

UC Davis

UC Davis Previously Published Works

Title

Deep Learning Based Joint PET Image Reconstruction and Motion Estimation

Permalink

<https://escholarship.org/uc/item/25p6191k>

Journal

IEEE Transactions on Medical Imaging, 41(5)

ISSN

0278-0062

Authors

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

Publication Date

2022-05-01

DOI

10.1109/tmi.2021.3136553

Peer reviewed

Deep Learning Based Joint PET Image Reconstruction and Motion Estimation

Tiantian Li, Mengxi Zhang, Wenyuan Qi, Evren Asma and Jinyi Qi, *Fellow, IEEE*

Abstract—Respiratory motion is one of the main sources of motion artifacts in positron emission tomography (PET) imaging. The emission image and patient motion can be estimated simultaneously from respiratory gated data through a joint estimation framework. However, conventional motion estimation methods based on registration of a pair of images are sensitive to noise. The goal of this study is to develop a robust joint estimation method that incorporates a deep learning (DL)-based image registration approach for motion estimation. We propose a joint estimation framework by incorporating a learned image registration network into a regularized PET image reconstruction. The joint estimation was formulated as a constrained optimization problem with moving gated images related to a fixed image via the deep neural network. The constrained optimization problem is solved by the alternating direction method of multipliers (ADMM) algorithm. The effectiveness of the algorithm was demonstrated using simulated and real data. We compared the proposed DL-ADMM joint estimation algorithm with a monotonic iterative joint estimation. Motion compensated reconstructions using pre-calculated deformation fields by DL-based (DL-MC recon) and iterative (iterative-MC recon) image registration were also included for comparison. Our simulation study shows that the proposed DL-ADMM joint estimation method reduces bias compared to the ungated image without increasing noise and outperforms the competing methods. In the real data study, our proposed method also generated higher lesion contrast and sharper liver boundaries compared to the ungated image and had lower noise than the reference gated image.

Index Terms—Deep learning, joint estimation, image reconstruction, motion correction, PET

I. INTRODUCTION

Positron emission tomography (PET) imaging is a noninvasive imaging modality that provides in vivo visualization of biochemical processes in a living body through the use of radiotracers. It is widely used in many clinical imaging applications, including oncology, cardiology and neurology. With the latest state-of-the-art whole-body PET scanners using detector crystals of about 3 mm in size [1], [2], the intrinsic spatial resolutions of PET scanners have been substantially improved. As a result, physiological motions, e.g., heart beating and respiratory motion, have become a limiting factor for PET spatial resolution in clinical practice. For example, Lu *et al.* has reported that respiratory motion caused an averaged displacement of 6.1 mm among 10 patient scans for a total of 323 regions of interest (ROIs) [3]. Therefore, correction for respiratory motion is essential for obtaining high-resolution PET images. One motion correction method is the event-by-event correction based on list-mode data processing. Motion is measured using a special motion detection hardware, such as a POLARIS [4] or Anzai AZ-733V system (Anzai Medical Co, Ltd., Tokyo, Japan), or using a data-driven approach [5], [6]. The event-driven motion compensation is performed by transforming the line-of-response (LOR) along which the event is measured to the position it would have been measured if the object had not moved [7]. However, the event-driven approach has two limitations. First, an event that is normally detected can exit the field of view (FOV) because of motion, which results in a loss of the event. Second, an event that is normally undetected may be detected because of the LOR transformation. Therefore, after correction for motion, some detected events may correspond to no actual detector pairs. These problems may lead to image artifacts and a decrease of signal-to-noise ratio (SNR) [8].

Another motion correction method is the image-driven approach. In this method, respiratory gating is used to divide PET data into several gates based on a respiratory motion signal that is either obtained externally or estimated from the PET data. Motion inside each gate is considered negligible. Because gated PET images can be very noisy due to low count statistics, one way to enhance the performance of gated PET is to estimate the motion information from the 4D gated PET images and then to perform motion compensated image reconstruction to utilize all gated data [3]. However, conventional motion estimation

This paper was submitted for review on December 31, 2020. This work is supported by the National Institutes of Health under Grant R01EB000194.

T. Li, M. Zhang and J. Qi are with Department of Biomedical Engineering, University of California, Davis, CA, USA (telephone: 1 530 7546142, e-mail: qi@ucdavis.edu).

W. Qi, E. Asma are with Canon Medical Research USA, Inc., Vernon Hills, IL, USA (e-mail: easma@mru.medical.canon).

methods based on registration of a pair of images are sensitive to noise and require high SNR for an accurate estimation. A better alternative is to estimate the PET activity image and patient motion simultaneously from the gated data through a joint estimation framework. Such a framework uses all measured PET data for image reconstruction and enables more accurate noise modeling in the motion estimation [9], [10]. In joint estimation methods, the objective function includes a set of parameters describing the image activity distribution and a set of deformation parameters describing the motion between a fixed frame and one or more moving frames. These two sets of parameters are estimated jointly. A challenge of the joint estimation method is that the resulting objective function is nonconvex and the optimization problem is difficult to solve. Some algorithms require a line search to guarantee convergence, which can be computationally intensive because of the involvement of a forward projection operation. Wang *et al.* proposed an image-space algorithm based on the idea of optimization transfer (OT) [11]. The algorithm uses the expectation-maximization (EM) surrogate function and does not require the forward projection operation in the estimation of motion parameters. However, the cost function is still nonconvex in the intensity-based non-rigid image registration problem, so the OT algorithm can be trapped in a local optimum and the solution is highly influenced by noise in the images. Burger *et al.* proposed a variational model for joint motion estimation and image reconstruction [12], and Chen *et al.* further extended this method to simulated tomographic images [13].

Recently, deep learning (DL) techniques have provided new approaches to image registration [14]–[19]. There are three widely used components of DL-based image registration [20], an encoder–decoder based architecture incorporating several hierarchical convolution layers for multi-scale feature extraction [16], [19], [21], a spatial transformer network (STN) [22] for spatial transformation, and a generative adversarial network (GAN) [23]–[25] where a generator predicts the deformation field and warps the moving image and the warped moving image is evaluated by a discriminator. One advantage of DL methods is that a pretrained network can incorporate prior knowledge of motion fields from training data into the motion estimation process. Previously, we proposed an unsupervised non-rigid image registration framework using DL to estimate deformation fields for respiratory gated images [21] and demonstrated that it could outperform an iterative image registration method.

In this paper, we propose a joint estimation method that incorporates the DL-based image registration network. The activity image and deformation field are formulated as the solution of a constrained optimization problem. The constrained optimization problem is solved by the alternating direction method of multipliers (ADMM) algorithm [26]. Each iteration of our proposed algorithm, which is referred to as DL-ADMM, consists of three steps: gated image reconstruction by the maximum *a posteriori* expectation maximization (MAP-

EM) algorithm, motion estimation by a regularized DL-based image registration, and regularized image fusion with motion compensation. The overall diagram of our proposed method is shown in Fig.1. The effectiveness of the algorithm is demonstrated using simulated and real data.

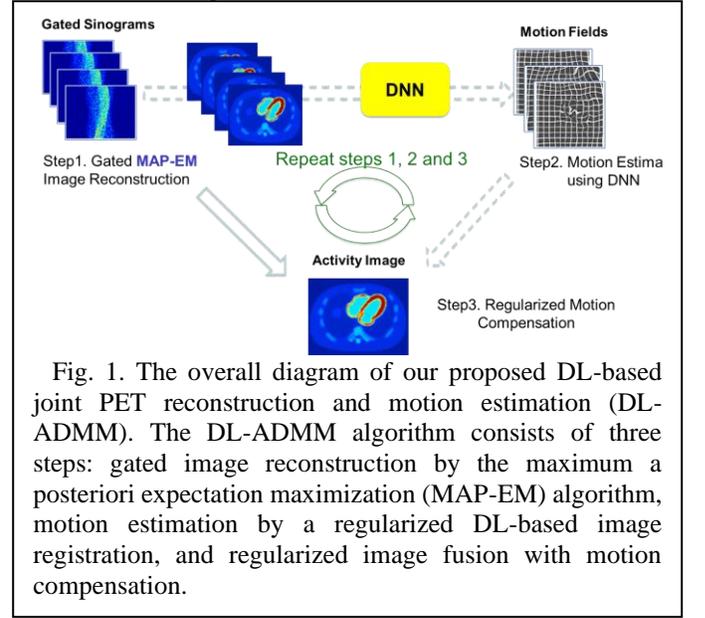


Fig. 1. The overall diagram of our proposed DL-based joint PET reconstruction and motion estimation (DL-ADMM). The DL-ADMM algorithm consists of three steps: gated image reconstruction by the maximum a posteriori expectation maximization (MAP-EM) algorithm, motion estimation by a regularized DL-based image registration, and regularized image fusion with motion compensation.

II. MATERIALS AND METHODS

A. Data model

Let us denote the PET activity image of a reference fixed gate by $\mathbf{x}_f \in \mathbb{R}^{N \times 1}$ and that of the m th moving gate by $\mathbf{x}_m \in \mathbb{R}^{N \times 1}$, where $m \in \{1, 2, \dots, K\}$ and K is the total number of moving gates. In statistically based PET image reconstruction, the measured data in each gate, $\mathbf{y}_m \in \mathbb{R}^{M \times 1}$, can be modeled as independent Poisson random variables and their mean $\bar{\mathbf{y}}_m \in \mathbb{R}^{M \times 1}$ are related to the corresponding activity image through

$$\bar{\mathbf{y}}_m = w_m \cdot \mathbf{N} \cdot \mathbf{A}_m \cdot \mathbf{P} \cdot \mathbf{x}_m + \mathbf{s}_m + \mathbf{r}_m, \quad (1)$$

where the (i, j) th element of $\mathbf{P} \in \mathbb{R}^{M \times N}$, $p_{i,j}$, denotes the probability of detecting an emission from pixel j , $j \in \{1, \dots, N\}$, at detector pair i , $i \in \{1, \dots, M\}$, $\mathbf{N} \in \mathbb{R}^{M \times M}$ and $\mathbf{A}_m \in \mathbb{R}^{M \times M}$ are diagonal matrices containing the normalization factors and attenuation factors, respectively, for the m th moving gate, $\mathbf{s}_m \in \mathbb{R}^{M \times 1}$ denotes the expectation of scattered events, $\mathbf{r}_m \in \mathbb{R}^{M \times 1}$ denotes the expectation of random events. The weighting factor w_m accounts for the duration of each gate with $\sum_m w_m + w_f = 1$. The log-likelihood function for the m th gate can be written as

$$L(\mathbf{y}_m | \mathbf{x}_m) = \sum_{i=1}^M y_m^i \log \bar{y}_m^i - \bar{y}_m^i - \log y_m^i!, \quad (2)$$

