Generation of structural MR images from amyloid PET: Application to MR-less quantification

[Running Title] Structural MR generation

Hongyoon Choi[1], Dong Soo Lee[1,2,3], for the Alzheimer's Disease Neuroimaging Initiative[*]

[1]*Department of Nuclear Medicine, Seoul National University College of Medicine, Seoul, Republic of Korea;*

[2]*Department of Molecular Medicine and Biopharmaceutical Sciences, Graduate School of Convergence Science and Technology, Seoul National University, Seoul, Republic of Korea*; [3]*Korea Brain Research Institute, Daegu, Korea*

[Correspondence and Reprint Request]

Dong Soo Lee, MD.,Ph.D.

Department of Nuclear Medicine, Seoul National University Hospital

28 Yongon-Dong, Jongno-Gu, Seoul, 110-744, Korea

Tel: 82-2-2072-2501, Fax: 82-2-2072-7690, E-mail: dsl@plaza.snu.ac.kr

[First author]

Hongyoon Choi, MD.,Ph.D.

Department of Nuclear Medicine, Seoul National University Hospital

28 Yongon-Dong, Jongno-Gu, Seoul, 110-744, Korea

E-mail: chy1000@snu.ac.kr, Tel: +822-2072-2802, Fax: +822-745-0345

[Word counts]   4984 words

Abstract

Structural magnetic resonance (MR) images concomitantly acquired with PET images can provide crucial anatomical information for precise quantitative analysis. However, in the clinical setting, not all the subjects have corresponding MR. Here, we developed a model to generate structural MR images from amyloid PET using deep generative networks. We applied our model to quantification of cortical amyloid load without structural MR.

Methods We used florbetapir PET and structural MR data of Alzheimer's Disease Neuroimaging Initiative database. The generative network was trained to generate realistic structural MR images from florbetapir PET images. After the training, the model was applied to the quantification of cortical amyloid load. PET images were spatially normalized to the template space using the generated MR and then standardized uptake value ratio (SUVR) of the target regions was measured by predefined regions-of-interests. A real MR-based quantification was used as the gold standard to measure the accuracy of our approach. Other MR-less methods, a normal PET template-based, multi-atlas PET template-based and PET segmentation-based normalization/quantification methods, were also tested. We compared performance of quantification methods using generated MR with that of MR-based and MR-less quantification methods.

Results Generated MR images from florbetapir PET showed visually similar signal patterns to the real MR. The structural similarity index between real and generated MR was 0.91 ± 0.04. Mean absolute error of SUVR of cortical composite regions estimated by the generated MR-based method was 0.04±0.03, which was significantly smaller than other MR-less methods (0.29±0.12 for the normal PET-template, 0.12±0.07 for multiatlas PET-template and 0.08±0.06 for PET segmentation-based methods). Bland-Altman plots revealed that the generated MR-based SUVR quantification was the closest to the SUVR values estimated by the real MR-based method.

Conclusion Structural MR images were successfully generated from amyloid PET images using deep generative networks. Generated MR images could be used as template for accurate and precise amyloid quantification. This generative method might be used to generate multimodal images of various organs for further quantitative analyses.

INTRODUCTION

Anatomical information of structural MR can help quantitative analysis of PET as well as guide detailed anatomical structures for functional images (*1,2*). Despite this usefulness, not all subjects have both PET and MR images in the clinical setting. For instance, either amyloid PET or structural MR could be acquired for subjects suspected to have cognitive decline in the clinic. In terms of quantitative analysis of PET, lack of structural information limits segmentation and accurate quantification.

A number of imaging biomarkers were developed to characterize Alzheimer's disease (AD) or to predict cognitive decline in mild cognitive impairment (MCI). In particular, on amyloid PET such as ^{11}C-Pittsburgh B and ^{18}F-florbetapir PET, a core pathologic marker, cortical amyloid deposition, can be quantitatively measured (*3,4*). Cortical amyloid load has been commonly quantified using standardized uptake value ratio (SUVR) between target and reference regions (*3,5*). In this quantification endeavor, accurate region segmentation is crucial for these regions, and structural MR has been used for the delineation of these regions (*6-8*). Typically, these methods have used MR for spatial normalization to template space and regional segmentation. The MR-based processing can be applied to the amyloid PET to measure the target and reference uptake of the radiotracers. Instead of structural MR, PET-template without MR has been attempted by direct spatial normalization to the PET template (*9*), however, this simpler trial resulted in biased estimation of SUVR. To overcome this bias, multiple PET templates-based processing was developed for MR-less quantification. It chose a PET template the most similar with the subject's image for normalization (*10,11*).

Until now, direct generation of structural MR from PET imaging has not yet been attempted. Since PET images have relatively low spatial resolution compared with structural MR, it has been challenging to directly generate mapping from PET to MR. In this study, we developed a model for generating structural MR images from amyloid PET. This model was trained by pairs of amyloid PET and MR images of AD, MCI patients and normal controls. We applied this model to quantification of cortical amyloid load without structural MR and compared this method with the MR-based quantification as a gold standard. Furthermore, other MR-less quantification methods were also compared.

MATERIALS AND METHODS

Patient Population

In this study, image data were collected from Alzheimer's Disease Neuroimaging Initiative (ADNI) (http://adni.loni.usc.edu) database. The ADNI was launched in 2003 as a public-private partnership, led by Principal Investigator Michael W. Weiner, MD, VA Medical Center and University of California San Francisco. ADNI included subjects from over 50 sites across the US and Canada. The primary goal of ADNI has been to develop combined biomarkers by testing whether serial imaging and biological markers, and clinical and neuropsychological assessment can be combined to measure the progression of mild cognitive impairment and early Alzheimer's disease. For up-to-date information, see http://www.adni-info.org.

