Inferring PET from MRI with pix2pix

Merel M. Jung[1,2], Bram van den Berg[2], Eric Postma[1,2], and Willem Huijbers[1,2]

[1] Tilburg University, Cognitive Science & Artificial Intelligence, Tilburg, NL
[2] Jheronimus Academy of Data Science, s'Hertogenbosch, NL w.huijbers@uvt.nl

Abstract. Medical image-to-image translation, using conditional Generative Adversarial Networks (cGANs), could be beneficial for clinical decisions when additional diagnostics scans are requested. The recently proposed pix2pix architecture provides an effective image-to-image translation method to study such medical use of cGANs. This study addresses the question to what extent pix2pix can translate a magnetic resonance imaging (MRI) scan of a patient into an estimate of a positron emission tomography (PET) scan of the same patient. We perform two image-to-image translation experiments using paired MRI and PET brain scans of Alzheimer's disease patients and healthy controls. In experiment 1, we train using data sliced in one dimension (the axial plane). In experiment 2, we train using augmented data sliced in all three dimensions (axial, sagittal and coronal). After training, the synthetically generated PET scans are compared to the actual ones. The results suggest that PET scans can be sufficiently and reliably estimated from MRI, with similar results using axial and augmented training. We conclude that image-to-image translation is a promising and potentially cost-saving method for making informed use of expensive diagnostic technology.

Keywords: Medical image analysis · Neuroimaging · Image synthesis · Generative adversarial networks · Deep neural networks · MRI2PET

1 Introduction

The soaring costs of health-care can be mitigated by an informed use of expensive diagnostic technology. In this paper, we focus on the use of magnetic resonance imaging (MRI) and positron emission tomography (PET) scans for health diagnostics. MRI and PET are versatile, but also expensive diagnostic tools, used for the diagnosis of cancer, cardiovascular and neurological disorders. The measurements obtained by MRI and PET provide different views on the inner brain processes and brain anatomy. Whereas MRI mostly uses the magnetic properties of tissue or blood to create various contrasts (e.g. T_1,T_2,T_2*-weighted images), PET uses radioactive-labeled tracers to bind to specific molecular targets (e.g. glucose, oxygen and amyloid-beta). MRI is also cheaper and far less invasive, compared to PET, since it does not depend on radiation.

The main question addressed in this study is: To what extent can a brain PET image be generated from the corresponding MRI image? If a PET scan -using fludeoxyglucose PET tracer-, can be reasonably approximated from its

corresponding MRI, it may be easier to determine, on the basis of the already acquired MRI scan, if performing an expensive PET-scan would be valuable or not.

We suspect that some degree of MRI to PET translation is feasible for the following reason. PET measures the brain's metabolism, whereas MRI measures the brain's anatomy. Our brain is always metabolically active, consuming approximately 20% of our daily energy budget [10] and these metabolic processes are closely linked to the anatomy of the brain. Hence, there is a relation between both measurements that may be exploited.

Machine learning methods, deep neural networks in particular, offer effective means to determine to what extent a PET image can be generated from its corresponding MRI image. Deep neural networks are revolutionizing the field of medical image analysis [8]. Prominent examples of successful applications are dermatologist-level classification of skin cancer [3] and tissue-based outcome prediction in colorectal cancer [2]. For our task, which involves translating an image in input format to an output format, the recently proposed pix2pix architecture [5] is highly appropriate. Simply training a deep neural network to translate MRI images to PET images is not feasible, because the L2 or L1 minimization is pixel-based which tends to favour averaged (i.e., blurry) output images [9].

Pix2pix is a general network for image-to-image translation which uses a conditional Generative Adversarial Network (cGAN) in which the minimization is image-based, rather than pixel-based (see, e.g., [5]). Conditional variants of Generative Adversarial Networks [4], simultaneously train a conditional generator and a discriminator. The generator is trained to generate images (in our case PET images) conditioned on input images (in our case the corresponding MRI images). The discriminator aims to classify whether the generated PET images are real or fake [5]. Before the introduction of pix2pix, the generation of synthetic PET images has been proposed as a solution for missing data in the context of brain-disease diagnosis [7]. Here we evaluate how well pix2pix can infer PET data from MRI in a mixed sample of healthy controls and Alzheimer's disease patients.

The outline of the remainder of the paper is as follows. First, we will describe our sample of MRI and PET scans and detail how these scans are preprocessed to transform the 3D volume-metric data into 2D slices, allowing the images to be used by the pix2pix architecture. We compare two methods for conversion. One sparse method, using data sliced in one dimension (axial data only), and a second method with data augmentation in three dimensions (i.e., axial, sagittal, and coronal). Next, we will show examples and evaluate the results of both the axial and the augmented MRI2PET models. Finally, we will discuss avenues to improve upon our results and the future application of MRI2PET.

