



Deep Learning Based Framework for Direct Reconstruction of PET Images

Zhiyuan Liu, Huai Chen, and Huafeng Liu (✉)

State Key Laboratory of Modern Optical Instrumentation, College of Optical Science and Engineering, Zhejiang University, Hangzhou, China
liuhf@zju.edu.cn

Abstract. In Positron Emission Tomography (PET), high radioactivity maps are essential to better understand the physiological processes associated with the disease. In this paper, we propose a deep learning based framework for PET image reconstruction from sinogram domain directly. In the framework, conditional Generative Adversarial Networks (cGANs) is constructed to learn a mapping from sinogram data to reconstructed image and generate a well-trained model. To verify the accuracy and robustness of the model, both Monte Carlo simulation data and real data are adopted in the test. The experimental results show that the proposed framework is of great robustness and the reconstructed image is much more accurate in comparison with the traditional methods.

Keywords: PET image · Direct reconstruction · cGANs

1 Introduction

Positron emission tomography (PET) is a nuclear medicine imaging technique, which has become an indispensable tool in clinical trials and medical practice. PET image reconstruction is essential and is to convert photon counting data (projection data) into radioactivity concentration distribution, which is an ill-posed inverse problem [1].

There have been plenty of studies on various iterative strategies based on statistical data model, such as maximum likelihood-expectation maximum (MLEM) [2], maximum a posteriori (MAP) [3], penalized weighted least square (PWLS) [4] and total variation (TV) [5], which have strong implications on the reliability of measured or derived system response, often called system matrix. In addition to direct reconstruction strategies, there have been dedicated efforts aimed at obtaining high quality PET images [6, 7] from the reconstructed images based on above mentioned methods. Gong et al. [8] proposed a reconstruction framework combining a residual convolutional network with ML. Zhu et al. [9] achieved reconstruction from sensor domain to image domain on PET data. However, these efforts are belonging to indirect reconstruction methods, which are from images to images. Only Ida et al. [10] paid attention to the direct reconstruction and adopt a deep encoder-decoder network to learn the mapping from the projection data to the reconstructed PET image. They mainly focused on the reconstruction speed instead of image quality. And as they mentioned, the work proposed was verified by the simulated as opposed to real clinical data.

method learns a mapping from sinogram y and random noise z to reconstructed image x , which can be stated as: $y = G \{x, z\}$.

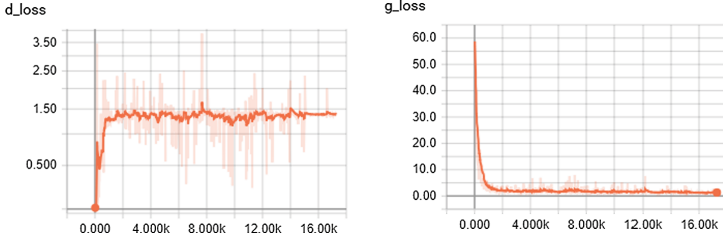


Fig. 2. The convergence curves of discriminator and generator

2.2 Direct Reconstruction Framework Based cGANs

Objective. Unlike traditional GANs, the cGANs also need a conditional variant to lead the training of the network. The conditional restriction can make the results closer to the intended goals. So the input includes noise z and measurement y , the objective of cGANs can be represented as

$$\begin{aligned} L_{cGANs}(D, G) &= \min_G \max_D V(D, G) \\ &= E_{x \sim p_{data}(x)} [\log D(x|y)] + E_{z \sim p_z(z)} [\log(1 - D(G(z|y)))] \end{aligned} \quad (2)$$

G denotes a generate model which captures the data distribution, D denotes a discriminator model which estimates the probability that a sample came from the training data rather than G . G tries to adjust parameters to minimize $\log(1 - D(G(z|y)))$, while D tries to maximize it, they are two adversarial models. The training according adversarial loss makes the generated images becoming clear, but doesn't mean good similarity with the established image. To enhance the accuracy in the low frequency information, L1 regularizer is introduced in the proposed model. λ is the weight on L1 term. Thus the entire objective function can be described as

$$G^* = \arg \min_G \max_D L_{cGANs}(D, G) + \lambda L_{L1}(G) \quad (3)$$

where $L_{L1}(G)$ is L1-norm, and represented as

$$L_{L1}(G) = E_{x, y \sim p_{data}(x, y), z \sim p_z(z)} [\|y - G(x, z)\|_1] \quad (4)$$

Direct Reconstruction Framework. As mentioned above, the basic framework includes two parts: discriminator and generator. Figure 1 illustrates the structure of the entire reconstruction networks.