261 subjects who underwent both florbetapir PET and structural MR as baseline studies were included for this study. Test set data were independent from the training data to develop and validate the model that generally compatible with various scanners at different sites. Thus, we divided the subsets according to image acquisition sites. Image data of training set were obtained from 10 sites and those of test set were obtained from 8 sites different from the training set. 163 pairs of PET and MR images were used for the training and 98 pairs of the images were used for the independent test set. The institutional review boards of all participating institutions approved imaging studies and all participants signed a written informed consent. Demographics and baseline clinical diagnosis of the subjects were summarized in Table 1.

Image Preprocessing for the Generative Model

Florbetapir PET images were downloaded with minimally preprocessed data to develop the generative model compatible with the image data acquired from various sites. The raw image data consisted of four 5-min frames 50–70 min after injection. We used the data of averaged images of all the 4 registered frames (*12*). Structural T1 images acquired concurrently with the baseline florbetapir images were used. High resolution 3D-T1 images were acquired in the sagittal plane.

Because the images were acquired from different sites, they have different matrix sizes as well as orientations. Voxel size of axial slices were ranged from 1.02×1.02 mm to 2.57×2.57 mm and slice thickness was ranged from 1.02 to 4.25 mm. Preprocessing was performed to fit the voxel size for the generative model, thus, the model used

minimally processed PET images of different scanners without normalization. To train the generative model, PET images were coregistered to corresponding MR images using rigid transformation under statistical parametric mapping (SPM8, www.fil.ion.ucl.ac.uk/spm). For PET images of test set, coregistration was not performed. We resliced the images to have the same voxel and matrix sizes. Voxel size of the resliced PET and MR images was $1.2 \times 1.0 \times 1.0$ mm^3. Matrix size of an axial slice of resliced images was 192×256. For the training data, axial slices of MR images out of range of PET scans were replaced by zero. After the reslicing, voxel values of PET and MR volumes of each subject were adjusted in the range of -1 to 1. Thus, the input image of the generative model was resliced PET images which have different orientations but same matrix size.

Adversarial Training for MR Image Generative Model

The MR generation model was based on image-to-image translation using generative adversarial network (GAN) model (*13*). The model has two convolutional neural networks, generator and discriminator (Fig. 1). Generator is trained to translate PET to MR images which cannot be discriminated from real MR. Discriminator is trained to discriminate real MR from generated MR by the generator. This adversarial training results in realistic MR image generation. Axial slices of coregistered PET and MR images were used for the training. Total axial slices for the training process were 32659 images. For the independent test and application to the quantitation of amyloid load, MR images of test set which included 98 subjects recruited from 8 sites different from training data were generated by corresponding PET images. Detailed methods and architectures of neural networks are described in **Supplementary Materials, Supplementary Table 1 and 2**.

Quantification of Amyloid Load

Cortical amyloid load was quantified for normalized florbetapir PET images of the test set. We compared the quantification results according to the different methods (Fig. 2). As a standard method, MR of each subject was segmented into gray matter, white matter and cerebrospinal fluids after image-intensity nonuniformity correction, and then nonlinear transformation parameters were calculated between the tissues of native space and the Montreal Neurological Institute. The transformation was applied to the corresponding PET. We applied the same normalization methods for generated MR.

As MR-less quantification methods, PET-template based normalization, using a normal PET template or multi-

atlas PET template, was also performed. Normal subjects' PET templates was generated by MR-based normalized PET images of normal controls. All subjects' PET images of native spaces were spatially normalized into the PET template. As a modified MR-less quantification methods, multi-atlas normalization, was adopted as a web-based tool, Capaibl (https://capaibl-milxcloud.csiro.au/) (*11,14*). Briefly, this method chose the best template among multiple PET templates by calculating similarity, and then a subject's PET image was registered to the normalized space. As another modified method, amyloid PET of each subject was used for tissue segmentation by directly inputting PET instead of MR for tissue segmentation algorithm of SPM8 and then nonlinear transformation was performed.

After the normalization using these five methods 1) authentic MR-based, 2) generated MR-based, 3) a normal PET template-based, 4) multi-atlas PET template-based, 5) PET segmentation-based, cortical uptake, SUVR, was scaled using the GM-masked cerebellum defined by automated anatomical labeling map. A template GM mask of SPM8 was identically applied regardless of normalization methods. Composite cortical SUVR values were computed using the mean SUVR in the GM-masked region consisting of the frontal, superior parietal, lateral temporal, and anterior and posterior cingulate regions. As uptake in specific regions, frontal, cingulate, superior parietal and lateral temporal SUVR were also obtained and compared.

Statistics

To measure the similarity between generated MR and real MR, structural similarity (SSIM) index was calculated (*15*).

$$\text{SSIM}(x,y) = \frac{(2\mu_x\mu_y + c_1)(2\sigma_{xy} + c_2)}{(\mu_x^2 + \mu_y^2 + c_1)(\sigma_x^2 + \sigma_y^2 + c_2)}$$

where μ and σ represent mean and standard deviation of image *x* and *y*. c_1 and c_2 represent two constant variables determined by pixel-value range. SSIM can have a value between -1 and 1, and SSIM=1 means that two images are the same. As PET and MRI have different field-of-view, tissues outside the brain are differently included in generated and real MR. Thus, SSIM was measured for the extracted brain of both images. One-way ANOVA was conducted to compare the effect of subjects' diagnosis on SSIM.

SUVR of MR-less methods were compared with those of authentic MR-based method using paired T-test. Mean absolute errors (MAEs) of SUVR for each MR-less methods were calculated by using the MR-based method as a

gold standard. MAEs of different methods were compared using paired T-test. Bland-Altman plots were drawn to evaluate the agreement between MR-less and MR-based methods.

RESULTS

Our model generated structural MR images using amyloid PET images. The processing time for a subject was approximately 11 seconds under GPU (NVIDIA® GTX 1080Ti) and 49 seconds under CPU (Intel® i7-7700). The examples of generated MR images are represented in Fig. 3. Generated MR image has similar signal intensity patterns with real MR. Note that MR images could be generated in both AD patient as well as normal subjects regardless of various tracer uptake patterns. Overall SSIM was 0.91 ± 0.04 (range: 0.77-0.98) for the brain. Subjects' diagnosis had no effect on SSIM (0.91 ± 0.04, 0.92 ± 0.04 and 0.91 ± 0.04 for AD, MCI and normal controls, respectively; F=0.68, p =0.52).

