

# 200x Low-dose PET Reconstruction using Deep Learning

Junshen Xu<sup>†</sup>, Enhao Gong<sup>†</sup>, John Pauly and Greg Zaharchuk<sup>\*</sup>

**Abstract**—Positron emission tomography (PET) is widely used in various clinical applications, including cancer diagnosis, heart disease and neuro disorders. The use of radioactive tracer in PET imaging raises concerns due to the risk of radiation exposure. To minimize this potential risk in PET imaging, efforts have been made to reduce the amount of radio-tracer usage. However, lowering dose results in low Signal-to-Noise-Ratio (SNR) and loss of information, both of which will heavily affect clinical diagnosis. Besides, the ill-conditioning of low-dose PET image reconstruction makes it a difficult problem for iterative reconstruction algorithms. Previous methods proposed are typically complicated and slow, yet still cannot yield satisfactory results at significantly low dose. Here, we propose a deep learning method to resolve this issue with an encoder-decoder residual deep network with concatenate skip connections. Experiments shows the proposed method can reconstruct low-dose PET image to a standard-dose quality with only two-hundredth dose. Different cost functions for training model are explored. Multi-slice input strategy is introduced to provide the network with more structural information and make it more robust to noise. Evaluation on ultra-low-dose clinical data shows that the proposed method can achieve better result than the state-of-the-art methods and reconstruct images with comparable quality using only 0.5% of the original regular dose.

**Index Terms**—Deep Learning (DL), Positron emission tomography (PET), low-dose PET reconstruction, Image enhancement/restoration, Denoising, Convolution Neural Network (CNN)

## I. INTRODUCTION

POSITRON emission tomography (PET) has a wide range of clinical applications, such as cancer diagnosis, tumor detection[1] and early diagnosis of neuro diseases[2], for its ability of cellular level imaging and high specificity. In order to acquire high quality PET image for diagnostic purpose, a standard dose of radioactive tracer should be injected to the subject which will lead to higher risk of radiation exposure damage. Usually, a PET scan or the a scan of PET/MR and PET/CT can expose patients with even more ionizing radiation than a scan using CT along. To minimize such risk, the well-known principle of ALARA (as low as reasonably achievable)[3] is adopted in clinical practice. In addition, lowering injected dose in PET can also result in reduction of imaging costs, shorter imaging time and may improve imaging logistics when fast-decaying tracers are

used. However, dose reduction will adversely affect PET image quality with lower Signal-to-Noise-Ratio (SNR), as shown in Fig. 1.

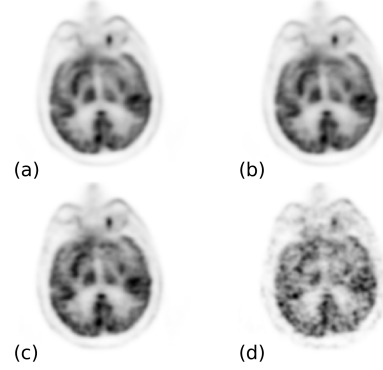


Fig. 1. PET images with normal dose and different levels of dose reduction. (a) standard-dose, (b) quarter-dose, (c) twentieth-dose, and (d) two-hundredth-dose.

To address this problem, many algorithms were proposed to improve the image quality for low-dose PET image. In general, these algorithms can be categorized into three categories: (1) iterative reconstruction algorithm, (2) image filtering and post-processing, and (3) machine learning.

Iterative reconstruction algorithms formulate the low-dose reconstruction problem as a convex optimization problem combining statistical model of the acquired data (i.e., sino-gram or listmode) and the regularization term to suppress noise. Wang *et al.* [4] proposed an iterative algorithm using a Total Variation (TV) regularization to reduce the noise of synthetic emission phantom with different photon counts. Although Iterative reconstruction algorithms are potentially most accurate since they consider the raw count information directly, they also have three main weaknesses. First, the substantial computational expenses interacting with all the acquired data make most of this kind of methods time-consuming. Second, iterative methods are typically vendor-specific, since different scanners may adopt different geometric configurations, data formats (e.g., time-of-flight (TOF) [5] and depth-of-interaction (DOI) [6]), and data correction procedures, which will significantly affect the raw data. Finally, in these methods, a predefined regularization term is need, which may leads to undesirable over-smoothing, artifacts or hallucinated textures.

As for image processing methods, several general-purpose image denoising algorithms, such as nonlocal means (NLM) [7] and block-matching 3D (BM3D) [8], are introduced into PET image denoising [9], [10]. Besides, Bagci *et al.* [11] combined singular value thresholding method and Stein's

This work was supported in part by —. Dagger indicates equal contribution of the first two authors. Asterisk indicates corresponding author.

<sup>†</sup>Junshen Xu is with the Department of Electrical Engineering, Stanford University, Stanford, CA, 94305, USA and the Department of Engineering Physics, Tsinghua University, Beijing, 100084 China

<sup>†</sup>Enhao Gong and John Pauly is with the Department of Electrical Engineering, Stanford University, Stanford, CA, 94305, USA

<sup>\*</sup>Greg Zaharchuk is with the Department of Radiology, Stanford University, Stanford, CA, 94305, USA

