Report on "MDPET: A Unified Motion Correction and Denoising Adversarial Network for Low-Dose Gated PET" (Zhou et al., 2021)

Saiffedine Barkia SAIFEDDINE.BARKIA@TELECOM-PARIS.FR
Gonzalo Reinoso GREINOSO@UCM.ES

Abstract

Low-dose positron emission tomography (PET) is of great interest given that it reduces radiation exposition for patients and clinicians. However, this dose reduction increases image noise levels in gated images. In order to deal with motion correction and noise in gated images at once, MDPET, a unified motion correction and denoising adversarial network for generating high-quality images from low-dose gated PET is proposed. Appart from providing a unified pipeline, comparative studies show that MDPET is able to improve motion estimation and denoising of motion compensated images.

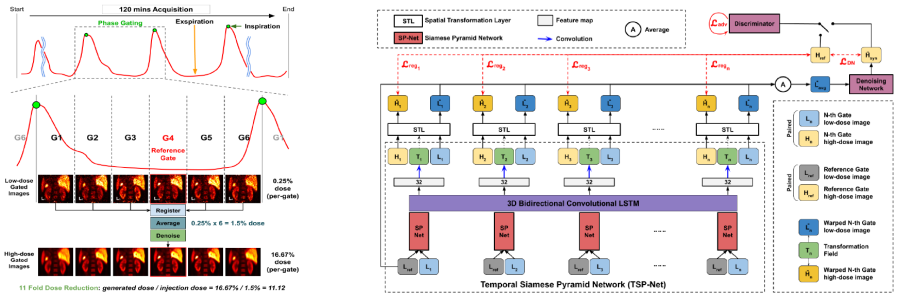
1. Background

Positron emission tomography (PET) is an imaging modality used in a wide range of medical fields that requires the injection of a radioactive tracer to visualize and measure changes in metabolic processes and in other physiological activities. Since it requires exposition of the patient to radiation, low-dose PETs are of great interest. Data acquisition takes between 10 and 20 minutes, making the reconstruction of the images challenging given respiratory motion. To address this issue, motion monitoring devices (Gucht et al., 2013b) are typically used to provide respiratory gating. However, even with these gated images, noise still poses an issue. Thus, the main problems to deal with are motion estimation and image denoising, which is aggravated by the reduction of the radioactive tracer dose.

Several works have attempted to tackle this issue using an initial image reconstruction for each gate followed by an image registration for motion estimation among different gates. Registration has been performed with conventional non-rigid registration algorithms (Pépin et al., 2014) (Pépin et al., 2014) (Chan et al., 2018) as well as with deep learning based methods (Li et al., 2020) (Balakrishnan et al., 2019). However, in these approaches, the noisy gated images could still lead to an inaccurate motion estimation, an issue aggravated by low-dose injections. Low-dose PET denoising efforts can be summarized into: conventional image post-processing (like (Maggioni et al., 2013) or (Maggioni et al., 2013)) and deep learning based methods like (Xiang et al., 2017) (Wang et al., 2018). Given the complex statistical characteristics of noise in medical imaging, deep learning has outperformed conventional methods. To date, deep learning low-dose PET denoising efforts amount to predicting standard-dose PET images from low-dose PET images using GANs (Goodfellow et al., 2014) (Ronneberger et al., 2015b) or other methods incorporating MR or CT scans (Xiang et al., 2017) (Chen et al., 2019) (Ronneberger et al., 2015a). However, the current state-of-the-art has not addressed motion estimation and denoising in low-dose respiratory gated PET.

1.1. Objectives and main contributions

The objective of this article is to propose MDPET: a **unified motion correction and denoising** adversarial network for generating motion-compensated low-noise images from low-dose gated PET data.



(a) (b)

Figure 1: Left: phase gated PET and the proposed method. The red curve guides the assignment of the detected events to each respiratory phase and generates 6 gated images. Right: overall structure of the MDPET.

