Title

Motion correction of respiratory-gated PET images using deep learning based image registration framework

Journal

Physics in Medicine and Biology, 65(15)

ISSN

0031-9155

Authors

Li, Tiantian
Zhang, Mengxi
Qi, Wenyuan
et al.

Publication Date

2020-08-07

DOI

10.1088/1361-6560/ab8688

Peer reviewed

HHS Public Access

Author manuscript

Phys Med Biol. Author manuscript; available in PMC 2021 July 30.

Published in final edited form as:

Phys Med Biol. ; 65(15): 155003. doi:10.1088/1361-6560/ab8688.

Motion Correction of Respiratory-Gated PET Images Using Deep Learning Based Image Registration Framework

Tiantian Li[a], **Mengxi Zhang**[a], **Wenyuan Qi**[b], **Evren Asma**[b], **Jinyi Qi**[a]

[a]Department of Biomedical Engineering, University of California, Davis, CA 95616, USA;

[b]Canon Medical Research USA, Inc., 706 N Deerpath Dr., Vernon Hills, IL 60061, USA

Abstract

Purpose: Artifacts caused by patient breathing and movement during PET data acquisition affect image quality. Respiratory gating is commonly used to gate the list-mode PET data into multiple bins over a respiratory cycle. Non-rigid registration of respiratory-gated PET images can reduce motion artifacts and preserve count statistics, but it is time consuming. In this work, we propose an unsupervised non-rigid image registration framework using deep learning for motion correction.

Methods: Our network uses a differentiable spatial transformer layer to warp the moving image to the fixed image and uses a stacked structure for deformation field refinement. Estimated deformation fields were incorporated into an iterative image reconstruction algorithm to perform motion compensated PET image reconstruction. We validated the proposed method using simulation and clinical data and implemented an iterative image registration approach for comparison. Motion compensated reconstructions were compared with ungated images.

Results: Our simulation study showed that the motion compensated methods can generate images with sharp boundaries and reveal more details in the heart region compared with the ungated image. The resulting normalized root mean square error (NRMS) was 24.3±1.7% for the deep learning based motion correction, 31.1±1.4% for the iterative registration based motion correction, and 41.9±2.0% for ungated reconstruction. The proposed deep learning based motion correction reduced the bias compared with the ungated image without increasing the noise level and outperformed the iterative registration based method. In the real data study, both motion compensated images provided higher lesion contrast and sharper liver boundaries than the ungated image and had lower noise than the reference gate image. The contrast of the proposed method based on the deep neural network was higher than the ungated image and iterative registration method at any matched noise level.

Conclusions: In this work, we proposed a motion correction method for respiratory-gated PET images using a deep learning based image registration framework. The proposed approach is very convenient to implement as it does not require the knowledge of the true deformation field for training the network. We validated the proposed method using simulation and clinical data and showed its ability to reduce motion artifacts while utilizing all gated PET data.

Disclosure

None

Keywords

deep learning; motion correction; PET; image reconstruction

1. Introduction

Artifacts caused by patient breathing and movement during positron emission tomography (PET) data acquisition affect image quality and lead to underestimation of tumor activity and overestimation of tumor volume (Kalantari *et al.*, 2016). Lesion detectability also suffers from motion blurring since small lesions are likely to remain undetected, which may result in misdiagnosis (Nehmeh *et al.*, 2002b). Respiratory gating has been used to gate list-mode PET data into multiple bins over a respiratory cycle based on either an external hardware or a data-driven self-gating technique (Chan *et al.*, 2017; Büther *et al.*, 2009). Within each time bin, the motion blurring is assumed to be negligible (Nehmeh *et al.*, 2002a). The motion-frozen images can be reconstructed gate-by-gate using the data from each bin. However, gated PET reconstructed images suffer from low signal-to-noise ratio since the count level is low in each gate. One of the widely used methods to reduce noise is to utilize events from all gates by incorporating a motion model into the reconstruction procedure. While the motion information can be obtained from high resolution anatomical images, e.g. computed tomography (CT) (Lamare *et al.*, 2007) or magnetic resonance imaging (MRI) (Fayad *et al.*, 2015), utilizing other image modalities always leads to multiple issues, such as extra time and cost, image co-registration, extra radiation dose from the CT scan and synchronization issues between the scanners. Accurate non-rigid registration based on gated PET images themselves is challenging due to their high noise levels and is also time-consuming. Recently, deep learning techniques have provided new approaches for either supervised (Sokooti *et al.*, 2017; Krebs *et al.*, 2017) or unsupervised (Balakrishnan *et al.*, 2018; Li and Fan, 2018; Lau *et al.*, 2019) image registration.

The supervised convolutional neural network (CNN) architecture aims to find a mapping between the learned image features and the deformation field that registers the training image pairs. Training these kinds of networks relies on the knowledge of the true deformation field; therefore, training pairs were usually simulated by warping existing images with artificially generated deformation fields (Sokooti *et al.*, 2017; Krebs *et al.*, 2017). For real data, the ground truth deformation field is usually substituted with the estimate from an iterative image registration algorithm (Liao *et al.*, 2017). However, in medical imaging, an accurate ground-truth deformation between image pairs is difficult to obtain which limits the application of supervised learning.

To address this difficulty, a spatial transformer network (STN) (Jaderberg *et al.*, 2015) has been proposed to warp images, which enables neural networks to perform unsupervised learning without knowing the true deformation field. The combination of stacked CNNs and STNs have been proposed recently to learn the image feature representations and a mapping between the image features and the deformation field at the same time (Balakrishnan *et al.*, 2018; Li and Fan, 2018; Lau *et al.*, 2019). The optimization is realized by minimizing an image dissimilarity metric between fixed and warped moving images with a proper

regularization. This unsupervised approach does not require ground truth for training the network, which makes it very convenient to implement. In this work, we propose an unsupervised non-rigid image registration framework using deep learning and incorporate the estimated deformation field into PET image reconstruction for motion correction. The rest of this paper is organized as follows. Section 2 introduces the theory of the unsupervised deep learning model and the motion compensated reconstruction. Section 3 describes the details of the simulation setup, patient data acquisition, and methods for quantitative data analysis. Results are shown in section 4. Finally, we discuss findings and limitations in section 5 and draw conclusions in section 6.

