Emerging Challenges of AI for Biomedical Image Analysis: A CBCT image noise reduction method based on cGAN

Weili Shi  
Changchun University of Science and Technology

Yu Miao (miaoyu@cust.edu.cn)  
Changchun University of Science and Technology

Chengcheng Zhang  
Changchun University of Science and Technology

Yifei Ding  
Changchun University of Science and Technology

Lemin Shi  
Changchun University of Science and Technology

Mingye Li  
RMIT University

Wei Wang  
Monash University

Research Article

Keywords: tooth segmentation, CBCT, oral scanning model, neural networks, image noise reduction

Posted Date: November 9th, 2022

DOI: https://doi.org/10.21203/rs.3.rs-2233402/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License. Read Full License
Emerging Challenges of AI for Biomedical Image Analysis: A CBCT image noise reduction method based on cGAN

Weili Shi1,2, Yu Miao1,2,*, Chengcheng Zhang1, Yifei Ding1, Lemin Shi1, Mingye Li3,4,*, Wei Wang4,5,*

1 School of Computer Science and Technology, Changchun University of Science and Technology, Changchun, Jilin 130022, China

2 Zhongshan Institute of Changchun University of Science and Technology, Zhongshan, Guangdong 528400, China

3 Department of Information Systems and Business Analytics, RMIT University, Melbourne, Victoria 3001, Australia

4 School of Computing and Information Systems, University of Melbourne, Parkville, Victoria 3010, Australia

5 Department of Software Systems and Cybersecurity, Monash University, Clayton, Victoria 3800, Australia

Correspondence can be addressed to Yu Miao: miaoyu@cust.edu.cn; Mingye Li: mingye.li1@unimelb.edu.au; and Wei Wang: wang.w5@monash.edu
Abstract

Background

As an imaging modality, cone beam CT (CBCT) is widely employed in dentistry, which can help dentists to observe tissues such as roots and jaws without confusion. CBCT has the advantages of convenience and low radiation dose, however low contrast and large noise are the critical issues in the images. These disadvantages make it difficult for dentists to identify target tissues accurately. Due to the differences in scanning methods and reconstruction algorithms between CBCT and multi-row detector spiral CT (MDCT), the current CT noise reduction models present significant shortcomings in terms of reducing the noise on CBCT images.

Results

In this paper, we propose a method of image noise reduction based on a conditional generative adversarial network to improve the quality of CBCT images. The normal-dose MDCT images are used as the ground truth images for model training to generate the denoised images. In order to increase the model’s sensitivity to the gradient information, a gradient loss function is involved in our proposed method. The verification experiments on the simulated data set and the real data set show that our model effectively generates denoised images as well as preserves the quality of the images. We compare the denoising effect between our model and other models with different loss functions. The scores by PSNR, MS-SSIM and GMSD show that our model presents better edge characteristics and a denoising effect.
Conclusion

Through the experiments, the performance of our improved cGAN for denoising in oral CBCT data are better than the compared with other models in terms of evaluation score and visual quality. Our model is more competitive for clinical application. Comparing with other models which only reduce simulated noise, we directly input CBCT images to reduce the noise.

Keywords: tooth segmentation, CBCT, oral scanning model, neural networks, image noise reduction

Background

With the development of CT equipment and technology, low-dose CT (LDCT) imaging is becoming more and more popular [1]. Cone beam CT (CBCT), as a representative of LDCT, some of CBCT’s advantages are lower dose, prompter imaging speed, and lower cost. However, due to its lower dose, the image quality is poor, and the noise increases during the reconstruction.

In recent years, machine learning has made rapid progress in image processing. It has been widely used in image noise reduction. For visible images, Kupyn et al. proposed the method named DeblurGAN [2] to remove motion blur in images. Its model improved Generative Adversarial Networks (GAN) [3] and used the discriminator of patchGAN [4]. Other research clusters used Generative Networks to reduce the noise for medical images. Yi et al. proposed Sharpness Aware Generative Adversarial Network (SAGAN), adding the detection network results into the loss function to train the network [5]. Cheng et al. proposed a novel CBCT-based generative architecture to improve the accuracy of anatomical structures [6]. Lei et al. introduced Dense block
and AG into CycleGAN [7]. Ma et al. designed a generative adversarial network that combines least squares, structural similarity and L1 loss for low-dose CT denoising [8]. Gu et al. proposed a novel CycleGAN architecture using a single generator, which can convert a low-dose CT image to a routine-dose CT image [9]. Chen Hu et al. proposed RED-CNN to reduce the noise for LDCT images [10], which was proved to improve the training speed compared to other models. Despite such advantages, some disadvantages such as low radiation dose and high noise level of CBCT images could not be ignored. An additional difficulty is that CBCT and Multi-Detector row spiral CT (MDCT) has different hardware systems and reconstruction algorithms. The current models, however, have some limitations such as noise misidentification and incomplete noise reduction.