We applied this model to quantitative assessment of amyloid PET images. We compared the quantification of cortical amyloid load using different normalization methods. MAEs of MR-less methods were computed by using the MR-based method as a gold standard (Table 2). MAE of SUVR of composite regions estimated by the generated MR-based method was 0.04 ± 0.03, which was significantly smaller than other MR-less methods (0.29 ± 0.12 for the normal PET-template, 0.12 ± 0.07 for multiatlas PET-template and 0.08 ± 0.06 for PET segmentation-based methods). MAEs of generated MR-based method for other region-of-interests were significantly lower than other methods (p < 0.0001 for all regions). SUVRs calculated by MR-based and MR-less methods of each subject was plotted (Fig. 4). Bland-Altman plots were also drawn to compare the methods (Supplementary Fig. 1). As shown in the figures, SUVRs of generated MR-based method most highly corresponded to those of the MR-based method, while PET-based methods (normal template-based, multiatlas-based and PET segmentation-based) showed relatively higher errors and biases. In particular, normal PET template-based and PET segmentation-based methods showed higher bias when a subject was AD. Mutiatlas-based method showed less bias than normal PET-template based method, however, still tended to underestimate SUVRs. SUVRs of different regions calculated by these 5 different methods were compared (Table 3). The normal PET template-based and PET segmentation-based methods showed significant underestimation compared with MR-based methods regardless of subjects' diagnosis as well as cortical regions-of-interest. The multiatlas PET-template based method also showed significantly lower SUVR in composite cortical,

frontal and lateral temporal regions. SUVRs calculated by generated MR-based method showed no significant difference with the gold standard in most brain regions of interests.

DISCUSSION

The MR generation from PET is challenging because PET has relatively less structural and textural information than MR. We employed a recently developed deep neural network model, GAN (*16*). In brief, GAN typically has two network components, generator and discriminator. The generator is trained to estimate mapping of realistic images from a few feature vectors with particular distribution and the discriminator is trained to discriminate between true and generated data. In our approach, we used the network translating PET images to MR images as the generator component. Generated MR was paired with the matched PET and entered into the discriminator. This type of image translation based on the adversarial training was recently reported for generating realistic images from image sketches (*13*). Since the image translation could be useful in medical images as multimodal images have provided different information. We extended to this image translation method to MR generation from PET and suggested a clinically feasible application to MR-less amyloid quantification.

Accurate cortical amyloid quantification is crucial in AD diagnosis as well as predicting future cognitive decline in MCI and early dementia patients (*17*). A simple normal PET template-based normalization resulted in considerable bias in SUVR calculation. AD patients showed underestimated SUVR compared with MR-based methods, which corresponded to the previous result (*9*). This bias could be caused by nonrigid registration error as it used intensity-based registration. For AD patients, high florbetapir uptake in the cortex tends to shift to the white matter of template atlas during normalization, which could lead to the underestimation (*9*). The biased results were much less in multiatlas PET template-based and PET segmentation-based methods, however, they also showed a trend of underestimation. Generated MR-based method was highly correlated with real MR-based normalization results as well as unbiased. Bland-Altman plots showed that SUVR measured by the generated MR-based method was the closest to SUVR of MR-based method. The deviation of SUVR from the MR-based SUVR was also lower than the other methods. In addition, MAE of generated MR-based method was significantly lower than other methods. The accurate MR-less quantification method enables the use of PET data without MR for large clinical trials, and it can be utilized clinically as a quantitative marker for predicting cognitive outcome.

Though the amyloid load was variable according to the subjects' diagnosis and the regions-of-interest, our model could generate pseudo-MR images irrespective of the uptake patterns. This is a crucial advantage of our GAN method in PET image quantification because common spatial normalization algorithms are used in intensity-based nonrigid transformation (*18*). As aforementioned, when regional tracer uptakes are variable between regions and between individuals according to the disease status, the normalization based on intensity-based nonrigid transformation could cause seriously biased results. Another examples were reported when the investigators applied normalization for dopamine transporter imaging, the results were biased in severe patients (*19,20*). Quantification of tracer accumulation using generated MR could substantially overcome these biases. Furthermore, our GAN model could be trained for other variety of PET imaging and the generated MR images could be easily used for further spatial normalization and region segmentation. We think that this GAN method might be also used for quantitative analysis of other organs. Several unsolved problems in multimodal imaging can be solved with this GAN method. For example, MR generation could be applied to lesion segmentation in cancer imaging and CT generation could be used for attenuation correction without real CT images (*21*). It could be also used in partial volume correction without anatomical images (*22,23*). As a future work, various validation studies are warranted by the application of generative network model to multimodal imaging.

Despite promising quantification results in our proposed GAN model, there are some limitations. Even though our model could generate realistic MR images, they could not reach the quality of the real MR images. Generated MR images tended to be blurred, making it difficult to visually distinguish between gray and white matter. Specifically, several generated MR images showed artifacts and inhomogeneous intensity across the brain (Supplementary Fig. 2). These artifacts could be caused by GAN which tend to generate noisy images compared to other generative models. In addition, model learning based on patches instead of whole image may cause inhomogeneous signal intensity. Therefore, in terms of amyloid load quantification, the benefit of generated MR could be a rough tissue segmentation for spatial normalization instead of definite brain tissue contrast. Nonetheless, as direct PET segmentation-based method showed significant bias for the quantification, the generated MR seemed to play a role in accurate MR-less processing. Further modification in network architectures and training processes might improve the image quality in the near future. In addition, our model was optimized for the training dataset, which shall limit the use of the generated MR images in the patients with seriously distorted architecture such as tumors or large tissue losses. The training data set in our study included all the subjects with various amount of

amyloid deposit and brain atrophy, MR generation for this clinically observed spectrum of florbetapir PET would be feasible. Another strength of this model was the compatibility of PET images obtained by various scanners as training was performed by minimally processed images acquired from various sites. We remind that MR generation was successfully achieved in test set regardless of PET images of different image quality due to various machines. Thus, we expect that our model could be applied to another florbetapir PET database. As a further work, the GAN model application to prospectively acquired PET images other than ADNI database will be proceeded for the establishment of generalized usage of this method.