The overall log-likelihood function of all gated PET data is

$$L(\mathbf{y} | \mathbf{x}) = \sum_{m=1}^K L(\mathbf{y}_m | \mathbf{x}_m) + L(\mathbf{y}_f | \mathbf{x}_f), \quad (3)$$

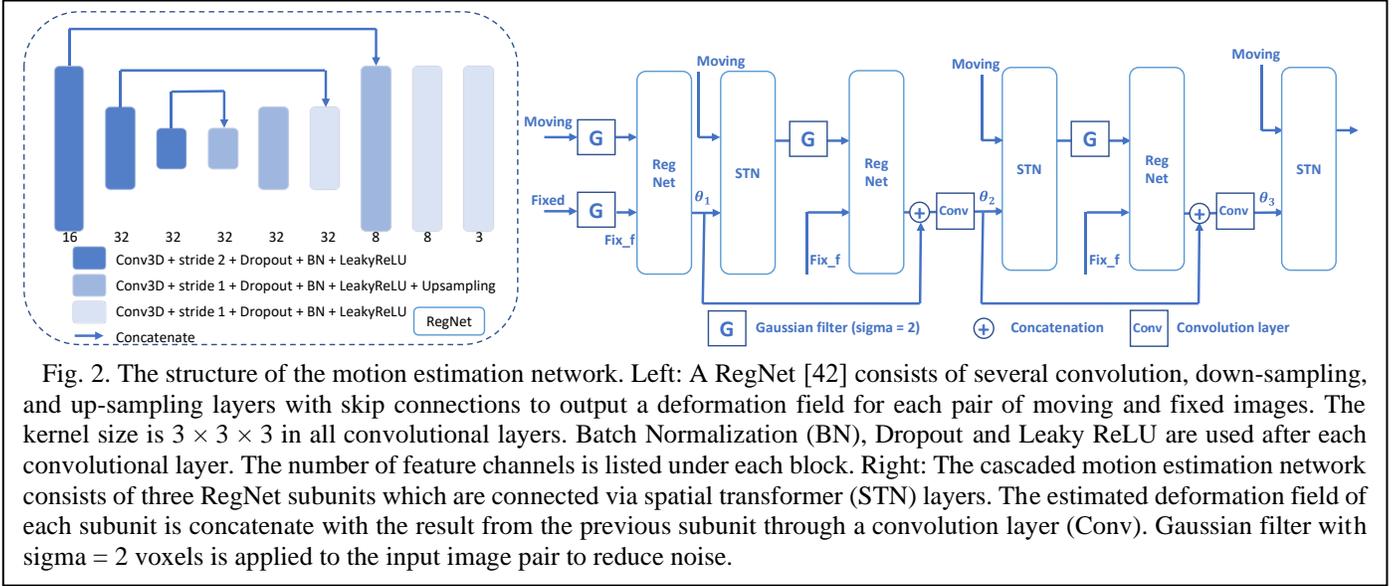


Fig. 2. The structure of the motion estimation network. Left: A RegNet [42] consists of several convolution, down-sampling, and up-sampling layers with skip connections to output a deformation field for each pair of moving and fixed images. The kernel size is $3 \times 3 \times 3$ in all convolutional layers. Batch Normalization (BN), Dropout and Leaky ReLU are used after each convolutional layer. The number of feature channels is listed under each block. Right: The cascaded motion estimation network consists of three RegNet subunits which are connected via spatial transformer (STN) layers. The estimated deformation field of each subunit is concatenate with the result from the previous subunit through a convolution layer (Conv). Gaussian filter with $\sigma = 2$ voxels is applied to the input image pair to reduce noise.

B. Motion deformation field estimation using a neural network

In this study, we apply our previously developed unsupervised deep learning model for deformable motion estimation [21]. The overall network architecture is shown in Fig. 2.

The network was trained using pairs of moving and fixed images, \mathbf{x}_m and \mathbf{x}_f , in an unsupervised manner by minimizing a penalized loss function:

$$\boldsymbol{\theta} = \arg \min_{\boldsymbol{\theta}} -CC(\mathbf{T}(\mathbf{x}_m, \boldsymbol{\theta}), \mathbf{x}_f) + \lambda \sum \|\nabla \boldsymbol{\theta}\|^2, \quad (4)$$

where CC denotes the cross correlation between the fixed and warped moving images, and λ is a regularization parameter. An L-2 norm regularizer on the gradients of the deformation field ($\boldsymbol{\theta}$) is applied to encourage a smooth $\boldsymbol{\theta}$.

After training, the network generates a deformation field ($\boldsymbol{\theta}_m$) for any pair of moving (\mathbf{x}_m) and fixed (\mathbf{x}_f) images. Here we denote this estimation by a non-linear function

$$\boldsymbol{\theta}_m = \mathbf{f}(\mathbf{x}_m, \mathbf{x}_f), \quad (5)$$

where $\mathbf{f}: \mathbb{R} \rightarrow \mathbb{R}$ represents the DL based motion estimation operator.

C. Joint estimation of activity image and motion

The joint estimation is formulated as a constrained optimization problem with moving frames related to a fixed frame via the deep neural network:

$$\mathbf{x} = \arg \max_{\{\mathbf{x}_m\}_{m=1}^K, \mathbf{x}_f} \sum_{m=1}^K L(\mathbf{y}_m | \mathbf{x}_m) + L(\mathbf{y}_f | \mathbf{x}_f), \quad (6)$$

$$s. t. \mathbf{x}_f = \mathbf{T}(\mathbf{x}_m, \boldsymbol{\theta}_m) \text{ and } \boldsymbol{\theta}_m = \mathbf{f}(\mathbf{x}_m, \mathbf{x}_f)$$

where $\mathbf{T}(\cdot, \cdot)$ denotes the deformation operator. The constrained optimization problem in (6) can be written in an augmented

Lagrangian format as

$$\begin{aligned} \mathbf{x} = \arg \max_{\{\mathbf{x}_m\}_{m=1}^K, \mathbf{x}_f} & \sum_{m=1}^K L(\mathbf{y}_m | \mathbf{x}_m) + L(\mathbf{y}_f | \mathbf{x}_f) \\ & - \sum_{m=1}^K \frac{\rho_m}{2} \|\boldsymbol{\beta} - \mathbf{T}(\boldsymbol{\alpha}_m, \boldsymbol{\theta}_m) \\ & + \mathbf{E}_m\|^2 \\ & - \sum_{m=1}^K \frac{\tau_m}{2} \|\boldsymbol{\theta}_m - \mathbf{f}(\boldsymbol{\alpha}_m, \boldsymbol{\beta}) \\ & + \mathbf{B}_m\|^2 \\ & - \sum_{m=1}^K \frac{\eta_m}{2} \|\mathbf{x}_m - \boldsymbol{\alpha}_m + \mathbf{C}_m\|^2 \\ & - \frac{\delta}{2} \|\mathbf{x}_f - \boldsymbol{\beta} + \mathbf{D}\|^2, \end{aligned} \quad (7)$$

which can be solved by the ADMM algorithm. This constrained optimization is decoupled into the following subproblems:

$$\mathbf{x}_m = \arg \max_{\mathbf{x}_m} L(\mathbf{y}_m | \mathbf{x}_m) - \frac{\eta_m}{2} \|\mathbf{x}_m - \boldsymbol{\alpha}_m + \mathbf{C}_m\|^2, \quad (8)$$

$$\mathbf{x}_f = \arg \max_{\mathbf{x}_f} L(\mathbf{y}_f | \mathbf{x}_f) - \frac{\delta}{2} \|\mathbf{x}_f - \boldsymbol{\beta} + \mathbf{D}\|^2, \quad (9)$$

$$\begin{aligned} \boldsymbol{\theta}_m = \arg \min_{\boldsymbol{\theta}_m} & \frac{\rho_m}{2} \|\boldsymbol{\beta} - \mathbf{T}(\boldsymbol{\alpha}_m, \boldsymbol{\theta}_m) + \mathbf{E}_m\|^2 \\ & + \frac{\tau_m}{2} \|\boldsymbol{\theta}_m - \mathbf{f}(\boldsymbol{\alpha}_m, \boldsymbol{\beta}) \\ & + \mathbf{B}_m\|^2, \end{aligned} \quad (10)$$

$$\begin{aligned} \boldsymbol{\alpha}_m = \arg \min_{\boldsymbol{\alpha}_m} & \frac{\rho_m}{2} \|\boldsymbol{\beta} - \mathbf{T}(\boldsymbol{\alpha}_m, \boldsymbol{\theta}_m) + \mathbf{E}_m\|^2 \\ & + \frac{\tau_m}{2} \|\boldsymbol{\theta}_m - \mathbf{f}(\boldsymbol{\alpha}_m, \boldsymbol{\beta}) + \mathbf{B}_m\|^2 \\ & + \frac{\eta_m}{2} \|\mathbf{x}_m - \boldsymbol{\alpha}_m + \mathbf{C}_m\|^2, \end{aligned} \quad (11)$$

$$\begin{aligned} \boldsymbol{\beta} = \arg \min_{\boldsymbol{\beta}} & \sum_{m=1}^K \frac{\rho_m}{2} \|\boldsymbol{\beta} - \mathbf{T}(\boldsymbol{\alpha}_m, \boldsymbol{\theta}_m) + \mathbf{E}_m\|^2 \\ & + \sum_{m=1}^K \frac{\tau_m}{2} \|\boldsymbol{\theta}_m - \mathbf{f}(\boldsymbol{\alpha}_m, \boldsymbol{\beta})\|^2 \\ & + \mathbf{B}_m\|^2 + \frac{\delta}{2} \|\mathbf{x}_f - \boldsymbol{\beta} + \mathbf{D}\|^2. \end{aligned} \quad (12)$$

