Combination of convolutional networks with a learning based ensemble strategy for hippocampus subfield segmentation

Guanghua He¹, Tianzhe Lu² and Hancan Zhu²*

¹College of International Finance and Trade, Zhejiang Yuexiu University of Foreign Languages, Shaoxing, 312000, China;
²School of Mathematics Physics and Information, Shaoxing University, Shaoxing, 312000, China

*Corresponding author:
Hancan Zhu, PhD. 900 ChengNan Road, Shaoxing University, Shaoxing, Zhejiang, China 312000.
E-mail: hancanzhu@yeah.net

Abstract

Background
Accurate segmentation of hippocampal subfields from magnetic resonance (MR) brain images is an important step for studying brain disorders, including epilepsy, Alzheimer’s Disease (AD) and Parkinson’s disease. However, it is a difficult task because of the low signal contrast and small structural size.

Methods
Many advanced convolutional networks have been proposed and have achieved state-of-the-art performances in various applications. To take advantage of these advanced convolutional networks, in this paper, we propose a learning based ensemble strategy to integrate the results of different convolutional networks for hippocampus subfield segmentation. Our ensemble strategy is implemented by using a convolutional network. We have validated the proposed method based on a publicly available dataset.

Results
The experiment results have showed that the proposed ensemble strategy can significantly improve the performance of each single convolutional network, and outperform the state-of-the-art hippocampus subfield segmentation method.

Conclusion
The proposed ensemble strategy is effective for combining multiple different convolutional networks in hippocampus subfield segmentation.

Background
Hippocampus is a bilateral brain structure, involved in many brain disorders, such as epilepsy, Alzheimer’s disease (AD), and Parkinson’s disease. It consists of several histologically and functionally specialized subfields: the subiculum (SUB), the cornu ammonis sectors (CA) 1–3, and the dentate gyrus (DG). The studies have shown that different diseases affect different subfields, which suggest that hippocampal subfields may provide more precise information for earlier disease diagnosis than using the whole hippocampus.

To characterize subfields for studying brain disorders, it is critically important to accurately segment subfields from magnetic resonance (MR) images. Manual segmentation is often treated as a gold standard, but it is time consuming and tedious, along with inter- and intra-observer variability. Thus, automatic segmentation of hippocampal subfields is desirable, especially for large-scale disease studies. Among the automatic image segmentation methods, multi-atlas based image segmentation (MAIS) methods have been applied in both the hippocampal segmentation and hippocampal subfield segmentation. In the MAIS methods, all atlas images are first registered to a target image. Then, the
corresponding atlas label maps are warped into the target image space and further fused to obtain the final segmentation for the target image. However, these MAIS methods often need accurate nonlinear registration between each atlas image and the target image, which is very time-consuming.

In recent years, deep neural networks have been widely applied in medical image segmentation, without the need of image registration and designing features. For example, in U-net was proposed for biomedical image segmentation, which adopted skip connections to connect the contracting path and the symmetric expanding path. In a volumetric convolutional network with mixed residual connections was proposed for prostate segmentation. In V-net was proposed for volumetric medical image segmentation, in which the residual connections was also used. In pooling-free fully convolutional networks with dense skip connections were proposed for semantic segmentation, which were applied to brain tumor segmentation. In deep voxelwise residual networks were proposed for brain segmentation from 3D MR images. In a fully convolutional network was proposed for quick and accurate segmentation of neuroanatomy. In generative adversarial networks were used for hippocampal subfields segmentation in brain MR images. In a dilated dense U-net for was proposed for infant hippocampus subfield segmentation. However, it is often difficult to select a suitable network for a specific task, since there are so many candidate segmentation networks.

To alleviate the difficulty of network selection and also take advantage of multiple advanced convolutional networks, in this paper, we propose a learning-based ensemble strategy to combine the results of multiple convolutional networks, for obtaining more accurate segmentation than the case of using just a single convolutional network. Specifically, we adopt three segmentation networks in our task, i.e., 1) Unet-like structure (Unet-like) 2) Unet-like structure with residual connections (ResUnet) 3) fully convolutional network with residual connections (ResFCN). We validate our proposed method for hippocampal subfield segmentation based on a publicly available dataset. The experimental results demonstrate that our proposed learning based ensemble strategy can significantly improve the performance of each single convolutional network, and also outperforms the state-of-the-art hippocampal subfield segmentation method.