2 Methods

This section describes the data, the preprocessing of the data, the pix2pix network, the postprocessing of the generated images, and the measures used to evaluate the results.

2.1 Data

We used data from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (adni.loni.usc.edu) [12]. We selected scans from 16 older adults (9 female) with an average age of 76.4 ($SD = 6.45$) who obtained their MRI and PET scans within one year, who where first enrolled in ADNI1. Half of the sample consisted of normal controls with a clinical dementia rating (CDR) sum of boxes of 0 ($SD = 0$) and an average mini mental state exam (MMSE) of 28.8 ($SD = 0.92$; $min = 28.0.$; $max = 30.0$). The other half consisted of Alzheimer's disease patients with an CDR sum of boxes of 4.1 ($SD = 1.74$; $min = 2.0$; $max = 7.0$) and an average MMSE of 21.9 ($SD = 2.23$; $min = 19.0$; $max = 26.0$).

From these older adults, we obtained the T1-weighted MRI scans, originally acquired with a 1.5 Tesla scanner at a resolution of $1.0\,mm^3$. We also obtained PET scans partially preprocessed by the ADNI consortium. The PET scans were acquired with a 185 MBq (5 mCi), dynamic 3D scan of six 5-min frames $30-60$ min post injection and a resolution of $2.0\,mm^3$. From the ADNI database, we obtained the coregistered average of these six frames that was re-sampled to a resolution of $1.5\,mm^3$.

2.2 Preprocessing

We preprocessed the MRI and PET images in five steps: registration, downsampling, slicing, image format conversion, and train-test set partitioning. *Registration:* We registered the MRI and PET scan using the McFLIRT software [6]. *Downsampling:* We down-sampled the MRI scans to the resolution of the PET scans, i.e., from $1.0\,mm^3$ to $1.5\,mm^3$, while retaining the larger field of view of the MRI. This resulted in 3D-volumes, for both MRI and PET, with the dimensions: $121 \times 145 \times 121$, where each voxel is $1.5\,mm^3$. *Slicing:* We sliced the registered MRI and PET volumes in three planes: sagittal, coronal and axial. *Image format conversion:* The sliced MRI and PET volumes were converted from the 3D NIFTI format to a 2D uncompressed JPEG format. *Train-test set partitioning:* The sliced data were divided per subject into a train (12 adults) and test set (4 adults) with an equal number of normal controls and Alzheimer's disease patients in each subset.

2.3 Pix2pix network

To generate the synthetic PET images we ran a Keras implementation of the pix2pix network with Tensorflow as backend[3]. We ran two experiments to evaluate the impact of a network trained on slices from a single axis compared to slices

[3] https://github.com/tjwei/GANotebooks/blob/master/pix2pix-keras.ipynb

from multiple axes on the quality of the generated images. First, we trained a pix2pix network using only the axial slices to translate from MRI to PET (i.e., 121 axial slices $\times 12$ adults = 1452 images in total). We labelled this the *axial MRI2PET model*. Second, we trained a pix2pix network with augmented data using slices in all three axes to translate MRI2PET (i.e., (121 sagittal +145 coronal +121 axial slices) $\times 12$ adults = 4644 images in total). We label this the *augmented MRI2PET model*. Both MRI2PET models were trained for 100 epochs.

2.4 Postprocessing

To compare the axial MRI2PET model and the augmented MRI2PET model, we generated synthetic PET slices from unseen MRI images in the test dataset (121 axial slices $\times 4$ adults = 484 images). Next, we reconstructed the 3D volumes from the axial slices generated by the two MRI2PET models. From the reconstructed volumes we extracted 121 sagittal images. These reconstructed volumes and sagittal images allow us to also evaluate MRI2PET translation of volumetric data.

2.5 Evaluation

Proper evaluation of the quality of generated images is an open problem (see, e.g., [5]). To evaluate both MRI2PET models, we calculated the mean squared error (MSE), the peak signal-to-noise ratio (PSNR) and the structural similarity (SSIM) index [11]. These measures quantify the difference between the synthetic PET and real PET scans for the two MRI2PET models. For the evaluation we used the middle slices of the PET scans generated with both models when conditioned with MRI images from the test set.

