Magn Reson Med Sci 2021; 20; 166–174
doi:10.2463/mrms.mp.2019-0199
Published Online: May 11, 2020

MAJOR PAPER

A Robust and Accurate Deep-learning-based Method for the Segmentation of Subcortical Brain: Cross-dataset Evaluation of Generalization Performance

Naoya Furuhashi, Shiho Okuhata, and Tetsuo Kobayashi*

Purpose: To analyze subcortical brain volume more reliably, we propose a deep learning segmentation method of subcortical brain based on magnetic resonance imaging (MRI) having high generalization performance, accuracy, and robustness.

Methods: First, local images of three-dimensional (3D) bounding boxes were extracted for seven subcortical structures (thalamus, putamen, caudate, pallidum, hippocampus, amygdala, and accumbens) from a whole brain MR image as inputs to the neural network. Second, dilated convolution layers, which input information of variable scope, were introduced to the blocks that make up the neural network. These blocks were connected in parallel to simultaneously process global and local information obtained by the dilated convolution layers. To evaluate generalization performance, different datasets were used for training and testing sessions (cross-dataset evaluation) because subcortical brain segmentation in clinical analysis is assumed to be applied to unknown datasets.

Results: The proposed method showed better generalization performance that can obtain stable accuracy for all structures, whereas the state-of-the-art deep learning method obtained extremely low accuracy for some structures. The proposed method performed segmentation for all samples without failing with significantly higher accuracy ($P < 0.005$) than conventional methods such as 3D U-Net, FreeSurfer, and Functional Magnetic Resonance Imaging of the Brain's (FMRIB's) Integrated Registration and Segmentation Tool in the FMRIB Software Library (FSL-FIRST). Moreover, when applying this proposed method to larger datasets, segmentation was robustly performed for all samples without producing segmentation results on the areas that were apparently different from anatomically relevant areas. On the other hand, FSL-FIRST produced segmentation results on the area that were apparently and largely different from the anatomically relevant area for about one-third to one-fourth of the datasets.

Conclusion: The cross-dataset evaluation showed that the proposed method is superior to existing methods in terms of generalization performance, accuracy, and robustness.

Keywords: *deep learning, segmentation, subcortical brain, cross-dataset evaluation*

Introduction

Volumetric analyses of each subcortical brain structure based on magnetic resonance imaging (MRI) have suggested a relationship between the subcortical brain volume and pathologic state of psychiatric and neurological diseases such as schizophrenia, autism, and Parkinson's disease.[1–4] For example, in individuals with schizophrenia, the subcortical brain volume of structures such as the hippocampus, amygdala, thalamus, accumbens, and pallidum are significantly altered as compared with those of healthy controls.[1,2]

To measure the volume of the subcortical brain based on MRI, each structure must be identified from a whole brain image. However, even when performed by experts, the manual segmentation of 3D MR images is time-consuming and labor intensive. Therefore, many studies use automatic segmentation software such as FreeSurfer (http://surfer.nmr.mgh.harvard.edu/) and Functional Magnetic Resonance Imaging of the Brain's (FMRIB's) Integrated Registration and Segmentation

Graduate School of Engineering, Kyoto University, Kyoto, Japan

*Corresponding author: Graduate School of Engineering, Kyoto University, A-Cluster, Kyoto-daigaku-katsura, Nishikyo-ku, Kyoto, Kyoto 615-8510, Japan. Phone: +81-75-383-2228, Fax: +81-75-383-2228, E-mail: kobayashi.tetsuo.2c@kyoto-u.ac.jp

Received: December 25, 2019 | Accepted: April 11, 2020

Tool in the FMRIB Software Library (FSL-FIRST; https://fsl.fmrib.ox.ac.uk/fsl/fslwiki/FIRST/).[1,2,4]

Compared with FreeSurfer, FSL-FIRST obtains higher segmentation accuracy with a smoother shape of the segmented area.[5,6] However, previous studies[7,8] indicated that FSL-FIRST produces segmentation results that are apparently and largely different from the anatomically relevant area when it is used on brains with shapes that are largely different from the standard brain in MNI space (we call these cases "failures"). In fact, the volumetric meta-analysis of subcortical brain[2] excluded some samples that FSL-FIRST failed to properly segment.

Recently, image analysis techniques using deep learning have been studied. Deep learning is a method of machine learning based on multilayer neural networks, and it is good for recognizing images, sounds, and natural languages. The convolutional neural network also reduces the calculation cost by reducing the number of learning parameters of the neural network to relatively few. Additionally, it processes input while maintaining a positional relationship of image elements.[9,10] Deep learning has also been used for image analysis for MR brain images, for classification of psychiatric and neurological diseases,[11,12] detection of cerebral microbleeds,[13] and segmentation of brain tumors and structures.[14–18]

It is important to evaluate the generalization performance of the technique used for subcortical structure segmentation based on MR brain images. Nevertheless, most studies[15–17] do not evaluate cross-dataset performance using plural supervised datasets. Although a previous study[18] showed state-of-the-art accuracy of subcortical brain segmentation when training and testing in the same dataset, it also showed that there was less accuracy in cross-dataset evaluation. This was because the intensity distribution of images obtained by different MRI scanners greatly differed, and the output distributions of the intermediate layer of the neural network that took them as input were largely different.[18] Accordingly, the deep learning segmentation methods in the previous studies have not been used for analyses of large datasets because of their limited generalization performance.

