Journal of Neuroscience Methods 274 (2016) 146–153

Contents lists available at ScienceDirect

Journal of Neuroscience Methods

journal homepage: www.elsevier.com/locate/jneumeth

Research article

Fast and robust segmentation of the striatum using deep convolutional neural networks

Hongyoon Choi [a,*,1], Kyong Hwan Jin [b,**,1]

[a] Department of Nuclear Medicine, Cheonan Public Health Center, Chungnam, Republic of Korea
[b] Biomedical Imaging Group, École Polytechnique Fédérale de Lausanne (EPFL), 1015 Lausanne, Switzerland

HIGHLIGHTS

- We describe a new method for stratium segmentation.
- We employ two serial deep convolutional neural networks (CNN).
- Segmentation accuracy of deep CNN is comparable with that of previous methods.
- The processing time of the new method is much faster than previous methods.

ARTICLE INFO

Article history:
Received 25 July 2016
Received in revised form
26 September 2016
Accepted 10 October 2016
Available online 21 October 2016

Keywords:
Segmentation
Striatum
Convolutional neural network
Deep learning
MRI

ABSTRACT

Background: Automated segmentation of brain structures is an important task in structural and functional image analysis. We developed a fast and accurate method for the striatum segmentation using deep convolutional neural networks (CNN).
New method: T1 magnetic resonance (MR) images were used for our CNN-based segmentation, which require neither image feature extraction nor nonlinear transformation. We employed two serial CNN, Global and Local
CNN: The Global CNN determined approximate locations of the striatum. It performed a regression of input MR images fitted to smoothed segmentation maps of the striatum. From the output volume of Global CNN, cropped MR volumes which included the striatum were extracted. The cropped MR volumes and the output volumes of Global CNN were used for inputs of Local CNN. Local CNN predicted the accurate label of all voxels. Segmentation results were compared with a widely used segmentation method, FreeSurfer.
Results: Our method showed higher Dice Similarity Coefficient (DSC) (0.893 ± 0.017 vs. 0.786 ± 0.015) and precision score (0.905 ± 0.018 vs. 0.690 ± 0.022) than FreeSurfer-based striatum segmentation ($p = 0.06$). Our approach was also tested using another independent dataset, which showed high DSC (0.826 ± 0.038) comparable with that of FreeSurfer.

Comparison with existing method Segmentation performance of our proposed method was comparable with that of FreeSurfer. The running time of our approach was approximately three seconds.
Conclusion: We suggested a fast and accurate deep CNN-based segmentation for small brain structures which can be widely applied to brain image analysis.

1. Introduction

Accurate segmentation of brain regions from magnetic resonance imaging (MRI) is important in order to find structural changes in various neuropsychiatric disorders (Ross and Tabrizi, 2011; Tinaz et al., 2011; Videbech and Ravnkilde, 2004). Furthermore, the segmentation of the brain structures is crucial in the analysis of functional imaging including positron emission tomography (PET) and functional MRI (fMRI). Segmentation of small brain regions such as striatum is a challenging task because of the variety of their intensity contrast, shape and size. Manual segmentation performed by experts is regarded as a gold standard, though it is time-consuming and impractical to use routinely. Moreover, manual segmentation could show person-to-person inconsistency. Thus, automated methods have been developed to delineate subcortical structures or small nuclei. Numerous automated methods

* Corresponding author at: Department of Nuclear Medicine, Cheonan Public Health Center, 234-1 Buldang-Dong, Seobuk-Gu, Cheonan, Republic of Korea.
** Corresponding author at: Biomedical Imaging Group, École Polytechnique Fédérale de Lausanne (EPFL), 1015 Lausanne, Switzerland.
E-mail addresses: chy1000@gmail.com (H. Choi), kyonghwan.jin@gmail.com (K.H. Jin).
[1] These authors contributed equally to this work.

http://dx.doi.org/10.1016/j.jneumeth.2016.10.007
0165-0270/

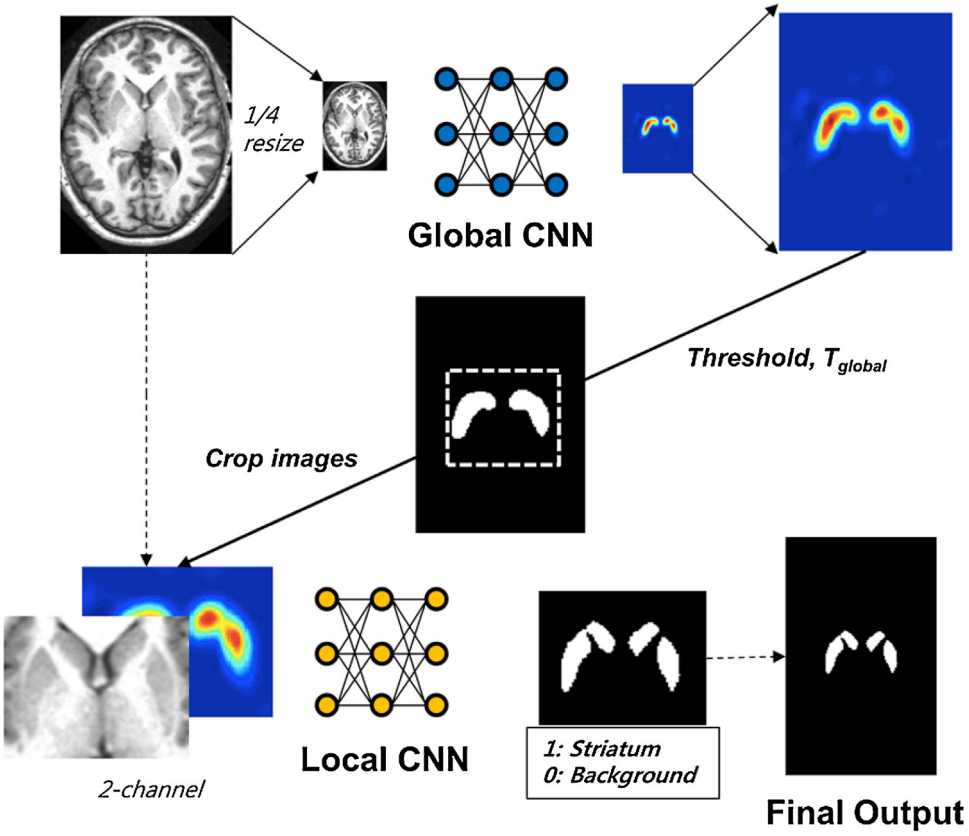

Fig. 1. Overview of the proposed segmentation method. Two serial deep convolutional neural networks (CNN), Global CNN and Local CNN were used for the segmentation of striatum. Reduced size MR volumes were inputted for the Global CNN to predict approximate location of the striatum. The output map of the Global CNN with a threshold value, T_{global}, generated the rough segmentation maps and bounding box. Cropped MR images and the output of Global CNN were used for the input of Local CNN. Local CNN classified the voxels for accurate segmentation.

