



Segmentation-guided Denoising Network for Low-dose CT Imaging

Zhenxing Huang^{a,1}, Zhou Liu^{b,1}, Pin He^b, Ya Ren^b, Shuluan Li^c, Yuanyuan Lei^d, Dehong Luo^b, Dong Liang^a, Dan Shao^e, Zhanli Hu^a, Na Zhang^{a,*}



^a Lauterbur Research Center for Biomedical Imaging, Shenzhen Institute of Advanced Technology, Chinese Academy of Sciences, Shenzhen 518055, China

^b Department of Radiology, National Cancer Center/National Clinical Research Center for Cancer/Cancer Hospital & Shenzhen Hospital, Chinese Academy of Medical Sciences and Peking Union Medical College, Shenzhen, 518116, China

^c Department of Medical Oncology, National Cancer Center/National Clinical Research Center for Cancer/Cancer Hospital & Shenzhen Hospital, Chinese Academy of Medical Sciences and Peking Union Medical College, Shenzhen, 518116, China

^d Department of Pathology, National Cancer Center/National Clinical Research Center for Cancer/Cancer Hospital & Shenzhen Hospital, Chinese Academy of Medical Sciences and Peking Union Medical College, Shenzhen, 518116, China

^e Department of Nuclear Medicine, Guangdong Provincial People's Hospital, Guangdong Academy of Medical Sciences, Guangzhou 510080, China

ARTICLE INFO

Article history:

Received 6 July 2022

Revised 6 October 2022

Accepted 21 October 2022

MSC:

41A05

41A10

65D05

65D17

Keywords:

Low-dose CT imaging

image denoising

structure semantic segmentation

ABSTRACT

Background: To reduce radiation exposure and improve diagnosis in low-dose computed tomography, several deep learning (DL)-based image denoising methods have been proposed to suppress noise and artifacts over the past few years. However, most of them seek an objective data distribution approximating the gold standard and neglect structural semantic preservation. Moreover, the numerical response in CT images presents substantial regional anatomical differences among tissues in terms of X-ray absorbency. **Methods:** In this paper, we introduce structural semantic information for low-dose CT imaging. First, the regional segmentation prior to low-dose CT can guide the denoising process. Second the structural semantical results can be considered as evaluation metrics on the estimated normal-dose CT images. Then, a semantic feature transform is engaged to combine the semantic and image features on a semantic fusion module. In addition, the structural semantic loss function is introduced to measure the segmentation difference. **Results:** Experiments are conducted on clinical abdomen data obtained from a clinical hospital, and the semantic labels consist of subcutaneous fat, muscle and visceral fat associated with body physical evaluation. Compared with other DL-based methods, the proposed method achieves better performance on quantitative metrics and better semantic evaluation. **Conclusions:** The quantitative experimental results demonstrate the promising performance of the proposed methods in noise reduction and structural semantic preservation. While, the proposed method may suffer from several limitations on abnormalities, unknown noise and different manufacturers. In the future, the proposed method will be further explored, and wider applications in PET/CT and PET/MR will be sought.

© 2022 Elsevier B.V. All rights reserved.

1. Introduction

Computed tomography (CT) is an important imaging modality and is extensively applied in modern hospitals and clinics in screening and diagnosis. Due to the health risks of high-dose X-ray radiation exposure[1], adaptive low-dose CT (LDCT) technologies have attracted wide interest for predicting normal-dose CT (NDCT) counterparts in research studies and clinical applications. However, a salient typical drawback of low-dose CT is the noise or artifact in images because of a weaker X-ray flux that compromises the diagnostic performance [2].

To address this ill-posed problem, many denoising methods have been developed to improve low-dose CT image quality. Generally, these methods can be divided into three categories: 1) sinogram-domain filtering methods [3–5], 2) iterative reconstruction methods [6–17], and 3) image postprocessing methods [18–25]. However, these methods are limited by their strong reliance on handcrafted regularization and hyperparameters, as well as high computational cost. The direct mapping relationship between NDCT images and LDCT images does not follow a clear rule, making it a challenging problem for a denoising model. For example, shortened sparse exposure views may lead to streak artifacts, and reduced tube current may result in mixed Gaussian and Poisson noise. Because LDCT images may not obey a certain distribution, the results generated by traditional methods may still suffer from artifacts. Compared with these traditional methods, deep

* Corresponding author.

E-mail address: na.zhang@siat.ac.cn (N. Zhang).

¹ These authors contributed equally to this work.

learning (DL)-based methods can take advantage of the LDCT domain due to their excellent feature extraction and nonlinear fitting capabilities based on convolutional neural networks (CNNs). In this study, much attention is paid to DL-based methods.

More recently, many DL-based methods have gained promising performance improvements in image denoising in the LDCT domain. Several DL-based methods adopt CNNs to obtain high-quality CT images from FBP images, such as basic convolutional neural networks [26], U-shaped networks [27,28], wavelet networks [29], residual encoder-decoder convolutional neural networks (RED-CNNs) [30], and deep cascade residual networks (DCRNs) [31]. Most of these methods are trained with the mean square error loss function, which may lead to oversmooth images. Subsequently, generative adversarial networks (GANs) [32–35] have been widely employed for LDCT imaging to address the limitation of voxelwise regression. In addition, Yin et al. [36] introduced a 3D residual convolution network in both the sinogram and image domains that achieves significantly improved performance by combining network processing in the two domains. Several model-based methods [37,38] seek to combine iterative reconstruction methods and neural networks to improve the interpretation of deep learning models. Li et al. [39] proposed self-attention leverage to capture a wide range of spatial information both intra- and interslices. On the other hand, prior information, such as anatomical information [40,41] and dose level [42], is introduced and improves the robustness of the trained models. Although previous DL-based methods have greatly improved the image quality of low-dose CT, most of these methods ignore regional anatomical differences.

In this paper, we proposed a segmentation-guided denoising network (SGDNet) based on 3D convolution and structural segmentation for loss CT imaging instead of introducing a probabilistic atlas prior for iterative solving algorithms in [43]. Considering X-ray absorbency and anatomical differences among the tissues, the semantic prior features from a structural semantic extraction network (SSE-LD) trained on low-dose CT images represent a pattern of adaptive spatial regional weights to guide the denoising process. Semantic feature transform technology is applied to transform semantic features into image features in each semantic fusion module. Moreover, a structural semantic loss function is designed to measure the semantic distance between the estimated normal-dose CT images and the gold standards. For image quality evaluation, segmentation accuracy is implemented by a trained structural semantic extraction network (SSE-ND) trained using the gold standard of normal dose CT images. On the other hand, most DL-based methods seek a mapping function to approximate the data distribution of the gold standard by different distance constraints, such as L1 or L2, and they often ignore direct structural semantic preservation. Therefore, we tend to explore a direct denoising approach to synthesize normal dose CT images with both semantic and distribution preservation in Fig. 1.

The contributions of this study are summarized in the following 3 categories:

- 1. Structural semantic information is introduced in the LDCT domain.** Considering substantial regional differences among tissues in terms of the absorbency of X-rays, structural semantic information is explored to improve the denoising performance. The semantic features extracted from LDCT images can guide the denoising process. Moreover, the semantic feature transform is proposed to combine the semantic and image features in each semantic fusion module.
- 2. A structural semantic loss function is designed for network training.** To measure the semantic distance of estimated NDCT images and the gold standard, a structural semantic loss function similar to perceptual loss is designed to improve the se-

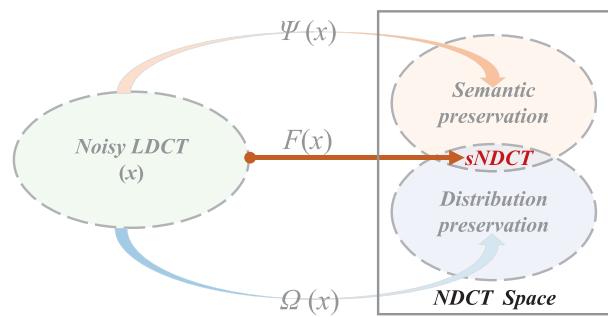


Fig. 1. The motivation of the proposed method is to obtain the estimated normal dose CT images equipped with both structural semantic and data distribution preservation. x denotes the input noisy LDCT images (x), and $sNDCT$ denotes synthetic NDCT images. $\Psi(x)$ denotes the semantic extraction function, $\Omega(x)$ denotes the distribution mapping function, and $F(x)$ denotes the promising function to estimate normal dose CT under conditional constraints of semantic and distribution preservation, respectively.

mantic evaluation results on synthetic NDCT images. This loss function depends on trained structural semantic models.

- 3. A semantic evaluation metric is applied for image quality assessment.** Considering that clinical applications, such as therapy, treatment as well as radiomics, usually contains image segmentation on normal dose CT images, we apply the quantitative semantic evaluation metric to measure the image quality of the estimated NDCT images.

This paper is organized as follows: Section 2 introduces the proposed segmentation-guided denoising methods. Section 3 elaborates the experimental details. Section 4 shows the quality of the results quantitatively under different evaluation metrics. Finally, a discussion and conclusions are given in Section 5 and Section 6.

2. Methods

In this section, the overview framework of the proposed segmentation-guided denoising network is elaborated. Importantly, the details of the semantic feature transform are described. Then, we present the structural semantic loss function, and multiple loss functions are introduced to improve the performance. Finally, parameter selection is described.