2. Theory

2.1 Unsupervised deep learning model for deformable motion estimation

The goal of this registration network is to predict a deformation field ($\boldsymbol{\theta}$) between a moving image ($\boldsymbol{x}_{\mathrm{m}}$) and a fixed image ($\boldsymbol{x}_f$) that minimizes a penalized loss function:

$$\boldsymbol{\theta} = argmin_{\boldsymbol{\theta}} - S\left(\boldsymbol{T}(\boldsymbol{x}_m, \boldsymbol{\theta}),\ \boldsymbol{x}_f\right) + \lambda\, U(\boldsymbol{\theta}), \tag{1}$$

Where $S(\cdot,\cdot)$ is an image similarity metric, the operator $\boldsymbol{T}(\boldsymbol{x}_{\mathrm{m}},\boldsymbol{\theta})$ deforms $\boldsymbol{x}_{\mathrm{m}}$ based on the deformation field $\boldsymbol{\theta}$, $U(\boldsymbol{\theta})$ is a regularization function on $\boldsymbol{\theta}$, and λ is a weighting factor.

In this study, we use cross correlation (CC) (Balakrishnan *et al.*, 2018) as the similarity metric. The CC between fixed and warped moving images is defined as:

$$CC\left(M, \boldsymbol{x}_f\right) = \sum\nolimits_{i\ \in\ \Omega} \left(\frac{\sum_{p_i} \left(x_f(p_i) - \bar{x}_f(i)\right)\left(M(p_i) - \overline{M}(i)\right)}{\sqrt{\sum_{p_i} \left(x_f(p_i) - \bar{x}_f(i)\right)^2 \sum_{p_i} \left(M(p_i) - \overline{M}(i)\right)^2}} \right), M = \boldsymbol{T}(\boldsymbol{x}_m, \boldsymbol{\theta}) \tag{2}$$

where p_i iterates over a 9 × 9 × 9 cubic volume around voxel i, $\bar{x}_f(i)$ and $\overline{M}(i)$ are the mean values inside the 9 × 9 × 9 cubic volume around voxel i in the two images, respectively, and Ω is the set containing all the voxel indices.

In order to obtain a smoothed $\boldsymbol{\theta}$, an L-2 norm regularizer on the gradients of the deformation field is applied. Then the loss function can be written as

$$Loss\left(\boldsymbol{x}_m,\ \boldsymbol{x}_f, \boldsymbol{\theta}\right) = \ - CC\left(\boldsymbol{T}(\boldsymbol{x}_m, \boldsymbol{\theta}),\ \boldsymbol{x}_f\right) + \lambda \sum \| \nabla \boldsymbol{\theta} \|^2 . \tag{3}$$

Our implementation of the network is inspired by the Voxelmorph (Balakrishnan *et al.*, 2018) and FlowNet 2.0 (Ilg *et al.*, 2017) structures. We employ a stacked framework which comprises of three subunits. Each subunit consists of a 9-layer encoder-decoder called "RegNet" and a STN. The RegNet consists of a series convolution, down-sampling, and up-sampling layers with concatenation as shown in Fig. 1. The kernel size is 3×3×3 in all convolutional layers. Batch Normalization (BN) (Ioffe and Szegedy, 2015), Dropout (Srivastava *et al.*, 2014) and Leaky Rectified Linear Unit activation (Leaky ReLU) (Maas *et*

al., 2013) are used after each convolutional layer. Unsupervised learning is realized by using the differentiable STN (Jaderberg *et al.*, 2015) to warp the moving image and by comparing the warped image to the fixed image. The warped image based on the intermediate estimate is then fed into the next subunit for refinement. The estimated deformation field of each subunit is combined with the result from the previous subunit through a convolution layer. We applied a Gaussian filter to the input gated PET images to reduce noise.

2.2 Motion compensated PET image reconstruction

Let $\boldsymbol{y}^k$, $k \in \{1,2,\ldots K\}$, be the measured PET data in the kth gate and $\bar{\boldsymbol{y}}^k$ be the expectation of the measurement. The log-likelihood of all gated PET data is

$$\boldsymbol{L}\left(\boldsymbol{y}^{\mathbf{1}}, \ldots, \boldsymbol{y}^{\boldsymbol{K}} | \boldsymbol{x}\right) = \sum_k \sum_i \left\{ y_i^k \, ln(\bar{y}_i^k) - \bar{y}_i^k \right\}. \tag{4}$$

The expectation $\bar{\boldsymbol{y}}^k$ is related to the reference gate PET image $\boldsymbol{x}$ through $\bar{\boldsymbol{y}}^k = w^k \cdot \mathbf{N} \cdot \mathbf{A}^k \cdot \boldsymbol{P} \cdot \boldsymbol{T}(\boldsymbol{x}, \boldsymbol{\theta}^k) + \boldsymbol{S}^k + \boldsymbol{r}^k$, where the (i, j)th element of $\boldsymbol{P} \in \mathbb{R}^{\mathrm{M}\times\mathrm{N}}$, $p_{i,j}$, denotes the probability of detecting an emission from pixel j, $j \in \{1,\ldots,N\}$, at detector pair i, $i \in \{1, \ldots,\mathrm{M}\}$, $N \in \mathbb{R}^{\mathrm{M}\times\mathrm{M}}$ and $\mathbf{A}^k \in \mathbb{R}^{\mathrm{M}\times\mathrm{M}}$ are diagonal matrices containing the normalization factors and attenuation factors, respectively, for the k^{th} gate, $\boldsymbol{s}^k \in \mathbb{R}^{\mathrm{M}\times 1}$ denotes the expectation of scattered events and $r^k \in \mathbb{R}^{\mathrm{M}\times 1}$ denotes the expectation of random events for the k^{th} gate, and $\boldsymbol{\theta}^k$ is the estimated deformation field from the reference gate to the k^{th} gate. The weighting factor w^k accounts for the duration of gate k with $\sum_k w^k = 1$.

The maximum likelihood expectation maximization (ML-EM) iteration is (Li *et al.*, 2006)

$$\boldsymbol{x}^{\boldsymbol{n}+\mathbf{1}} = \frac{\boldsymbol{x}^{\boldsymbol{n}}}{\sum_{\boldsymbol{k}} w^k \cdot \boldsymbol{T}^*\left(\boldsymbol{u}^k, \boldsymbol{\theta}^k\right)} \cdot \sum_{\boldsymbol{k}} w^k \cdot \boldsymbol{T}^*\left(\boldsymbol{P}^T \frac{\boldsymbol{y}^k}{w^k \cdot \boldsymbol{P} \cdot \boldsymbol{T}\left(\boldsymbol{x}^n, \boldsymbol{\theta}^k\right) + \frac{\boldsymbol{s}^k + \boldsymbol{r}^k}{\boldsymbol{N} \cdot \boldsymbol{A}^{\boldsymbol{k}}}}, \boldsymbol{\theta}^k\right), \tag{5}$$

Where $\boldsymbol{u}^k$ denotes the sensitivity image with $u_j^k = \sum_i [\mathbf{N} \cdot \mathbf{A}^k \cdot \mathbf{P}]_{ij}$, and the multiplications and divisions between the vectors are performed element-wise. The update procedure begins with deforming current image from the reference gate to other gates. After the deformation, a standard forward projection and back projection are performed to get the error image of that gate. All error images are deformed back to the reference gate and summed together. Finally, the current image is updated using the summation of deformed error images.

