Parameter Constrained Transfer Learning for Low Dose PET Image Denoising

Yu Gong, Yueyang Teng*, Hongming Shan, Taohui Xiao, Ming Li, Guodong Liang, Ge Wang, Fellow, IEEE, and Shanshan Wang, Senior Member, IEEE

Abstract—Positron emission tomography (PET) is widely used in clinical practice. However, the potential risk of PET-associated radiation dose to patients needs to be minimized. With reduction of the radiation dose, the resultant images may suffer from noise and artifacts which compromises the diagnostic performance. In this paper, we propose a parameter-constrained generative adversarial network with Wasserstein distance and perceptual loss (PC-WGAN) for low-dose PET image denoising. This method makes two main contributions: i) a PC-WGAN framework is designed to denoise low-dose PET images without compromising structural details; and ii) a transfer learning strategy is developed to train PC-WGAN with parameters being constrained, which has major merits; namely, making the training process of PC-WGAN efficient and improving the quality of denoised images. The experimental results on clinical data show that the proposed network can suppress image noise more effectively while preserving better image fidelity than three selected state-of-the-art methods.

Index Terms—Deep learning, low-dose PET, parameter constraint, transfer learning, image quality

I. INTRODUCTION

Positron emission tomography (PET) is an advanced clinical imaging technology in the field of nuclear medicine. As a functional imaging technique, PET has many merits in neurology [1], oncology [2] and cardiology [3]. For a PET scan, a patient needs to be injected with some radioactive tracer; for example, fludeoxyglucose (F-18) which emits positrons that annihilate in the patient body to emit paired $\gamma$ photons. Less radioactive tracer means less cost and less risk, making PET scanning safer to patients and staff [4]. In recent years, researchers have tried to reduce the dose of radioactive tracers used in PET scans [5]. PET dose reduction follows the well-known guiding principle of ALARA (as low as reasonably achievable) [6]. Due to various physical degradation factors and low coincident-photon counts detected [7], reducing the dose of radioactive tracers affects the final image quality significantly. Therefore, advanced image processing algorithms are desirable to denoise low-dose PET images.

Classical PET image denoising algorithms can be divided into two categories: iterative reconstruction algorithms and image post-processing algorithms. An iterative reconstruction algorithm combines the statistical model of data noise as a regularization term to suppress the noise in a reconstructed image. For instance, Wang et al. [8] proposed a patch-based regularization method for iterative image reconstruction. Ehrhardt et al. [9] proposed randomized optimization for PET reconstruction aided by a large class of non-smooth priors including total variation, total generalized variation, various different physical constraints, and so on. Iterative reconstruction shows an excellent denoising ability but there are limitations as well; for example, iterative reconstruction is computationally intensive and may induce additional artifacts. On the other hand, image post-processing after reconstruction is computationally efficient compared to iterative reconstruction. Over the past years, there are many excellent image post-processing algorithms published, such as block-matching 3D [10] and non-local means [11]. Although denoising through image post-processing may improve the image quality substantially, over-smoothing and residual artifacts are often observed in a denoised image.

Recently, deep learning has achieved extraordinary results in the field of medical imaging, such as for correction [12], segmentation [13], reconstruction [4],[14], diagnosis [15], and so on. The statistical characteristics of noise in medical images are too complex and difficult to model. Deep learning can solve this problem very well due to its powerful learning ability to model image noise driven by big data. Therefore, deep learning based medical image noise reduction has led to state of the art results, clearly outperforming traditional methods. For example, Chen et al. [16] used a residual convolutional neural network to improve the quality of low-dose CT images greatly. Shan et al. [17] proposed the conveying path-based convolutional encoder-decoder network with 3D convolution (CPCE-3D) for low-dose CT denoising. Xie et al. [18] proposed to use a generative adversarial network (GAN) to produce remarkable denoising effects.

Currently, most denoising methods are in the 2D domain which means to utilize 2D features only. In reality, experienced radiologists look at only a single slice but also adjacent slices for analysis. For denoising, there is more prior information in the 3D domain than that in 2D domain [17]. Therefore, it is advisable to target 3D features for image denoising. However, 3D networks are more memory demanding and more
challenging in a training process than 2D networks [19].

In this paper, we propose a 3D parameter-constrained generative adversarial network with Wasserstein distance and perceptual loss (PC-WGAN) for low-dose PET images denoising. The generator of PC-WGAN is designed for noise reduction. It has a encoder-decoder structure as shown in Fig. 1. Typically, the 3D decoder network consists of 3D upsampling operators. In this paper, we use 3D deconvolution operators to replace 3D up-sampling operators. Similar to a 3D decoder network, the 2D decoder network consists of 2D deconvolution operators. 2D and 3D encoder networks consist of 2D and 3D convolution operators respectively. In the proposed network, we first use 3D convolution operators and then use 2D convolution operators to combine features in the 3D and 2D domains. It bridges the gap between the 3D and 2D feature spaces. To reduce the training difficulty of WGAN and improve the authenticity of denoised images, we use a transfer learning strategy in the training process. First, we use structural similarity index (SSIM) as the loss function to train the generator of the proposed network separately. In the next training phase, we use the parameter of the trained model to initialize the generator without involving the Xavier initializer [20]. Our initialization method constrains the parameters of the proposed network to reduce the training difficulty of the proposed network and regularize the denoising process heuristically. The experimental results on clinical images show that the proposed network can suppress more noise while preserving more details than three state-of-the-art methods.