CONCLUSION

We developed a model for generating structural MR from florbetapir PET using deep generative networks. Generated brain MR images were similar with real MR images and they were successfully applied to MR-less quantification of cortical amyloid load. Cortical florbetapir uptake measured using generated MR was the closest to that measured using real MR among MR-less quantification methods. As our model generating MR from brain PET and application to PET quantification are one of the feasible applications for the image generation, we expect that the model will be used for various imaging modalities as well as applications for developing quantitative imaging biomarkers.

## Competing Interests

Acknowledgement

REFERENCES

1. Werner P, Barthel H, Drzezga A, Sabri O. Current status and future role of brain PET/MRI in clinical and research settings. *Eur J Nucl Med Mol Imaging.* 2015;42:512.

2. Pichler BJ, Kolb A, Nägele T, Schlemmer H-P. PET/MRI: paving the way for the next generation of clinical multimodality imaging applications. *J Nucl Med.* 2010;51:333-336.

3. Clark CM, Schneider JA, Bedell BJ, et al. Use of florbetapir-PET for imaging β-amyloid pathology. *JAMA.* 2011;305:275-283.

4. Johnson KA, Minoshima S, Bohnen NI, et al. Appropriate use criteria for amyloid PET: a report of the Amyloid Imaging Task Force, the Society of Nuclear Medicine and Molecular Imaging, and the Alzheimer's Association. *J Nucl Med.* 2013;54:476-490.

5. Lopresti BJ, Klunk WE, Mathis CA, et al. Simplified quantification of Pittsburgh Compound B amyloid imaging PET studies: a comparative analysis. *J Nucl Med.* 2005;46:1959-1972.

6. Landau SM, Breault C, Joshi AD, et al. Amyloid-β imaging with Pittsburgh compound B and florbetapir: comparing radiotracers and quantification methods. *J Nucl Med.* 2013;54:70-77.

7. Fleisher AS, Chen K, Liu X, et al. Using positron emission tomography and florbetapir F 18 to image cortical amyloid in patients with mild cognitive impairment or dementia due to Alzheimer disease. *Arch Neurol.* 2011;68:1404-1411.

8. Jack Jr CR, Lowe VJ, Senjem ML, et al. 11C PiB and structural MRI provide complementary information in imaging of Alzheimer's disease and amnestic mild cognitive impairment. *Brain.* 2008;131:665-680.

9. Edison P, Carter S, Rinne JO, et al. Comparison of MRI based and PET template based approaches in the quantitative analysis of amyloid imaging with PIB-PET. *Neuroimage.* 2013;70:423-433.

10. Lundqvist R, Lilja J, Thomas BA, et al. Implementation and validation of an adaptive template registration method for 18F-flutemetamol imaging data. *J Nucl Med.* 2013;54:1472-1478.

11. Bourgeat P, Villemagne VL, Dore V, et al. Comparison of MR-less PiB SUVR quantification methods. *Neurobiol Aging.* 2015;36:S159-S166.

12. Jagust WJ, Landau SM, Koeppe RA, et al. The Alzheimer's Disease Neuroimaging Initiative 2 PET Core: 2015. *Alzheimers Dement.* 2015;11:757-771.

13. Isola P, Zhu J-Y, Zhou T, Efros AA. Image-to-image translation with conditional adversarial networks. *arXiv preprint arXiv:161107004.* 2016.

14. Zhou L, Salvado O, Dore V, et al. MR-less surface-based amyloid assessment based on 11C PiB PET. *PLoS One.* 2014;9:e84777.

15. Wang Z, Bovik AC, Sheikh HR, Simoncelli EP. Image quality assessment: from error visibility to structural similarity. *IEEE Trans Image Process.* 2004;13:600-612.

16. Goodfellow I, Pouget-Abadie J, Mirza M, et al. Generative adversarial nets. Paper presented at: Adv Neural Inf Process Syst, 2014.

17. Albert MS, DeKosky ST, Dickson D, et al. The diagnosis of mild cognitive impairment due to Alzheimer's disease: Recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. *Alzheimers Dement.* 2011;7:270-279.

18. Crum WR, Hartkens T, Hill D. Non-rigid image registration: theory and practice. *Br J Radiol.* 2004;77:S140-S153.

19. Kim JS, Cho H, Choi JY, et al. Feasibility of computed tomography-guided methods for spatial normalization of dopamine transporter positron emission tomography image. *PLoS One.* 2015;10:e0132585.

20. Kas A, Payoux P, Habert M-O, et al. Validation of a standardized normalization template for statistical parametric mapping analysis of 123I-FP-CIT images. *J Nucl Med.* 2007;48:1459-1467.

21. Hofmann M, Pichler B, Schölkopf B, Beyer T. Towards quantitative PET/MRI: a review of MR-based attenuation correction techniques. *Eur J Nucl Med Mol Imag.* 2009;36:93-104.