3. Data Generation and Evaluation

3.1. Simulation study

Twenty-two XCAT phantoms (Segars *et al.*, 2010) with various organ sizes and genders (11 Male and 11 Female) were generated (10 for training, 1 for validation and 11 for testing). Each modeled a respiratory motion amplitude between 2 and 4 cm with a period of 5 sec.

The respiratory cycle was divided into 8 gates, each with a matched attenuation map. Activity parameters included a 5% variation to simulate the population difference. A Canon PET/CT scanner geometry was simulated using the SimSET Monte-Carlo toolkit (Harrison *et al.*, 1993). The scanner consists of 40 detector blocks arranged in a ring of diameter 90.9 cm. Each block contains 16×48 lutetium-based scintillation crystals. The individual crystal size is $4 \times 4 \times 12$ mm^3. We simulated a 200 MBq ^{18}F-FDG injection and a 20 min PET scan starting from 1 hour post-injection (Zhang *et al.*, 2014). We only use true coincidences for reconstruction, since we assume perfect scatter and random correction. For motion estimation, gated PET images were first reconstructed using the ML-EM algorithm (50 iterations) with normalization and attenuation corrections. The image matrix size was 128×128×48 with a voxel size of 4.08× 4.08 ×4.08 mm^3. Sample images are shown in Fig. 2. The end-inspiration phase was chosen as the reference gate. In total 70 (7 moving gates for each of the 10 phantoms) 3D training pairs were generated, each containing 48 axial slices.

The network was implemented using Keras 2.2.4 with a Tensorflow 1.5.0 backend. We use the adaptive moment estimation (ADAM) optimizer with a learning rate of 0.005 and batch size of 1. Moving-reference image pairs were fed into the network for training and the network was trained with 3000 epochs. After the training, deformation fields ($\boldsymbol{\theta}$) between any pair of images can be estimated by feeding the moving and fixed images into the network.

For comparison, we also implemented an iterative image registration with regularization to encourage the deformation to be invertible (Chun and Fessler, 2009b) using a publicly available Bspline toolbox[1]. We used the default weighted-least-squares similarity measure. The number of iterations was chosen to be 200 based on the visual assessment of the deformation field. Motion compensated reconstructions were performed by running the ML-EM algorithm in (5) for 50 iterations using deformation fields estimated either from the neural network or the iterative registration software.

3.2. Real data study

A patient dataset was obtained from the Canon whole-body TOF PET/CT scanner (230 MBq ^{18}F-FDG injection, 90 minutes uptake). Two 50% overlapping bed positions were acquired, at 14 mins per bed. The list-mode data were divided into 7 respiratory gates based on an externally measured respiratory signal (Heinz *et al.*, 2015). Events in irregular breathing cycles were rejected (bed 1: 9.9%, bed 2: 22.0%). Gated PET data were first reconstructed with ML-EM (30 iterations) to estimate the motion fields, which were then fed into the motion compensated reconstruction for 50 iterations. The image matrix size was 152×152×48 with a voxel size of 4.08× 4.08 ×4.08 mm^3. The normalization factors were computed based on a uniform cylinder scan. The attenuation factors were obtained from a helical CT scan. The attenuation map was not gated and was used for all gates. Randoms were estimated using the delayed window method. Scatters were estimated using the single-scatter simulation. The reconstructed gated images are shown in Fig. 3. Due to the difference

[1]Part of Michigan Image Reconstruction Toolbox (MIRT) from http://web.eecs.umich.edu/~fessler/code/index.html

between real data and simulation, the network was fine-tuned using the first-bed data with 500 epochs and then applied to the second-bed data for evaluation.

3.3. Evaluation figures of merit

For the simulation study, we reconstructed the reference gate with 8x counts (same as the ungated data) and used it as the ground truth for quantitative evaluation. We calculate the normalized root mean square error (NRMS) between different reconstructions and the ground truth:

$$\mathbf{NRMS} = \frac{1}{\|\bar{\boldsymbol{x}}\|_2}\sqrt{\sum_{i=1}^{N}|x_i - \bar{x}_i|^2}, \tag{6}$$

Where $\boldsymbol{x}$ denotes the ungated image or a motion compensated reconstructed image and $\bar{\boldsymbol{x}}$ denotes the ground truth. N denotes the number of voxels in the image.

For region of interest (ROI) quantification, we calculated the bias compared with the original phantom in the left and right myocardium regions and the standard deviation (STD) in the lung background. The percentage difference relative to the mean was used for both bias and STD.

For the real data study, a lesion ROI was drawn on the reference gate image for quantification. Due to the lack of a ground truth, a contrast-noise curve was used for the evaluation. The contrast was calculated by taking the ratio between the mean of the lesion ROI and the mean of a background ROI in the liver. The background noise was calculated as the standard deviation of the liver ROI over its mean.

4. Results

4.1. Simulation study

Fig. 4 shows a coronal slice of the reconstructed images of a test phantom with and without (ungated) motion correction. The ungated image looked blurry in the cardiac region and near the liver-lung boundary, while the motion compensated method using either the deep neural network or the iterative registration for deformation field estimation can generate images with sharp boundaries and reveal more details in the heart region. We plotted the bias-variance curves of the left and right myocardium ROIs in Fig. 1. All the methods were plotted by 10-MLEM-iteration intervals except the reference gate which also included 1 to 9 iterations for better comparison. In all cases, STD increased with increasing iterations. For both ROIs, the reference gate image exhibited much higher noise than either the ungated image or the motion corrected images. Both motion corrected reconstructions reduced the bias compared with the ungated image without increasing the noise level and the deep learning approach outperformed the iterative registration method with less bias.

For all 11 test phantoms, we computed the NRMS of reconstructed images at the 30th MLEM iteration and the results are shown in Table 1. The mean and standard deviation of the NRMS values were 24.3±1.7% for the deep learning based motion correction, 31.1±1.4% for the iterative registration based motion correction, 41.9±2.0% for ungated

reconstruction, and 42.6±4.0% for the reference gate reconstruction. Clearly the proposed deep learning based method achieved the best performance.