The contributions of this paper are as follows:

i) A network PC-WGAN is designed for low-dose PET image denoising. It maps the distribution of low-dose PET images to that of normal-dose images to reduce noise and keep details as much as feasible. In the proposed network, the use of mixed 2D and 3D convolution and deconvolution operators is effective in synergizing 2D and 3D features and improving the fidelity of denoised images;

ii) A transfer learning strategy is adapted to constrain the parameters of PC-WGAN and improve the quality of denoised images.

II. METHODS

A. Low Dose PET Denoising

Let $x$ denote a low-dose PET image and $y$ denote the corresponding normal-dose image. Seeking a function $G(\cdot)$ which maps the low-dose PET image to the normal-dose image is the task of denoising process.

$$G : x \rightarrow y$$

If we take $x$ as a sample from the low-dose PET image distribution $p_L$ and $y$ from the normal-dose PET image distribution $p_N$. The function $G(\cdot)$ maps samples from $p_L$ into a certain denoising distribution $p_g$. We could make $p_g$ close to $p_N$ by varying the function $G(\cdot)$. In other words, we treat the denoising process as the convertor which could move one data distribution to another.

B. PC-WGAN

As shown in Eq. (1), the main task for low-dose PET images denoising is to construct the function $G(\cdot)$. Motivated by it, we use deep learning with powerful nonlinear fitting ability. Specially, we design a network named PC-WGAN to approximate $G(\cdot)$.

The introduction of PC-WGAN consists of two parts: network structure and objective function.

1) Network Structure: The overall structure of the proposed network is shown in Fig. 2. There are three parts in it.
The first part is the generator. Although it has the similar shape to U-Net [21] and V-net [22], the difference is to use mixed 2D and 3D convolution operator instead of pure 2D convolution operators in U-net or pure 3D convolution operators in V-net. Denoising process of the generator is shown in Fig. 1. After stacking, several continuous low-dose PET images processed by 3D convolution, 2D convolution, 2D deconvolution and 3D deconvolution operators in turn to produce denoising images. The architecture of the generator is shown in Fig 4. There are sixteen layers in the generator include four 3D convolutional layers, four 2D convolutional layers, four 2D deconvolutional layers and four 3D deconvolutional layers. Similar to that discarded detail caused by the down-sampling of 2D convolution could be recovered by 2D deconvolution, we used 3D deconvolution to recover the discarded detail caused by 3D convolution. Since there are sixteen layers in the proposed network, we introduced the residual compensation [16] similar to deep residual learning [23], [24]. It can prevent the training difficulty caused by the gradient diffusion.

The outputs between the layers of the generator follow rule of LIFO (Last In First Out). For example, the output of the first layer was superimposed with the output of the last layer to obtain the final result. This shortcut connection was used except the 4th 2D convolution layer. The number of kernels of each layer of the generator is shown in Table I. Following the common practice in the deep learning community [25], small 3×3 kernels were used in each convolutional and deconvolutional layer. The stride of convolution and deconvolution was a constant value of 1. Zero-padding was not used in the network. ReLU was used as the activation function after each layer.

TABLE I

| Layer | # Kernel | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
|-------|----------|---|---|---|---|---|---|---|---|
|       |          | 64| 64| 128|128|256|256|512|512|
| Layer |          | 9 | 10| 11 |12 |13 |14 |15 |16 |
|       |          | 512|256|256|128|128|64 |64 |1  |

The second part is discriminator. There are eight layers in discriminator including six 2D convolution layers and two fully connected layers. Leaky ReLU was chosen as the activation function for the discriminator and it was used for each layer except the last layer. Small 3×3 kernels were used in discriminator too. The convolutional stride of the odd layer was constant 1 and the convolutional stride of the even layer was constant 2. The number of kernels and outputs of each layer is shown in Table II. The WGAN [26] framework consists of the generator and the discriminator described above.

The third part is perceptual feature extractor. Similar to [27], [28], we use pre-trained VGG-19 network as the perceptual feature extractor. There are sixteen convolutional layers in VGG-19 and we choose the output of the 16th convolutional layer as the extracted perceptual feature. The generated image from the generator and the corresponding normal-dose image are sent into VGG-19 for extracting features in the high-dimensional feature space.

2) Objective Function: Shan et al. [17] and Xie et al. [18] proved that the potential denoising ability of GAN is better than CNN. Inspired by their excellent results, we optimize the proposed network in the framework of WGAN. In addition to the adversarial loss, we use additional perceptual loss into the objective function.

