Cognitive Computation (2019) 11:630–643
https://doi.org/10.1007/s12559-019-09645-z

Multi-target Interactive Neural Network for Automated Segmentation of the Hippocampus in Magnetic Resonance Imaging

Beibei Hou [1,2] · Guixia Kang [1,2] · Ningbo Zhang [1] · Kui Liu [1,2]

Received: 14 February 2018 / Accepted: 30 April 2019 / Published online: 23 July 2019

Abstract

The hippocampus has been recognized as an important biomarker for the diagnosis and assessment of neurological diseases. Convenient and accurate automated segmentation of the hippocampus facilitates the analysis of large-scale neuroimaging studies. This work describes a novel technique for hippocampus segmentation in magnetic resonance images, in which interactive neural network (Inter-Net) is based on 3D convolutional operations. Inter-Net achieves the interaction through two aspects: one is the compartments, which builds an exponential ensemble network that integrates numerous short networks together when forward propagation. The other is the pathways, which realizes inter-connection between feature extraction and restoration. In addition, a multi-target architecture is proposed by designing multiple objective functions in terms of evaluation index, information theory, and data distribution. The proposed architecture is validated in fivefold cross-validation on the Alzheimer's Disease Neuroimaging Initiative (ADNI) dataset, where the mean Dice similarity indices of 0.919 (± 0.023) and precision of 0.926 (± 0.032) for the hippocampus segmentation. The running time is approximately 42.1 s from reading the image to outputting the segmentation result in our computer configuration. We compare the experimental results of a variety of methods to prove the effectiveness of the Inter-Net and contrast integrated architectures with different objective functions to illustrate the robustness of the fusion. The proposed framework is general and can be easily extended to numerous tissue segmentation tasks while it is tailored for the hippocampus.

Keywords Hippocampus · Interactive neural network · Magnetic resonance images · Multi-target · Objective function

Introduction

Cognitive dysfunction usually brings about the nervous system disease which is caused by structural changes in the brain [1, 2]. Especially, the hippocampus plays a vital role in human memory and learning [3–5]. Atrophy and volume reduction of the hippocampus have been shown to be observable characteristics for the detection of mild cognitive impairment (MCI) and Alzheimer's disease (AD) [6]. In neuroimaging studies, magnetic resonance imaging (MRI) is often used for the volumetric assessment of the hippocampus [7], which has become an indispensable tool for disease monitoring [8], diagnosis [9], treatment [10], and prognosis [11]. As such, the quantitative analysis of the hippocampus in MRI is of great significance in clinical to better understand the inter-individual variability of subject neuroanatomy.

So far, manual segmentation of brain tissues is still regarded as the optimal standard [12], in spite of the fact that it is an extremely tedious and time-consuming operation [13]. Intra- and inter-rater volume variability appear upon the wider variability which is inherent to manual hippocampal segmentation [14]. It is noticeable that the gray level of the hippocampus in MRI verges on the adjacent structures, such as the amygdala, thalamus, and caudate nucleus [15]. There is no conspicuous boundary between the hippocampus and neighboring regions (Fig. 1), which increases the difficulty of the hippocampus segmentation. Consequently, a consistent and faithful automatic segmentation methodology for hippocampus is essential for improving the reliability of tissue segmentation as well as relieving the workload of radiologists.

Over the past decades, automatic segmentation of hippocampus has received widespread attention due to its importance in clinical applications. There are also various

✉ Guixia Kang
gxkang@bupt.edu.cn

[1] Beijing University of Posts and Telecommunications, No.10 Xitucheng Road, Haidian District, Beijing, China

[2] Wuxi BUPT Sensory Technology and Industry Institute CO.LTD, No.97 Linghu Road, Newly District, Wuxi, China

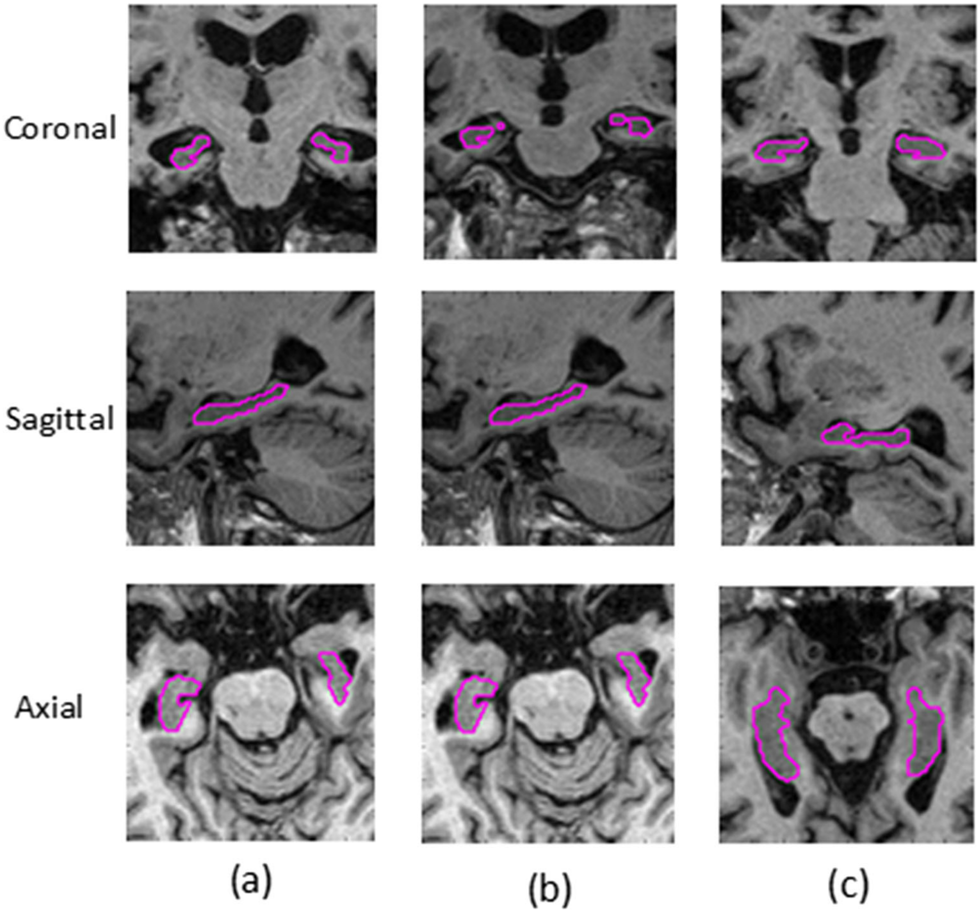

Fig. 1 Patients with different degrees of illness is selected to display the contrast in the manually segmented hippocampus (purple circle). Left to right are normal control (NC) (a), MCI (b), and AD (c)

retrospective studies which compared the strengths and weaknesses of various techniques of automatic hippocampal segmentation [16–18]. In general, these methodologies can be divided into three categories.