22. Rousset O, Rahmim A, Alavi A, Zaidi H. Partial volume correction strategies in PET. *PET clinics.* 2007;2:235-249.

23. Soret M, Bacharach SL, Buvat I. Partial-volume effect in PET tumor imaging. *J Nucl Med.* 2007;48:932-945.

Figure legends
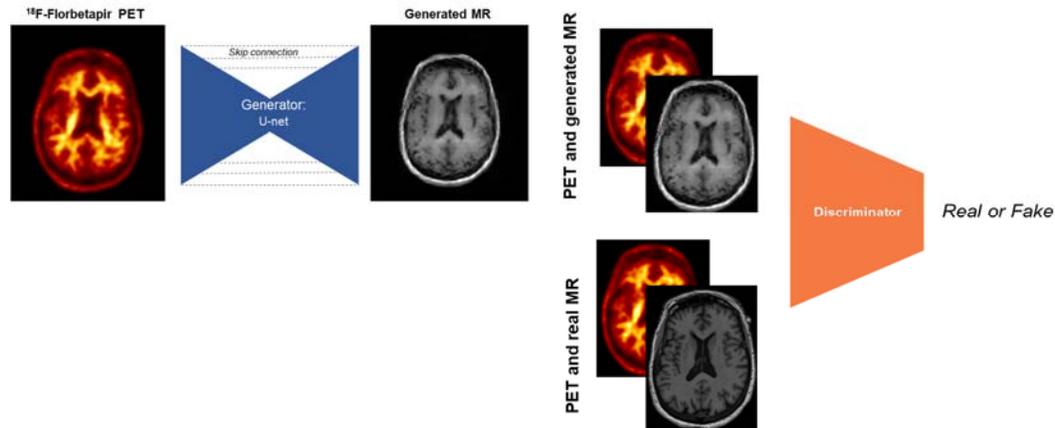
Figure 1. Adversarial training for the MR generation network. A generative adversarial network (GAN) consists of multiple convolutional and deconvolutional layers to translate florbetapir PET to structural MR images. Training of the network was aimed at generating MR images which cannot be distinguished from real images. In contrast, another discriminator network was trained to distinguish real MR from generated MR images. They competed in the entire training process.
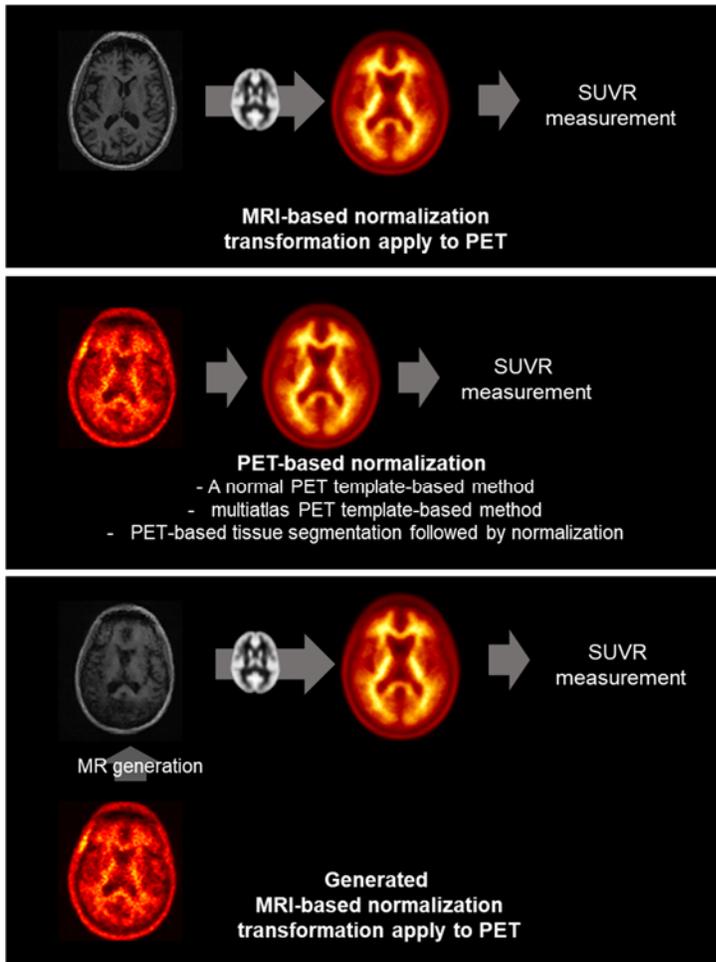
Figure 2. Amyloid PET quantification using different methods. We applied the MR generation model to quantification of amyloid PET. As a gold standard method, the MR-based normalization was employed. PET images were coregistered to corresponding MR and then nonrigid transformation of MR was performed for spatial normalization. Predefined cortical and reference regions were used for calculating standardized uptake value ratio (SUVR). For a normal PET template-based method, an averaged florbetapir PET images of normal controls were used as a template and then all the PET images were directly normalized to this template. Multiatlas PET template-based quantification chose the most similar PET template to a subject's PET image among various PET templates with different tracer uptake patterns and then the images were normalized to the selected templates. In addition, as a modified method, PET was directly used for the tissue segmentation and the segmented tissues were normalized into the template space. As an application of our GAN model, generated MR images were spatially normalized to the MR template and the corresponding PET images were transformed to the template space. We compared the SUVRs measured by these four different normalization methods.
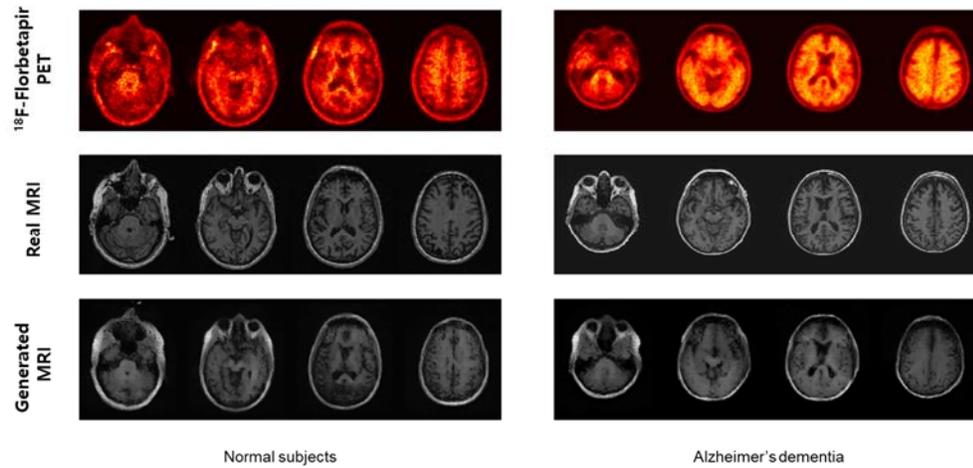
Figure 3. Examples of generated MR images. After the training, MR images were generated from amyloid PET images of independent test set. Regardless of subjects' diagnosis, MR images were generated and similar signal patterns were observed to the corresponding real MR images. Quantitative similarity measured by structural similarity index measurement between real and generated brain was 0.91 ± 0.04.
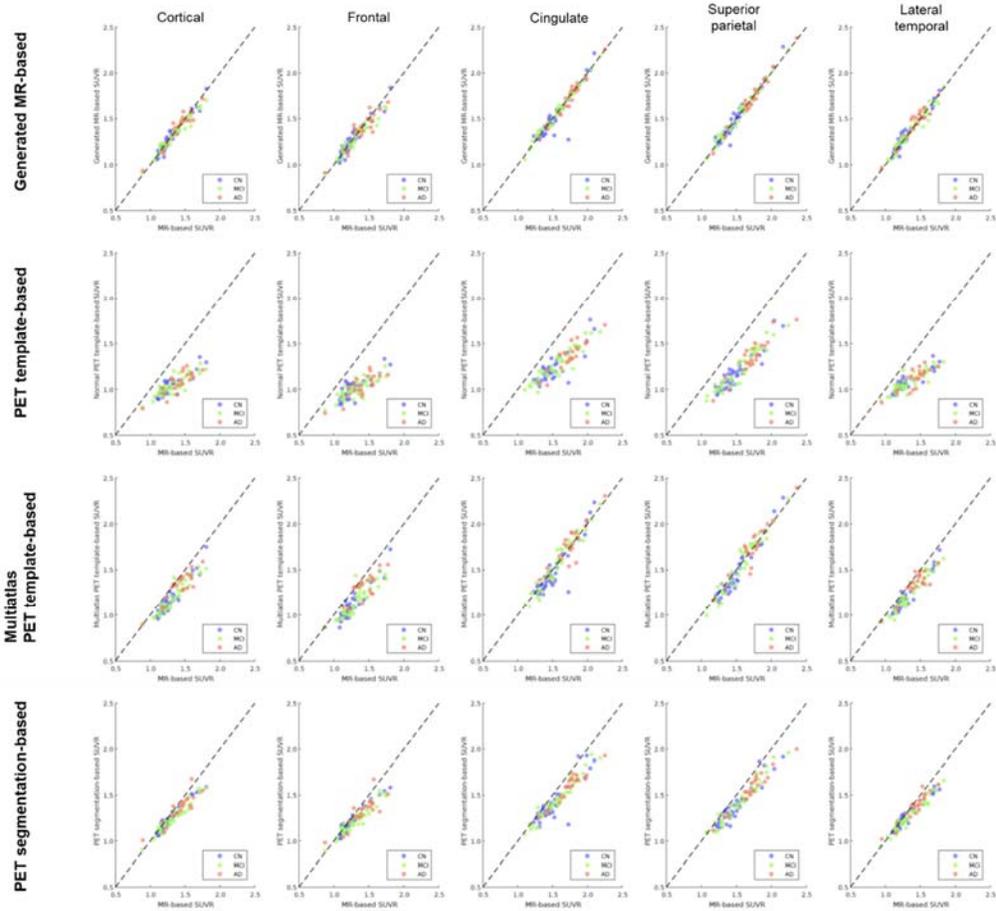
Figure 4. Scatter plots of SUVRs calculated by different normalization methods. SUVRs measured by MR-less methods were compared with MR-based quantification results. Generated MR-based SUVR quantification results were highly correlated with the MR-based quantification results. However, a normal PET template-based method showed biased results. The multiatlas PET template-based and PET segmentation-based methods showed less biased results than the normal PET template-based method, however, relatively higher error than the generated MR-based method (AD: Alzheimer's disease; MCI: mild cognitive impairment, CN: Controls).