Subproblems (8, 9) correspond to MAP-EM PET image reconstruction using a quadratic regularization. The MAP image estimates can be found by the OT method [11], [27]. Using the EM surrogate function, the surrogate objective functions for subproblems (8, 9) are

$$P(\mathbf{x}_m | \mathbf{x}_m^n) = \sum_{j=1}^N p_{m,j} (\hat{x}_{m,j,EM}^{n+1} \log x_{m,j} - x_{m,j}) - \frac{\eta_m}{2} (x_{m,j} - \alpha_{m,j} + C_{m,j})^2, \quad (13)$$

$$P(\mathbf{x}_f | \mathbf{x}_f^n) = \sum_{j=1}^N p_{f,j} (\hat{x}_{f,j,EM}^{n+1} \log x_{f,j} - x_{f,j}) - \frac{\delta}{2} (x_{f,j} - \beta_j + D_j)^2, \quad (14)$$

where $p_{m,j} = \sum_{i=1}^M [W_m \cdot \mathbf{N} \cdot \mathbf{A}_m \cdot \mathbf{P}]_{ij}$ and $\hat{x}_{m,j,EM}^{n+1}$ is given by

$$\hat{x}_{m,j,EM}^{n+1} = \frac{x_{m,j}}{p_{m,j}} \sum_{i=1}^M p_{i,j} \frac{y_{m,i}}{[\mathbf{P} \cdot \mathbf{x}_m + \frac{\mathbf{s}_m + \mathbf{r}_m}{W_m \mathbf{N} \cdot \mathbf{A}_m}]_i}. \quad (15)$$

The update equations for the j th pixel of \mathbf{x}_m and \mathbf{x}_f by maximizing (13) and (14), respectively, are

$$x_{m,j}^{n+1} = \frac{1}{2} \left[\left(\alpha_{m,j}^n - C_{m,j}^n - \frac{p_{m,j}}{\eta_m} \right) + \sqrt{\left(\alpha_{m,j}^n - C_{m,j}^n - \frac{p_{m,j}}{\eta_m} \right)^2 + 4 \frac{p_{m,j}}{\eta_m} \hat{x}_{m,j,EM}^{n+1}} \right], \quad (16)$$

$$x_{f,j}^{n+1} = \frac{1}{2} \left[\left(\beta_j^n - D_j^n - \frac{p_{f,j}}{\delta} \right) + \sqrt{\left(\beta_j^n - D_j^n - \frac{p_{f,j}}{\delta} \right)^2 + 4 \frac{p_{f,j}}{\delta} \hat{x}_{f,j,EM}^{n+1}} \right]. \quad (17)$$

The minimization of the subproblems (10-12) are regularized weighted least squares (RWLS) problems with a quadratic regularization. A gradient descent approach can be used to solve these subproblems. We get the following update equations by taking the derivative of the objective functions of each subproblem with respect to the unknown variable:

$$\begin{aligned} \theta_{m,j}^{n+1} = \theta_{m,j}^n - lr_{\theta} & \left\{ -\rho_m \sum_{u=1}^N [\boldsymbol{\beta}^n - \mathbf{T}(\boldsymbol{\alpha}_m^n, \boldsymbol{\theta}_m^n)] \right. \\ & + \mathbf{E}_m^n \mathbf{1}_u \left[\frac{\partial \mathbf{T}(\boldsymbol{\alpha}_m^n, \boldsymbol{\theta}_m^n)}{\partial \theta_{m,j}} \right]_u \\ & \left. + \tau_m (\boldsymbol{\theta}_m^n - \mathbf{f}(\boldsymbol{\alpha}_m^n, \boldsymbol{\beta}^n) + \mathbf{B}_m^n)_j \right\}, \end{aligned} \quad (18)$$

$$\begin{aligned} \alpha_{m,j}^{n+1} = \alpha_{m,j}^n - lr_{\alpha} & \left\{ -\rho_m [\mathbf{T}^*(\boldsymbol{\beta}^n - \mathbf{T}(\boldsymbol{\alpha}_m^n, \boldsymbol{\theta}_m^{n+1}) \right. \\ & + \mathbf{E}_m^n, \boldsymbol{\theta}_m^{n+1})]_j \\ & - \tau_m \left[\sum_{u=1}^N (\boldsymbol{\theta}_m^{n+1} - \mathbf{f}(\boldsymbol{\alpha}_m^n, \boldsymbol{\beta}^n) \right. \\ & \left. + \mathbf{B}_m^n)_u \frac{\partial \mathbf{f}(\boldsymbol{\alpha}_m^n, \boldsymbol{\beta}^n)_u}{\partial \alpha_{m,j}} \right] \\ & \left. - \frac{\eta_m}{2} (\mathbf{x}_m^{n+1} - \alpha_m^n + \mathbf{C}_m^n)_j \right\}, \end{aligned} \quad (19)$$

$$\begin{aligned} \beta_j^{n+1} = \beta_j^n - lr_{\beta} & \left\{ \sum_{m=1}^K \rho_m (\boldsymbol{\beta}^n - \mathbf{T}(\boldsymbol{\alpha}_m^{n+1}, \boldsymbol{\theta}_m^{n+1}) + \mathbf{E}_m^n)_j \right. \\ & - \sum_{m=1}^K \tau_m \left[\sum_{u=1}^N (\boldsymbol{\theta}_m^{n+1} - \mathbf{f}(\boldsymbol{\alpha}_m^{n+1}, \boldsymbol{\beta}^n) \right. \\ & \left. + \mathbf{B}_m^n)_u \frac{\partial \mathbf{f}(\boldsymbol{\alpha}_m^{n+1}, \boldsymbol{\beta}^n)_u}{\partial \beta_j} \right] \\ & \left. - \delta (\mathbf{x}_f^{n+1} - \boldsymbol{\beta}^n + \mathbf{D}^n)_j \right\}, \end{aligned} \quad (20)$$

where lr_{θ} , lr_{α} , lr_{β} are the step sizes of each gradient update, $\mathbf{T}(\cdot, \cdot)$ denotes the forward transformation from the fixed gate to the moving gates, and $\mathbf{T}^*(\cdot, \cdot)$ is the transpose of the transformation for back warping the fixed gate to the moving gates.

The RWLS problems in (10-12) are non-convex and can have multiple local minima. Therefore, the solutions for the updates in (18-20) highly depend on the implementation of the optimization processes and initialization. In order to assign a good initialization for $\boldsymbol{\theta}_m$, we use the gated Maximum Likelihood Expectation Maximization (ML-EM) reconstructed images to estimate the initial values of $\boldsymbol{\theta}_m$. For $\boldsymbol{\alpha}_m$ and $\boldsymbol{\beta}$ initialization, we first obtained a motion compensated reconstruction \mathbf{x}_{ini} [21], and used \mathbf{x}_{ini} to initialize $\boldsymbol{\beta}$ and warped \mathbf{x}_{ini} based on $\boldsymbol{\theta}_m$ to initialize $\boldsymbol{\alpha}_m$.

The final step of the ADMM algorithm is the update of the Lagrange multipliers as the following:

$$\mathbf{E}_m^{n+1} = \mathbf{E}_m^n + \boldsymbol{\beta}^{n+1} - \mathbf{T}(\boldsymbol{\alpha}_m^{n+1}, \boldsymbol{\theta}_m^{n+1}), \quad (21)$$

$$\mathbf{B}_m^{n+1} = \mathbf{B}_m^n + \boldsymbol{\theta}_m^{n+1} - \mathbf{f}(\boldsymbol{\alpha}_m^{n+1}, \boldsymbol{\beta}^{n+1}), \quad (22)$$

$$\mathbf{C}_m^{n+1} = \mathbf{C}_m^n + \mathbf{x}_m^{n+1} - \boldsymbol{\alpha}_m^{n+1}, \quad (23)$$

$$\mathbf{D}^{n+1} = \mathbf{D}^n + \mathbf{x}_f^{n+1} - \boldsymbol{\beta}^{n+1}. \quad (24)$$

Algorithm 1. Algorithm for DL-based joint PET reconstruction and motion estimation

Input: Maximum iteration number: MaxIt; number of gates: K; penalty parameters: ρ_m, η_m, τ_m , and δ ; image initialization: $\mathbf{x}_{ini}, \mathbf{x}_m, \mathbf{x}_f$.

1. Initialize the auxiliary images and Lagrange multipliers: $\alpha_m^1 = \mathbf{T}^*(\mathbf{x}_{ini}, \boldsymbol{\theta}_m), \boldsymbol{\beta}^1 = \mathbf{x}_{ini}, \mathbf{E}_m^1 = \mathbf{B}_m^1 = \mathbf{C}_m^1 = \mathbf{D}^1 = \mathbf{0}$
 2. **for** $n = 1$ to MaxIt **do**
 3. Update image \mathbf{x}_f^{n+1} by MAP-EM using (17)
 4. **for** $m = 1$ to K **do**
 5. Update image \mathbf{x}_m^{n+1} by MAP-EM using (16)
 6. Update motion field $\boldsymbol{\theta}_m^{n+1}$ using (18)
 7. Update auxiliary image α_m^{n+1} using (19)
 8. **end for**
 9. Update auxiliary image $\boldsymbol{\beta}^{n+1}$ by (20)
 10. Update the Lagrange multipliers $\mathbf{E}_m^{n+1}, \mathbf{B}_m^{n+1}, \mathbf{C}_m^{n+1}, \mathbf{D}^{n+1}$ using (21-24)
 11. **end for**
- Output:** $\boldsymbol{\beta}^{n+1}$

The overall algorithm is summarized in Algorithm 1.

D. Implementation details and reference methods

The motion estimation network was implemented in Keras 2.2.4 with Tensorflow 1.5.0 and trained on a NVIDIA GTX 1080Ti GPU. The adaptive moment estimation (ADAM) optimizer [28] with the default parameter settings was used. The learning rate was set to 0.0005 and batch size was set to 1. The cost function was calculated between the warped moving image and the fixed image using (4). For the proposed joint estimation framework, the hyperparameters ρ_m, η_m and δ were set to 1 and τ_m was set to 0.4 and they were fixed during each iteration. The first-order gradient of $f(\alpha_m, \boldsymbol{\beta})$ with respect to α_m and $\boldsymbol{\beta}$ in subproblem (19, 20) was implemented using the tf.gradient function in the TensorFlow on GPU. However, this gradient operation is time consuming. To speed up the reconstruction, we further evaluate the performance when $\tau_m = 0$ using simulation datasets. Table 1 summarizes the methods and pre-determined hyper-parameters for the reconstruction included in this study.

We compared the proposed DL-ADMM joint estimation algorithm with the monotonic iterative joint estimation method

using OT with the EM surrogate function[11]. Below we refer to this method as ‘‘Iterative-joint estimation’’ method. For a fair comparison, we initialized the motion field using the result from an iterative image registration [29] using a publicly available Bspline toolbox¹. We used the default weighted-least-squares similarity measure. Gaussian post-filtering with sigma = 2 voxels was applied to suppress gated image noise. The number of iterations was chosen to be 200 based on visual assessment of the deformation field. Furthermore, we also performed motion compensated reconstruction using the motion field pre-determined either using the DL method or the iterative registration software. We refer to these two methods as ‘‘DL-MC recon’’ and ‘‘Iterative-MC recon’’.