Firstly, atlas-based methods were utilized as non-parametric regression models. They map the coordinates of tissues from the atlas image to the target image through registration [19, 20]. Subsequently, multi-atlas [21, 22] approaches were introduced to improve the accuracy of segmentation, and probabilistic atlas [23, 24] methods were suggested because of high computational cost. Secondly, traditional machine learning methods assigned labels to image through handcraft characteristics. Yongfu Hao et al. [25] proposed local label learning strategies to estimate the segmentation of subject images using k-nearest neighbor (k-NN) and support vector machine (SVM) methods based on image intensity and texture features. Additionally, the automatic brain structure segmentation (ABSS) [26] which was a two-stage architecture performed well in the hippocampus segmentation task [17]. Thirdly, data-driven approaches extracted high-level characteristics automatically and realized tasks in an end-to-end manner. Convolutional neural network (CNN), fully convolutional network (FCN) [27], and U-net [28] models were proposed in succession for image segmentation. It is worth mentioning that the FCN label the input image densely with the fully convolutional layer. A case in point is Kamnitsas et al. [29], as an integrated technique between FCNs and conditional random fields (CRFs) was proposed to further improve the segmentation accuracy. Liu X et al. [30] designed an RPP model through the incorporation of residual connections and pyramid pooling into the FCN framework. Although RPP has a great effect on road detection, it was not applicable in medical image segmentation because of the computer memory limitation. In view of segmentation accuracy and operation complexity, we consider the data-driven approach represented by CNNs to achieve automatic segmentation of hippocampus owing to its flexibility.

Additionally, ensemble learning achieves more accurate, stable, and robust results by combining multiple single models. Especially, multi-view learning makes use of the difficulty of learning data in different views and exerts interactions between views, complementary advantages, and collaborative learning. Hessian multiset canonical correlations (HesMCC) [31, 32] for multi-view feature analysis is presented, taking the advantage of Hessian and provides superior extrapolating capability and finally leverage the classification performance. A novel multi-view attention network (MuVAN) [33] is proposed to learn finegrained attentional representations by constructing a hybrid focus procedure. Recently, multi-view convolutional networks [34, 35] are proposed, for which the features are combined using a dedicated fusion method for false positive reduction. Based on the previous parallel network, this paper designs a multi-target integrated network to characterize the consistency of predicted output and ground truth from multi-view.

In this paper, the hippocampus is considered as a whole, without meticulous delimitations of the left or right hippocampal areas [24, 25, 36, 37]. We filter dataset with hippocampus annotated expertly from the Alzheimer's Disease Neuroimaging Initiative (ADNI) dataset [38] and illustrate the relevance distribution between the extent AD and hippocampal volume regarding the left and right volume. As shown in Fig. 2, whether it is the left, the right, or the total hippocampal volume, the density distribution curve regarding the degree of disease is similar. Therefore, the overall hippocampus is used as a research goal to achieve automatic segmentation.

To segment the hippocampus accurately and efficiently, a multi-target interactive neural network is proposed for automatic hippocampus segmentation in MRI. We compare the proposed various objective functions with the blended architecture to validate the effectiveness of multi-target fusion. Simulation results indicate that the performance of the volume correlation between the automatic output and manual segmentation improved greatly although the improvement in the accuracy is not obvious. Main contributions of this work are summarized as follows:

(1) 3D interactive network (Inter-Net) is designed by drawing on the notion of the U-net and residual network. The Inter-Net is capable of directly mapping volumetric data into a corresponding score volume within a single forward propagation. Besides, the segmentation accuracy increases to some extent through the interaction of compartments and pathways.

(2) Multi-target integrated architecture is proposed to perform robust hippocampus segmentation. To the best of our knowledge, we are one of the founding pioneers who combine the advantages of multiple objective functions.

(3) To ensure the validity and authority, the segmentation performance of our proposed method is measured in voxel-, volume-, and distance-based metric.

The remainder of this paper is organized as follows. We detail the proposed method in the "Method" section and report the experimental results in the "Experiments and Results" section. The "Discussion" section further analyzes the key issues of the proposed method and discusses future directions. The conclusions are drawn in the "Conclusion" section.

Method

In this section, the hippocampal segmentation is presented as an optimization problem. Simply put, we find a suitable function that maps the input image I to the corresponding binary tissue mask M. Main contributions of our work are the network architecture and the multi-target mind, which are outlined in the "Model Architecture" and "Objective Function" sections, respectively.