Adversarial loss is introduced by the loss function of WGAN. Arjovsky et al. [26] used Wasserstein distance to estimate the difference between \( p_g \) and \( p_N \). The generator of WGAN is the function \( G(\cdot) \) that we are looking for. Let \( D(\cdot) \) denote the discriminator. The definition of Wasserstein distance is shown as follows:

\[
W(p_g, p_N) = \mathbb{E}_{y \sim p_N}[D(y)] - \mathbb{E}_{\tilde{x} \sim p_g}[D(\tilde{x})] \tag{2}
\]

where \( \mathbb{E}_a \) denotes the expectation when variable \( a \) follows distribution \( b \). Gradient penalty [29] was introduced to improve the stability of WGAN. The loss function of WGAN is shown as Eq. [3]

\[
L_{WGAN} = \mathbb{E}_{y \sim p_N}[D(y)] - \mathbb{E}_{\tilde{x} \sim p_g}[D(\tilde{x})] - \lambda \mathbb{E}_{\tilde{x} \sim p_a}[\|\nabla_{\tilde{x}} D(\tilde{x})\|_2 - 1]^2 \tag{3}
\]

Here, \( \lambda \) is a constant weighting parameter of gradient penalty. \( \tilde{x} = \epsilon \cdot \tilde{x} + (1 - \epsilon) \cdot y \) denotes uniformly sampling along straight lines connecting pairs of denoising images and normal-dose images, \( \epsilon \) is an interval of \( [0, 1] \) and \( p_a \) is a distribution for \( \tilde{x} \).

The perceptual similarity measure proposed in [30] and [27] compares the difference between images using the feature extracted by perceptual extractor in the high-dimensional feature space rather than pixel space. The perceptual loss is then defined as:

\[
L_{Perceptual} = \frac{1}{N} ||\phi(\tilde{x}) - \phi(y)||_2^2 \tag{4}
\]

where \( N \) denotes the batch size and \( \phi(\cdot) \) denotes the perceptual feature extracting process. Perceptual loss helps to prevent over-smoothing, blurred edges, etc. in the generated images which helps the proposed network to generate the images meeting the visual requirement. In addition, perceptual loss has a similar effect to regularization term. It limits the generation ability of WGAN. This limitation prevents the generation of...
sham texture to ensure the credibility of the denoising image generated by WGAN.

The objective function of the discriminator is defined as follows:

$$L_D = \arg \min_D \mathbb{E}_{\tilde{x} \sim p_g}[D(\tilde{x})] - \mathbb{E}_{y \sim p_N}[D(y)] + \lambda \mathbb{E}_{\hat{x} \sim p_{\hat{x}}}[\left(\|\nabla_{\hat{x}} D(\hat{x})\|_2 - 1\right)^2]$$  \hspace{1cm} (5)

The objective function of the generator is defined as follows:

$$L_G = \arg \min_G - \mathbb{E}_{\tilde{x} \sim p_g}[D(\tilde{x})] + \lambda_p L_{Perceptual}$$  \hspace{1cm} (6)

where $\lambda_p$ denotes the constant weighting parameter of perceptual loss.

C. Transfer Learning Strategy

Although the use of Wasserstein distance and gradient penalty reduces the training difficulty of GAN, the convergence problem in the training process is still not solved completely. Transfer learning is generally defined as the ability of a system to utilize knowledge learned from one task to another task that shares some common characteristics [31]. Shan et al. [17] proposed to transfer a pre-trained CPCE-2D model to get a CPCE-3D model. It improved the denoising performance and reduced the training difficulty of CPCE-3D. It is obvious that there are many common characteristics between the networks with the same structure and target but training by different loss function. In consideration of the generator of the proposed network is a simple CNN network that is designed to use low-dose PET images to generate denoising images which are as close to the corresponding normal-dose images as possible. Whether it is under the framework of GAN. In addition, the training process of a single CNN network is more simple than that of GAN. Therefore, we trained the generator without the framework of GAN as the pre-trained model. It improves the efficiency of pre-trained process and is also the difference with the work of Shan et al. [17].

Firstly, we tried a number of loss functions, including mean square error (MSE), SSIM, perceptual loss and the mixture of them to train the generator of PC-WGAN separately. Then the parameters of the trained generator were used to initialize the generator in the joint training process of PC-WGAN instead of Xavier initializer [20]. Due to the training difficulty of CNN is low, this kind of initialization is efficient. It provides a good starting point to prevent PC-WGAN to fall into mode collapse. The training stability of the proposed network has been improved by doing so.

On the other hand, GAN was originally designed to produce the image that mimics the real thing. Here comes a problem, if the generator of GAN gets the best performance, it would produce the non-existing texture which can confuse the doctor. It is totally unacceptable for medical image. Similar to the regularization term used in iterative reconstruction, we could use transfer learning strategy to limit PC-WGAN which prevents the generation of non-existing texture. The different pre-trained model we use, the different final network we get.