III. EXPERIMENTAL SETUP

A. Simulation study

We generated twenty-two voxelized phantoms with various organ sizes (11 Male and 11 Female) using the 4D extended cardiac-torso phantom (XCAT Version 2.0) [30] following the same procedure described in [21]. The scales of the long- and short-axes and height were used to control the body size which scales the entire phantom in different regions of the body (training dataset: long axis scale: 0.99 ± 0.02 ; short axis scale: 1.02 ± 0.12 ; height scale: 1.03 ± 0.11). To better evaluate the generalization ability of our trained model in terms of the body size, we included a wider range variation in the testing dataset (long axis scale: 1.01 ± 0.03 ; short axis scale: 1.18 ± 0.18 ; height scale: 0.97 ± 0.13) than in the training dataset to simulate outliers. Ten phantoms were used for training, one for validation and eleven for testing. The maximum displacement of the diaphragm was set to 4 cm during respiration with a period of 5 sec. For each phantom, we simulated 8 gated activity images with matched attenuation maps. To model the variations between subjects, activity parameters of different tissues included a 5% variation. We conducted our Monte-Carlo simulation using the SimSET package [31] and modeled a Canon Celesteion PET/CT scanner [32] (Canon Medical Corporation, Tochigi, Japan) geometry. A 20 min PET scan was simulated with 5.4 mCi ¹⁸F-FDG injection starting from 1 hour post-injection [33]. To focus on the comparison of motion compensation, only true coincidences were included in reconstruction by assuming perfect scatter and random

Table 1. List of the methods and pre-determined hyper-parameters for the reconstruction methods included in this study.

Methods	Regularizer on motion	Regularizer on image	Hyperparameters
DL-ADMM	DL-based	DL-based	$\rho_m = \eta_m = \delta = 1$ and $\tau_m = 0$ or 0.4 Simulation: $lr_\theta = 0.1, lr_\alpha = lr_\beta = 0.5$ Real data: $lr_\theta = 1e^{-4}, lr_\alpha = lr_\beta = 3e^{-4}$
Iterative joint estimation	N/A	N/A	Simulation: $lr_\theta = 0.1$; Real data: $lr_\theta = 1e^{-4}$
DL-MC recon	Gradient of motion field	N/A	$\lambda = 1$
Iterative-MC recon	Local invertibility	N/A	$\lambda = 4$

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

corrections.

To generate the training data for motion estimation, gated PET data were reconstructed using the ML-EM algorithm for 30 iterations with a voxel size of $4.08 \times 4.08 \times 4.08$ mm³. The reconstructed images were stored in a $128 \times 128 \times 48$ matrix. The end-inspiration phase (Gate 4) was chosen as the fixed gate. To further improve the network performance, fixed-fixed training pairs (two identical images) were also included in the training datasets. A total of 80 (8 gates \times 10 phantoms) 3D training pairs were generated. Moving-fixed image pairs were fed into the network for training and the network training ended after 1500 epochs. The training process took a few days. After the training, deformation fields (θ) between any pair of images can be estimated by feeding a pair of moving and fixed images into the network, which takes only 8 s per image pair.

For the DL-ADMM algorithm, x_m and x_f were initialized using 10 iteration gated ML-EM reconstructions (Fig. 3). We initialized θ_m by the deformation fields estimated by the pretrained network from the gated ML-EM reconstructed images. For α_m and β initialization, we first ran DL-MC recon for 10 iterations and used it as the initial image for β , in addition we warp β using θ_m as the initial image for α_m .

For quantitative assessment, we reconstructed the reference gate with 8x counts, which corresponds to the result with perfect motion compensation, and used it as the ground truth. The normalized root mean square error (NRMS) were calculated by (25) as a measure of the differences between reconstructions with and without motion compensation and the ground truth. We excluded the top and bottom three noisy slices which have low count statistics due to the low sensitivity at the axial edge of the scanner field of view.

$$\text{NRMS} = \frac{1}{\|\bar{x}\|_2} \sqrt{\sum_{i=1}^N |x_i - \bar{x}_i|^2}, \quad (25)$$

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

For ROI quantification, we calculated the bias versus background standard deviation (STD) trade off curves. The bias was computed in the left and right myocardium regions and the STD was calculated in the lung region.

B. Real data study

A real respiratory gated dataset was obtained from a Canon Celesteion TOF PET/CT scanner. A patient was injected with 6.2 mCi ¹⁸F-FDG. List mode data were acquired for two bed positions sequentially, each with 14 min starting from 90 minutes post-injection. Equal phase-based gating scheme based on an externally measured respiratory signal using an Anzai AZ-733V system (Anzai Medical Co, Ltd., Tokyo, Japan) was applied to divide the list mode data into 7 respiratory gates. Events in irregular breathing cycles were excluded. We found that the patient did not breathe in the same way for the two bed positions because there were more events rejected in the 2nd bed position than in the 1st bed position (bed 1: 9.9%, bed 2: 22.0%). The resulting gated images from two bed positions also had different motion amplitude ranges which affected the lesion shape and contrast. Thus, the data from the two beds were reconstructed and analyzed separately.

Gated PET data were first reconstructed using the ML-EM method for 10 iterations for DL-ADMM initialization. The voxel size was $4.08 \times 4.08 \times 4.08$ mm³ and the image array size is $152 \times 152 \times 48$. The projection matrix used a simplified geometric projection matrix with an image blurring matrix estimated by point source measurements [34]. A uniform cylinder scan was used to obtain normalization factors and a co-registered helical CT scan acquired at the reference gate was initially used to compute the attenuation factors. Scatters were

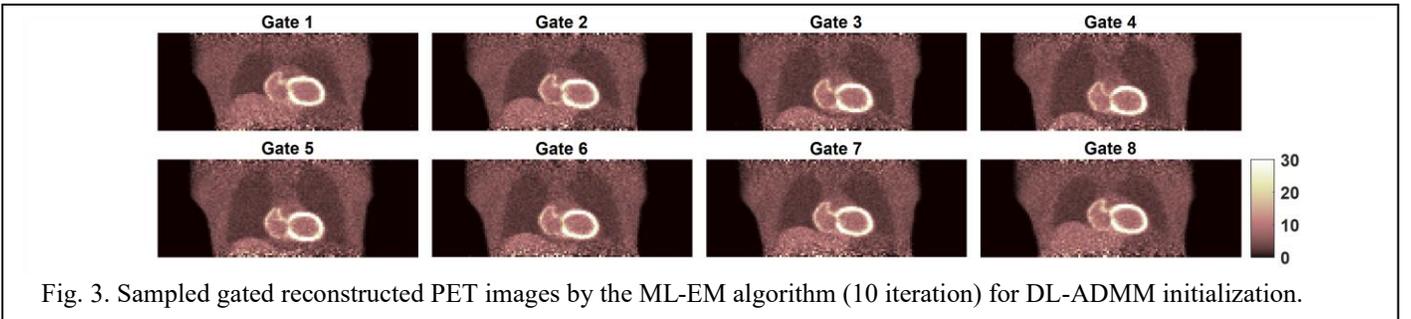


Fig. 3. Sampled gated reconstructed PET images by the ML-EM algorithm (10 iteration) for DL-ADMM initialization.