The main contributions of the article with respect to the current state-of-the-art are: **a unified pipeline for denoising and motion compensation, improved denoising performance and improved motion compensation performance**.

2. Problem formulation

It is desirable to develop a method that directly estimates the motion from original low-dose gated images. Once an accurate motion estimation is obtained, the gated images can be registered to a reference low-dose gated image and averaged to generate a motion-compensated PET image, which deals with part of the noise. Then, the image can be introduced into a deep network for denoising. Figure 1a shows a representation of the gating procedure and the summarized pipeline performed by the proposed MDPET and figure 1b shows MDPET's overall structure. Its structure consists of a Temporal Siamese Pyramid motion estimation network (TSP-Net), a denoising network, and a discriminator. Each SP-Net predicts the transformation field between the source image and the reference image. After registering all the source low-dose gated images with the reference low-dose gated image, the average image is fed into the denoising network for generation of the final motion-compensated denoised PET image.

2.1. Mathematical formulation

By introducing some notations, the problem addressed by MDPET can be mathematically formulated in a simple manner. Let H_i, $L_i \in \mathbb{R}^{hxwxd}$ where i stands for the i-th of the 6 gates. Gate 4 is used as the reference gate, i.e. the gate with least intra-gate motion. Thus the following notation is adopted $H_4 = H_{ref}$ and $L_4 = L_{ref}$.

First, the objective is to obtain a motion-compensated low-dose average image out of the L_i with $i \in \{1, 2, 3, 4, 5, 6\}$. **Step 1**: a motion estimation model P_{TSP} parametrized by θ_{TSP} estimates T_i, the transformation fields between L_i and L_{ref} for $i \in \{1, 2, 3, 5, 6\}$, i.e., $T_1, ..., T_6 = P_{TSP}(L_1, ..., L_6; L_{ref}, \theta_{TSP})$. **Step 2**: a motion-compensated low-dose averaged image L_{avg} is obtained by transforming each L_i with the obtained T_i and averaging, i.e., $L_{avg} = \frac{1}{6}(L_{ref} + \sum_{n \neq ref} T_i \circ L_i)$.

Second, a denoising model P_{DN} parametrized by θ_{DN} is used to denoise L_{avg} and generate a high quality PET image, H_{syn}, i.e., $H_{syn} = P_{DN}(L_{avg}; \theta_{DN})$.

3. Methodology

In order to build high-quality motion-compensated low-noise reconstructions and generate accurate motion estimates straight from low-dose gated pictures, the authors propose an end-to-end architecture which is mainly divided into two parts. The motion estimation model P_{TSP} and the denoising model H_{syn}. The details of the architecture are illustrated in figure 2.

For the training phase, 6 gated images at the low-dose level and 6 gated images at the high-dose level of each of the patients are used. The aim is to generate, from the low-dose images, a smooth motion estimation and images like the ones obtained in the high-dose setting.

As stated in the previous section, for the motion estimation model, the first goal is to estimate a set of transformations T_n between all L_n and L_{ref} so that we can use them to generate L_{avg}. In order to further denoise L_{avg}, they pass it to the denoising model. Now, we will proceed to detail each part separately.

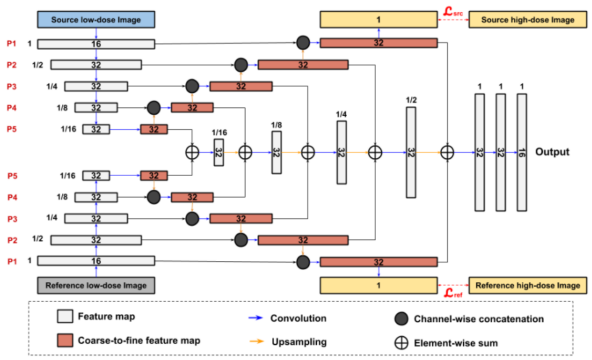


Figure 2: Design of the Siamese Pyramid Network

Motion Estimation Network:

For this part of the model, the authors built a Temporal Siamese Pyramid Network (TSP-Net) by merging many units (one for each gate) of Siamese Pyramid Network (SP-NET) and a Bidirectional Long-short Term memory (BiConvLSTM).