2.1. Structural semantic information

Previous works have sought an objective data distribution approximating the gold standard and have ignored structural semantic preservation. In addition, tissues show substantial regional anatomical differences in terms of the absorbency of X-rays. Then, the structural semantic information is explored in the LDCT denoising problem. This structural semantic information can be highly useful in three areas of the whole framework: 1) the structural semantic features extracted from LDCT images can be considered prior information to guide the denoising process; 2) the semantic distance on synthetic NDCT images and gold standards is an excellent loss measure function to train the whole network model; and 3) the evaluation metric of semantic accuracy can be employed for image quality assessment.

For the semantic labels in this study, abdominal tissue regions, including subcutaneous fat, muscle and visceral fat, are employed as the gold standard masks. In clinical applications, body mass index (BMI) is usually utilized in body physical evaluation and is defined as the ratio of height to weight squared. Compared with routine body mass index, these tissue regions (subcutaneous fat, muscle and visceral fat) can play a crucial role in more accurate body

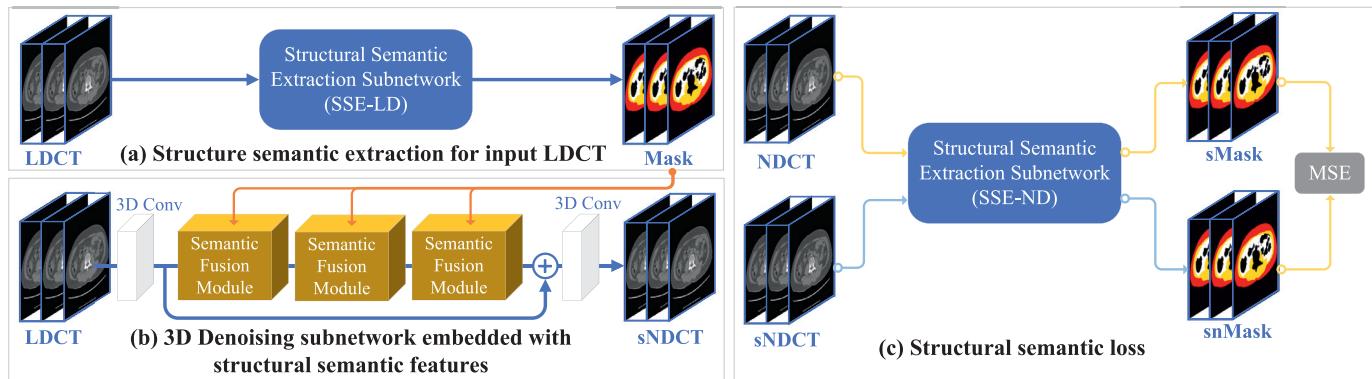


Fig. 2. Overview framework and structural semantic loss of the proposed method. $sNDCT$ denotes the estimated normal dose CT images. $Mask$ denotes the gold standard of segmentation labels, $sMask$ denotes the estimated segmentation results on NDCT images, and $snMask$ denotes the estimated segmentation results on preestimated NDCT images.

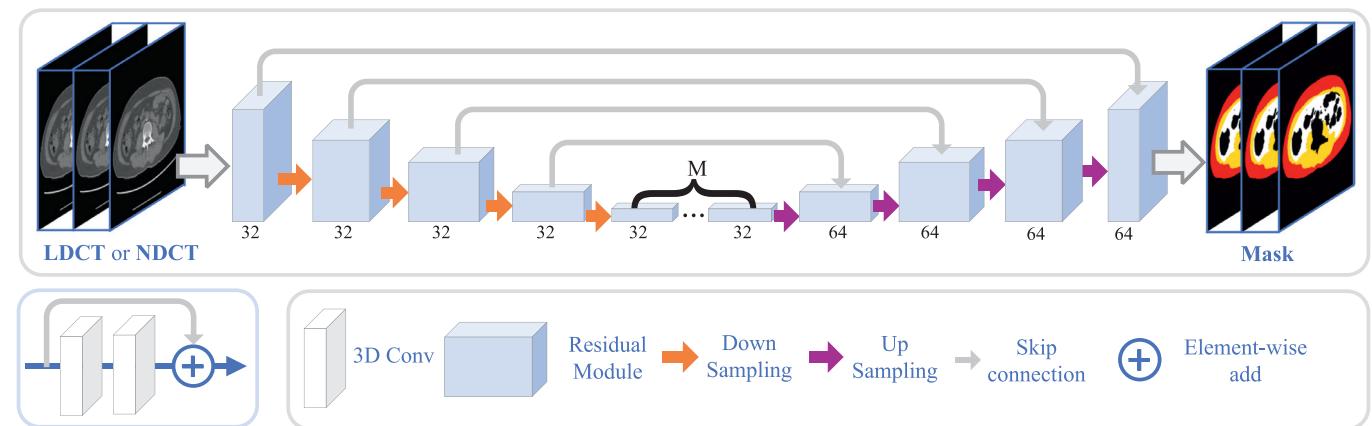


Fig. 3. Structural semantic extraction (SSE) subnetwork. $Mask$ denotes the shared gold standard of segmentation labels, including subcutaneous fat, muscle and visceral fat. M denotes the number of 3D residual modules and is fixed at 3 in this study.

physical evaluation for routine physical examination and radiotherapy, as well as prognosis assessment.

2.2. Segmentation-guided denoising network (SGDNet)

The main framework consists of two paths: 1) a structural semantic extraction subnetwork for low-dose CT (SSE-LD) in Fig. 2(a) and 2) a 3D denoising subnetwork embedded with semantic features in Fig. 2(b). Moreover, structural semantic loss is defined to measure the semantic distance between the estimated normal dose CT and the gold standard in Fig. 2(c). The SSE-LD subnetwork aims to extract structural semantic regions for subcutaneous fat, muscle and visceral fat.

2.2.1. The SSE-LD subnetwork

conducts a segmentation task to predict 3 kinds of structural semantic regions (subcutaneous fat, muscle and visceral fat), as shown in Fig. 3. This subnetwork adopts a 3D convolutional encoder-decoder network architecture. Both the encoder and decoder parts contain the residual module, and skip connections are employed. Concatenation is used to ensure the diversity of contextual feature information. In addition, the downsampling operation is implemented by a convolution with a stride fixed at 2, and the upsampling operation applies the deconvolution layer to enlarge the image size twice.

2.2.2. 3D denoising subnetwork

The goal of this subnetwork is to estimate the ultimate normal dose CT images embedded with structural semantic informa-

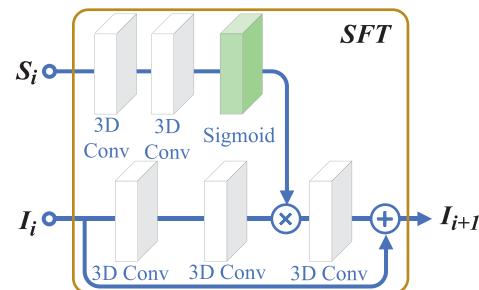


Fig. 4. Semantic fusion module. S_i denotes the estimated structural semantic results based on the SSE-LD subnetwork. SFT denotes the semantic feature transform that transforms the estimated semantic results to image feature maps I_i from LDCT images in each semantic fusion module.

tion from the SSE-LD subnetwork. The first and last 3D convolutional layers expand or squeeze channels to fit the input image size. The core components of this subnetwork are cascaded semantic fusion modules that combine structural semantic and image features through semantic feature transform (SFT) in Fig. 4. Inspired by the spatial feature transform in [44], adaptive affine scaling transformation is conducted on the semantic features extracted by the SSE-LD subnetwork.

2.3. Loss Functions

Given Y as NDCT gold standards, X denotes the corresponding LDCT counterpart. The inverse mapping function F between

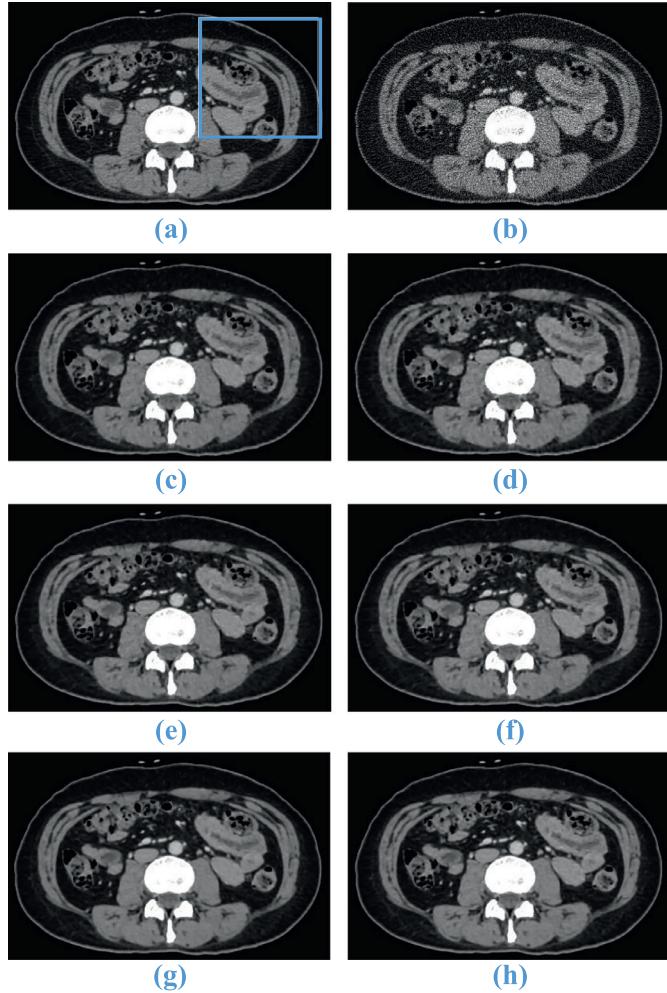


Fig. 5. Visual results of different methods. This abdomen display window is [−340, 460] HU. The blue box denotes the selected ROI. (a) NDCT. (b) LDCT. (c) CNN. (d) RED-CNN. (e) SACNN. (f) Baseline. (g) SGDNet-s(Ours). (h) SGDNet(Ours).