III. EXPERIMENTS

A. Experimental Data

We used 2,920 pairs of 256×256 clinical PET images from 8 anonymous patients scanned by Neusoft NeuSight EWN. The images with a scanning time of 75s were used as the low-dose PET images and the images with a scanning of 150s were used as the normal-dose PET images. We selected 2,190 pairs of PET images from 6 patients randomly as training data sets and 730 pairs of PET images of the remaining 2 patients as validation sets. Due to the obtainment of medical big data is difficult, we used the following operations to augment the data set and saved computational resources: randomly crop 43,800 pairs patches of size 64×64 from the training data and perform rotation and contrast transformation operations on the input of proposed network.
B. Implementation Details

Just like the training process of other GANs, we trained $G(\cdot)$ and $D(\cdot)$ separately by fixing one and updating the other. We used Adam algorithm [32] to optimize the proposed network. The hyper parameters of Adam algorithm were set as $\alpha = 1e^{-4}$, $\beta_1 = 0.9$ and $\beta_2 = 0.999$. The weight of the gradient penalty $\lambda$ is fixed at 10 as suggested in [29]. The weight of the perceptual loss $\lambda_p$ is fixed at 10. The number of iterations is 20,000 times. The networks were implemented in Python 3.6 with the TensorFlow 1.4 [33]. A NVIDIA QUADRO M5000 GPU was used.

C. Pre-trained Network

We used different loss functions to train the generator of PC-WGAN separately. We designed a total of five loss functions. Except $L_{\text{Perceptual}}$ is shown in Eq. (4), the other loss functions are expressed as follows:

$$L_{\text{MSE}} = \frac{1}{N} \| \tilde{x} - y \|_2^2$$
$$L_{\text{SSIM}} = \frac{1}{N} \frac{2\mu_{\tilde{x}}\mu_y + c_1}{(\mu_{\tilde{x}}^2 + \mu_y^2 + c_1)(\sigma_{\tilde{x}}^2 + \sigma_y^2 + c_2)}$$
$$L_{\text{SSIM,MSE}} = \sqrt{1 + L_{\text{MSE}} \cdot (1 - L_{\text{SSIM}})}$$

where $\mu_a$ denotes the mean of $a$, $\sigma^2_a$ denotes the variance of $a$ and $\sigma_{ab}$ denotes the covariance of $a$ and $b$. $c_1 = (k_1L)^2$ and $c_2 = (k_2L)^2$ are constant to maintain stability, where $L \in [0, 1]$, $k_1 = 0.01$ and $k_2 = 0.03$.

In order to demonstrate the denoising results intuitively, we present them in the way of 3D reconstruction. Denoising results of networks with different loss function for the test sets are shown in Fig. 5. The zoomed abdomen area of the patient’s PET image is shown in Fig. 6 for viewing the texture detail.

First of all, it is clear that the quality of the low-dose PET image is much lower than that of the normal-dose PET image. There is much noise in the low-dose PET image which affects the observation of patient’s organs and tissues. Overall, each network achieved different degrees of denoising effect for the low-dose PET image. As shown in Fig. 5(b) the denoising effect of the network with MSE as the loss function (M) is excessive. There is little residual noise in the denoising result by using MSE as the loss function. However, a lot of structural details were discarded. As can be seen from Fig. 5(b) is that the texture of the tissue is loss seriously. The SSIM value between the low-dose image and the normal-dose image is higher than 0.95. Although the optimization space for the network with SSIM as the loss function (S) is small, the denoising result is significant as shown in Fig. 5(c). As shown in Fig. 5(c), the problem of over-smoothing in denoising images has been suppressed by using SSIM as the loss function. There are many texture details preserved after denoising. The denoising result of the network with perceptual loss as the loss function (P) has the best visual effect compared to the denoising result of the other networks. As shown in Fig. 5(d), it looks the closest to the normal-dose PET image. The cost of visual improvement is that there is much residual noise in the denoising image.

In order to combine the advantage of MSE, SSIM and perceptual loss, we combined them to form loss functions. As can be seen in the Fig. 5(e), simultaneous optimization of MSE and SSIM suppressed excessive denoising and relieved the loss of the detail of the denoising image. Compared to the denoising result of the network with mixed MSE and SSIM as the loss function (MS), the visual performance of the network with mixed MSE, SSIM and perceptual loss (MPS) is better. However, residual noise can be observed in the denoising result as shown in Fig. 5(f). It is similar to the result of P.

According to the denoising result of networks with different loss functions, we can summaries the following rules. First, MSE as the loss function guarantees the denoising ability of the network. Second, SSIM as the loss function makes the boundary of the images generated by the network sharpen. Third, perceptual loss improves the visual effect of denoising images. For a pre-trained model, we concern more about the detail protection than noise suppression since it could be adjusted after transferring to the framework of WGAN. Therefore, we chose S as the pre-trained model for transfer learning.

D. Parameter Analysis

Although perceptual loss is beneficial to improve the visual effect of the denoising image. The residual noise is often observed in the denoising image. Therefore, while introducing perceptual loss as the regularization, we introduce $\lambda_p$ as the
weight to control it. When \( \lambda_p \) is set to 0, it means that perceptual loss is not used in the training process of the proposed network. As \( \lambda_p \) increases, the effect of perceptual loss will be more and more obvious. If \( \lambda_p \) is set to \( \infty \), there is no use of adversarial loss. We set \( \lambda_p \) to 0, 0.01, 0.1, 1, 10 and \( \infty \) to train models and then apply them to the test data sets for parameter analysis. Due to the denoising effect of the network with additional perceptual loss, we chose Peak Signal to Noise Ratio (PSNR), SSIM and MSE as the indicators for evaluating the quality of the generated image instead of evaluating it for image. PSNR is based on error sensitivity to evaluate the image quality. A bigger PSNR relates to the less noise ratio of the image. SSIM is an index that measures the similarity between two images. PSNR is based on error sensitivity to evaluate the image quality. A bigger PSNR relates to the less noise ratio of the image. SSIM is an index that measures the similarity between two images. When it equals to 1, two images are completely identical. Similarly, when it equals to 0, two images are completely different. MSE calculates the mean square value of image pixel difference between the noisy image and the denoising image. Normalized the indicators of two test sets was calculated and shown in Figs. 7 and 8. It is obvious that network, it results in much residual noise in the denoising image. It is obvious that the denoising performance of mixed adversarial loss and perceptual loss is better than that of adversarial loss or perceptual loss alone. As shown in Figs. 7 and 8, the denoising images generated by the network with \( \lambda_p = 10 \) has the highest PSNR, SSIM and lowest MSE value compared with that of the other network. As discussed above, \( \lambda_p \) in the proposed network be assigned to 10.

