Journal of Digital Imaging (2020) 33:816–825
https://doi.org/10.1007/s10278-020-00321-5

ORIGINAL PAPER

Semantic Segmentation of White Matter in FDG-PET Using Generative Adversarial Network

Kyeong Taek Oh[1] · Sangwon Lee[2] · Haeun Lee[1] · Mijin Yun[2] · Sun K. Yoo[1]

Published online: 10 February 2020

Abstract

In the diagnosis of neurodegenerative disorders, F-18 fluorodeoxyglucose positron emission tomography/computed tomography (^{18}F-FDG PET/CT) is used for its ability to detect functional changes at early stages of disease process. However, anatomical information from another modality (CT or MRI) is still needed to properly interpret and localize the radiotracer uptake due to its low spatial resolution. Lack of structural information limits segmentation and accurate quantification of the ^{18}F-FDG PET/CT. The correct segmentation of the brain compartment in ^{18}F-FDG PET/CT will enable the quantitative analysis of the ^{18}F-FDG PET/CT scan alone. In this paper, we propose a method to segment white matter in ^{18}F-FDG PET/CT images using generative adversarial network (GAN). The segmentation result of GAN model was evaluated using evaluation parameters such as dice, AUC-PR, precision, and recall. It was also compared with other deep learning methods. As a result, the proposed method achieves superior segmentation accuracy and reliability compared with other deep learning methods.

Keywords GAN · Deep learning · FDG-PET · ADNI · White matter segmentation

Introduction

Segmentation of the brain compartment such as gray matter (GM), white matter (WM), and cerebrospinal fluid (CSF) for the quantification of tissue volume and functional analysis of different structures is of great importance for research and clinical studies using magnetic resonance imaging (MRI) of the brain [1]. For MRI, various approaches and open source software packages have been used for brain segmentation and volumetric quantification. Recently, deep learning models have been used in developing algorithms for segmentation of brain structures in anatomical images [2–5].

Of the deep learning-related algorithms, generative adversarial network (GAN) model has revealed excellent performance in image generation tasks, including image-to-image translation, text-to-image synthesis, semantic segmentation, and low to high resolution translation [6]. GAN consists of two networks which has a generator and a discriminator. The generator learns a mapping function to create similar output to real data. The discriminator learns how to differentiate the generated data from the original data. After the concept of adversarial learning was introduced, various GAN models were applied for automatic segmentation of medical images with excellent results. Mondal et al. showed a higher performance for segmenting brain structure adopting feature matching loss than conventional adversarial training approaches [7]. Dai et al. used GAN to segment organ in chest X-ray and described that the interplay of the generator and discriminator can correct the shape inconsistency [8]. Izadi et al. used GAN to segment skin lesions and verified that the adversarial training helps to refine the boundary precision compared with u-net alone [9]. Others proposed conditional GAN with pix2pix [10] framework for semantic segmentation of tumors from MR images [10, 11].

✉ Mijin Yun
YUNMIJIN@yuhs.ac

✉ Sun K. Yoo
SUNKYOO@yuhs.ac

Kyeong Taek Oh
okt2704@yuhs.ac

Sangwon Lee
LSW0423@yuhs.ac

Haeun Lee
DLGK0205@yuhs.ac

[1] Department of Medical Engineering, Yonsei University College of Medicine, Seoul, South Korea

[2] Department of Nuclear Medicine, Yonsei University College of Medicine, Seoul, South Korea

F-18 fluorodeoxyglucose positron emission tomography (18F-FDG PET/CT) is a functional imaging modality which measures changes of glucose metabolism in the brain [12]. As a parameter of functionality and density of synapse, the detection of metabolic changes allows the diagnosis of neurodegenerative diseases at early stages [13]. It provides the severity, extent, and location of disease which are important clues for the identification of subtypes, staging, and prognostication of neurodegenerative diseases. Compared with MRI, there are few studies that applied deep learning for brain PET/CT. Wang et al. proposed a method to estimate high-quality full-dose PET images from low-dose PET images using 3D conditional GAN [14]. Choi et al. proposed a method to generate MR images from amyloid PET using conditional GAN with pix2pix framework [15]. For segmentation, Blanc-Durand et al. used 3D u-net shaped convolutional neural network to segment lesion of F-18 fluoroethyltyrosine (18F-FET) PET in cerebral gliomas [16]. So far, no studies have applied GAN framework to segment brain compartment using 18F-FDG PET/CT.

18F-FDG PET/CT evaluates cortical or subcortical neuronal metabolic activity of the brain and the assessment of the white matter pathologies depends on anatomical imaging modalities such as MRI [17]. The potential values of extracting the white matter from 18F-FDG PET/CT have not been evaluated for the quantitative evaluation of various brain diseases. In this study, we proposed a GAN model to segment the white matter compartment of the brain using 18F-FDG PET/CT images.

Methods

The learning structure of the GAN model used in this study was shown in Fig. 1. To estimate the segmentation map M which showed the white matter region in the image when ^{18}F-FDG PET/CT image I was given, we let a set of given images be $I = \{I_1, \ldots, I_n\}$ and then the set of segmentation maps according to a given image was labeled as $M = \{M_1, \ldots, M_n\}$. The mapping function of the image of 18F-FDG PET/CT to white matter segmentation map was defined as $F: I \rightarrow M$ in which the mapping function F was designed as a GAN model.

Data Set

18F-FDG PET/CT and MRI data were collected from Alzheimer's disease neuroimaging initiative (ADNI) database to train the GAN model. ADNI is designed to develop combined biomarkers for early detection and to track progression of Alzheimer's disease which includes data from more than 50 sites across the USA and Canada. For this study, we used data from 192 subjects who have both 18F-FDG PET/CT and MRI. Test and validation set independent from the training set were used to verify the performance of the GAN model. Of the 192 data, 154 were used for training set, 19 were used for validation set, and 19 were used for test set. Table 1 summarized the patients in the training set, validation set, and test set.