The SP-NET shares the same network parameters of the network with each other, they are meant to generate features so that in the next stage (along with BiConvLSTM) will be used to calculate the transformations between each low-dose gated image and the reference gate. Those features are generated using coarse-to-fine pyramid features from those pairs (L_{ref} and L_n).More precisely, the authors used 2 3D-UNet, as illustrated in figure 2, in each SP-Net to generate 5 levels of pyramid features to learn to generate those types of features along with the denoising purpose to obtain robust features representations. At the end of this architecture, the finest decoded features are fed into one 3D channel to generate an image with the same dimensions as the low source input (same as the high-level dose image) for both the reference and the gate image . Since the goal is to generate something that looks similar to the high-dose level images, in this part of the architecture, the authors regulated the output (symbolized by $\hat{H}$) with the MSE loss with respect to the high-dose level images: $L_{SP_n} = L_{ref} + L_{src_n} = \frac{1}{|H|}\sum_p [H_{ref}(p) - \hat{H_{ref}}(p)]^2 + \frac{1}{|H|}\sum_p [H_{src_n}(p) - \hat{H_{src_n}}(p)]^2$.

We find it necessary here to add a regularization (L2 for instance) in order to avoid any possible way of over-fitting. We think that adding this type of regularization would not harm the model, at the contrary, it will improve it, especially in certain critical situations.

As mentioned before, this part also does some denoising which is achieved by the UNet decoder since it reduces the noise in feature representation. All of this is done for both the reference and the gate image (at the low level), that's why two 3D UNets are used so at the end, the authors fused the coarse to fine pyramid features of both these parts and decoded them to generate features to predict the final transformations (which are 3 in the end according to the paper). To further improve the modelization of the transformations, the authors integrated the BiConvLSTM part to capture the adjacent and the non-adjacent SP-Net's features which can give more information such as the patterns in the respiratory cycle (feature patterns that are correlated over time in general). This is useful for the motion estimation step both in forward and backward temporal directions.

At the end, all these concatenated features (32 channels generated after each SP-Net + Bi-ConvLSTM) are fed to a convolutional block with an output of 3 channels to predict the final transformation T_n for each gate.

Those transformations, despite being computed only based on the low-level dose images, are also applied to the high-level dose images. This is controlled by the regression loss which is a sum of the MSE and, this time, they used a penalty term (smoothing) for the regularization.

$$L_{reg} = \sum_n L_{sim_n} + \lambda L_{smooth_n} = \sum_n (\frac{1}{|H|} \sum_p [H_{ref}(p) - [T_n \circ H_n](p)]^2 + \lambda \sum_p ||\nabla T_n(p)||^2)$$

The denoising network :

At the end of the previous TSP-Net, the L_{avg} image has been successfully computed. This formulation indicates a motion-compensated low-dose gated image. The authors believe that despite the fact that this formulation reduces significantly the noise, it is possible to denoise it even further by feeding this image into a denoising network. Once more, they used the UNet architecture for this model but they claim that any other denoising netwok should work fine too. This network is being trained based on two types of losses, the usual denoising loss: $L_{DN} = \frac{1}{|H|} \sum_p [H_{ref}(p) - G(L_{avg}(p)]^2$ and an adversarial loss (used with a discriminator) like the one we find in GANS to obtain realistic denoised images $L_{adv} = [D(H_{ref}) - 1]^2 + [D(G(L_{avg})]^2$. We find that this is the most important part of this network especially in medical imaging where we don't want to produce synthetic images.For that reason, the authors emphasized on that and compared the results when trained with and without the adversarial loss, which is the most important in this network.