Model Architecture

Borrowing from the innovation of self-encoder [39, 40] and residual network [41, 42], the interactive neural network (Inter-Net) is designed. It achieves the interaction through two aspects: one is through the compartments, which builds an exponential ensemble network that integrates numerous short networks together when propagation is forwarded. Second is through the pathways, which realizes interconnected pathways between feature extraction and restoration. Inter-Net encourages feature reuse throughout the network, which makes the transfer of features and gradients more

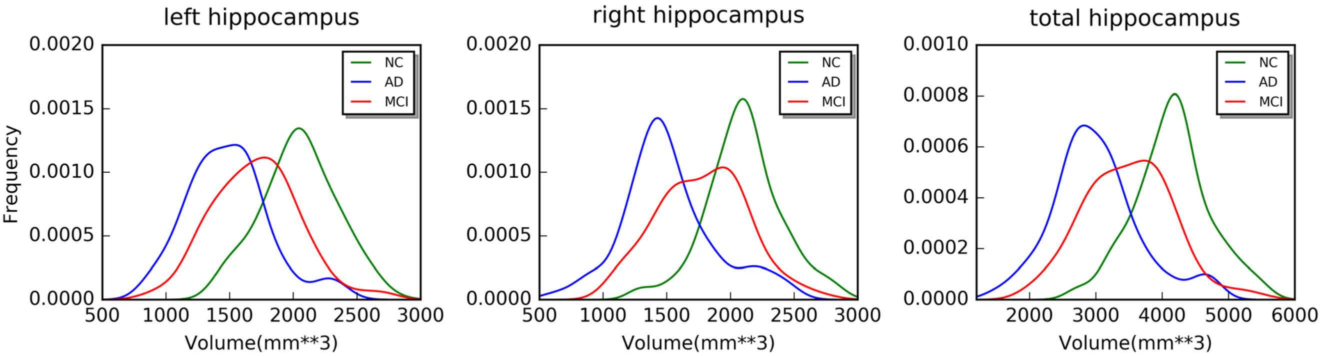

Fig. 2 Density map on hippocampal volume. From left to right, the pictures depict the relevant distribution between the hippocampal volume and the extent of the disease in left, right, and total hippocampus level

efficient and the network is easier to train. Inter-Net is constructed with the interactive block and the sampling block, where the specific routine is shown in Fig. 3.

Similarly, the Inter-Net, as designed in Fig. 3a, is constructed with the interactive block (Fig. 3b) and sampling block (Fig. 3c). Similarly, the interactive block recognizes the interactive delivery of extracted characteristics while the sampling block achieves the change in the numbers of channels and down-sampling operations.

Next, we describe the working principles of this architecture in details.

(a) Interactive Block

The biggest difference between the Inter-Net and the residual network [54] lies in the way of information interaction. As for the Inter-Net, interactive block which is shown in Fig. 3b cooperates with the characteristics of layers by $h(x^{(l)}) = x^{(l-2)}$, instead of $h(x^{(l)}) = x^{(l)}$ in the residual network, where $x^{(l-2)}$ is the input feature to the $(l-2)$th interactive unit. Its specific formula is as follows:

$$x_j^{(l)} = \max\left(0, w_{ij}^{(l)} * x_i^{(l-1)} + w_{ij}^{(l-2)} * x_i^{(l-3)} + b_j^{(l)}\right), \tag{1}$$

where $w_{ij}^{(l)}, l \in [1, L]$ denotes the trainable convolution filter that connects the feature maps of contiguous layers, L is the total number of layers in the network and l represents the index of a convolutional layer, $x_j^{(l)}, i \in [1, F]$ represents the channel j in layer l, F is the number of filters in current layer, $b_j^{(l)}$ represents the trainable bias terms, $*$ denotes the convolution operation in the same border mode, and max(0, ·) means that we need to adopt rectified linear units (ReLU) [43] as an activation function.

(b) Sampling Block

As the name suggests, sampling blocks, shown in Fig. 3c, characterize the number of channels with a concurrent fully convolutional layer. It is composed of two concurrent convolutional actions; one is used to extract features in a normal receptive field while another realizes fully connected to the previous layer.

$$y_j^{(l)} = \max\left(0, w_{ij}^{(l)} * x_i^{(l-1)} + \lambda^{(l)} \cdot x_i^{(l-1)} + c_j^{(l)}\right), \tag{2}$$

where $\lambda^{(l)}$ denotes the sharing weights between the $(l-1)$th and the lth layer in a fully connected manner.