The generator adapts U-net structure which consists of two parts: (1) Encoder: contracting path (shown at the left-hand side) to compress the input image and get context information; (2) Decoder: symmetrical expanding path (shown at the right-hand side) to expand the path and locate accurately. The strategy of discriminator is

PatchGAN [12] which does not judge whether the target image is real or fake by comparing all the pixels of the whole image, but by classifying the authenticity of each $N \times N$ patch in the image. And as for N , it can be set much smaller than the image size for a smaller PatchGAN has fewer parameters and needs less computational cost. And to optimize our networks, we use the Adam solver. The number of the steps to apply to discriminator k is set to $k = 1$, which means the discriminator runs once, then the generator runs on, the two work alternately.

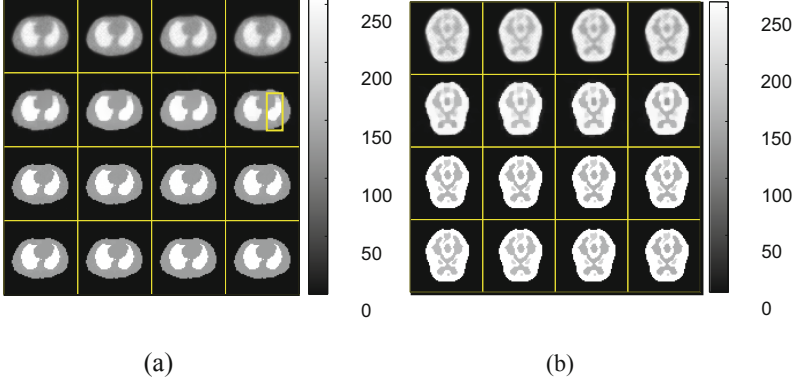


Fig. 3. Reconstruction results (a). Reconstruction results for Zubal thorax phantom with 40-min scanning and 1×10^7 counting rate. From top to bottom: MLEM results, TV results, cGANs results, ground truth. From left to right: the 3rd, 7th, 12th, 18th frame. (b) Reconstruction results for Hoffman brain with 70-min scan and 5×10^7 counting rate. From top to bottom: MLEM results, TV results, cGANs results, ground truth. From left to right: the 4th, 9th, 14th, 18th frame.

3 Experiments and Results

3.1 Simulation Experiments

Monte Carlo Simulations and Datasets. In order to verify the accuracy and robustness of the framework comprehensively, we generate a series of simulated data of the three phantoms using Monte Carlo simulations, including Zubal thorax phantom with ^{64}Cu -ATSM, Hoffman brain phantom with ^{18}F -FDG and Zubal brain phantom with ^{11}C -Acetate. Each phantom consists of three sampling times of data. For each sampling time, we set three levels of counting rates 5×10^6 , 1×10^7 , 5×10^7 . For different counting rates, each reconstructed dynamic image has 18 frames with a resolution of 64×64 pixels. Here the simulated PET scanner we adopt is Hamamatsu SHR74000.

Evaluation Metrics. For simulation data, the most remarkable advantage is the ground truth is known, so the reconstruction results can be quantified. In this paper, both bias and variance are adopted for comparison:

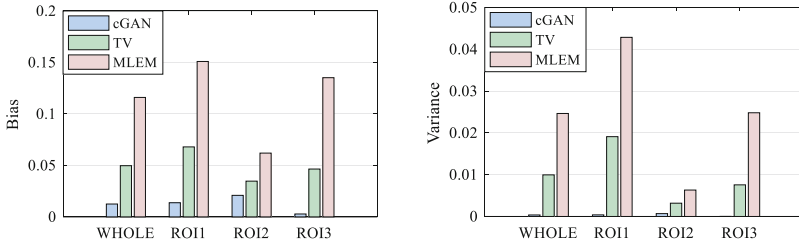


Fig. 4. Bias and variance of whole images, ROI1, ROI2 and ROI3 of Zubal thorax phantom.

$$bias = \frac{1}{n} \sum_{i=1}^n \frac{|\hat{x}_i - x_i|}{x_i} \quad (5)$$

$$variance = \frac{1}{n-1} \sum_{i=1}^n \left(\frac{\hat{x}_i - \bar{x}_i}{x_i} \right)^2 \quad (6)$$

where n denotes the overall number of pixels in ROI, \hat{x}_i denotes the reconstructed value at voxel i , x_i denotes the true value at voxel i , \bar{x}_i is the mean value of x_i .

