE	$0.891 \pm 0.054$	$0.045 \pm 0.031$	$26.943 \pm 3.144$
CNCL-U-Net	<b><math>0.936 \pm 0.024</math></b>	<b><math>0.040 \pm 0.018</math></b>	<b><math>27.793 \pm 3.013</math></b>
CNCL-DnCNN	$0.927 \pm 0.013$	$0.042 \pm 0.031$	$27.789 \pm 2.948$
CNCL-SRDenseNet	$0.931 \pm 0.024$	$0.042 \pm 0.015$	$27.676 \pm 2.980$



**Fig. 15.** Comparison of lesion display before and after CNCL-U-Net denoising in a prostate T2 MR slice. (a) Full-sampling MR. (b) 4 fold under-sampling MR before denoising. (c) After CNCL-U-Net denoising. (d) PET image overlayed upon the MR image.

CT dataset, we compared the CNCL-based networks with Q-AE here. The quantitative and qualitative results are shown in Table VII and Fig. S6 in the supplementary material, respectively. Obviously, the CNCL-based networks generated fewer noise residues than Q-AE.

Prostate MR data (approved by the Institutional Review Board of Peking University Cancer Hospital, No. 2017KT94) have also been denoised using the CNCL-U-Net trained on the abdomen MR dataset, to confirm the generalization capability of the CNCL strategy. The Prostate MR data were obtained on the same PET/MR scanner as the abdomen MR data. However, for the purpose of specific lesion diagnosis, the MR sequence of the prostate images was T2WI, instead of WFI in the training dataset. However, this data distribution difference caused no significant issues in the prostate MR denoising validation, as shown in Fig. 15. Quantitatively, CNCL-U-Net increased SSIM from 0.405 to 0.494, decreased RMSE from 50.533 to 41.955, and improved PSNR from 25.445 dB to 26.505 dB, on the complete prostate MR test set. Through the denoising model, the image quality of the MR images was drastically enhanced, and the structures and intensity were restored with fidelity. A clinician could thus confidently

assert the malignancy with decreased T2 signal on the right peripheral zone, which was also verified by the merged PET image. The superior generalization ability of CNCL could facilitate its future clinical translation.

These results demonstrate that the proposed CNCL strategy is expected to equip with good generalization capability to the “unseen” medical images, and has potential toward clinical applications.

## V. CONCLUSION

In this work, we proposed a simple yet effective CNCL strategy for medical image denoising, in which two predictors were used to learn the content and the noise of the image dataset complementarily. The implemented CNCL-based networks were verified over three medical image datasets (CT, MR, and PET). The results demonstrated that the CNCL-based networks outperformed state-of-the-art methods both qualitatively and quantitatively. Furthermore, the universality and generalization capability of the CNCL strategy were also investigated. In summary, the proposed CNCL strategy has demonstrated its capability in medical image denoising, showing certain potential toward clinical applications.

## REFERENCES

- [1] S. V. Mohd Sagheer and S. N. George, “A review on medical image denoising algorithms,” *Biomed. Signal Process. Control*, vol. 61, Aug. 2020, Art. no. 102036.
- [2] F. E. Turkheimer, N. Boussion, A. N. Anderson, N. Pavese, P. Piccini, and D. Visvikis, “PET image denoising using a synergistic multiresolution analysis of structural (MRI/CT) and functional datasets,” *J. Nucl. Med.*, vol. 49, no. 4, pp. 657–666, 2008.
- [3] M. H. Al-Mallah, A. Aljzeeri, M. Alharthi, and A. Alsaileek, “Routine low-radiation-dose coronary computed tomography angiography,” *Eur. Heart J. Suppl.*, vol. 16, pp. B12–B16, Nov. 2014.
- [4] K. J. Murphy and J. A. Brunberg, “Adult claustrophobia, anxiety and sedation in MRI,” *Magn. Reson. Imag.*, vol. 15, no. 1, pp. 51–54, Jan. 1997.
- [5] D. L. Donoho, “Compressed sensing,” *IEEE Trans. Inf. Theory*, vol. 52, no. 4, pp. 1289–1306, Apr. 2006.
- [6] S. D. Voss, G. H. Reaman, S. C. Kaste, and T. L. Slovis, “The ALARA concept in pediatric oncology,” *Pediatric Radiol.*, vol. 39, no. 11, p. 1142, 2009.
- [7] J. V. Manjón, J. Carbonell-Caballero, J. J. Lull, G. García-Martí, L. Martí-Bonmatí, and M. Robles, “MRI denoising using non-local means,” *Med. Image Anal.*, vol. 12, no. 4, pp. 514–523, 2008.
- [8] P. F. Feruglio, C. Vinegoni, J. Gros, A. Sbarbati, and R. Weissleder, “Block matching 3D random noise filtering for absorption optical projection tomography,” *Phys. Med. Biol.*, vol. 55, no. 18, p. 5401, 2010.
- [9] A. M. Mendrik, E.-J. Vönken, A. Rutten, M. A. Viergever, and B. van Ginneken, “Noise reduction in computed tomography scans using 3-D anisotropic hybrid diffusion with continuous switch,” *IEEE Trans. Med. Imag.*, vol. 28, no. 10, pp. 1585–1594, Oct. 2009.
- [10] O. Ronneberger, P. Fischer, and T. Brox, “U-Net: Convolutional networks for biomedical image segmentation,” in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.* Cham, Switzerland: Springer, 2015, pp. 234–241.
- [11] Y. Lu *et al.*, “A learning-based material decomposition pipeline for multi-energy X-ray imaging,” *Med. Phys.*, vol. 46, no. 2, pp. 689–703, Feb. 2019.
- [12] M. Geng *et al.*, “PMS-GAN: Parallel multi-stream generative adversarial network for multi-material decomposition in spectral computed tomography,” *IEEE Trans. Med. Imag.*, vol. 40, no. 2, pp. 571–584, Feb. 2021.
- [13] H. Chen *et al.*, “Low-dose CT with a residual encoder-decoder convolutional neural network,” *IEEE Trans. Image Process.*, vol. 36, no. 12, pp. 2524–2535, Dec. 2017.
- [14] F. Fan *et al.*, “Quadratic autoencoder (Q-AE) for low-dose CT denoising,” *IEEE Trans. Med. Imag.*, vol. 39, no. 6, pp. 2035–2050, Jun. 2020.
- [15] H. Shan *et al.*, “3-D convolutional encoder-decoder network for low-dose CT via transfer learning from a 2-D trained network,” *IEEE Trans. Med. Imag.*, vol. 37, no. 6, pp. 1522–1534, Jun. 2018.

