Generation of Amyloid PET Images via Conditional Adversarial Training for Predicting Progression to Alzheimer's Disease

Yu Yan[1(✉)], Hoileong Lee[1], Edward Somer[2], and Vicente Grau[1]

[1] Institute of Biomedical Engineering, Oxford University, Oxford, UK
yu.yan@linacre.ox.ac.uk
[2] GE Healthcare, London, UK

Abstract. New positron emission tomography (PET) tracers could have a substantial impact on early diagnosis of Alzheimer's disease (AD) and mild cognitive impairment (MCI) progression, particularly if they are accompanied by optimised deep learning methods. To realize the full potential of deep learning for PET imaging, large datasets are required for training. However, dataset sizes are restricted due to limited availability. Meanwhile, most of the AD classification studies have been based on structural MRI rather than PET. In this paper, we propose a novel application of conditional Generative Adversarial Networks (cGANs) to the generation of ^{18}F-florbetapir PET images from corresponding MRI images. Furthermore, we show that generated PET images can be used for synthetic data augmentation, and improve the performance of 3D Convolutional Neural Networks (CNN) for predicting progression to AD. Our method is applied to a dataset of 79 PET images, obtained from Alzheimer's Disease Neuroimaging Initiative (ADNI) database. We generate high quality PET images from corresponding MRIs using cGANs, and we evaluate the quality of generated PET images by comparison to real images. We then use the trained cGANs to generate synthetic PET images from additional MRI dataset. Finally we build a 152-layer ResNet to compare the MCI classification performance using both traditional data augmentation method and our proposed synthetic data augmentation method. Mean Structural Similarity (SSIM) index was 0.95 ± 0.05 for generated PET and real PET. For MCI progression classification, the traditional data augmentation method showed 75% accuracy while the synthetic data augmentation improved this to 82%.

Keywords: Alzheimer's disease · Mild cognitive impairment
Data augmentation · Generative adversarial network · Deep learning
PET · MRI

1 Introduction

In recent years, amyloid positron emission tomography (PET) imaging has been applied in some medical imaging problems such as Alzheimer's disease classification and detection of amyloid plaques [1, 2]. The first PET tracer used to image β-amyloid

plaques was ^{11}C-Pittsburgh-Compound-B (PiB) [3]. Due to the limited availability of ^{11}C-PiB with its short half-life, ^{18}F-labelled alternatives have been developed, which allow off-site production and regional distribution. ^{18}F-flutemetamol, ^{18}F-florbetapir and ^{18}F-florbetaben have recently been approved by the US Food and Drug Administration (FDA) for clinical use. Abnormal uptake in grey matter causes a disruption of the characteristic white matter pattern caused by non-specific white matter binding [4]. These scans are generally interpreted visually.

A separate group from healthy volunteers (HV) and patients with probable Alzheimer's disease (pAD), mild cognitive impairment (MCI) is an intermediate cognitive state between normal aging and dementia. Subjects with MCI, especially MCI involving memory problems, are more likely to develop AD and other dementias [5]. According to this progression, MCI subjects can subsequently be classified as progressive MCI (pMCI) or stable MCI (sMCI) [6].

Many deep learning methods have been proposed to classify different AD stages based on high dimensional features extracted from various neuroimaging biomarkers. Meanwhile, the focus for AD classification has gradually evolved from classification between healthy control and disease patients to classification between pMCI and sMCI. In a recent paper on MCI classification, Kim et al. developed a deep learning-based method for classifying tau-pet imaging patterns. MCI subjects were split into three subgroups with the Louvain method. This method discriminated subgroups 1 and 2 with accuracy 90.91%, and 80.49% for subgroups 2 and 3 [7].

A big challenge in the medical imaging field is how to cope with small datasets and limited amount of annotated samples [8]. One promising solution inspired by game theory for image synthesis is known as Generative Adversarial Networks (GANs) [9]. The method is based on the idea of training two networks, a generator and a discriminator simultaneously with competing losses. In the past few years, different variations of GANs have been applied to generate realistic natural images, and recently, the popularity of using GANs to generate medical images have also increased [10]. For example, Frid et al. [11] proposed a CNN based classification framework to classify different CT images, where GANs was used to generate high quality 2D liver lesion ROIs from a vector of 100 random numbers. The classification performance using only traditional data augmentation yielded 78.6% sensitivity and 88.4% specificity. By adding the synthetic data augmentation the results increased to 85.7% sensitivity and 92.4% specificity. Recently, Madani et al. [12] used a GAN to generate the 2D chest X-ray images from random noise, and the generated data were subsequently used to train a CNN to classify images for cardiovascular abnormalities.