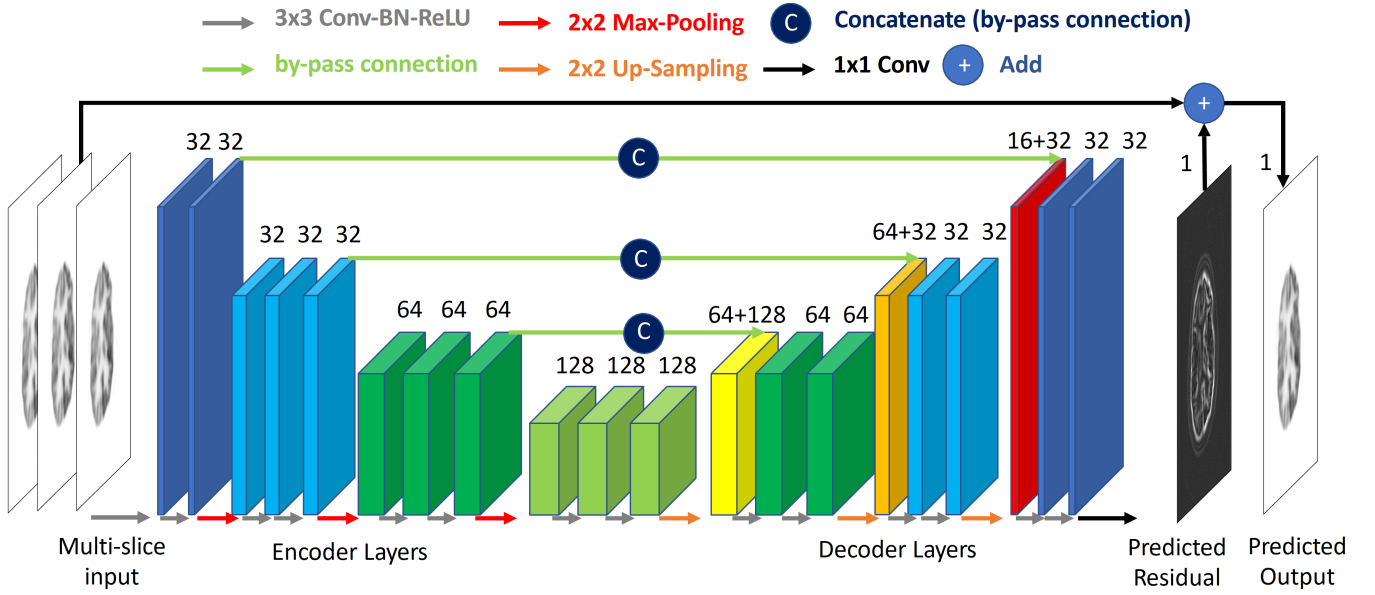


Fig. 2. Overall architecture of our proposed network.

unbiased risk estimate to denoise PET image. Based on the multi-scale Curvelet and Wavelet analysis, Pogam *et al.* [12] proposed a method to denoise PET image while preserving image resolution and quantification.

Another important category is the data-driven machine learning methods such as mapping-based sparse representation [13], semi-supervised triple dictionary [14], and multi-level canonical correlation analysis [15]. Instead of denoise the low-dose PET image directly, machine learning methods utilize paired low-dose and standard-dose images to train models that can predict standard-dose images from low-dose inputs.

Recently, deep learning attracts a lot of attention in computer vision applications, yields much better results compared with traditional methods, and achieves human-level performance in some tasks such as image classification [16] and face verification [17]. Several key factors contribute to the success of deep learning methods: (1) acceleration of parallel computation due to modern powerful GPUs that make it possible to train models with large amount of parameters [18], (2) larger datasets are released boosting more open source research and training, e.g., ImageNet [19], and (3) new efficient neural network structures, e.g., convolution neural network (CNN) which utilizes weight sharing and local connection [20]. In addition, deep learning methods are also successfully applied to the category of low-level vision problems including image denoising [21], super resolution [22], and image restoration [23], etc., achieving state-of-the-art results.

Although these methods mainly focus on natural image processing, several efforts have been made to apply these promising methods to medical image analysis. U-Net [24] is a fully convolutional network for medical image segmentation which consists of a contracting path and an expansive path to extract features at different resolution. To regain the lost res-

olution information, U-Net also employs skip connection to concatenate corresponding contracting and expansive steps. Inspired by U-Net, Han *et al.* [25] proposed a multi-scale CNN to remove streaking artifacts in sparse-view CT images, using residual learning. WaveNet [26], which is also used for low-dose X-ray CT reconstruction, adopts a similar structure combined with multi-scale wavelet transformation as feature augmentation for input data. In the field of low-dose PET reconstruction, compared with low-dose CT reconstruction, there are few researches on low-dose PET image denoising that utilize deep learning methods. Xiang *et al.* [27] proposed a deep learning method to predict standard-dose PET images from low-dose PET images and corresponding MR T1 images with an auto-context convolution network which tries to refine the prediction results step by step.

In terms of dose reduction factor (DRF), methods in [9], [10], [11], [12] are used to denoising standard-dose images (DRF = 1) while methods in [13], [14] try to reconstruct standard-dose images from quarter-dose images (DRF=4). However, to the best of our knowledge, there is no work that reconstructs low-dose PET images with higher DRF.

In this paper, we propose a deep learning method to reconstruct standard-dose PET images from ultra-low-dose images (99.5% reduction or DRF=200), using a fully convolutional encoder-decoder residual deep network model. To our best knowledge, this is the first time a deep learning method is proposed and demonstrated for enabling ultra-low-dose PET reconstruction at such a high reduction factor and with in-vivo PET datasets.

## II. METHOD

### A. Dataset and experiments setup

PET/MRI images from nine patients with glioblastoma (GBM) were acquired on a simultaneous time-of-flight enabled PET/MRI system (SIGNA, GE Healthcare) with stan-

dard dose of 18F-fluorodeoxyglucose (FDG) (370 MBq). Images were acquired for about 40 min, beginning 45 min after injection. We stored the raw count listmode datasets for each scan and then generate synthesized low-dose raw data at DRF = 200 by simply randomly selecting 0.5% of the count events, spread uniformly over the entire acquisition period. Then we reconstruct PET images from the acquired data at DRF = 1 (standard full dose) and DRF = 200 (target low dose) using standard OSEM methods (28 subsets, 2 iterations).

The size of each reconstructed 3D PET data is  $256 \times 256 \times 89$ . There are slices of air at the top and bottom, which are removed. To avoid over fitting, data augmentation is adopted during the training process to simulate a larger dataset. Before fed into the network, images are randomly flipped along x and y axes and transposed.

### B. Deep Learning based low-dose PET reconstruction

The goal of this work is to train a model to learn to reconstruct from the DRF = 200 image to DRF = 1 reconstruction.

1) *Network structure*: As shown in Fig. 2, the proposed fully convolutional network is based on an encoder-decoder structure with symmetry concatenate connections between corresponding stages, which is inspired by the U-Net structure but modified for image synthesis task instead of segmentation task. Specifically, each stage consists of convolution with  $3 \times 3$  kernels, batch normalization, and rectified linear unit (ReLU). The downsampling and upsampling between stages are done by  $2 \times 2$  max pooling and bilinear interpolation respectively. By downsampling and then upsampling the image, the network can extract the multi-scale and high-level features from the image. The low-dose PET image reconstruction task is similar to image denoising which is within the category of low-level vision problems and are susceptible to resolution loss if only a encoding-decoding procedure is used. Therefore, concatenate connections are added to preserve local information and resolution of the image.

2) *Residual learning*: Residual learning[28] is first introduced into CNN as a technique to avoid performance degradation when training very deep CNNs. It shows by separating the identity and the residual part, the neural network can be trained more effectively and efficiently. Originally, residual learning is used in image recognition task [28] and Zhang *et al.* [29] proposed DnCNN which is the first denoising convolution network using residual learning. It is showed in [25], using persistent homology analysis, that the residual manifold of CT artifacts has a much simpler structure. Our network also employs the residual learning technique, by adding a residual connection from input to output directly, i.e., instead of learning to generate standard-dose PET images directly, the network tries to learn the difference between standard-dose images outputs and low-dose images inputs. Our study shows that residual learning can also lead to a significant improvement

in network performance for low-dose PET reconstruction problem.

