Joint low-count PET/CT segmentation and reconstruction with paired variational neural networks

Lim, Hongki, Dewaraja, Yuni, Fessler, Jeffrey

Hongki Lim, Yuni K. Dewaraja, Jeffrey A. Fessler, "Joint low-count PET/CT segmentation and reconstruction with paired variational neural networks," Proc. SPIE 11312, Medical Imaging 2020: Physics of Medical Imaging, 113120U (16 March 2020); doi: 10.1117/12.2543252

SPIE.

Event: SPIE Medical Imaging, 2020, Houston, Texas, United States

Joint low-count PET/CT segmentation and reconstruction with paired variational neural networks

Hongki Lim[a], Yuni K. Dewaraja[b], and Jeffrey A. Fessler[a]

[a]Electrical Engineering and Computer Science, University of Michigan, Ann Arbor, MI
[b]Radiology, University of Michigan, Ann Arbor, MI

ABSTRACT

Most existing fully automatic or semi-automatic medical imaging segmentation methods start from reconstructed images. However, a framework for joint segmentation and image reconstruction can be beneficial because both tasks can be mutually dependent. Better segmentation can improve image reconstruction and vice-versa. We propose to perform joint PET image reconstruction and fully automatic PET image segmentation, using the CT image from a PET-CT scanner as a given input. Within a unified framework, our proposed method generates a PET image and a segmentation mask utilizing two connected trained networks: 1) a network dedicated to denoising the PET image with boundary information from the segmentation network. While reconstructing the PET image, the algorithm exploits the denoised image recovered from the trained network. 2) a segmentation network dedicated to estimating the lesion and background (e.g., liver) masks using PET/CT information. A boundary indicator image is generated based on the gradients of segmentation masks. We simulated extremely low-count PET, typical for Y-90 imaging, where traditional segmentation and reconstruction methods tend to perform poorly. For PET reconstruction, proposed method using true boundary improves CNR (RMSE) by 28.9 % (49.1%) and 16.8 % (13.2%) compared to EM and proposed method without using boundary. For multi-modal segmentation, our proposed method improved global Dice score in tumor by 70.6% compared to our proposed segmentation framework using only CT information.

Keywords: PET/CT, Joint, Segmentation, Reconstruction, Neural network, Y-90

1. INTRODUCTION

Lesion segmentation has been identified as a factor that leads to significant variability in Y-90 radioembolization dosimetry.[1] PET-based segmentation is particularly challenging due to the inherent noise and poor spatial resolution. The (radiologist defined) manual segmentation on morphological images currently used for dosimetry is time consuming, labor intensive and has high intra- and inter-observer variability.[2,3] Most existing fully automatic or semi-automatic medical imaging segmentation is typically performed after reconstruction and typically uses a single modality. This sequential approach leads to propagation of errors from noisy reconstructions to the segmentation step. Furthermore, using only a single modality does not fully exploit the information from dual-modality systems like PET/CT. Hence, fully automatic joint segmentation-reconstruction using multi-modality images, specifically the CT image from a PET-CT scanner, as proposed here is highly desirable. Previously, for PET/MRI a joint regularization with a common edge function[4] was proposed using a traditional (non-learning) algorithm that has the potential to fail at low counts, typical for Y-90 PET. We propose a deep-learning based unified PET/CT reconstruction and fully automatic segmentation method that exploits the mutual dependency of the two tasks; better estimation in one can benefit the other.