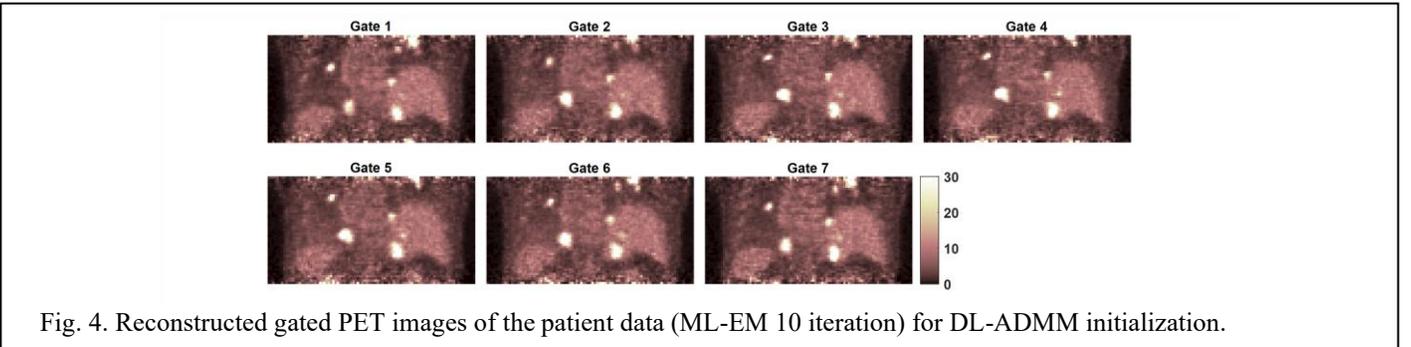
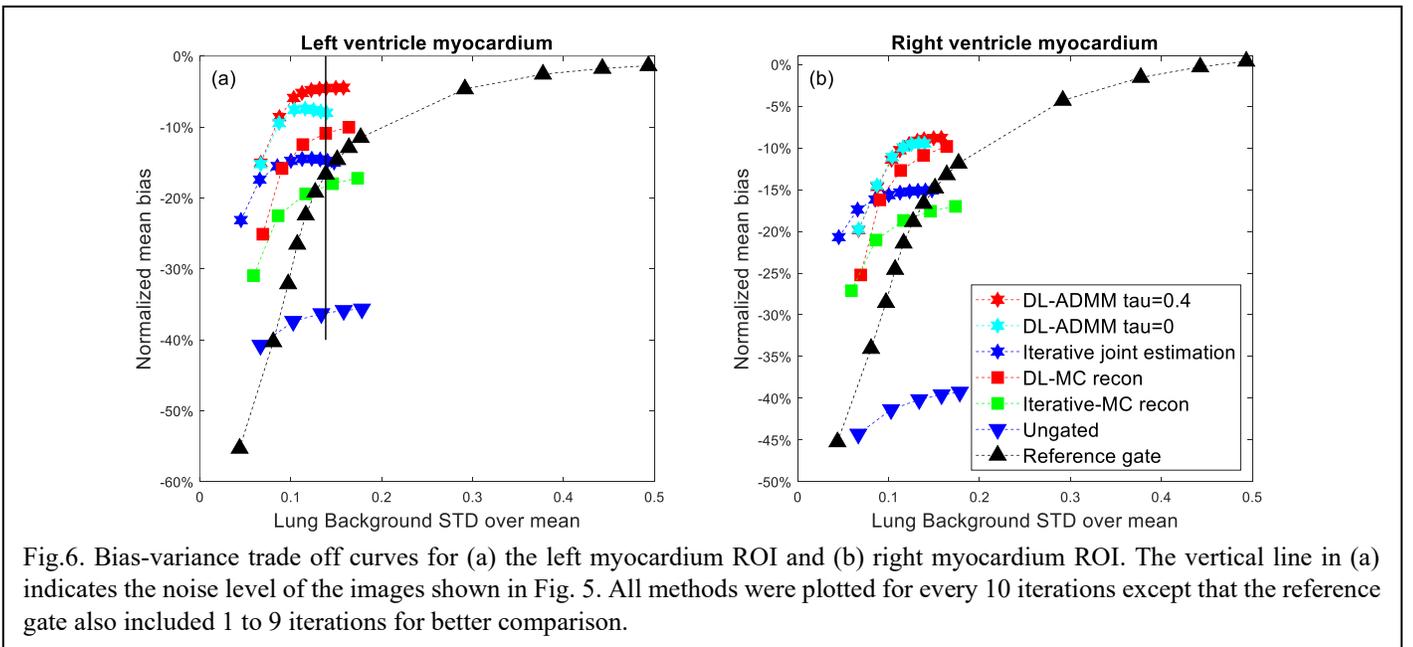
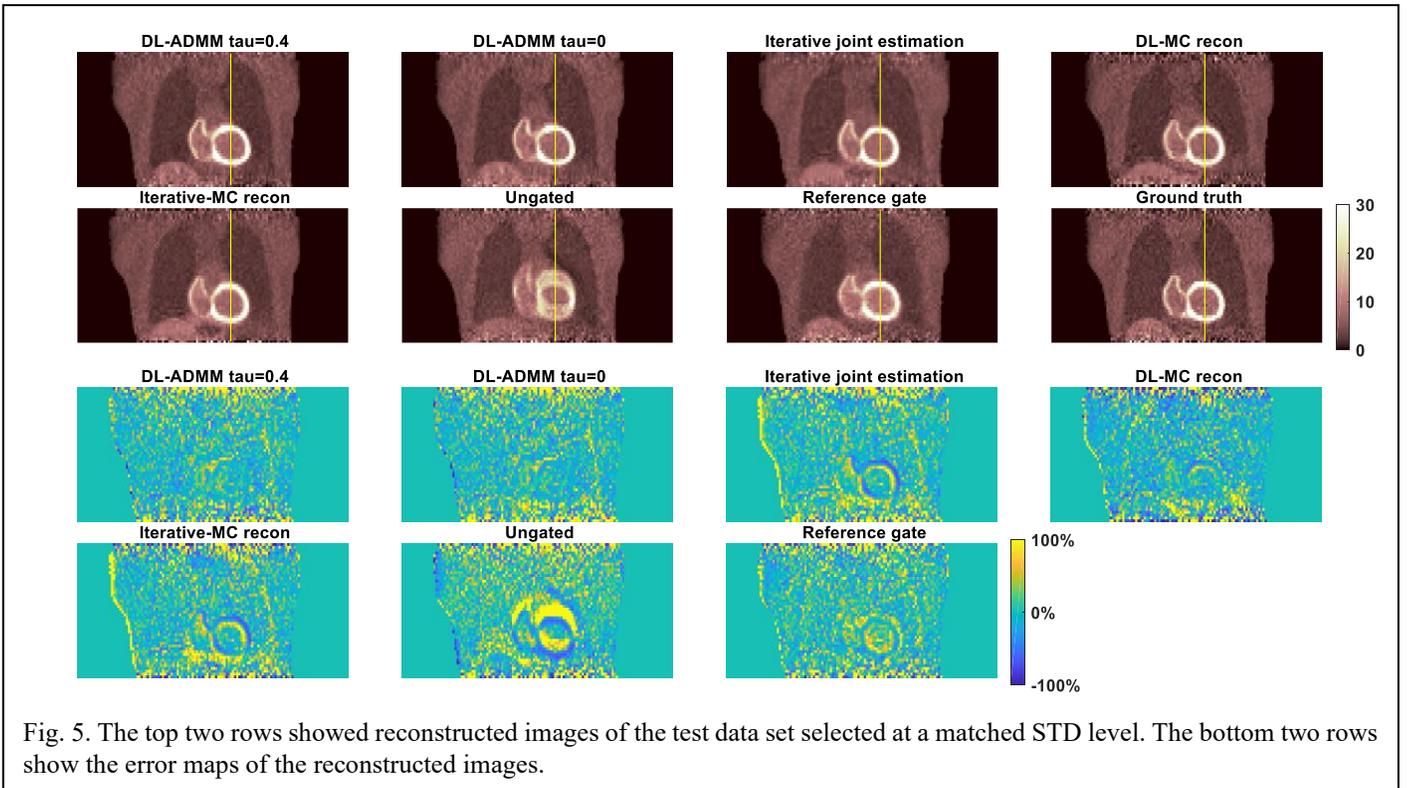


Fig. 4. Reconstructed gated PET images of the patient data (ML-EM 10 iteration) for DL-ADMM initialization.



estimated using the single scatter simulation algorithm and randoms were estimated using the delayed random sinogram. The reconstructed gated images are shown in Fig. 4. For the DL-ADMM algorithm, \mathbf{x}_m and \mathbf{x}_f were initialized using 10 iteration gated ML-EM reconstruction. Due to the difference between the real data and simulation data, the network was fine-tuned for 200 epochs using the patient data and its output was used to initialize the motion field θ_m . The fine tuning took 20 mins for all gated images for each bed position. While the fine-tuning is necessary here because we have only one set of real patient scan, we do not expect to fine-tune the network for every

patient scan after we process a sufficient number of patient images. These estimated motion fields were then used to deform the CT image to each moving gate to obtain a phase-matched attenuation map for attenuation correction. We ran DL-MC recon for 10 iterations for β initialization and warped β using θ_m for α_m initialization.

A lesion ROI was drawn on the reference gate image (Gate 3) for quantification. Due to the lack of the ground truth, a contrast-noise curve was used for comparison. 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

Table 2. NRMS of reconstructed images by different methods at the matched noise level.

Testing Phantom #	DL-ADMM $\tau_m = 0.4$	DL-ADMM $\tau_m = 0$	Iterative joint estimation	DL-MC recon	Iterative-MC recon	Ungated	Reference gate
#1	23.1%	23.9%	33.3%	26.2%	34.3%	42.8%	27.1%
#2	27.4%	27.6%	31.3%	30.6%	31.8%	38.0%	28.9%
#3	27.3%	28.3%	33.8%	35.5%	35.5%	45.7%	31.1%
#4	23.3%	24.4%	34.0%	33.9%	33.9%	42.2%	27.5%
#5	27.5%	28.3%	33.8%	28.1%	35.3%	46.1%	30.7%
#6	25.2%	25.6%	32.4%	27.4%	33.4%	38.6%	28.2%
#7	25.6%	26.2%	33.4%	27.7%	35.5%	42.1%	29.1%
#8	29.1%	30.1%	34.2%	31.4%	36.9%	41.9%	31.7%
#9	30.7%	31.3%	34.2%	33.5%	35.2%	45.7%	32.4%
#10	26.8%	27.6%	32.3%	31.4%	34.1%	38.1%	29.0%
#11	25.1%	25.4%	32.9%	28.0%	34.8%	41.7%	28.7%
Mean	26.5%	27.2%	33.2%	30.3%	34.6%	42.1%	29.5%
STD	2.2%	2.2%	0.9%	2.9%	1.3%	2.8%	1.7%

background noise was calculated as the standard deviation of the liver ROI over its mean.

IV. RESULTS

A. Simulation study

The reconstructed images of the test phantom by different methods are shown in Fig. 5. The reconstructions were chosen at a matched STD level (see Fig. 6). It shows that the proposed DL-ADMM joint estimation method can generate images with sharper boundaries in the myocardium region compared with other methods. The error maps revealed that the iterative joint estimation and iterative-MC recon suffer from inaccurate motion estimation leading to high quantification errors near the myocardial boundary. Clearly, DL-ADMM joint estimation method reduced the motion artifacts. We also plotted the bias-variance curves of the left and right myocardium ROIs in Fig. 6. All the methods were plotted for every 10 iteration except the reference gate which also included 1 to 9 iterations for better comparison. The proposed DL-ADMM algorithm reduces the bias compared with the ungated image without increasing the noise level and outperforms the other methods, especially for the right ventricle myocardium region, which is more susceptible to motion blurring. Fig. 7 compares the line profiles through the myocardium. The yellow lines in Fig. 5 indicate the position of the vertical profile. The average FWHMs of the two peaks are 1.99, 2.01, 2.18, 2.11, 2.38, 5.35 and 1.95 (pixels) for the DL-ADMM $\tau_m = 0.4$, DL-ADMM $\tau_m = 0$, iterative joint estimation, DL-MC recon, iterative-MC recon, ungated and ground truth, respectively, which indicated that the profile from the proposed DL-ADMM is the closest to the ground truth.

The algorithms were further evaluated based on their NRMS performance. The resulting NRMS calculated at a matched noise level were $26.5 \pm 2.2\%$, $33.2 \pm 0.9\%$, $30.3 \pm 2.9\%$, and $34.6 \pm 1.3\%$ for the DL-ADMM joint estimation, iterative joint

estimation, DL-MC recon, and iterative-MC recon, respectively, and $42.1 \pm 2.8\%$ for ungated reconstruction (Table 2). Clearly the proposed DL-ADMM joint estimation achieved the best performance compared with all the other methods. We observed that the NRMS of the DL-ADMM increased from $26.5 \pm 2.2\%$ to $27.2 \pm 2.2\%$ if we set $\tau_m = 0$, but the computational time could be reduced from about 8 mins to 1.5 mins per iteration.