Accuracy. The Zubal thorax datasets and the Hoffman brain datasets are divided according to sampling time and counting rates respectively. The 20 min scanning dataset and 30 min scanning dataset belong to training set which contains a total of 108 PET images and corresponding sinogram images. The 40 min scanning data are choosing to be testing sets including 54 image pairs. For the Hoffman brain datasets, the 1×10^7 counting rate data is selected to be the testing data. The rest two kinds of counting rates data formed a training set. The curves shown in Fig. 2 demonstrate the convergence of the algorithm that we proposed. The left curve shows the convergence of the discriminator; the right curve represents the convergence of the generator. The X-axis indicates the loss value, and the Y-axis indicates the value of iteration steps. It is obvious that both generator and discriminator converge to a value quickly and tend to be stable. We extract the 3rd, 7th, 12th, 18th frame of 1×10^7 counting rate from the testing dataset to exhibit reconstructed results which can be seen in Fig. 3. We also compared our method with both MLEM and TV algorithm.

As can be seen in Fig. 3, the reconstruction results of the proposed method are highly consistent with the ground truth in both terms of sharp edges and high pixel values of ROIs. The results of MLEM method contain too much noise and artifacts due to the chessboard effect, even the boundaries could not be seen clearly. The TV method provides a clearer and sharper result comparing with MLEM. However, the reconstruction effect in ROI3 area is really poor, as indicated in the yellow rectangular box.

Figure 4 shows the mean value of bias and variance of all 18 frames. It can be seen that no matter bias or variance, the cGANs framework absolutely gets a good score. And TV method shows better performance than MLEM method. Comparing with another two areas, the reconstruction results of ROI2 are unsatisfactory, the values of bias and variance of three methods are all higher than other areas and the two

parameters curve are pretty fluctuant. We consider it maybe because of the ROI2 area is too small to distinguish from the other two parts.

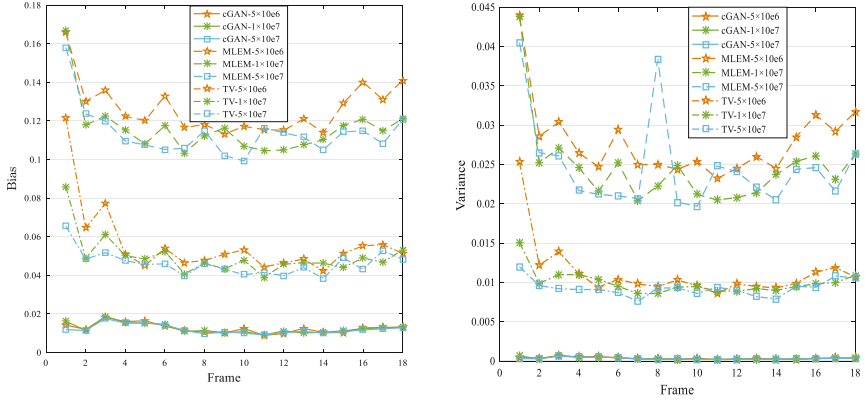


Fig. 5. Robustness verification for different counting rates (a) Bias (b) Variance (Color figure online)

We also extract 4 frames of Hoffman brain datasets to demonstrate the accuracy of the proposed framework. The reconstruction images are shown in Fig. 3(b), and the detailed quantitative results can be seen in Table 1. The bias of cGANs method several times smaller than the other two methods. The variance property of our methods are even nearly 100 times better than that of other methods.

Table 1. Hoffman brain reconstruction evaluation results

Frame	Method	Bias				Variance			
		Total	ROI1	ROI2	ROI3	Total	ROI1	ROI2	ROI3
4th	MLEM	0.1034	0.1667	0.0611	0.0823	0.0187	0.0391	0.0063	0.0107
	TV	0.1008	0.0829	0.0710	0.1485	0.0216	0.0148	0.0130	0.0369
	cGANs	0.0150	3.3e-5	0.0250	0.0199	4.3e-4	1.3e-7	7.8e-4	5.2e-4
9th	MLEM	0.0879	0.1603	0.0647	0.0388	0.0154	0.0365	0.0073	0.0024
	TV	0.0784	0.0629	0.0837	0.0886	0.0131	0.0096	0.0163	0.0134
	cGANs	0.0141	3.7e-5	0.0324	0.0097	4.5e-4	1.5e-7	0.0012	1.6e-4
14th	MLEM	0.0904	0.1566	0.0724	0.0421	0.0159	0.0359	0.0087	0.0031
	TV	0.0800	0.0575	0.0762	0.1064	0.0147	0.0087	0.0172	0.0184
	cGANs	0.0328	5.0e-5	0.0869	0.0114	0.0026	2.0e-7	0.0077	2.0e-4
18th	MLEM	0.0952	0.1576	0.0794	0.0485	0.0167	0.0355	0.0109	0.0037
	TV	0.0814	0.0859	0.0660	0.0922	0.0136	0.0135	0.0129	0.0143
	cGANs	0.0294	3.7e-5	0.0557	0.0326	0.0015	1.5e-7	0.0032	0.0012