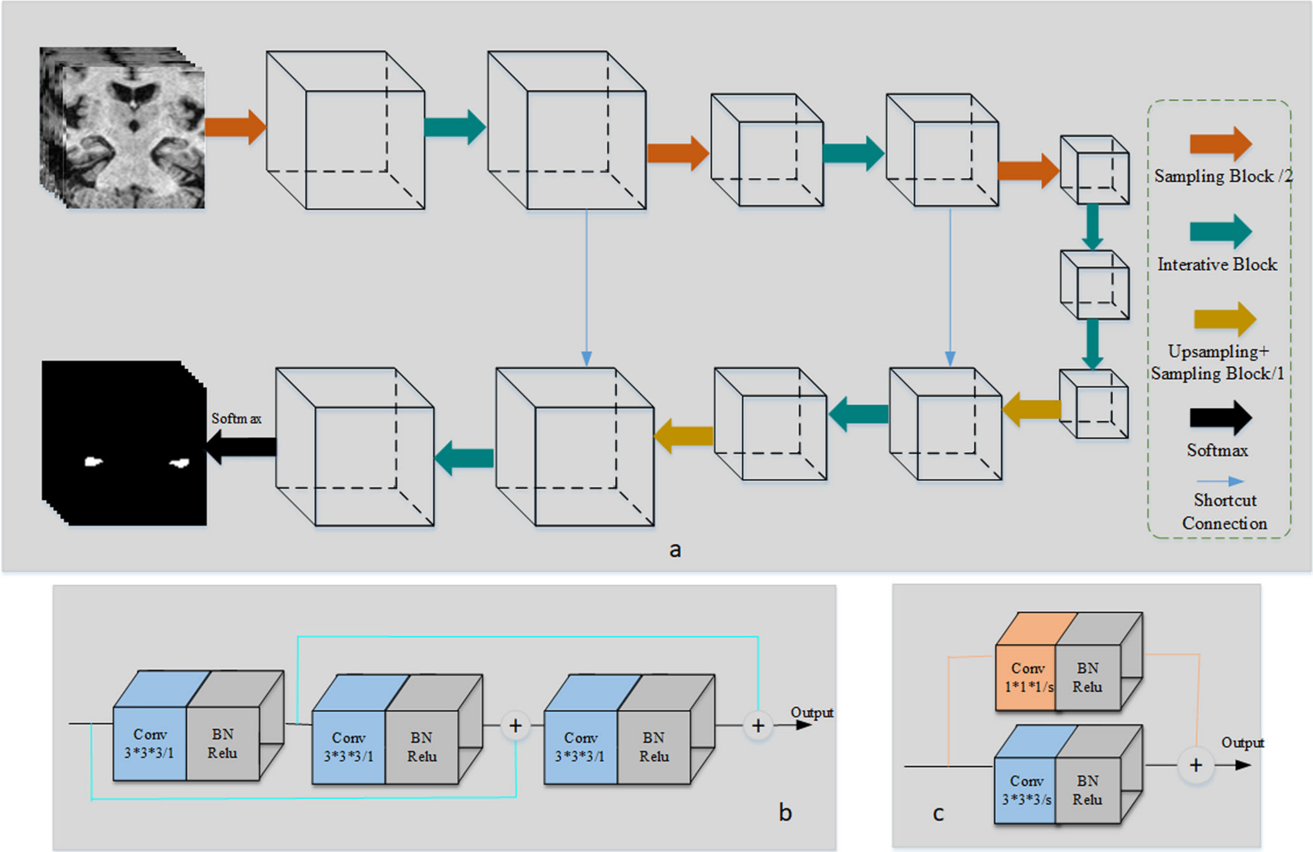

Fig. 3 An overview of the proposed interactive network (Inter-Net). **a** The structure of Inter-Net which consists of two interactive pathways between the upper and subcortical. **b**, **c** The Inter-Net component modules interactive block and sampling block, respectively

In regard to downsampling, we can set the strides (s) for both parallel pathways. In the subcortical pathway, the module is only responsible for changing the number of channels in its current condition due to the absence of downsampling.

(c) Discriminative Layer

In the last layer, the prediction of the voxel is achieved by fully connected convolutional layer with the activation function of Softmax. It outputs a matrix with values in the range between 0 and 1, which represents the probability that each pixel belongs to the tissue in turn. The computational process is as follows:

$$o_1 = \frac{e^{\theta_i^T x^{(L-1)}}}{\sum\limits_{j=1}^{2} e^{\theta_j^T x^{(L-1)}}}, \tag{3}$$

where θ_i, $i \in [1, k]$ denotes the adjustable parameters in a fully connected network, k indicates the total number of categories, and 2 is specific to the hippocampal segmentation issue.

In addition, the Inter-Net can be seen as a segmentation network composed of an upper pathway and a subcortical pathway. The increasingly abstracted image features can be learned in the upper pathway as well as the subcortical pathway, which calculates the probabilistic segmentation at the voxel level. Furthermore, contextual image information is crucial for maintaining the accuracy of image segmentation either it is in the upper or subcortical pathway. The higher-lower level features are concatenated in a shortcut connection manner, realizing the interconnections of the upper and subcortical pathways.

Objective Function

In this paper, we model hippocampal segmentation as an optimization issue, finding an appropriate function f which minimizes the loss function. It can be represented in the following formula:

$$\hat{f} = \arg\min_{f \in \mathcal{F}} \sum_{n} E(f_n, f(I)), \tag{4}$$

where $\mathcal{F}$ indicates the set of all possible functions for segmentation, E is an error measure, which is also known as an objective function, calculating the dissimilarity between manual segmentation and automated segmentation, and $\hat{f}$ is the most suitable function corresponding specific objective function E.

The objective function E is a destination that explores the relationship between input images I and the objective mask. From an engineering perspective, the objective is the criteria that evaluate the performance of the system. In a practical manner, a single evaluative criterion will reduce the robustness of the system, which is too monotonous to illustrate the effect of the model. Therefore, we attempt to design three objective functions from different considerations. The specific explanations are as follows:

(a) Probabilistic Similarity Objective Function (PSF)

Probabilistic similarity index (PSI) is evolved from the most common dice similarity coefficient, which represents a proximity measure between probability segmentation output and binary reference segmentation. On the account of satisfying the definition of minimizing the objective function, we construct PSF by subtracting PSI from 1 because PSI is greater than 0 and less than 1. Then, it can be described as follows:

$$E_{\mathrm{PSF}} = 1 - \frac{2 \times \sum\limits_{x} pg}{\sum\limits_{x} p^2 + \sum\limits_{x} g^2} = 1 - \mathrm{PSI}, \tag{5}$$

where p denotes the probabilistic matrix corresponding to the current output, and g indicates the ground truth for current samples. The value is between 0 and 1, which is an alternative map of dice similarity coefficient.

(b) Cross-Entropy Objective Function (CEF)

Cross-entropy, the most prevalent theory in various applications currently, is evolved from information theory. It is an effective tool for calculating linguistic disambiguation, which measures the similarity between ground truth g and predicted markers p. The formula is as follows:

$$E_{\mathrm{CEF}} = -\frac{1}{n} \sum_{x} [g \ln p + C \cdot (1-g) \ln(1-p)]. \tag{6}$$

Compared with a typical binary cross-entropy cost function, an extra parameter C, balance factor, is used to compensate for the imbalance in the class gap. In the hippocampal segmentation task, $C = 5$ is set to achieve a more accurate segmentation.

(c) Poisson Distribution Objective Function (PDF)

The majority of things in real life comply with the Poisson distribution that is the limiting distribution for a normal approximation to a binomial where the probability goes to zero and the number of trials goes to infinity. E_{PDF} comes from Poisson distribution and measures the degree to which the predicted distribution deviates from the expected distribution. In this paper, it is computed by

$$E_{\mathrm{PDF}} = \frac{1}{n} \sum_{x} (p - g \log p). \tag{7}$$

From the description of formula, E_{PDF} focuses more on voxels marked as 1 in ground truth, which makes up the imbalance of categories precisely.

(d) Model Fusion in Multi-Target Function

A simple overlay of the above objective function will decrease the performance of the model conversely due to dimensional inconsistency. It has been greatly demonstrated that integration always improves the performance of models more or less, such as multi-view [35, 37]. Likewise, we construct the parallel multi-target architecture by integrating the above objective functions, presented in Fig. 4.

$$P(\hat{y}|I) = \sum_i \lambda_i p_i(\hat{y}|I; W_i), \tag{8}$$

where $P(\hat{y}|I)$ is the fused prediction probabilistic matrix corresponding to the input image I through the entirety of the network. λ_i represents the ratio in which architecture with the ith objective function accounts for the entire network, respectively. In this paper, we determine λ_i through a grid search strategy and discover $\lambda_1 = 0.5$, $\lambda_2 = 0.3$, $\lambda_3 = 0.2$, realizing the best combination among objective networks mentioned above.