Y and X is used to achieve the goal of noise and artifact reduction. The proposed method learns NDCT images $Y = \{Y_1, Y_2, \dots, Y_N\}$ from their LDCT counterparts $X = \{X_1, X_2, \dots, X_N\}$ based on a 3D convolutional end-to-end network. In addition, LDCT and NDCT images are equipped with shared structural semantic labels $S = \{S_1, S_2, \dots, S_N\}$.

Before training denoising models, the pretrained SSE-LD and SSE-ND are needed. Next, the loss function of segmentation and denoising models are described.

2.3.1. Segmentation model

To obtain the trained segmentation model, the most commonly used loss functions, including dice coefficient loss and focal loss, are introduced [45]. For the multiclass segmentation, the Dice loss is formulated as:

$$L_{Dice} = \frac{1}{C} \sum_{c=1}^C - \frac{2 \sum_{n=1}^N S_{nc} \hat{S}_{nc}}{\sum_{n=1}^N S_{nc}^2 + \sum_{n=1}^N \hat{S}_{nc}^2}, \quad (1)$$

where C denotes the number of structural semantic labels and is fixed as 4 in this study.

The focal loss [46] is a dynamically scaled cross-entropy loss used to address the class imbalance problem for segmentation tasks. The focal loss for these studies is formulated as:

$$L_{Focal} = \frac{1}{C} \sum_{c=1}^C \sum_{n=1}^N (1 - \hat{S}_{nc})^\gamma S_{nc} \log \hat{S}_{nc}, \quad (2)$$

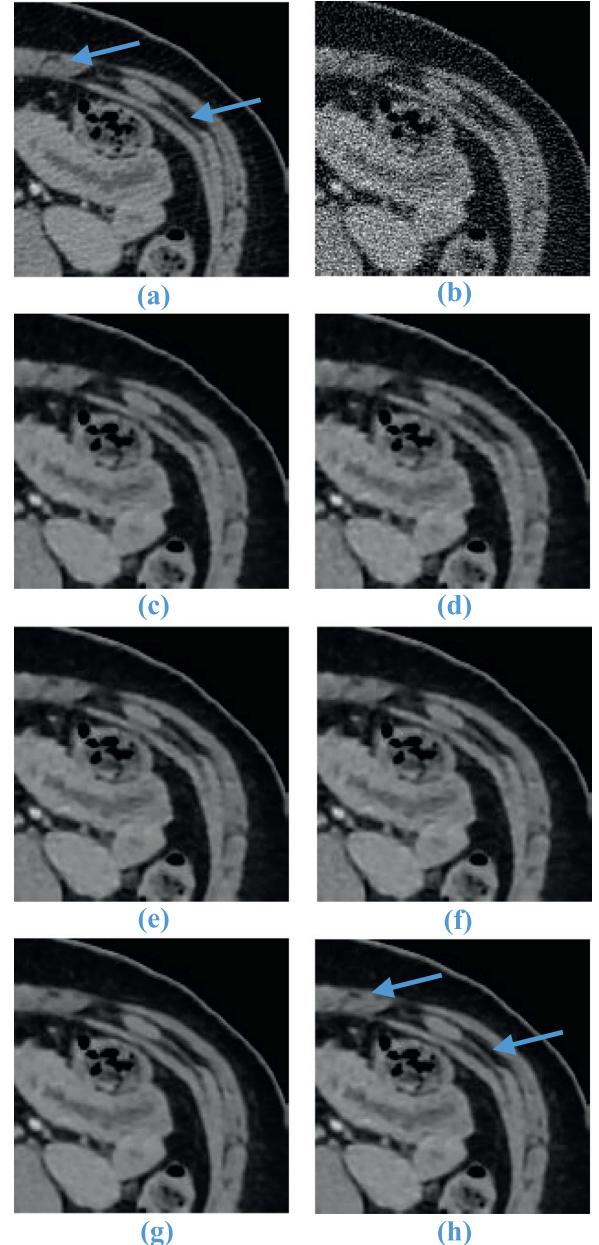


Fig. 6. Zoomed ROI of the red rectangle in Fig. 5. This abdomen display window is [−340, 460] HU. The blue arrows point to the obvious visual difference of the proposed method. (a) NDCT. (b) LDCT. (c) CNN. (d) RED-CNN. (e) SACNN. (f) Baseline. (g) SGDNet-s(Ours). (h) SGDNet(Ours).

where γ is a tunable parameter and fixed as 2. Then, the total loss function of the segmentation model is formulated as:

$$L_{Segmentation} = L_{Dice} + \alpha L_{Focal}, \quad (3)$$

where α is the balance factor and fixed as 0.5 following [45].

2.3.2. Denoising model

For the sake of the distribution and semantic preservation of synthetic normal-dose CT images, the mean absolute error and structural semantic loss are joined to the framework training. To obtain more high-frequency details, we utilize the MAE, that is also widely known as the L1-norm, and is given by:

$$L_{MAE} = \frac{1}{N} \sum_{n=1}^N \|F(X_n) - Y_n\|_1, \quad (4)$$

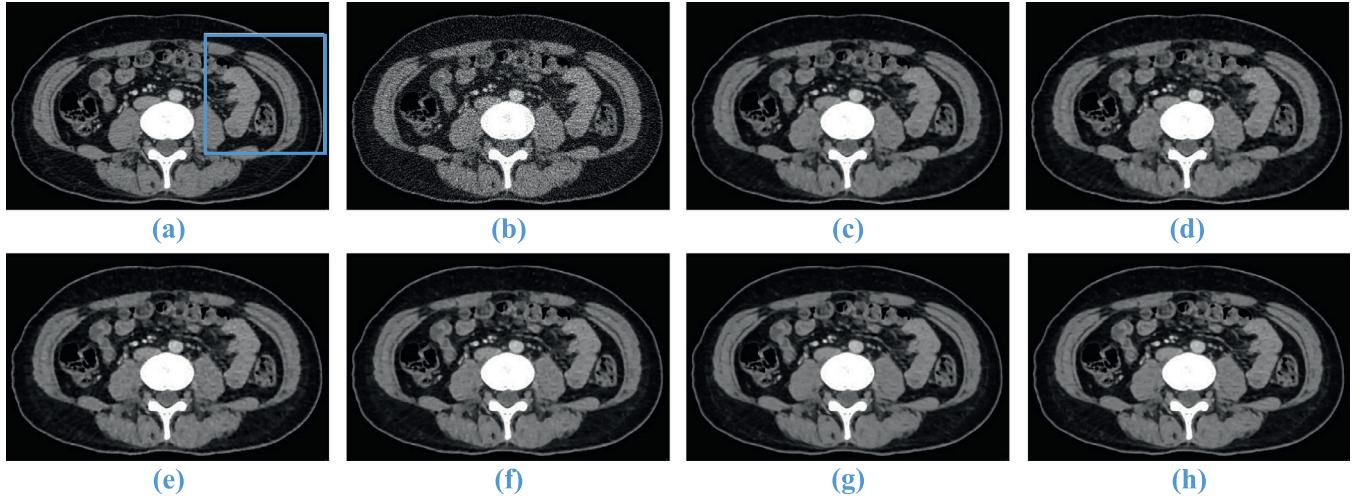


Fig. 7. Visual results of different methods. This abdomen display window is $[-340, 460]$ HU. The blue box denotes the selected ROI. (a) NDCT. (b) LDCT. (c) CNN. (d) RED-CNN. (e) SACNN. (f) Baseline. (g) SGDNet-s(Ours). (h) SGDNet(Ours).

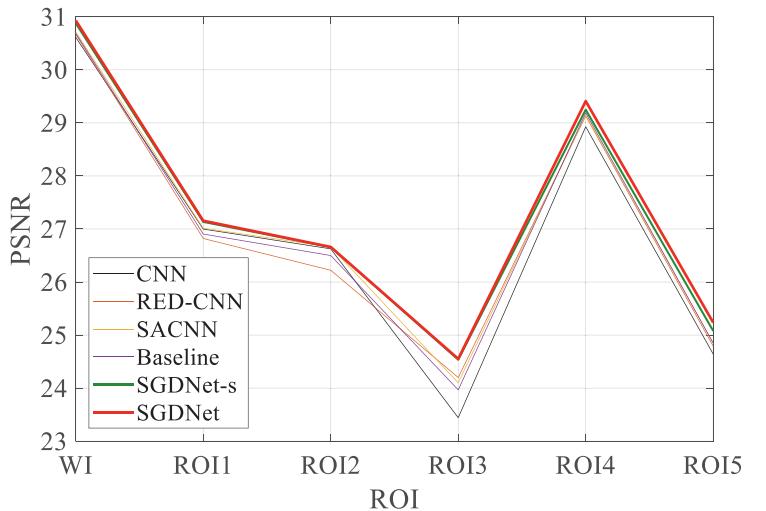
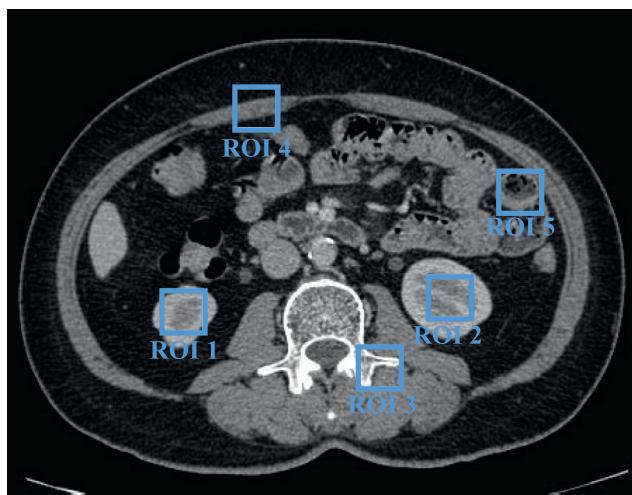


Fig. 8. Performance comparison of different methods over 5 ROIs in terms of PSNR. The blue boxes define selected ROIs.