In general, the availability of MRI is much higher than PET for a number of reasons. PET scanners are expensive to buy and operate, and thus less common. PET scans require subjects are exposed to ionising radiation during the test. More importantly, the number of test datasets is very limited when newly developed PET tracers are being tested. In this case, we aim at compensating this imbalance between available MRI and PET images by using a limited dataset.

To the best of our knowledge, direct generation of 3D amyloid PET imaging from structural MR has not yet been attempted. In this study, we focus on the application of conditional GANs to generate high quality volumetric florbetapir PET images from

corresponding MRI images. In this way, we expect the natural variability in MRI scans and the image characteristics in PET to be combined. We also build a 152-layer ResNet classification model to distinguish pMCI and sMCI subjects, and quantify the difference in performance caused by the addition of this synthetic datasets in training. The summary of data generation model in this work is shown in Fig. 1.
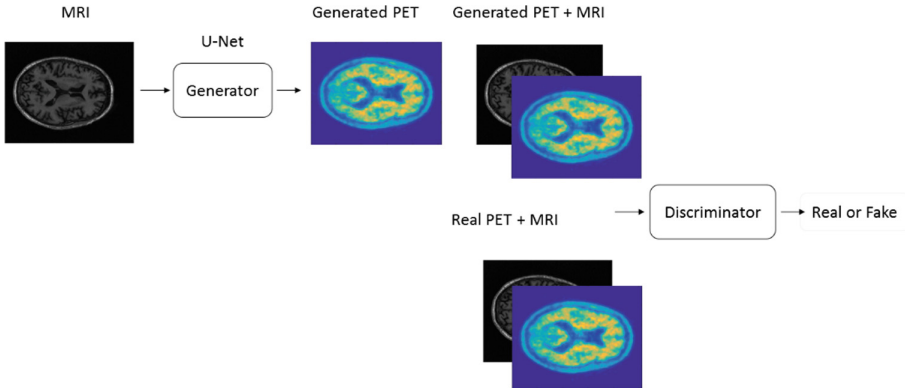

Fig. 1. Summary of data generation procedure using a conditional GANs

2 Materials and Methods

2.1 Data and Pre-processing

All image data were acquired from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (adni.loni.usc.edu). ADNI aims to improve clinical trials for the prevention and treatment of Alzheimer's disease. To date, over 1000 scientific publications have used ADNI data. ADNI has been running since 2004 and is currently funded until 2021.

In this study, 50 sMCI and 29 pMCI florbetapir images with corresponding T1 MRI were obtained from the ADNI database (set A). A second group of 29 T1 MRI images (21 of them with corresponding PET) from a different pMCI group were also downloaded (set B) and used independently. More details about the use of these datasets in training/validating/testing the cGANs and ResNet are provided in the relevant sections below. All the florbetapir images were pre-processed: MRI and PET scans from each subject were co-registered, and the PET scan was then reoriented into a standard $160 \times 160 \times 96$ voxel image grid, comprising 1.5 mm cubic voxels. This image grid was oriented such that the anterior-posterior axis of the subject is parallel to the AC-PC line. The MRI images have dimensions $256 \times 256 \times 196$ with a voxel size of 1 mm $\times$ 1 mm $\times$ 1.2 mm.