have used nonlinear registration of a given image to an atlas (Gouttard et al., 2007; Iosifescu et al., 1997; Van Leemput, 2009), however atlas-based methods could not segment details of margins due to a resolution of deformation fields. FreeSurfer software (Fischl, 2012), which provides multiclass segmentation based on Markov random field, also used a probabilistic atlas. Accordingly, it is used as a spatial constraint before the segmentation into classes by Markov random field (MRF). Instead of atlas-based approach, FIRST used Bayesian appearance modeling trained by the voxel intensity and possible shape variations (Patenaude et al., 2011).

We aimed to develop a rapid and robust automated segmentation method for the striatum using deep convolutional neural network (CNN). Deep CNN is a type of machine learning which recently has resulted in stunning performance in the computer vision (Krizhevsky et al., 2012; LeCun et al., 2015). The deep learning approach does not need to model features acquired from data by hand as it can learn the target operator for a specific purpose without prior knowledge. Neither handcrafted feature selection nor complicated image processing such as nonlinear registration is required for the deep CNN. As a result, deep CNN has been applied to medical image segmentation recently. For instance, knee cartilage segmentation using deep CNN showed better performance than previous methods based on handcrafted features (Prasoon et al., 2013). Brain tissues have been accurately segmented into gray matter, white matter and cerebrospinal fluid using deep CNN (Moeskops et al., 2016; Zhang et al., 2015).

In this paper, we describe the CNN-based striatum segmentation approach. Unlike typical brain tissue segmentation, which is usually a global process for all the regions of the brain, voxels of the striatum always occupy very small proportion of the whole brain. Accordingly, the approach should be different from the previous CNN-based brain segmentation approaches (Moeskops et al., 2016; Zhang et al., 2015). Specifically, these previous approaches used a whole image as an input for CNN to segment large regions of interests. Inputting whole image for segmentation of a small region could cause errors due to locally similar regions with the target region as well as unnecessary computation burden. To solve such specific problems, we employed two serial networks called Global CNN and Local CNN. The delicate segmentation process performed by Local CNN could use small portion of the brain image rather than the whole image. Our proposed method was simple for the application and showed very fast and accurate results. Complicated image processing and a priori hypothesis for the segmentation were unnecessary. Because of the strength of deep CNN that can hierarchically learn representations from data, the network architecture we designed could learn features for the segmentation from raw T1 MR images.

2. Materials and methods

2.1. Dataset

Twenty subjects from Open Access Series of Imaging Studies (OASIS) dataset (Marcus et al., 2007) publicly available on Mindboggle-101 (http://www.mindboggle.info) were used (Klein and Tourville, 2012). The dataset includes T1-weighted brain MRI data acquired from 20 healthy subjects (Age 23.4 ± 4.0; 8 males

and 12 females; 20 right-handed). The voxel size of the MR images was $1 \times 1 \times 1$ mm and the matrix size was $160 \times 256 \times 256$. It also includes manually edited cortical labels and manually segmented subcortical labels performed by Neuromorphometrics (http://www.neuromorphometrics.com/). This manually labeled striatum was used as a ground-truth for the training and testing phases. We divided the dataset into two groups, training/validation and test set: 15 MRI data were used for training and validation and the performance of the segmentation was measured from the 5 MRI test dataset.

We exploited another independent dataset, Internet Brain Segmentation Repository (IBSR) data. The dataset included 18 T1-weighted MR images from healthy subjects (Age 38.0 ± 22.4; 14 males and 4 females) with manual segmentations of 43 individual structures. Images and the segmentation maps were provided by the center for Morphometric Analysis at Massachusetts General Hospital (http://www.cma.mgh.harvard.edu/ibsr). The voxel size of the IBSR T1 images was $0.94 \times 1.5 \times 0.94$ mm and the matrix size was $256 \times 128 \times 256$.

2.2. Overview of the approach based on deep CNN

An overview of our approach for segmentation of the striatum is presented in Fig. 1. The striatum was automatically segmented by two serial networks: Global CNN and Local CNN. Firstly, T1-weighted MR images were analyzed by Global CNN to determine approximate location of the striatum. Then, MR images were cropped according to the results of Global CNN. The cropped MR images and the outputs of Global CNN were used for the input volumes of Local CNN to determine whether voxels in the cropped volumes were striatum or not. Accordingly, Global CNN could make an approximation of the segmentation and Local CNN could draw the detailed target structure. We summarize the details of our approach in Algorithm 1.

Algorithm 1. Serial convolutional neural network.