However, there are some ways to improve generalization performance. A commonly used data augmentation technique increases the number of samples by randomly enlarging, reducing, rotating, or adding noise to the image to provide sample diversity.[10] Transfer learning is a technique in which a small target dataset is adapted on the basis of a model that was already learned with another dataset.[19,20] However, the technique requires at least one target datum with ground truth, which is difficult to create for 3D medical image segmentation such as MR images. In addition, a considerable amount of training data is required when performing transfer learning on a large dataset composed of a plurality of datasets obtained by many MRI scanners.

In this study, we proposed a method using a new neural network based on MRI that enables high accuracy and generalization performance for the segmentation of seven subcortical structures (thalamus, putamen, caudate, pallidum, hippocampus, amygdala, and accumbens) of the human brain. Moreover, the method is so robust to abnormal brain shape that it segments all samples without failing.

Using this method, we evaluated cross-dataset generalization performance using the local 3D bounding box containing each structure instead of a whole brain image as input. As a result, generalization performance and robustness were improved because the neural network was less susceptible to brain shape and image quality. In addition, the neural network processes global and local information by dilated convolution[21] in parallel for accurate segmentation.

Materials and Methods

Proposed method

Bounding box extraction

We extracted a local bounding box image containing each structure from a whole brain image. We targeted seven subcortical brain structures: the thalamus, putamen, caudate, pallidum, hippocampus, amygdala, and accumbens (Fig. 1).

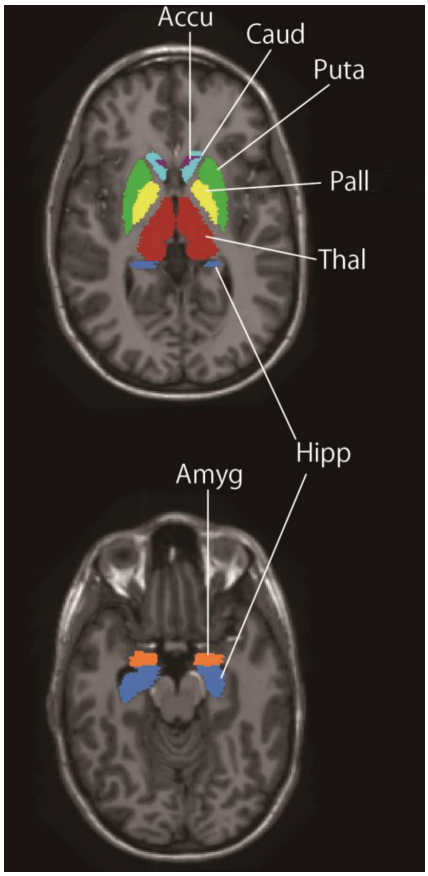

Fig. 1 Seven subcortical structures (accumbens, caudate, putamen, pallidum, thalamus, hippocampus, and amygdala) represented on two axial images. Thal, thalamus; Puta, putamen; Caud, caudate; Pall, pallidum; Hipp, hippocampus; Amyg, amygdala; Accu, accumbens.

Other segmentation methods[15,16,18] input a whole brain image into the neural networks by converting a 3D image into multiple 2D images. This is done because the whole 3D brain image would exceed the capacity of a general computer's memory, and converting the image to 2D images reduces memory usage.

Dolz et al.[17] used 3D bounding boxes sampled randomly from larger brain regions as input, but this random technique made learning difficult and inefficient. In our new approach, we extracted the local 3D bounding box contained in each structure, as shown in Fig. 2a. This not only reduced the need for as much computer memory while maintaining the 3D structure but also specified a target structure. It also reduced the variety of the image and contributed to improved generalization performance.

The 3D bounding box of each structure was extracted independently on the basis of the center coordinates of the structures. We specified the center coordinates using FreeSurfer for automation, although we could have done this with visual inspection. The sizes of the 3D bounding boxes for each structure were determined to match the size of the structure by using the number of axial × coronal × saggital voxels. The results were as follows: thalamus, 40 × 48 × 48; putamen, 40 × 56 × 48; caudate, 32 × 64 × 56; pallidum, 32 × 32 × 32;