E. Denoising results

In order to demonstrate the denoising performance of the proposed network, we also trained RED-CNN [16], 200x-CNN [34] and CPCE-3D [17]. In addition, the performance of PC-WGAN without transfer learning (PC-WGAN-D) was compared with PC-WGAN with transfer learning (PC-WGAN-T). The details of the above networks is shown in Table III.

![Diagram](image)

Fig. 6. Zoomed ROI of red rectangle in Fig.5. (a) low-dose image, (b) the network with MSE as the loss function (M), (c) the network with perceptual loss as the loss function (P), (d) the network with mixed MSE and SSIM as the loss function (MS), (f) the network with mixed MSE, perceptual loss and SSIM as the loss function (MSP) and (g) normal-dose image.

![Diagram](image)

Fig. 7. Normalized indicators of networks with different \( \lambda_p \in \{0, 0.01, 0.1, 1, 10, \infty \} \) on test set 1.

![Diagram](image)

Fig. 8. Normalized indicators of networks with different \( \lambda_p \in \{0, 0.01, 0.1, 1, 10, \infty \} \) on test set 2.

| Name         | Number of layers | Size of input | Number of input |
|--------------|------------------|---------------|-----------------|
| RED-CNN      | 10               | 1 × 64 × 64   | 43,800          |
| 200x-CNN     | 15               | 3 × 64 × 64   | 43,800          |
| CPCE-3D      | 11               | 9 × 64 × 64   | 43,800          |
| PC-WGAN-D    | 16               | 9 × 64 × 64   | 43,800          |
| PC-WGAN-T    | 16               | 9 × 64 × 64   | 43,800          |
effect of the proposed network since there are many structural details in it.

All networks demonstrated significant denoising effect as shown in Figs. 9 and 11. Both of 200x CNN and RED-CNN use a structure of encoder-decoder and residual compensation. The denoising effect of RED-CNN is obvious. Although there is less noise in the denoising image than that in the low-dose PET image, excessive denoising makes many structural details loss in the denoising image as shown in Figs. 10(b) and 12(b). It is caused by using MSE as the loss function. Xu et al. [34] used l1-norm instead of MSE as the loss function of 200x-CNN and added adjacent two slices for denoising. The denoising result of 200x-CNN shows that the over-smoothing phenomenon of the organ is suppressed. As shown in Figs. 10(c) and 11(c), the boundary of the organ in the denoising result of 200x-CNN is clearer than RED-CNN. The shortcoming of using l1-norm as loss function is that there is a little residual noise in the denoising image. Compared with the normal-dose image, there is some room for improvement of texture preservation of REC-CNN and 200x-CNN. The tissue shown in Figs. 10(g) and 12(g) has fine texture. Texture loss means that minor pathological changes may be difficult to detect.

Compared with RED-CNN and 200x-CNN, CPCE-3D improves the quality of the denoising image by using adversarial loss and 3D convolution. Overall, the visually effect of the organ in the denoising result of CPCE-3D is more close to the normal-dose image than those of RED-CNN and 200x-CNN. As shown in Figs. 9(d) and 11(d), the boundary of organ in the denoising result of CPCE-3D is clearly. It proves that adversarial loss and 3D convolution is helpful to improve denoising ability. However, the tiny distortion of the shape can be seen in Figs. 10(d) and 12(d). Since both RED-CNN and CPCE-3D are networks for CT image denoising, the experimental results show that their denoising effects for clinical PET images are not ideal.

The denoising effect of PC-WGAN-D has a certain improvement compared with the denoising effect of CPCE-3D by introducing 3D deconvolution. As shown in Figs. 10(e) and 12(e), the tiny distortion of the shape in the denoising of CPCE-3D is suppressed. We can easily find that the tissue in the zoomed ROI images generated by PC-WGAN has more texture details and the boundaries are much clearer than that generated by the other networks which confirms that 3D deconvolution can help PET image denoising networks protect the detail of the denoising image. The denoising result of PC-WGAN-D proves that the structure of the proposed network is suitable for low-dose PET image denoising. Compared with 200x-CNN, PC-WGAN-D processes nine images at the same time so the information obtained is much more than three images so the denoising result of PC-WGAN-D is better than 200x-CNN. As can be seen from Figs. 10(e) and 12(e), the texture in the denoising result of PC-WGAN-D is more close to the normal-dose PET image than those of RED-CNN, 200x-CNN and CPCE-3D. There is much structural information in the denoising image.