4.2. Real data study

Fig. 6 shows a coronal slice of the reconstructed motion corrected and ungated images in comparison to the reconstructed image of the reference gate. Both motion compensated images provided higher lesion contrast and sharper liver boundaries than the ungated image and had lower noise than the reference gate image. For quantitative comparison, we plotted the contrast-noise curves for a lung lesion in Fig. 1(a). As expected, the contrast of the proposed method based on the deep neural network was higher than that of the ungated image and iterative registration method at any matched background STD level.

4.3. Computational time

Deep learning is a very fast registration approach after a one-time training process, which took a few days on an Nvidia GeForce GTX 1080ti GPU. Depending on the image size, the iterative registration time of a single pair of images took about 15 to 20 mins, while for the deep learning method it only took 8 secs to register a test pair of images. The runtimes of the neural network and iterative registration are compared in Fig. 1(b), with the neural network being about 100 times faster than the iterative method. This fast registration of PET images significantly speeds up image analysis and processing pipelines, which can facilitate novel directions in respiratory and cardiac motion correction and body movement during long time dynamic PET scans.

5. Discussion

In this work, we have demonstrated the feasibility of incorporating 4D deformation fields from deep learning into motion compensated PET image reconstruction. In addition to the computational speed advantage, we note a higher registration accuracy for the deep learning method compared with the iterative registration method. This could be attributed to the training of the deep neural network using an ensemble of image pairs, which improved the robustness of the image registration. Our methods rely on an unsupervised spatial registration model, which does not need ground truth. Compared with supervised learning which uses ground truth deformation fields from iterative registration, unsupervised approaches have the potential to achieve better registration accuracy. In addition, to account for both potential large displacements and fine deformations between images, the stacked architecture was used to estimate coarse-to-fine deformation fields, where the front layer estimates a coarse deformation field with large displacements, and the back layers provide fine deformation field. Although our network was trained based on respiratory motions, it appears to be capable of handling other kinds of motion to some degree. To demonstrate this, we superimposed a moderate bulk body motion (1–3 voxels) on the simulated respiratory motion and fed the images into the network without any retraining. The RMSE between the predicted warped image and the fixed image increased by 1.3% for 1-voxel shift and 13.7% for 3-voxel shift compared to the results without the bulk body motion. Obviously, motion fields having a larger deviation from the training data would require either retraining or fine tuning as we did for the patient data. This problem could also be

solved by increasing the training data size to better match the real data and it remains a part of future work in our research.

Since in this study we focused on the motion estimation, we used motion-independent correction factors for attenuation, random and scattered events. The expectation of the scattered and random events were estimated based on ungated emission data. The randoms estimation is the least sensitive to motion, because it was determined by the singles rates of the detectors. Scatter distribution is also less affected by respiratory motion than true coincidences because it has a much smoother distribution than the true events. The correction factor that is the most sensitive to motion is the attenuation factor. Approaches to compensate for motion in attenuation factors have been proposed (Alessio *et al.*, 2007; Lu *et al.*, 2018) and can be combined with the deep learning based motion estimation method proposed here. Modeling the motion for attenuation map and scattered events will be included in the future work.

One limitation of our registration network is that the invertibility of deformations is not considered. Recently, Dalca et al. proposed a deep learning architecture which can estimate probabilistic diffeomorphic deformations that is differentiable and invertible, and thus preserve topology (Dalca et al., 2018). This new network can also be incorporated in the motion compensated image reconstruction framework.

Furthermore, due to the lack of gated patient data, we only validate our proposed method using one patient with 50% overlapped bed positions. The fine-tuned model might be overfitted to the motion characteristics and tracer distribution specific to this patient. To address this concern, we also deployed our fine-tuned model on another patient scan with three respiratory gated phases for deformation fields estimation (Fig. 8). This dataset has a variable count distribution across the three gates with 11% in gate 1, 31% in gate 2, and 58% in gate 3. The image reconstruction parameters were the same as those given in Section 3.2. Motion compensated reconstructions were performed using deformation fields estimated either from the deep learning or the iterative registration. Fig. 9 show the reconstructed images for the motion corrected and ungated images with comparison to the reconstructed image of the reference gate. Higher lesion contrast and sharper liver boundary were observed for deep learning based method. The iterative registration however failed to capture the lesion motion near the liver boundary since the motion amplitude for this patient was relatively small (1~2 voxels). The contrast-noise curve for the deep learning based registration was also higher than the iterative based registration and unaged reconstruction (Fig. 10). Since reference gate contains over 50% of the counts, the benefit from motion compensation is not significant. Further investigation with systematic analysis on a large population of patients is needed to validated our proposed method.

It also worth noting that our deformation fields are estimated from the gated PET images before the motion compensated image reconstruction. Although there have been many works on joint estimation of the image and motion fields during image reconstruction (Wang and Qi, 2015; Chun and Fessler, 2009a), the nonconvex optimization problem in the joint estimation methods is very challenging and it is hard to guarantee monotonicity of the

algorithm. Incorporating the deep learning method in a joint estimation framework with guaranteed convergence is of great research interest.

6. Conclusion

In this work, we proposed a motion correction method of respiratory-gated PET images using a deep learning-based image registration framework. It does not require ground truth for training the network, which makes it very convenient to implement. We validated the proposed method using simulation and clinical data and showed its ability to reduce the motion artifacts while utilizing all gated PET data.

Acknowledgments

This work is supported in part by the National Institutes of Health under Grant R01EB000194.

References

Alessio AM, Kohlmyer S, Branch K, Chen G, Caldwell J and Kinahan P 2007 Cine CT for attenuation correction in cardiac PET/CT J Nucl Med 48 794–801 [PubMed: 17475969]

Balakrishnan G, Zhao A, Sabuncu MR, Guttag J and Dalca AV 2018 An Unsupervised Learning Model for Deformable Medical Image Registration Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition 9252–60

Büther F, Dawood M, Stegger L, Wübbeling F, Schäfers M, Schober O and Schäfers KP 2009 List mode–driven cardiac and respiratory gating in pet J Nucl Med 50 674–81 [PubMed: 19372491]

Chan C, Onofrey J, Jian Y, Germino M, Papademetris X, Carson RE and Liu C 2017 Non-rigid event-by-event continuous respiratory motion compensated list-mode reconstruction for PET IEEE transactions on medical imaging 37 504–15 [PubMed: 29028189]

Chun SY and Fessler JA 2009a Joint image reconstruction and nonrigid motion estimation with a simple penalty that encourages local invertibility *Medical Imaging 2009*: Physics of Medical Imaging 7258 72580U

Chun SY and Fessler JA 2009b A simple regularizer for B-spline nonrigid image registration that encourages local invertibility IEEE J Sel Top Signal Process 3 159–69 [PubMed: 21179392]