Data Preprocessing

Preprocessed 18F-FDG PET/CT images downloaded from ADNI were used to train the GAN model. The raw image data consisted of six 5-min frames for 30–60 min after injection. Each image was co-registered to the first acquired image (the image acquired from 30 to 35 min after the injection) and the co-registered images were averaged. The preprocessed images were created by re-orienting the averaged images to a normalized space.

For the MRI data, structural T1 images acquired concurrently with 18F-FDG PET/CT images were used. Unlike 18F-FDG PET/CT images, MR data have different voxel sizes and orientations. The voxel size in the coronal slice was in the range of $0.93 \times 1.18\text{mm}^2$ to $1.31 \times 1.22\text{mm}^2$, and slice thickness was in the range of 0.92 to 1.31 mm. In order to match the images with different voxel sizes and phases to the normalized space, images were normalized to the space defined by the International Consortium for Brain Mapping (ICBM) template.

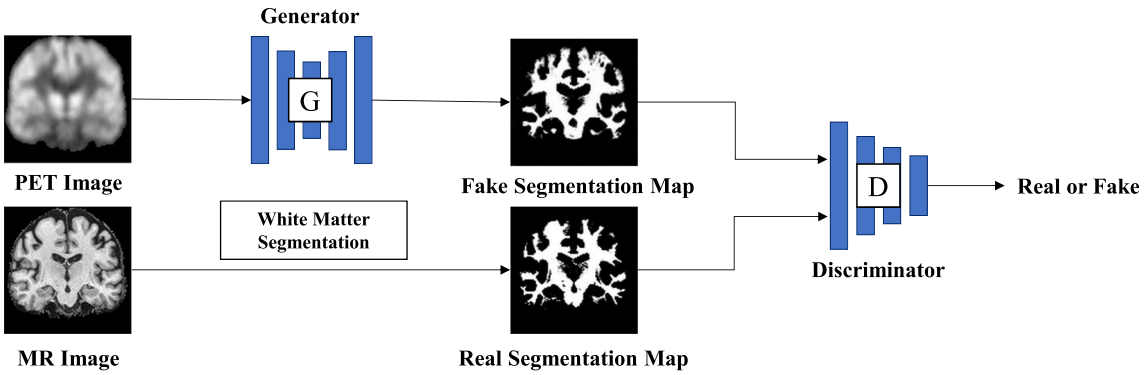

Fig. 1 Adversarial training for the segmentation map generation network

Table 1 Demographics of training and test dataset

	Training dataset (n = 154)	Validation dataset (n = 19)	Test dataset (n = 19)
Age	77.4 ± 6.2 (60.0–90.0)	78.2 ± 8.2 (64.0–87.0)	77.7 ± 5.5 (64.0–87.0)
Sex (M:F)	87:64	12:7	8:11

To avoid non-specific information of the non-brain region, only brain region was extracted from the 18F-FDG PET/CT and MRI that underwent spatial normalization which includes affine transformation and warping. Then, 18F-FDG PET/CT was co-registered to MRI. The voxel size of co-registered 18F-FDG PET/CT and MRI was $1.50 \times 1.50 \times 1.50$ mm^3. For training, the voxel values outside the range of the FDG-PET/CT image in the co-registered MRI were replaced by zero. Next, the segmentation map of the white matter was extracted from MRI to compare with the segmentation map generated from GAN model. The spatial normalization, brain segmentation, co-registration, and segmentation map extraction in preprocessing were performed using statistical parametric mapping (SPM) 12 [18].

Architectural Design

The GAN model was based on the structure of the image-to-image translation GAN [10] model which is called pix2pix. This model consisted of two convolution networks as shown in Fig. 1, corresponding to generator and discriminator, respectively. The generator was trained to convert the 18F-FDG PET/CT image to a segmentation map which was to be indistinguishable from the real segmentation map. The discriminator was trained to distinguish the generated segmentation map from the real segmentation map. Through adversarial training of generators and discriminators, the generator generated realistic segmentation maps. Figure 2 showed the structure of the generator and discriminator.

Residual Block

Each residual block consisted of two convolution layers and each of them was followed by the batch-normalization layer (Fig. 2). The rectifier linear unit (ReLU) was for the activation function of the first convolution layer as proposed by He et al. [19] to reduce the effect of vanishing gradient problem and to accelerate the speed of training of the deep networks. In the residual block, the kernel size of the convolution layer was 3×3, and the size of the input feature map as well as the output feature map were constant by using the reflect padding. The input and output channels of the convolution layer were 256.

Generator

The generator was made of 6 convolution layers and 6 residual blocks. The first convolution layer 1 had a kernel size of 7×7 with reflect padding and a stride of 1. The kernel size for convolution layer 2 and 3 was 3×3 with zero-padding and a