To measure the semantic distance between synthetic NDCT images and gold standards, a structural semantic loss function similar to perceptual loss is designed to improve the semantic evaluation results on the predicted NDCT images.:

$$L_{SSL} = \frac{1}{N} \sum_{n=1}^N \|\psi(F(X_n)) - \psi(Y_n)\|_2^2, \quad (5)$$

ψ denotes the structural segmentation network (SSE-ND) trained on the NDCT images. Finally, the total loss function of the denoising model is given by:

$$L_{Denoising} = L_{MAE} + \beta L_{SSL}, \quad (6)$$

where β is the balance factor.

2.4. Parameter Selection

For the segmentation model, the U-net network is employed, and 3D residual convolutional layers are used to extract feature maps. The input of this segmentation model is a continual multislice $H \times W \times D$, and the output is a $H \times W \times D \times C$ estimated semantic mask, where H and W denote the height and width of the image shape, D denotes the depth of multislices and C denotes

the number of segmentation labels. The encoder and decoder consist of 5 3D residual convolution modules, and the filter size is $(3 \times 3 \times 3)$. For the down- and upsampling layers, the filter size is fixed at $(1 \times 2 \times 2)$, and the stride is fixed at $(1, 2, 2)$ to expand or squeeze the image size without changing the depth D . The channel number of the 3D convolution layer for the encoder is 32, and it is fixed at 64 for the decoder with concatenation operations. Finally, the output channel is fixed as C to output semantic results.

For the denoising model, the 3D convolution layer adopts a filter size of $(3 \times 3 \times 3)$, and the channel number is fixed at 32 to obtain considerable savings in CUDA memory and reduce the computation time. The input of this denoising model is a continual multislice $H \times W \times D$, and the output is a $H \times W \times D$ estimated denoised result.

3. Experiments

In this section, the performance of our method is validated. First, the clinical data studies from 35 patients and CT scanner parameter configurations are described. Next, training implementation details and evaluation metrics are described.

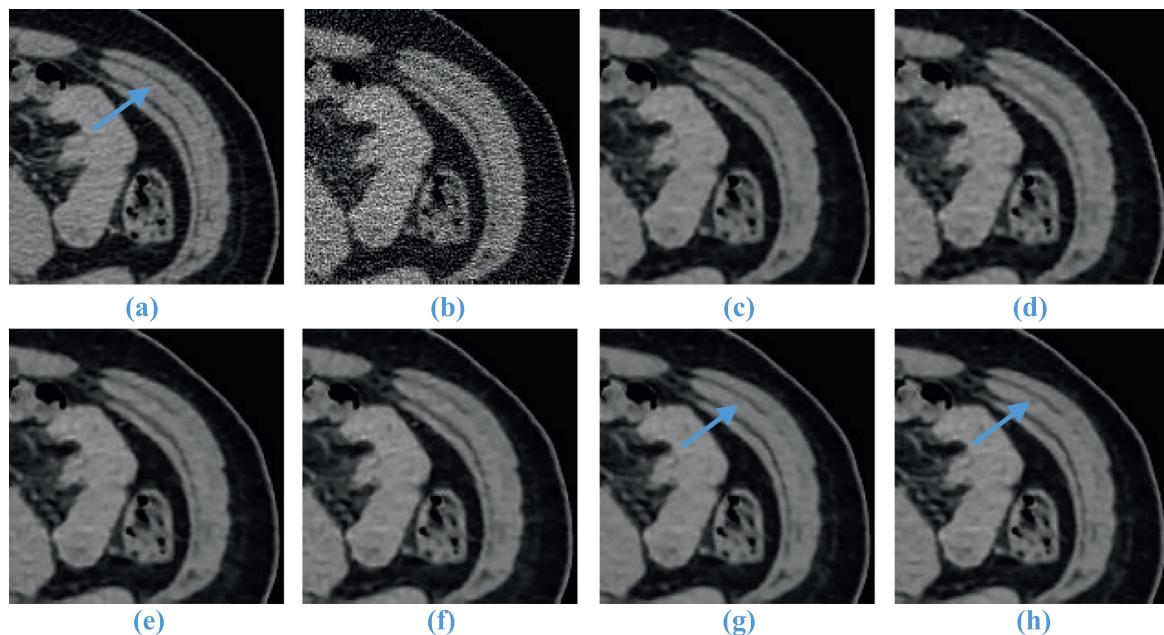


Fig. 9. Zoomed ROI of the red rectangle in Fig. 7. This abdomen display window is $[-340, 460]$ HU. The blue arrows point to the obvious visual difference of the proposed method. (a) NDCT. (b) LDCT. (c) CNN. (d) RED-CNN. (e) SACNN. (f) Baseline. (g) SGDNet-s(Ours). (h) SGDNet(Ours).

3.1. Patient Data Studies

Clinical abdomen data from 35 patients (15 male, 20 female, ages ranging from 32 to 71) were collected, and the dataset was acquired using a Siemens CT scanner (SOMATOM Definition). CT images from 28, 2 and 5 patients were selected as training, validation and test datasets, respectively. The number of abdominal CT images for each patient ranged from 18 to 22. The scan tube voltage is 120 kV, the tube current ranges from 150 to 600 mA, and the thickness is set as 1 mm. The reconstruction matrix is 512×512 . To avoid multiple scans for patients, we conducted the simulation process to obtain low-dose CT images through the MRIT toolbox [9], and the simulated low-dose CT images were reconstructed using the FBP algorithm similar to [47–49]. For segmentation labels, the abdomen data are divided into subcutaneous fat, muscle and visceral fat marked by radiologists, because these tissues play an important role in determining the physical health of the human body.

3.2. Training Implementation Details

The number of samples for clinical patients was over 700 slices. Data augmentation strategies are used on the training data, such as random horizontal or horizontal flipping and random rotation by 90° , 180° , and 270° , to increase the sample diversity. Training paired samples are randomly cropped into 64×64 image patches. To reduce the computational cost and CUDA memory, the depth of the continual multislices D is set to 3. The batch size was set to 32 for segmentation and denoising models. We trained 1,000 epochs for each model, and the initial learning rate was 0.0001, which was halved every 200 epochs for the three training datasets. The loss function is minimized via an ADAM optimizer with $\beta_1 = 0.9$ and $\beta_2 = 0.999$. We applied PyTorch to implement our models on an Ubuntu 18.04 system with an Nvidia Quadro RTX 8000 GPU.

3.3. Evaluation metrics

To evaluate the performance of the proposed method, popular metrics, including the peak signal-to-noise ratio (PSNR), structural

similarity index measure (SSIM) and normalized mean square error (RMSE), are calculated. The paired student t-test was introduced, and an alpha of 0.05 was used as the cutoff for significance. If the p value is less than 0.05, we reject the null hypothesis that there is no difference between the means and conclude that a significant difference does exist.

Furthermore, we introduce the semantic evaluation metrics for the estimated normal dose CT images. Specifically, the dice similarity is commonly used to quantify segmentation accuracy. In addition, the absolute error ratio of fat to muscle is introduced to further describe the segmentation performance.

4. Results

In this section, we evaluate the performance of the proposed SGDNet-s on individual test datasets from 5 patients. The statistical properties of the PSNR, SSIM and RMSE common evaluation metrics are employed to calculate the quantitatively evaluated the different methods. Moreover, the quantitative semantic metric and clinical ratio of fat-to muscle are used to measure the image quality of the estimated NDCT images.

To better illustrate the performance, common popular CNN [26], RED-CNN [30] and SACNN [39] are used as comparison methods. For a fair comparison, these methods are reimplemented on our datasets and adopt 3D convolutional neural networks. The training implementations are similar to those of the proposed models. For the proposed methods, the baseline model is the 3D convolutional neural network with 8 residual modules, and the primary version of SGDNet fused only with structural semantic features is named SGDNet-s. Then, the latter version SGDNet is trained jointly with the structural semantic loss functions. Both of these proposed models (SGDNet-s and SGDNet) are embedded with segmentation information.