Given notations
(N_x, N_y, N_z) : size of input volume
(B_x, B_y, B_z) : size of bounding box
$*$: 3D convolution
$Z(\cdot)$: zero − padding
$R(\cdot)$: ReLU
$T_\theta(\cdot)$: thresholding by θ. $\left(\text{Default value} : \theta = 0.2\right)$
$D_N(\cdot)$: erasing small clusters less than k (default value : $k = 500$)
$B(\cdot)$: conversion into binary map
$C(\cdot)$: cropping to make bounding box
$A_{ch}(\cdot)$: augmentation along channel direction
$S_{max}(\cdot)$: softmax function
Input
$W_G^i(\cdot)$: i^{th}layer Global CNN
$W_L^i(\cdot)$: i^{th}layer Local CNN
$X \in \mathbb{R}^{N_x \times N_y \times N_z}$: $T1$weighted volume
Output
$O_L \in \mathbb{R}^{N_x \times N_y \times N_z}$: Binary segmentation map
Phase 1–Global CNN
Initial setting
$X \rightarrow X_d \in \mathbb{R}^{N_x/4 \times N_y/4 \times N_z/4}$
$O_0 = X_d$
for $i = 1 : I$ **do**
$\hat{o}_i \longleftarrow Z\left(W_G^i * O_{i-1}\right)$
if $i = I$ **then**
$O_i \longleftarrow \hat{o}_i$
else
$O_i \longleftarrow R\left(\hat{o}_i\right)$
end if
end for
$\hat{o}_G \longleftarrow T_{0.2}(O_I)$
$O_G \longleftarrow D_{500}\left(\hat{o}_G\right), O_G \in \mathbb{R}^{N_x/4 \times N_y/4 \times N_z/4}$
Phase 2 − Local CNN
Initial setting
$B = B(O_G)$
M $\longleftarrow$ upscaling B by trilinear interpolation, $M \in \mathbb{R}^{N_x \times N_y \times N_z}$
$I_X \longleftarrow C(X), I_X \in \mathbb{R}^{B_x \times B_y \times B_z}$
$I_M \longleftarrow C(M), I_M \in \mathbb{R}^{B_x \times B_y \times B_z}$
$O_0 = A_{ch}(I_X, I_M)\ O_0 \in \mathbb{R}^{B_x \times B_y \times B_z} \times 2$,
for $i = 1 : I$ **do**
$\hat{o}_i \longleftarrow Z\left(W_L^i * O_{i-1}\right)$
if $i = I$ **then**
$O_i \longleftarrow S_{max}(\hat{o}_i)$
else
$O_i \longleftarrow R\left(\hat{o}_i\right)$
end if
end for
$O_L(\iota_C) \longleftarrow O_I, \iota_C$ is index set for a bounding box

2.3. Global CNN

As the feature of images could be automatically learned from raw data by CNN, only simple preprocessing steps for MR images were required. MR images were rescaled to a range of 0 to 255. Because Global CNN was aimed to find approximate location of the striatum, we reduced the matrix size of MR images from $160 \times 256 \times 256$ to $40 \times 64 \times 64$. These resized MR volumes were used as input images for the Global CNN.

Target image for the network was a map of manually labeled striatum, where the voxels located at the region of striatum was 1 and those of background was 0. For the training of global CNN, size of target image was also resized into $40 \times 64 \times 64$. Thus, the size of the target image matrix was identical to the input matrix. After the resizing process, Gaussian filter with the $3 \times 3 \times 3$ convolutional kernel was applied. This smoothing process for the target image of Global CNN helps the loss function to converge. We used this processed label map as a target image for the Global CNN.

Global CNN consisted of six 3-D convolutional layers (Fig. 2A). Types of convolutional filters were the same for all the layers: size of the filters was $3 \times 3 \times 3$ and the number of channels was 64. Each convolutional layer was followed by Rectified Linear Unit (ReLU) activation layer except the last convolutional layer. Because the matrix size of the output of convolutional layers is reduced, we applied zero-padding along all 3 dimensional axes. Thus, the size of the input matrix is maintained after series of convolutional layers.

Training of the Global CNN was performed by 13 images out of 15 training/validation set. The remaining 2 images were used for validation. Training phase of the network was aimed to minimize the difference between the output of the network corresponding to a predicted striatum label map and the processed target image corresponding to the ground-truth. The difference was measured by the mean square error, $mean\left[\frac{1}{2}\left(Y - f\left(X;\Theta\right)\right)^2\right]$, where Y is the target image, $f(X;\Theta)$ is the output of the network, and Θ is the learned convolutional neural network for segmentation purpose.

The CNN was implemented using a deep learning library, MatConvNet (Version 1.0-beta 16) (Vedaldi and Lenc, 2015), based on MATLAB. Training was conducted by stochastic gradient descent (SGD) algorithm. SGD is a type of a gradient descent algorithm commonly used to train large networks including deep CNN (Bousquet and Bottou, 2008). At each update of the weights in the SGD algorithm, instead of considering all the training data to compute the gradient of the error function, only a mini-batch of training data is used. In our data, the gradient was computed by a target image and an output of Global CNN to train the weights of the network. $\Delta w_L(t)$ is the update of weights of layer L at iteration t, $\Delta w_L(t) = -\alpha \frac{\partial E}{\partial w_L} + m\Delta w_L(t-1)$, where α and m are the learning rate and the momentum, respectively and E is error function. For the training of Global CNN, the momentum parameter was set to 0.9

and the learning rate was 0.001. 100 epochs of the training process using all the training data were iteratively conducted.

2.4. Input images for local CNN

Local CNN was aimed to segment striatum accurately. Because striatum occupies a small proportion of the whole brain, inputting all voxels of the whole brain into CNN could result in unnecessary computation and inaccurate segmentation. To solve this issue, we cropped the input volume using the output of global CNN. A threshold value, T_{global}, to crop the approximate striatum was applied to the output of Global CNN. T_{global} was set to 0.2, which sufficiently include entire striatum of the training set. In the approximate stratum segmentation map, the clusters smaller than a predefined cluster number threshold ($k_{cluster} = 500$) were regarded as background. A 3-dimensional bounding box which includes the approximate segmentation was drawn. This bounding box was exploited to crop the original MR volume and the output of Global CNN (Fig. 2B). These two cropped volumes were used as the input of Local CNN.

2.5. Architecture of local CNN

The architecture of Local CNN was described in Fig. 2B. It was similar to Global CNN which consisted of six 3-D convolutional layers. Also, size of the filters was $3 \times 3 \times 3$ and the number of channels was 64. As the cropped MR volume and the output of Global CNN were used, input volumes had two channels. For preprocessing, the intensity of the voxels of cropped MR volumes was normalized to have zero mean and one standard deviation. Mean and standard deviation were computed from voxels of each cropped MR image. Zero-padding along all 3 dimensional axes was also applied similar to Global CNN to maintain the size of the matrix after series of convolutional layers. After the last convolutional layer, 64 feature maps were connected to the two feature maps which corresponded to the scores for the striatum and the background. In this layer, a dropout function was applied. This layer randomly drops the connections with predefined probability. Dropout makes the network reduce overfitting and learn more robust features. We set the output of each neuron to zero with probability 0.5. A softmax layer was followed to discriminate two labels. The network minimized the cross entropy loss between the predicted label and the ground truth label made from manually segmented striatum.