Robustness. In this section, two phantoms are adopted to verify the influence of different factors. The Zubal thorax dataset is used to validate the robustness of the algorithm to different counting rates. The Zubal head dataset is used to verify that the proposed framework can predict the following tracer concentration distribution by learning the concentration distribution when the tracer is just injected. The reconstruction results of different counting rates are shown in Fig. 5. It is obvious that the three solid curves at the bottom of graph (a) and graph (b) have the highest coincidence which means the cGANs method is least affected by the counting rates. The blue line with box labeled, the green line with snow labeled and the orange line with star labeled indicates the counting rates 5×10^7 , 1×10^7 and 5×10^6 respectively. With the counting rate increasing, both variance and bias of MLEM and TV methods decrease distinctly. So we can conclude that these two methods are greatly influenced by the counting rate. We speculate that this is because both methods are based on probabilistic statistical model, so when the counting rate increases, the probabilistic statistical characteristics of data can be better guaranteed. Comparing with the other two methods, the proposed framework has minimum variance and deviation and maximum stability.

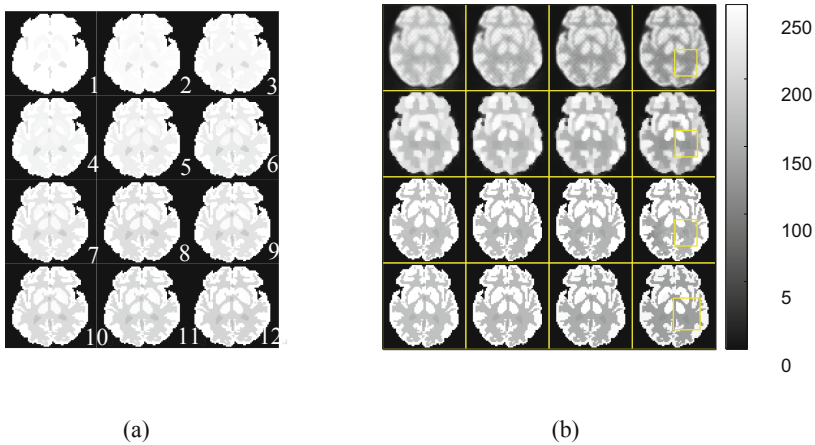


Fig. 6. Robustness verification graphs for different frames. (a) Concentration distribution of training images (the first 12 frames). (b) Reconstruction results for Zubal head phantom with a 70-min scan and 5×10^7 counting rate. From top to bottom: MLEM results, TV results, cGANs results, ground truth. From left to right: the 13th, 14th, 16th, 18th frame.

The Zubal head phantom is used to verify whether the proposed algorithm has the effect of predicting the reconstructed images of the last 6 frames by training the first 12 frames. The concentration distribution of the first 12 frames are shown in Fig. 6(a). We choose 4 frames from the subsequent 6 frames to exhibit the reconstruction results as shown in Fig. 6(b).

It can be seen that our results are highly consistent with the true concentration distribution. Even for the tiny area, such as ROI3 can be reconstructed clearly, which has been circled with yellow rectangular boxes in Fig. 6.

3.2 Real Data Experiments

We collected brain PET projection data of 10 real patients with the Hamamatsu SHR22000. Each patient has 10 frames of sinogram image with 192×192 pixels in size, and the corresponding reconstructed images were generated by traditional reconstruction method, filtered-back-projection (FBP). We randomly selected 7 patients as training dataset which consisted of 70 image pairs. The rest 30 sinogram images of the other 3 patients would use as testing data. In the experiments, the epoch $n = 200$.