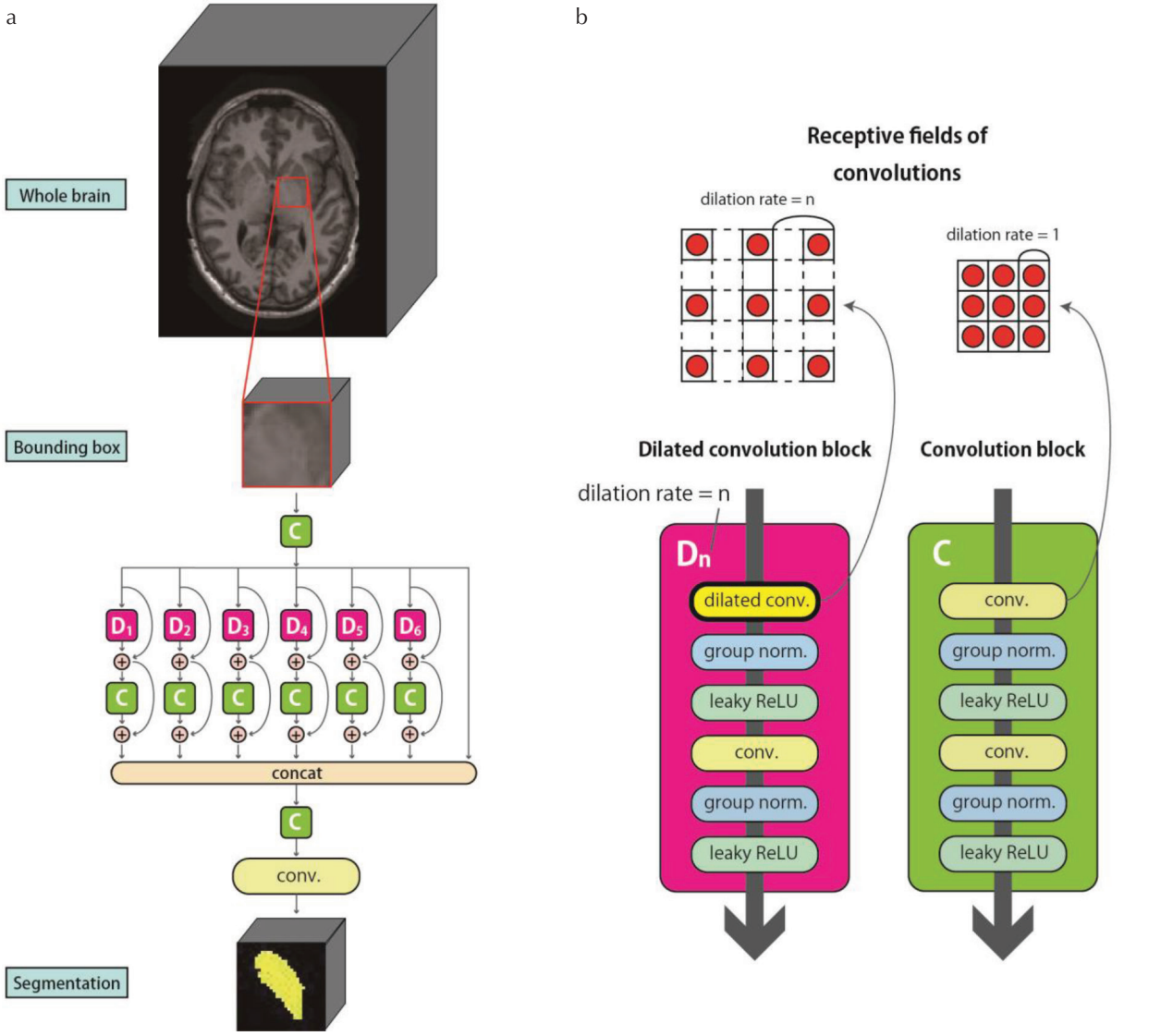

Fig. 2 (**a**) Flow chart of proposed method. (**b**) Dilated convolution block, convolution block, and their receptive fields. The blocks in (**b**) are connected as shown in (**a**). The dilated convolution block, represented as D_n, has the dilated convolution with dilation rate set to n. Concat represents a concatenation of inputs. The input image is processed by one convolution block and then input to the six dilated convolution blocks, the immediate summation, and the concat layer. Arrays connected in the concat layer are processed in the convolution block and convolution layer, and a segmentation image is output. The dilation rates of the dilated convolutions are set to 1–6. The filter size of the intermediate convolution layers is 3 × 3 × 3 and that of the last convolution layer is 1 × 1 × 1. The number of channels of intermediate layers is 32.

hippocampus, 48 × 48 × 56; amygdala, 32 × 24 × 32; and accumbens, 24 × 24 × 24.

We used this method to determine the sizes of the 3D bounding boxes because imbalance in the number of segmentation labels would impede neural network learning. The small variety of location, volume, and angle of structure in the bounding box does not matter because of the augmentation technique when actual implication of the proposed method to the clinical images.

Neural network

Dilated convolution[21] was used as a part of the convolutional layers that composed the neural network, and it was performed with an arbitrary voxel interval (dilation rate). Receptive fields, shown at the top of Fig. 2b, affect a normal convolution (right) and the dilated convolution (left) with 3 × 3 (3 × 3 × 3 in the case of 3D convolution) filters. In the figure, circles represent the pixels (voxels in the case of 3D images) to be convolved. The larger the dilation rate the convolution has, the wider the receptive field, meaning it can obtain more global information. When the dilation rate is 1, the dilated convolution is equivalent to the normal convolution. In this study, 3D dilated convolutions with different dilation rates were incorporated into the neural network in parallel. The dilation rates were determined by validation session.

As shown in Fig. 2, the neural network was built on the basis of blocks that consisted of multiple layers. The network introduced shortcut connections[22] that made it easy to propagate gradient in a backward manner. In addition, we used group normalization[23] to normalize the input and suppress the output from becoming too small or too large. This facilitated learning by preventing gradient vanishing or explosion. A leaky rectified linear unit,[24] applied after linear transformation, was used for the activation function. We made the dilated convolution block and convolution block shown in Fig. 2b. They were connected in parallel as shown in Fig. 2a, and thus, global and local information were obtained by a plurality of dilated convolution blocks that were simultaneously processed. Subsequently, they were integrated via the concat layer.

MRI datasets

To validate the proposed method, we used supervised and unsupervised MRI datasets.

Supervised datasets

First, we applied our method to two supervised MRI datasets, which consisted of both T_1-weighted and ground truth images for segmentation of seven subcortical brain structures.

The first dataset was the Internet Brain Segmentation Repository (IBSR; https://www.nitrc.org/projects/ibsr/). This consists of images of 18 healthy subjects (14 males and four females, 7–71 years old) and expert-labeled segmentations of their subcortical brains. T_1-weighted images were obtained by 1.5T MRI scanner (Sonata, Siemens, Munich, Germany; Signa, General Electric, Boston, MA, USA). The voxel sizes were 0.9375 × 0.9375 × 1.5, 0.837 × 0.837 × 1.5 mm^3, and 1.0 × 1.0 × 1.5 mm^3.