For training local CNN, 13 images were also used for the training and 2 images were used for the validation. Local CNN was trained for 100 epochs. The momentum parameter was set to 0.9 and the initial learning rate was 1×10^{-3}. The learning rate was logarithmically decreased to 1×10^{-5}.

2.6. Testing for striatum segmentation

After the training, 5 MR images independent from training/validation set were used for the striatum segmentation. A given T1 MR image was resized to have $64 \times 64 \times 40$ matrix size as input for Global CNN. Using the output of Global CNN and T_{global}, a roughly segmented striatum could be generated. From this segmented area, bounding box for Local CNN was drawn and MR volumes and the output of Global CNN were cropped. After the intensity normalization, Local CNN could generate a segmentation map for the striatum. As previously described, because Global CNN could analyze the approximate margin of the striatum from global contextual features, the number of voxels analyzed by Local CNN was much smaller than those of whole brain.

As another independent dataset, the segmentation process was tested using 18 MR images from IBSR dataset. The size of a given image of IBSR dataset was reduced to 1/4 of the original for Global CNN. Local CNN analyzed the cropped MR images by the generated approximate segmentation map.

2.7. Evaluation

The performance of the segmentation was evaluated by Dice Similarity Coefficient (DSC). The DSC measures the overlap between two different segmentation results (Dice, 1945). It is defined as twice the size of the overlap divided by the sum of sizes of the two segmentation results: $DSC = \frac{2V_{A \cap B}}{V_A + V_B}$ (A : *Automatic segmentation*; B : *Ground Truth*). DSC was measured by the final output of the two serial CNN compared with the manual segmentation result. We also measured precision and recall scores. The precision is the number of true positive voxels divided by the sum of true positives and false positives. The recall is the score of the number of true positive divided by true positives and false negatives (Raghavan et al., 1989). The recall represents sensitivity of the segmentation.

$$\text{Precision} = \frac{V_{A \cap B}}{V_A};\ \text{Recall} = \frac{V_{A \cap B}}{V_B}$$

(A : *Automatic segmentation*; B : *Ground Truth*)

We compared our CNN framework with a publicly available automatic segmentation tool, FreeSurfer (Fischl, 2012). FreeSurfer provides automatic brain segmentation using T1-weighted images. FreeSurfer was conducted with its default options on the T1-weighted volumes of the test set. FreeSurfer automatically segmented subcortical structures based on probabilistic information. DSC was also evaluated for the outputs of FreeSurfer and manual segmentation results. DSC, precision and recall scores were statistically compared using Wilcoxon-signed rank test.

3. Results

An example of the segmentation result was shown in Fig. 3. Global CNN could provide a score map for striatum segmentation (Fig. 3A). The approximate segmentation using the Global CNN with a threshold, T_{global}, could determine the location of the patches for the input of Local CNN. Cropped MR images and the output of Global CNN passed the Local CNN, and the fine-tuned outputs from Local CNN keep the dimensions the same as the original MR image matrix. Total computation time for the segmentation was 3 s under GPU (NVIDIA GeForce Titan X) and 1.5 min under CPU (1.7 GHz Intel Core i5). Training took 2 h for Global and Local CNN under GPU, respectively.

The segmentation results of the proposed method were quantitatively evaluated by difference from the manual segmentation results (Fig. 4). Segmentation results using FreeSurfer were also evaluated and compared with our results. The proposed method showed higher DSC than that from FreeSurfer segmentation results. The difference was statistically a marginal trend toward significance ($p = 0.06$). Mean $\pm$ SD of DSC was 0.893 ± 0.017 and 0.786 ± 0.015 for CNN-based segmentation and FreeSurfer, respectively. The CNN-based approach showed a trend of higher Precision score than FreeSurfer (0.893 ± 0.017 and 0.690 ± 0.022 for CNN and FreeSurfer, respectively; $p = 0.06$). The CNN-based approach showed relatively lower recall score than FreeSurfer segmentation, however, the difference was statistically insignificant (0.882 ± 0.031 and 0.914 ± 0.036 for CNN and FreeSurfer, respectively: $p = 0.31$). It suggested that FreeSurfer segmentation included slightly more true-positive voxels and much more false-positive voxels than CNN-based approach, which resulted in lower DSC.

Segmentation was additionally tested using IBSR dataset as an independent dataset. Though the networks were trained by OASIS