B. Real data study

The reconstructed images of the real patient data from 2 bed positions are shown in Fig. 8 and Fig. 9. The contrast versus noise curves of a lung lesion (marked by the red circle) in 2 beds are plotted in Fig. 10. Visually, the proposed DL-ADMM method produced images with sharper boundary than the ungated image and other methods. A vertical line profile was drawn across the highlighted lesion and our proposed DL-

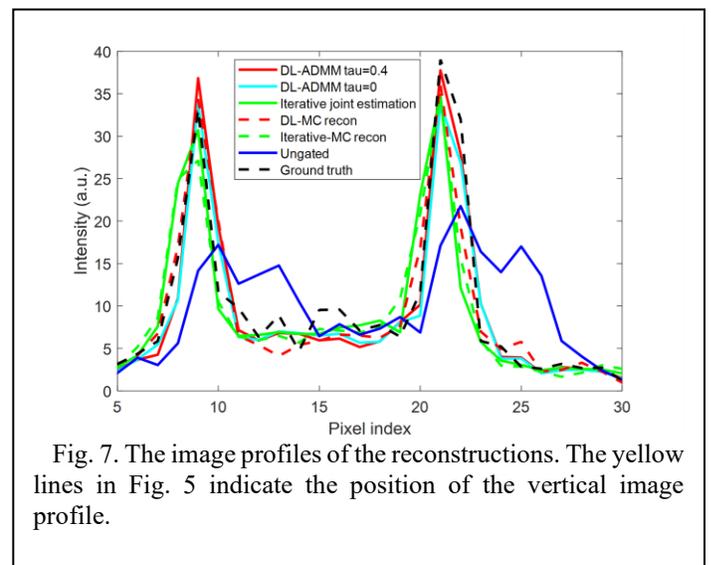


Fig. 7. The image profiles of the reconstructions. The yellow lines in Fig. 5 indicate the position of the vertical image profile.

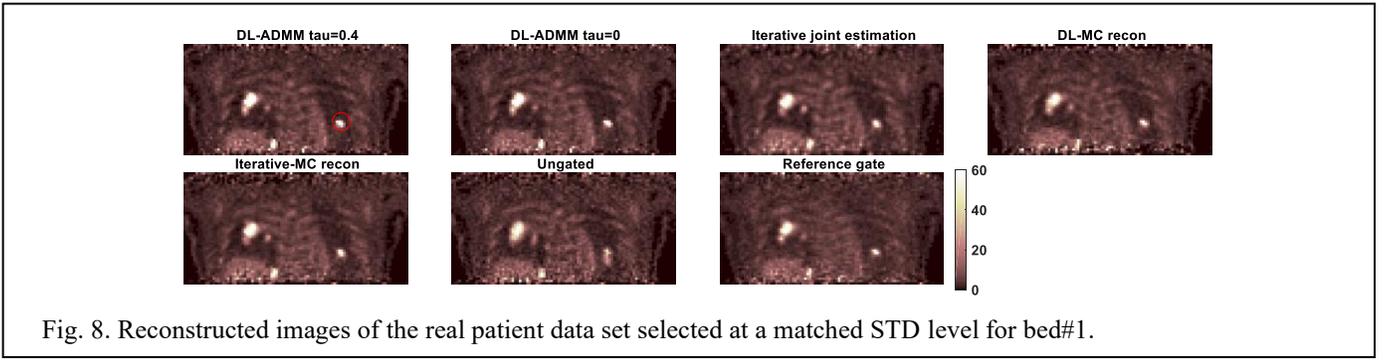


Fig. 8. Reconstructed images of the real patient data set selected at a matched STD level for bed#1.

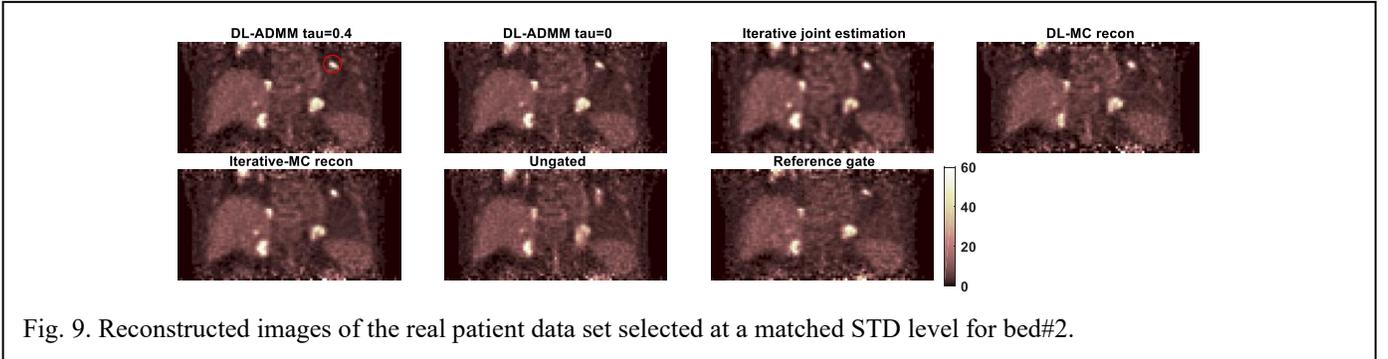


Fig. 9. Reconstructed images of the real patient data set selected at a matched STD level for bed#2.

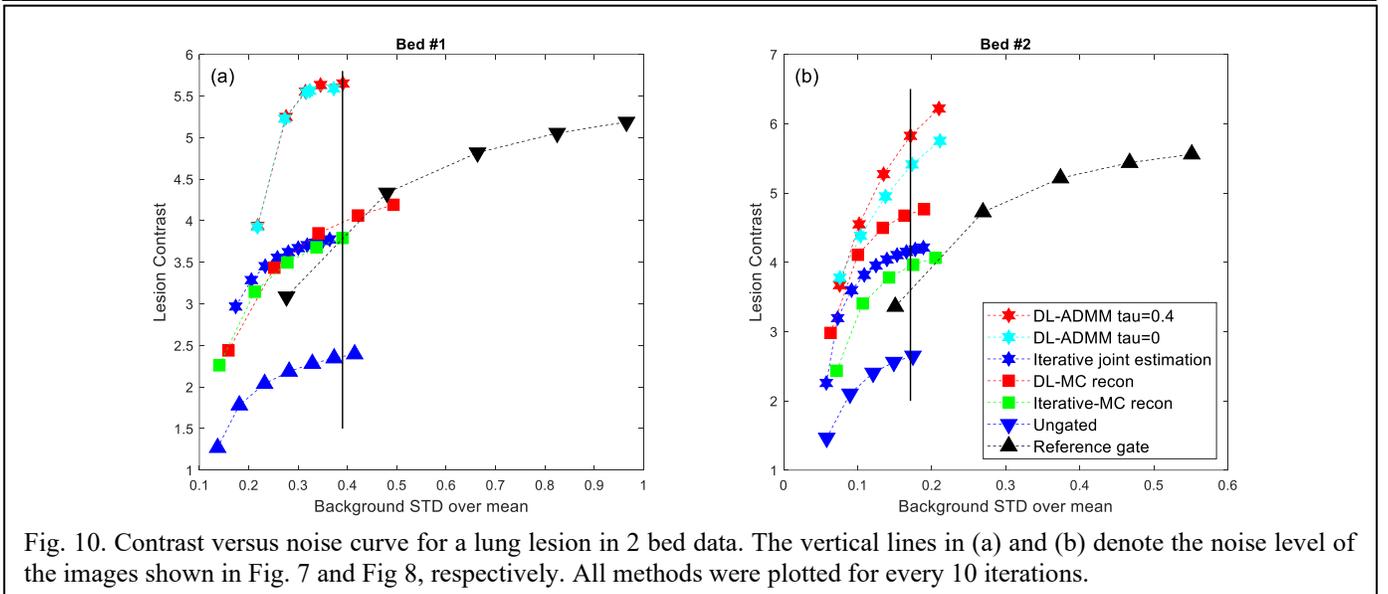


Fig. 10. Contrast versus noise curve for a lung lesion in 2 bed data. The vertical lines in (a) and (b) denote the noise level of the images shown in Fig. 7 and Fig 8, respectively. All methods were plotted for every 10 iterations.

ADMM produced narrow peaks (Fig. 11). The lesion contrast vs. noise plots also show that the proposed method achieved the highest lesion contrast among all motion correction methods at any matched noise level.

V. DISCUSSION

In this study, we proposed a robust joint estimation method that incorporates a DL-based image registration approach which leverages a cascaded network structure to estimate a coarse-to-fine motion field. To our knowledge, this is the first study incorporating a DL-based registration model into a joint estimation framework and demonstrating better performance than the iterative joint estimation approach. Also, our results are in accordance with reports in the literature that joint estimation

framework yields better performance [9]–[11] compared with embedding pre-calculated motion estimation methods [21],[29] into motion compensated reconstruction frameworks. Compared with the DL-MC recon [21], motion field estimation in the proposed DL-ADMM motion estimation method is estimated using updated reconstructed images from MAP-EM with better image quality, which would lead to better motion estimation. To demonstrate this point, we show the deformation fields of the DL-ADMM initialization and the final outputs of DL-ADMM with $\tau = 0$ and $\tau = 0.4$ for the XCAT phantom simulation in Fig. 12. We also compute the average L2 norm of the residue error between the estimated deformation field and the ground truth as a function of DL-ADMM iteration. Fig. 12(g) shows that the residue error decreases as the algorithm iterates and the residue error is lower with $\tau = 0.4$ than with

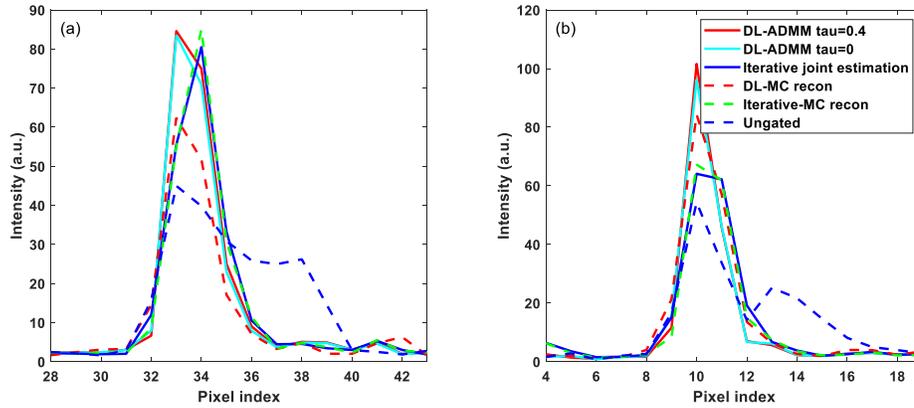


Fig. 11. The image profiles of the reconstructions for different method (a) bed 1 and (b) bed 2.