In this paper, we propose an image noise reduction method based on cGAN to improve the quality of CBCT images. We use the oral CBCT images as input and the MDCT images as the ground truth images. We train the network to generate the corresponding MDCT image to reduce the noise in CBCT. We propose a gradient loss function, which is a part of the generator G loss function, to make the model more sensitive to image gradient and to improve the quality of the generated images.

Results

In this section, we discuss the performance of our model in terms of noise reduction for CBCT data. The noise reduction of simulated data is compared with Poisson noise and real data for RED-CNN [10] and SAGAN [5]. The choice of hyperparameters and the adjusting of the model are discussed as well.
The experiment data include CBCT and the head MDCT of five volunteers from the Second Hospital of Jilin University. CBCT's tube potential is 90kV, current is 10mA, and slice thickness is 0.3mm; MDCT's tube potential is 120kV, current is 232mA, and slice thickness is 0.9mm. The direction and position of the data was deviated due to the fact that it came from a different scanning device. We first preprocessed the data. We used iso-surface to 3D reconstruct the data, used a point cloud registration technology based on SAC-IA and ICP to register the 3D model, and resampled data again to obtain the experiment data of the same direction and the same slice.

We used the CBCT image as the input of generator $G$ and the MDCT image as the ground truth image to reduce the noise in the CBCT image. The experiment uses 620 images as training data and 150 images as test data. The batch size is 1, the patch is 256, the number of training is 300, and the learning rate was kept at $10^{-4}$ for the first 150 epochs and decreased linearly to 0 for the last 150 epochs. The experiment system is Ubuntu 16.04, the CPU is 20-core Intel Xeon E5-2698, and the model that is based on tensorflow1.14 is trained on a tesla V100.

**Image quality evaluation method**

We used the peak signal-to-noise ratio (PSNR), multi-level structural similarity (MS-SSIM), and gradient amplitude similarity deviation (GMSD) to evaluate the quality of the generated image.

PSNR is a full reference image quality evaluation standard. It is calculated as follows:

$$\text{PSNR} = 10 \log_{10} \left( \frac{(2^n - 1)^2}{\text{MSE}} \right)$$  \hspace{1cm} (1)
where, MSE represents the mean square error between the current image X and the reference image Y, MSE is calculated as follows:

$$MSE = \frac{1}{H \times W} \sum_{i=1}^{H} \sum_{j=1}^{W} (X(i,j) - Y(i,j))^2$$  \hspace{1cm} (2)$$

where W represents the length and H represents height of the image, n represents the number of pixel bit, and the unit of PSNR is dB. The larger the value, the smaller the distortion.

MS-SSIM[16]is the improved structural similarity proposed by Wang et al., Which is closer to the subjective quality evaluation results than SSIM. SSIM is calculated as follows:

$$SSIM(X,Y) = \frac{L(M(X,Y)) \cdot C_j(X,Y) \cdot S_j(X,Y)}{\prod_{j=1}^{M}[C_j(X,Y)]^{\beta_j}[S_j(X,Y)]^{\gamma_j}}$$  \hspace{1cm} (3)$$

MS-SSIM adds scale information to the calculation of the SSIM, scales down the original image to varying degrees and calculates the contrast factor and structure factor on each size image. The calculation method is shown as follows:

$$MSSSIM(X,Y) = [L_M(X,Y)]^{\alpha_5} \prod_{j=1}^{M}[C_j(X,Y)]^{\beta_j}[S_j(X,Y)]^{\gamma_j}$$  \hspace{1cm} (4)$$

X is current image and Y is the reference image, L(X, Y) is the brightness contrast factor, C (X, Y) is the contrast factor of contrast, and S (X, Y) is the structure contrast factor, $\beta_1 = \gamma_1 = 0.0448$, $\beta_2 = \gamma_2 = 0.2856$, $\beta_3 = \gamma_3 = 0.3001$, $\beta_4 = \gamma_4 = 0.2363$, $\alpha_5 = \beta_5 = \gamma_5 = 0.1333$, M represents the scale factor of image reduction. L, C and S are calculated as follows:

$$L(X,Y) = \frac{2u_Xu_Y + C_1}{u_X^2 + u_Y^2 + C_1}$$  \hspace{1cm} (5)$$
\[ C(X,Y) = \frac{2\sigma_X\sigma_Y + C_2}{\sigma_X^2 + \sigma_Y^2 + C_2} \]  

\[ S(X,Y) = \frac{\sigma_{XY} + C_3}{\sigma_X\sigma_Y + C_3} \]  

(6)  

(7)  

The \( u_X \) and \( u_Y \) are the image means, the \( \sigma_X \) and the \( \sigma_Y \) are the image variances, and \( \sigma_{XY} \) is the covariance. The higher the MS-SSIM value, the more similar the two image structures.