The second dataset was the Medical Image Computing and Computer Assisted Intervention (MICCAI) Multi-Atlas Labeling Challenge 2012 (https://my.vanderbilt.edu/masi/workshops/). This consists of images of 35 healthy subjects (13 males and 22 females, 19–90 years old) and expert-labeled segmentations of their subcortical brains. The subjects were divided into two groups of 15 and 20 people each to evaluate generalization performance using the MICCAI Multi-Atlas Labeling Challenge 2012. T_1-weighted images were obtained by 1.5T MRI scanner (Vision, Siemens) using magnetization-prepared rapid acquisition of gradient echo (MPRAGE). The voxel size was 1.0 × 1.0 × 1.0 mm^3.

Unsupervised datasets

The two unsupervised MRI datasets consisted of T_1-weighted images without ground truth images for segmentation of subcortical brain structures. These unsupervised datasets were used to qualitatively evaluate whether proposed segmentation could be performed for more data obtained by various MRI scanners.

The first dataset was the Open Access Series of Imaging Studies 3 (OASIS-3; https://www.oasis-brains.org). It consisted of images of 609 healthy controls and 489 Alzheimer's disease patients (487 males and 611 females, 42–95 years old) and was collected at The Charles F. and Joanne Knight Alzheimer Disease Research Center (Department of Neurology at Washington University School of Medicine). The total number of images was 2057 because most of the subjects underwent multiple image acquisitions. T_1-weighted images were obtained by 1.5T MRI scanners (Vision or Sonata, Siemens) using MPRAGE. The voxel sizes were 1.0 × 1.0 × 1.3 and 1.0 × 1.0 × 1.0 mm^3. In addition, T_1-weighted images are obtained by 3T MRI scanners (Tim Trio, BioGraph mMR PET-MR, Siemens) using MPRAGE. The voxel sizes were 1.0 × 1.0 × 1.0 and 1.2 × 1.1 × 1.1 mm^3.

The second dataset was the Information eXtraction from Images (IXI; https://brain-development.org/ixi-dataset/), which consisted of images of 619 healthy subjects (277 males and 342 females, 20–107 years old). The dataset was collected at three hospitals in London: the Hammersmith Hospital using a 3T MRI scanner (Intera; Philips, Amsterdam, The Netherlands), Guy's Hospital using a 1.5T MRI scanner (Gyroscan Intera; Philips), and the Institute of Psychiatry using a 1.5T MRI scanner (Signa; General Electric). The voxel sizes were 0.94 × 0.94 × 1.2 mm^3.

Evaluation metrics

The segmentation accuracy was evaluated by the dice coefficient as in previous studies.[14–18] The dice coefficient is expressed by the following equation:

$$\mathrm{dice}(V_1, V_2) = \frac{2|V_1 \cap V_2|}{|V_1| + |V_2|}$$

where V_1 and V_2 are the two segmentation areas and $|V|$ is the volume of area V. The dice coefficient ranges between 0 and 1. It is 0 when the two areas do not overlap at all and 1 when they overlap completely. In this study, we evaluated the dice coefficient of the ground truth segmentation area and the estimated segmentation area.

Experiments

For all samples, the left and right brain structures were treated without distinction by inverting the left brain structures to the right. This resulted in doubling the number of samples of each structure.

We used a part of the MICCAI dataset, which consisted of the 15 people used in the MICCAI competition to validate hyperparameters such as the dilation rates. The validation data were not used following testing.

For evaluation of the generalization performance, we conducted two cross-dataset experiments.

In the first experiment, to compare our method with two deep learning methods, 3D U-Net[25] and Kushibar,[18] we conducted (1) training with the IBSR dataset and testing with the MICCAI dataset and (2) training with the MICCAI dataset and testing with the IBSR dataset. 3D U-Net[25] is a popular network for 3D segmentation. We input the bounding box to the network. Kushibar's method, based on a simple convolutional neural network inputting 2D whole brain images, is a state-of-the-art method of subcortical brain segmentation when training and testing in the same dataset. We cited the dice coefficient score of the cross-dataset evaluation used in Kushibar.[18] The augmentation was not performed to match the conditions of the experiment.

In the second experiment, to compare our method with the conventional methods using FreeSurfer and FSL-FIRST, we conducted (1) training with the augmented IBSR dataset and testing with the MICCAI dataset and (2) training with the augmented MICCAI dataset and testing with the IBSR dataset. In this case, the training datasets were augmented by randomly enlarging, reducing, and translating images. In addition, Gaussian noise was added to the images.

Furthermore, the proposed method and FSL-FIRST were applied to unsupervised OASIS-3 and IXI datasets to qualitatively evaluate whether segmentation was correctly performed for more data without depending on MRI scanners. For the proposed method, we utilized the preceding model trained with the augmented datasets in the second experiment.

To optimize the neural network, the Adam method[26] was used. The learning rate was set to 1×10^{-5}. A batch size was 2 because of the memory capacity. Training ended when the dice coefficient decreased after 1000 iterations. In these experiments, the TensorFlow (https://www.tensorflow.org/) framework on Python was used. We used a personal computer with Intel Core i7-9750H 2.60 GHz central processing unit (CPU) and a NVIDIA RTX 2060 (6 GB memory) graphics processing unit. Training the proposed network for all structures took about 2 days. In the proposed method, FreeSurfer took about 5 h to segment the structures for extracting bounding boxes from whole brain image on the single core of the CPU, and the proposed network took <1 s to segment each structure from the bounding box. On the other hand, FSL-FIRST took about 5 min to segment all structures from whole brain image on the single core of the CPU, and Kushibar's method was reported to take <5 min.[18]

Results

Comparison with the deep learning methods

Figure 3a shows the dice coefficient of the IBSR dataset segmented by Kushibar's method,[18] 3D U-Net,[25] and the proposed method trained with the MICCAI dataset for seven subcortical brain structures. Kushibar's method showed extremely low dice coefficients, < 0.2 for the thalamus, whereas the proposed method and 3D U-Net showed dice coefficients of ≥ 0.6 for all structures. As a result of the Wilcoxon signed-rank test, the dice coefficients of the proposed method were significantly higher than those of 3D U-Net for the caudate, pallidum, and hippocampus and significantly lower for the accumbens ($P < 0.005$).