Unlike other works where they train architectures separately, as mentioned before, the authors trained both these networks in an end-to-end fashion where the propagation of the gradients of the denoising part back-propagate until the motion estimation parts also.

So that at the end we have the following loss where the hyperparameters are empirically set here.:

$$L_{tot} = \lambda_{DN} L_{DN} + \lambda_{adv} L_{adv} + \lambda_{reg} L_{reg} + \lambda_{SP} \sum nL_{SP_n}$$

4. Validation

MDPET performance was evaluated on 28 pancreas studies where all the PET, high-dose and low-dose, were obtained with the same scanner. External respiratory motion was tracked using Aznai respiratory gating system (Gucht et al., 2013a).

The authors performed a four-fold cross validation on the MDPET performance where each fold consisted of 7 studies. During each validation, 21 studies were used for training and 7 studies were used for testing. The evaluation was performed on all 28 studies with 6 gated images in each study. We believe that this is a bad practice since they test performance on their training and validation data. They might have been motivated by data scarcity, however, they should have reserved an independent test set or evaluated performance as an average of their cross validations. One could argue that, since they use the same validation procedure to compare performance with other approaches, the results are still comparable, however, such a validation scheme could be suggesting an outstanding performance to a very overfitted model. Thus, we believe that a fair comparison should be done with a proper validation strategy.

The metric used for motion estimation evaluation was Normalized Mean Absolute Error (NMAE) between the reference high-dose gated image and the transformed high-dose gated images. The results were compared against VoxelMorph (VM) (Balakrishnan et al., 2019), Siamese Adversarial Network (SAN) (Zhou et al., 2020), and a non-deep learning based Non-Rigid B-spline Registration (NRB) (Papademetris et al., 2006). For denoising evaluation, they were computed Peak Signal-to-Noise Ratio (PSNR), Structural Similarity Index (SSIM), and NMAE between final synthetic

high-dose image and the reference high-dose gated image. First, TSP-Net was trained with $\lambda_{DN} = \lambda_{adv} = 0$. Then, the denoising network was trained using the predicted averaged images from the pre-trained TSP-Net and its denoising ground-truth. Finally, the pre-trained TSP-Net and denoising network were loaded into MDPET to train in an end-to-end fashion.

5. Results

5.1. Motion estimation

Qualitative and quantitative results show that MDPET can accurately predict and deform each low-dose gated image to the reference low-dose gated image (L_4). The use of the estimated deformations (T_n) significantly prevents misalignments in the gated images. The results of the proposed MDPET were compared with those of VoxelMorph (Balakrishnan et al., 2019), Non-Rigid B-spline Registration (Papademetris et al., 2006), and Siamese Adversarial Network (Zhou et al., 2020) obtaining superior motion estimation results by MDPET. Figure 3a summarizes these results. An ablation study in terms of motion estimation was conducted. Performance was checked first without Adversarial Learning and then without both BiConvLSTM and Adversarial Learning. Results were significantly better when incorporating all of them.

5.2. Denoising

After motion prediction, L_{avg} was inputted into the denoising network. Figure 3b shows a quantitative comparison of performance against other two-stage processing methods; NRB+UNet, VM+UNet, SAN+UNet, NRB+GAN, VM+GAN and SAN+GAN. MDPET motion estimation and denoising demonstrated a performance superior to the others methods tried with mean NMAE, SSIM, and PSNR. An ablation study was as well performed which showed that BiConvLSTM and adversarial learning further improves performance with regards to the baseline MDPET.