Tables

Table 1. Demographics and clinical diagnosis of training and test dataset

	Training Dataset (n = 163)	Test Dataset (n=98)
Age	73.2 ± 6.7 (56.0 - 90.0)	72.2 ± 5.9 (57.0 - 88.0)
Sex (M:F)	75:88	53:45
Diagnosis (AD:MCI:NC)	34:80:49	21:41:36

Table 2. Errors of standardized uptake value ratio of florbetapir measured by MR-less methods.

	Mean absolute error compared with MR-based method			
	Generated MR-based	Normal PET template-based	Multiatlas PET template-based	PET segmentation-based
Cortical	0.04±0.03	0.29±0.12	0.12±0.07	0.08±0.06
Frontal	0.05±0.04	0.31±0.13	0.15±0.08	0.09±0.07
Cingulate	0.04±0.05	0.31±0.12	0.07±0.06	0.13±0.08
Superior parietal	0.03±0.03	0.33±0.10	0.07±0.05	0.13±0.07
Lateral temporal	0.04±0.03	0.26±0.12	0.11±0.06	0.07±0.05

Table 3. Standardized uptake value ratio of cortical florbetapir PET measured by different methods.

		Amyloid SUVR quantification methods				
		MR-based	**Generated MR-based**	**Normal PET template-based**	**Multiatlas PET template-based**	**PET segmentation-based**
Normal Controls	Cortical	1.29±0.18	**1.28±0.17**	1.04±0.11**	1.17±0.17**	1.24±0.14**
	Frontal	1.27±0.18	**1.26±0.18**	1.00±0.12**	1.12±0.17**	1.22±0.14**
	Cingulate	1.52±0.23	**1.50±0.24**	1.24±0.17**	1.48±0.25*	1.41±0.20**
	Superior parietal	1.49±0.23	**1.49±0.25**	1.17±0.18**	1.45±0.26**	1.38±0.20**
	Lateral temporal	1.30±0.18	1.32±0.18*	1.07±0.10**	1.20±0.17**	1.26±0.14**
MCI	Cortical	1.34±0.20	1.32±0.18*	1.05±0.10**	1.21±0.18**	1.26±0.15**
	Frontal	1.32±0.20	1.30±0.17*	1.05±0.10**	1.16±0.17**	1.23±0.14**
	Cingulate	1.57±0.26	**1.57±0.25**	1.27±0.17**	**1.57±0.27**	1.45±0.22**
	Superior parietal	1.55±0.27	**1.55±0.27**	1.22±0.21**	**1.53±0.30**	1.42±0.23**
	Lateral temporal	1.35±0.22	**1.35±0.21**	1.08±0.11**	1.23±0.19**	1.28±0.17**
Alzheimer's dementia	Cortical	1.41±0.19	**1.41±0.18**	1.08±0.12**	1.32±0.16**	1.35±0.15**
	Frontal	1.40±0.20	**1.38±0.18**	1.04±0.13**	1.29±0.17**	1.31±0.15**
	Cingulate	1.74±0.21	**1.73±0.22**	1.37±0.16**	**1.76±0.22**	1.59±0.16**
	Superior parietal	1.73±0.25	**1.72±0.26**	1.35±0.21**	**1.74±0.27**	1.57±0.20**
	Lateral temporal	1.43±0.18	**1.45±0.18**	1.12±0.12**	1.34±0.15**	1.38±0.15**