Figure 3b shows the dice coefficients of the MICCAI dataset segmented using Kushibar's method,[18] 3D U-Net,[25] and the proposed method trained with the IBSR dataset for seven subcortical brain structures. In this case as well, Kushibar's method showed extremely low dice coefficients, < 0.3 on some structures such as pallidum and accumbens, whereas the proposed method and 3D U-Net showed dice coefficients of ≥ 0.6 for all structures. As a result of the Wilcoxon signed-rank test, the dice coefficients of the proposed method were significantly higher than those of 3D U-Net for the thalamus, putamen, caudate, pallidum, hippocampus, and accumbens and significantly lower for the amygdala ($P < 0.005$).

Wilcoxon signed-rank tests between Kushibar's method and the proposed method or 3D U-Net were not performed because the dice coefficient of each sample found using Kushibar's method was not published.

Comparison with conventional segmentation methods

Figure 3c shows the dice coefficients of the IBSR dataset segmented by FreeSurfer, FSL-FIRST, and the proposed method trained with the augmented MICCAI dataset. Examples of the ground truth segmentation and their estimated segmentation are shown in Fig. 4a. As a result of the Wilcoxon signed-rank test, the dice coefficient of the proposed method was significantly higher than those of FreeSurfer for the thalamus, putamen, caudate, pallidum, accumbens, and all structures as one group ($P < 0.005$). Compared with FSL-FIRST, the dice coefficient of the proposed method was significantly higher for the caudate and accumbens and significantly lower for the thalamus and amygdala ($P < 0.005$). For the IBSR dataset, all samples were segmented by FSL-FIRST without the process failing.

