IMAGE CORRECTION IN EMISSION TOMOGRAPHY USING DEEP CONVOLUTION NEURAL NETWORK

*Tomohiro Suzuki and Hiroyuki Kudo**

Department of Computer Science, Graduate School of Systems and Information Engineering,
University of Tsukuba, Tennoudai 1-1-1, Tsukuba 305-8573, Japan

ABSTRACT

We propose a new approach using Deep Convolution Neural Network (DCNN) to correct for image degradations due to statistical noise and photon attenuation in Emission Tomography (ET). The proposed approach first reconstructs an image by the standard Filtered Backprojection (FBP) without correcting for the degradations followed by inputting the degraded image into DCNN to obtain an improved image. We consider two different scenarios. The first scenario inputs an ET image only into DCNN, whereas the second scenario inputs a pair of degraded ET image and CT/MRI image to improve accuracy of the correction. The simulation result demonstrates that both the scenarios can improve image quality compared to the FBP without correction, and, in particular, accuracy of the second scenario is comparable to that of the standard iterative reconstruction such as Maximum Likelihood Expectation Maximization (MLEM) and Ordered-Subsets EM (OSEM) methods. The proposed method is able to output an image in very short time, because it does not rely on iterative computations.

Index Terms— Convolution neural network, emission tomography, image reconstruction, attenuation correction

1. INTRODUCTION

Traditionally, image reconstruction in Computed Tomography (CT) has been performed using analytical reconstruction or statistical (iterative) reconstruction. Very recently, the third approach based on deep learning has been investigated in CT community, which is beginning to demonstrate novel advantages in terms of image quality and computation time [1-4]. However, the current research has been still limited mainly to applications to low-dose x-ray CT, sparse-view x-ray CT, and noise reduction in Emission Tomography (ET), *i.e.* PET and SPECT. In this paper, we propose an application of deep learning to image correction in ET. In ET, image degradations due to statistical noise and photon attenuation inside the body occur so that these effects need to be corrected during the image reconstruction [5-7]. Currently, this task has been performed by using iterative reconstruction such as Maximum Likelihood Expectation Maximization (MLEM) and Ordered-Subsets EM (OSEM) methods [5-7]. The major drawbacks of this approach are 1) long computation time due to the iteration and 2) necessity of a rather complicated image processing step to construct an attenuation map of γ-ray photons from a CT/MRI image.

In this paper, we propose the following new approach, which aims at correcting for the effects of noise/attenuation by using deep learning. It also possesses a future potential to correct for other physical degradations such as detector blurring and photon scatters. First, an ET image is reconstructed by using the standard Filtered Backprojection (FBP) method without noise/attenuation correction. This degraded image is inputted to U-Net to obtain a corrected output image [8]. The training of U-Net is performed by using a large dataset of degraded FBP images, corresponding ideal (answer) images, and additional CT/MRI images of the same patient to accurately learn the process of attenuation correction. Using the CT/MRI image also matches to a recent trend of image reconstruction in ET, which utilizes information of CT/MRI to improve the reconstruction [5-7].

We have performed a simulation study using 13,000 2-D PET-CT images (augmented by decomposing a 3-D image into a number of 2-D images) acquired with a Philips scanner to diagnose soft-tissue sarcoma under good measurement conditions, which are disclosed in TCIA (The Cancer Image Archive) site [9]. The dataset used in the simulation was prepared by numerically simulating imaging process of SPECT including the Poisson noise and the attenuation. We used 12,000 examples for the training of U-Net and 1,000 examples for the performance test. We compared the proposed approach to the FBP method without noise/attenuation correction and the standard MLEM/OSEM iterative reconstruction with attenuation correction and regularization for smoothing. The result demonstrates that the proposed method is able to produce comparable images to those of MLEM/OSEM + smoothing very fast without using iterative computations. We also confirmed that the use of multi-modality image, *i.e.* pair of SPECT and CT, is very important to accurately learn the process of attenuation correction.

*This work was partially supported by JSPS KAKENHI (Grant Number 15K06103) and the JST-ERATO project (Grant Number JPMJER1403).