3) *Using multi-slice as input*: Using only the low-dose image as input for the neural network may not provide enough information to reconstruct the standard-dose counterpart. As shown in fig, the noise due to dose reduction cannot be fully eliminated by the network for the network may have insufficient information to distinguish noise from brain structure. To address this problem, we use multi-slice input instead of single-slice input, i.e., adjoining slices are stacked as different input channels. In general, the multi-slice inputs can be regarded as a kind of feature augmentation. Since the structure of the brain is deterministic, adjoining slices may share similar structure while having different noise which is random. Thus, combining different slices as input can provide the network with 2.5D structural information that can be used to distinguish random noise from the consistent structure. One example is illustrated in Fig. 3, in the low-dose PET image there is a black noise in the zoomed part, which cannot be eliminated but hallucinated as a structure by the network trained with single-slice input. However, the network trained with three-slice input can achieve better results, as shown in Fig. 3(d). Training 2.5D multi-slice inputs is different from training with 3D convolution network since the former solution uses depth-wise operation of 3D convolution which has fewer parameters and higher efficiency.

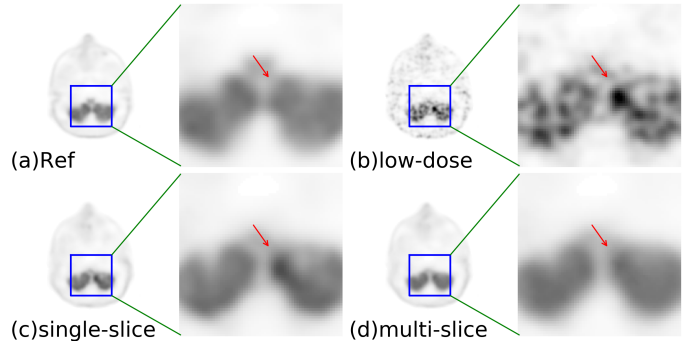


Fig. 3. Effect of multi-slice input. (a) standard-dose, (b) 200x low-dose, (c) reconstructed result using single-slice input, and (d) reconstructed result using multi-slice input (3 slices).

4) *Selection of loss functions*: The mean squared error (MSE) or  $L_2$  loss is still the most popular choice of loss function in training networks for image restoration problems, e.g., super resolution or denoising. The use of MSE as loss function is under the assumption of additive white Gaussian noise, which should be independent of the local features of the image. However, this is not valid for low-dose PET reconstruction in general. Since the intensity of PET image reflects the activity distribution of tracer in the subject and the noise results from dose reduction is related to the counting of each detector, noise and spatial information are not independent. Besides, the MSE loss may be not suitable for task related to clinical evaluation for it relates poorly to the human visual system and produces splotchy artifacts[30].

Besides the traditional MSE, there are other loss functions that can be used to measure image similarity between reconstructed image and the ground-truth image. The  $L_1$  loss is the mean absolute error of two images which can be defined as

$$L^{l_1} = \frac{1}{NM} \sum_{i=1}^N \sum_{j=1}^M |x_{ij} - y_{ij}| \quad (1)$$

where  $N$ ,  $M$  are number of rows and columns of the image respectively, while  $x_{ij}$  and  $y_{ij}$  denote the intensity at pixel  $(i, j)$  in the two images. To measure the structural and perceptual similarity, structural similarity index (SSIM[31]), and multi-scale the structural similarity index (MS-SSIM[32]) are proposed and can be estimated as

$$L^{\text{SSIM}} = \frac{1}{NM} \sum_{i=1}^N \sum_{j=1}^M (1 - \text{SSIM}(i, j)) \quad (2)$$

$$L^{\text{MS-SSIM}} = \frac{1}{NM} \sum_{i=1}^N \sum_{j=1}^M (1 - \text{MS-SSIM}(i, j)) \quad (3)$$

where

$$\begin{aligned} \text{SSIM}(i, j) &= \frac{(2\mu_x\mu_y + C_1)}{(\mu_x^2 + \mu_y^2 + C_1)} * \frac{(2\sigma_{xy} + C_2)}{(\sigma_x^2 + \sigma_y^2 + C_2)} \\ &= l(i, j) * cs(i, j) \end{aligned} \quad (4)$$

$$\text{MS-SSIM}(i, j) = l_K(i, j) * \prod_{k=1}^K cs_j(i, j) \quad (5)$$

$C_1$  and  $C_2$  are constants.  $\mu_x$ ,  $\mu_y$ ,  $\sigma_x$ ,  $\sigma_y$ , and  $\sigma_{xy}$  are the image statistics calculated in the patch centered at pixel  $(i, j)$ .  $K$  is the number of level of multi-scale.

recent researches[30], [33] suggested that  $L_1$ , SSIM, MS-SSIM are more perceptually preferable in image generative model. Among these three alternatives, the  $L_1$  loss can not only avoid the patchy artifact brought by  $L_2$  loss but add almost no overhead in back propagation compared with SSIM and MS-SSIM. Therefore, the  $L_1$  loss is selected as a loss function for training procedure in the following experiments.

### C. Computation environment and hardware settings

All the computation works were done on a Ubuntu server with 2 NVIDIA GTX 1080Ti GPUs. The proposed network is implemented in TensorFlow. The RMSprop optimizer is used in our experiments with a learning rate initialized by  $1 \times 10^{-3}$  which slowly decreases down to  $2.5 \times 10^{-4}$ . The network was trained for 120 epochs. Convolution kernel were initialized with truncated Gaussian distributions with zero mean and standard deviation 0.02. All biases are initialized with zero.

### D. Evaluation and similarity metrics

To evaluate the performance of the proposed method and demonstrate it can generalize for new dataset, especially for new patient data with different pathology, we used the leave-one-out cross validation (LOOCV). For each of the patient dataset, we generated the full-dose reconstruction using the model trained only on the other eight patients. We used the statistics of LOOCV results to quantify the generalization error of the proposed model. To quantitatively evaluate image quality, three similarity metrics are used in our experiment, including the normalized root mean square error (NRMSE), peak signal to noise ratio (PSNR) and SSIM. SSIM is defined in equation 4, while NRMSE and PSNR are defined as follows.

$$\text{NRMSE} = \sqrt{\frac{\sum_{i=1}^N \sum_{j=1}^M (x_{ij} - y_{ij})^2}{\sum_{i=1}^N \sum_{j=1}^M y_{ij}^2}} \quad (6)$$

$$\text{PSNR} = 20 * \log_{10}\left(\frac{\text{MAX}}{\sqrt{\text{MSE}}}\right) \quad (7)$$

where MAX is the peak intensity of the image. To better match the metric computation to the real clinical assessment, all the similarity metrics were computed after applying a brain mask estimated using image support.