4.1. Visual results

As shown in Fig. 5, the visual results of 6 methods are compared, and the proposed method SGDNet can generate images with more edge details. In Fig. 6, the zoomed region of interest is se-

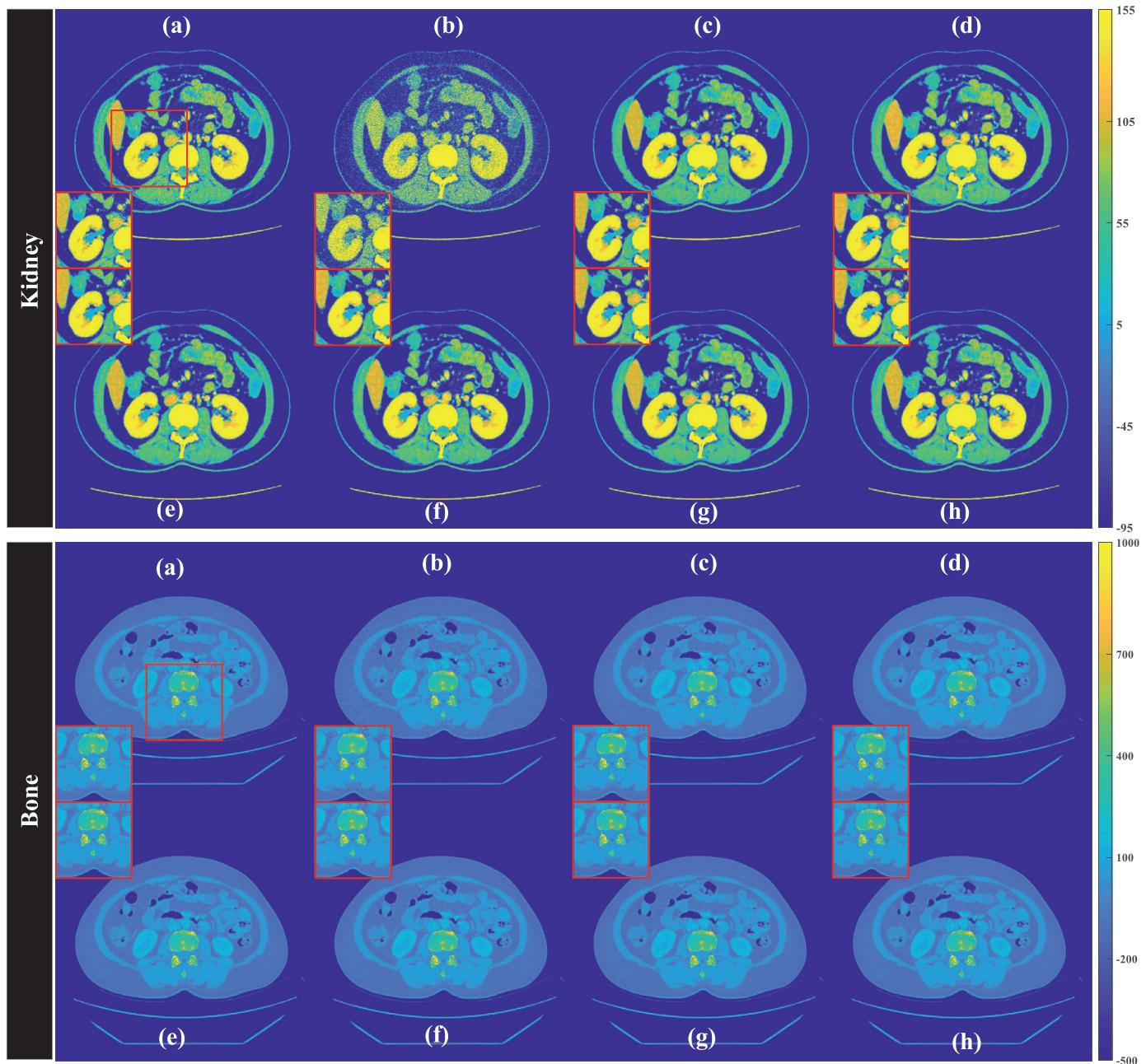


Fig. 10. Visual results of different methods under display windows of kidney and bone. The red box denotes the selected ROI. (a) NDCT. (b) LDCT. (c) CNN. (d) RED-CNN. (e) SACNN. (f) Baseline. (g) SGDNet-s(Ours). (h) SGDNet(Ours).

Table 1
Quantitative results (Mean \pm Std) on test datasets in terms of PSNR, SSIM and RMSE. The highest scores are highlighted in bold.

Methods	PSNR	SSIM	RMSE
LDCT	22.45 \pm 0.78	0.7522 \pm 0.0419	0.1757 \pm 0.0068
CNN	32.13 \pm 1.08*	0.9021 \pm 0.0268*	0.0249 \pm 0.0031*
RED-CNN	32.24 \pm 1.11*	0.9034 \pm 0.0269*	0.0246 \pm 0.0032*
SACNN	32.29 \pm 1.13*	0.9027 \pm 0.0268*	0.0245 \pm 0.0032*
Baseline	32.26 \pm 1.11*	0.9031 \pm 0.0267*	0.0246 \pm 0.0032*
SGDNet-s	32.49 \pm 1.16*	0.9050 \pm 0.0268*	0.0239 \pm 0.0032*
SGDNet	32.54 \pm 1.16*	0.9066 \pm 0.0263*	0.0238 \pm 0.0032*

* denotes $P < 0.05$, corresponding to a significant difference.

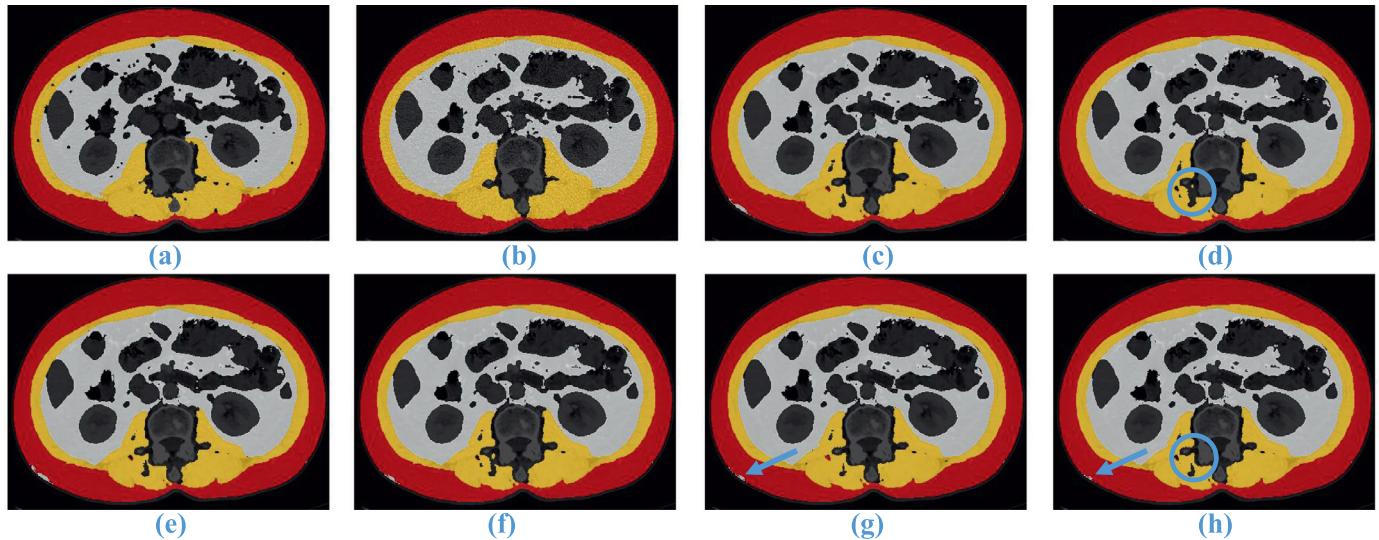


Fig. 11. Visual segmentation results of different methods implemented by SSE-ND. (a) GT. (b) NDCT. (c) CNN. (d) RED-CNN. (e) SACNN. (f) Baseline. (g) SGDNet-s(Ours). (h) SGDNet(Ours).

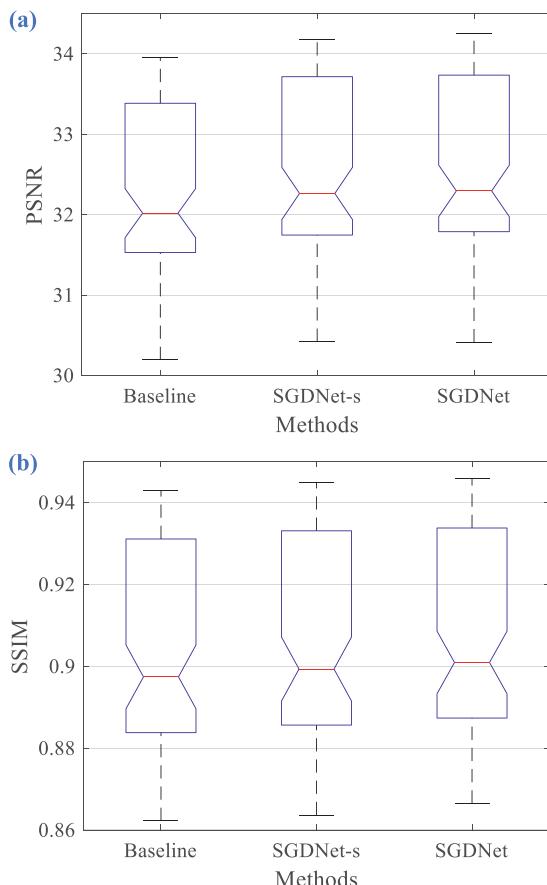


Fig. 12. Boxplot of statistical results for baseline, SGDNet-s and SGDNet in terms of PSNR and SSIM. The red lines denote the median values.

lected and marked in the blue box in Fig. 5. The same gap between subcutaneous fat marked by the blue arrow is blurred in the CNN, RED-CNN and SACNN images and can be observed clearly in SGDNet-s and SGDNet images. Another example is the connected slim fat that is evident in the SGDNet image.

In Fig. 7, the subcutaneous fat gap on the left in our methods is closer to the reference NDCT image. In addition, the magnified re-

gion of interest of the subcutaneous fat gap on the right is shown in Fig. 9. The same gap between the subcutaneous fat marked by the blue arrow is blurred in the CNN, RED-CNN, SACNN and baseline images and can be observed clearly in SGDNet-s and SGDNet images.