This work is supported by NIH-NIBIB grant R01EB022075.
Hongki Lim: hongki@umich.edu
Yuni K. Dewaraja: yuni@med.umich.edu
Jeffrey A. Fessler: fessler@umich.edu

Medical Imaging 2020: Physics of Medical Imaging, edited by Guang-Hong Chen, Hilde Bosmans, Proc. of SPIE
Vol. 11312, 113120U · © 2020 SPIE · CCC code: 1605-7422/20/$21 · doi: 10.1117/12.2543252

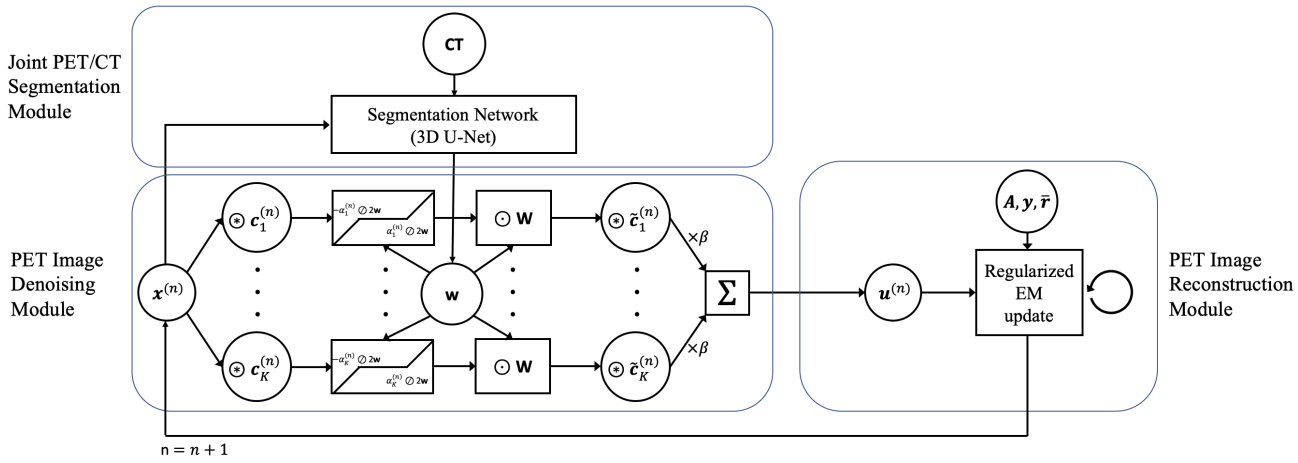

Figure 1. High-level block diagram of our proposed method

2. METHODS

2.1 Joint PET/CT reconstruction and segmentation framework

We propose to perform joint PET image reconstruction and PET/CT segmentation (optimizing CT reconstruction is not a focus of this work). Our proposed joint segmentation-reconstruction framework exploits the capabilities of deep learning networks. The proposed unified framework is inspired by the following variational problem:

$$\hat{\boldsymbol{x}} = \arg\min_{\boldsymbol{x}} f(\boldsymbol{x}) + \beta \mathsf{R}(\boldsymbol{x}; \boldsymbol{w}), \tag{1}$$

where $\boldsymbol{x}$ is the unknown PET image, $f(\boldsymbol{x})$ is the Poisson negative log-likelihood for measurement $\boldsymbol{y}$ and estimated measurement means $\bar{\boldsymbol{y}}(\boldsymbol{x}) = \boldsymbol{A}\boldsymbol{x} + \bar{\boldsymbol{r}}$, the matrix $\boldsymbol{A}$ denotes the system model, $\bar{\boldsymbol{r}}$ denotes the mean background events such as scatter and random coincidences, $\mathsf{R}(\boldsymbol{x}; \boldsymbol{w})$ is the joint regularization term, $\boldsymbol{w}$ is a boundary indicator image (zero-valued at boundary) given by segmentation network utilizing both CT and PET modalities, and β is the regularization parameter. The regularization term is composed of convolutional operations followed by a thresholding operation to promote sparsity:

$$\mathsf{R}(\boldsymbol{x}; \boldsymbol{w}) = \sum_{k=1}^{K} \|\boldsymbol{c}_k * \boldsymbol{x} - \boldsymbol{z}_k\|_{\boldsymbol{W}}^2 + \alpha_k \|\boldsymbol{z}_k\|_1 , \tag{2}$$

where $\boldsymbol{W} = \mathrm{diag}\{\boldsymbol{w}\}$, $\{\boldsymbol{c}_k : k = 1, ..., K\}$ is a set of convolution filters, $*$ denotes convolution operation, $\{\boldsymbol{z}_k : k = 1, ..., K\}$ is a set of sparse codes, and K is the number of filters. The diagonal weighting matrix W is designed to avoid smoothing across boundaries between different regions.[5,6]