Dalca AV, Balakrishnan G, Guttag J and Sabuncu MR 2018 Unsupervised learning for fast probabilistic diffeomorphic registration International Conference on Medical Image Computing and Computer-Assisted Intervention 729–38

Fayad H, Schmidt H, Wuerslin C and Visvikis D 2015 Reconstruction-incorporated respiratory motion correction in clinical simultaneous PET/MR imaging for oncology applications J Nucl Med 56 884–9 [PubMed: 25908830]

Harrison R, Haynor D, Gillispie S, Vannoy S, Kaplan M and Lewellen T 1993 A public-domain simulation system for emission tomography-photon tracking through heterogeneous attenuation using importance sampling J Nucl Med 34 60

Heinz C, Reiner M, Belka C, Walter F and Söhn M 2015 Technical evaluation of different respiratory monitoring systems used for 4D CT acquisition under free breathing J Appl Clin Med Phys 16 334–49 [PubMed: 26219016]

Ilg E, Mayer N, Saikia T, Keuper M, Dosovitskiy A and Brox T 2017 Flownet 2.0: Evolution of optical flow estimation with deep networks Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition 2 6

Ioffe S and Szegedy C 2015 Batch normalization: Accelerating deep network training by reducing internal covariate shift arXiv preprint arXiv:1502.03167

Jaderberg M, Simonyan K and Zisserman A 2015 Spatial transformer networks Advances in neural information processing systems 2017–25

Kalantari F, Li T, Jin M and Wang J 2016 Respiratory motion correction in 4D-PET by simultaneous motion estimation and image reconstruction (SMEIR) Phys Med Biol 61 5639 [PubMed: 27385378]

Krebs J, Mansi T, Delingette H, Zhang L, Ghesu FC, Miao S, Maier AK, Ayache N, Liao R and Kamen A 2017 Robust non-rigid registration through agent-based action learning International Conference on Medical Image Computing and Computer-Assisted Intervention 344–52

Lamare F, Carbayo ML, Cresson T, Kontaxakis G, Santos A, Le Rest CC, Reader A and Visvikis D 2007 List-mode-based reconstruction for respiratory motion correction in PET using non-rigid body transformations Phys Med Biol 52 5187 [PubMed: 17762080]

Lau T, Luo J, Zhao S, Chang EI and Xu Y 2019 Unsupervised 3D End-to-End Medical Image Registration with Volume Tweening Network arXiv preprint arXiv:1902.05020

Li H and Fan Y 2018 IEEE 15th International Symposium on Biomedical Imaging 1075–8

Li T, Thorndyke B, Schreibmann E, Yang Y and Xing L 2006 Model-based image reconstruction for four-dimensional PET Med Phys 33 1288–98 [PubMed: 16752564]

Liao R, Miao S, de Tournemire P, Grbic S, Kamen A, Mansi T and Comaniciu D 2017 An artificial agent for robust image registration Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence

Lu Y, Fontaine K, Mulnix T, Onofrey JA, Ren S, Panin V, Jones J, Casey ME, Barnett R and Kench P 2018 Respiratory motion compensation for PET/CT with motion information derived from matched attenuation-corrected gated PET data J Nucl Med 59 1480–6 [PubMed: 29439015]

Maas AL, Hannun AY and Ng AY 2013 Rectifier nonlinearities improve neural network acoustic models Proceedings of Machine Learning Research 30 3

Nehmeh S, Erdi Y, Ling C, Rosenzweig K, Squire O, Braban L, Ford E, Sidhu K, Mageras G and Larson S 2002a Effect of respiratory gating on reducing lung motion artifacts in PET imaging of lung cancer Med Phys 29 366–71 [PubMed: 11929020]

Nehmeh SA, Erdi YE, Ling CC, Rosenzweig KE, Schoder H, Larson SM, Macapinlac HA, Squire OD and Humm JL 2002b Effect of respiratory gating on quantifying PET images of lung cancer J Nucl Med 43 876–81 [PubMed: 12097456]

Segars W, Sturgeon G, Mendonca S, Grimes J and Tsui BM 2010 4D XCAT phantom for multimodality imaging research Med Phys 37 4902–15 [PubMed: 20964209]

Sokooti H, de Vos B, Berendsen F, Lelieveldt BP, Išgum I and Staring M 2017 3D Convolutional Neural Networks Image Registration Based on Efficient Supervised Learning from Artificial Deformations International Conference on Medical Image Computing and Computer-Assisted Intervention 232–9

Srivastava N, Hinton G, Krizhevsky A, Sutskever I and Salakhutdinov R 2014 Dropout: a simple way to prevent neural networks from overfitting J Mach Learn Res 15 1929–58

Wang G and Qi J 2015 A monotonie image-space algorithm for joint PET image reconstruction and motion estimation IEEE Nuclear Science Symposium and Medical Imaging Conference (NSS/MIC) 1–3

Zhang X, Zhou J, Wang G, Poon J, Cherry S, Badawi R and Qi J 2014 Feasibility study of micro-dose total-body dynamic PET imaging using the EXPLORER scanner J Nucl Med 55 269