**Experiments and Results**

**Dataset and Pre-processing**

The proposed method was validated on a publicly available dataset (https://www.nitrc.org/projects/mni-hisub25). The dataset contains 25 subjects with manual labels for hippocampal subfields (CA1-3, CA4/DG and SUB). Fig. 1 shows the manual labels of a selected subject. Each subject consists of an isotropic 3D-MPRAGE T1-weighted image (TR=3000 ms; TE=4.32 ms; TI=1500 ms; flip angle=7°; matrix size=336×384; FOV =201×229 mm²; 240 axial slices with 0.6 mm slice thickness resulting in 0.6×0.6×0.6 mm³ voxels; acquisition time=16.48 min) and an anisotropic 2D T2-weighted TSE image (TR =10810 ms; TE=81 ms; flip angle=119°; matrix size=512×512; FOV=203×203 mm²; 60 coronal slices angled perpendicular to the hippocampal long axis, slice thickness of 2 mm, resulting in 0.4×0.4×2.0 mm³ voxels; acquisition time=5.47min). All images underwent automated correction for intensity non-uniformity, intensity standardization, and linear registration to the MNI152 space, and were resampled to a resolution of 0.4×0.4×0.4 mm³.

To facilitate the processing, we identified a bounding box to remove non-hippocampus region, with the box being large enough to cover the whole hippocampus. Then, we applied a histogram matching to
the cropped images such that brain tissues have similar intensity levels across all the subjects. To leverage the limited data, we left-right flipped each training image to double the number of training subjects.

![Cropped Images](image)

**Fig. 1.** Manual labels of a selected subject for hippocampal subfields (CA1-3, CA4/DG and SUB). The red indicates CA1-3; the blue indicates CA4/DG; and the green indicates SUB.

**Experimental Details**

Five-fold cross validation was used in the experiment. In each fold, we selected 20 subjects for training, in which 15 subjects were used for training Unet-like, ResUnet and ResFCN, and 5 subjects were used for training FuseNet. The rest of 5 subjects were used for testing. As the number of training subjects is limited, we randomly extracted patches from each training subject, instead of using the whole images as the input for each network. The patch size was set to $32 \times 32 \times 32$. Since both T1w and T2w images are available, we concatenated the extracted T1w and T2w image patches as input to Unet-like, ResUnet and ResFCN. The cross-entropy loss was used in each network.

The networks were trained by Adam method with a mini-batch size of 15, and implemented with caffe. For Unet-like, ResUnet and ResFCN, the learning rate was initially set to 0.0001, and was decreased by a factor of $\gamma = 0.1$ every 10000 iterations. For FuseNet, the learning rate was initially set to 0.0001, and was decreased by a factor of $\gamma = 0.25$ every 4000 iterations. We used a weight decay of 0.0005 and a momentum of 0.9 for all networks.

**Results**

Table 1 reports the Dice coefficients (mean ± std) of the segmentation results obtained by different networks including Unet-like, ResUnet, ResFCN and FuseNet. For each subfield, the best performance is indicated in bold. It can be observed that ResUnet achieves better segmentation results than Unet-like, suggesting that the residual connection is helpful when used in the Unet-like. Also, it can be found that the long-skip connection is still useful in residual networks by comparing the results of ResUnet and ResFCN. By integrating the results of all these three networks by FuseNet, the most accurate segmentation results are obtained, which are better than the best results obtained by a single network, ResUnet. Fig. 2 shows the hippocampal subfield segmentation for a randomly selected subject obtained by manual segmentation and different networks, respectively. From Fig. 2, we can find that local segmentation errors (indicated by black circles) introduced by each single network can be effectively corrected by ensemble learning.
We also compare our proposed method with a state-of-the-art hippocampal subfield segmentation method. The results are listed in Table 2. Note that, for fair comparison, we used the published results of HIPS as reported in. It shows that our proposed method FuseNet outperforms HIPS method, especially for segmenting the CA4/DG subfield which is the most difficult task.