A traditional optimization approach for solving (1) would be to use a block coordinate descent algorithm that alternatively updates $\{\boldsymbol{z_k}\}$ and $\boldsymbol{x}$:

$$\boldsymbol{z}_k^{(n+1)} = \arg\min_{\boldsymbol{z}_k} \left\| \boldsymbol{c}_k * \boldsymbol{x}^{(n)} - \boldsymbol{z}_k \right\|_{\boldsymbol{W}}^2 + \alpha_k \|\boldsymbol{z}_k\|_1 = \mathcal{T}(\boldsymbol{c}_k * \boldsymbol{x}^{(n)}, \alpha_k \oslash 2\boldsymbol{w}) \tag{3}$$

$$\boldsymbol{x}^{(n+1)} = \arg\min_{\boldsymbol{x}} f(\boldsymbol{x}) + \beta \Big(\sum_{k=1}^{K} \left\| \boldsymbol{c}_k * \boldsymbol{x} - \boldsymbol{z}_k^{(n+1)} \right\|_{\boldsymbol{W}}^2 \Big), \tag{4}$$

where $\mathcal{T}(\cdot, \cdot)$ is the element-wise soft thresholding operator:

$$\mathcal{T}(\boldsymbol{t}, \boldsymbol{q})_j := \mathrm{sign}(t_j) \max(|t_j| - q_j, 0). \tag{5}$$

When we solve (4), we do not penalize the filtered image at jth voxel when $w_j = 0$. Since $\boldsymbol{w}$ is a boundary indicator image (zero-valued at edge), our formulation enforces taking no *priori* information (such as similarity between adjacent voxels) on the edge (where $w_j = 0$) area, hence promoting edge-preserving. Inspired by this edge-preserving regularization method with convolution operators, our proposed framework alternatively updates the reconstructed image $\boldsymbol{x}$ and the denoised image $\boldsymbol{u}$ with convolution filters $\boldsymbol{c}_k^{(n)}$ and soft-thresholding values $\alpha_k^{(n)}$ trained at each iteration:

$$\boldsymbol{u}^{(n+1)} = \sum_{k=1}^{K} \tilde{\boldsymbol{c}}_k^{(n)} * \left(\boldsymbol{W}\left(\mathcal{T}(\boldsymbol{c}_k^{(n)} * \boldsymbol{x}^{(n)}, \alpha_k^{(n)} \oslash 2\boldsymbol{w})\right)\right) \tag{6}$$

$$\boldsymbol{x}^{(n+1)} = \arg\min_{\boldsymbol{x}} f(\boldsymbol{x}) + \beta \left\| \boldsymbol{x} - \boldsymbol{u}^{(n+1)} \right\|_2^2, \tag{7}$$

where $\tilde{\boldsymbol{c}}_k^{(n)}$ is flipped version of $\boldsymbol{c}_k^{(n)}$. (3)-(4) is identical to (6)-(7) when $\boldsymbol{c}_k^{(n)}$ is fixed at each nth iteration and $\sum_{k=1}^{K} \boldsymbol{C}_k^T \boldsymbol{W} \boldsymbol{C}_k = \boldsymbol{I}$ where $\boldsymbol{C}_k$ is sparse matrix satisfying $\boldsymbol{C}_k \boldsymbol{x} \iff \boldsymbol{c}_k * \boldsymbol{x}$. In the training phase, we train the set of filters $\{\boldsymbol{c}_k : \boldsymbol{c}_1, ..., \boldsymbol{c}_K\}$ and soft-thresholding values $\{\alpha_k : \alpha_1, ..., \alpha_K\}$ to map the previously estimated image to high quality image (e.g., true image if possible) at each iteration:

$$\{\hat{\boldsymbol{c}}_1^{(n)}, \ldots, \hat{\boldsymbol{c}}_K^{(n)}\}, \{\hat{\alpha}_1^{(n)}, \ldots, \hat{\alpha}_K^{(n)}\} = \arg\min_{\{\boldsymbol{c}_k\}, \{\alpha_k\}} \left\| \boldsymbol{x}_{\text{true}} - \sum_{k=1}^{K} \tilde{\boldsymbol{c}_k} * \left(\boldsymbol{W}\left(\mathcal{T}(\boldsymbol{c}_k * \boldsymbol{x}^{(n)}, \alpha_k \oslash 2\boldsymbol{w})\right)\right) \right\|_2^2. \tag{8}$$

Fig. 1 illustrates equation (6)-(7). To generate the boundary image $\boldsymbol{w}$, we use multimodal data (PET/CT). As shown in Fig. 1, using given CT image and sequentially updated PET image $\boldsymbol{x}^{(n)}$, the segmentation network generates the segmentation masks. A boundary image $\boldsymbol{w}$ is generated based on the gradients of segmentation mask. For segmentation network, we implemented 3-D version of U-Net.[7] Input of segmentation network is four dimensional array (channel×image depth×image height×image width). First channel is CT image and second channel is PET image.

2.2 Experimental setup: PET simulation

We use LiTS (Liver Tumor Segmentation) dataset[8] to show the efficacy of our proposed method. Among 130 training samples in 3-D, we use 15 samples for training segmentation network and PET denoising module, and 4 samples for testing the performance. We use label images of LiTS dataset for generating true PET image. We changed the tumor value to 5 to set the tumor-to-liver ratio as 5:1 a typical value in patients following Y-90 radioembolization. We downsample the CT sized label image (512×512×CT slice#) to PET sized image (128×128×PET slice#) with a voxel size 4.0×4.0×4.0 (mm^3). We simulated the extremely low-count scan with total true coincidences and random fraction based on numbers from patient PET imaging performed after radioembolization.[9] Fig. 2 shows CT and label image provided by the dataset and PET and boundary image that we generate based on label image.

2.3 Experimental setup: Training denoising and segmentation networks

2.3.1 PET denoising network: Sparse convolutional autoencoder

We train 3D convolutional filters and thresholding values in each iteration using PyTorch[10] deep-learning library. We train a 10 outer-iteration where each outer-iteration has $K = 192$ sets of thresholding values and convolutional encoding/decoding filters. We set the size of each filter as $3 \times 3 \times 3$. We use Adam optimization method to train the network with learning rate of 10^{-2} for encoding filters, 10^{-3} for decoding filters, and 10^{-1} for thresholding values. We use 400 epochs to train the denoising network at each outer-iteration.

2.3.2 Segmentation network: U-Net

For training U-Net, we also use Adam optimization method with learning rate of 10^{-2}. We set the number of convolutional filter channels of first layer of encoder as 32 with four times of contraction and four times of expansion.