GMSD was proposed by Zhang et al. [17] in 2014. It evaluates the structural distortion of local images by the local gradient amplitude, and uses the standard deviation of local image quality to measure the global image quality, the calculation method is:

\[ GMSD = \sqrt{\frac{1}{N} \sum_{i=1}^{N} (GMS(i) - GMSM)^2} \]  

(8)  

GMS is the similarity of the gradient amplitude and is calculated as follows, GMSM is the mean of the regional gradient field.

\[ GMS(i) = \frac{2m_r(i)m_d(i) + c}{m_r^2(i) + m_d^2(i) + c} \]  

(9)  

The \( m_r \) represents the gradient magnitude of the current image and the \( m_d \) represents the gradient magnitude of the ground truth image.

\[ m_r(i) = \sqrt{(r \otimes h_x)^2(i) + (r \otimes h_y)^2(i)} \]  

(10)  

\[ m_d(i) = \sqrt{(d \otimes h_x)^2(i) + (d \otimes h_y)^2(i)} \]  

(11)  

The \( h_x \) and \( h_y \) are the operators of Prewitt. The smaller the GMSD value, the closer the current image is to the ground truth image.
Simulation data set

We added Poisson noise to the MDCT to simulate the LDCT data. The number of images, size, and network parameters were kept consistent with the real dataset. The objective evaluation scores for each model in the test data are given in Table 1. We found that RED-CNN was slightly better at denoising the simulated dataset than our model. The results also show that SAGAN has a lower overall score because it incorrectly identifies low-density regions as errors, but the main tissues, such as teeth, are clear. Figure shows the results of the different model on simulated data.

| Classification | Model     | Psnr  | Ms-ssim | Gmsd |
|----------------|-----------|-------|---------|------|
| All            | Noise image | 39.531 | 0.996   | 0.000 |
|                | RED-CNN   | 45.886 | **0.998** | 0.000 |
|                | SAGAN     | 28.279 | 0.936   | 0.011 |
|                | Our model | 43.835 | **0.998** | 0.000 |

Figure 1: The results of the different model on simulated data. (a) CBCT data. (b) RED-CNN result. (c) SAGAN result. (d) Our model’s result. (e) MDCT data.

Real data set

The real dataset uses CBCT images as LDCT images and their corresponding MDCT images as ground truth images. We divided the data into three groups, discussing the denoise results of the various models on the crown, root and jaw respectively.
Figure shows the results in the test set. It shows that the images trained by RED-CNN are blurred. However, the images trained with the SAGAN model are clearer and insensitive to noise compared to RED-CNN model. Some noises have not been removed, and the part of the low-density bone is misidentified as noise.

Figure 2: Qualitative comparison of each model on the test set. (a) crown part. (b) root part. (c) jaw part. number 1 is the original CBCT image, number 2 is the denoise image by RED-CNN, number 3 is the denoise image by SAGAN, and number 4 is denoise image by our model, number 5 is the MDCT image.

To evaluate the performance of the proposed approach, we conduct experiments on four pars: crown part, root part, jaw part and all parts. Table 2 presents the quantitative performance results of various methods on 4 parts. It can be seen that our approach achieves better results than other methods on all of the classification.

We constructed the error images according to the average of absolute difference between all generated images and ground truth images pixel by pixel in the test set.
Table 2: Evaluation scores of RED-CNN, SAGAM and our model in real data.

| Classification | Model   | Psnr  | Ms-ssim | Gmsd  |
|----------------|---------|-------|---------|-------|
| Crown part     | CBCT    | 18.536| 0.742   | 0.033 |
|                | RED-CNN | 21.249| 0.867   | 0.027 |
|                | SAGAN   | 18.241| 0.792   | 0.039 |
|                | Our model | **22.834** | **0.895** | **0.022** |
| Root part      | CBCT    | 22.366| 0.795   | 0.024 |
|                | RED-CNN | 24.477| 0.897   | 0.020 |
|                | SAGAN   | 21.561| 0.83    | 0.031 |
|                | Our model | **25.957** | **0.920** | **0.016** |
| Jaw part       | CBCT    | 22.435| 0.776   | 0.031 |
|                | RED-CNN | 23.328| 0.851   | 0.031 |
|                | SAGAN   | 21.899| 0.791   | 0.041 |
|                | Our model | **24.034** | **0.872** | **0.026** |
| All parts      | CBCT    | 21.112| 0.771   | 0.029 |
|                | RED-CNN | 23.018| 0.872   | 0.026 |
|                | SAGAN   | 20.567| 0.804   | 0.037 |
|                | Our-model | **24.275** | **0.896** | **0.021** |

The representative example is shown in Figure 3. It can be clearly seen that SAGAN method shows obvious wrong texture features, RED-CNN method shows less wrong texture features than SAGAN method, and our model hardly shows wrong texture features.