p-values are obtained by paired t-tests compared with the MR-based method. * $p < 0.05$; ** $p < 0.01$. Bold texts

represent no significant difference with the MR-based method.

Structural MR images generation from amyloid PET: Application to MR-less quantification

Supplementary Materials and Methods

Generative Adversarial Model for MR Image Generation

The model was based on recent works for image-to-image translation using generative adversarial model (*1*). It successfully applied to tasks for synthesizing complex images from simple representations, for instance, translating sketches to photos and colorization of black-and-white images (*1*). We applied this approach to translate PET images to structural MR images. The training process uses generative adversarial network. The generator G, a convolutional neural network based on U-net (*2*), is trained to translate PET (x) to MR (y) images which cannot be distinguished from real images. In contrast, the discriminator D, another convolutional neural network, attempts to minimize the misclassification error for distinguishing real pairs of PET/MR images from synthetic pairs of PET/MR images (Fig. 1). Specifically, the objective of the model is

$$L_{Adv}(G, D) = \mathbb{E}_{x,y \sim p(x,y)}[\log D(x, y)] + \mathbb{E}_{x \sim p(x)}[1 - \log D(x, G(x))]$$

where $\mathbb{E}_{x,y \sim p(x,y)}$ represents the expectation that PET (x) and MR (y) images are sampled from the probability distribution of real pairs *p(x,y)*. $\mathbb{E}_{x,y \sim p(x,y)}[\log D(x, y)]$ is maximized when $D(x, y) = 1$ as the output of D has the range of 0 to 1. $\mathbb{E}_{x \sim p(x)}$ represents the expectation that PET is sampled from the probability distribution of *p(x)*. The posterior part is maximized when $D(x, G(x))$=0, while it is minimized when G successfully fools the D, $D(x, G(x))$=1. Thus, the training of D is aimed to maximize the L_{GAN}, while G attempts to minimize the L_{GAN}. The adversarial objective function was additionally combined with the loss for pixel-based regression represented by L1-distance between real and generated images.

$$L_{L1}(G, D) = \mathbb{E}_{x,y \sim p(x,y)}(\| y - G(x) \|_1)$$

Thus, the final objective function is $L(G, D) = L_{Adv} + \alpha L_{L1}$. α is a factor which determines the contribution of two types of losses. In our model, $\alpha = 100$ was used.

Training and Testing for MR Image Generation

Neural networks consisted of generator and discriminator. Specific parameters of the neural networks are described in Supplementary Tables. The generator consisted of multiple convolutional layers with skipped connections, which was based on U-net (*2*). Axial slices of PET images with 192 x 256 matrix size were subsampled to half-sized feature maps iteratively. These convolutional layers finally produced feature maps with 3 x 4 matrices followed by iterative deconvolutional layers which finally produce original sized matrices. Each input of deconvolutional layer was connected to same sized feature matrices produced by a previous convolutional

layer. Discriminator consisted of 6 convolutional layers and finally produced 6 x 8 matrix to perform patch-based learning.

Training was performed by PET and MR images of 169 subjects. Input of the model was an axial slice image of PET scan. Total axial slices of 169 subjects were 32659 images. The network was trained using the Adam optimizer (*3*). Testing was performed by images of 98 subjects independent from the training data. Row PET images were resliced to have voxel size of $1.2 \times 1.0 \times 1.0$ mm^3 as aforementioned. Generator produced synthetic MR slices corresponding to PET slices.

Supplementary References

1. Isola P, Zhu J-Y, Zhou T, Efros AA. Image-to-image translation with conditional adversarial networks. *arXiv preprint arXiv:161107004.* 2016.

2. Ronneberger O, Fischer P, Brox T. U-net: Convolutional networks for biomedical image segmentation. Paper presented at: International Conference on Medical Image Computing and Computer-Assisted Intervention, 2015.