- [16] X. Yi and P. Babyn, "Sharpness-aware low-dose CT denoising using conditional generative adversarial network," *J. Digit. Imag.*, vol. 31, no. 5, pp. 655–669, Oct. 2018.
- [17] M. Li, W. Hsu, X. Xie, J. Cong, and W. Gao, "SACNN: Self-attention convolutional neural network for low-dose CT denoising with self-supervised perceptual loss network," *IEEE Trans. Med. Imag.*, vol. 39, no. 7, pp. 2289–2301, Jul. 2020.
- [18] Q. Yang *et al.*, "Low-dose CT image denoising using a generative adversarial network with wasserstein distance and perceptual loss," *IEEE Trans. Med. Imag.*, vol. 37, no. 6, pp. 1348–1357, Jun. 2018.
- [19] D. Lee, J. Yoo, and J. C. Ye, "Deep artifact learning for compressed sensing and parallel MRI," 2017, *arXiv:1703.01120*. [Online]. Available: <http://arxiv.org/abs/1703.01120>
- [20] Y. Han, J. Yoo, H. H. Kim, H. J. Shin, K. Sung, and J. C. Ye, "Deep learning with domain adaptation for accelerated projection-reconstruction MR," *Magn. Reson. Med.*, vol. 80, no. 3, pp. 1189–1205, Sep. 2018.
- [21] C. M. Hyun, H. P. Kim, S. M. Lee, S. Lee, and J. K. Seo, "Deep learning for undersampled MRI reconstruction," *Phys. Med. Biol.*, vol. 63, no. 13, Jun. 2018, Art. no. 135007.
- [22] D. Jiang, W. Dou, L. Vosters, X. Xu, Y. Sun, and T. Tan, "Denoising of 3D magnetic resonance images with multi-channel residual learning of convolutional neural network," *Jpn. J. Radiol.*, vol. 36, no. 9, pp. 566–574, Sep. 2018.
- [23] M. Kidoh *et al.*, "Deep learning based noise reduction for brain mr imaging: Tests on phantoms and healthy volunteers," *Magn. Reson. Med. Sci.*, vol. 19, no. 3, p. 195, 2020.
- [24] L. Xiang *et al.*, "Deep auto-context convolutional neural networks for standard-dose PET image estimation from low-dose PET/MRI," *Neurocomputing*, vol. 267, pp. 406–416, Dec. 2017.
- [25] A. Sano, T. Nishio, T. Masuda, and K. Karasawa, "Denoising PET images for proton therapy using a residual U-Net," *Biomed. Phys. Eng. Exp.*, vol. 7, no. 2, Mar. 2021, Art. no. 025014.
- [26] Y. Wang *et al.*, "3D conditional generative adversarial networks for high-quality PET image estimation at low dose," *NeuroImage*, vol. 174, pp. 550–562, Jul. 2018.
- [27] K. Kim *et al.*, "Penalized PET reconstruction using deep learning prior and local linear fitting," *IEEE Trans. Med. Imag.*, vol. 37, no. 6, pp. 1478–1487, Jun. 2018.
- [28] W. Lu *et al.*, "An investigation of quantitative accuracy for deep learning based denoising in oncological PET," *Phys. Med. Biol.*, vol. 64, no. 16, Aug. 2019, Art. no. 165019.
- [29] B. Zhou, Y.-J. Tsai, X. Chen, J. S. Duncan, and C. Liu, "MDPET: A unified motion correction and denoising adversarial network for low-dose gated PET," *IEEE Trans. Med. Imag.*, early access, Apr. 28, 2021, doi: [10.1109/TMI.2021.3076191](https://doi.org/10.1109/TMI.2021.3076191).
- [30] F. Hashimoto, H. Ohba, K. Ote, A. Kakimoto, H. Tsukada, and Y. Ouchi, "4D deep image prior: Dynamic PET image denoising using an unsupervised four-dimensional branch convolutional neural network," *Phys. Med. Biol.*, vol. 66, no. 1, Jan. 2021, Art. no. 015006.
- [31] H. Sun, L. Peng, H. Zhang, Y. He, S. Cao, and L. Lu, "Dynamic PET image denoising using deep image prior combined with regularization by denoising," *IEEE Access*, vol. 9, pp. 52378–52392, 2021.
- [32] Y. Lei *et al.*, "Whole-body PET estimation from low count statistics using cycle-consistent generative adversarial networks," *Phys. Med. Biol.*, vol. 64, no. 21, Nov. 2019, Art. no. 215017.