![Error images of different models. (a) error image by RED-CNN. (b) error image by SAGAN. (c) error image by our model.](image)

Model analysis

Hyperparameter selection and model adjustment
In this section, we demonstrate the effectiveness of the long skip link through the contrast experiments and discuss the hyperparameter selection and the effect of the pre-processing.

**Long skip link**

We respectively tested the quality of the images generated by the model without skip connections and with one skip connection. The scores shown in Table 3 indicate that the quality of the images generated by the model with the skip-in link added was improved.

| Method                                         | PSNR  | MS-SSIM | GMSD  |
|-----------------------------------------------|-------|---------|-------|
| Our model                                     | 24.275| 0.896   | 0.021 |
| One skip link model after the first down sampling | 24.041| 0.894   | 0.022 |
| No skip link model                            | 23.978| 0.893   | 0.022 |

**The number of Residual Blocks**

The depth of the model is determined by the number of residual blocks. We compared several depth models with residual block number counts of 6, 8, 10, 12, and 14. The results are depicted in Figure 4. It can be observed that increasing the depth of the model within a given range improves the quality of the produced image. However, as the depth continued to increase, the quality of the generated image decreased. Meanwhile, as the model depth increases, the training time linearly increase.
The effects of different noise on noise reduction

Contrast models in oral CBCT images are less effective in denoising due to different tissue types and CT imaging methods. The current LDCT denoising model is mainly used for soft tissues such as the abdominal cavity and the brain. These tissues have similar CT values at adjacent points, and they are continuous. However, the stomatology tests include examinations of bones, teeth, and gums. The density of bone cancellous and bone dense is uneven, and the CT values at the edge of bones change suddenly, which are distinctly different from the soft tissues and thus affect the denoising effect.

MDCT scan method is multi-row detector scanning layer-by-layer, while CBCT is completely different, using a conical beam emitter for circular projection around the projected object and different reconstruction algorithms. It can also be concluded CBCT contains more kinds of noise compared to other models. The radial noise shown
in Figure (a) is related to its position in the scanned area. For this type of noise, the
SAGAN model has the best denoising effect, while RED-CNN and our model still have
some noise. Figure (b) focuses on comparing the effect of CT partial volume effect on
the denoised images, and the results show that there is a significant volume effect on
the edges of the teeth in the CBCT image. The results also demonstrate that our model
outperformed SAGAN and RED-CNN in terms of noise reduction, for example,
SAGAN and RED-CNN did not completely eliminate noise. Figure 5(c) shows the
region of strong noise, where our model outperforms the other models in terms of noise
removal. In summary, the model proposed in this paper identifies noise more accurately
than the comparison model. However, our model does not completely remove the radial
noise, when compared with other models.

Figure 5: The impact of various noises on CBCT and the denoise effect of each model.
(a) Radial noise. (b) Volume effect. (c) Strong noise area. The number 1 is the original CBCT
image, number 2 is denoised by RED-CNN number 3 is denoised by SAGAN, and number 4
is denoised by our model, number 5 is the MDCT image.
It can be concluded from the experiments that RED-CNN, while performing well in objective evaluation metrics such as PSNR, is less visually effective and more blurred than other models. To evaluate the effect of noise removal, we did Gaussian filtering and mean filtering with a window size of 3 on the MDCT images with Poisson noise, and calculated the image quality of the results. Table 4 shows the experiment results. The results indicate that for Poisson noise, the above filtering improves the objective evaluation score of the images to some extent, and that image blurring should be avoided when training the model with denoising.

Table 4: Image quality after Gaussian filtering and mean filtering.

| Method                | PSNR  | MS-SSIM | GMSD |
|-----------------------|-------|---------|------|
| MDCT with Poisson noise | 39.531| 0.996   | 0.000|
| Mean filtering        | 42.424| 0.996   | 0.000|
| Gaussian filtering    | 43.961| 0.997   | 0.000|

Effectiveness of gradient loss

We discuss the effect of the generator G loss function on the quality of the generated images. As loss functions, we employ several loss functions and their combinations. The comparative findings are displayed in Figure 6.
Figure 6: The results of different loss function. (1) Our model. (2) Perceptual loss combined with adversarial loss. (3) Gradient loss combined with MSE. (4) Gradient loss combined with MSE and perceptual loss. (5) Gradient loss combined with MSE and adversarial loss. (6) Perceptual loss combined with adversarial loss and MSE. (7) Perceptual loss combined with adversarial loss and gradient loss.