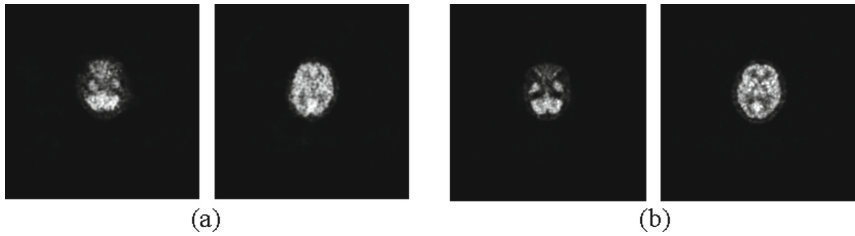


Fig. 7. Reconstruction results of real data. (a) Reconstructed image generated by cGANs. (b) The reconstructed images by FBP.

As seen in Fig. 7, the reconstructed images that produce by our method has a similar contour structure to the ground truth. But the details of the image are not clear enough. We think this may be because the noise contained in the real data is more complex.

4 Conclusion

A deep learning strategy has been derived for reconstruction of activity distribution from PET emission sinograms. In the framework, we rely on conditional generative adversarial networks (cGANs) is to represent a mapping from sinogram domain to image domain. Simulated phantoms, as well as real patient data, are used to evaluate the proposed strategy.

Acknowledgements. This work is supported in part by the National Natural Science Foundation of China (No: U1809204, 61525106, 61427807, 61701436), by the National Key Technology Research and Development Program of China (No: 2017YFE0104000, 2016YFC1300302), and by Shenzhen Innovation Funding (No: JCYJ20170818164343304, JCYJ20170816172431715).

References

1. Qi, J., Leahy, R.M.: Iterative reconstruction techniques in emission computed tomography. *Phys. Med. Biol.* **51**(15), R541–R578 (2006). <https://doi.org/10.1088/0031-9155/51/15/R01>
2. Shepp, L.A., Vardi, Y.: Maximum likelihood reconstruction for emission tomography. *IEEE Trans. Med. Imaging* **1**(2), 113–122 (1982). <https://doi.org/10.1109/TMI.1982.4307558>

3. Levitan, E., Herman, G.T.: A maximum a posteriori probability expectation maximization algorithm for image reconstruction in emission tomography. *IEEE Trans. Med. Imaging* **6** (3), 185–192 (1987). <https://doi.org/10.1109/TMI.1987.4307826>
4. Zhou, J., Luo, L.M.: Sequential weighted least squares algorithm for PET image reconstruction. *Digit. Sig. Proc.* **16**(6), 735–745 (2006). <https://doi.org/10.1016/j.dsp.2006.08.006>
5. Cabello, J., Torres-Espallardo, I., Gillam, J.E., Rafecas, M.: PET reconstruction from truncated projections using total-variation regularization for hadron therapy monitoring. *IEEE Trans. Nucl. Sci.* **60**(5), 3364–3372 (2013). <https://doi.org/10.1109/TNS.2013.2278121>
6. Xie, S., et al.: Artifact removal using improved GoogLeNet for sparse-view CT reconstruction. *Sci. Rep.* **8**(1), 6700–6709 (2018). <https://doi.org/10.1038/s41598-018-25153-w>
7. Kim, K., et al.: Penalized PET reconstruction using deep learning prior and local linear fitting. *IEEE Trans. Med. Imaging* **37**(6), 1478–1487 (2018). <https://doi.org/10.1109/TMI.2018.2832613>
8. Gong, K., et al.: Iterative PET image reconstruction using convolutional neural network representation. *IEEE Trans. Med. Imaging* **38**(3), 675–685 (2019). <https://doi.org/10.1109/tmi.2018.2869871>
9. Zhu, B., Liu, J.Z., Cauley, S.F., Rosen, B.R., Rosen, M.S.: Image reconstruction by domain-transform manifold learning. *Nature* **555**(7697), 487–492 (2018). <https://doi.org/10.1038/nature25988>
10. Häggström, I., Schmidtlein, C.R., Campanella, G., Fuchs, T.J.: DeepPET: a deep encoder–decoder network for directly solving the PET image reconstruction inverse problem. *Med. Image Anal.* **54**, 253–262 (2019). <https://doi.org/10.1016/j.media.2019.03.013>
11. Mirza, M., Osindero, S.: Conditional generative adversarial nets (2014)
12. Isola, P., Zhu, J., Zhou, T., Efros, A.A.: Image-to-image translation with conditional adversarial networks. In: Paper Presented at the 30th IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017, pp. 5967–5976 (2017). <https://doi.org/10.1109/cvpr.2017.632>