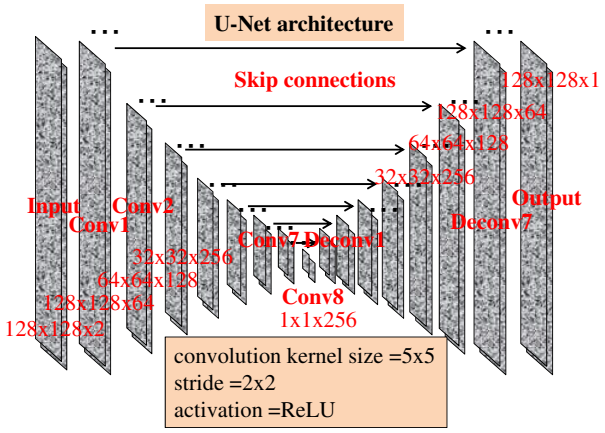

Fig. 1. Structure of U-Net adopted in this work.

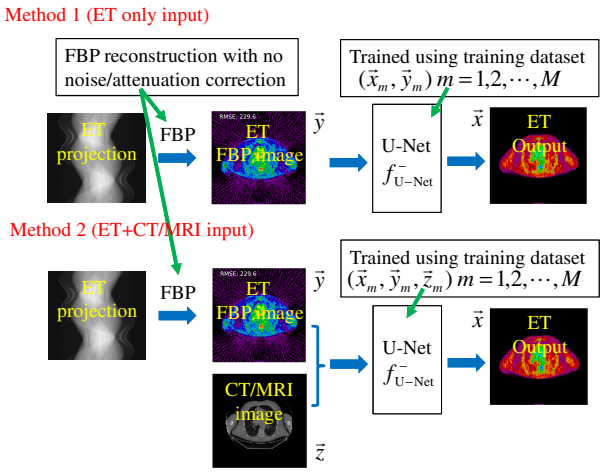

Fig. 2. Principle of the two proposed methods. Method 1 uses an ET image only as an input of U-Net, whereas Method 2 uses a pair of ET and CT/MRI images.

2. PROPOSED METHOD

2.1. Problem Formulation

First, we formulate the image correction problem in ET. Let $\vec{x} = (x_1, x_2, \ldots, x_J)^T$ denote an ideal ET image with no image degradations due to the statistical noise and the photon attenuation. Let $\vec{y} = (y_1, y_2, \ldots, y_J)^T$ denote a degraded ET image which was reconstructed by using the standard FBP method without noise/attenuation correction. Then, the relation between $\vec{x}$ and $\vec{y}$ is expressed as

$$\vec{y} = f(\vec{x}), \tag{1}$$

where $f(\cdot)$ denotes some nonlinear operator which represents the degradation process including the noise/attenuation. If the degradation process $f(\cdot)$ is known, an image $\vec{x}$ can be recovered from the degraded image $\vec{y}$ by a correction (or approximate inverse) operator $f^-(\cdot)$ as

$$\vec{x} = f^-(\vec{y}). \tag{2}$$

The aim of this work is to achieve the image correction in Eq. (2) by using Deep Convolution (or Convolutional) Neural Network (DCNN) as described in Section 2.2.

2.2. Proposed Method

In this work, we use U-Net known as a standard DCNN developed for the purpose of image conversion tasks such as denoising, segmentation, and superresolution [8]. The structure of U-Net is shown in Fig. 1. It consists of multiple convolution layers, multiple deconvolution layers, and skip connections introduced to bypass operations in the deep layers. The downsampling is performed between contiguous convolution layers, and the upsampling is performed between contiguous deconvolution layers. The operation performed in each layer is expressed by the following nonlinear operation.

$$x^{(o)}_{u,v} = \mathrm{ReLU}\Big(\sum_{p=-S}^{S}\sum_{q=-S}^{S} w_{p,q} x^{(i)}_{u-p,v-q} + b_{u,v}\Big), \tag{3}$$

where $\mathrm{ReLU}(\cdot)$ denotes the activation function called the Rectified Linear Unit. By training U-Net with a set of training images consisting of a pair of ideal ET image and corresponding degraded image, we are able to obtain the desirable correction function $f^-(\cdot)$. Of course, there is no guarantee that U-Net possesses an enough representation power to approximate the exact inverse function $f^{-1}(\cdot)$ with sufficient accuracy, but the simulation result shown in Section 3 is encouraging.