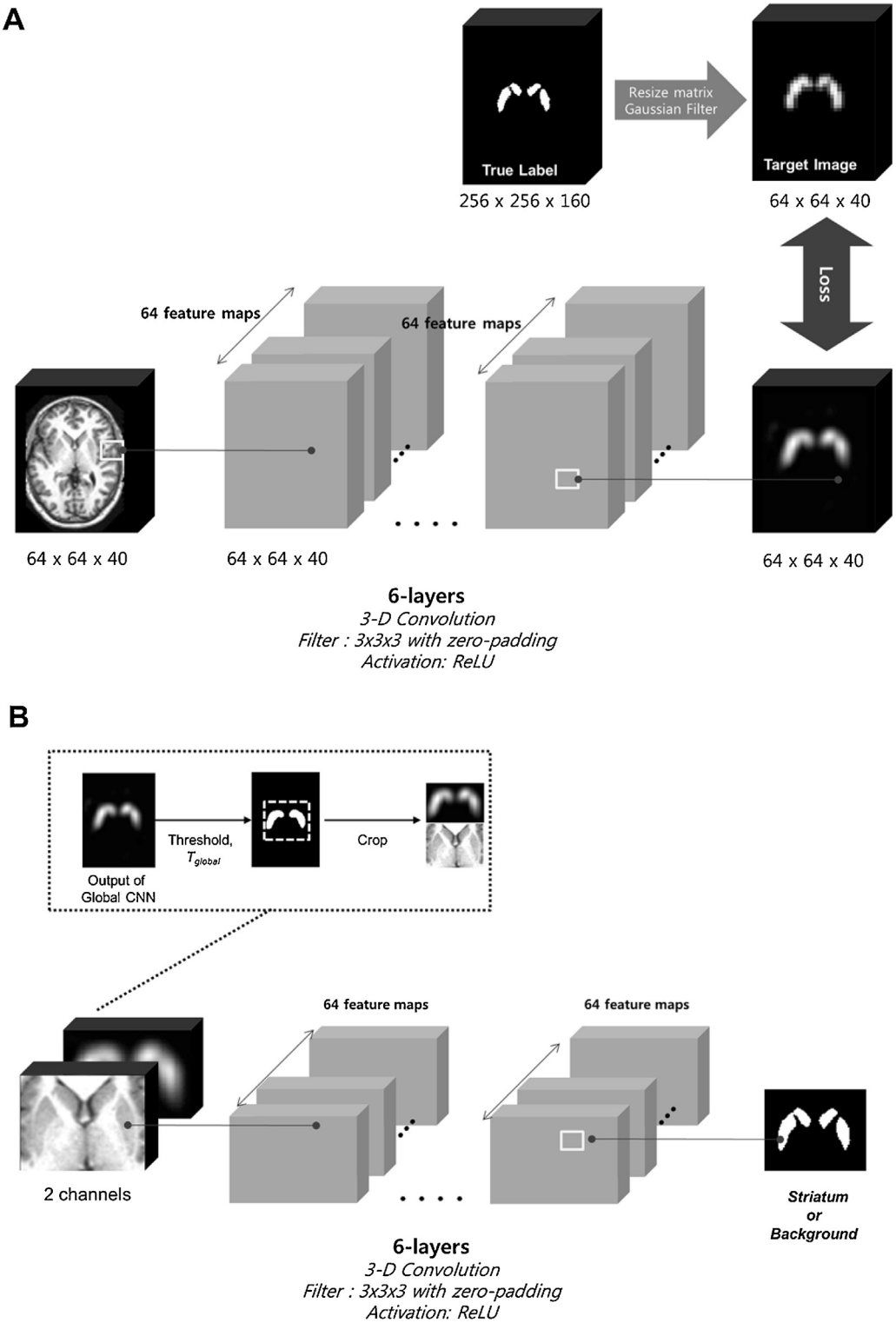

Fig. 2. Network architectures. Two serial networks consisted of 3-dimensional convolutional filters. (A) For Global CNN, 6 pairs of layers, a convolutional filter and Rectified Linear Unit (ReLU) activation, were repeatedly applied. The target of the network was processed manually segmentation map with Gaussian filter. As the size of output was same with the size of input matrix, zero-padding was applied for each layer. (B) Using the output of global CNN, bounding box for the striatum was drawn and MR volumes were cropped. These cropped MR volumes and the output of Global CNN were used for the inputs of Local CNN. The architecture of Local CNN was similar with Global CNN, 6 convolutional layers with ReLU activation. After the serial convolutional layers, softmax function was applied to determine whether a given voxel was included in the striatum or not. Parameters for the layers are summarized in the figure.

dataset which has different MR protocols including image size from IBSR dataset, our approach could generate the segmentation map. DSC of the CNN-based approach was 0.826 ± 0.038, which was not significantly different from that of FreeSurfer (0.827 ± 0.022; $p = 0.76$). Precision score was significantly higher in CNN-based approach (0.917 ± 0.028 vs. 0.790 ± 0.038; $p < 0.001$), while Recall score was significantly lower (0.756 ± 0.066 vs. 0.870 ± 0.049; $p < 0.001$). It suggests that the CNN-based approach showed more

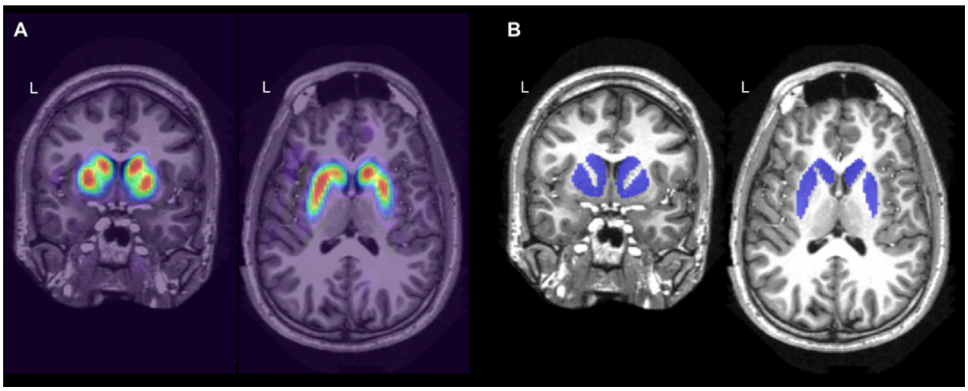

Fig. 3. An example of segmentation results. (A) The output of Global CNN. Global CNN produced a score map for the striatum segmentation as the target of Global CNN was a smoothed volume of manual segmentation map. (B) The output of Local CNN was reconstructed to the segmentation map. (L: left-side).