In Fig. 10, the proposed SGDnet depicts more textured structures on the kidney organs compared with other deep learning-based methods in the first row. In the second row, the highlighting and internal structure of bone density suggests that our method is closer to the ground truth.

To further demonstrate the visualization effect of the synthetic LDCT images, the semantic results fused with synthetic images are shown in Fig. 11. For the segmentation of the subcutaneous fat marked by the blue arrow, RED-CNN achieves the best semantic mask, and the proposed SGDNet is better than CNN, SACNN and baseline. However, RED-CNN performs poorly in the muscle semantic part marked by the blue circle, and the protruding edges are not preserved.

4.2. Quantitative results

In Table 1, we present the quantitative results obtained for all LDCT images that were not included among the training datasets. These quantitative results demonstrate the superior performance of our proposed SGDNet. The SGDNet gains an improvement of over 10.1 dB in terms of PSNR versus the LDCT images. In addition, the SGDNet achieves an improvement of more than 0.2 dB in PSNR versus SACNN.

As shown in Fig. 9, 5 regions of interest are selected, and statistical quantitative results on PSNR are compared. Based on observation, the proposed SGDNet and SGDNet show the best performance in all 5 ROIs. Specifically, evident improvement is observed in ROI3.

In Table 2, the quantitative segmentation results are compared. These quantitative results demonstrate the superior overall dice of our proposed SGDNet-s and SGDNet. SGDNet shows the best segmentation performance in dice on subcutaneous fat. In addition, SGDNet-s achieves the smallest fat-to-muscle error rate.

4.3. Ablation Analysis

In this part, ablation studies are conducted to validate the role of the two key components for the denoising effectiveness of the

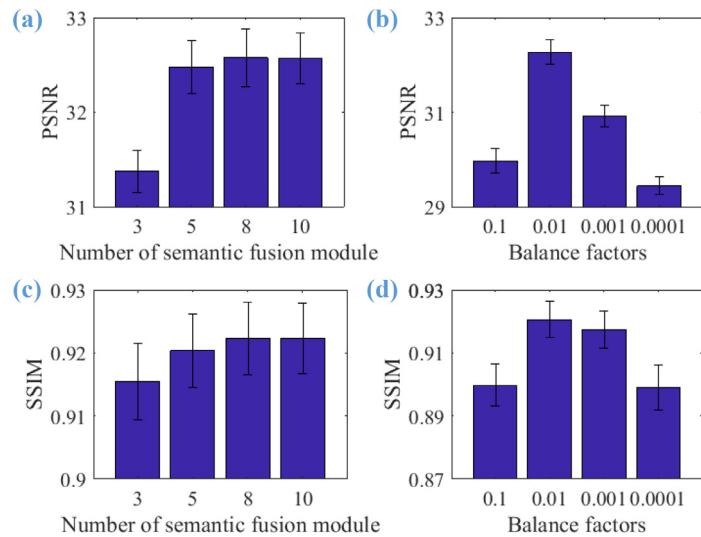


Fig. 13. Parameter selection based on the number of semantic fusion modules and balance factors.

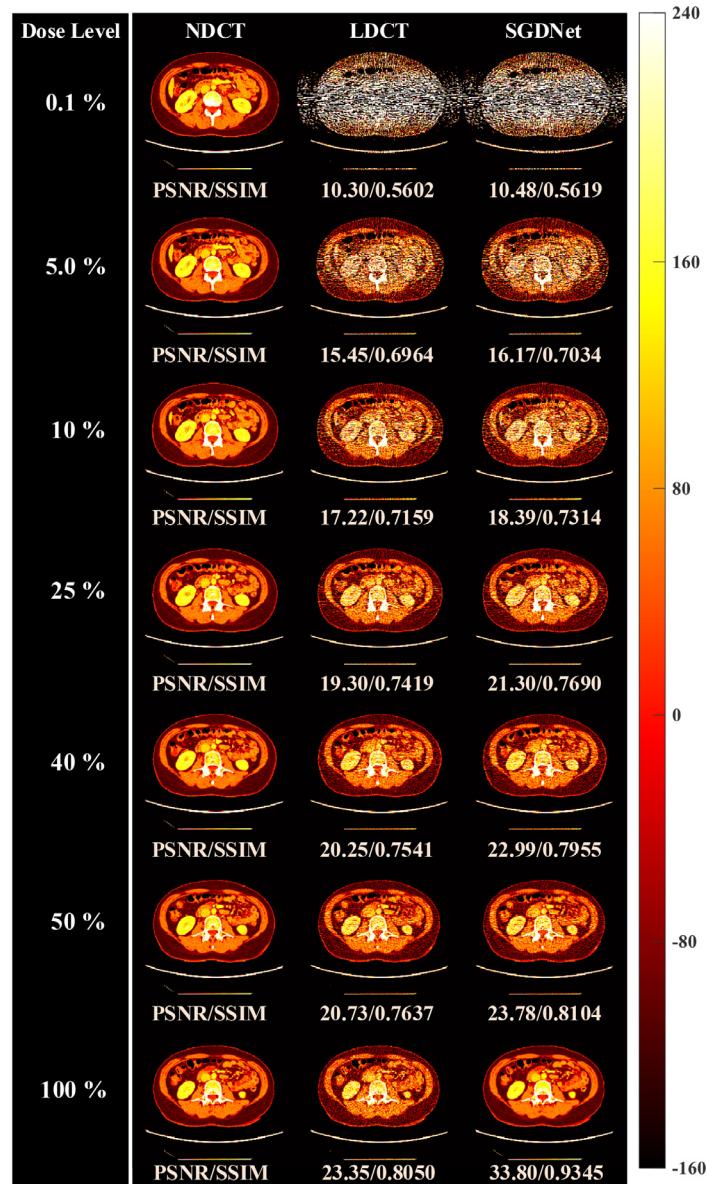


Fig. 14. Visual results of the proposed method under different dose levels. We simulated multiple low-dose images with different grades based on the simulated number of photons and fixed the current experimental low-dose level at 100%.

Table 2

Quantitative segmentation results (Mean \pm Std) of different methods in terms of the dice and absolute error ratio of fat-to-muscle. The highest scores are highlighted in bold.

Methods	Dice on subcutaneous fat	Dice on muscle	Dice on visceral fat	Overall dice	Overall absolute error ratio of fat-to-muscle
Segmentation GT	1.0000 \pm 0.0000	1.0000 \pm 0.0000	1.0000 \pm 0.0000	1.0000 \pm 0.0000	0.00% (2.8340 \pm 0.0333)
SSE-NDCT	0.9884 \pm 0.0018	0.9539 \pm 0.0055	0.9404 \pm 0.0094	0.9805 \pm 0.0015	5.00% (2.9758 \pm 0.0313)
CNN	0.9846 \pm 0.0024*	0.9490 \pm 0.0057*	0.9204 \pm 0.0106*	0.9758 \pm 0.0018*	7.30% (3.0410 \pm 0.0375*)
RED-CNN	0.9848 \pm 0.0021*	0.9441 \pm 0.0103*	0.9227 \pm 0.0102*	0.9751 \pm 0.0018*	8.88% (3.0858 \pm 0.0585*)
SACNN	0.9847 \pm 0.0022*	0.9476 \pm 0.0079*	0.9200 \pm 0.0111*	0.9756 \pm 0.0019*	7.23% (3.0390 \pm 0.0434*)
Baseline	0.9853 \pm 0.0021*	0.9455 \pm 0.0096*	0.9207 \pm 0.0107*	0.9753 \pm 0.0018*	8.08% (3.0631 \pm 0.0542*)
SGDNet-s	0.9854 \pm 0.0016*	0.9509 \pm 0.0061*	0.9209 \pm 0.0110*	0.9762 \pm 0.0019*	6.46% (3.0172 \pm 0.0360*)
SGDNet	0.9863 \pm 0.0017*	0.9503 \pm 0.0058*	0.9212 \pm 0.0110*	0.9762 \pm 0.0019*	7.21% (3.0382 \pm 0.0415*)

* denotes $P < 0.05$, corresponding to a significant difference.

proposed model. The effectiveness of structural semantic features and structural semantic loss function will be evaluated using PSNR and SSIM. To illustrate these cases, the evaluation results are calculated for the baseline, SGDNet-s and SGDNet.

4.3.1. Effectiveness of structural semantic features

To validate the effectiveness of the structural semantic features, we focus on the performance comparison between baseline and SGDNet-s. In Table 1, the quantitative results of SGDNet-s are better than those of the baseline. In addition, the boxplot based on the statistical data in terms of PSNR and SSIM is shown in Fig. 12. Compared with baseline, the integration of structural semantic features provides denoising performance improvement.

4.3.2. Effectiveness of structural semantic loss

For the denoising performance of SGDNet, we validate the effectiveness of the structural semantic loss function compared with SGDNet-s. In Table 1 and Fig. 12, the quantitative results for SGDNet show better PSNR and SSIM than those for SGDNet-s, demonstrating the denoising effectiveness of structural semantic loss. Moreover, the quantitative results on 5 local ROIs in Fig. 8 also prove that training with the structural semantic loss function provides denoising performance improvement.