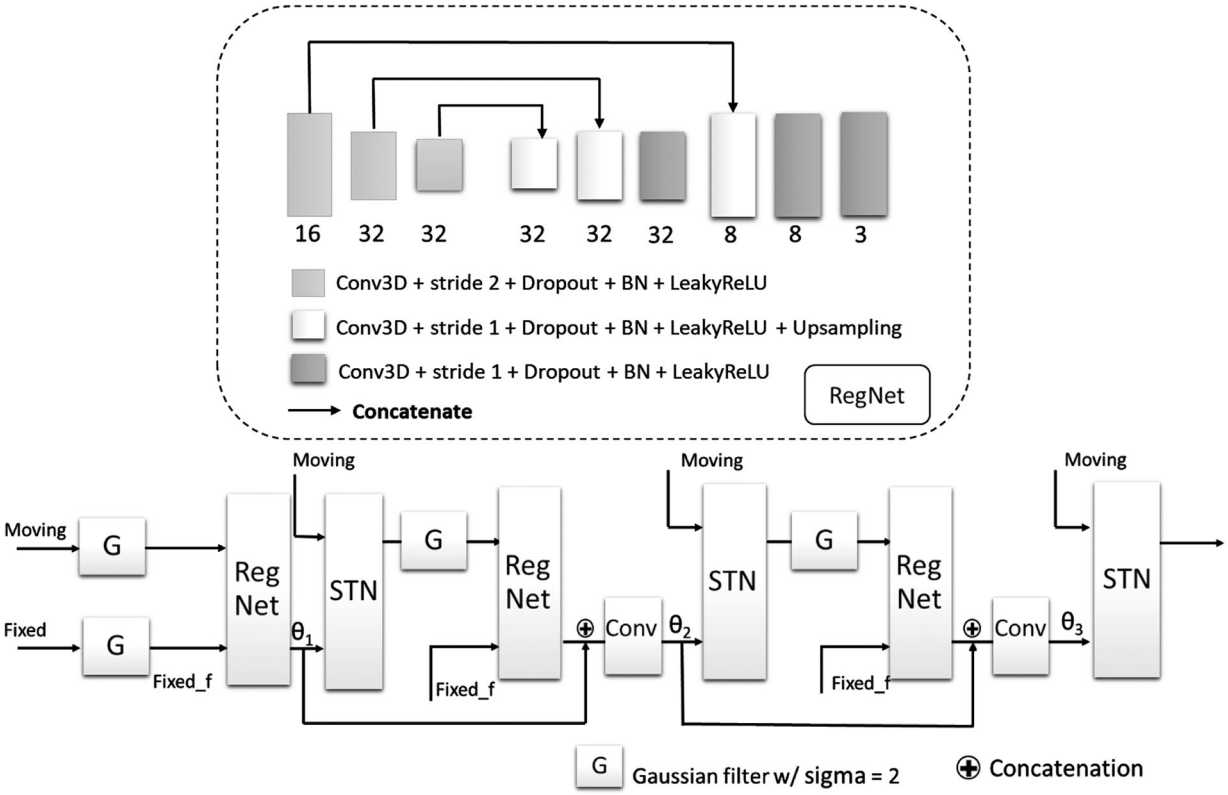

Fig.1.
The structure of the proposed network.

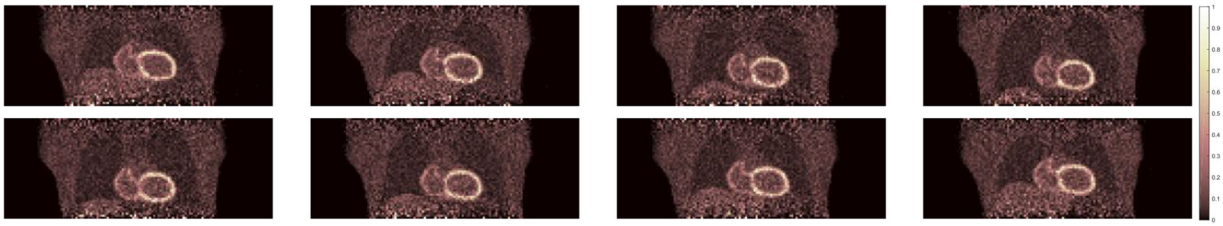

Fig. 2.
Sampled reconstructed MLEM images (50 iteration) for each gate.

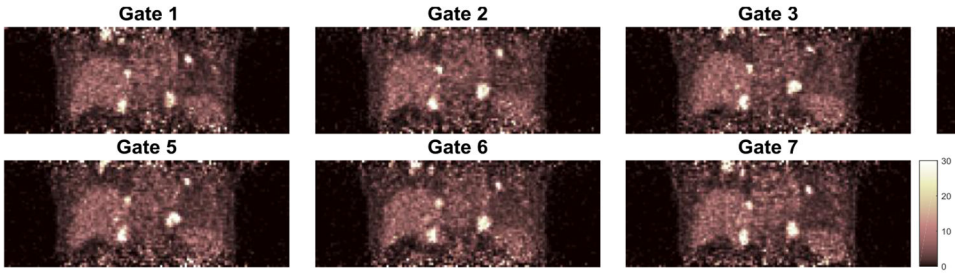

Fig. 3.
Reconstructed gated PET images of the patient data (MLEM 30 iteration).

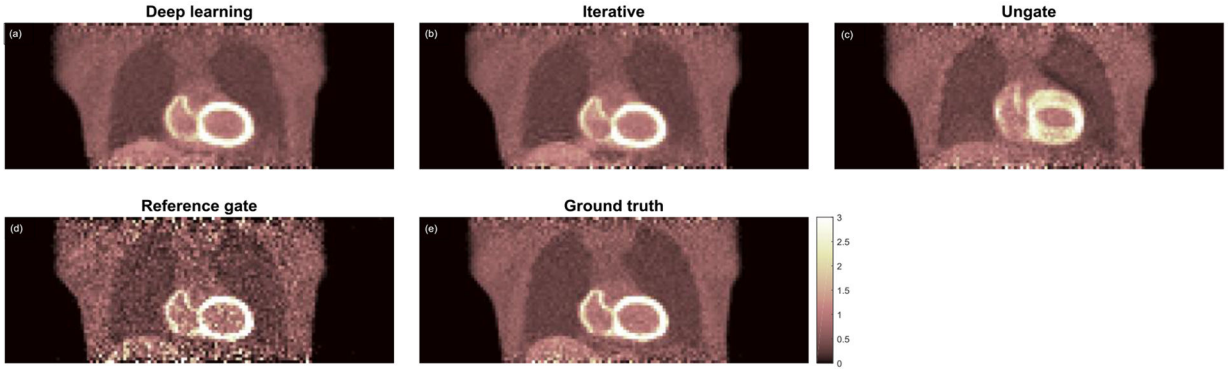

Fig.4.

Reconstructed images of the simulated data set (MLEM 30 iteration). (a) Motion compensated reconstruction using deep neural network for motion estimation, (b) motion compensated reconstruction using iterative registration for motion estimation, (c) ungated reconstruction, (d) reference gate, and (e) ground truth (no motion).

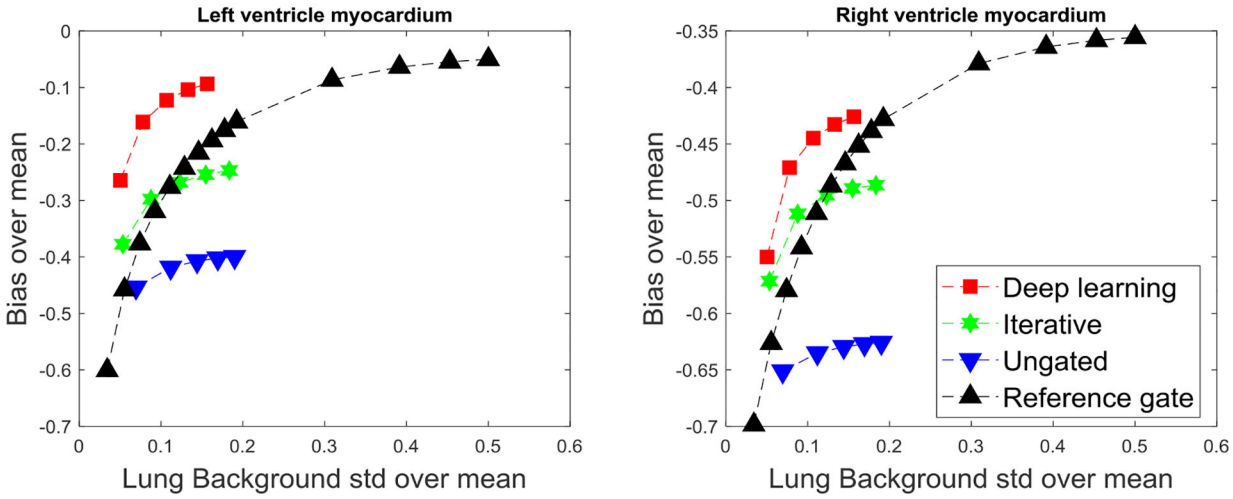

Fig.5.