## III. RESULTS

### A. Comparison with other methods

We compared our proposed method against three state-of-the-art denoising methods in low-dose PET reconstruction, including NLM [7], BM3D [8] and auto-context network [27] (AC-Net). Cross validation is conducted to evaluate these methods.

Fig. 4 shows the average performance in terms of NRMSE, PSNR and SSIM of all the subjects, while Fig. 5 gives the scores of these three metrics for all 9 subjects in the leave-one-out testing.

To examine perceptual image quality, two representative slices are selected from different subjects. The quantitative metrics in terms of NRMSE, PSNR and SSIM of the selected slices are listed in Table I. The reconstruction results, zoomed tumors, and the corresponding error map are visually illustrated in Fig. 6, Fig. 7, and Fig. 8 respectively.

TABLE I  
QUANTITATIVE RESULTS ASSOCIATED WITH DIFFERENT  
ALGORITHMS FOR REPRESENTATIVE SLICES.

		slice A			slice B	
	NRMSE	PSNR	SSIM	NRMSE	PSNR	SSIM
low-dose	0.228	30.05	0.917	0.214	29.58	0.899
NLM	0.153	33.50	0.958	0.142	33.13	0.947
BM3D	0.145	32.98	0.961	0.189	30.70	0.926
AC-Net	0.147	33.83	0.960	0.137	33.50	0.958
Proposed	<b>0.124</b>	<b>35.33</b>	<b>0.974</b>	<b>0.106</b>	<b>35.66</b>	<b>0.969</b>

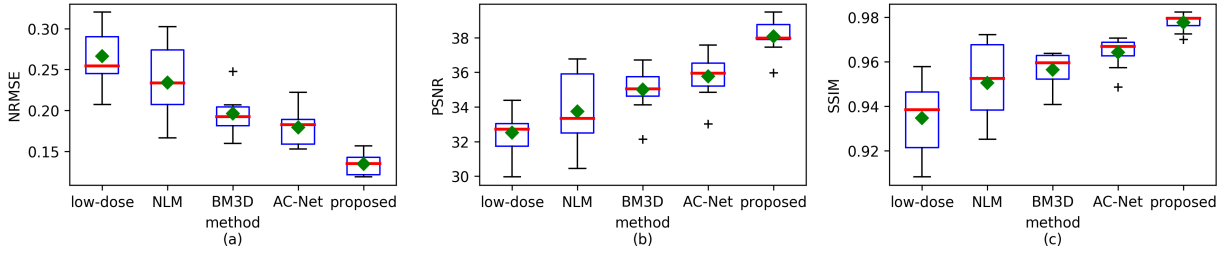


Fig. 4. Comparison of the averaged performance and similarity metrics of different methods for low-dose reconstruction, where green diamonds denote means.

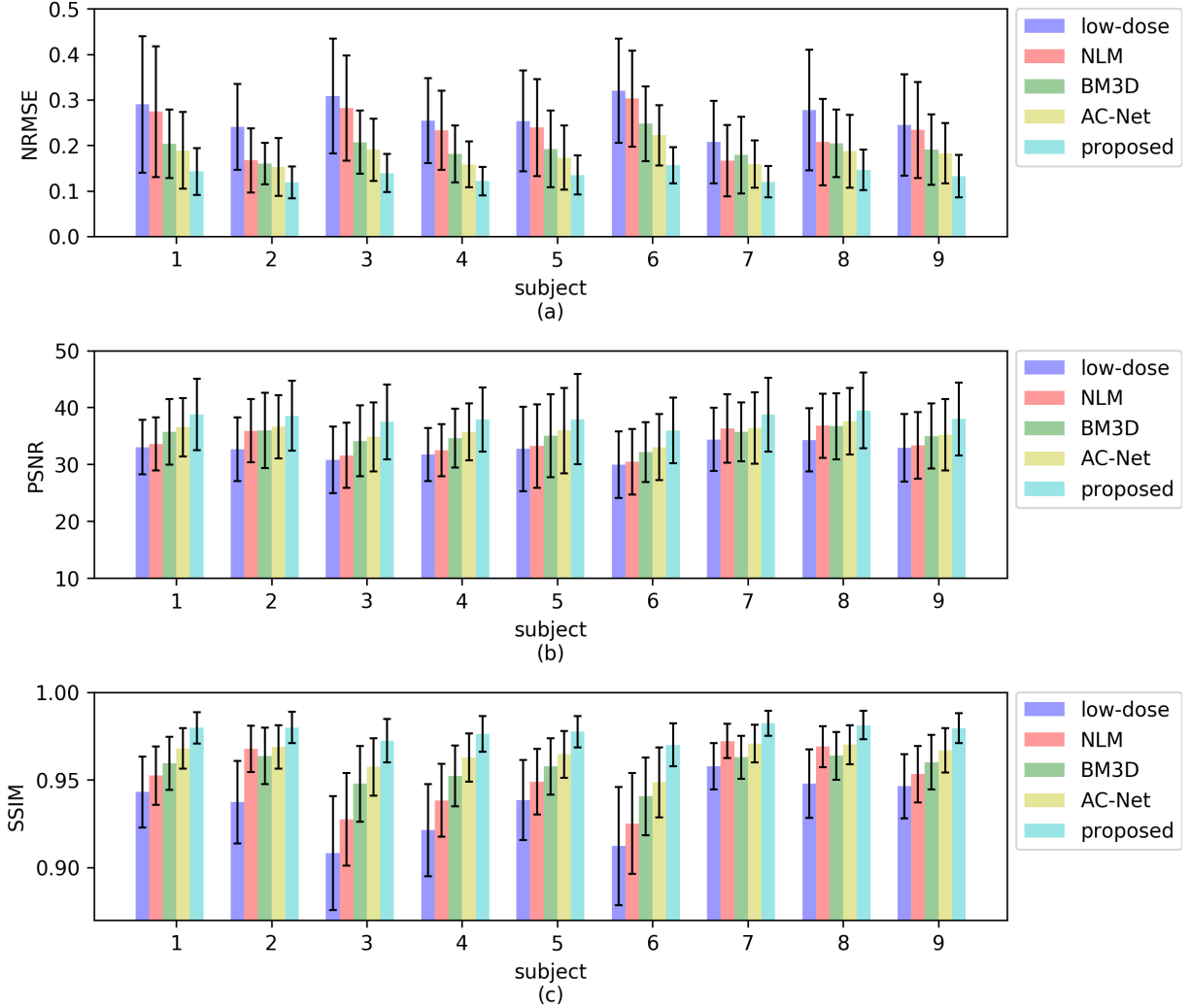


Fig. 5. Quantitative comparison between our proposed method and previous method using LOOCV.