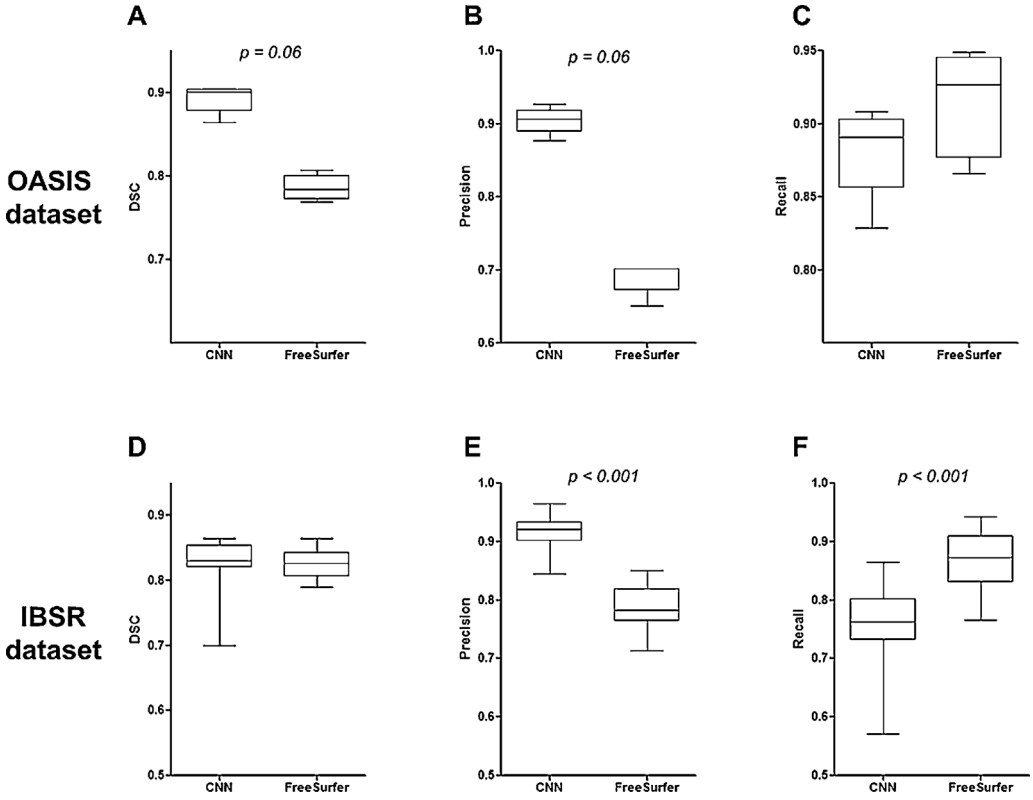

Fig. 4. Quantitative indices of segmentations. Dice Similarity Coefficient (DSC), precision and recall scores were calculated for the test dataset. The results were compared with those of FreeSurfer. Test was performed on the OASIS dataset (A–C) and IBSR dataset (D–E). (A) The CNN-based approach showed a trend of higher DSC than FreeSurfer ($p = 0.06$). (B) The CNN-based approach also showed a trend of higher Precision score than FreeSurfer ($p = 0.06$) (C) Recall score of the CNN-based approach was slightly lower than that of FreeSurfer ($p = 0.31$). (D) CNN-based approach applied to IBSR dataset as an independent dataset from the training data. DSC of the CNN-based approach was comparable with FreeSurfer ($p = 0.76$). (E) Precision score of the CNN-based approach was significantly higher than that of FreeSurfer ($p < 0.001$). (F) Recall score of FreeSurfer was significantly higher than that of the CNN-based approach ($p < 0.001$).

false-negative but less false-positive segmentation compared with FreeSurfer.

An example of segmented striatum in a subject is shown in Fig. 5. As shown in the figure, our CNN-based segmentation was mostly overlapping with the ground-truth segmentation and showed fewer false-positive labeling.

4. Discussion

In this study, we developed a novel framework for the striatum segmentation using deep CNN. The segmentation process was fast, and demonstrated segmentation performance comparable to FreeSurfer. We simply designed the network architecture without feature selections by hands. Image processing such as nonlinear transformation was not required for our approach. The framework was trained and validated by 15 T1 MR raw image data and their manual segmentation maps. DSC of our proposed method was higher than that of FreeSurfer for OASIS dataset and similar with that of FreeSurfer for IBSR dataset. Moreover, the running time for the segmentation was only a few seconds using GPU (GTX TITAN) (approximately, 1.5 min using CPU, 1.7 GHz Intel Core i5), whereas FreeSurfer takes approximately 10 h for one subject.

Our approach provided accurate segmentation and was faster than previous methods including FreeSurfer. When the segmenta-

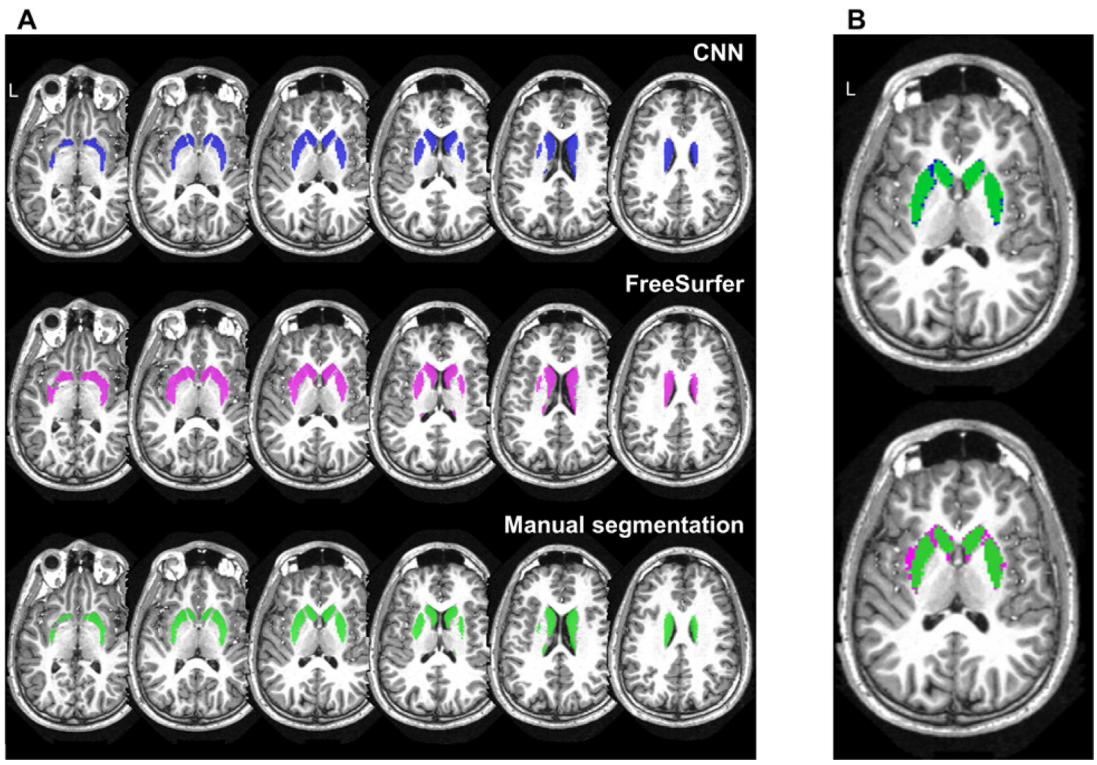

Fig. 5. Comparison of striatum segmentation results. (A) Multiple slice view images present examples of the CNN-based approach, FreeSurfer and manual segmentation. Compared with the CNN-based approach, FreeSurfer showed erroneous classification particularly in the basal area of striatum. (B) Two automated segmentation results were overlapped with the manual segmentation. The CNN-based segmentation showed less erroneously classified voxels than FreeSurfer-based segmentation.