NMAE	Gate 1	Gate 2	Gate 3	Gate 4	Gate 5	Gate 6	GPU sec	CPU sec
w/o registration	0.1846	0.1066	0.0653	-	0.0659	0.1360	0	0
NRB [40]	0.1564↓*	0.1347↑*	0.1212↑*	-	0.1288↑*	0.1339↓†	-	1489
VM [9]	0.1362↓*	0.1202↑*	0.1126↑*	-	0.1144↑*	0.1232↓*	2.1	220
SAN [37]	0.1298↓*	0.0882↓*	0.0682↑†	-	0.0751↑*	0.1103↓*	4.3	423
Ours	0.1098↓*	0.0749↓*	0.0582↓*	-	0.0619↓*	0.0908↓*	0.54	59

(a)

Method–mean(std)	NMAE	SSIM	PSNR
✗REG+✗DN	.1712(.0225)	.9018(.0175)	25.87(1.87)
✓NRB+✗DN	.1174(.0198)	.9424(.0096)	28.97(1.79)
NRB+UNet	.1166(.0177)	.9479(.0068)	29.49(1.85)
NRB+GAN	.1147(.0179)	.9489(.0071)	29.66(1.91)
✓VM+✗DN	.1165(.0130)	.9431(.0080)	28.98(1.90)
VM+UNet	.1125(.0124)	.9480(.0052)	29.43(1.99)
VM+GAN	.1128(.0130)	.9490(.0061)	29.48(1.98)
✓SAN+✗DN	.1401(.0187)	.9191(.0154)	27.99(1.49)
SAN+UNet	.1062(.0122)	.9498(.0061)	30.31(1.87)
SAN+GAN	.1036(.0117)	.9503(.0061)	30.87(1.79)
✓Ours+✗DN	.1383(.0185)	.9193(.0153)	28.14(1.46)
Ours	.0883(.0133)	.9669(.0054)	32.28(1.89)

(b)

Figure 3: Left: comparison of different motion estimation methods for different gates. Right: comparison of denoising performance on different motion compensated images.

6. Conclusion

A unified motion correction and denoising network for low-dose PET images is proposed and shown superior in performance to alternative approaches. However, we believe that the validation procedure undermines the comparison with other approaches and thus, the results should be taken with a grain of salt. It is also worth mentioning that the denoising result is still worse than the one obtained in the high-dose setting. Further work could focus on the use of different state-of-the-art denoising and motion estimating sub-networks (Xiang et al., 2017) and the incorporation of perceptual loss (Hu et al., 2021).

References

Guha Balakrishnan, Amy Zhao, Mert R. Sabuncu, John Guttag, and Adrian V. Dalca. Voxelmorph: A learning framework for deformable medical image registration. *IEEE Transactions on Medical Imaging*, 38(8):1788–1800, 2019. doi: 10.1109/TMI.2019.2897538.

C. Chan, J. Onofrey, Y. Jian, M. Germino, X. Papademetris, R. E. Carson, and C. Liu. Non-Rigid Event-by-Event Continuous Respiratory Motion Compensated List-Mode Reconstruction for PET. *IEEE Trans Med Imaging*, 37(2):504–515, 02 2018.

Kevin T. Chen, Enhao Gong, Fabiola Bezerra de Carvalho Macruz, Junshen Xu, Athanasia Boumis, Mehdi Khalighi, Kathleen L. Poston, Sharon J. Sha, Michael D. Greicius, Elizabeth Mormino, John M. Pauly, Shyam Srinivas, and Greg Zaharchuk. Ultra–low-dose 18f-florbetaben amyloid pet imaging using deep learning with multi-contrast mri inputs. *Radiology*, 290(3):649–656, 2019. doi: 10.1148/radiol.2018180940. URL https://doi.org/10.1148/radiol.2018180940. PMID: 30526350.

Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial nets. In Z. Ghahramani, M. Welling, C. Cortes, N. Lawrence, and K. Q. Weinberger, editors, *Advances in Neural Information Processing Systems*, volume 27. Curran Associates, Inc., 2014. URL https://proceedings.neurips.cc/paper/2014/file/5ca3e9b122f61f8f06494c97b1afccf3-Paper.pdf.