### B. Contributions of skip connections

We explored the contribution to reconstruction from different skip connection components in the network. There are two types of skip connections in our proposed network. One is the residual connection from input to output, and the other is the concatenating connections between corresponding encoder and decoder layers. To evaluate the effect of these two types of skip connection on the network performance, four different models are trained and tested, i.e., (1) with both types of skip connection, (2) with only concatenate connec-

tion, (3) with only residual connection, and (4) without any skip connection. Fig. 9 shows the different testing loss of these four models during training and the quantitative results of cross validation are illustrated in Fig. 10.

### C. Contributions of multi-slice inputs

As mentioned above, we used multi-slice input to combine information from adjoining slices so that the network can more accurately generate reconstruction with less noise and artifact while robustly preserve original structure and details.

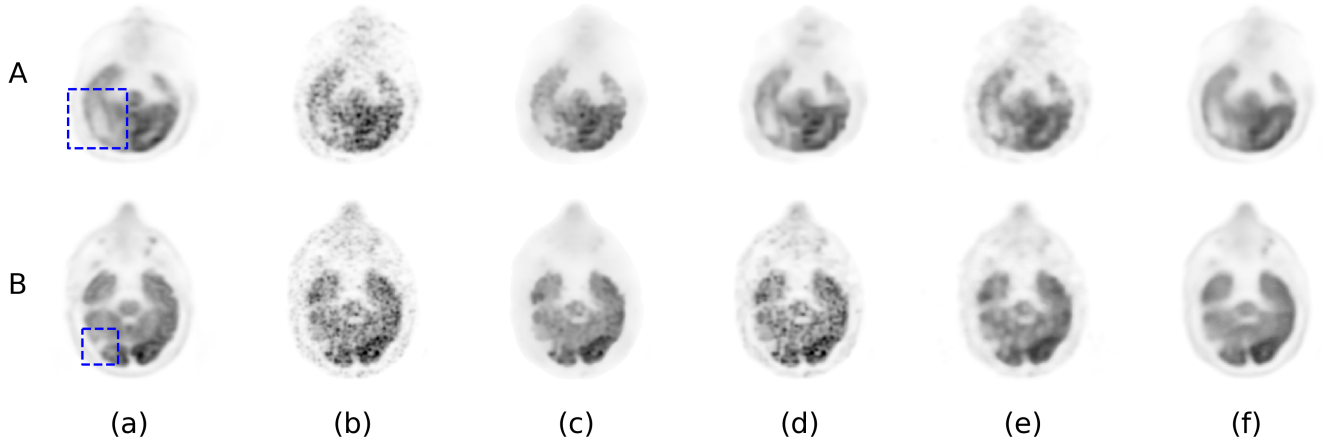


Fig. 6. Results from different methods for comparison. (a) standard-dose, (b) low-dose, (c) NLM, (d) BM3D, (e) AC-Net, and (f) proposed.

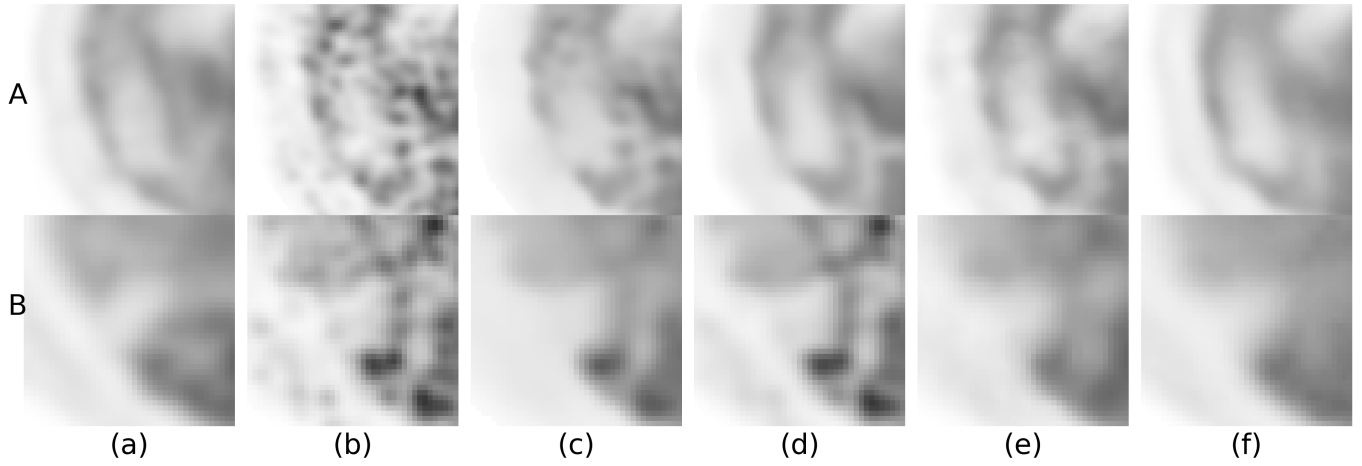


Fig. 7. Zoomed part in Fig. 6. (a) standard-dose, (b) low-dose, (c) NLM, (d) BM3D, (e) AC-Net, and (f) proposed.