To obtain reference values, we defined baselines for our three measures (MSE, PSNR and SSIM). Therefore, we calculated the MSE, PSNR and SSIM by comparing each (axial) PET image, at the middle slice, to all other (axial) PET images in the training set, excluding the comparison to itself. See Table 1 for the baseline values of MSE, PSNR and SSIM.

3 Results

We present the results of the axial and augmented MRI2PET models in two parts. In the first part (3.1) we report the quantitative results obtained with both models. In the second part (3.2), the results are qualitatively assessed by visualization of axial and sagittal reconstructions.

3.1 Quantitative evaluation of synthetic PET images

Table 1 lists the results for the axial and augmented model, as quantified by the three evaluation metrics (MSE, PSNR, and SSIM). For both models and the

baseline, the mean, standard deviation (std) and range (min/max) are given. We observe that the mean MSE (smaller is better) is smaller than the mean of the baseline and that all PSNR and SSIM values (larger is better) are larger than their respective baseline values. The differences between the axial and augmented models are minor. Figure 1 displays the results of table 1 in box-whiskers plots.

Table 1. Comparison of the real and synthetic PET slices on different evaluation metrics for the axial and augmented MRI2PET models. Baseline values are provided for comparison.

	MSE			PSNR		
	axial	augmented	baseline	axial	augmented	baseline
mean	22.98	24.36	26.76	73.94	73.35	67.83
std	4.01	1.71	4.23	1.52	0.63	1.48
min	18.44	21.91	15.73	72.13	72.86	65.40
max	27.99	25.74	34.91	75.75	74.25	72.32

	SSIM		
	axial	augmented	baseline
mean	0.93	0.93	0.85
std	0.01	0.01	0.02
min	0.92	0.92	0.81
max	0.95	0.94	0.91

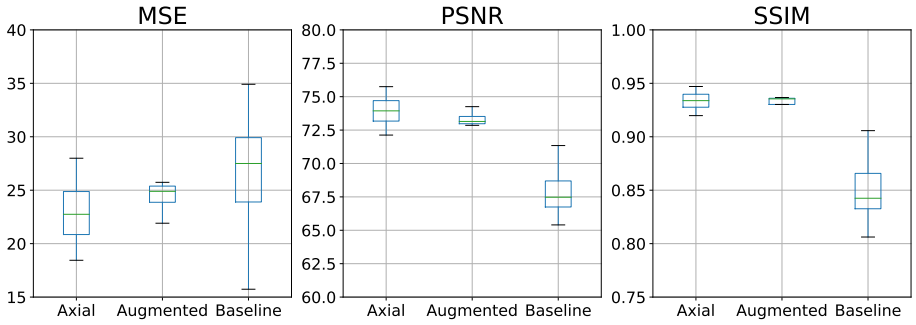

Fig. 1. Boxplots that compare the real and synthetic PET slices on three evaluation metrics (MSE, PSNR and SSIM) for the axial and augmented MRI2PET models. Baseline values are provided for comparison.

3.2 Qualitative evaluation of synthetic PET images

We illustrate the synthetic PET slices generated by both models in figure 2. The top row shows the results for the axial MRI2PET model, the bottom row those

for the augmented MRI2PET model. For each row, the MRI image that serves as input to the model is shown in the left panel (labelled MRI). The second panel from the left (real PET) shows the ground truth. The third panel (synthetic PET) displays the generated PET image. Finally, the fourth panel (synthetic-real PET (MAE)) shows the difference (mean absolute error (MAE)) between the generated and real PET. Figure 3 illustrates the results for the reconstructed sagittal view. Both figures illustrate the relatively small reconstruction errors as reflected in the similarities of the real and synthetic PET images. The differences in quality between the images generated by the axial and the augmented model are minor.