Bias-variance trade off curves for (a) the left myocardium ROI and (b) right myocardium ROI.

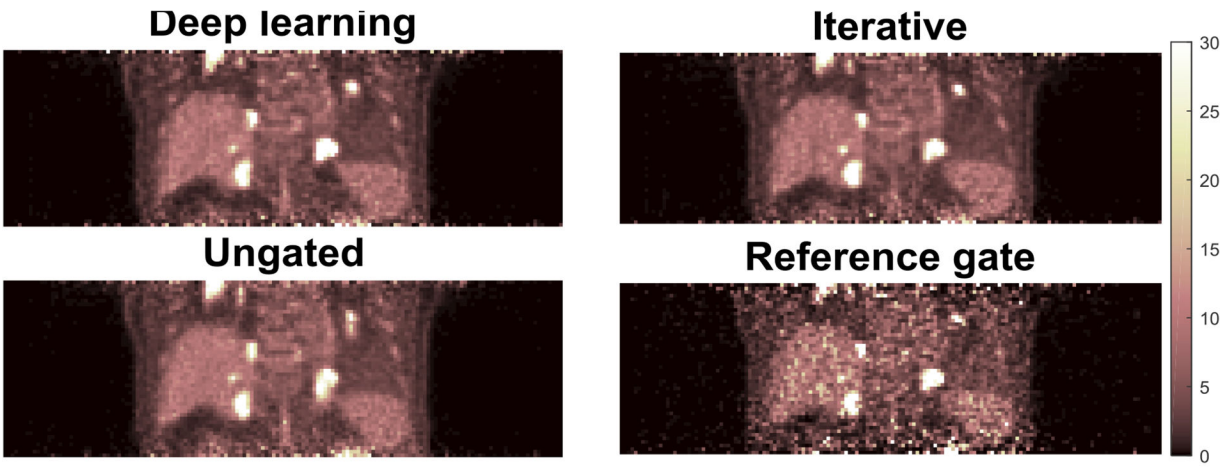

Fig. 6.

Reconstructed images of the real patient data (MLEM 30 iteration). (a) Motion compensated reconstruction using deep learning for motion estimation, (b) motion compensated reconstruction using iterative registration for motion estimation, (c) ungated image, and (d) the reference gate image.

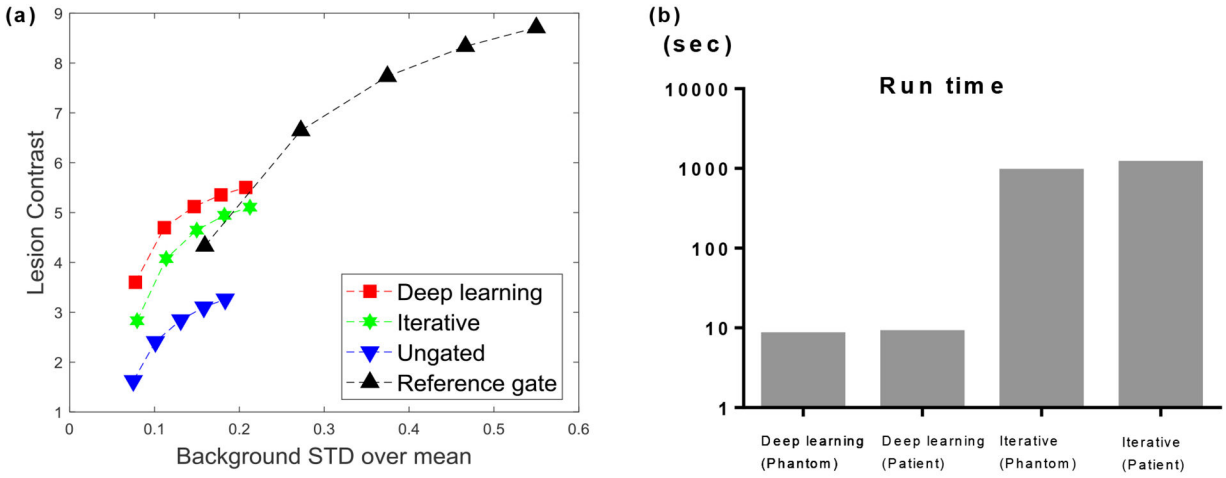

Fig.7.

(a) Contrast versus noise curve for the lung lesion. (b) Registration runtimes of the neural network (deep learning) and iterative registration method for a pair of images. Run time was plotted on a logarithmic scale.

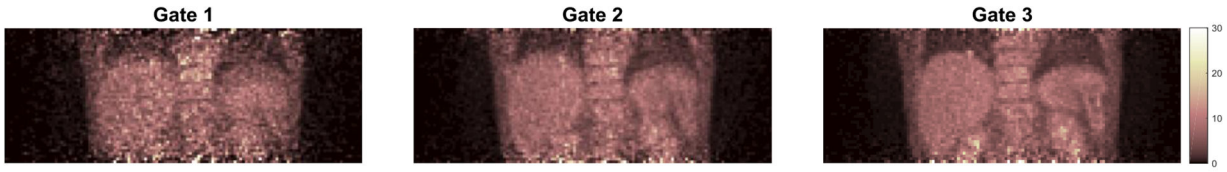

Fig. 8.
Reconstructed gated PET images of the additional real patient data (MLEM 30 iteration).