2.2   Amyloid PET Generation with Conditional Adversarial Training

GANs are generative networks that learn a mapping from random noise to output image. They are composed of two networks, a generator and a discriminator, trained in an adversarial way. The goal of the generator is to generate synthetic images, while the discriminator, evaluates them for authenticity. In conditional GANs, the generator learns a mapping between an input and an output image [13]. In this study, the generator is a U-Net based convolutional neural network with skip connections [14]. The discriminator is a convolutional Markovian discriminator (PatchGAN), which only penalizes structure at the scale of image patches. During the GANs training process, the generated PET was paired with the corresponding MRI and entered into the discriminator. The loss function of the conditional GANs is:

$$L_{cGAN}(G, D) = E_{x,y}[logD(x, y)] + E_x[\log(1 - D(x, G(x)))] \tag{1}$$

where x are MRI images and y are PET images. The first term is maximized when $D(x, y) = 1$, and the second is maximized when the $D(x, G(x)) = 0$, while it is minimised when $D(x, G(x)) = 1$, i.e. discriminator is not able to distinguish the generated images and real images. The generator G tries to minimize this objective against an adversarial discriminator D that tries to maximize it. In addition, conditional GANs also add an L1 loss term:

$$L_{L1}(G) = E_{x,y}(\|y - G(x)\|) \tag{2}$$

Therefore the complete form of loss function is:

$$L_{total}(G, D) = L_{cGAN}(G, D) + \varepsilon L_{L1}(G) \tag{3}$$

where ε is used to adjust the contribution of L1 loss, and it is set to 100 in the experiments reported here.

In order to measure the similarity between generated PET and real PET, We used SSIM due to its combination of errors in image contrast and overall structure [15, 16]. The structural similarity index (SSIM) was calculated as:

$$SSIM(x, y) = \frac{(2\mu_x\mu_y + C_1)(2\sigma_{xy} + C_2)}{(\mu_x^2 + \mu_y^2 + C_1)(\sigma_x^2 + \sigma_y^2 + C_2)} \tag{4}$$

where μ_x, μ_y, σ_x, σ_y, σ_{xy} are the local means, standard deviations, and cross-covariance for images x, y. C_1 and C_2 are regularization constants determined by pixel value range. The SSIM = 1 meaning that the two images are identical.

In this work, we used the 29 paired PET and MRI images from the pMCI group in set A to train the conditional GANs, and subsequently apply the mapping to the unseen 29 MRI images in set B to generate 29 synthetic PET images, thus doubling the size of

the pMCI dataset. In set B, 21 subjects had available PET scans, which we used to test the cGANs by calculating SSIM values. For the implementation of the cGANs architecture we used the Keras framework. The experiment was conducted on computer cluster equipped with NVIDIA GeForce GTX 1080 Ti GPU.

2.3   MCI Progression Classification Architecture Using 3D ResNet

Deep Residual Network (ResNet) [17] is arguably one of the most important developments in the deep learning area in the last few years. ResNet makes it possible to train up to thousands of layers and still achieves competitive performance with fast convergence. The core concept of ResNet is introducing an identity shortcut connection that skips one or more layers.

ResNet have been used successfully for 3D image segmentation as in VoxResNet, where the authors use identity mappings as skip connections [18]. In our work, the ResNet architecture was modified based on the identity mappings version [19] that refines the residual block with a pre-activation variant.

The main difference between our network and the identity mappings version is the number of dimensions of convolutional kernels and pooling. Our ResNet architecture has 152 layers containing 50 3-layer blocks. The three layers are $1\times1\times1$, $3\times3\times3$, $1\times1\times1$ convolutions, where the $1\times1\times1$ layers are responsible for reducing and then increasing dimensions, leaving the $3\times3\times3$ layer a bottleneck with smaller input and output dimensions, as detailed in Table 1. Down-sampling is performed by conv3_1, conv4_1, conv5_1 with a stride of 2.

Table 1. 152-ResNet architecture for pMCI and sMCI classification

Layer name	152-layer
Conv1	$7 \times 7 \times 7$, 64, stride 2
Conv2_x	$3 \times 3 \times 3$ max pool, stride 2
	$\begin{bmatrix} 1 \times 1 \times 1 & 64 \\ 3 \times 3 \times 3 & 64 \\ 1 \times 1 \times 1 & 256 \end{bmatrix} \times 3$
Conv3_x	$\begin{bmatrix} 1 \times 1 \times 1 & 128 \\ 3 \times 3 \times 3 & 128 \\ 1 \times 1 \times 1 & 512 \end{bmatrix} \times 8$
Conv4_x	$\begin{bmatrix} 1 \times 1 \times 1 & 256 \\ 3 \times 3 \times 3 & 256 \\ 1 \times 1 \times 1 & 1024 \end{bmatrix} \times 36$
Conv5_x	$\begin{bmatrix} 1 \times 1 \times 1 & 512 \\ 3 \times 3 \times 3 & 512 \\ 1 \times 1 \times 1 & 2048 \end{bmatrix} \times 3$
	Average pool, 2-d fc, softmax