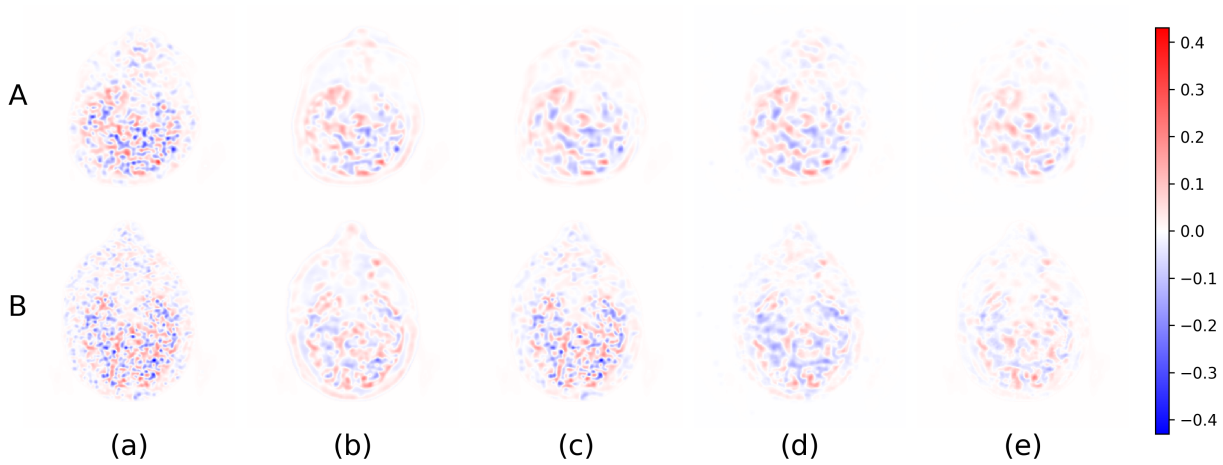


Fig. 8. Error map of Fig. 6. (a) low-dose, (b) NLM, (c) BM3D, (d) AC-Net, and (e) proposed.



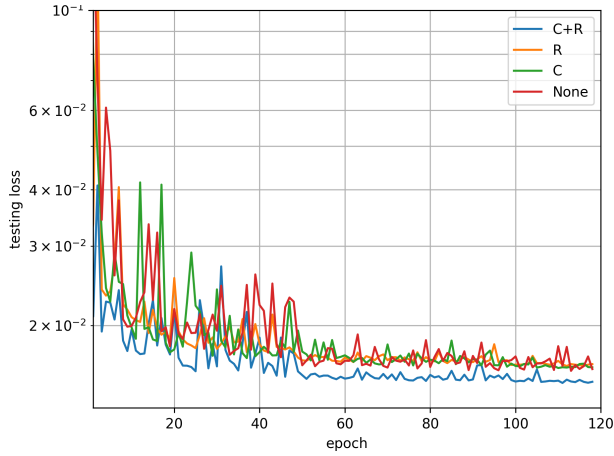


Fig. 9. Testing loss of networks with different types of skip connection. R means residual connection and C means concatenate connection.

To study the limit of this technique, networks with different numbers of input slices (1,3,5,7) are trained and their results are compared, shown in Fig. 11.

Fig. 12 shows results of three similarity metrics of networks trained with different numbers of slices as 2.5D inputs in the network. The evolutions of the three metrics all validate the performance gain of the proposed method using more input slice number. Compared with single-slice input, three-slice input can provide significantly better results. However, the performance gain of the network, by continuously adding slices more than 3 slices, is not as significant. Similar phenomenon can be seen in Fig11. (d)-(f) contain details that are missing or blurred in (c). However, (d)-(f) are perceptually similar.

#### D. Depth of network

To optimize our proposed network, experiments are conducted to evaluate the impact of depth of our model on network performance. Two hyper-parameters are used to control the depth of our network, namely number of pooling layers ( $n_p$ ) and number of convolutions between two poolings ( $n_c$ ). The strategy of grid search is adopted. In our experiment,  $n_p$  varies from 2 to 5 while  $n_c$  varies from 1 to 3. The results are shown in Fig. 13, which suggest that  $n_p = 3$  and  $n_c = 2$  is the best architecture in our study.

### IV. DISCUSSION

#### A. Compare with other methods

Quantitative results in Fig. 4 and Fig. 5 show that the propose method demonstrated the best performance in all nine subjects in the data set, compared with other methods that we have tested. From the visual results, it also suggests that our proposed method have highest image quality. NLM produces patchy artifact in the image as shown in Fig 6(c). Both BM3D and AC-Net cannot fully remove the noise in low-dose image and tend to over-blur the image without recover important details, as illustrated in Fig. 6(d) and (e). Same conclusion can also be drawn from the error map in

Fig. 8. In addition, our proposed method can achieve the best perceptual result in the region of GBM, as shown in Fig. 7.

In terms of computational costs, although deep learning requires long time for training, their efficiency in inference can easily outperforms traditional methods due to efficient implementation with Tensorflow and parallelization on GPUs. Time consumptions of each method for a  $256 \times 256$  image are listed in Table II. Compared with other methods, the proposed solution is not only more accurate but also more efficient.

It is the encoder-decoder structure that enable the network to adopt more parameters and channels to extract higher level features while reducing computation time, compared with single-scale model used in AC-Net.

TABLE II  
TESTING TIME (PER IMAGE) FOR EACH METHOD.

Method	Average Speed/Image (ms)
NLM(CPU)	1180
NLM(GPU)	63
BM3D(CPU)	680
BM3D(GPU)	232
AC-Net(GPU)	27
Proposed(GPU)	<b>19</b>

#### B. Benefits from concatenation and residual skip connections

As the result shown in Fig. 10, the model with both types of skip connections obviously achieves the best performance. However, for model with only one type of skip connection, their performances are close to that of the model without skip connection, or even worse. There results indicates that these two kinds of connections are not independent.

#### C. Benefits from 2.5D augmentation

Here we compared both quantitatively and qualitatively on the reconstruction using different options for combining multi-slice inputs. Detailed structures in Fig.11(c) are blurred during the denoising process, while they are preserved in Fig. 11(d)-(f), which shows the benefits from multi-slice inputs.

Since resolution of our 3D PET data along z axial direction is worse than within axial image. Stacking a few slices along z axis can recover the 3D spatial relationship. Here we showed significant performance improvement from the 2.5D slice with augmentation by only using 3 slices, however the performance is not further improved by using more slices as inputs. This result is consistent with the assumption that the structural similarity of different slices persists until the relationship and redundancy one can leverage between slices vanish eventually due to distance.

### V. CONCLUSION

In this paper, a deep fully convolutional network was proposed for ultra-low-dose PET reconstruction, where multi-scale encoder-decoder architecture, concatenate connections and residual learning are adopted.

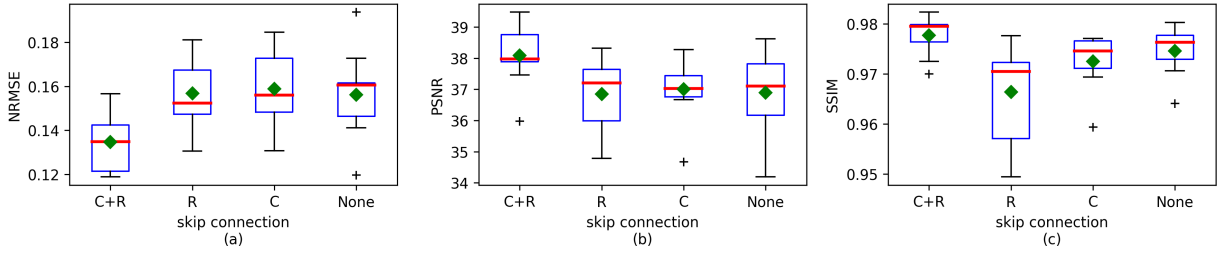


Fig. 10. Similarity metrics of network with different types of skip connection. R means residual connection and C means concatenate connection.