Comparing the results in Figure 6(2) and (3), it can be seen that the "gradient loss combined with MSE" model performs better on PSNR than the "perceptual loss combined with antagonistic loss" model, but worse on GMSD and MS-SSIM. We combined perceptual loss with gradient loss and MSE, and antagonistic loss with gradient loss and MSE. The results are shown in Figure 6(4) and (5), respectively. For the model with adversarial loss, the generated image quality is poor, and it appears fake details. We also compared the “perceptual loss combined with adversarial loss and MSE” model with “perceptual loss combined with adversarial and gradient loss” model. Figure 6(6) and (7) show the results. It can be observed that our model obtains the best results with a combination of the four loss functions as the final function.
Figure 7: The effect of gradient loss. (a) CBCT image. (b) The model with gradient loss. (c) The model without gradient loss. (d) MDCT image.

We compared the effect between the models without gradient loss and the model with gradient loss on the edge information of the generated image. Figure 6(1) and (6) show the quantitative results. In the test data, the model with gradient loss has better edge characteristics, as shown in Figure 7.

Figure 8 shows the difference of objective evaluation score between the model with gradient loss and the model without gradient loss. We compared the difference between the PSNR and MS-SSIM scores of each image in the test set. And formula (21), (22) define $\Delta \text{PSNR}$ and $\Delta \text{MS-SSIM}$ respectively:

$$\Delta \text{PSNR} = \text{PSNR}_{\text{with gradient loss model}} - \text{PSNR}_{\text{without gradient loss model}}$$ (12)

$$\Delta \text{MS-SSIM} = \text{MS-SSIM}_{\text{with gradient loss model}} - \text{MS-SSIM}_{\text{without gradient loss model}}$$ (13)

It can be observed that the median of $\Delta \text{PSNR}$ is larger than 0, and the first quartile is slightly lower than 0 in Figure 8, while the median and lower quartile of $\Delta \text{MS-SSIM}$ are both greater than 0. It can be concluded that the test image gradient loss improves the quality of the generated image. But there are also some images, the model without gradient loss denoised better yet.
The effects of registration

The denoising effect of our proposed model depends on the input paired matching images, so the 3D image registration effect plays an important role in the denoising effect. We achieved different registration effects by adjusting the parameters in the registration algorithm based on SAC-IA and ICP, and further compared the effects of registration on denoising, as shown in Figure 9.

Figure 9: Effect of registration on teeth. (a) Accurate registration effect. (b) Inaccurate registration effect.
Figure 10: The difference in denoising effect. (a) Result with accurate registration effect. (b) Result with inaccurate registration effect. The number 1 is a CBCT image, number 2 is the denoised image, number 3 is a MDCT image corresponding to number 1.

According to Figure 10, we can clearly see that there is some noise interference at the upper end of the teeth in b, and the image quality is not clear enough, while the upper end of the teeth in a is clearer, the denoising effect is better than b. Therefore, we can conclude that the precise registration effect can be helpful for subsequent denoising experiments.

Conclusions

We propose an improved cGAN for denoising in oral CBCT data. We use adversarial and content loss with gradient loss to train the model. Through the experiments, the results of our model are better than the compared with other models.

Our model has a higher evaluation score and better visual quality. It is satisfied to retain bone cancellous. Also, it can show the edge of the image more obviously. Our model
is more competitive for clinical application. Comparing with other models which only reduce simulated noise, we directly input CBCT images to reduce the noise.

However, our model still has some limitations. First, it also exists noise in MDCT. We need to consider its effect on the training and evaluation. Second, the scores of PSNR and other indicators may be inconsistent with the visual effect in the evaluation for medical images. It is necessary to further optimize the quantitative evaluation method of CT image quality. Finally, for all methods based on deep learning, it needs to be trained for specific dose levels, window widths, and window levels. Our model only reduces noise for high-density bones and teeth in oral CBCT data. The denoise performance of other parts and other tissues remains to be evaluated.

Methods

cGAN has been widely used in many kinds of image processing [18]. We improved the generator structure of cGAN using the Unet-like structure and enhanced gradient propagation through skip connection. We have designed a gradient loss function that better reflect the noise effect on the image gradient. Minimizing the gradient loss can retain the image edge information while reducing the noise.

cGAN

The original GAN consists of generator G and discriminator D. Both of them are trained through adversarial training process. G receives the noise signal as input and generates samples. D receives real samples and the generated samples, and attempts to differentiate between these two samples. The goal of training G is to create samples that are extremely similar to the real samples. Training D is to accurately identify whether the samples originate from the generated samples or the real samples. This
training process is analogous to a min-max game between $G$ and $D$.

$$\min_G \max_D V(D, G) = E_{x \sim P_r} \left[ \log(D(x)) \right] + E_{z \sim P_g} \left[ \log(1 - D(G(z))) \right]$$ \hspace{1cm} (14)

where, the input $x$ of $D$ satisfies real data distribution $P_r$, and input $z$ of $G$ which is the random noise satisfies distribution $P_g$.