Next, we explain the proposed method in detail. In Fig. 2, we show the structure of the two proposed methods, *i.e.* Method 1 and Method 2. In Method 1, we first reconstruct an image by using the standard FBP without noise/attenuation correction to obtain a degraded image $\vec{y}$, which is inputted to trained U-Net to obtain an improved image with noise/attenuation correction. The training of U-Net is performed as follows. In applications to image reconstruction, use of real images for the training would not be practical due to the lack of sufficient number of images in many situations, so that we use a collection of artificially generated images through a numerical simulation of PET/SPECT imaging process. More details of the simulation process will be explained in Section 3. The training of U-Net is performed by minimizing the Mean Squares Error (MSE) loss function with L^2 regularization to avoid the overfitting, which is expressed as

$$L(\vec{W}) = \lambda \parallel \vec{W} \parallel^2 + \sum_{m=1}^{M} \parallel \vec{x}_m - f^-_{\text{U-Net}}(\vec{W})(\vec{y}_m) \parallel^2, \tag{4}$$

where $\vec{W}$ denotes the weights $w_{p,q}$ and the biases $b_{u,v}$ of U-Net, $f^{-}_{\mathrm{U-Net}}(\vec{W})(\cdot)$ is the input-output function of U-Net, and $(\vec{x}_m, \vec{y}_m)(m = 1, 2, \ldots, M)$ is a set of training images. We used the standard backpropagation incorporating Batch Normalization and Adam to facilitate the training [10,11]. The drawback of Method 1 is that accuracy of the attenuation correction is not enough, because it does not use any information of the attenuation map inside the body, *i.e.* γ-ray transmission scan image or CT/MRI image. Normally, for the purpose of attenuation correction by the iterative reconstruction, either CT/MRI image or transmission scan image is utilized. In Method 2, to improve the accuracy of attenuation correction closer to that of the iterative reconstruction, a CT/MRI image of the same patient is inputted to the U-Net in addition to the degraded ET image. In other words, U-Net with the multi-modality input and the single ET output is used. A current trend of the iterative reconstruction in ET is to use information of CT/MRI to improve the reconstruction in terms of not only the attenuation but also the noise. Method 2 matches to this trend, from which we expect that Method 2 significantly improves the accuracy of noise/attenuation correction. In Method 2, the training of U-Net is performed by using a collection of set of ideal image $\vec{x}_m$, degraded image $\vec{y}_m$, and CT/MRI image $\vec{z}_m$ by minimizing the following modified loss function.

$$L(\vec{W}) = \lambda \parallel \vec{W} \parallel^2 + \sum_{m=1}^{M} \parallel \vec{x}_m - f^{-}_{\mathrm{U\text{-}Net}}(\vec{W}) \begin{pmatrix} \vec{y}_m \\ \vec{z}_m \end{pmatrix} \parallel^2 . \quad (5)$$

3. SIMULATION STUDY

We have performed a simulation study using 2-D 13,000 PET-CT images (augmented by decomposing a 3-D image into a number of 2-D images) acquired with a Philips scanner to diagnose soft-tissue sarcoma under good measurement conditions, which are disclosed in TCIA site [9]. The dataset used in the simulation was prepared by numerically simulating SPECT tomographic imaging process including the Poisson noise and the photon attenuation. We used 12,000 examples for the training and 1,000 examples for the performance test. We compared Method 1 and Method 2 with the FBP method without correction and the standard iterative reconstructions by MLEM/OSEM with attenuation correction and regularization for smoothing. The simulation mainly consists of 1) generation of training images, 2) training of U-Net, and 3) application to validation images together with evaluation in terms of the MSE error metric and the CNR (Contrast-Noise-Ratio) metric. Below, we explain each step in detail.

3.1. Generation of Training Images

We constructed a set of training images through the simulation of SPECT imaging process. First, each ideal image $\vec{x}_m(m = 1, 2, \ldots, 12000)$ is forward projected with the attenuation followed by adding the Poisson noise to obtain the corresponding projection data. Then, the projection data was reconstructed by the FBP method using Shepp-Logan filter to obtain each degraded image $\vec{y}_m(m = 1, 2, \ldots, 12000)$. In SPECT, the degree of photon attenuation inside the body strongly depends on the energy of γ-ray photons used for imaging. We constructed four different datasets corresponding to the four Scenarios A1, A2, B1, and B2, which differ in terms of energies (120keV, 100keV) and availability of CT image. The summary of the created four datasets is specified in Table 1.

Table 1. Setup of the four scenarios in the simulation study (A1,A2,B1,B2).