The PC-WGAN model obtained by transfer learning is the best denoising network in this paper. From a visual point of view, the image generated by PC-WGAN-T is the closest to the normal-dose PET image. Thanks to the effect of transfer learning, the denoising result has improved significantly than PC-WGAN-D. More structural information in the denoising image than that of the other networks. As shown in Figs. 10(f) and 12(f), the boundaries and textures of the tissue in the image using the proposed network are excellent. Due to the use of perceptual loss, the denoising image of PC-WGAN-T has nice visual performance. In addition, we don’t find that PC-WGAN-T produces the texture which doesn’t exist in the low-dose PET image. It is the benefit of the use of transfer learning and perceptual loss.
Fig. 11. PET image of the patients head denoising by different networks. (a) Low-Dose Image, (b) RED-CNN, (c) 200x-CNN, (d) CPCE-3D, (e) PC-WGAN-D, (f) PC-WGAN-T and (g) Normal-Dose Image.

Fig. 12. Zoomed ROI of red rectangle in Fig.11. (a) Low-Dose Image, (b) RED-CNN, (c) 200x-CNN, (d) CPCE-3D, (e) PC-WGAN-D, (f) PC-WGAN-T and (g) Normal-Dose Image.

In addition to the analysis from the visual perspective of the denoising image, we chose PSNR, SSIM, Standard Deviation (SD) and MEAN as the indicators for evaluating the quality of the generated image. We calculated average of PSNR, SSIM, MEAN and SD of the images generated from the test data set. In terms of the two indicators of PSNR and SSIM, our network achieved the best results as can be seen in Table IV. Although the SSIM between the low-dose image and the normal dose image in this subject is high, but the image generated by the proposed network still improves SSIM based on the low-dose image. The most important point is that the denoising effect of the proposed network is stabler than other networks.

The average of MEAN and SD of denoising images is shown in Table V. Regardless of MEAN or SD, we try to make the generated image as close to the normal dose image as possible. As shown in Table V, the value of MEAN and SD of the image generated by the proposed network is the closest to that of the normal dose image. The best MEAN means that both noise reduction and detail preservation are excellent in the denoising process of PC-WGAN-T. The best SD means that the smoothness of the denoising image is close to the normal-dose image. Over-smoothing phenomenon has been suppressed in the result of PC-WGAN-T.

IV. DISCUSSIONS AND CONCLUSION

The novel features of the proposed network include the well-designed structure aided by detail-capturing loss terms and constrained parameters through transfer learning.

It is important to decrease the loss of details while denoising low-dose PET images. We have introduced 3D deconvolutional operators to recover the details discarded by 3D convolution. Then, we have adjusted the shortcut connections between the corresponding layers to make the best trade-off between noise suppression and detail preservation. There are not only 3D but also 2D convolutional operators. It is shown practical to use both of 3D contextual and planar information for low-dose PET images denoising.

Generating a normal-dose image from a low-dose image is the ultimate goal. However, achieving it perfectly is impossible, so in practice we only try to generate the image that is as close to the normal-dose image as possible. Denoising results of [17] and [18] give an excellent quality, taking the

### TABLE IV

| Methods  | Low-Dose | RED-CNN | 200x-CNN | CPCE-3D | PC-WGAN-D | PC-WGAN-T | \(\text{PSNR}\) |
|----------|----------|---------|----------|---------|-----------|-----------|-----------------|
|          | 35.6716  | 39.3597 | 39.997   | 38.7194 | 39.6548   | 40.691    |                 |
| \(\text{SSIM}\) | 0.978766 | 0.986028 | 0.974495 | 0.984793 | 0.986349 | 0.987976 |                 |

### TABLE V

| Methods  | Low-Dose | RED-CNN | 200x-CNN | CPCE-3D | PC-WGAN-D | PC-WGAN-T | Normal-Dose |
|----------|----------|---------|----------|---------|-----------|-----------|-------------|
| MEAN     | 5.61279  | 6.60765 | 7.21692  | 7.15337 | 6.83989   | 7.01298   | 6.71938     |
| SD       | 19.6496  | 22.752  | 24.5644  | 23.8008 | 23.2165   | 23.8092   | 23.0788     |
advantages of GAN. However, the authenticity of the images generated by GAN can be problematic if not well constrained. In addition, the training of GAN is very challenging. To address these issues, we have used a transfer learning strategy coupled with the perceptual loss, and demonstrated a success.

In transfer learning, we use the parameters of a pre-trained CNN model to initialize the parameters of PC-WGAN. The reasonable starting point transferred from the pre-trained model makes the training process of PC-WGAN well informed. Its effect is similar to that of regularization. Such a constraint applied on PC-WGAN can help prevent it from producing non-existing texture, since we can control the performance of PC-WGAN through selecting the pre-trained model. On the other hand, the perceptual loss not only improves the visual performance but also guarantees the authenticity of denoised images.