The original GAN, on the other hand, may suffer from problems such as model collapse and gradient disappearance [19]. Arjovsky suggested the Wasserstein GAN [19] to address these issues by using quantitative training metrics and substantially avoiding model collapse. Wasserstein GAN improves the stability of training compared with GAN [11].

Furthermore, to address the randomness of the input and output of GAN [12], cGAN was proposed. cGAN specifies the mapping from input to output by adding constraint $y$ into generator $G$ and discriminator $D$, respectively:

$$\min_G \max_D V(D, G) = E_{x \sim P_r} \left[ \log(D(x|y)) \right] + E_{z \sim P_g} \left[ \log(1 - D(G(z|y))) \right]$$ \hspace{1cm} (15)

In the noise reduction process, constraint $y$ is the ground truth image. The ground truth images with noise are put into generator $G$, while both the images generated by the generator $G$ and the ground truth images are put into discriminator $D$. Through the adversarial training, the generator $G$ continuously generates images that are closer to the ground truth image and eventually outputs the denoised image.

**Models**

We propose an improved cGAN to solve noise misidentification and incomplete noise reduction problems existing in current cGAN models for CBCT images.
The generator G uses Unet-like structure in this paper which is similar to DeblurGAN [2]. Figure 11 shows the network structure. As an advantage over the direct convolution on the original image, Auto Encoder (AE) structure shortens training time by doing image transformation on a smaller size through down sampling and then restoring the original image size through deconvolution. Long skip connections can help increase the number of feature channels, thus making it easier to propagate texture information from the original image [13].

![Figure 11: The structure of generator G.](image)

We use residual blocks in the residual net (ResNet) [14] to do the image transformation on the reduced size. Figure 12 shows the residual block structure. The residual blocks help the gradient propagation, prevent the gradient from disappearing, and avoid the degradation of the network when deepening. In our model, the number of residual blocks is 10.
The discriminator $D$ structure of our model is consistent with that of patchGAN proposed by Isola et al [4]. The generator's loss function is the sum of content and adversarial loss:

$$
\mathcal{L} = \mathcal{L}_{adv} + \mathcal{L}_{content}
$$

The adversarial loss is a loss function in the original GAN. Its goal is to generate images that approximate the ground truth image with high precision.

The content loss has three parts: the perceptual loss, the L2 loss (mean square error, MSE), and the gradient loss.

$$
\mathcal{L}_{content} = \lambda_1 * \mathcal{L}_{perceptual} + \lambda_2 * \mathcal{L}_{L2} + \lambda_3 * \mathcal{L}_{gradient}
$$

In recent research, the output of deepen network expressed the abstract features [15]. The perceptual loss is mainly focused on how to restore the general content of the generated image [4]. In our paper, we calculate the perceptual loss from the output of third activated function on the third layer of VGG19.
Although L2 loss is the traditional method for content loss, it offers several advantages, including smoothness, ease of finding derivatives, and stability of solutions. However, when there is the only one optimization objective, the generated image is blurred.

We propose the gradient loss, which makes the model more sensitive to image gradient and edge information. The gradient vector of image function $f(x, y)$ at the point $(x, y)$ is defined as:

$$\nabla f(x, y) = \left[ G_x, G_y \right]^T = \left[ \frac{\partial f}{\partial x}, \frac{\partial f}{\partial y} \right]^T$$ (18)

The digital image can be expressed by a two-dimensional discrete function. We use the approximation differential derivatives in horizontal and vertical directions. For the gradient $G$ of the digital image at the point $(i, j)$, we have:

$$\text{grad} = [\text{grad}_x, \text{grad}_y]^T$$ (19)

$$\text{grad}_x = f(i + 1, j) - f(i - 1, j)$$ (20)

$$\text{grad}_y = f(i, j + 1) - f(i, j - 1)$$ (21)

The difference vector measures how far the generated images are from the ground truth images. The gradient loss, defined as the mean of the squares of the difference vectors, as follows:

$$L_{\text{gradient}} = \frac{1}{W \times H} \sum \| \text{grad}(G_{ij}) - \text{grad}(\tilde{G}_{ij}) \|^2$$ (22)

where $G_{x_{ij}}$ and $G_{y_{ij}}$ represent the gradient of the generated image, $\tilde{G}_{x_{ij}}$ and $\tilde{G}_{y_{ij}}$ represent the gradient of the ground truth image at point $(i, j)$, with $W$, $H$ being the length and height of the image respectively.
In medical images, both edge and noise are high-frequency components. It depends on the gradient information when doing the edge detecting and segmentation. The noise greatly effects the accuracy. L2 loss only concerns the difference value of each pixel. It can’t reflect the effect of noise on the gradient. Gradient loss solves this problem.