tion was tested using OASIS dataset, DSC and precision score was higher and recall score was slightly lower than those of FreeSurfer, which suggested that the CNN-based method included fewer false-positive labeling. Using IBSR dataset, the segmentation accuracy of our approach was comparable with FreeSurfer, though the architecture was only trained by 13 MR images. Moreover, the CNN-based approach was much faster. Such a fast computation speed could be advantageous to volumetric studies as well as functional studies. For example, fast striatal volume measurement could be useful in a clinical setting as a diagnostic feature of Huntington disease (Rosas et al., 2001). Quantification of dopamine transporter imaging commonly used in diagnosis of movement disorders requires fast and accurate striatal segmentation (Choi et al., 2014; Nobili et al., 2013). Functional connectivity study especially focusing on the corticostriatal connectivity which affected cognitive functions needs accurate striatal segmentation (Ystad et al., 2011). The approach based on CNN-based segmentation might facilitate clinical studies of a large brain imaging cohort as well as several structural and functional brain imaging studies that require accurate segmentation results.

In this study, we used two serial 3-dimensional CNN architectures for the segmentation. Until now, CNN-based segmentation studies have employed patch extraction for the input of the networks (Kleesiek et al., 2016; Moeskops et al., 2016; Zhang et al., 2015). These studies were mainly aimed at skull stripping or brain tissue segmentation. Unlike these studies, as the striatum was a small structure, inputting all the voxels to CNN was unnecessary and could result in errors due to the existence of similar structures to the striatum. In addition, it laid a computation burden. Furthermore, previous segmentation approaches have limitations that each voxel was classified solely by the features of small patches around the voxels. Recently developed deep CNN architectures have exploited global contextual feature for segmentation process as well as the local features (Havaei et al., 2016; Kamnitsas et al., 2015). These combined global and local contextual features could improve overall accuracy as well as reduce computational burden while they exploited large receptive field. Our approach was also similar to these combined network architectures for learning global and local contextual features. Accordingly, Global CNN in our study exploited global contextual features as MR volumes were subsampled. It generated a score map of striatum segmentation and a bounding box to reduce number of voxels for the classification in Local CNN.

Global CNN used a loss function based on mean square error for the regression rather than classification. This approach, minimization of the least squares, was also employed in deep learning application for various types of image processing (Cheng et al., 2015; Dong et al., 2014). Furthermore, 3-dimensional MR images need different methods from recent segmentation methods of 2-dimensional images (e.g. (Long et al., 2015)) due to their complexity originated from scalability. As we used 3-D convolutional filters for both the Global CNN and Local CNN, 3-dimensional information around a given voxel could be used. In short, efficient computation was achieved by Global CNN by reducing the number of voxels for the analysis, and 3-D convolutional filters could help accurately predict true label.

The features of the striatum were automatically learned from the raw data using the network architecture. This automatic feature learning is an advantage of deep neural network. Accordingly, image features or image processing specific to the striatum were not used in our approach. However, as a limitation of our work, various types of network architecture for the segmentation are possible modified from our suggested. As a proof-of-concept study of the 3-D CNN for segmentation of a small brain structure, the architecture of the network was empirically designed by monitoring the errors of the validation set. Adjustments of our proposed network including number of nodes and layers and other settings such as threshold values are possible, though we tested several changes in our proposed architecture. As a future work, optimization of the network

architecture could improve the performance. Our approach could be flexibly used in the segmentation of other small structures. As our architecture and the parameters (e.g. threshold values) were set for the striatum segmentation, the design should be additionally optimized and re-trained for other region-of-interests.

One of the issues regarding deep CNN is number of training samples. In general, CNN needs a large number of training data. Our method did not show better segmentation performance measured by IBSR dataset than FreeSurfer. In addition, while our approach was dependent on the training using manual segmentation, another recently developed automatic segmentation tool based on multimodal images (MIST) did not require a manually labeled training set (Visser et al., 2016), which could be flexible for various types of MR data. It also showed higher accuracy for the striatum segmentation than conventional methods though we could not compare it with our approach as our dataset has only T1-weighted images. Nevertheless, our proposed network was only trained by a limited number of samples (*i.e.* 13 MR volumes). It might limit the performance in different dataset. Further training using various MR data from different data sets could improve the performance by using more degrees of freedoms. Moreover, our approach required T1-weighted images only, while MIST requires multimodal images for the segmentation. Modification and re-training using various dataset could also generalize the segmentation in MR data acquired by different machines and protocols.

5. Conclusion

We developed a novel striatum segmentation framework using deep CNN. The segmentation performance of our method was comparable with conventional approaches such as FreeSurfer, while it was faster. Two serial CNN architectures were used in our approach for efficient segmentation. These two CNN structures are composed of a global feature encoder and a local feature analyzer. As deep CNN could automatically learn features for the segmentation, manual a priori feature extraction was unnecessary. Because of the speed and accuracy of our CNN-based approach, it can be widely applied to various clinical and neuroscience fields combined with multimodal functional images.

Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Competing interests

None declared.