We trained our classification model using 50 sMCI and 29 pMCI real florbetapir images from set A. A 10-fold cross-validation was applied to the whole dataset. We tested three scenarios. In the first one only real images were used with no

augmentation. The second one included traditional augmentation, which was done at each epoch. Specifically, the random rotation range is set to 20°, and images will be randomly flipped horizontally and vertically. These two experiments used 65 samples for training, 6 for validation, and 8 for testing. The third experiment used our cGANs augmented dataset, including the additional 29 PET images generated from the MRI scans in set B, resulting in 89 samples for training, 8 for validation, and 11 for testing. For training we used a batch size of 1 with a learning rate of 0.0001 for 100 epochs. We used Keras to implement our MCI classification framework. The experiment was performed on computer cluster with NVIDIA GeForce GTX 1080 Ti GPU.

3   Results and Discussion

3.1   Data Generation

In this study, we used volumetric MRI images to generate 3D PET images to enlarge pMCI group. Examples of real and generated PET images, with their corresponding SSIM values, are shown in Fig. 2. As can be seen from Fig. 2, the generated PET and real PET contain similar signal patterns. The mean SSIM obtained was 0.95 ± 0.05. Figure 3 shows generated PET images obtained from MRI scans for which the corresponding real PET was not available.

3.2   Classification Results

Classification results for pMCI against sMCI using 152-ResNet are shown in Table 2. We computed both the area under the receiver operating characteristic curve (AUC) and the accuracy (ACC). Three different cases are compared: classification with a network trained using only the real images, with no augmentation (top row); with a network trained using traditional augmentation (middle row) and using our synthetic images based augmentation method (bottom row).

Table 2. Classification ROC AUC and accuracy (mean $\pm$ std) with 152-ResNet

	AUC	ACC
sMCI vs pMCI (real images)	0.71 ± 0.08	0.63 ± 0.11
sMCI vs pMCI (traditional augmentation)	0.77 ± 0.11	0.75 ± 0.09
sMCI vs pMCI (real + generated images)	$\mathbf{0.81 \pm 0.07}$	$\mathbf{0.82 \pm 0.12}$

As can be seen from Table 2, the classification score for sMCI against pMCI using real PET images achieved accuracy 0.63, and with the aid of traditional data augmentation, the accuracy raised to 0.75. As we expected, the highest accuracy was obtained by using our proposed synthetic augmentation method, achieving an improvement of 7% over the traditional augmentation.
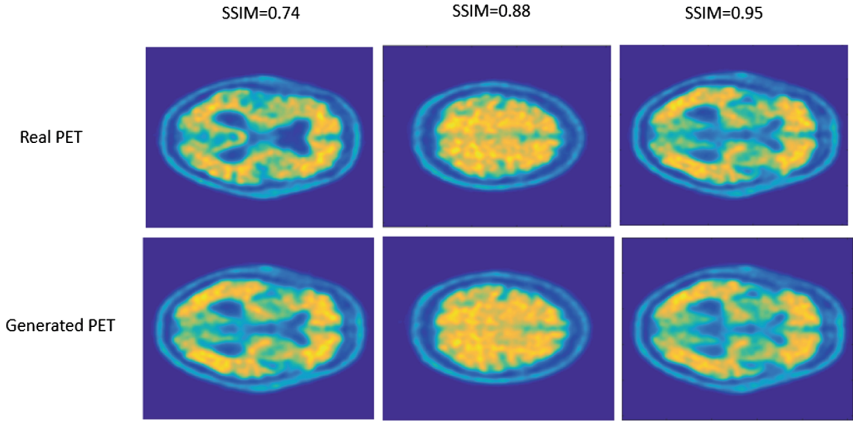
Fig. 2. Examples of real PET and generated PET images presented in different axial slices with SSIM score
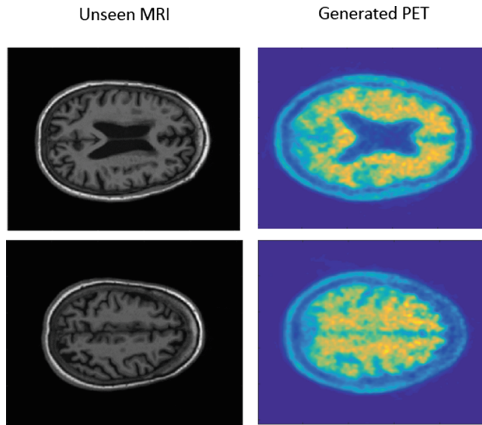
Fig. 3. Examples of unseen MRI and corresponding generated PET, corresponding real PET is not available