Data	γ-ray energy	Input of U-Net	SPECT projection data
A1	120(keV)	SPECT only	512(bins)×128(angles) average counts=4.3×10^7
A2	120(keV)	SPECT+CT	512(bins)×128(angles) average counts=4.3×10^7
B1	100(keV)	SPECT only	512(bins)×128(angles) average counts=2.7×10^7
B2	100(keV)	SPECT+CT	512(bins)×128(angles) average counts=2.7×10^7

Table 2. The values of MSE metric per pixel averaged over 1,000 examples and the CNR metric for each method.

	FBP (A)	U-Net (A1)	U-Net (A2)	MLEM (A)	OSEM (A)
MSE	2.89	1.48	1.18	0.78	0.90
CNR	4.19	3.52	6.10	5.98	6.31
	FBP (B)	U-Net (B1)	U-Net (B2)	MLEM (B)	OSEM (B)
MSE	3.40	1.97	1.38	0.82	0.94
CNR	3.10	3.44	5.40	5.87	5.11

3.2. Training of U-Net and Application to Testing Images

We used Tensorflow framework with a GPU (NVidia Tesla K80) for the implementation of DCNN. Before the training, contrasts of all the images were normalized in such a way that the intensity range becomes $[-1, 1]$. The training was performed by using the MSE loss functions in Eqs. (4) and (5) with Adam (the learning rate parameters $\alpha = 0.0001, \beta_1 = 0.5, \beta_2 = 0.999$ and the regularization parameter $\lambda = 1.0 \times 10^{-7}$) [9,10]. The number of epochs in the learning process to reach to a saturation of validation loss was 10, and the computation time was approximately 10 (hours) for each dataset.

For each scenario (A1,A2,B1,B2), trained U-Net was applied to 1,000 testing images. The processing time to process

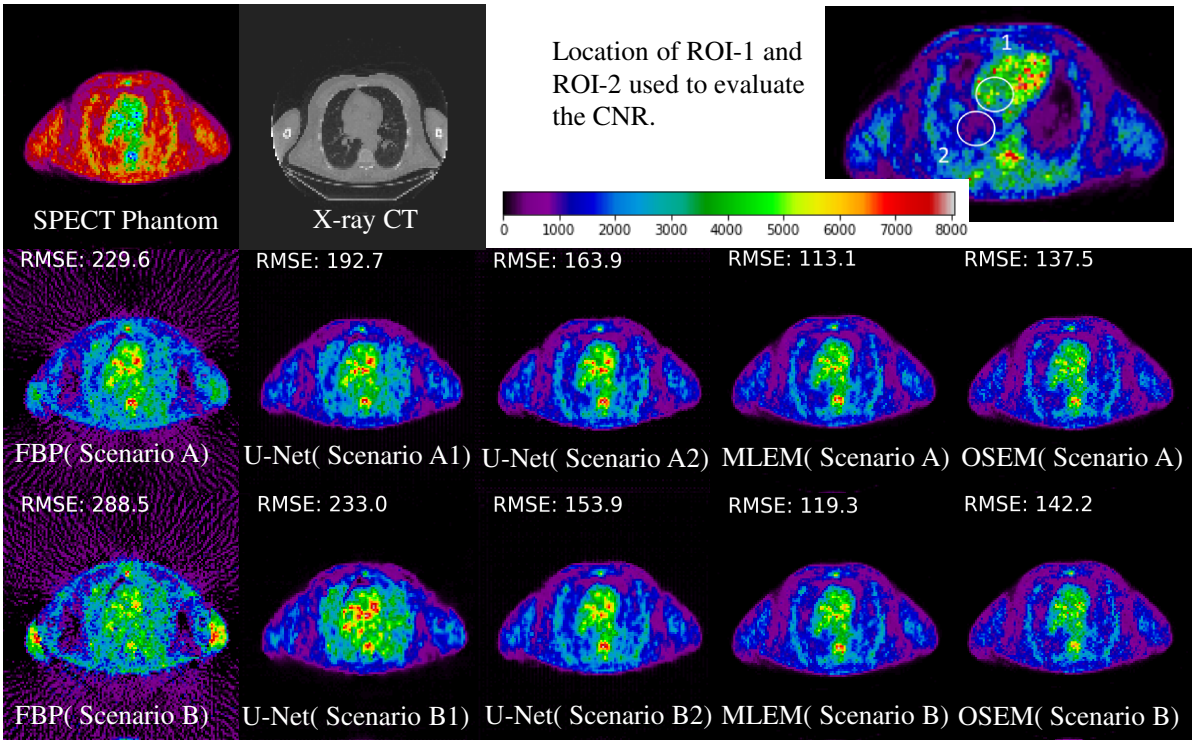

Fig. 3. Comparison of reconstructed images by various methods (FBP, U-Net using SPECT only input, U-Net using SPECT+CT input, MLEM, OSEM) for the scenarios A and B.

each example was approximately 0.1 (s), which was dramatically shorter compared to the iterative reconstruction.