References

Bousquet, O., Bottou, L., 2008. The tradeoffs of large scale learning. Advances in neural information processing systems. 161-168.
Cheng, Z., Yang, Q., Deep, Sheng B., 2015. Colorization. Proceedings of the IEEE International Conference on Computer Vision, 415–423.
Choi, H., Cheon, G.J., Kim, H.J., Choi, S.H., Lee, J.S., Kim, Y.I., Kang, K.W., Chung, J.K., Kim, E.E., Lee, D.S., 2014. Segmentation-based MR attenuation correction including bones also affects quantitation in brain studies: an initial result of 18F-FP-CIT PET/MR for patients with parkinsonism. J. Nucl. Med. 55, 1617–1622.
Dice, L.R., 1945. Measures of the amount of ecologic association between species. Ecology 26, 297–302.
Dong, C., Loy, C.C., He, K., Tang, X., 2014. Learning a deep convolutional network for image super-resolution. In: Computer Vision–ECCV 2014. Springer, pp. 184–199.
Fischl, B., 2012. FreeSurfer. NeuroImage 62, 774–781.
Gouttard, S., Styner, M., Joshi, S., Smith, R.G., Hazlett, H.C., Gerig, G., 2007. Subcortical structure segmentation using probabilistic atlas priors Medical Imaging. International Society for Optics and Photonics, 65122J-J-11.
Havaei, M., Davy, A., Warde-Farley, D., Biard, A., Courville, A., Bengio, Y., Pal, C., Jodoin, P.M., Larochelle, H., 2016. Brain tumor segmentation with deep neural networks. Med. Image Anal. 35, 18–31.
Iosifescu, D.V., Shenton, M.E., Warfield, S.K., Kikinis, R., Dengler, J., Jolesz, F.A., McCarley, R.W., 1997. An automated registration algorithm for measuring MRI subcortical brain structures. Neuroimage 6, 13–25.
Kamnitsas, K., Chen, L., Ledig, C., Rueckert, D., Glocker, B., 2015. Multi-Scale 3D Convolutional Neural Networks for Lesion Segmentation in Brain MRI. Ischemic Stroke Lesion Segmentation, 13.
Kleesiek, J., Urban, G., Hubert, A., Schwarz, D., Maier-Hein, K., Bendszus, M., Biller, A., 2016. Deep MRI brain extraction: a 3D convolutional neural network for skull stripping. Neuroimage 129, 460–469.
Klein, A., Tourville, J., 2012. 101 labeled brain images and a consistent human cortical labeling protocol. Front. Neurosci. 6, 171.
Krizhevsky, A., Sutskever, I., Hinton, GE., 2012. Imagenet classification with deep convolutional neural networks. Advances in neural information processing systems, 1097–1105.
LeCun, Y., Bengio, Y., Hinton, G., 2015. Deep learning. Nature 521, 436–444.
Long, J., Shelhamer, E., Darrell, T., 2015. Fully convolutional networks for semantic segmentation. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 3431–3440.
Marcus, D.S., Wang, T.H., Parker, J., Csernansky, J.G., Morris, J.C., Buckner, R.L., 2007. Open Access Series of Imaging Studies (OASIS): cross-sectional MRI data in young, middle aged, nondemented, and demented older adults. J. Cogn. Neurosci. 19, 1498–1507.
Moeskops, P., Viergever, M.A., Mendrik, A.M., de Vries, L.S., Benders, M.J., Isgum, I., 2016. Automatic segmentation of MR brain images with a convolutional neural network. Ieee Transactions on Medical Imaging.
Nobili, F., Naseri, M., De Carli, F., Asenbaum, S., Booij, J., Darcourt, J., Ell, P., Kapucu, O., Kemp, P., Svarer, C., Morbelli, S., Pagani, M., Sabri, O., Tatsch, K., Tossici-Bolt, L., Sera, T., Vander borght, T., van laere, K., varrone, A., 2013. Automatic semi-quantification of [123I]FP-CIT SPECT scans in healthy volunteers using BasGan version 2: results from the ENC-DAT database. Eur. J. Nucl. Med. Mol. Imaging 40, 565–573.
Patenaude, B., Smith, S.M., Kennedy, D.N., Jenkinson, M., 2011. A Bayesian model of shape and appearance for subcortical brain segmentation. Neuroimage 56, 907–922.
Prasoon, A., Petersen, K., Igel, C., Lauze, F., Dam, E., Nielsen, M., 2013. Deep feature learning for knee cartilage segmentation using a triplanar convolutional neural network Medical image computing and computer-assisted intervention: MICCAI. International Conference on Medical Image Computing and Computer-Assisted Intervention 16, 246–253.
Raghavan, V., Bollmann, P., Jung, G.S., 1989. A critical investigation of recall and precision as measures of retrieval system performance. ACM Trans. Inform. Syst. 7, 205–229.
Rosas, H.D., Goodman, J., Chen, Y.I., Jenkins, B.G., Kennedy, D.N., Makris, N., Patti, M., Seidman, L.J., Beal, M.F., Koroshetz, W.J., 2001. Striatal volume loss in HD as measured by MRI and the influence of CAG repeat. Neurology 57, 1025–1028.
Ross, C.A., Tabrizi, S.J., 2011. Huntington's disease: from molecular pathogenesis to clinical treatment. Lancet Neurol. 10, 83–98.
Tinaz, S., Courtney, M.G., Stern, C.E., 2011. Focal cortical and subcortical atrophy in early Parkinson's disease. Mov. Disord. 26, 436–441.
Van Leemput, K., 2009. Encoding probabilistic brain atlases using Bayesian inference. IEEE Trans. Med. Imaging 28, 822–837.
Vedaldi, A., Lenc, K., 2015. MatConvNet: convolutional neural networks for matlab. In: Proceedings of the 23rd Annual ACM Conference on Multimedia Conference, ACM, pp. 689–692.
Videbech, P., Ravnkilde, B., 2004. Hippocampal volume and depression: a meta-analysis of MRI studies. Am. J. psychiatry 161, 1957–1966.
Visser, E., Keuken, M.C., Douaud, G., Gaura, V., Bachoud-Levi, A.C., Remy, P., Forstmann, B.U., Jenkinson, M., 2016. Automatic segmentation of the striatum and globus pallidus using MIST: Multimodal Image Segmentation Tool. Neuroimage 125, 479–497.
Ystad, M., Hodneland, E., Adolfsdottir, S., Haasz, J., Lundervold, A.J., Eichele, T., Lundervold, A., 2011. Cortico-striatal connectivity and cognition in normal aging: a combined DTI and resting state fMRI study. Neuroimage 55, 24–31.
Zhang, W., Li, R., Deng, H., Wang, L., Lin, W., Ji, S., Shen, D., 2015. Deep convolutional neural networks for multi-modality isointense infant brain image segmentation. Neuroimage 108, 214–224.