Training

Throughout the recent years, a rich body of methods have been proposed for automatically choosing learning rates, such as adaptive learning rate (Adadelta) [44], adaptive moment estimation (Adam) [45], root mean square prop (RMSprop) [46], which sets an adaptive learning rate for parameters through gathering various statistical knowledge of the partial derivatives during the iteration. In our experiments, we compare a variety of adaptive optimization functions and adopt the most robust optimizer, Adam, with a learning rate of 0.001 and other parameters with the default.

Dropout [47] is set to 0.2 in the middle layer and the fully convolutional layer. Taking into account the impact of the number of parameters, the size of all convolution kernels in the Inter-Net is set to 3 × 3 × 3 with the exception of the fully convolutional operation. This thought is borrowed from the visual geometry group (VGG-Net) [48] and can decrease the trainable parameters without affecting the receptive field [49]. Additionally, early stopping with patience = 5, batch-normalization with epsilon = 1e-6, and momentum = 0.9 are utilized to avoid overfitting. We implement the proposed architecture on Python based on the deep learning library of Keras, utilizing TensorFlow backend and a GPU of NVIDIA TITAN X.

Experiments and Results

Dataset

ADNI is a global research effort of various coinvestigators from extensive private companies and academic institutions. It was launched as a public-private partnership in 2003, led by Principal Investigator Michael W. Weiner, MD. The initial purpose of ADNI has been to investigate and develop treatments that mitigate or halt the progression of AD through biological markers or clinical and neuropsychological assessment. These could be combined to measure the progression of MCI and early AD. While being established, subjects have been recruited from over 50 locations across Canada and the USA. In this article, all images utilized in the preparation were downloaded from the ADNI LONI Image Data Archive (adni.loni.usc.edu).

Specifically, T1-weighted 1.5T MR image is considered from the ADNI1 database. To ensure consistency in the segmentation of the hippocampus, scans with the hippocampal mask are selected with the voxel size 1.25 × 1.25 × 1.2 mm, repetition time (TR) = 2300 ms, and echo time (TE) = 3 ms. As a result, we screen 550 scans that were accompanied by manual hippocampal segmentation results. The demographic information includes sex, age, weight, and Mini-Mental State Examination (MMSE) listed in Table Table 1.

Considering the relatively fixed position and a very small volume of hippocampus in the entire brain, we crop the

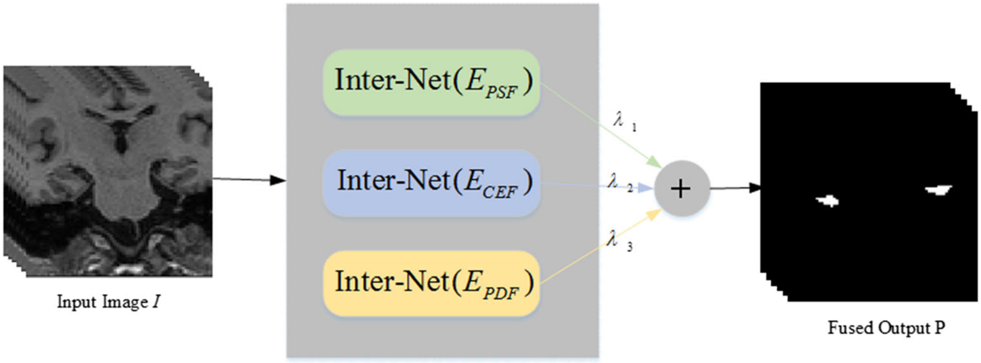

Fig. 4 Multi-target integrated architecture used in realizing high-precision segmentation

Table 1 Fundamental dataset information (mean ± std)

	M/F	Age	Weight	MMSE	Number
NC	76/86	76.15 ± 5.49	71.05 ± 14.91	29.13 ± 1.06	162
MCI	156/126	76.33 ± 6.86	74.47 ± 13.81	26.05 ± 2.82	282
AD	58/48	76.55 ± 5.89	70.45 ± 11.94	21.87 ± 3.59	106

volume of interest (VOI) of the hippocampus into each scan. In addition, computer memory limitation also makes it necessary to clip target images. After many attempts, the cropping box of 76^3 is defined.

Evaluation Metrics

A series of volumetric quality measures are adopted to assess the performance of the above model quantitatively. For binary classification problems, the confusion matrix is the most basic and commonplace method of evaluation. By definition, a confusion matrix is composed of true positive (TP), false positive (FP), false negative (FN), and true negative (TN). Several combinations of the above indices are widely used to measure the overlap and correlation between automatic A and manual M segmentation results. To thoroughly evaluate the effectiveness of the proposed method, four evaluation indicators are defined from the aspects of voxel, volume, and distance.

Firstly, dice similarity coefficient (DSC) is the most popular factor to quantify the segmentation accuracy [50]. It is a voxel-based level and describes a normalized overlap coefficient between the gold standard and automated segmentation, which is defined as follows:

$$\mathrm{DSC} = \frac{2\mathrm{Vol}(MA)}{\mathrm{Vol}(M) + \mathrm{Vol}(A)} = \frac{2\mathrm{TP}}{2\mathrm{TP} + \mathrm{FP} + \mathrm{FN}}, \tag{9}$$

where MA denotes the intersection of manual M and automated A segmentation, that is, the overlapping voxels between the gold standard and the automated segmentation, and Vol(·) represents a volume formula, that is, how many voxels belong to the hippocampal structures.

Secondly, precision (PRE) is a description of the random error and is a measure of statistical variability. It is the repeatability of successive measurements under the same conditions, characterizing the size of the random error during the measurement process.

$$\mathrm{PRE} = \frac{\mathrm{Vol}(MA)}{\mathrm{Vol}(M)} = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FP}}. \tag{10}$$

PRE is defined as the ratio of the number of the true positive voxel to the sum of both true positive and false positive size. The larger the value is, the higher the accuracy of the model is, and the better it performs.