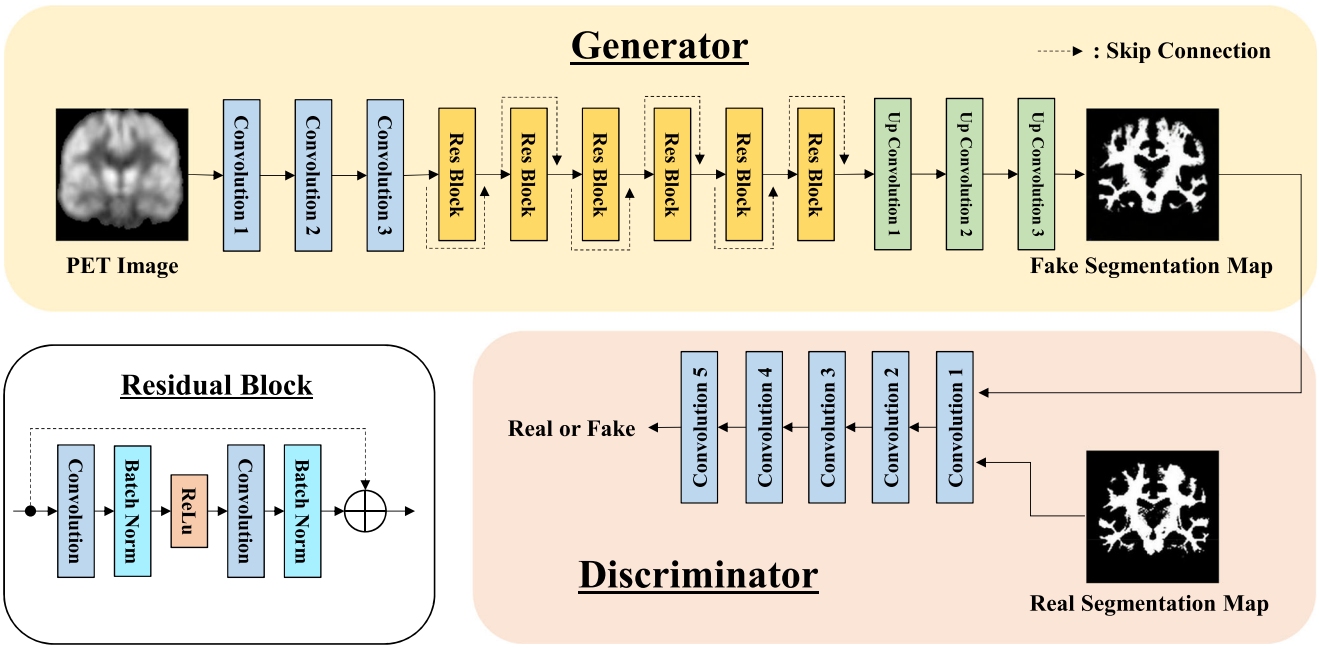

Fig. 2 Architecture of generator and discriminator

stride of 2 to down-sample the spatial dimension of the feature map output. The output channels of convolution layers 2 and 3 were 128 and 256, respectively. The kernel size of the up-convolution layers 2 and 3 following the residual block was 3 × 3 with zero-padding and a stride of 2. Unlike the first convolution layer, the latter up-convolution layer doubled the spatial dimension of the feature map, which was reduced by convolution operation. The last layer, up-convolution layer 3, had a kernel size of 7 × 7 with reflect padding and a stride of 1 to generate an image of the same size as the input image. The output channels of the up-convolution layers 1, 2, and 3 were 128, 64, and 1, respectively. After every convolution layer except the last layer, there was a batch-normalization layer followed by ReLU as an activation function. In up-convolution layer 3, the hyperbolic tangent function was used as the activation function.

Discriminator

The discriminator consisted of 5 convolution layers in which the kernel size was 4 × 4 with zero-padding and stride of 2 to down-sample the spatial dimension of the output feature map. After each convolution layer except for the last one, there was a batch-normalization layer and followed by leaky ReLU. The sigmoid function was used as the activation function of the last layer. The output channels were 64, 128, 256, 512, and 1 from convolution layer 1 to 5 in order.

Loss Function

There was optimization of two loss functions to train GAN model. The first was the GAN loss(L_GAN) that occurred when the discriminator tried to distinguish the segmentation map generated by the generator. The second was the L1 loss(L_1) which was pixel-based regression loss expressed by the L1-distance between the generated segmentation map and the actual segmentation map.

Generator of GAN model, G, was trained to convert 18F-FDG PET/CT image (I) to segmentation map (M) which was hard to distinguish from real segmentation map. On the other hand, the discriminator, D, was trained to reduce the misclassification error of the real segmentation map and the segmentation map generated by the generator. This adversarial training was expressed as Eq. (1).

$$L_{\mathrm{GAN}}(G,D) = \mathbb{E}_{I,M\sim p(I,M)}[\log D(I,M)] + \mathbb{E}_{I\sim p(I)}[\log D(I,G(I))] \quad (1)$$

In Eq. (1), $\mathbb{E}_{I,M\sim p(I,M)}$ represented the expected value at which 18F-FDG PET/CT(I) and segmentation map (M) was to be sampled in the probability distribution $p(I,M)$. $\mathbb{E}_{I,M\sim p(I,M)}[\log D(I,M)]$ is the maximum when $D(I,M) = 1$ since the output of D is in the range of 0 to 1. $\mathbb{E}_{I\sim p(I)}$ represents the expectation value that PET(I) to be sampled from the probability distribution $p(I)$. $\mathbb{E}_{I\sim p(I)}[\log D(I,G(I))]$ is maximized when $D(I,G(I)) = 0$ and minimized when G successfully deferred D. Thus, training of D aims to maximize L_{GA6N} and G tries to minimize L_{GAN}.

L1 loss(L_1) calculated L1-distance between M and $G(I)$, which is expressed as Eq. (2).

$$L_{L1}(G) = \mathbb{E}_{I,M\sim p(I,M)}\left[\|M - G(I)\|_1\right] \quad (2)$$

The two loss functions were combined into one loss function, which was shown in Eq. (3). In Eq. (3), α was a weight parameter to determine the weight of each loss function. In this paper, α was set to 100.

$$L = L_{GAN} + \alpha \cdot L_{L1} \quad (3)$$

Model Learning

To optimize the GAN model, we applied the hyper parameter proposed by previous method [10] to train the model. To train the model, a minibatch stochastic gradient descent (SGD) was used and the batch size was set to 1. Adaptive moment estimation (Adam) was used as optimizer, learning rate was set to 0.0002, momentum parameters were set to $\beta_1 = 0.5$ and $\beta_2 = 0.999$. The GPU used to train the model was NVIDIA Geforce GTX 1080 Ti.

The GAN model was trained using preprocessed 18F-FDG PET/CT images and segmentation maps representing white matter region in MRI. Coronal slices of co-registered 18F-FDG PET/CT images MR images were used for training. The total coronal slices for the training set were 1694 images. Two hundred nine coronal slices were used as validation set. Also, 209 coronal slices were used as test set. The size of the model input image was 256 × 256, and the size of the feature map through the encoding path was 64 × 64. The output of the model was reconstructed as an original image with a size of 256 × 256 through an up-convolution process.