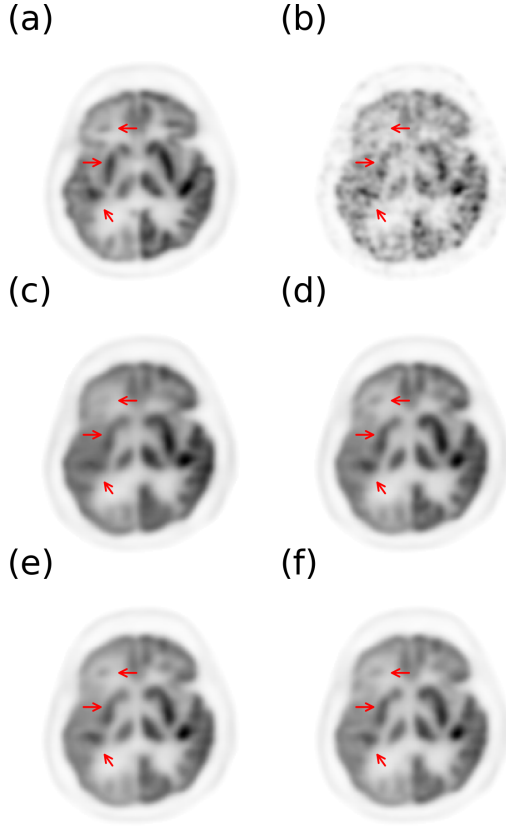


Fig. 11. Reference, inputs and reconstruction results using model with different settings in multi-slice inputs. (a) standard-dose, (b) low-dose, (c) single slice, (d) three slices, (e) five slices, and (f) seven slices

The results showed the proposed method has superior performance in reconstructing high-quality PET images and generating comparable quality as from normal-dose PET images. The method significantly reduces noise while robustly preserve resolution and detailed structures.

In addition, we demonstrated how different components of the proposed method contribute to the improved performance: the design of loss function, 2.5D multi-slice inputs as well as concatenating and residual skip connections, etc. Detailed quantitative and qualitative comparison proved the proposed method can better preserve structure and avoid hallucination due to noise and artifacts.

With extensive comparison, our proposed method achieves significantly better reconstruction compared with previous methods from ultra-low-dose PET data from 0.5% of the

regular dose, potentially enabling safer and more efficient PET scans.

## VI. ACKNOWLEDGE

The authors would like to thank Dr. Mohammad Mehdi Khalighi for his technical assistance in data acquisition and processing.

## REFERENCES

- [1] K. Ono, R. Ochiai, T. Yoshida, M. Kitagawa, J. Omagari, H. Kobayashi, and Y. Yamashita, "The detection rates and tumor clinical/pathological stages of whole-body FDG-PET cancer screening," *Annals of Nuclear Medicine*, vol. 21, pp. 65–72, 1 2007.
- [2] K. Herholz, E. Salmon, D. Perani, J. C. Baron, V. Holthoff, L. Frölich, P. Schönknecht, K. Ito, R. Mielke, E. Kalbe, and others, "Discrimination between Alzheimer dementia and controls by automated analysis of multicenter FDG PET," *Neuroimage*, vol. 17, no. 1, pp. 302–316, 2002.
- [3] S. D. Voss, G. H. Reaman, S. C. Kaste, and T. L. Slovits, "The ALARA concept in pediatric oncology," *Pediatric Radiology*, vol. 39, p. 1142, 9 2009.
- [4] Chenye Wang, Zhenghui Hu, Pengcheng Shi, and Huafeng Liu, "Low dose PET reconstruction with total variation regularization," in *2014 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pp. 1917–1920, IEEE, 8 2014.
- [5] W. W. Moses, "Time of flight in PET revisited," *IEEE Transactions on Nuclear Science*, vol. 50, pp. 1325–1330, 10 2003.
- [6] J. Ohi and H. Tonami, "Investigation of a whole-body DOI-PET system," *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, vol. 571, no. 1, pp. 223–226, 2007.
- [7] A. Buades, B. Coll, and J.-M. Morel, "A non-local algorithm for image denoising," in *Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on*, vol. 2, pp. 60–65, 2005.
- [8] K. Dabov, A. Foi, V. Katkovnik, K. Egiazarian, and others, "Image denoising with block-matching and 3 D filtering," in *Proceedings of SPIE*, vol. 6064, p. 606414, 2006.
- [9] S. Yu and H. H. Muhammed, "Noise type evaluation in positron emission tomography images," in *Biomedical Engineering (IBIOMED), International Conference on*, pp. 1–6, 2016.
- [10] D. Joyita and R. M. A. N. D. L. Q. Leahy, "Non-Local Means Denoising of Dynamic PET Images," *PLOS ONE*, vol. 8, pp. 1–15, 7 2013.
- [11] U. Bagci and D. J. Mollura, "Denoising PET Images Using Singular Value Thresholding and Stein's Unbiased Risk Estimate," *Medical image computing and computer-assisted intervention : MICCAI ... International Conference on Medical Image Computing and Computer-Assisted Intervention*, vol. 16, no. 0 3, pp. 115–122, 2013.
- [12] A. L. Pogam, H. Hanzouli, M. Hatt, C. C. L. Rest, and D. Visvikis, "Denoising of PET images by combining wavelets and curvelets for improved preservation of resolution and quantitation," *Medical Image Analysis*, vol. 17, no. 8, pp. 877 – 891, 2013.
- [13] Y. Wang, P. Zhang, L. An, G. Ma, J. Kang, F. Shi, X. Wu, J. Zhou, D. S. Lalush, L. Weili, and D. Shen, "Predicting standard-dose PET image from low-dose PET and multimodal MR images using mapping-based sparse representation," *Physics in Medicine & Biology*, vol. 61, no. 2, p. 791, 2016.