Thirdly, relative volume difference (RVD) is defined to measure the reliability of the automated hippocampal segmentation through evaluating volumetric, which is defined as follows:

$$\mathrm{RVD} = \frac{|\mathrm{Vol}(M) - \mathrm{Vol}(A)|}{\mathrm{Vol}(M)} = \frac{|\mathrm{TN} - \mathrm{FP}|}{\mathrm{TP} + \mathrm{TN}}. \tag{11}$$

RVD measures the volumetric difference between two images that are superimposed on each other, so that the lower its grade, the more reliable it is. However, there is no information concerning the overlap of the segmentation. In the most extreme cases, the automatic segmentation may achieve the same total tissue size without any voxels in overlapping.

Lastly, the root mean square (RMS) depicts the degree of resemblances of resulting hippocampal segmentation that is masked in a statistical method. It is a distance-based metric by using surface-to-surface geometrics which reflect the magnitude of a varying quantity.

$$\mathrm{RMS} = \sqrt{\frac{1}{|S_A| + |S_M|} \times \left(\sum_{a \in S(A)} d^2(a, S_M) + \sum_{m \in S(M)} d^2(m, S_A) \right.}, \tag{12}$$

where S_A and S_M are indicated as the set of surface voxels of automated segmentation A and mask M, respectively, $d(a,S_M)$ denotes the nearest distance from point "a" to the surface S_A. RMS assesses the imperfection of the automated to the manual segmentation based on surface distance.

The Effectiveness of Network Architecture

In this paper, we perform extensive ablation studies on the selected dataset (fivefold cross validation) to investigate the efficacy of proposed architecture on hippocampal segmentation.

Our ultimate goal is to generate a binary matrix which is rooted in the probabilistic output; 1 denotes a tissue voxel and 0 to the opposite. Therefore, we choose the mean DSC [51] as the fixed criteria, which should be maximized throughout the training process, and the same criteria should be utilized for the validation set. In this experiment, we discover that $t_{\mathrm{PSF}} = 0.5$, $t_{\mathrm{CEF}} = 0.67$, $t_{\mathrm{PDF}} = 0.45$, and $t_I = 0.45$ were the thresholds of each network. The results of the evaluation of each network are indicated in the experimental part of Table 2.

To ascertain whether the interactive block is available for tissue segmentation, we compare the segmentation results of different neural network architectures, using PSF as a benchmark shown in Eq. (1). In this study, we regard Inter-Net architecture as the standard and construct a generic convolutional network (CNN-Net) and a residual network

Table 2 Comparison of the segmentation accuracy of different experiments and other methods

	DSC	PRE	RVD	RMS	Number of volumes tested
Experiment					550
Inter-Net (E_{PSF})	0.913 ± 0.025	0.908 ± 0.0368	0.046 ± 0.039	0.492 ± 0.123	
Inter-Net (E_{CEF})	0.904 ± 0.029	0.912 ± 0.049	0.058 ± 0.049	0.514 ± 0.110	
Inter-Net (E_{PDF})	0.903 ± 0.033	0.886 ± 0.054	0.063 ± 0.057	0.496 ± 0.100	
Integrated	0.919 ± 0.023	0.926 ± 0.032	0.040 ± 0.036	0.464 ± 0.082	
Other Networks					550
CNN-Net (E_{PSF})	0.845 ± 0.0447	0.867 ± 0.094	0.145 ± 0.105	1.033 ± 0.859	
Res-Net (E_{PSF})	0.909 ± 0.0257	0.897 ± 0.037	0.048 ± 0.037	0.504 ± 0.184	
Other methods					
MAGeT [21]	0.869	0.894	–	0.48 ± 0.09	60
LEAP [55]	0.848 ± 0.031	–	–	–	796
MvU-Net [37]	0.895	–	–	–	110

(Res-Net). Figure 5 reveals the comparison results among Inter-Net, CNN-Net, and Res-Net, where the specific experimental results are also listed in the middle part of Table 2. Obviously, the interactive neural network outperforms Res-Net and CNN-Net, with high accuracy as well as small variance.

It is not a trivial task to perform an impartial comparison of the accuracy of different automatic techniques. Although Ghanei et al. [52] accomplished a DSC of 0.94, it was not convincing because it conducted only one sample test [17]. To increase the comparability of methods, we chose three typical techniques in the published articles, namely multiple automatically generated templates (MAGeT), learning embeddings for atlas propagation (LEAP), and multi-view U-ConvNet (MvU-Net), using the same data source ADNI as our experiments. In addition, the number of volumes tested is also shown in the last column to indicate the credibility of our comparison. The last part of Table 2 manifests the comparison results of different methods.

Looking at the three-part evaluations of Table 2, it is clear that the Inter-Net has an excellent performance in hippocampal segmentation whether in DSC, PRE, RVD, or RMS. For different objective functions, EPSF performs the best on DSC and RVD, EPDF focuses on RMS, and ECEF achieves high-precision. This is especially due to the metrics that have been optimized after integrating the multi-target loss function; its performance is the best and the most stable. Comparing the first and the third rows of Table 2, our model still outperforms the other three comparison techniques even before integrated.

Volumetric Correlation

DSC index is not sufficient to represent the level of segmentation system exactly though it has become the most popular indicator for assessing the performance of segmentation. Bland-Altman plots (Fig. 6; Table 3) assess the agreement between the automated segmentation and the gold standard by investigating the existence of any systematic differences with fixed bias. There is no doubt that the right bottom panel shows the densest, smallest standard deviation of differences. The points outside the limits of agreement (mean ± 1.96SD) indicate that the Inter-Net with E_{PSF} (left, upper) and the integrated model (right, bottom) segment the hippocampus overly, that is, a tendency to segment larger tissue than the standard.