Axel Gucht, Benjamin Serrano, Florent Hugonnet, Benoît Paulmier, Nicolas Garnier, and Marc Faraggi. Impact of a new respiratory amplitude-based gating technique in evaluation of upper abdominal pet lesions. *European journal of radiology*, 83, 11 2013a. doi: 10.1016/j.ejrad.2013.11.010.

Axel Gucht, Benjamin Serrano, Florent Hugonnet, Benoît Paulmier, Nicolas Garnier, and Marc Faraggi. Impact of a new respiratory amplitude-based gating technique in evaluation of upper abdominal pet lesions. *European journal of radiology*, 83, 11 2013b. doi: 10.1016/j.ejrad.2013.11.010.

Zhanli Hu, Hengzhi Xue, Qiyang Zhang, Juan Gao, Na Zhang, Sijuan Zou, Yueyang Teng, Xin Liu, Yongfeng Yang, Dong Liang, Xiaohua Zhu, and Hairong Zheng. Dpir-net: Direct pet image reconstruction based on the wasserstein generative adversarial network. *IEEE Transactions on Radiation and Plasma Medical Sciences*, 5(1):35–43, 2021. doi: 10.1109/TRPMS.2020.2995717.

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*, 65(15):155003, 07 2020.

M. Maggioni, V. Katkovnik, K. Egiazarian, and A. Foi. Nonlocal transform-domain filter for volumetric data denoising and reconstruction. *IEEE Trans Image Process*, 22(1):119–133, Jan 2013.

X. Papademetris, M. P. Jackowski, N. Rajeevan, M. DiStasio, H. Okuda, R. T. Constable, and L. H. Staib. BioImage Suite: An integrated medical image analysis suite: An update. *Insight J*, 2006: 209, 2006.

Audrey Pépin, Joël Daouk, Pascal Bailly, Sébastien Hapdey, and Marc-Etienne Meyer. Management of respiratory motion in pet/computed tomography: the state of the art. *Nuclear medicine communications*, 35(2):113–122, Feb 2014. ISSN 1473-5628. doi: 10.1097/MNM.0000000000000048. URL https://pubmed.ncbi.nlm.nih.gov/24352107. 24352107[pmid].

Audrey Pépin, Joël Daouk, Pascal Bailly, Sébastien Hapdey, and Marc-Etienne Meyer. Management of respiratory motion in pet/computed tomography: the state of the art. *Nuclear medicine communications*, 35(2):113—122, February 2014. ISSN 0143-3636. doi: 10.1097/mnm.0000000000000048. URL https://europepmc.org/articles/PMC3868022.

Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In Nassir Navab, Joachim Hornegger, William M. Wells, and Alejandro F. Frangi, editors, *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*, pages 234–241, Cham, 2015a. Springer International Publishing. ISBN 978-3-319-24574-4.

Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. *CoRR*, abs/1505.04597, 2015b. URL http://arxiv.org/abs/1505.04597.

Y. Wang, B. Yu, L. Wang, C. Zu, D. S. Lalush, W. Lin, X. Wu, J. Zhou, D. Shen, and L. Zhou. 3D conditional generative adversarial networks for high-quality PET image estimation at low dose. *Neuroimage*, 174:550–562, 07 2018.

L. Xiang, Y. Qiao, D. Nie, L. An, Q. Wang, and D. Shen. Deep Auto-context Convolutional Neural Networks for Standard-Dose PET Image Estimation from Low-Dose PET/MRI. *Neurocomputing*, 267:406–416, Dec 2017.

Bo Zhou, Yu-Jung Tsai, and Chi Liu. Simultaneous denoising and motion estimation for low-dose gated pet using a siamese adversarial network with gate-to-gate consistency learning, 09 2020.

Bo Zhou, Yu-Jung Tsai, Xiongchao Chen, James S. Duncan, and Chi Liu. Mdpet: A unified motion correction and denoising adversarial network for low-dose gated pet. *IEEE Transactions on Medical Imaging*, 40(11):3154–3164, 2021. doi: 10.1109/TMI.2021.3076191.