**Table 1.** Comparison between different networks in term of Dice coefficients (mean±std) for each structure segmentation. Best results are shown in bold. (* indicates that FuseNet achieves significant improvement over the corresponding method in the Wilcoxon signed rank tests with p-value<0.001).

| Structure   | Unet-like | ResFCN  | ResUnet | FuseNet  |
|-------------|-----------|---------|---------|----------|
| CA1-3       | 0.916±0.011* | 0.912±0.011* | 0.919±0.010  | **0.920±0.011** |
| CA4/DG      | 0.871±0.021* | 0.869±0.021* | 0.876±0.017  | **0.878±0.018** |
| SUB         | 0.885±0.016* | 0.884±0.017* | 0.887±0.017  | **0.890±0.017** |
| Hippocampus | 0.960±0.005* | 0.955±0.005* | 0.960±0.005  | **0.961±0.005** |

**Table 2.** Comparison with a state-of-the-art method in term of Dice coefficients (mean±std) for each structure segmentation. Better results are shown in bold.

| Structure   | CA1-3 | CA4/DG | SUB     | Hippocampus |
|-------------|-------|--------|---------|-------------|
| HIPS        | 0.916±0.015 | 0.862±0.034 | 0.886±0.021 | 0.960±0.006 |
| FuseNet     | **0.920±0.011** | **0.878±0.018** | **0.890±0.017** | **0.961±0.005** |

**Discussion**

The idea of the proposed learning based combination strategy was inspired by the multi-atlas label fusion, which was used to fuse the warped label maps from multi-atlas, while the proposed method was used for fusing the results of multiple 3D deep convolutional networks. Compared to multi-atlas segmentation methods, the proposed method do not need the nonlinear registration, and will be much faster in the testing stage. A recent work has proposed three Ensemble-Nets for hippocampal segmentation. The purpose of their ensemble learning is to combine the label decisions from 2D multi-views, such that the complementary information residing in the multiple label probabilities
produced by their 2D networks can be captured. By contrast, our ensemble strategy is to tackle the difficulty of network model selection by fusing the results of different 3D networks with additional convolutional network.

According to the results, we also find that ResUnet outperforms Unet-like and ResFCN, which demonstrates both the long-skip connection and the short-skip connection (residual connection) are useful. With the long-skip connection, high-resolution features from the contracting path are combined with upsampled features, which can supplement the information that is lost during pooling. The residual connection can promote information propagation, accelerate the convergence, and improve performance.

Conclusion

In this paper, we have proposed a learning based ensemble strategy to combine the results of three different 3D deep convolutional networks, namely, 1) Unet-like, 2) ResUnet and 3) ResFCN for hippocampal subfield segmentation. The experimental results show that the learning based ensemble strategy can significantly improve the performance of each single network, and outperforms the state-of-the-art hippocampal subfield segmentation method.

Methods

In this section, we first introduce three different convolutional networks, namely Unet-like, ResUnet and ResFCN. Then, we present a learning based ensemble strategy to combine the results of these three networks.

Unet-like Structure

The Unet-like structure is shown in Fig. 3. It consists of a contracting path and an expansive path. The contracting path repeats the application of two $3 \times 3 \times 3$ convolutions and one $2 \times 2 \times 2$ max pooling operation with stride 2. Correspondingly, the expansive path repeats the application of one $4 \times 4 \times 4$ deconvolution with stride 2, and two $3 \times 3 \times 3$ convolutions. Each convolution is followed by a batch normalization and a rectified linear unit (ReLU). Different from the original Unet, padded convolution layers are used to maintain the spatial dimension, and also feature maps in the contracting path are added to the corresponding feature maps with the same resolution in the expansive path with long-skip connections. This element-wise summation can reduce the dimension of features compared to conventional concatenation.