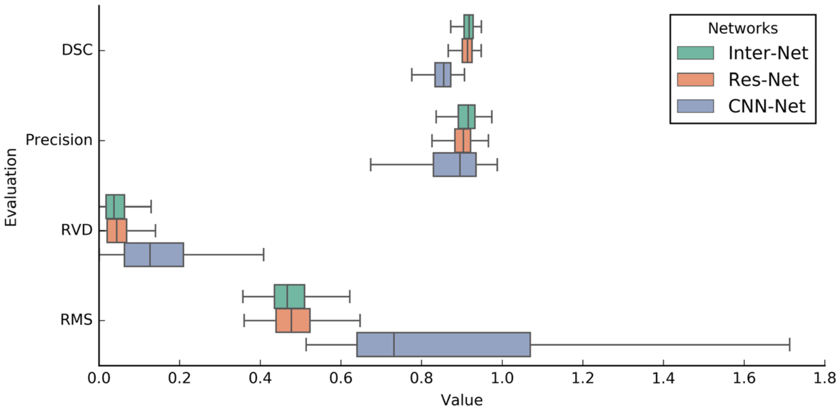

Fig. 5 Graphical box depiction of networks of the Inter-Net, the Res-Net, and the CNN-Net with PSF as the objective function by quartile

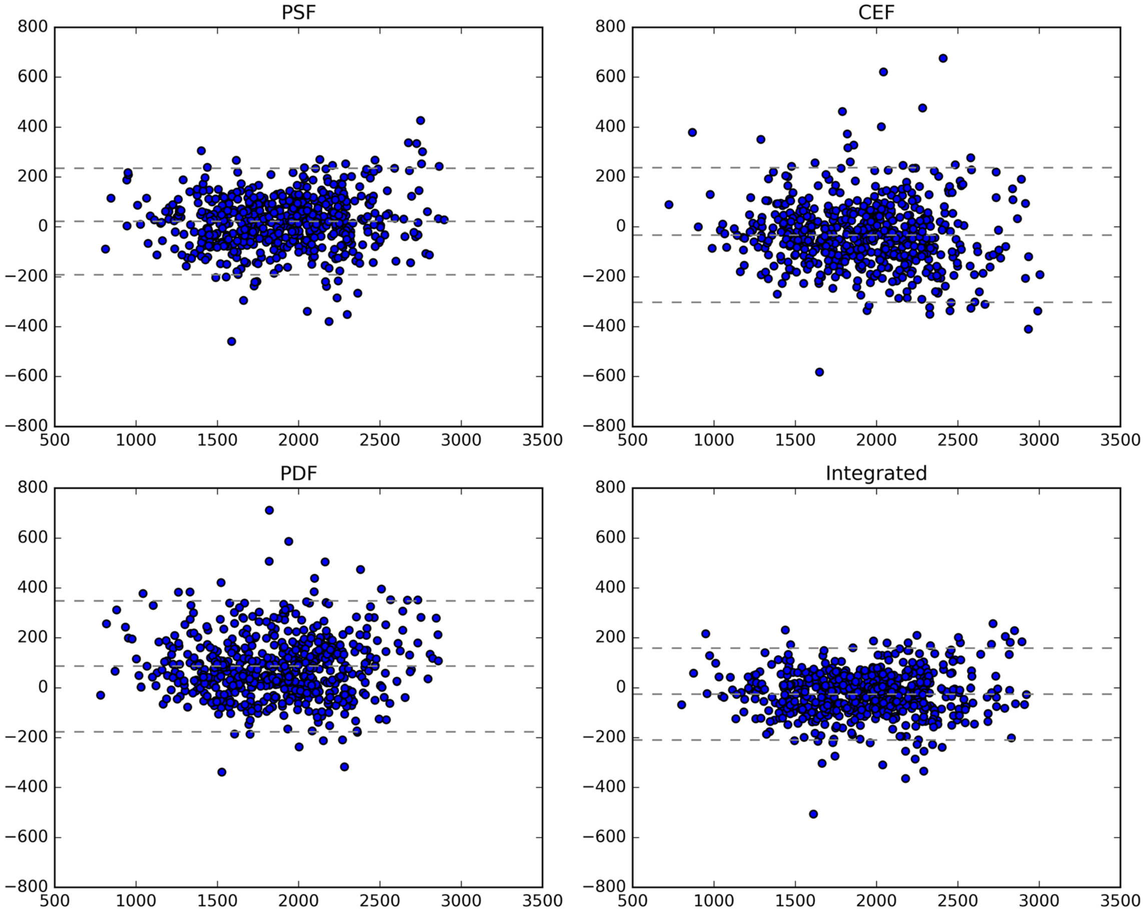

Fig. 6 Bland-Altman plots for volumetric comparisons between hippocampal volumes resulting from manual labeling versus the results of following objective models: PSF (upper, left), CEF (upper, right), PDF (bottom, left), and integrated (bottom, right)

Likewise, the extracted volumes are compared through linear regression per subject in Fig. 7. From the perspective of fitting curve alone, the integrated network and Inter-Net with E_{CEF} have similar performance, which is closer to $y = x$. Combining with the scatter plot distribution and the fitting curve, a closer relationship in total hippocampal volume between all methods and manual label volumes can be revealed in the subtle gaps. It is also notable that the allocation of the scatter is more concentrated around the fitting curve with the lowest deviation.

In terms of volumetric correlation, the integrated network performs the best, which combines the advantages of the high correlation of E_{PSF} and the closest fitting curve on E_{CEF}.

Discussion

In current clinical routines, the hippocampus has been regarded as an important biomarker for the assessment of neurological disease. And the manual notation is still the ideal standard for hippocampal segmentation. With an aim to relieve the workload of radiologists while also reducing the inconsistencies with intra- and inter-volume variability, a deep learning computational architecture known as the multi-target integrated interactive neural network is designed to segment the hippocampus in a data-driven manner. It is an end-to-end system which embeds hippocampal feature extraction along with characteristics revivification into a hierarchical network. Next, a shortcut connection is introduced to realize interactive information between high- and low-level characteristics. Furthermore, we pay special attention to the gradient dispersion and propose the interactive block which is borrowed from the residual network, achieving better segmentation results.