Evaluation of Segmentation Results

In order to verify the performance of the proposed method, we compared the proposed method with the method (pix2pix_unet method) replacing the generator part with the u-net structure instead of the residual block, the method (h_dense_unet method) which used dense block which is composed of repetitive densely connected building blocks [20], and the method (u-net method) using the convolution network of the conventional u-net structure [21]. For the evaluation of various methods, the h_dense_unet and the u-net model were trained by changing the input size and using

zero-padding to maintain the input image size. Other parameters were set as the same as the original paper. We also compared the segmentation map generated by the proposed method and the method used for comparison with segmentation result of SPM in MRI which is used as ground truth. In addition, the precision-recall curve was compared to evaluate the performance of the proposed method.

Segmentation Quality Analysis

The generated images were first visually inspected for segmentation quality. Thirty samples in the evaluation set were randomly selected. The generated segmentation map from the proposed method, pix2pix_unet method, and u-net method of the randomly selected samples was anonymized, then presented by series number to five observers. The segmentation status of each segmentation map was determined. The segmentation result of SPM in MRI was treated as the ground truth. For each segmentation map, the observer assigned a segmentation quality score in a three-point scale: 1, over-estimated; 2, under-estimated; 3, adequate.

Evaluation Parameter

To evaluate the performance of the GAN model, area under the curve of precision-recall metrics (AUC-PR) and dice similarity coefficient (DSC) metrics [22] were used. AUC-PR produced a confusion matrix between ground truth and segmentation results. The confusion matrix was mainly used as an index to evaluate the performance of an algorithm. The DSC measured the similarity of spatial coincidence between the ground truth and segmentation results. The precision and recall used to calculate AUC-PR were defined by Eqs. (4) and (5), and the DSC matrix was defined by Eq. (6). F_P, F_N, and T_P used in the equation represented false positive, false negative, and true positive, respectively.

$$\text{Precision} = \frac{T_P}{T_P + F_P} \tag{4}$$

$$\text{Recall} = \frac{T_P}{T_P + F_N} \tag{5}$$

$$\text{Dice} = \frac{2 \times T_P}{2 \times T_P + F_P + F_N} \tag{6}$$

Statistical Evaluation

The Kruskal–Wallis test was performed between the methods to verify whether the differences between the evaluation parameter are statistically significant. The sample used in the Kruskal–Wallis test consists of evaluation parameters calculated from the segmentation map generated by each model using a test set. In addition, Dunn's multiple comparison test was performed to verify whether there was statistically significant difference in the evaluation parameters between each method.

Results

Segmentation Quality Analysis

Figure 3 showed the white matter ground truth and segmentation results of various methods in different conditions. All segmentation results were shown in red. In most accurate case, all the methods segmentation result was visually similar to ground truth. In the case with median DSC value, u-net method showed poor segmentation result by over-segmenting white matter regions while others showed visually similar results. In least accurate case, pix2pix_unet method showed poor result by segmenting less regions than the ground truth. In contrast, the h_dense_unet and the u-net method also showed poor result by over-segmenting white matter regions. However, the proposed method showed good segmentation result in least accurate case.

The segmentation quality scores assigned by each observer to each of the segmentation maps are shown in Fig. 4. For the proposed method, 78% of the segmentation results scored adequate. The pix2pix_unet method had fewer segmentation results with adequate (31%) and had more segmentation results with under-estimated (49%). For the h-dense-unet method, 63% of the segmentation results scored adequate and 27% of the segmentation results scored over-estimated. For the u-net method, most of the segmentation results scored over-estimated (93%). The mean value ± standard deviation (SD) was 2.6 ± 0.7 in the proposed method, 2.1 ± 0.7 in the pix2pix_unet method, and 1.1 ± 0.4 in the u-net method.

Quantitative Analysis of Evaluation Parameters

To compare the performance of each method, the evaluation parameters (precision, recall, dice, and AUC-PR) were calculated. Figure 5 showed the scores of evaluation parameters for each method. Table 2 summarized the results of Kruskal–Wallis test and Dunn's multiple comparison test between methods for each evaluation parameter.

For precision, the mean value ± SD was 0.821 ± 0.036 in the proposed method, 0.778 ± 0.054 in the pix2pix_unet method, 0.778 ± 0.054 in the pix2pix_unet method, 0.699 ± 0.039 in the h_dense_unet method, and 0.603 ± 0.048 in the u-net method, respectively ($p < 0.0001$). Also, the differences between all the methods were statistically significant for the precision.

For recall, the mean value ± SD of the recall of the proposed method was 0.814 ± 0.029, while the values for the pix2pix_unet method, h_dense_unet method, and u-net