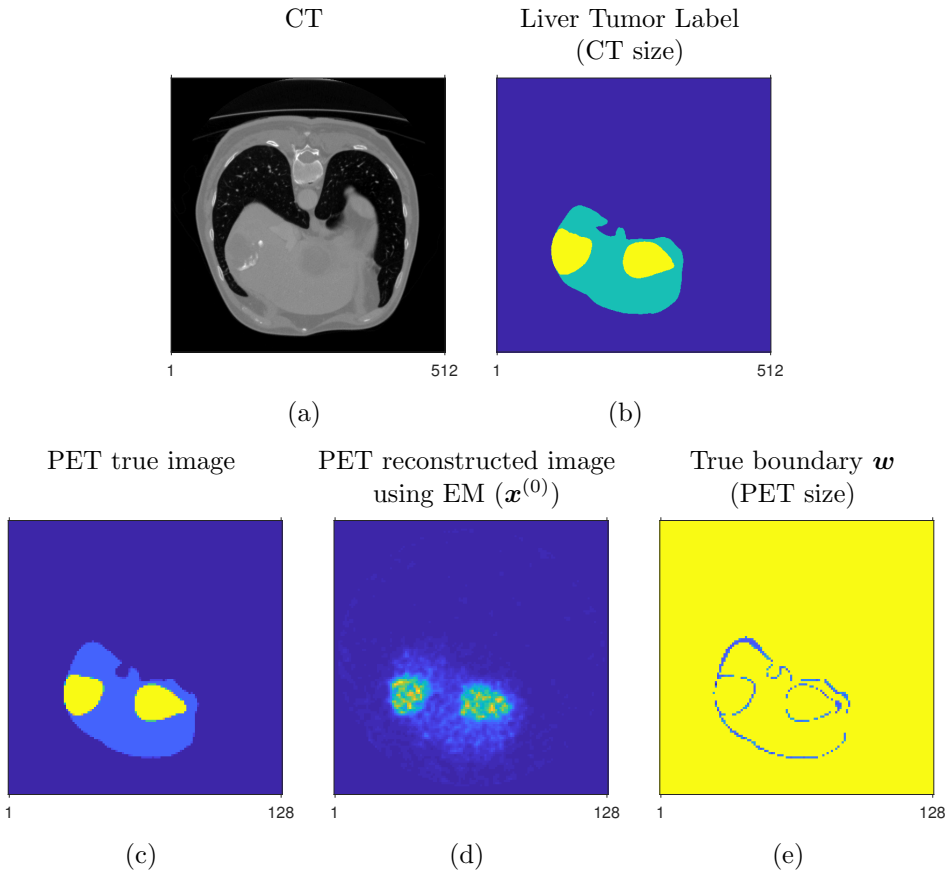

Figure 2. LiTS dataset provides CT image (a) in HU unit and label image (b) (tumor:2, liver:1) corresponding to CT. We generated PET image (d) simulating Y-90 PET after radioembolization. We set the tumor-to-liver ratio as 5:1. Image (e) shows true boundary based on label image. Boundary is zero-valued. Slice thickness of PET-sized images (c)-(e) is 4 times that of CT-sized images (a)-(b).

2.4 Evaluation metrics

We evaluate PET reconstruction performance with contrast to noise ratio (CNR) and root mean squared error (RMSE):

$$\text{CNR} = \frac{C_{\text{Lesion}} - C_{\text{Liver}}}{STD_{\text{Liver}}}$$

$$\text{RMSE } (\%) = \sqrt{\frac{\sum_j (\boldsymbol{x}_{\text{true}}[j] - \hat{\boldsymbol{x}}[j])^2}{J_{\text{FOV}}}} \times 100,$$

where C_{VOI} is mean counts in the volume of interest (VOI), STD_{Liver} is standard deviation between voxel values in background liver, and J_{FOV} is the total number of voxels in field of view (FOV). We evaluate segmentation performance with global Dice similarity coefficient:[8]

$$\text{Dice}(A, B) = \frac{2|A \cap B|}{|A| + |B|},$$

where A is a estimated segmentation mask and B is a true segmentation mask.

3. RESULTS

For evaluation of PET reconstruction, we compare the proposed joint framework (using true or estimated boundary) to the standard EM (1 subset), proposed PET reconstruction method without using boundary information.[11]