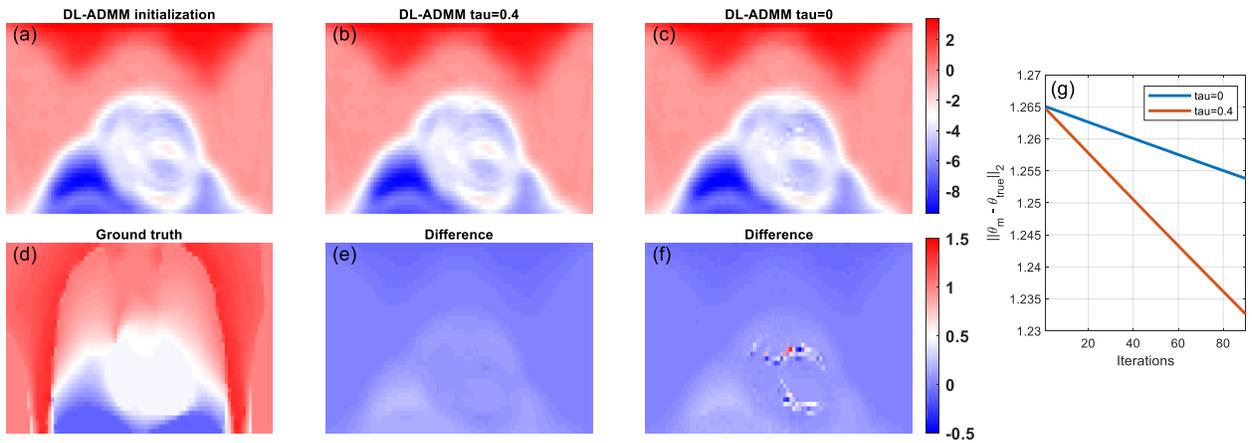


Fig 12. Deformation fields along the z-direction for (a) DL-ADMM initialization, (b) DL-ADMM $\tau = 0.4$ output, (c) DL-ADMM $\tau = 0$ output, (d) ground truth from XCAT program. Difference between the initialization and the output from (e) DL-ADMM $\tau=0.4$ or (f) DL-ADMM $\tau=0$. (g) L2 norm of the difference between the motion vector estimated by DL-ADMM ($\tau=0.4$ and $\tau=0$) and the ground truth as a function of DL-ADMM iteration.

$\tau = 0$. From the difference deformation fields between the initialization and final outputs (Fig. 12 (e) and (f)), we can see the result from DL-ADMM $\tau = 0$ is also noisier than that from DL-ADMM $\tau = 0.4$. We note that the difference between the estimated deformation and the ground truth is large in regions with a uniform activity distribution because the network training relies on image registration based on the PET image activity and there no features available in a uniform region for the estimation of the deformation field. However, such difference has little impact on PET image reconstruction. To improve the motion estimation, Susenburger *et al* extended the VoxelMorph network to accommodate lung sliding motion [35]. The same network can be easily incorporated in the joint estimation framework.

Although we can apply image denoising on the noisy gated images before motion estimation, it discards the temporal information provided by other gates. Our proposed joint estimation directly estimates the motion compensated image utilizing all gated sinograms which is more statistically efficient. For the iterative joint estimation [11], even though the OT algorithm is guaranteed to increase the objective function

monotonically, the cost function is nonconvex in the intensity-based non-rigid image registration problems, so it can be easily trapped in local minima and also slow in convergence rate. In our proposed DL-ADMM method, the deep neural network is trained using an ensemble of image pairs, which improved the performance of the image registration.

We note that Zhou *et al.* devised a unified motion correction and denoising adversarial network (MDPET) for low dose gated PET images [24], [25]. This study employed a recurrent layer for temporal motion feature learning and showed its ability to generate high-quality motion compensated PET images as a post-processing tool. Instead of post-processing the low count gated reconstructed images, we incorporate prior information from training pairs by embedding the pretrained network into a constrained image reconstruction algorithm which has been shown to be more robust and can provide better image quality [36], [37]. The DL-registration performance would also be affected by the mismatch in noise level, spatial resolution, and deformation magnitude between training and test datasets [38]. We observed that the STD is reduced when we include the network into joint estimation which indicates

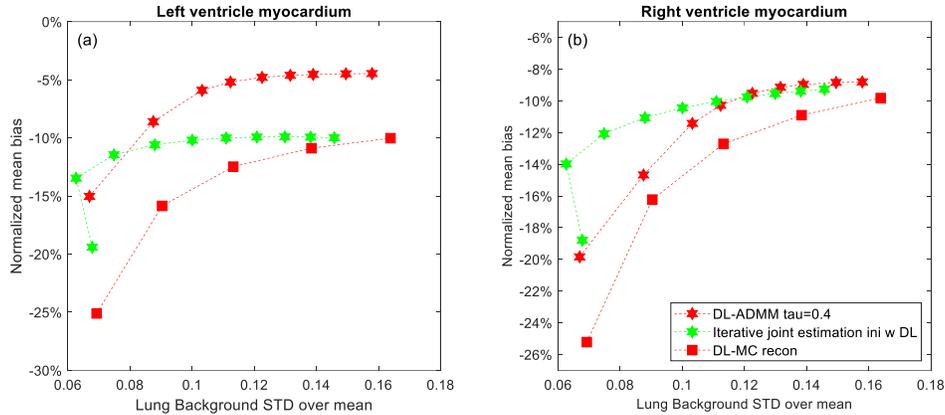


Fig 13. Bias-variance trade off curves for (a) the left myocardium ROI and (b) right myocardium ROI. All methods were plotted for every 10 iterations. The iterative joint estimation is initialized using the DL estimated motion field.

that DL-ADMM can reduce the variations by adding constraints. Another difference is that our network is unsupervised and does not require ground truth for training the network, while MDPET requires high-dose gated PET images as training labels, which are not readily available, to train an image denoising network.

In non-rigid image registration, the deformation field with high degrees of freedom can lead to unrealistic transformation on results such as folding and expanding in the absence of appropriate constraints. Thus, transformation using unconstrained motion fields is a large source of error in intensity-based registration. There have been some methods proposed using constraints to encourage the invertibility [29]. However, most of these methods require very high memory or computation complexity, making them poorly suited for dealing with multiple images simultaneously in the motion-compensated image reconstruction. Besides, for the heart sliding motion against the lungs throughout the cardiac cycle and respiration-induced sliding motion of the lungs against the chest wall, this constraint does not hold well. Large registration errors were observed at the region where the deformation fields were not continuous and they consistently appeared in all testing phantoms, which led to larger RMSE for the iterative based method. Therefore, the improved performance of the proposed DL-ADMM joint estimation method can be attributed to the aid of the constraint on the transformation from moving gated images to a fixed image via the deep neural network. Similar to our method, Öktem *et al* proposed a joint estimation method that incorporated a DL based on probabilistic diffeomorphic deformation model which is differentiable and invertible. However, their algorithm does not find the maximum of a joint objective function and cannot guarantee the convergence [39].

Given the nonconvex nature of the problem, the different performances between the DL-ADMM and iterative joint estimation could be due to different initializations. To further study the effect of different joint estimation methods, we used the same motion estimation from DL-MC for initialization. The bias versus background STD trade off curves are shown in Fig. 13. We found that the DL-MC based initialization did improve

the performance of the iterative joint estimation, but the resulting bias is still greater than that of our proposed method.

One drawback of the proposed DL-ADMM algorithm is that there are number of hyperparameters. In theory, the hyperparameters only affect the convergence speed of the ADMM algorithm and do not affect the final solution. However, in practice, they still influence the reconstructed image as the algorithm is stopped with a finite number of iterations. As a proof of the concept, we did not optimize the choices of hyperparameters for each dataset. Further fine tuning those hyperparameters could improve the results. Better optimization methods and more effective initialization strategies also deserve further investigations. Moreover, due to the lack of gated patient data, we only validate our proposed method using one patient with 50% overlapped bed positions. Further investigation with systematic analysis on a large population of patients is needed to validate our proposed method.

In this paper, we focused on incorporating DL-based motion estimation in the joint estimation framework. There are also other works on applying DL to PET image reconstruction or denoising [36], [37], [40], [41]. Although these DL techniques can generate PET images with good quality even at low-dose situations, the ability to utilize all counts from multiple gates is always beneficial. Therefore, the DL-denoising techniques are complementary to the motion compensation framework that we proposed here. The DL-ADMM algorithm can be combined with the DL-based denoising methods to further improve the PET image quality and it will be in our future work.

VI. CONCLUSION

In this work, we proposed a joint estimation framework incorporating deep learning-based image registration for motion estimation. We validated the proposed method using simulation and clinical data. The results showed that the proposed method can reduce motion artifacts while utilizing all gated PET data and outperform existing motion compensated reconstruction methods. Further validation using patient data will be performed in the future.

VII. ACKNOWLEDGMENTS

This work was supported in part by NIBIB under grant R01EB000194. The authors would like to thank Dr. Xuezhu Zhang for the help in SimSET simulations.