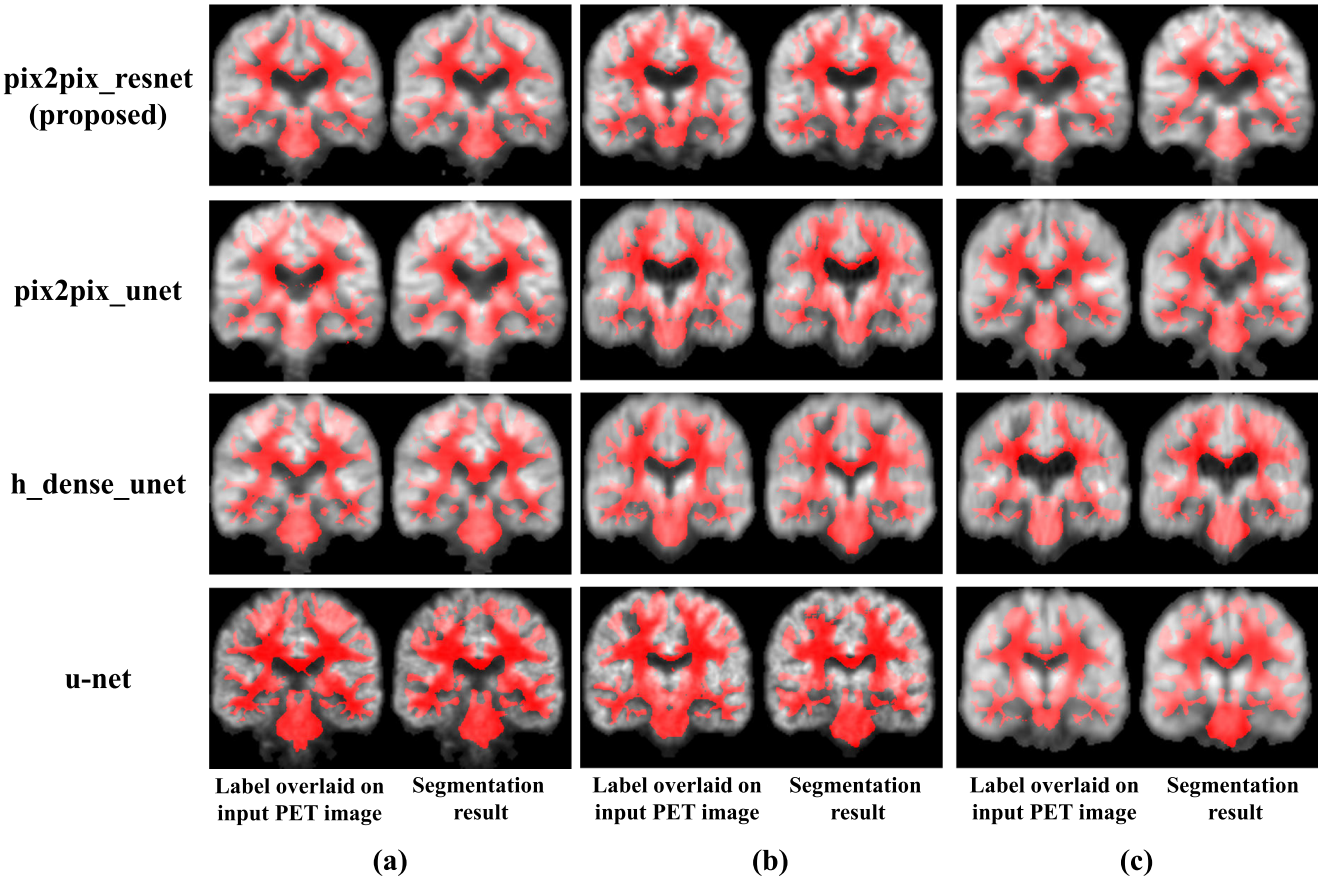

Fig. 3 White matter ground truth and segmentation results of various methods. **a** Most accurate case. **b** Case with median DSC value. **c** Least accurate case

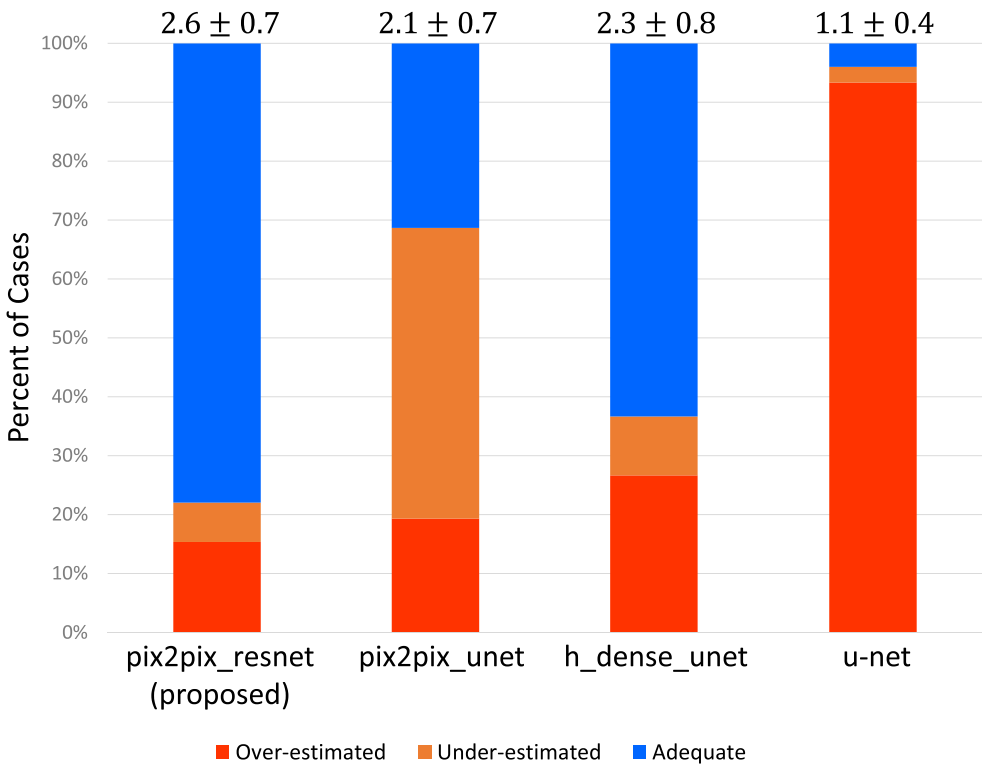

Fig. 4 Segmentation quality scores (1 = over-estimated, 2 = under-estimated, 3 = adequate; mean scores and standard deviation of all readings displayed at top of each bar) assigned by the five observers