Residual Connection

Residual connection is introduced in the deep residual network. It has been demonstrated that the residual connection can promote information propagation and accelerate the convergence. To exploit the advantage of the residual connection, we group every two convolutional layers with a residual connection in the Unet-like structure, and obtain the residual Unet-like structure (ResUnet). To investigate the influence of long-skip connections in residual networks, we also present a fully convolutional structure with residual connections (ResFCN), by removing the long-skip connections from ResUnet.
Learning based Ensemble Strategy

To fully take advantage of these advanced nets, we propose a learning based ensemble strategy to combine the predictions $p_{i,k}$ from each network with trainable weighting coefficients $\alpha_{i,k}, i \in \{1, \ldots, m\}, k \in \{1, \ldots, K\}$,

$$p_k = \sum_{i=1}^{m} \alpha_{i,k} p_{i,k}$$  \hspace{1cm} (1)

where $p_{i,k}$ is the probability map for the $k^{th}$ hippocampal subfield estimated by the $i^{th}$ network, $p_k$ is the fused probability map for the $k^{th}$ hippocampal subfield. The weighting coefficients $\alpha_{i,k}$ are determined with the training subjects.

Suppose multiple networks have been trained. By applying these trained networks to the training subjects, we can obtain the probability maps of hippocampal subfields for each training subject $p_{i,k}^{(j)}$, $j = 1, \ldots, N$, where $p_{i,k}^{(j)}$ is the probability map of the $k^{th}$ hippocampal subfield obtained by the $i^{th}$
network for the \(j^{th}\) training subject. Thus, we can construct a training probability set 
\[
\{(p_{1,j}^{(j)}, \ldots, p_{1,k}^{(j)}, \ldots, p_{m,j}^{(j)}, \ldots, p_{m,k}^{(j)}, y^{(j)}_j)\}, j = 1, \ldots, N,
\]
where \(y^{(j)}_j\) is the target hippocampal subfield label among \(K\) classes. Based on this training probability set, we can obtain the values of weighting coefficients \(\alpha_{i,k}\) by solving the following optimization problem:
\[
\text{argmin}_{\alpha} \text{Loss} = -\frac{1}{N} \sum_{j=1}^{N} \sum_{k=1}^{K} 1\{y^{(j)} = k\} \log \left( \sum_{i=1}^{m} \alpha_{i,k} p_{i,k}^{(j)} \right).
\]  \hspace{1cm} (2)

The optimization problem (2) is considered as a neural network with \(1 \times 1 \times 1\) convolutions, which is named as FuseNet. The architecture of FuseNet is shown in Fig. 4. In the FuseNet, we concatenate probability maps for each subfield that are obtained from different networks, and then fuse them with a \(1 \times 1 \times 1\) convolution.

Once the weighting coefficients \(\alpha_{i,k}, i \in \{1, \ldots, m\}, k \in \{1, \ldots, K\}\) are obtained, we can get the fused probability map for each hippocampal subfield with Eq. (1). Then, the final hippocampal subfield label of voxel \(x\) can be obtained by
\[
s(x) = \text{argmax}_k p_k(x), \ k \in \{1, \ldots, K\}.
\]  \hspace{1cm} (3)

**Evaluation**

We evaluated the image segmentation results based on Dice coefficient (Dice), which is used to measure the relative volumetric overlap between the automated segmentation and the manual segmentation \(28\). By denoting \(A\) as the manual segmentation, \(B\) as the automated segmentation, and \(V(X)\) as the volume of segmentation \(X\), Dice is defined as:
\[
\text{Dice} = 2 \frac{V(A \cap B)}{V(A) + V(B)}
\]

**Availability of data and materials**

The data used during the study are publicly available (https://www.nitrc.org/projects/mni-hisub25) \(22\).

**Abbreviations**

MR: Magnetic Resonance  
AD: Alzheimer’s Disease  
SUB: The Subiculum  
CA: The Cornu Ammonis  
MAIS: Multi-Atlas based Image Segmentation  
Unet-like: Unet-like Structure  
ResUnet: Unet-like Structure with Residual Connections  
ResFCN: Fully Convolutional Network with Residual Connections  
ReLU: Rectified Linear Unit  
Dice: Dice coefficient