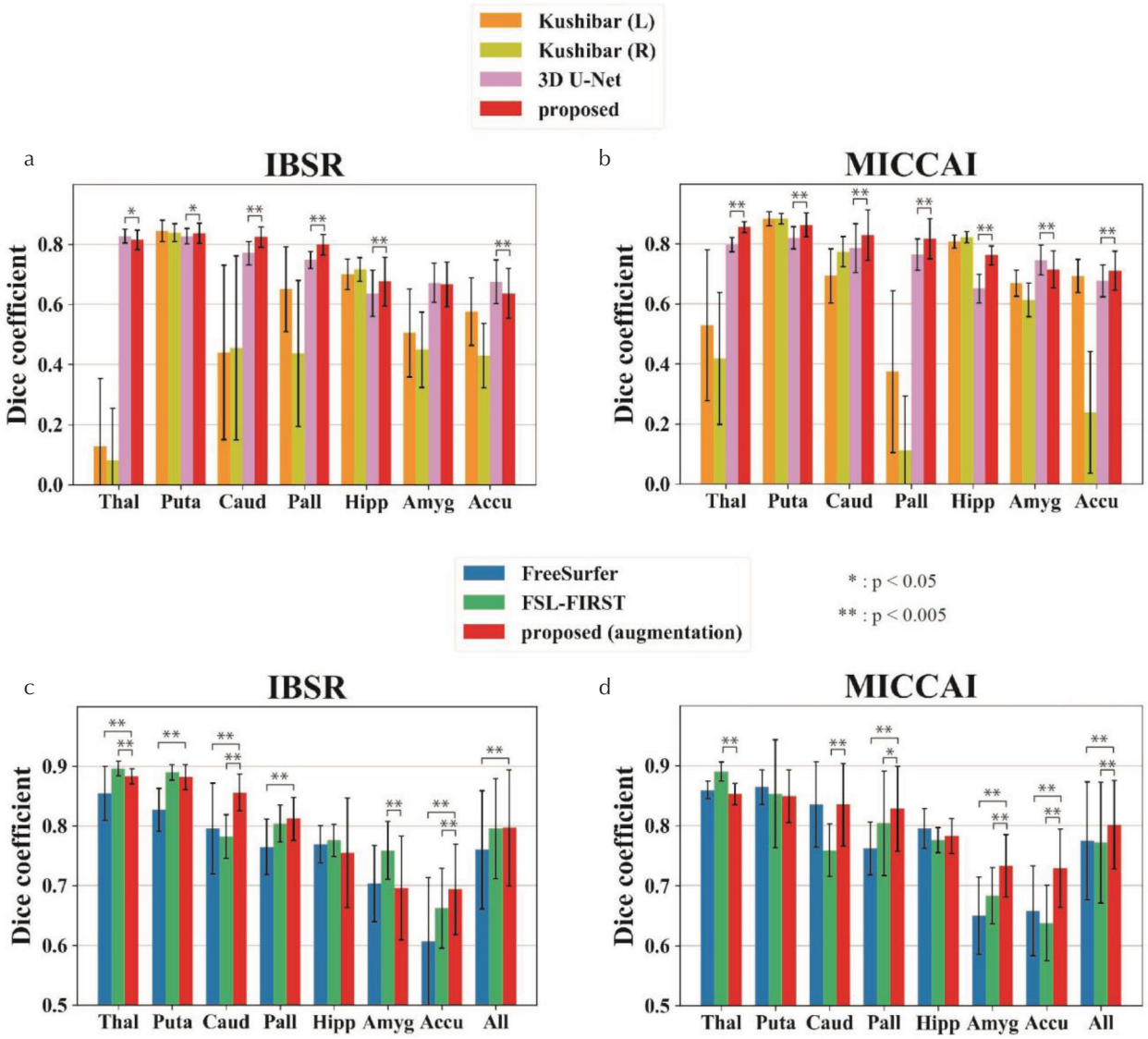

Fig. 3 (**a**) The dice coefficient of the Internet Brain Segmentation Repository (IBSR) dataset segmented by Kushibar's method (L, structures in left hemisphere; R, structures in right hemisphere), 3D U-Net, and the proposed method trained with the MICCAI dataset. (**b**) The dice coefficient of the Medical Image Computing and Computer Assisted Intervention (MICCAI) dataset segmented by Kushibar's method, 3D U-Net, and the proposed method trained with the IBSR dataset. (**c**) The dice coefficient of the IBSR dataset segmented by FreeSurfer, Functional Magnetic Resonance Imaging of the Brain's Integrated Registration and Segmentation Tool in the FMRIB Software Library (FSL-FIRST), and the proposed method trained with the augmented MICCAI dataset. (**d**) The dice coefficient of the MICCAI dataset segmented by FreeSurfer, FSL-FIRST, and the proposed method trained with the augmented IBSR dataset. Wilcoxon signed-rank tests of the differences between Kushibar's method and the proposed method or 3D U-Net, FreeSurfer, and FSL-FIRST were not performed. Thal, thalamus; Puta, putamen; Caud, caudate; Pall, pallidum; Hipp, hippocampus; Amyg, amygdala; Accu, accumbens.

Figure 3d shows the dice coefficient of the MICCAI dataset segmented by FreeSurfer, FSL-FIRST, and the proposed method trained with the augmented IBSR dataset. As a result of the Wilcoxon signed-rank test, the dice coefficient of the proposed method was significantly higher than that of FreeSurfer for the pallidum, amygdala, accumbens, and all structures as one group ($P < 0.005$). Compared with FSL-FIRST, the dice coefficient of the proposed method was significantly higher for the caudate, amygdala, accumbens, and all structures as one group and was significantly lower in the thalamus ($P < 0.005$). For FSL-FIRST, 13 out of 35 samples were excluded because the segmentation process failed.

Application to unsupervised datasets

Figure 4b shows examples of the segmentation results from the proposed method in the unsupervised OASIS-3 and IXI datasets. The proposed method produced anatomically relevant segmentation results for all the target structures for all samples successfully. On the other hand, FSL-FIRST produced apparently and largely different segmentation from