For example, for the sequence $a_n = [1,2,3,4,5,6]$, if we define the gradient as $a_i = a_{i+1} - a_i$, the gradient sequence is $\Delta a_n = [1,1,1,1,1]$. Inserting a noise of amplitude 1 at the places 3 and 4 in the sequence $a_n$, the possible results are shown in Table 5.

| No. | Sequence   | MSE | Gradient sequence | Gradient loss |
|-----|------------|-----|-------------------|---------------|
| 1   | [1,2,2,3,5,6] | 2   | [1,0,1,2,1]       | 2             |
| 2   | [1,2,4,5,5,6] | 2   | [1,2,1,0,1]       | 2             |
| 3   | [1,2,2,5,5,6] | 2   | [1,0,3,0,1]       | 6             |
| 4   | [1,2,4,3,5,6] | 2   | [1,2, -1,2,1]     | 6             |

It can be observed that No. 1 and No. 2 roughly maintain the overall gradient, while No. 3 appears a large change, and No. 4 appears the opposite gradient. These four sets of data have the same L2 loss. The gradient loss better reflects how noises impact the gradient information. The bigger the gradient changes, the greater the loss value.

The gradient loss is the mean square for gradient difference vector length. It is more sensitive to points with larger gradient errors in the image and has a greater impact on edge detection and segmentation, while the points with smaller gradient errors do not. The drawback of gradient loss is that it cannot perceive the brightness change of the whole image, and it is insensitive to the synchronous change of the continuous areas. We combine the L2 loss and the gradient loss. The experiments in the Results and Discussion section show that our method works better than using these two loss functions alone.
List of abbreviations

AI: Artificial intelligence

CBCT: Cone beam computed tomography

MDCT: Multi detector computed tomography

LDCT: Low dose computed tomography

PSNR: Peak signal-to-noise ratio

MS-SSIM: Multi-level structural similarity

GMS: Gradient magnitude similarity

GMSD: Gradient magnitude similarity deviation

MSE: Mean Square Error

GAN: Generative Adversarial Networks

cGAN: Conditional Generative Adversarial Networks

SAGAN: Sharpness Aware Generative Adversarial Network

RED-CNN: Residual Encoder-Decoder Convolutional Neural Network

Declarations

Ethics approval and consent to participate

Not applicable.
Consent for publication

Not applicable.

Availability of data and material

The datasets used and analysed during the current study are available from the corresponding author on reasonable request.

Competing interests

The authors declare that they have no competing interests.

Funding

This work is supported by the Project of the Science & Technology Development Program of Jilin Province, China (No. 20200403127SF).

Authors’ contributions

Conceptualization, W.S. and Y.M.; methodology, W.S., Y.M., C.Z., and Y.D.; software, W.S. and L.S.; formal analysis, W.S. and Y.D.; validation, M.L. and W.W.; investigation, W.S., C.Z., and Y.D.; data curation, L.S.; writing – original draft preparation, W.S. and Y.M.; writing – review and editing, M.L. and W.W.; visualization, W.S., L.S., and M.L.; supervision, Y.M.; project administration, Y.D. and M.L. All authors read and approved the final manuscript.

Acknowledgements

Not applicable.
Authors’ information

W.S. is an Associate Professor in the School of Computer Science and Technology, Changchun University of Science and Technology, China. His main research theme is computer-aided diagnosis. He has published in the Computational and Mathematical Methods in Medicine, International Journal of Pattern Recognition and Artificial Intelligence, IEEE ACCESS, and other outlets.

Y.M. is an Associate Professor in the School of Computer Science and Technology, Changchun University of Science and Technology, China. His main research theme is computer-aided diagnosis. He has published in the Computational and Mathematical Methods in Medicine, International Journal of Pattern Recognition and Artificial Intelligence, IEEE ACCESS, and other outlets.

C.Z. studied for a master’s degree in the School of Computer Science and Technology of Changchun University of Science and Technology from 2019 to 2022.

Y.D. studied for a master’s degree in the School of Computer Science and Technology of Changchun University of Science and Technology from 2017 to 2019.

S.L. is a PhD Candidate in computer science and technology at Changchun University of Science and Technology, China. He also serves as a lecturer at the Department of Biomedical Engineering of the university. His main research theme is computer-aided diagnosis.