4 Conclusion

We developed a model for generating florbetapir PET from structural MR using deep generative networks, with generated data showing a high similarity to real corresponding PET. The generated data were then used for data augmentation for MCI classification on a limited dataset. We compared the synthetic augmentation method with a traditional augmentation method, and the synthetic augmentation outperformed the traditional augmentation. Future work will focus on using multi-modality imaging biomarkers for CNN classification.

References

1. Cattell, L., et al.: Classification of amyloid status using machine learning with histograms of oriented 3D gradients. NeuroImage Clin. **12**, 990–1003 (2016)
2. Nordberg, A.: PET imaging of amyloid in Alzheimer's disease. Lancet Neurol. **3**(9), 519–527 (2004)
3. Klunk, W.E., et al.: Imaging brain amyloid in Alzheimer's disease with Pittsburgh compound-B. Ann. Neurol. **55**(3), 306–319 (2004)
4. Knesaurek, K., et al.: Comparison of SUVr calculations in amyloid PET brain imaging. J. Nuclear Med. **55**, 1873 (2014)
5. Villemagne, V.L., et al.: Amyloid β deposition, neurodegeneration, and cognitive decline in sporadic Alzheimer's disease: a prospective cohort study. Lancet Neurol. **12**(4), 357–367 (2013)
6. Thurfjell, L., et al.: Combination of biomarkers: PET 18F flutemetamol imaging and structural MRI in dementia and mild cognitive impairment. Neurodegenerative Diseases **10**(1–4), 246–249 (2012)
7. Kim, J., et al.: Mild Cognitive impairment classification using deep learning. Alzheimer's Dement. J. Alzheimer's Assoc. **13**(7), 1070–1071 (2017)
8. Greenspan, H., et al.: Guest editorial deep learning in medical imaging: overview and future promise of an exciting new technique. IEEE Trans. Med. Imaging **35**(5), 1153–1159 (2016)
9. Goodfellow, I., et al.: Generative adversarial nets. In: Advances in Neural Information Processing Systems (2014)
10. Odena, A., et al.: Conditional image synthesis with auxiliary classifier GANs. arXiv (2016)
11. Frid, M., et al.: GAN-based synthetic medical image augmentation for increased CNN performance in liver lesion classification. arXiv (2018)
12. Madani, A., et al.: Chest x-ray generation and data augmentation for cardiovascular abnormality classification. International Society for Optics and Photonics (2018)
13. Isola, P., et al.: Image-to-image translation with conditional adversarial networks. arXiv (2017)
14. Ronneberger, O., et al.: U-net: convolutional networks for biomedical image segmentation. In: International Conference on Medical Image Computing and Computer-Assisted Intervention (2015)
15. Wang, Z., et al.: Image quality assessment: from error visibility to structural similarity. IEEE Trans. Image Process. **13**(4), 600–612 (2004)
16. Choi, H., Dong, S.L.: Generation of structural MR images from amyloid PET: application to MR-less quantification. J. Nucl. Med. **59**(7), 1111–1117 (2017)
17. He, K., et al.: Deep residual learning for image recognition. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (2016)
18. Chen, H., et al.: Voxresnet: deep voxelwise residual networks for volumetric brain segmentation. arXiv (2016)
19. He, K., et al.: Identity mappings in deep residual networks. In: European Conference on Computer Vision (2016)