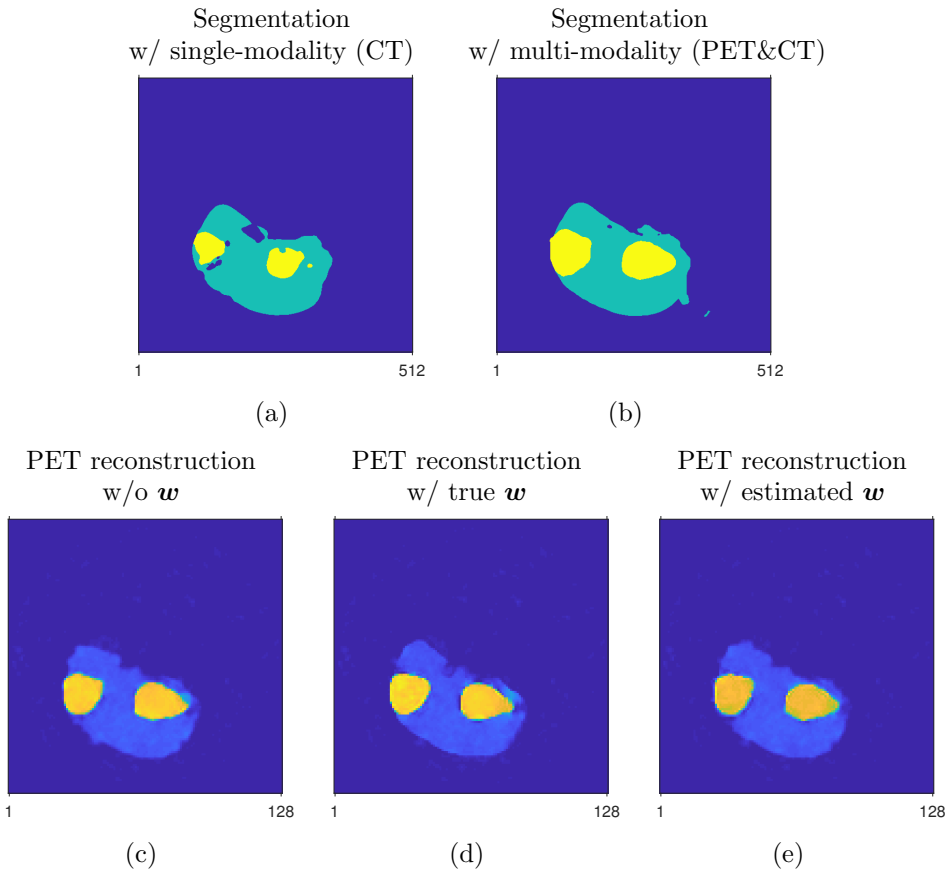

Figure 3. Segmentation results (a)-(b) and reconstructed images (c)-(e) from different reconstruction and segmentation methods.

$\mathsf{R}(\boldsymbol{x};\boldsymbol{w})$ in (2) is replaced by

$$\mathsf{R}(\boldsymbol{x}) = \sum_{k=1}^{K} \|\boldsymbol{c}_k * \boldsymbol{x} - \boldsymbol{z}_k\|_2^2 + \alpha_k \|\boldsymbol{z}_k\|_1 . \tag{9}$$

For evaluation of segmentation, we compare the proposed method to the segmentation network using only single modality (CT). The proposed method with or without boundary information significantly improves overall reconstruction performance over the standard reconstruction method. See Table 1 and Fig. 3(c)-(e). Table 1 shows that proposed method achieves best results in all evaluation metrics when true boundary is given. In particular, proposed method using true boundary improves CNR (RMSE) by 28.9 % (49.1%) and 16.8 % (13.2%) compared to EM and proposed method without using boundary. Our proposed method gives more clear improvement in segmentation. See Table 2 and Fig. 3(a)-(b). Our proposed method improves tumor Dice score by 70.6% compared to the segmentation method using only single modality (CT).

4. DISCUSSION AND CONCLUSION

We presented a framework that jointly reconstructs and segments the PET images aided by CT images from multimodal PET/CT systems. This paper is a proof-of-concept study to demonstrate that segmentation and reconstruction tasks can be mutually beneficial to each other. The proposed joint multi-modality framework achieves significant qualitative and quantitative improvements over the standard image reconstruction and the segmentation utilizing single modality. However, further investigation of the potential challenges including impact of misregistration between modalities and inconsistency in activity distribution (tumor-to-liver ratio,

Table 1. Evaluation result on PET reconstruction. Numbers shown here are averaged over 4 testing samples.