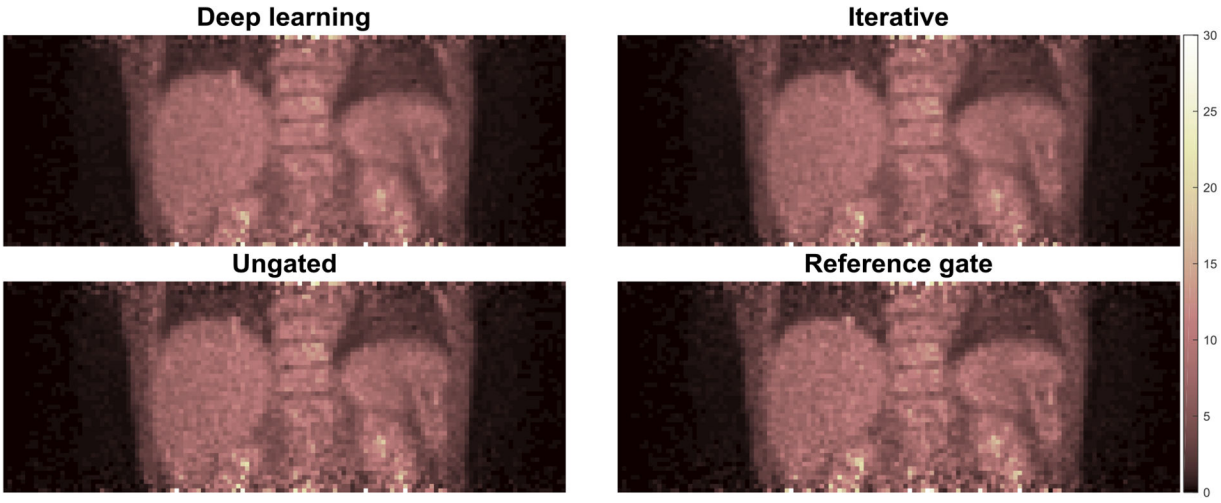

Fig. 9.

Reconstructed images of the additional real patient data (MLEM 30 iteration). (a) Motion compensated reconstruction using deep learning for motion estimation, (b) motion compensated reconstruction using iterative registration for motion estimation, (c) ungated image, and (d) the reference gate image.

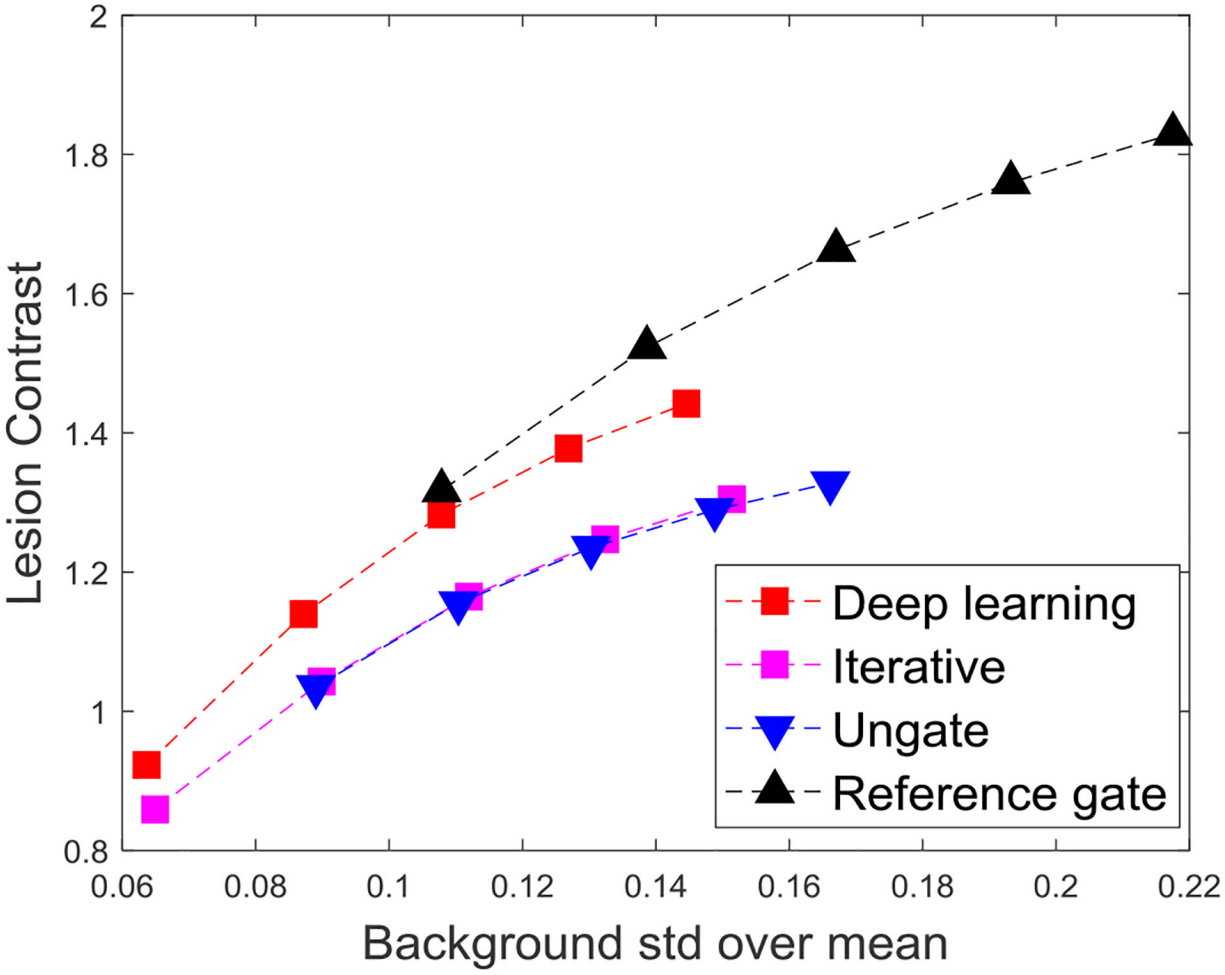

Fig.10.
Contrast versus noise curve for the lesion at liver boundary.

Table 1.

NRMS of reconstructed images for different methods at the 30th MLEM iteration.

Test Phantom #	Deep learning	Iterative	Ungated	Reference gate
#1	23.9%	34.3%	44.0%	35.1%
#2	25.2%	28.1%	40.2%	44.2%
#3	23.0%	30.7%	40.9%	40.9%
#4	22.2%	31.2%	43.7%	37.4%
#5	24.3%	30.8%	45.3%	43.5%
#6	22.7%	30.5%	39.2%	41.7%
#7	22.8%	31.3%	42.1%	41.7%
#8	25.4%	31.9%	41.0%	48.8%
#9	28.0%	31.4%	43.6%	49.1%
#10	26.3%	30.9%	38.9%	44.6%
#11	23.0%	30.7%	41.9%	40.9%
Mean	24.3%	31.1%	41.9%	42.6%
STD	1.7%	1.4%	2.0%	4.0%