REFERENCES

- [1] J. Van Sluis, J. De Jong, J. Schaar, W. Noordzij, P. Van Snick, R. Dierckx, R. Borra, A. Willemsen, and R. Boellaard, "Performance characteristics of the digital biograph vision PET/CT system," *J. Nucl. Med.*, vol. 60, no. 7, pp. 1031–1036, 2019.
- [2] S. Chen, T. Cao, Y. Dong, S. Liu, G. Yang, Y. Zhao, X. Chen, L. Hu, and H. Shi, "NEMA NU2-2012 Performance Measurements of the United-Imaging uPMR790: A HD TOF Simultaneous PET/MR System," *J. Nucl. Med.*, vol. 59, no. supplement 1, p. 215, 2018.
- [3] Y. Lu, K. Fontaine, T. Mulnix, J. A. Onofrey, S. Ren, V. Panin, J. Jones, M. E. Casey, R. Barnett, P. Kench, R. Fulton, R. E. Carson, and C. Liu, "Respiratory motion compensation for PET/CT with motion information derived from matched attenuation-corrected gated PET data," *J. Nucl. Med.*, vol. 59, no. 9, pp. 1480–1486, 2018.
- [4] B. J. Lopresti, "Implementation and performance of an optical motion tracking system for high resolution brain pet imaging," *IEEE Trans. Nucl. Sci.*, vol. 46, no. 6 PART 3, pp. 2059–2067, 1999.
- [5] T. Feng, J. Wang, Y. Dong, J. Zhao, and H. Li, "A Novel Data-Driven Cardiac Gating Signal Extraction Method for PET," *IEEE Trans. Med. Imaging*, vol. 38, no. 2, pp. 629–637, 2019.
- [6] Y. Lu, M. Naganawa, T. Toyonaga, J. D. Gallezot, K. Fontaine, S. Ren, E. M. Revilla, T. Mulnix, and R. E. Carson, "Data-driven motion detection and event-by-event correction for brain pet: Comparison with vicra," *J. Nucl. Med.*, vol. 61, no. 9, pp. 1397–1403, 2020.
- [7] P. M. Bloomfield, T. J. Spinks, J. Reed, L. Schnorr, A. M. Westrip, L. Livieratos, R. Fulton, and T. Jones, "The design and implementation of a motion correction scheme for neurological PET," *Phys. Med. Biol.*, vol. 48, no. 8, pp. 959–978, 2003.
- [8] A. Rahmim, O. Rousset, and H. Zaidi, "Strategies for Motion Tracking and Correction in PET," *PET Clin.*, vol. 2, no. 2, pp. 251–266, 2007.
- [9] S. Y. Chun and J. A. Fessler, "Joint image reconstruction and nonrigid motion estimation with a simple penalty that encourages local invertibility," in *Medical Imaging 2009: Physics of Medical Imaging*, 2009, vol. 7258, p. 72580U.
- [10] M. Blume, A. Martinez-Möller, A. Keil, N. Navab, and M. Rafecas, "Joint reconstruction of image and motion in gated positron emission tomography," *IEEE Trans. Med. Imaging*, vol. 29, no. 11, pp. 1892–1906, 2010.
- [11] G. Wang and J. Qi, "A monotonic image-space algorithm for joint PET image reconstruction and motion estimation," in *2015 IEEE Nuclear Science Symposium and Medical Imaging Conference, NSS/MIC 2015*, 2016, pp. 1–3.
- [12] M. Burger, H. Dirks, and C. B. Schönlieb, "A variational model for joint motion estimation and image reconstruction," *SIAM J. Imaging Sci.*, vol. 11, no. 1, pp. 94–128, 2018.
- [13] C. Chen, B. Gris, and O. Oktem, "A new variational model for joint image reconstruction and motion estimation in spatiotemporal imaging," *SIAM J. Imaging Sci.*, vol. 12, no. 4, pp. 1686–1719, 2019.
- [14] H. Sokooti, B. de Vos, F. Berendsen, M. Ghafforian, S. Yousefi, B. P. F. Lelieveldt, I. Isgum, and M. Staring, "3D Convolutional Neural Networks Image Registration Based on Efficient Supervised Learning from Artificial Deformations," *arXiv Prepr. arXiv1908.10235*, 2019.
- [15] J. Krebs, T. Mansi, H. Delingette, L. Zhang, F. C. Ghesu, S. Miao, A. K. Maier, N. Ayache, R. Liao, and A. Kamen, "Robust non-rigid registration through agent-based action learning," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*, 2017, vol. 10433 LNCS, pp. 344–352.
- [16] G. Balakrishnan, A. Zhao, M. R. Sabuncu, A. V. Dalca, and J. Guttag, "An Unsupervised Learning Model for Deformable Medical Image Registration," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2018, pp. 9252–9260.
- [17] H. Li and Y. Fan, "Non-rigid image registration using self-supervised fully convolutional networks without training data," in *Proceedings - International Symposium on Biomedical Imaging*, 2018, vol. 2018-April, pp. 1075–1078.
- [18] A. Rahmim, J. C. Cheng, S. Blinder, M. L. Camborde, and V. Sossi, "Statistical dynamic image reconstruction in state-of-the-art high-resolution PET," *Phys. Med. Biol.*, vol. 50, no. 20, pp. 4887–4912, 2005.
- [19] S. Zhao, T. Lau, J. Luo, E. I. C. Chang, and Y. Xu, "Unsupervised 3D End-to-End Medical Image Registration with Volume Tweening Network," *IEEE J. Biomed. Heal. Informatics*, vol. 24, no. 5, pp. 1394–1404, 2020.
- [20] X. Chen, A. Diaz-Pinto, N. Ravikumar, and A. F. Frangi, "Deep learning in medical image registration," *Prog. Biomed. Eng.*, vol. 3, no. 1, p. 012003, 2021.
- [21] T. Li, M. Zhang, W. Qi, E. Asma, and J. Qi, "Motion correction of respiratory-gated PET images using deep learning based image registration framework," *Phys. Med. Biol.*, vol. 65, no. 15, p. 155003, 2020.
- [22] M. Jaderberg, K. Simonyan, A. Zisserman, and K. Kavukcuoglu, "Spatial transformer networks," in *Advances in Neural Information Processing Systems*, 2015, vol. 2015-Janua, pp. 2017–2025.
- [23] J. Fan, X. Cao, Z. Xue, P. T. Yap, and D. Shen, "Adversarial similarity network for evaluating image alignment in deep learning based registration," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*, 2018, vol. 11070 LNCS, pp. 739–746.

- [24] B. Zhou, Y. J. Tsai, and C. Liu, "Simultaneous Denoising and Motion Estimation for Low-Dose Gated PET Using a Siamese Adversarial Network with Gate-to-Gate Consistency Learning," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*, 2020, vol. 12267 LNCS, pp. 743–752.
- [25] B. Zhou, Y. J. Tsai, X. Chen, J. S. Duncan, and C. Liu, "MDPET: A Unified Motion Correction and Denoising Adversarial Network for Low-dose Gated PET," *IEEE Trans. Med. Imaging*, vol. 40, no. 11, pp. 3154–3164, 2021.
- [26] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein, *Distributed optimization and statistical learning via the alternating direction method of multipliers*, vol. 3, no. 1. Foundations and Trends in Machine Learning, 2010.
- [27] G. Wang and J. Qi, "Penalized likelihood PET image reconstruction using patch-based edge-preserving regularization," *IEEE Trans. Med. Imaging*, vol. 31, no. 12, pp. 2194–2204, 2012.
- [28] D. P. Kingma and J. L. Ba, "Adam: A method for stochastic optimization," *3rd Int. Conf. Learn. Represent.*, 2015.
- [29] S. Y. Chun and J. A. Fessler, "A simple regularizer for B-spline nonrigid image registration that encourages local invertibility," *IEEE J. Sel. Top. Signal Process.*, vol. 3, no. 1, pp. 159–169, 2009.
- [30] W. P. Segars, G. Sturgeon, S. Mendonca, J. Grimes, and B. M. W. Tsui, "4D XCAT phantom for multimodality imaging research," *Med. Phys.*, vol. 37, no. 9, pp. 4902–4915, 2010.
- [31] R. L. Harrison, D. R. Haynor, S. B. Gillispie, S. D. Vannoy, M. S. Kaplan, and T. K. Lewellen, "A public-domain simulation system for emission tomography-photon tracking through heterogeneous attenuation using importance sampling," in *Journal of Nuclear Medicine*, 1993, vol. 34, no. 5, pp. P60–P60.
- [32] T. Kaneta, M. Ogawa, N. Motomura, H. Iizuka, T. Arisawa, A. Hino-Shishikura, K. Yoshida, and T. Inoue, "Initial evaluation of the Celesteion large-bore PET/CT scanner in accordance with the NEMA NU2-2012 standard and the Japanese guideline for oncology FDG PET/CT data acquisition protocol version 2.0," *EJNMMI Res.*, vol. 7, no. 1, p. 83, 2017.
- [33] X. Zhang, J. Zhou, G. Wang, J. Poon, S. Cherry, R. Badawi, and J. Qi, "Feasibility study of micro-dose total-body dynamic PET imaging using the EXPLORER scanner," *J. Nucl. Med.*, vol. 55, no. supplement_1, pp. 269-, 2014.
- [34] J. Zhou and J. Qi, "Efficient fully 3D list-mode TOF PET image reconstruction using a factorized system matrix with an image domain resolution model," *Phys. Med. Biol.*, vol. 59, no. 3, pp. 541–559, 2014.
- [35] M. Susenburger, P. Paysan, R. Savjani, J. Maier, P. Igor, and M. Kachelrieß, "DeepSLM: Image Registration Aware of Sliding Interfaces for Motion-Compensated Reconstruction," in *Proceedings of the 16th Virtual International Meeting on Fully 3D Image Reconstruction in Radiology and Nuclear Medicine*, 2021, pp. 72–75.
- [36] K. Gong, J. Guan, K. Kim, X. Zhang, J. Yang, Y. Seo, G. El Fakhri, J. Qi, and Q. Li, "Iterative PET image reconstruction using convolutional neural network representation," *IEEE Trans. Med. Imaging*, vol. 38, no. 3, pp. 675–685, 2019.
- [37] Z. Xie, R. Baikejiang, T. Li, X. Zhang, K. Gong, M. Zhang, W. Qi, E. Asma, and J. Qi, "Generative adversarial network based regularized image reconstruction for PET," *Phys. Med. Biol.*, vol. 65, no. 12, p. 125016, 2020.
- [38] M. D. Ketcha, T. De Silva, R. Han, A. Uneri, S. Vogt, G. Kleinszig, and J. H. Siewerdsen, "Effect of statistical mismatch between training and test images for CNN-based deformable registration," *J. Med. Imaging*, vol. 6, no. 4, p. 28, 2019.
- [39] O. Öktem, C. Pouchol, and O. Verdier, "Spatiotemporal PET Reconstruction Using ML-EM with Learned Diffeomorphic Deformation," in *International Workshop on Machine Learning for Medical Image Reconstruction*, 2019, vol. 11905 LNCS, pp. 151–162.
- [40] J. Ouyang, K. T. Chen, E. Gong, J. Pauly, and G. Zaharchuk, "Ultra-low-dose PET reconstruction using generative adversarial network with feature matching and task-specific perceptual loss," *Med. Phys.*, vol. 46, no. 8, pp. 3555–3564, 2019.
- [41] T. Yokota, K. Kawai, M. Sakata, Y. Kimura, and H. Hontani, "Dynamic PET image reconstruction using nonnegative matrix factorization incorporated with deep image prior," in *Proceedings of the IEEE International Conference on Computer Vision*, 2019, vol. 2019-October, pp. 3126–3135.
- [42] G. Balakrishnan, A. Zhao, M. R. Sabuncu, J. Guttag, and A. V Dalca, "Voxelmorph: a learning framework for deformable medical image registration," *IEEE Trans. Med. Imaging*, vol. 38, no. 8, pp. 1788–1800, 2019.