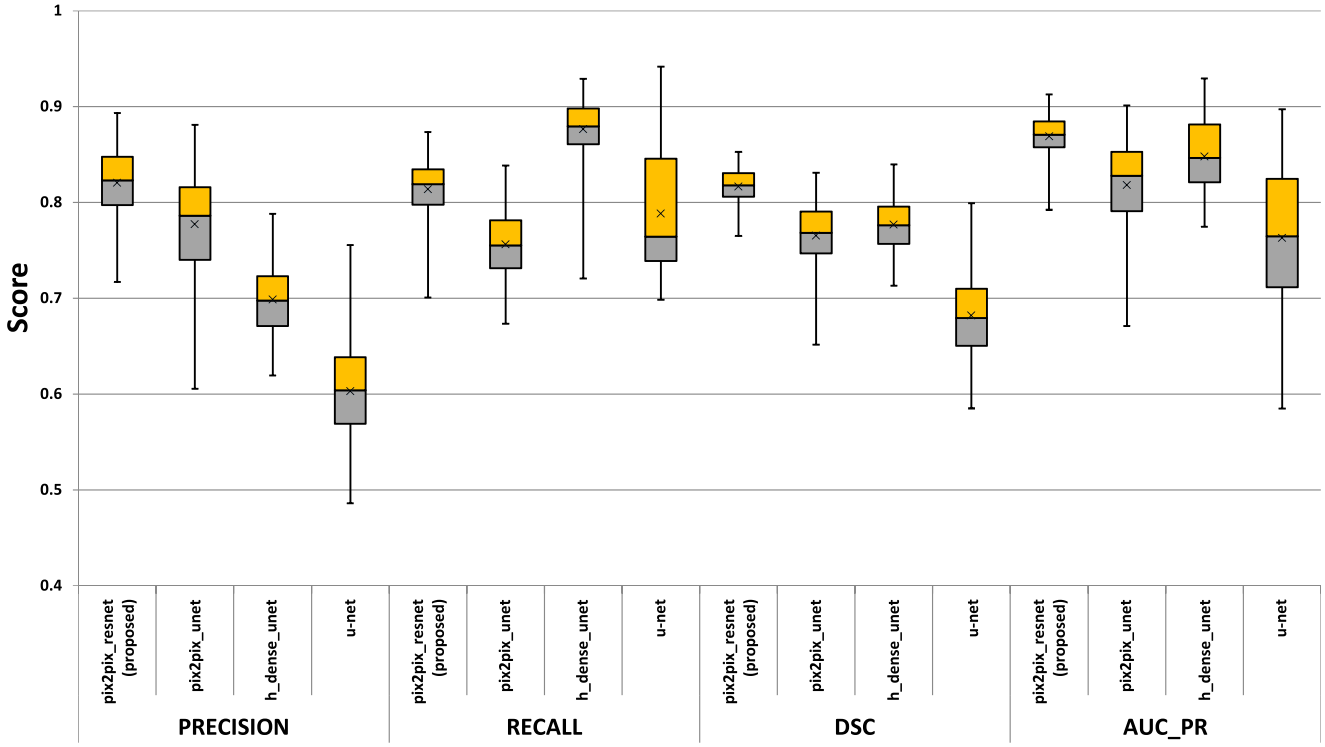

Fig. 5 Boxplot of evaluation parameters between various methods

method were 0.756 ± 0.029, 0.877 ± 0.029, and 0.789 ± 0.062, respectively($p < 0.0001$). There was a statistically significant difference between all the methods for the recall.

For the DSC, the mean value ± SD was 0.817 ± 0.018 in the proposed method, 0.766 ± 0.034 in the pix2pix_unet method, 0.777 ± 0.028 in the h_dense_unet method, and 0.682 ± 0.044 in the u-net method ($p < 0.0001$). There was a statistically

Table 2 Comparison of mean difference between groups using Kruskal–Wallis test

Evaluation parameter	Methods	Mean rank	df	*p* value	Kruskal–Wallis statistic
Precision	PR	674.85	4	<0.0001	644.0
	PU	548.81			
	HU	330.61			
	UN	119.74			
PR vs PU	PR vs HU	PR vs UN	PU vs HU	PU vs UN	HU vs UN
Yes	Yes	Yes	Yes	Yes	Yes
Recall	PR	440.75	4	< 0.0001	456.8
	PU	206.17			
	HU	692.14			
	UN	334.95			
PR vs PU	PR vs HU	PR vs UN	PU vs HU	PU vs UN	HU vs UN
Yes	Yes	Yes	Yes	Yes	Yes
Dice	PR	690.01	4	< 0.0001	571.1
	PU	401.79			
	HU	453.70			
	UN	128.50			
PR vs PU	PR vs HU	PR vs UN	PU vs HU	PU vs UN	HU vs UN
Yes	Yes	Yes	No	Yes	Yes
AUC-PR	PR	618.55	4	< 0.0001	329.4
	PU	360.28			
	HU	486.20			
	UN	208.98			
PR vs PU	PR vs HU	PR vs UN	PU vs HU	PU vs UN	HU vs UN
Yes	Yes	Yes	Yes	Yes	Yes

**PR*, pix2pix_resnet method (proposed), *PU*, pix2pix_unet method, *HU*, h_dense_unet method, *UN*, u-net method

significant difference between all the methods except (pix2pix_unet vs h_dense_unet) for the DSC.

For AUC-PR, the mean value ± SD of the recall of the proposed method was 0.869 ± 0.021, while the values for the pix2pix_unet method, h_dense_unet method, and u-net method were 0.819 ± 0.048, 0.848 ± 0.038, and 0.763 ± 0.072, respectively($p < 0.0001$). Like other parameters, there was a statistically significant difference between all the methods for the AUC-PR. Figure 6 shows the precision-recall curve using the test set for each model. In Fig. 6, it can be seen that the proposed method showed the best performance, followed by h_dense_unet, pix2pix_unet and u-net.

Discussion

For segmentation of brain compartment, conditional GAN with pix2pix framework was used to generate a segmentation map of the white matter compartment on 18F-FDG PET/CT images. This method has the advantage that it works very strongly when paired data is prepared. We also compared proposed method with the other deep learning method using visual analysis and different evaluation parameters.

For the visual analysis, five observers assigned a segmentation quality score. The higher score means the segmentation quality is adequate. The proposed method achieved the highest score and showed the best segmentation result. The pix2pix_unet method under-estimated white matter region in most of the 18F-FDG PET/CT image and achieved lower score. The h_dense_unet method achieved better score than the pix2pix_unet method by scoring "adequate" more than the pix2pix_unet. The u-net method achieved the lowest score by over-estimating the white matter region.