- [33] L. Zhou, J. D. Schaefferkoetter, I. W. K. Tham, G. Huang, and J. Yan, "Supervised learning with cyclegan for low-dose FDG PET image denoising," *Med. Image Anal.*, vol. 65, Oct. 2020, Art. no. 101770.
- [34] K. Zhao *et al.*, "Study of low-dose PET image recovery using supervised learning with CycleGAN," *PLoS ONE*, vol. 15, no. 9, Sep. 2020, Art. no. e0238455.
- [35] Y. Gong *et al.*, "Parameter-transferred wasserstein generative adversarial network (PT-WGAN) for low-dose PET image denoising," *IEEE Trans. Radiat. Plasma Med. Sci.*, vol. 5, no. 2, pp. 213–223, Mar. 2021.
- [36] H. Chen *et al.*, "Low-dose CT via convolutional neural network," *Biomed. Opt. Exp.*, vol. 8, no. 2, pp. 679–694, 2017.
- [37] K. Isogawa, T. Ida, T. Shiodera, and T. Takeguchi, "Deep shrinkage convolutional neural network for adaptive noise reduction," *IEEE Signal Process. Lett.*, vol. 25, no. 2, pp. 224–228, Feb. 2018.
- [38] S. Izadi, Z. Mirikhraji, M. Zhao, and G. Hamarneh, "WhiteNNer-blind image denoising via noise whiteness priors," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. Workshop (ICCVW)*, Oct. 2019, pp. 476–484.
- [39] H. Liao, W.-A. Lin, S. K. Zhou, and J. Luo, "ADN: Artifact disentanglement network for unsupervised metal artifact reduction," *IEEE Trans. Med. Imag.*, vol. 39, no. 3, pp. 634–643, Aug. 2019.
- [40] A. Abdelhamed, S. Lin, and M. S. Brown, "A high-quality denoising dataset for smartphone cameras," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 1692–1700.
- [41] K. Zhang, W. Zuo, Y. Chen, D. Meng, and L. Zhang, "Beyond a Gaussian Denoiser: Residual learning of deep CNN for image denoising," *IEEE Trans. Image Process.*, vol. 26, no. 7, pp. 3142–3155, Jul. 2017.
- [42] T. Tong, G. Li, X. Liu, and Q. Gao, "Image super-resolution using dense skip connections," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Oct. 2017, pp. 4799–4807.
- [43] P. Isola, J.-Y. Zhu, T. Zhou, and A. A. Efros, "Image-to-image translation with conditional adversarial networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jul. 2017, pp. 1125–1134.
- [44] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," in *Proc. Int. Conf. Mach. Learn.*, 2015, pp. 448–456.
- [45] A. Paszke *et al.*, "Pytorch: An imperative style, high-performance deep learning library," in *Proc. Adv. Neural Inf. Process. Syst.*, 2019, pp. 8026–8037.
- [46] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: From error visibility to structural similarity," *IEEE Trans. Image Process.*, vol. 13, no. 4, pp. 600–612, Apr. 2004.
- [47] T. Chai and R. Draxler, "Root mean square error (RMSE) or mean absolute error (MAE)," *Geosci. Model Develop. Discuss.*, vol. 7, no. 1, pp. 1525–1534, 2014.
- [48] Q. Huynh-Thu and M. Ghanbari, "Scope of validity of PSNR in image/video quality assessment," *Electron. Lett.*, vol. 44, no. 13, pp. 800–801, Jun. 2008.
- [49] Y. Zhang, Y. Tian, Y. Kong, B. Zhong, and Y. Fu, "Residual dense network for image super-resolution," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 2472–2481.
- [50] C. Ledig *et al.*, "Photo-realistic single image super-resolution using a generative adversarial network," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 4681–4690.
- [51] A. Abdelhamed, M. Afifi, R. Timofte, and M. S. Brown, "NTIRE 2020 challenge on real image denoising: Dataset, methods and results," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. Workshops*, Jun. 2020, pp. 496–497.
- [52] E. A. AlBadawy, A. Saha, and M. A. Mazurowski, "Deep learning for segmentation of brain tumors: Impact of cross-institutional training and testing," *Med. Phys.*, vol. 45, no. 3, pp. 1150–1158, Mar. 2018.