3.3. Results

For each scenario (A1,A2,B1,B2), we have compared the results obtained by the proposed methods to those obtained by the FBP without noise/attenuation correction and MLEM/OSEM with attenuation correction and regularization for smoothing. The regularization in MLEM/OSEM was performed by smoothing the image after every iteration by using an empirically optimized Gaussian filter of $\sigma = 0.15$ (pixels). In Fig. 3, we show examples of the reconstructed images for each scenario. It can be observed that the cupping artifacts and the streak artifacts occurred in the FBP method without correction were significantly reduced by the proposed methods. The cupping artifacts mainly originate from the effect of attenuation, and the streak artifacts mainly originate from the effect of statistical noise. In particular, the scenarios using the CT images (A2,B2) produced accurate images, which are closer to those by MLEM/OSEM reconstruction using information of exact attenuation maps. We also performed a numerical evaluation by using the MSE metric. The MSE values averaged over 1,000 examples for each scenario and each method are summarized in Table 2.

We also performed the evaluation of CNR metric using two region-of-interests, *i.e.* ROI-1 (organ region) and ROI-2 (cavity region) indicated in Fig. 3 (top-right). This metric is often used to evaluate image quality in tomographic imaging, which is defined by

$$\mathrm{CNR} = \frac{(\text{Mean of ROI-1}) - (\text{Mean of ROI-2})}{[(\text{SD of ROI-1}) + (\text{SD of ROI-2})]/2}, \qquad (6)$$

where SD denotes the standard deviation. The value of CNR is large when the image contrast is large and the statistical noise level is low. In Table 2, we summarize the CNR values for various cases. The result demonstrates that the use of CT image (A2,B2) significantly improves the CNR value compared to the SPECT only case (A1,B1).

4. CONCLUSION

We proposed a new approach using DCNN to correct for image degradations in ET. Unlike the standard iterative reconstruction, the proposed method allows us to correct for the image degradations due to the noise/attenuation in ET without iterative computations, so that the computation time can be significantly reduced.

5. REFERENCES

[1] H.Chen, Y.Zhang, W.Zhang *et al.* "Low-dose CT via deep neural network," Biomed.Opt.Express, Vol.8, pp.679-694, 2017.

[2] K.H.Jin, M.T.McCann, E.Froustey *et al.* "Deep convolutional neural network for inverse problems in imaging," IEEE Trans.Image Process., Vol.26, pp.4509-4522, 2017.

[3] K.Kim, D.Wu, K.Gong *et al.* "Penalized PET reconstruction using deep learning prior and local linear fitting," IEEE Trans.Med.Imaging, Vol.37, pp.1478-1487, 2018.

[4] J.Xu, E.Gong, J.Pauly *et al.* "200x low-dose PET reconstruction using deep learning," arXiv, 1712.04119, 2017.

[5] H.Zaidi "Quantitative analysis in nuclear medicine imaging," Springer Verlag, 2005.

[6] J.A.Patton and T.G.Turkington "SPECT/CT physical principles and attenuation correction," J.Nucl.Med.Tech., Vol.2008, pp.1-10, 2008.

[7] P.E.Kinahan, D.W.Townsend, T.Beyer *et al.* "Attenuation correction for a combined 3D PET/CT scanner," Med.Phys., Vol.25, pp.2046-2053, 1998.

[8] O.Ronneberger, P.Fischer, and T.Brox "U-Net: convolutional networks for biomedical image segmentation," arXiv, 1505.04597, 2015.

[9] K.Clark, B.Vendt, K.Smith *et al.* "The cancer imaging archive (TCIA): maintaining and operating a public information repository," J.Digital Imaging, Vol.26, pp.1045-1057, 2013.

[10] S.Ioffe and C.Szegedy "Batch normalization: accelerating deep network training by reducing internal covariate shift," arXiv, 1502.03167, 2015.

[11] D.P.Kingma and J.Ba "Adam: a method for stochastic optimization," arXiv, 1712.05877, 2017.