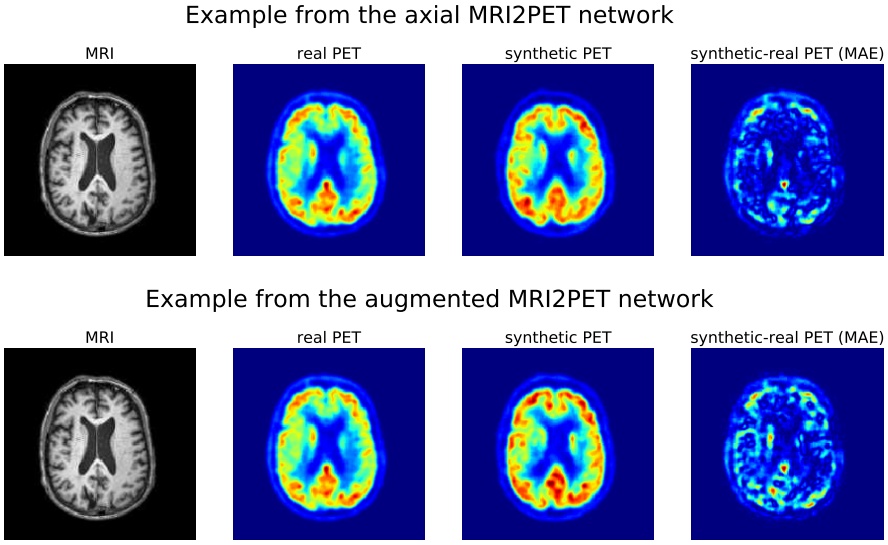

Fig. 2. Synthetic PET slice in the axial plane of the same subject generated by the axial MRI2PET model (above) and the augmented MRI2PET model (below). From left to right: T1 MRI, real PET, synthetic PET, mean absolute error between real and synthetic PET.

4 General Discussion

This paper demonstrates that inferring PET images from MRI images may be feasible, despite the fact that MRI and PET rely on very different techniques to image the brain. Our MRI2PET models exploit the shared structure between brain anatomy and brain metabolism, to generate PET images. The results of our evaluation indicate that the MRI2PET network trained on augmented data sliced in all three dimensions (i.e., axial, sagittal and coronal) compared to axial data only does not improve the quality of the generated PET images.

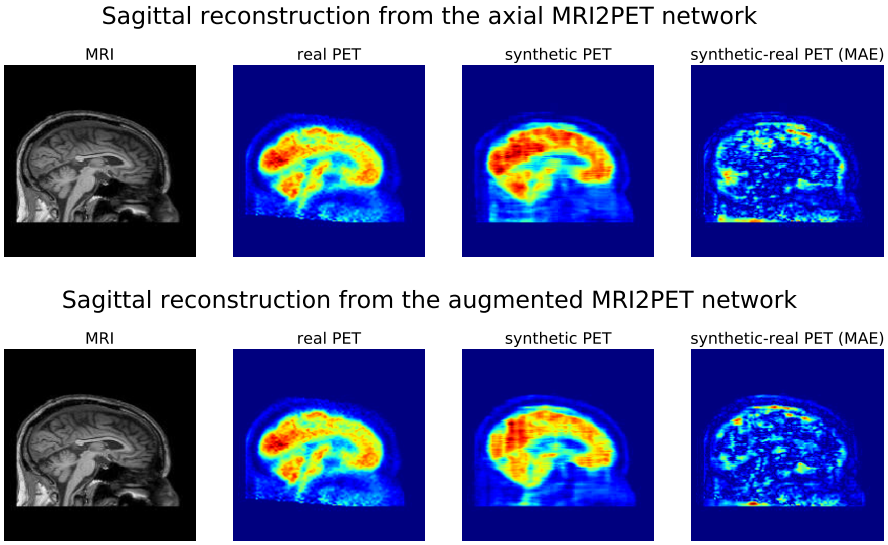

Fig. 3. Comparison of the same sagittal PET slices after reconstruction from 2D axial synthetic PET images into a 3D volume generated from either the axial and augmented MRI2PET network.

Three points of improvement of our current work are: (i) the evaluation of the generated images, (ii) the reconstruction of 3D volumes, and (iii) the limited size of the dataset. Below, we discuss each of these limitations and propose improvements. (i) As stated before, the evaluation of the quality of generated images is an unsolved problem. We evaluated the quality by using measures that are essentially pixel-based. There are two ways to improve the evaluation. The first is to rely on medical experts that score the generated PET images. The second is to use the generated images as input for a classification task. (ii) The MRI2PET models essentially use 2D convolution for a 3D problem. An obvious improvement would be to adapt the pix2pix method to incorporate 3D convolution filters, rather than 2D ones. However, the number of weight parameters that a 3D model will need to learn compared to a 2D model increases substantially. At the same time the number of examples decreases drastically when a model is trained on entire volumes instead of individual slices. As a result, the amount of training data needs to be expanded. Further research will be necessary to evaluate whether a 3D model will improve performance. (iii) The relatively modest size and homogeneity of the dataset may not reflect the true power of the MRI2PET approach. Although, we included normal controls and Alzheimer's disease patients, ADNI participants have been screened for other clinical abnormalities. As a result, there are no scans that include large regional abnormalities due to past strokes or brain tumors. Nevertheless, clinically MRI (and PET) scans are also used to diagnose cardiovascular disease and brain

cancer. Given a rich heterogeneous dataset to train a MRI2PET model, we might create synthetic PET scans that include brain tumors visible on both MRI and PET. These synthetic scans might inform the radiologist on where a tumor can be seen, prior to the acquisition of a PET scan. Further research will be necessary to determine if abnormal metabolism, due to brain tumors, can be generated via a MRI2PET model.