Of the different evaluation parameters, DSC can have a value from 0 to 1, and closer to 1 meant that the segmentation map generated by the model was similar to the ground truth by having fewer false positives and false negatives. The proposed method achieved the highest DSC and showed the best segmentation result.

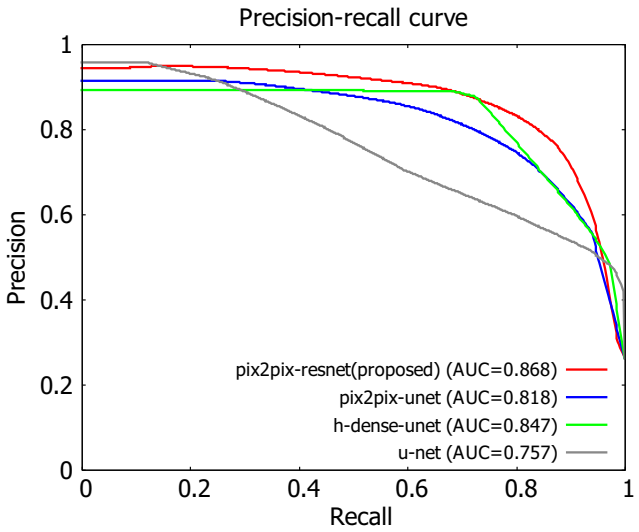

Fig. 6 Precision-recall curve of various methods

The h_dense_unet method achieved high score in recall but it scored low in precision. This means that the h_dense_unet method segments not only the white matter region but also the non-white matter region. Consistent with this, in the segmentation quality scores, the over-estimated ratio is quite high in h_dense_unet method. Nevertheless, unlike the u-net method, the white matter region was well segmented, and the DSC and AUC-PR were calculated to be high.

Since MRI clearly shows the anatomical information of the brain structure, many researches have segmented the brain structure on MRI. The segmentation of the white matter using intensity-based and statistical-based k-means methods was 0.714 and 0.808, respectively. The segmentation results using intensity-based and statistical-based on the fuzzy c-means method were 0.79 and 0.864, respectively [23]. Recent research results have reported that the method using deep learning outperforms prior methods and classical machine learning algorithms. For the classical machine learning algorithms using support vector machine (SVM) and random forest (RF) classifiers scores 0.769 and 0.831, respectively [24]. In contrast, the method using convolutional neural network (CNN) achieved a dice score of 0.864 and the method using multi-fully convolutional networks (mFCNs) achieved 0.887 [25]. However, the dice score of the proposed method was 0.817, which was quite good considering the low resolution of 18F-FDG PET/CT.

AUC-PR was used as a statistical value when comparing the performance of different algorithms [26]. AUC-PR can have a value from 0 to 1, and a higher value meant that the algorithm had better performance. The proposed method scored the highest AUC-PR. We also compared precision-recall curve of the different methods used in this study. A precision-recall curve closer to (1,1) in the coordinates meant better algorithm performance. As a result, the proposed method showed the best performance, followed by pix2pix_unet method and u-net method.

Unlike other parameters, recall confirmed that the pix2pix_unet method was calculated to be lower than that of the u-net method. Recall represented how well the model segmented the actual white matter region. The low value of recall means that the white matter region segmented by the model is smaller than the real white matter region. In Fig. 3, u-net method segments white matter more than ground truth and recall value is calculated to be high. However, segmentation results using the u-net method have many false positives that result in low precision.

18F-FDG PET/CT images were co-registered with MRI during preprocessing. This is because MRI shows more accurate anatomical indices than 18F-FDG PET/CT. The reason for co-registration was that the images obtained from different

modalities might have the same anatomical region, but the coordinate of the region might be different due to different geometrical scaling. If these coordinates were different, accurate segmentation results cannot be obtained. In addition, when the 18F-FDG PET/CT image was co-registered with MRI, the voxel value deviating from the brain region was replaced with zero. The reason for this was to prevent the GAN models from being trained in regions where the brain region was not mapped.

GAN model was trained to segment only white matter among brain structures. The volume change of white matter has been reported in aging, psychosis, and multiple sclerosis [5, 27, 28]. Also, white matter changes were observed in patients with Alzheimer's disease with extensive gray matter atrophy [29]. More importantly, white matter hyperintensities (WMH) have been associated with increased risk of vascular dementia and decreased cognitive abilities [30]. So far, the quantitative access of WMH is possible with only MRI. In this study, we were able to segment the white matter with relatively low information density by removing the cortex regions in 18F-FDG PET/CT. The information on metabolic volume change of the white matter extracted from 18F-FDG PET/CT may have potential values for the quantitative evaluation of various brain diseases associated with white matter volume change. In addition, other deep learning methods for the purpose of image-to-image translation to create WMH on FLAIR T2 images from our segmented white matter images on FDG PET/CT will help assess subcortical white matter-related changes related to vascular dementia.

Conclusions

In this paper, we used conditional GAN with pix2pix framework to generate a segmentation map for the white matter compartment in 18F-FDG PET/CT images. The segmentation results of the proposed method showed excellent performance mimicking the ground truth images of MRI compared with several commonly used deep learning methods. Further studies are needed to elucidate the clinical implications of FDG PET/CT based white matter segmentation in brain research.

Acknowledgments This research was supported by the Brain Research Program through the National Research Foundation of Korea(NRF) funded by the Ministry of Science and ICT (NRF-2018M3C7A1056898).