5. Discussion

Although the denoising effectiveness of the proposed methods has been validated based on the ablation studies and experimental results, several aspects must still be considered further. First, we pay more attention to denoising performance instead of semantic evaluation results due to the core low-dose CT imaging task. Based on the quantitative results in Table 1 and Table 2, the denoising performance is better for SGDNet, while SGDNet-s may achieve superiority in structural segmentation. Both of these proposed models are equipped with structural semantic information, proving that the segmentation-guided features can achieve semantic preservation for low-dose CT images. The semantic evaluation needs further validation for increased amount of clinical data. In addition, 3 semantic labels (subcutaneous fat, muscle and visceral fat) for abdominal CT images are employed for segmentation tasks considering the clinical application of body physical evaluation. In fact, semantic labels may be different for different clinical applications.

We validate the performance on the number of semantic fusion and balance factors in terms of PSNR and SSIM. As shown in Fig. 13, the first column shows the evaluation results on 3, 5, 8 and 10 semantic fusion modules; the second column shows evaluation results under different balance factors (0.1, 0.001, 0.0001 and 0.00001) on the validation data. The proper number of semantic fusion modules is 8, and the balance factor for β is 0.01. If the data are acquired using equipment from different manufacturers, the parameter configuration may need to be adjusted appropriately

for different tasks. If the data are acquired using different manufacturers, the parameter configuration may need to be adjusted appropriately for different tasks. Moreover, the noise intensity of low-dose CT images also affects the parameter configuration, resulting in ultimate denoising performance. As shown in Fig. 14, as the dose decreases, the performance of the proposed method also decreases, even though it can achieve a certain denoising effect. In particular, image restoration at ultralow doses will be very poor. For different noise intensities, our previous work [42] proposed a dose-aware network to address the low-dose CT imaging problem under different noise levels. In addition, our previous work [40,41] proposed adaptive networks considering the anatomical differences under different anatomical sites implemented with an attention mechanism. In the future, we will combine multiple strategies to conduct the denoising task of low-dose CT imaging.

It must be acknowledged that abnormalities of CT images have a huge impact on the segmentation network. In this work, CT images were collected from healthy volunteers for accurate body physical evaluation for routine physical examination and radiotherapy, as well as prognosis assessment. Considering these abnormalities, higher quality segmentation labels need to be provided by radiologists, and more experimental CT data need to be augmented to improve the robust of the proposed method in future works.

Incorrect segmentation will also affect the ultimate denoising network performance. The segmentation information is adaptively introduced to the denoising subnetwork implementation by an attention mask. The weight of the attention mask can effectively filter several incorrect segmentation data and preserve a large number of features of the original denoising subnetwork. In addition, a structural semantic loss function is designed for network training. To measure the semantic distance of estimated NDCT images and the gold standard, the structural semantic loss function similar to perceptual loss is designed to improve the semantic evaluation results on synthetic NDCT images. This loss function combined with the mean square error will result in better denoising performance during the training process. The semantic evaluation metric could be suggested on wide range of applications for image quality assessment, including tumor therapy [50,51].

In the future, we may explore other clinical applications with certain structural semantic labels to validate the proposed method. Moreover, multimodal applications, such as PET/CT and PET/MR, can be explored in combination with structural semantic information. For example, the structural semantic features from MR or CT can be excellent prior information for low-count PET. Furthermore, it could play an advantage in optical imaging and be applied to improve image resolution [52,53].

6. Conclusion

In conclusion, most previous works seek an objective data distribution approximating the gold standard and neglect structural semantic preservation for low-dose CT imaging. Therefore, we ex-

plore a direct denoising approach to synthesize normal-dose CT images with both semantic and distribution preservation. First, the structural semantic features represent a pattern of adaptive spatial regional weights to guide the denoising process. Then, a structural semantic loss function is designed to measure the semantic distance between the estimated normal dose CT images and the gold standards to improve the denoising performance. Furthermore, we introduce the semantic evaluation metrics for the synthetic NDCT images. The quantitative experimental results illustrate the promising performance of the proposed SGDNet-s and SGDNet in noise reduction and structural semantic preservation. However, the proposed method may suffer from several limitations under abnormalities, unknown noise and different manufacturers, and possible solutions are explored to alleviate these dilemmas. In the future, the proposed method will seek improvement and take advantage of multimodal applications, such as PET/CT and PET/MR.

Declaration of Competing Interest

- All authors have participated in (a) conception and design, or analysis and interpretation of the data; (b) drafting the article or revising it critically for important intellectual content; and (c) approval of the final version.
- This manuscript has not been submitted to, nor is under review at, another journal or other publishing venue.
- The authors have no affiliation with any organization with a direct or indirect financial interest in the subject matter discussed in the manuscript.
- The following authors have affiliations with organizations with direct or indirect financial interest in the subject matter discussed in the manuscript.

Acknowledgments

This work was supported by the National Natural Science Foundation of China (62101540), the Shenzhen Science and Technology Program (RCBS20210706092218043), the China Postdoctoral Science Foundation (2022M713290), and the Guangdong Innovation Platform of Translational Research for Cerebrovascular Diseases of China. The authors would like to thank the editor and anonymous reviewers for their constructive comments and suggestions. All authors declare that they have no known conflicts of interest in terms of competing financial interests or personal relationships that influence or are relevant to the work reported in this paper.