**References**

1. Akhondi-Asl, A., Jafari-Khouzani, K., Elisevich, K. & Soltanian-Zadeh, H. Hippocampal volumetry for lateralization of temporal lobe epilepsy: automated versus manual methods. *NeuroImage* **54**, S218-S226 (2011).
2 Duvernoy, H. M. *The human hippocampus: functional anatomy, vascularization and serial sections with MRI*. (Springer Science & Business Media, 2005).
3 Mueller, S. G. *et al.* Systematic comparison of different techniques to measure hippocampal subfield volumes in ADNI2. *NeuroImage: Clinical* **17**, 1006-1018 (2018).
4 Romero, J. E., Coupé, P. & Manjón, J. V. HIPS: A new hippocampus subfield segmentation method. *NeuroImage* **163**, 286-295 (2017).
5 Wang, H. *et al.* Multi-atlas segmentation with joint label fusion. *Pattern Analysis and Machine Intelligence, IEEE Transactions on* **35**, 611-623 (2013).
6 Pipitone, J. *et al.* Multi-atlas segmentation of the whole hippocampus and subfields using multiple automatically generated templates. *NeuroImage* **101**, 494-512 (2014).
7 Zhu, H., Tang, Z., Cheng, H., Wu, Y. & Fan, Y. Multi-atlas label fusion with random local binary pattern features: Application to hippocampus segmentation. *Scientific reports* **9**, 1-14 (2019).
8 Heckemann, R. A., Hajnal, J. V., Aljabar, P., Rueckert, D. & Hammers, A. Automatic anatomical brain MRI segmentation combining label propagation and decision fusion. *NeuroImage* **33**, 115-126 (2006).
9 Iglesias, J. E. & Sabuncu, M. R. Multi-atlas segmentation of biomedical images: A survey. *Medical image analysis* **24**, 205-219 (2015).
10 Shen, D., Wu, G. & Suk, H.-I. Deep learning in medical image analysis. *Annual review of biomedical engineering* **19**, 221-248 (2017).
11 Litjens, G. *et al.* A survey on deep learning in medical image analysis. *Medical image analysis* **42**, 60-88 (2017).
12 Ronneberger, O., Fischer, P. & Brox, T. U-net: Convolutional networks for biomedical image segmentation. *International Conference on Medical Image Computing and Computer-Assisted Intervention*, 234-241 (2015).
13 Yu, L., Yang, X., Chen, H., Qin, J. & Heng, P.-A. Volumetric ConvNets with Mixed Residual Connections for Automated Prostate Segmentation from 3D MR Images. *AAAI*, 66-72 (2017).
14 Milletari, F., Navab, N. & Ahmadi, S.-A. V-net: Fully convolutional neural networks for volumetric medical image segmentation. *3D Vision (3DV), 2016 Fourth International Conference on*, 565-571 (2016).
15 McKinley, R., Jungo, A., Wiest, R. & Reyes, M. Pooling-Free Fully Convolutional Networks with Dense Skip Connections for Semantic Segmentation, with Application to Brain Tumor Segmentation. *International MICCAI Brainlesion Workshop*, 169-177 (2017).
16 Chen, H., Dou, Q., Yu, L., Qin, J. & Heng, P.-A. VoxResNet: Deep voxelwise residual networks for brain segmentation from 3D MR images. *NeuroImage* **170**, 446-455 (2018).
17 Roy, A. G., Conjeti, S., Navab, N., Wachinger, C. & Initiative, A. s. D. N. QuickNAT: A fully convolutional network for quick and accurate segmentation of neuroanatomy. *NeuroImage* **186**, 713-727 (2019).
18 Shi, Y., Cheng, K. & Liu, Z. Hippocampal subfields segmentation in brain MR images using generative adversarial networks. *Biomedical engineering online* **18**, 5 (2019).
19 ZHU, H. *et al.* Dilated Dense U-net for Infant Hippocampus Subfield Segmentation. *Frontiers in Neuroinformatics* **13**, 30 (2019).
20 He, K., Zhang, X., Ren, S. & Sun, J. in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* **770-778** (2016).
Acknowledgments

Not applicable

Funding

This work was supported in part by National Natural Science Foundation of China (Nos. 61602307, 61877039), Natural Science Foundation of Zhejiang Province (No. LY19F020013), and Research Project of the Department of Education of Zhejiang Province (No. Y201839295).

Authors’ Contributions

Guanghua He developed the method and wrote the paper; Tianzhe Lu analyzed the data; Hancan Zhu supervised the data analysis and revised the manuscript. All authors read and approved the final manuscript.

Corresponding author

Correspondence to Hancan Zhu.

Ethics approval and consent to participate

Not applicable

Consent for publication

Authorization has been granted to publish the results.

Competing interests

The authors declare that they have no conflict of interest.