Method	CNR	RMSE
EM	8.72	14.36
Proposed method w/o boundary	9.62	8.41
Proposed method w/ true boundary	11.2	7.30
Proposed method w/ estimated boundary	9.89	8.41

Table 2. Evaluation result on segmentation. We evaluate on 4 testing samples combined into a single volume.

Method	Dice tumor	Dice liver
Proposed method using single-modality (CT)	0.51	0.92
Proposed method using multi-modality (PET/CT)	0.87	0.93

uniformity in background liver) is required. Moreover, in Y-90 imaging, anatomical and emission image based lesion boundaries can differ significantly not only due to variability in segmentation, but also due to tumor biology (micro-spheres are delivered based on blood flow, not metabolized). We plan to investigate whether including such cases in our training set would enable the CNN to learn those outlier cases. Future work also includes training and testing on measurement datasets where the activity distribution is non-uniform and misregistration can be significant as well as training neural networks for segmentation and denoising together with a weighted combination of loss functions.

REFERENCES

[1] Mikell, J. K., Kaza, R. K., Roberson, P. L., Younge, K. C., Srinivasa, R. N., Majdalany, B. S., Cuneo, K. C., Owen, D., Devasia, T., Schipper, M. J., et al., "Impact of Y-90 PET gradient-based tumor segmentation on voxel-level dosimetry in liver radioembolization," *EJNMMI physics* **5**(1), 31 (2018).

[2] Foster, B., Bagci, U., Mansoor, A., Xu, Z., and Mollura, D. J., "A review on segmentation of positron emission tomography images," *Computers in biology and medicine* **50**, 76–96 (2014).

[3] Hatt, M., Lee, J. A., Schmidtlein, C. R., Naqa, I. E., Caldwell, C., De Bernardi, E., Lu, W., Das, S., Geets, X., Gregoire, V., et al., "Classification and evaluation strategies of auto-segmentation approaches for PET: Report of AAPM task group no. 211," *Medical physics* **44**(6) (2017).

[4] Zhang, Y. and Zhang, X., "PET-MRI joint reconstruction with common edge weighted total variation regularization," *Inverse Problems* **34**(6), 065006 (2018).

[5] Fessler, J. A., Clinthorne, N. H., and Rogers, W. L., "Regularized emission image reconstruction using imperfect side information," *IEEE Transactions on Nuclear Science* **39**(5), 1464–1471 (1992).

[6] Alessio, A. M. and Kinahan, P. E., "Improved quantitation for PET/CT image reconstruction with system modeling and anatomical priors," *Medical physics* **33**(11), 4095–4103 (2006).

[7] Ronneberger, O., Fischer, P., and Brox, T., "U-net: Convolutional networks for biomedical image segmentation," *International Conference on Medical image computing and computer-assisted intervention* , 234–241, Springer (2015).

[8] Bilic, P., Christ, P. F., Vorontsov, E., Chlebus, G., Chen, H., Dou, Q., Fu, C.-W., Han, X., Heng, P.-A., Hesser, J., et al., "The liver tumor segmentation benchmark (LiTS)," *arXiv preprint arXiv:1901.04056* (2019).

[9] Lim, H., Dewaraja, Y. K., and Fessler, J. A., "A PET reconstruction formulation that enforces non-negativity in projection space for bias reduction in Y-90 imaging," *Phys. Med. Biol.* **63**, 035042 (Feb. 2018).

[10] Paszke, A., Gross, S., Chintala, S., Chanan, G., Yang, E., DeVito, Z., Lin, Z., Desmaison, A., Antiga, L., and Lerer, A., "Automatic differentiation in PyTorch," (2017).

[11] Lim, H., Chun, I. Y., Dewaraja, Y. K., and Fessler, J. A., "Improved low-count quantitative PET reconstruction with a variational neural network," *arXiv preprint arXiv:1906.02327* (2019).