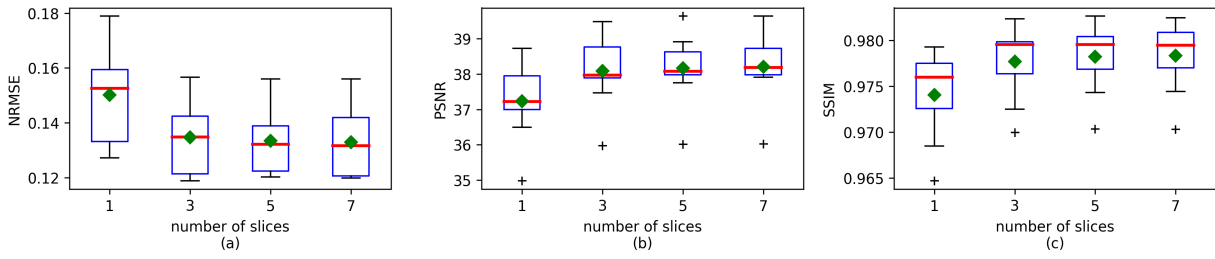


Fig. 12. Similarity metrics for networks trained with different numbers of input slices

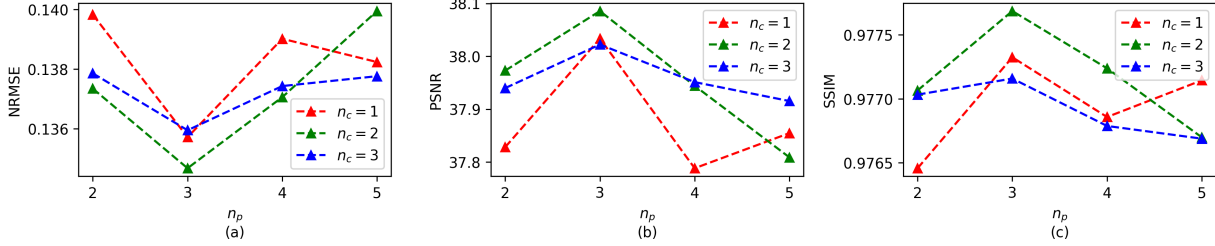


Fig. 13. Performance of networks without different depth evaluated with average NRMSE, PSNR and SSIM over all the 9 subjects in our study.

- [14] Y. Wang, G. Ma, L. An, F. Shi, P. Zhang, D. S. Lalush, X. Wu, Y. Pu, J. Zhou, and D. Shen, "Semisupervised Triple Dictionary Learning for Standard-Dose PET Image Prediction Using Low-Dose PET and Multimodal MRI," *IEEE Transactions on Biomedical Engineering*, vol. 64, pp. 569–579, 3 2017.
- [15] L. An, P. Zhang, E. Adeli, Y. Wang, G. Ma, F. Shi, D. S. Lalush, W. Lin, and D. Shen, "Multi-Level Canonical Correlation Analysis for Standard-Dose PET Image Estimation," *IEEE Transactions on Image Processing*, vol. 25, pp. 3303–3315, 7 2016.
- [16] K. He, X. Zhang, S. Ren, and J. Sun, "Delving Deep into Rectifiers: Surpassing Human-Level Performance on ImageNet Classification," in *The IEEE International Conference on Computer Vision (ICCV)*, 12 2015.
- [17] Y. Taigman, M. Yang, M. Ranzato, and L. Wolf, "DeepFace: Closing the Gap to Human-Level Performance in Face Verification," in *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 6 2014.
- [18] S. Chetlur, C. Woolley, P. Vandermersch, J. Cohen, J. Tran, B. Catanzaro, and E. Shelhamer, "cuDNN: Efficient Primitives for Deep Learning," *CoRR*, vol. abs/1410.0759, 2014.
- [19] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, "Imagenet: A large-scale hierarchical image database," in *Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on*, pp. 248–255, 2009.
- [20] Y. LeCun, B. Boser, J. S. Denker, D. Henderson, R. E. Howard, W. Hubbard, and L. D. Jackel, "Backpropagation applied to handwritten zip code recognition," *Neural computation*, vol. 1, no. 4, pp. 541–551, 1989.
- [21] J. Xie, L. Xu, and E. Chen, "Image denoising and inpainting with deep neural networks," in *Advances in Neural Information Processing Systems*, pp. 341–349, 2012.
- [22] C. Dong, C. C. Loy, K. He, and X. Tang, "Learning a deep convolutional network for image super-resolution," in *European Conference on Computer Vision*, pp. 184–199, 2014.
- [23] X. Mao, C. Shen, and Y.-B. Yang, "Image restoration using very deep convolutional encoder-decoder networks with symmetric skip connections," in *Advances in Neural Information Processing Systems*, pp. 2802–2810, 2016.
- [24] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional Networks for Biomedical Image Segmentation," in *Medical Image Computing and Computer-Assisted Intervention MICCAI 2015: 18th International Conference, Munich, Germany, October 5-9, 2015, Proceedings, Part III* (N. Navab, J. Hornegger, W. M. Wells, and A. F. Frangi, eds.), pp. 234–241, Cham: Springer International Publishing, 2015.
- [25] Y. Han, J. J. Yoo, and J. C. Ye, "Deep Residual Learning for Compressed Sensing CT Reconstruction via Persistent Homology Analysis," *CoRR*, vol. abs/1611.06391, 2016.
- [26] E. Kang, J. Min, and J. C. Ye, "WaveNet: a deep convolutional neural network using directional wavelets for low-dose X-ray CT reconstruction," *CoRR*, vol. abs/1610.09736, 2016.
- [27] L. Xiang, Y. Qiao, D. Nie, L. An, W. Lin, Q. Wang, and D. Shen, "Deep auto-context convolutional neural networks for standard-dose PET image estimation from low-dose PET/MRI," *Neurocomputing*, 2017.
- [28] 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*, pp. 770–778, 2016.
- [29] K. Zhang, W. Zuo, Y. Chen, D. Meng, and L. Zhang, "Beyond a gaussian denoiser: Residual learning of deep cnn for image denoising," *IEEE Transactions on Image Processing*, 2017.
- [30] H. Zhao, O. Gallo, I. Frosio, and J. Kautz, "Loss functions for image restoration with neural networks," *IEEE Transactions on Computational Imaging*, vol. 3, no. 1, pp. 47–57, 2017.
- [31] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: from error visibility to structural similarity," *IEEE transactions on image processing*, vol. 13, no. 4, pp. 600–612, 2004.
- [32] Z. Wang, E. P. Simoncelli, and A. C. Bovik, "Multiscale structural similarity for image quality assessment," in *Signals, Systems and Computers, 2004. Conference Record of the Thirty-Seventh Asilomar Conference on*, vol. 2, pp. 1398–1402, 2003.
- [33] K. Ridgeway, J. Snell, B. D. Roads, R. S. Zemel, and M. C. Mozer, "Learning to generate images with perceptual similarity metrics," *arXiv preprint arXiv:1511.06409*, 2015.