References

- [1] D.J. Brenner, E.J. Hall, Computed tomographyan increasing source of radiation exposure, *New England Journal of Medicine* 357 (22) (2007) 2277–2284.
- [2] C. Jiang, N. Zhang, J. Gao, Z. Hu, Geometric calibration of a stationary digital breast tomosynthesis system based on distributed carbon nanotube x-ray source arrays, *Plos one* 12 (11) (2017) e0188367.
- [3] M. Balda, J. Hornegger, B. Heismann, Ray contribution masks for structure adaptive sinogram filtering, *IEEE transactions on medical imaging* 31 (6) (2012) 1228–1239.
- [4] A. Manduca, L. Yu, J.D. Trzasko, N. Khaylova, J.M. Kofler, C.M. McCollough, J.G. Fletcher, Projection space denoising with bilateral filtering and ct noise modeling for dose reduction in ct, *Medical physics* 36 (11) (2009) 4911–4919.
- [5] J. Wang, T. Li, H. Lu, Z. Liang, Penalized weighted least-squares approach to sinogram noise reduction and image reconstruction for low-dose x-ray computed tomography, *IEEE transactions on medical imaging* 25 (10) (2006) 1272–1283.
- [6] E.Y. Sidky, X. Pan, Image reconstruction in circular cone-beam computed tomography by constrained, total-variation minimization, *Physics in Medicine & Biology* 53 (17) (2008) 4777.
- [7] Y. Zhang, W. Zhang, Y. Lei, J. Zhou, Few-view image reconstruction with fractional-order total variation, *JOSA A* 31 (5) (2014) 981–995.
- [8] Y. Zhang, Y. Wang, W. Zhang, F. Lin, Y. Pu, J. Zhou, Statistical iterative reconstruction using adaptive fractional order regularization, *Biomedical optics express* 7 (3) (2016) 1015–1029.
- [9] Y. Zhang, W.-H. Zhang, H. Chen, M.-L. Yang, T.-Y. Li, J.-L. Zhou, Few-view image reconstruction combining total variation and a high-order norm, *International Journal of Imaging Systems and Technology* 23 (3) (2013) 249–255.
- [10] J. Huang, Y. Zhang, J. Ma, D. Zeng, Z. Bian, S. Niu, Q. Feng, Z. Liang, W. Chen, Iterative image reconstruction for sparse-view ct using normal-dose image induced total variation prior, *PLoS one* 8 (11) (2013) e79709.
- [11] Z. Hu, Y. Zhang, J. Liu, J. Ma, H. Zheng, D. Liang, A feature refinement approach for statistical interior ct reconstruction, *Physics in Medicine & Biology* 61 (14) (2016) 5311.
- [12] Y. Chen, D. Gao, C. Nie, L. Luo, W. Chen, X. Yin, Y. Lin, Bayesian statistical reconstruction for low-dose x-ray computed tomography using an adaptive-weighting nonlocal prior, *Computerized Medical Imaging and Graphics* 33 (7) (2009) 495–500.
- [13] J. Ma, H. Zhang, Y. Gao, J. Huang, Z. Liang, Q. Feng, W. Chen, Iterative image reconstruction for cerebral perfusion ct using a pre-contrast scan induced edge-preserving prior, *Physics in Medicine & Biology* 57 (22) (2012) 7519.
- [14] Y. Zhang, Y. Xi, Q. Yang, W. Cong, J. Zhou, G. Wang, Spectral ct reconstruction with image sparsity and spectral mean, *IEEE transactions on computational imaging* 2 (4) (2016) 510–523.
- [15] M. Green, E.M. Marom, N. Kiryati, E. Konen, A. Mayer, Efficient low-dose ct denoising by locally-consistent non-local means (lc-nlm), in: *International conference on medical image computing and computer-assisted intervention*, Springer, 2016, pp. 423–431.
- [16] H. Shangguan, Q. Zhang, Y. Liu, X. Cui, Y. Bai, Z. Gui, Low-dose ct statistical iterative reconstruction via modified mrf regularization, *Computer methods and programs in biomedicine* 123 (2016) 129–141.
- [17] J.-F. Cai, X. Jia, H. Gao, S.B. Jiang, Z. Shen, H. Zhao, Cine cone beam ct reconstruction using low-rank matrix factorization: algorithm and a proof-of-principle study, *IEEE transactions on medical imaging* 33 (8) (2014) 1581–1591.
- [18] M. Katsura, I. Matsuda, M. Akahane, J. Sato, H. Akai, K. Yasaka, A. Kunimatsu, K. Ohtomo, Model-based iterative reconstruction technique for radiation dose reduction in chest ct: comparison with the adaptive statistical iterative reconstruction technique, *European radiology* 22 (8) (2012) 1613–1623.
- [19] Y. Chen, X. Yin, L. Shi, H. Shu, L. Luo, J.-L. Coatrieux, C. Toumoulin, Improving abdomen tumor low-dose ct images using a fast dictionary learning based processing, *Physics in Medicine & Biology* 58 (16) (2013) 5803.
- [20] Y. Lu, J. Zhao, G. Wang, Few-view image reconstruction with dual dictionaries, *Physics in Medicine & Biology* 57 (1) (2011) 173.
- [21] X.-Y. Cui, Z.-G. Gui, Q. Zhang, H. Shangguan, A.-H. Wang, Learning-based artifact removal via image decomposition for low-dose ct image processing, *ieee transactions on nuclear science* 63 (3) (2016) 1860–1873.
- [22] Z. Li, L. Yu, J.D. Trzasko, D.S. Lake, D.J. Blezek, J.G. Fletcher, C.H. McCollough, A. Manduca, Adaptive nonlocal means filtering based on local noise level for ct denoising, *Medical physics* 41 (1) (2014) 011908.
- [23] H. Zhang, J. Ma, J. Wang, Y. Liu, H. Lu, Z. Liang, Statistical image reconstruction for low-dose ct using nonlocal means-based regularization, *Computerized Medical Imaging and Graphics* 38 (6) (2014) 423–435.
- [24] P.F. Feruglio, C. Vinegoni, J. Gros, A. Sbarbati, R. Weissleder, Block matching 3d random noise filtering for absorption optical projection tomography, *Physics in Medicine & Biology* 55 (18) (2010) 5401.
- [25] D. Kang, P. Slomka, R. Nakazato, J. Woo, D.S. Berman, C.-C.J. Kuo, D. Dey, Image denoising of low-radiation dose coronary ct angiography by an adaptive block-matching 3d algorithm, in: *Medical Imaging 2013: Image Processing*, volume 8669, International Society for Optics and Photonics, 2013, p. 86692G.
- [26] H. Chen, Y. Zhang, W. Zhang, P. Liao, K. Li, J. Zhou, G. Wang, Low-dose ct via convolutional neural network, *Biomedical optics express* 8 (2) (2017) 679–694.
- [27] O. Ronneberger, P. Fischer, T. Brox, U-net: Convolutional networks for biomedical image segmentation, in: *International Conference on Medical image computing and computer-assisted intervention*, Springer, 2015, pp. 234–241.
- [28] H. Li, K. Mueller, Low-dose ct streak artifacts removal using deep residual neural network, in: *Proceedings of Fully 3D conference*, volume 2017, 2017, pp. 191–194.
- [29] E. Kang, J. Min, J.C. Ye, A deep convolutional neural network using directional wavelets for low-dose x-ray ct reconstruction, *Medical physics* 44 (10) (2017) e360–e375.
- [30] H. Chen, Y. Zhang, M.K. Kalra, F. Lin, Y. Chen, P. Liao, J. Zhou, G. Wang, Low-dose ct with a residual encoder-decoder convolutional neural network, *IEEE transactions on medical imaging* 36 (12) (2017) 2524–2535.
- [31] Z. Huang, Z. Chen, G. Quan, Y. Du, Y. Yang, X. Liu, H. Zheng, D. Liang, Z. Hu, Deep cascade residual networks (dcrn): Optimizing an encoder-decoder convolutional neural network for low-dose ct imaging, *IEEE Transactions on Radiation and Plasma Medical Sciences* (2022).
- [32] Z. Hu, C. Jiang, F. Sun, Q. Zhang, Y. Ge, Y. Yang, X. Liu, H. Zheng, D. Liang, Artifact correction in low-dose dental ct imaging using wasserstein generative adversarial networks, *Medical physics* 46 (4) (2019) 1686–1696.
- [33] J.M. Wolterink, T. Leiner, M.A. Viergever, I. Işgum, Generative adversarial networks for noise reduction in low-dose ct, *IEEE transactions on medical imaging* 36 (12) (2017) 2536–2545.
- [34] Q. Yang, P. Yan, Y. Zhang, H. Yu, Y. Shi, X. Mou, M.K. Kalra, Y. Zhang, L. Sun, G. Wang, Low-dose ct image denoising using a generative adversarial network with wasserstein distance and perceptual loss, *IEEE transactions on medical imaging* 37 (6) (2018) 1348–1357.
- [35] F. Jiao, Z. Gui, Y. Liu, L. Yao, P. Zhang, Low-dose ct image denoising via frequency division and encoder-dual decoder gan, *Signal, Image and Video Processing* (2021) 1–9.
- [36] X. Yin, Q. Zhao, J. Liu, W. Yang, J. Yang, G. Quan, Y. Chen, H. Shu, L. Luo, J.-L. Coatrieux, Domain progressive 3d residual convolution network to improve low-dose ct imaging, *IEEE transactions on medical imaging* 38 (12) (2019) 2903–2913.

- [37] D. Wu, K. Kim, G. El Fakhri, Q. Li, Iterative low-dose ct reconstruction with priors trained by artificial neural network, *IEEE transactions on medical imaging* 36 (12) (2017) 2479–2486.
- [38] K.H. Jin, M.T. McCann, E. Froustey, M. Unser, Deep convolutional neural network for inverse problems in imaging, *IEEE Transactions on Image Processing* 26 (9) (2017) 4509–4522.
- [39] M. Li, W. Hsu, X. Xie, J. Cong, W. Gao, Sacnn: self-attention convolutional neural network for low-dose ct denoising with self-supervised perceptual loss network, *IEEE transactions on medical imaging* 39 (7) (2020) 2289–2301.
- [40] Z. Huang, X. Liu, R. Wang, Z. Chen, Y. Yang, X. Liu, H. Zheng, D. Liang, Z. Hu, Learning a deep cnn denoising approach using anatomical prior information implemented with attention mechanism for low-dose ct imaging on clinical patient data from multiple anatomical sites, *IEEE Journal of Biomedical and Health Informatics* 25 (9) (2021) 3416–3427.
- [41] Z. Huang, X. Liu, R. Wang, J. Chen, P. Lu, Q. Zhang, C. Jiang, Y. Yang, X. Liu, H. Zheng, et al., Considering anatomical prior information for low-dose ct image enhancement using attribute-augmented wasserstein generative adversarial networks, *Neurocomputing* 428 (2021) 104–115.
- [42] Z. Huang, Z. Chen, J. Chen, P. Lu, G. Quan, Y. Du, C. Li, Z. Gu, Y. Yang, X. Liu, et al., Danet: dose-aware network embedded with dose-level estimation for low-dose ct imaging, *Physics in Medicine & Biology* 66 (1) (2021) 015005.
- [43] E.A. Rashed, H. Kudo, Probabilistic atlas prior for ct image reconstruction, *Computer Methods and Programs in Biomedicine* 128 (2016) 119–136.
- [44] X. Wang, K. Yu, C. Dong, C. Change Loy, Recovering realistic texture in image super-resolution by deep spatial feature transform, in: *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2018.
- [45] W. Zhu, Y. Huang, L. Zeng, X. Chen, Y. Liu, Z. Qian, N. Du, W. Fan, X. Xie, Anatomynet: deep learning for fast and fully automated whole-volume segmentation of head and neck anatomy, *Medical physics* 46 (2) (2019) 576–589.
- [46] T.-Y. Lin, P. Goyal, R. Girshick, K. He, P. Dollár, Focal loss for dense object detection, in: *Proceedings of the IEEE international conference on computer vision*, 2017, pp. 2980–2988.
- [47] Z. Huang, Z. Chen, Q. Zhang, G. Quan, M. Ji, C. Zhang, Y. Yang, X. Liu, D. Liang, H. Zheng, et al., Cagan: a cycle-consistent generative adversarial network with attention for low-dose ct imaging, *IEEE Transactions on Computational Imaging* 6 (2020) 1203–1218.
- [48] J. Liu, Y. Zhang, Q. Zhao, T. Lv, W. Wu, N. Cai, G. Quan, W. Yang, Y. Chen, L. Luo, et al., Deep iterative reconstruction estimation (dire): approximate iterative reconstruction estimation for low dose ct imaging, *Physics in Medicine & Biology* 64 (13) (2019) 135007.
- [49] H. Gupta, K.H. Jin, H.Q. Nguyen, M.T. McCann, M. Unser, Cnn-based projected gradient descent for consistent ct image reconstruction, *IEEE transactions on medical imaging* 37 (6) (2018) 1440–1453.
- [50] H. Huang, C. Dong, M. Chang, L. Ding, L. Chen, W. Feng, Y. Chen, Mitochondria-specific nanocatalysts for chemotherapy-augmented sequential chemoreactive tumor therapy, in: *Exploration*, volume 1, Wiley Online Library, 2021, pp. 50–60.
- [51] L. Tu, Z. Liao, Z. Luo, Y.-L. Wu, A. Herrmann, S. Huo, Ultrasound-controlled drug release and drug activation for cancer therapy, in: *Exploration*, volume 1, Wiley Online Library, 2021, p. 20210023.
- [52] K. Wanderi, Z. Cui, Organic fluorescent nanoprobes with nir-iib characteristics for deep learning, in: *Exploration*, volume 2, Wiley Online Library, 2022, p. 20210097.
- [53] J. Chen, T. Chen, Q. Fang, C. Pan, O.U. Akakuru, W. Ren, J. Lin, A. Sheng, X. Ma, A. Wu, Gd₂O₃/b-tio₂ composite nanoprobes with ultra-high photoconversion efficiency for mr image-guided nir-ii photothermal therapy, in: *Exploration*, Wiley Online Library, p. 20220014.