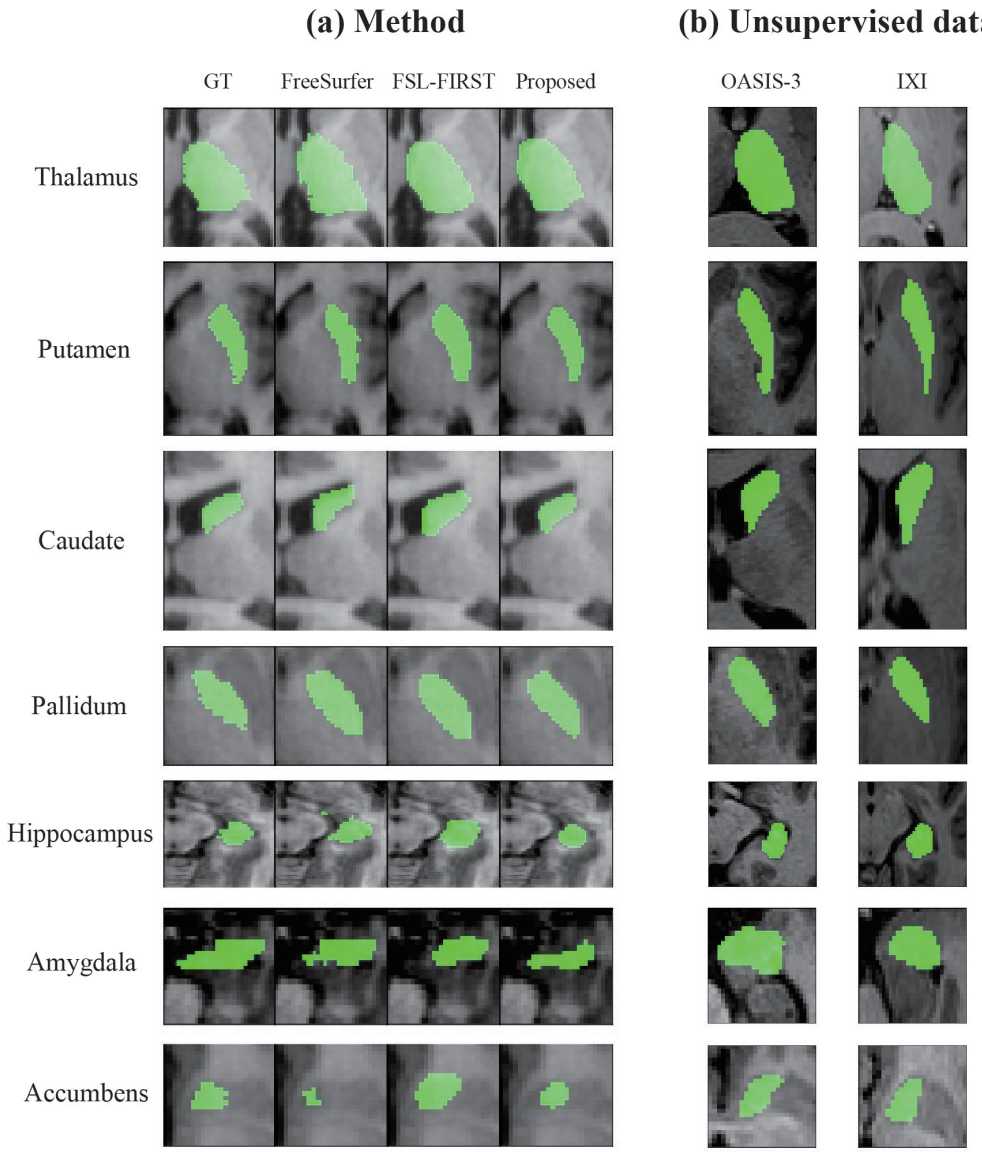

Fig. 4 (**a**) Ground truth (GT), FreeSurfer, Functional Magnetic Resonance Imaging of the Brain's Integrated Registration and Segmentation Tool in the FMRIB Software Library (FSL-FIRST), and proposed method segmentation. (**b**) Proposed method segmentation of unsupervised dataset.

anatomically relevant areas in 488 out of 2057 samples for the OASIS-3 dataset and 190 samples out of 619 images for the IXI dataset. In the "failure" cases, dice coefficients of the segmentation area and ground truth had to be zero.

Discussion

The proposed method and 3D U-Net[25] with bounding boxes as input showed better generalization performance as compared with the state-of-the-art Kushibar's deep learning method using 2D whole brain images as input. Kushibar's method showed extremely low dice coefficient for certain structures. It might be caused by confusion of left and right structures because of the use of the whole brain image as input.[18] On the other hand, the segmentation results of the methods that deal with the bounding box as input showed stable dice coefficients for all structures. There are two reasons why the methods reached significant generalization performance.

First, the proposed method was less susceptible to the differences in brain shape and image quality among datasets because it used local bounding boxes that contained each structure instead of a whole brain image. Second, the proposed method directly input 3D images of bounding boxes without dividing them into multiple 2D images. By doing so, the proposed method maintained 3D information of the structure, and therefore, it is less susceptible to head position during MRI acquisition. Thus, inputting local 3D bounding boxes containing each structure not only reduced required computer memory but also improved generalization performance.

It is important to evaluate generalization performance in tasks that deal with a small amount of data, such as medical images. Moreover, medical images such as MR images greatly vary in quality depending on the type of scanner,

even with the same T_1-weighted images.[18] Therefore, sufficient generalization performance cannot be obtained when the neural network is trained with one small dataset. Namely, the same performance is not expected to be achieved when testing on other datasets. When performing segmentation of medical images, subjects and imaging devices are usually different from those used with training datasets. It is necessary to evaluate cross-dataset generalization performance in cases when the properties of datasets are largely different or when datasets comprise a small number of samples, such as in medical image segmentation.

With the proposed method, highly accurate segmentation was possible. As shown in Fig. 3, dice coefficients of the proposed method were significantly higher than those of 3D U-Net[25] and FreeSurfer for many structures. In addition, those of the proposed method were comparable with those of FSL-FIRST. This was because the global and local information obtained by dilated convolution were processed in parallel and integrated without down sampling. 3D U-Net also integrated the global and local information by skip connection, but its pooling layer reduced information by down sampling. In addition, this also suggests that whole brain information is not necessarily indispensable for segmenting the subcortical structures. The image of the bounding box that includes each target structure may be sufficient for highly accurate segmentation.

The proposed method would be highly robust to the images of various subjects and MRI scanners. There was no sample that failed in anatomically relevant segmentation not only in IBSR and MICCAI but also in larger OASIS-3 and IXI datasets. The robustness is achieved by less variation in the local bounding box images containing each structure and by specifying and reducing the input area. On the other hand, FSL-FIRST failed segmentation in about one-third to one-fourth of the MICCAI, OASIS-3, and IXI datasets. FSL-FIRST tends to fail in segmentation for brain images that are largely different from the standard brain because individual brains are linearly transformed into the standard brain in the process.[7] Because human brains are very different from person to person, a certain number of samples usually fail during the segmentation process of FSL-FIRST. These samples had to be removed from the preceding statistical analyses of brain images, which may have caused an analysis bias. On the basis of these results, the proposed method improved robustness in brain MR image analyses compared with FSL-FIRST.

In this study, we extracted a bounding box for each structure by specifying the center coordinates of the structures using FreeSurfer, which takes a relatively long time for automation, although it was possible on the basis of visual inspection. On the basis of the recently developed neural networks of object detection,[13,27] bounding boxes containing each structure can be detected from the whole brain MRI without utilizing FreeSurfer. The object detection network may be combined with the segmentation network of the present study to realize end-to-end, automated, and fasten segmentation for all subcortical structures.

Conclusion

In this study, we proposed a method that segments seven subcortical structures based on MRI. The network devised for the proposed method input a local 3D bounding box for each structure and processed it via dilated convolutions in parallel. According to cross-dataset evaluations, the proposed method was found to be superior to currently used methods from the three viewpoints of generalization performance, accuracy, and robustness. It enables highly reliable analysis of subcortical brain volume of individuals with psychiatric and neurological diseases.