3. Kingma D, Ba J. Adam: A method for stochastic optimization. *arXiv preprint arXiv:14126980.* 2014.

Supplementary Figures
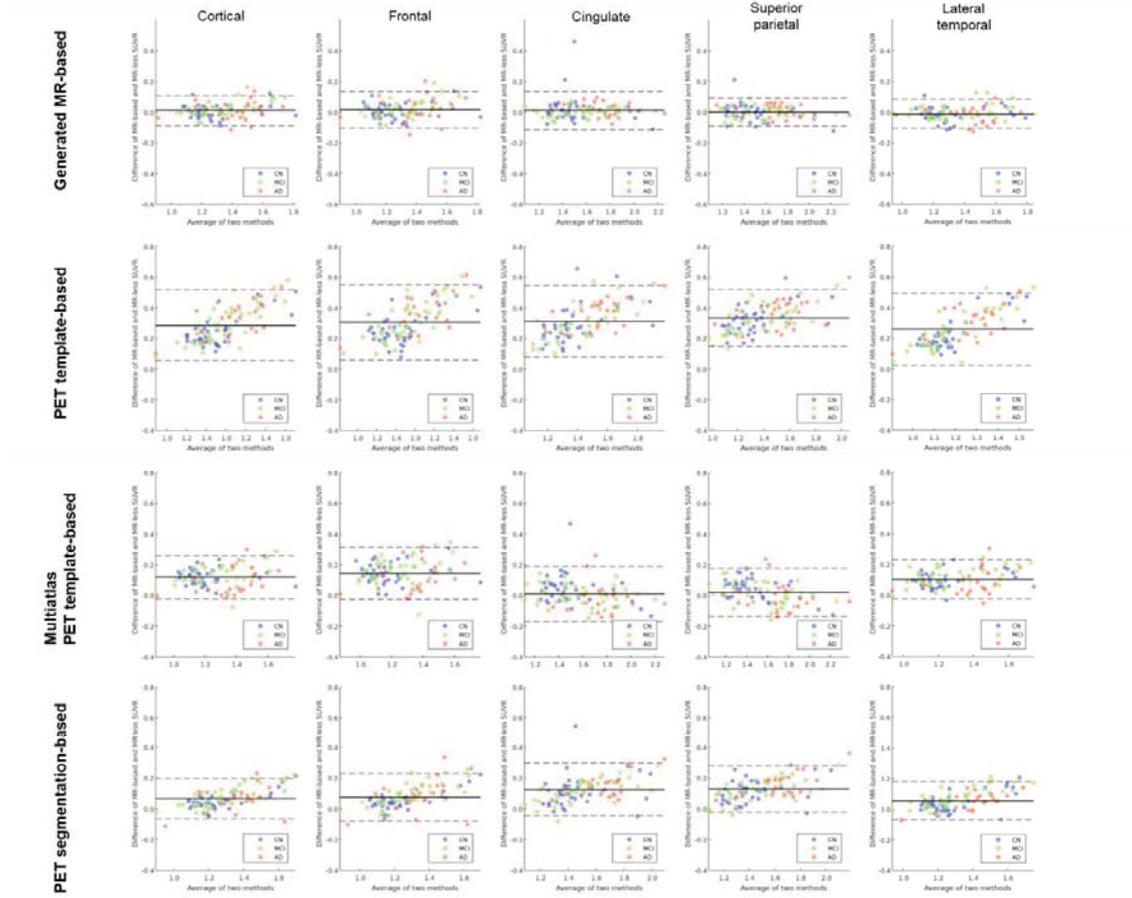
Supplementary Fig. 1. Bland-Altman plots for different MR-less methods. Bland-Altman plots were drawn for the comparison between the MR-less and MR-based quantification methods. SUVRs calculated using the generated MR-based method showed the most accurate and precise quantification results compared with the other three PET template-based methods. Of note, the normal PET template-based and PET segmentation-based methods showed a trend of higher bias when amyloid load was increased.
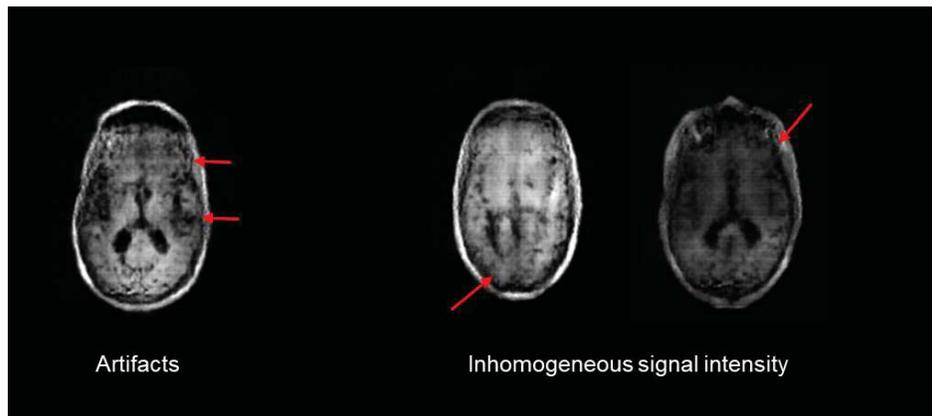
Supplementary Fig. 2. Artifacts of generated MR. Generated MR showed artifacts which could not be found in true MR images. In particular, irregular and inhomogeneous signal intensities were found in several generated MR images.

Supplementary Table 1. Network architectures for the generative network, U-net.

	Filter size	Filter number	Activation	Subsampling	Output Size	
Convolutional layer 1	3x3	64		1/2	96x128	Input: Axial slices of PET image
Convolutional layer 2	3x3	128		1/2	48x64	
Convolutional layer 3	3x3	256		1/2	24x32	
Convolutional layer 4	3x3	512		1/2	12x16	
Convolutional layer 5	3x3	512		1/2	6x8	
Convolutional layer 6	3x3	512	LeakyReLU activation	1/2	3x4	
Deconvolution layer 1	3x3	512		2	6x8	Concatenation to Convolutional layer 5
Deconvolution layer 2	3x3	512		2	12x16	Concatenation to Convolutional layer 4
Deconvolution layer 3	3x3	256		2	24x32	Concatenation to Convolutional layer 3
Deconvolution layer 4	3x3	128		2	48x64	Concatenation to Convolutional layer 2
Deconvolution layer 5	3x3	64		2	96x128	Concatenation to Convolutional layer 1
Deconvolution layer 6	3x3	1	Tanh activation	2	192x256	Output: Axial slices of PET image

Supplementary Table 2. Network architectures for the discriminative network.

	Filter size	Filter number	Activation	Subsampling	Output Size	
Convolutional layer 1	5x5	64		1/2	96x128	Input: Pairs of PET and MR images
Convolutional layer 2	5x5	64		1/2	48x64	
Convolutional layer 3	5x5	128	LeakyReLU activation	1/2	24x32	
Convolutional layer 4	5x5	256		1/2	12x16	
Convolutional layer 5	5x5	512		1/2	6x8	
Convolutional layer 6	5x5	1	Sigmoid activation	-	6x8	Output: 1 for real MR, 0 for generated MR