While the training difficulty of CNN is low, that of GAN is high. The transfer learning strategy overcomes the training difficulty of GANs. Based on common characteristics of underlying images, we transfer a pre-trained CNN model to GAN. It makes the training process of GAN efficient. This kind of transfer learning is efficient and heuristic. The parameters of the pre-trained network proves a good starting point that is close to the optimal point for PC-WGAN and makes the optimization easy. It is worth noting that the performance of the proposed network with knowledge transferred from the pre-trained network is better than that of the proposed network trained directly. The experimental results described above have demonstrated that the use of transfer learning can reduce the training difficulty and improve the performance of the proposed network simultaneously.

We believe that the effect of network-based image denoising is related to the width and depth. How to optimize the the network topology is worthy to further study. With the development of GAN, more and more well-structured GAN networks have emerged. Among them, WGAN is not necessarily the best. Therefore, there is still much room for development of denoising methods for low-dose PET imaging by developing more advanced networks than what has been proposed here. Finally, the performance of supervised PET image denoising via deep learning critically depends on the amount of labeled data, and self-supervised learning can help further. The use of self-supervised learning for PET image denoising will be our next step.

In conclusion, we have proposed a parameter-constrained generative adversarial network with Wasserstein distance and perceptual loss (PC-WGAN) for low-dose PET image denoising. The experimental results on clinical data show that the proposed network can suppress image noise more effectively while preserving better image fidelity than than three selected state-of-the-art methods. Further work is in progress for improvements in the performance of low-dose PET denoising.

V. ACKNOWLEDGMENT

This work was supported by the National Natural Science Foundation of China (61601450, 61871371, 8183005661671441), Natural Science Foundation of Liaoning Province of China (20170540321), Science and Technology Planning Project of Guangdong Province (2017B020227012, 2018B010109009), the Basic Research Program of Shenzhen (JCYJ20180507182400762) and Youth Innovation Promotion Association Program of Chinese Academy of Sciences (2019351). The authors wish to express gratitude to the Shenyang Neusoft-medical and the patients who provided data.

REFERENCES

[1] R. N. Gunn, M. Slifstein, G. E. Searle, and J. C. Price, “Quantitative imaging of protein targets in the human brain with PET,” *Physics in Medicine & Biology*, vol. 60, no. 22, p. R363, 2015.
[2] T. Beyer, D. W. Townsend, T. Brun, P. E. Kinahan, M. Charron, R. Roddy, J. Jerin, J. Young, L. Byars, R. Natt et al., “A combined PET/CT scanner for clinical oncology,” *Journal of nuclear medicine*, vol. 41, no. 8, pp. 1369–1379, 2000.
[3] J. Machac, “Cardiac positron emission tomography imaging,” in *Seminars in nuclear medicine*, vol. 35, no. 1. Elsevier, 2005, pp. 17–36.
[4] C. Wang, Z. Hu, P. Shi, and H. Liu, “Low dose PET reconstruction with total variation regularization,” in *2014 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*. IEEE, 2014, pp. 1917–1920.
[5] F. Fahey, G. El Fakhri, and M. King, “Novel approaches to dose optimization in nuclear medicine,” in *MEDICAL PHYSICS*, vol. 46, no. 6. WILEY 111 RIVER ST, HOBOKEN 07030-5774, NJ USA, 2019, pp. E285–E283.
[6] D. J. Ettenner and E. J. Hall, “Computed tomographyan increasing source of radiation exposure,” *New England Journal of Medicine*, vol. 357, no. 22, pp. 2277–2284, 2007.
[7] K. Gong, J. Guan, K. Kim, X. Zhang, J. Yang, Y. Seo, G. El Fakhri, J. Qi, and Q. Li, “Iterative PET image reconstruction using convolutional neural network representation,” *IEEE transactions on medical imaging*, vol. 38, no. 3, pp. 675–685, 2018.
[8] W. Wang and J. Qi, “ Penalized likelihood PET image reconstruction using patch-based edge-preserving regularization,” *IEEE transactions on medical imaging*, vol. 31, no. 12, pp. 2194–2204, 2012.
[9] M. J. Ehhardt, P. Markiewicz, and C.-B. Schönlieb, “Faster PET reconstruction with non-smooth priors by randomization and preconditioning,” arXiv preprint arXiv:1808.07150, 2018.
[10] K. Babov, A. Foi, V. Katkovnik, and K. Egiaziarian, “Image denoising with block-matching and 3D filtering,” in *Image Processing: Algorithms and Systems, Neural Networks, and Machine Learning*, vol. 6064. International Society for Optics and Photonics, 2006, p. 606414.
[11] A. Buades, B. Coll, and J.-M. Morel, “A non-local algorithm for image denoising,” in *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR’05)*, vol. 2. IEEE, 2005, pp. 60–65.
[12] K. Gong, J. Yang, K. Kim, G. El Fakhri, Y. Seo, and Q. Li, “Attenuation correction of PET/MR using cycle-consistent adversarial network,” *Journal of Nuclear Medicine*, vol. 60, no. supplement 1, pp. 171–171, 2019.
[13] L. Zhao, Z. Lu, J. Jiang, Y. Zhou, Y. Wu, and Q. Feng, “Automatic nasopharyngeal carcinoma segmentation using fully convolutional networks with auxiliary paths on dual-modality PET-CT images,” *Journal of digital imaging*, vol. 32, no. 3, pp. 462–470, 2019.
[14] K. Gong, D. Wu, K. Kim, J. Yang, T. Sun, G. El Fakhri, Y. Seo, and Q. Li, “Mapem-net: an unrolled neural network for fully 3D PET image reconstruction,” in *15th International Meeting on Fully Three-Dimensional Image Reconstruction in Radiology and Nuclear Medicine*, vol. 11072. International Society for Optics and Photonics, 2019, p. 1107200.
[15] Y. Xie, Y. Yu, T. Thamm, E. Gong, J. Ouyang, C. Huang, S. Christensen, G. Albers, and G. Zaharchuk, “Deep learning-based penumbra estimation using DWI and ASL for acute ischemic stroke patients,” *Journal of Cerebral Blood and Metabolism*, vol. 39, pp. 622–622, 2019.
[16] H. Chen, Y. Zhang, M. K. Kalra, F. Lin, Y. Chen, P. Liao, J. Zhou, and G. Wang, “Low-dose ct with a residual encoder-decoder convolutional neural network,” *IEEE transactions on medical imaging*, vol. 36, no. 12, pp. 2524–2535, 2017.
[17] H. Shan, Y. Zhang, Q. Yang, U. Kruger, M. K. Kalra, L. Sun, W. Cong, and G. Wang, “3-D convolutional encoder-decoder network for low-dose CT via transfer learning from a 2-D trained network,” *IEEE transactions on medical imaging*, vol. 37, no. 6, pp. 1522–1534, 2018.
[18] Z. Xie, R. Baikejiang, K. Gong, X. Zhang, and J. Qi, “Generative adversarial networks based regularized image reconstruction for PET,” in 15th International Meeting on Fully Three-Dimensional Image Reconstruction in Radiology and Nuclear Medicine, vol. 11072. International Society for Optics and Photonics, 2019, p. 110720P.