Acknowledgments

This work was supported by a Grant-in-Aid for Research (15H01813 and 19K12756) from the Ministry of Education, Culture, Sports, Science, and Technology (MEXT), Japan.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

References

1. van Erp TG, Hibar DP, Rasmussen JM, et al. Subcortical brain volume abnormalities in 2028 individuals with schizophrenia and 2540 healthy controls via the ENIGMA consortium. Mol Psychiatry 2016; 21:547–553.
2. Okada N, Fukunaga M, Yamashita F, et al. Abnormal asymmetries in subcortical brain volume in schizophrenia. Mol Psychiatry 2016; 21:1460–1466.
3. Goldman S, O'Brien LM, Filipek PA, Rapin I, Herbert MR. Motor stereotypies and volumetric brain alterations in children with autistic disorder. Res Autism Spectr Disord 2013; 7:82–92.
4. Geevarghese R, Lumsden DE, Hulse N, Samuel M, Ashkan K. Subcortical structure volumes and correlation to clinical variables in Parkinson's disease. J Neuroimaging 2015; 25:275–280.
5. Morey RA, Selgrade ES, Wagner HR, Huettel SA, Wang L, McCarthy G. Scan-rescan reliability of subcortical brain volumes derived from automated segmentation. Hum Brain Mapp 2010; 31:1751–1762.
6. González-Villà S, Oliver A, Valverde S, Wang L, Zwiggelaar R, Lladó X. A review on brain structures segmentation in magnetic resonance imaging. Artif Intell Med 2016; 73:45–69.
7. Feng X, Deistung A, Dwyer MG, et al. An improved FSL-FIRST pipeline for subcortical gray matter segmentation to study abnormal brain anatomy using quantitative susceptibility mapping (QSM). Magn Reson Imaging 2017; 39:110–122.
8. Amann M, Andělová M, Pfister A, et al. Subcortical brain segmentation of two dimensional T1-weighted data sets

with FMRIB's Integrated Registration and Segmentation Tool (FIRST). Neuroimage Clin 2015; 7:43–52.
9. Lecun Y, Bottou L, Bengio Y, Haffner P. Gradient-based learning applied to document recognition. Proc IEEE 1998; 86:2278–2324.
10. Krizhevsky A, Sutskever I, Hinton GE. ImageNet classification with deep convolutional neural networks. Neural Inf Process Syst, Lake Tahoe, 2012; 1097–1105.
11. Pinaya WHL, Gadelha A, Doyle OM, et al. Using deep belief network modelling to characterize differences in brain morphometry in schizophrenia. Sci Rep 2016; 6:38897.
12. Wada A, Tsuruta K, Irie R, et al. Differentiating Alzheimer's disease from dementia with Lewy bodies using a deep learning technique based on structural brain connectivity. Magn Reson Med Sci 2019; 18:219–224.
13. Dou Q, Chen H, Yu L, et al. Automatic detection of cerebral microbleeds from MR images via 3D convolutional neural networks. IEEE Trans Med Imaging 2016; 35:1182–1195.
14. Li H, Li A, Wang M. A novel end-to-end brain tumor segmentation method using improved fully convolutional networks. Comput Biol Med 2019;108:150–160.
15. Shakeri M, Tsogkas S, Ferrante E, et al. Sub-cortical brain structure segmentation using F-CNN'S. IEEE 13th International Symposium on Biomedical Imaging, IEEE, Prague, Czech Republic, 2016; 269–272.
16. Mehta R, Sivaswamy J. M-net: a convolutional neural network for deep brain structure segmentation. IEEE 14th International Symposium on Biomedical Imaging, IEEE, Melbourne, VIC, Australia, 2017; 437–440.
17. Dolz J, Desrosiers C, Ben Ayed I. 3D fully convolutional networks for subcortical segmentation in MRI: a large-scale study. Neuroimage 2018; 170:456–470.
18. Kushibar K, Valverde S, González-Villà S, et al. Supervised domain adaptation for automatic sub-cortical brain structure segmentation with minimal user interaction. Sci Rep 2019; 9:6742.
19. The workshop, Neural Inf Process Syst. Learning to Learn: Knowledge Consolidation and Transfer in Inductive Systems. Available at: http://plato.acadiau.ca/courses/comp/dsilver/NIPS95_LTL/transfer.workshop.1995.html (Published 1995).
20. Pan SJ, Yang Q. A survey on transfer learning. IEEE Trans Knowl Data Eng 2010; 22:1345–1359.
21. Yu F, Koltun V. Multi-scale context aggregation by dilated convolutions. The International Conference on Learning Representations, San Juan, 2016. https://arxiv.org/abs/1511.07122
22. He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, 2016; 770–778. https://arxiv.org/abs/1512.03385
23. Wu Y, He K. Group normalization. European Conference on Computer Vision, Munich, 2018. https://arxiv.org/abs/1803.08494
24. Maas AL, Hannun AY, Ng AY. Rectifier nonlinearities improve neural network acoustic models. International Conference on Machine Learning, Sydney, 2013; 3. https://ai.stanford.edu/~amaas/papers/relu_hybrid_icml2013_final.pdf
25. Çiçek Ö, Abdulkadir A, Lienkamp SS, Brox T, Ronneberger O. 3D U-Net: learning dense volumetric segmentation from sparse annotation. International Conference on Medical Image Computing and Computer-assisted Intervention, Athens, 2016; 424–432. https://arxiv.org/abs/1606.06650
26. Kingma DP, Ba J. Adam: a method for stochastic optimization. The International Conference on Learning Representation, San Diego, 2015. https://arxiv.org/abs/1412.6980
27. Zhao ZQ, Zheng P, Xu ST, Wu X. Object detection with deep learning: a review. IEEE Trans Neural Networks Learn Syst 2019; 30:3212–3232.