References

1. Wang Y, Wang Y, Zhang Z, Xiong Y, Zhang Q, Yuan C, Guo H: Segmentation of gray matter, white matter, and CSF with fluid and white matter suppression using MP2RAGE. J Magn Reson Imaging 48(6):1540–1550, 2018
2. LeCun Y, Bengio Y, Hinton G: Deep learning. Nature 521(7553): 436–444, 2015
3. Guha Roy A, Conjeti S, Navab N, Wachinger C: QuickNAT: A fully convolutional network for quick and accurate segmentation of neuroanatomy. NeuroImage 186:713–727, 2019
4. Akkus Z, Galimzianova A, Hoogi A, Rubin DL, Erickson BJ: Deep learning for brain MRI segmentation: State of the art and future directions. J Digit Imaging 30(4):449–459, 2017
5. Walterfang M, McGuire PK, Yung AR, Phillips LJ, Velakoulis D, Wood SJ, Suckling J, Bullmore ET, Brewer W, Soulsby B, Desmond P, McGorry PD, Pantelis C: White matter volume changes in people who develop psychosis. Br J Psychiatry 193(3):210–215, 2008
6. Yi X, Walia E, and Babyn P: Generative adversarial network in medical imaging: A review. arXiv preprint arXiv vol. 1809.07294 2018
7. Mondal AK, Dolz J, and Desrosiers C: Few-shot 3D multi-modal medical image segmentation using generative adversarial learning
8. Dai W, Dong N, Wang Z, Liang X, Zhang H, and Xing EP: Scan: Structure correcting adversarial network for organ segmentation in chest x-rays.. Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support. Springer, Cham, 2018, pp.263–273
9. Izadi S, Mirikharaji Z, Kawahara J, and Hamarneh G: Generative adversarial networks to segment skin lesions. 2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018). IEEE, pp.881–884. 2018
10. Isola P, Zhu JY, Zhou T, and Efros AA: Image-to-image translation with conditional adversarial networks. in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017, pp. 1125–1134
11. Kohl S, Bonekamp D, Schlemmer HP, Yaqubi K, Hohenfellner M, Hadaschik B, and Maier-Hein K: Adversarial networks for the detection of aggressive prostate cancer
12. Meltzer CC, Leal JP, Mayberg HS, Wagner, Jr HN, Frost JJ: Correction of PET data for partial volume effects in human cerebral cortex by MR imaging. J Comput Assist Tomogr 14(4):561–570, 1990
13. Petrella JR, Coleman RE, Doraiswamy PM: Neuroimaging and early diagnosis of Alzheimer disease: A look to the future. Radiology 226(2):315–336, 2003
14. Wang Y, Yu B, Wang L, Zu C, Lalush DS, Lin W, Wu X, Zhou J, Shen D, Zhou L: 3D conditional generative adversarial networks for high-quality PET image estimation at low dose. NeuroImage 174:550–562, 2018
15. Choi H, Lee DS: Generation of structural MR images from amyloid PET: Application to MR-less quantification. J Nucl Med 59(7): 1111–1117, 2018
16. Blanc-Durand P, Van Der Gucht A, Schaefer N, Itti E, Prior JO: Automatic lesion detection and segmentation of 18F-FET PET in gliomas: A full 3D U-Net convolutional neural network study. PloS One 13(4):e0195798, 2018
17. Foster B, Bagci U, Mansoor A, Xu Z, Mollura DJ: A review on segmentation of positron emission tomography images. Comput Biol Med 50:76–96, 2014
18. Nie B, Liu H, Chen K, Jiang X, Shan B: A statistical parametric mapping toolbox used for voxel-wise analysis of FDG-PET images of rat brain. PloS One 9(9):e108295, 2014
19. He K, Zhang X, Ren S, and Sun J: Deep residual learning for image recognition, pp. 770–778
20. Li X, Chen H, Qi X, Dou Q, Fu CW, Heng PA: H-DenseUNet: Hybrid densely connected UNet for liver and tumor segmentation from CT volumes. IEEE Trans Med Imaging 37(12):2663–2674, 2018
21. Ronneberger O, Fischer P, and Brox T: U-net: Convolutional networks for biomedical image segmentation, in International

Conference on Medical Image Computing and Computer-Assisted Intervention, Cham, 2015, pp. 234–241
22. Dice LR: Measures of the amount of ecologic association between species. Ecology 26(3):297–302, 1945
23. Deepa V, Benson CC, Lajish VL: Gray matter and white matter segmentation from MRI brain images using clustering methods. Int Res J Eng Technol (IRJET) 2(8):913–921, 2015
24. Zhang W, Li R, Deng H, Wang L, Lin W, Ji S, Shen D: Deep convolutional neural networks for multi-modality isointense infant brain image segmentation. NeuroImage 108:214–224, 2015
25. Nie D, Wang L, Gao Y, Shen D: Fully convolutional networks for multi-modality isointense infant brain image segmentation. Proc IEEE Int Symp Biomed Imaging 2016:1342–1345, 2016
26. Boyd K, Santos Costa V, Davis J, and Page CD: Unachievable region in precision-recall space and its effect on empirical evaluation. *Proceedings of the ...* International Conference on Machine Learning *International Conference on Machine Learning,* vol. 2012, pp. 349, 2012
27. Beheshti I, Maikusa N, and Matsuda H: Effects of aging on brain volumes in healthy individuals across adulthood. Neurological sciences : official journal of the Italian Neurological Society and of the Italian Society of Clinical Neurophysiology, 2019
28. Sastre-Garriga J, Ingle GT, Chard DT, Cercignani M, Ramio-Torrenta L, Miller DH, Thompson AJ: Grey and white matter volume changes in early primary progressive multiple sclerosis: A longitudinal study. Brain 128(Pt 6):1454–1460, 2005
29. Brown MS, Stemmer SM, Simon JH, Stears JC, Jones RB, Cagnoni PJ, Sheeder JL: White matter disease induced by high-dose chemotherapy: Longitudinal study with MR imaging and proton spectroscopy. AJNR Am J Neuroradiol 19(2):217–221, 1998
30. Habes M, Erus G, Toledo JB, Zhang T, Bryan N, Launer LJ, Rosseel Y, Janowitz D, Doshi J, Van der Auwera S, von Sarnowski B, Hegenscheid K, Hosten N, Homuth G, Volzke H, Schminke U, Hoffmann W, Grabe HJ, Davatzikos C: White matter hyperintensities and imaging patterns of brain ageing in the general population. Brain 139(Pt 4):1164–1179, 2016