M.L. is a PhD Candidate at the Department of Information Systems and Business Analytics, RMIT University. He also serves as a sessional lecturer and project supervisor at the School of Computing and Information Systems and the Department
of Management and Marketing of The University of Melbourne. His main research
theme is AI-enabled digital transformation. His work has been published in journals
such as The Computer Journal and presented at conferences such as Australasian
Conference on Information Systems.

W.W. is a PhD Candidate at the Department of Software Systems and Cybersecurity,
Monash University. She also serves as a tutor and project supervisor at the School of
Computing and Information Systems of The University of Melbourne. Her main
research interests are software engineering and digital health. Her work has been
published in journals such as the International Journal of Environmental Research and
Public Health and presented at conferences such as Hawaii International Conference on
System Sciences.

References

[1] M. Li, W. Hsu, X. Xie et al., “SACNN: Self-Attention Convolutional Neural Network for
Low-Dose CT Denoising With Self-Supervised Perceptual Loss Network,” IEEE
Transactions on Medical Imaging, vol. 39, no. 7, pp. 2289-2301, 2020.

[2] O. Kupyn, V. Budzan, M. Mykhailych et al., “DeblurGAN: blind motion deblurring using
conditional adversarial networks,” IEEE Conference on Computer Vision and Pattern
Recognition (CVPR), pp. 8183–8192, 2018.

[3] Goodfellow I J, Pouget-Abadie J, Mirza M et al., “Generative Adversarial Networks,”
Advances in Neural Information Processing Systems, vol.3, pp. 2672-2680, 2014.

[4] P. Isola, J.-Y. Zhu, T. Zhou et al., “Image-to-image translation with conditional adversarial
networks,” IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp.
1125–1134, 2017.

[5] Yi X, Babyn P, “Sharpness-Aware Low-Dose CT Denoising Using Conditional
Generative Adversarial Network,” Journal of Digital Imaging, 2018.
[6] Cheng H, Liu J, Liu J et al., “Generating Synthesized CT from Cone-Beam Computed Tomography (CBCT) Using Artifact Disentanglement Network for Image-Guided Radiotherapy (IGRT),” Proceedings of 2020 Chinese Intelligent Systems Conference, vol. II, 2020.

[7] Lei Y, Wang T, Harms J et al., “CBCT-based synthetic MRI generation for CBCT-guided adaptive radiotherapy,” Workshop on Artificial Intelligence in Radiation Therapy. Springer, Cham, pp. 154-161, 2019.

[8] Y. Ma, B. Wei, P. Feng et al., “Low-Dose CT Image Denoising Using a Generative Adversarial Network With a Hybrid Loss Function for Noise Learning,” IEEE Access, vol. 8, pp. 67519-67529, 2020.

[9] J. Gu, J. C. Ye, “AdaIN-Based Tunable CycleGAN for Efficient Unsupervised Low-Dose CT Denoising,” IEEE Transactions on Computational Imaging, vol. 7, pp. 73-85, 2021.

[10] Chen H, Zhang Y, Kalra M K et al., “Low-Dose CT with a Residual Encoder-Decoder Convolutional Neural Network (RED-CNN),” IEEE Transactions on Medical Imaging, pp. 1-1, 2017.

[11] Arjovsky M, Chintala S, Bottou L. “Wasserstein generative adversarial networks,” International conference on machine learning. PMLR, pp. 214-223, 2017.

[12] Mirza M, Osindero S, “Conditional Generative Adversarial Nets,” Computer Science, pp. 2672-2680, 2014.

[13] Ronneberger O, Fischer P, Brox T, “U-Net: Convolutional Networks for Biomedical Image Segmentation,” 2015.

[14] He K, Zhang X, Ren S et al., “Deep residual learning for image recognition,” Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 770-778, 2016.

[15] Ledig C, Theis L, Huszár F et al., “Photo-realistic single image super-resolution using a generative adversarial network,” Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 4681-4690, 2017.

[16] Wang Z, Simoncelli E P, Bovik A C, “Multi-scale structural similarity for image quality assessment,” Proc IEEE Asilomar Conference on Signals, 2003.
[17] Xue W, Zhang L, Mou X et al., “Gradient Magnitude Similarity Deviation: A Highly Efficient Perceptual Image Quality Index,” IEEE Transactions on Image Processing, vol. 23, no. 2, pp. 684-695, 2014.

[18] Ezeme O M , Mahmoud Q H , Azim A . Design and Development of AD-CGAN: Conditional Generative Adversarial Networks for Anomaly Detection[J]. IEEE Access, 8.

[19] M. Arjovsky, S. Chintala, and L. Bottou. Wasserstein gan. arXiv preprint arXiv:1701.07875, 2017.