5 Conclusion and Future Work

The results of our experiments lead us to conclude that image-to-image translation is a promising and potentially cost saving method for making informed use of expensive diagnostic technology. In our future work, we intend to explore the aforementioned three improvements. In addition, we will include the recently-proposed end-to-end MedGAN image translation architecture [1] in our investigations.

Acknowledgements

The Titan Xp used for this research was donated by the NVIDIA Corporation to WH. Data collection and sharing for this project was funded by the Alzheimer's Disease Neuroimaging Initiative (ADNI) (National Institutes of Health Grant U01 AG024904) and DOD ADNI (Department of Defense award number W81X WH-12-2-0012). ADNI is funded by the National Institute on Aging, the National Institute of Biomedical Imaging and Bioengineering, and through contributions from pharmaceutical companies. ADNI data are disseminated by the Laboratory for Neuro Imaging at the University of Southern California.

References

1. Armanious, K., Yang, C., Fischer, M., Küstner, T., Nikolaou, K., Gatidis, S., Yang, B.: MedGAN: Medical Image Translation using GANs. arXiv.org (2018)
2. Bychkov, D., Linder, N., Turkki, R., Nordling, S., Kovanen, P.E., Verrill, C., Walliander, M., Lundin, M., Haglund, C., Lundin, J.: Deep learning based tissue analysis predicts outcome in colorectal cancer. Scientific Reports **8**(1), 3395 (2018). https://doi.org/10.1038/s41598-018-21758-3
3. Esteva, A., Kuprel, B., Novoa, R.A., Ko, J., Swetter, S.M., Blau, H.M., Thrun, S.: Dermatologist-level classification of skin cancer with deep neural networks. Nature **542**(7639), 115–118 (2017). https://doi.org/10.1038/nature21056
4. Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., Bengio, Y.: Generative adversarial nets. In: Advances in Neural Information Processing Systems. pp. 2672–2680 (2014)
5. Isola, P., Zhu, J.Y., Zhou, T., Efros, A.A.: Image-to-Image Translation with Conditional Adversarial Networks. arXiv.org (2016)
6. Jenkinson, M., Bannister, P., Brady, M., Smith, S.: Improved Optimization for the Robust and Accurate Linear Registration and Motion Correction of Brain Images. NeuroImage **17**(2), 825–841 (2002)

7. Li, R., Zhang, W., Suk, H.I., Wang, L., Li, J., Shen, D., Ji, S.: Deep learning based imaging data completion for improved brain disease diagnosis. In: International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI). pp. 305–312. Springer (2014)
8. Litjens, G., Kooi, T., Bejnordi, B.E., Setio, A.A.A., Ciompi, F., Ghafoorian, M., van der Laak, J.A., Van Ginneken, B., Sánchez, C.I.: A survey on deep learning in medical image analysis. Medical Image Analysis **42**, 60–88 (2017)
9. Pathak, D., Krahenbuhl, P., Donahue, J., Darrell, T., Efros, A.A.: Context encoders: Feature learning by inpainting. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). pp. 2536–2544 (2016)
10. Vaishnavi, S.N., Vlassenko, A.G., Rundle, M.M., Snyder, A.Z., Mintun, M.A., Raichle, M.E.: Regional aerobic glycolysis in the human brain. Proceedings of the National Academy of Sciences **107**(41), 17757–17762 (2010)
11. Wang, Z., Bovik, A.C., Sheikh, H.R., Simoncelli, E.P.: Image quality assessment: from error visibility to structural similarity. IEEE Transactions on Image Processing **13**(4), 600–612 (2004)
12. Weiner, M.W., Veitch, D.P., Aisen, P.S., Beckett, L.A., Cairns, N.J., Cedarbaum, J., Green, R.C., Harvey, D., Jack, C.R., Jagust, W., Luthman, J., Morris, J.C., Petersen, R.C., Saykin, A.J., Shen, L., Schwarz, A., Toga, A.W., Trojanowski, J.Q., Initiative, A.D.N.: 2014 Update of the Alzheimer's Disease Neuroimaging Initiative: A review of papers published since its inception. Alzheimer's & Dementia **11**(6), e1–e120 (2015)