[19] T. Ni, L. Xie, H. Zheng, E. K. Fishman, and A. L. Yuille, “Elastic boundary projection for 3D medical image segmentation,” in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2019, pp. 2109–2118.

[20] X. Glorot and Y. Bengio, “Understanding the difficulty of training deep feedforward neural networks,” in Proceedings of the thirteenth international conference on artificial intelligence and statistics, 2010, pp. 249–256.

[21] O. Ronneberger, P. Fischer, and 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.

[22] F. Milletari, Navab, and S.-A. Ahmadi, “V-net: Fully convolutional neural networks for volumetric medical image segmentation,” in 2016 Fourth International Conference on 3D Vision (3DV). IEEE, 2016, pp. 565–571.

[23] K. He, X. Zhang, S. Ren, and J. Sun, “Deep residual learning for image recognition,” in Proceedings of the IEEE conference on computer vision and pattern recognition, 2016, pp. 770–778.

[24] R. K. Srivastava, K. Greff, and J. Schmidhuber, “Training very deep networks,” in Advances in neural information processing systems, 2015, pp. 2377–2385.

[25] S. Srinivas, R. K. Sarvadevabhatla, K. R. Mopuri, N. Prabhu, S. S. Kruthiventi, and R. V. Babu, “An introduction to deep convolutional neural nets for computer vision,” in Deep Learning for Medical Image Analysis. Elsevier, 2017, pp. 25–52.

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

[27] J. Johnson, A. Alahi, and L. Fei-Fei, “Perceptual losses for real-time style transfer and super-resolution,” in European conference on computer vision. Springer, 2016, pp. 694–711.

[28] Q. Yang, P. Yan, Y. Zhang, H. Yu, Y. Shi, X. Mou, M. K. Kalra, Y. Zhang, L. Sun, and G. Wang, “Low-dose CT image denoising using a generative adversarial network with wasserstein distance and perceptual loss,” IEEE transactions on medical imaging, vol. 37, no. 6, pp. 1348–1357, 2018.

[29] I. Gulrajani, F. Ahmed, M. Arjovsky, V. Dumoulin, and A. C. Courville, “Improved training of wasserstein gans,” in Advances in neural information processing systems, 2017, pp. 5767–5777.

[30] A. Dosovitskiy and T. Brox, “Generating images with perceptual similarity metrics based on deep networks,” in Advances in neural information processing systems, 2016, pp. 658–666.

[31] S. J. Pan and Q. Yang, “A survey on transfer learning,” IEEE Transactions on knowledge and data engineering, vol. 22, no. 10, pp. 1345–1359, 2009.

[32] D. P. Kingma and J. Ba, “Adam: A method for stochastic optimization,” arXiv preprint arXiv:1412.6980, 2014.

[33] M. Abadi, A. Agarwal, P. Barham, E. Brevdo, Z. Chen, C. Citro, G. S. Corrado, A. Davis, J. Dean, M. Devin et al., “Tensorflow: Large-scale machine learning on heterogeneous distributed systems,” arXiv preprint arXiv:1603.04467, 2016.

[34] J. Xu, E. Gong, J. Pauly, and G. Zaharchuk, “200x low-dose PET reconstruction using deep learning,” arXiv preprint arXiv:1712.04119, 2017.