From an engineering perspective, the objective function is a mathematical expression of designed variables, reflecting some specific purpose to be pursued. In this paper, we design three objective functions from the aspects of evaluation index,

Table 3 Bland-Altman estimations for segmentation methods

Models	Mean	2.5% limit	97.5% limit	SD
Inter-Net (E_{PSF})	21.735	− 191.886	235.356	108.990
Inter-Net (E_{CEF})	− 32.647	− 302.455	237.161	137.657
Inter-Net (E_{PDF})	85.847	− 176.385	348.079	133.792
Integrated	− 25.475	− 209.294	158.345	93.785

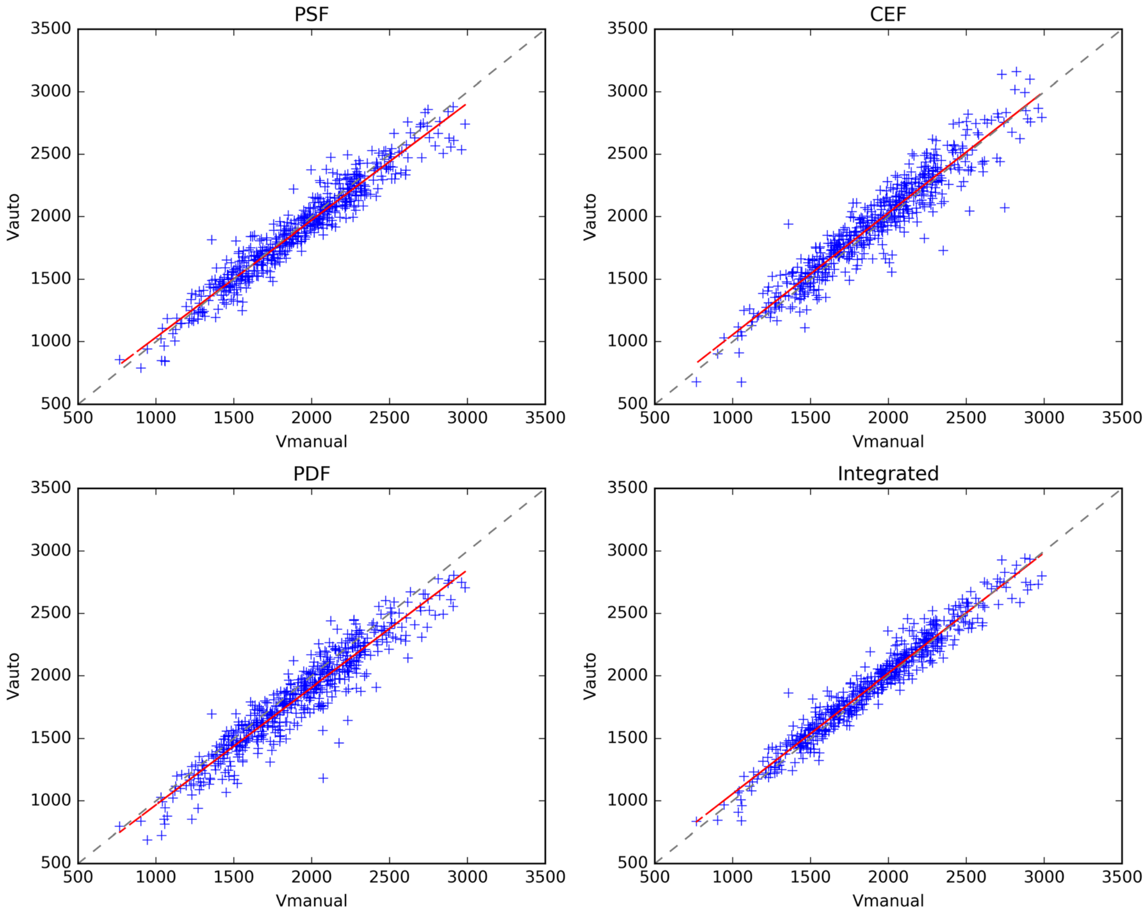

Fig. 7 Hippocampal volumes (mm^3) estimated by the manual and automated methods for the hippocampus and their linear regression

information theory, and data distribution. Then, we determine the corresponding weight coefficient through the grid search to maximize the DSC in the training dataset and construct the global segmentation network by $\lambda_{\mathrm{DSF}} = 0.5$, $\lambda_{\mathrm{CEF}} = 0.3$, $\lambda_{\mathrm{PDF}} = 0.2$. Analysis of voxel-, volume-, and distance-based metrics determines that the combination of the above objective function obtained a better overall performance of segmentation. The various evaluation indexes of different objective function networks are compared in the violin graph (see Fig. 8). It is revealed that the integrated network has varying degrees of improvement in each indicator and behaves more concentrated (the wider graphics in the middle line), which enhances the robustness of the system.

In order to profoundly explore the relationship between disease and model performance, we take AD as an example. Figure 9 presents the distribution of scatter density for disease group. As can be seen from this figure, the evaluation values of NC, MCI, and AD are all distributed evenly. Outliers are mostly caused by AD patients (orange part of the figure). There are two possible reasons for this phenomenon. One is the fact that the number of AD samples is less than that of MCI and NC, which leads to the model automatically deviating to the NC and MCI group during training. Another is the atrophy of the hippocampus which blurs the boundaries between the hippocampus and the surrounding tissue, increasing the difficulty of segmentation.

Relatively, a few studies analyzed the automated hippocampal segmentation regarding the left and right hippocampus as a whole. In the open toolbox about segmentation program, FreeSurfer provides multiclass segmentation based on Markov random field and probabilistic atlas. It is shown to have greater accuracy than other toolboxes especially in the head and tail portions of the hippocampus [15]. To further assimilate auxiliary segmentation results of the proposed approach, we take five samples from 550 scans randomly and perform the hippocampal segmentation using FreeSurfer 6.0. A comparison of the manual and automatic hippocampal probabilistic segmentation from a representative view is presented in Fig. 10. In the sagittal view, it is easy to measure the segmentation performance in details because of morphology. The second, third, and fourth columns represent the probabilistic segmentation results corresponding to three objective functions; the last column is the probabilistic results produced by FreeSurfer and purple circles the final segmentation area. Obviously, our model produces better segmentation results than FreeSurfer. After being integrated, the architecture

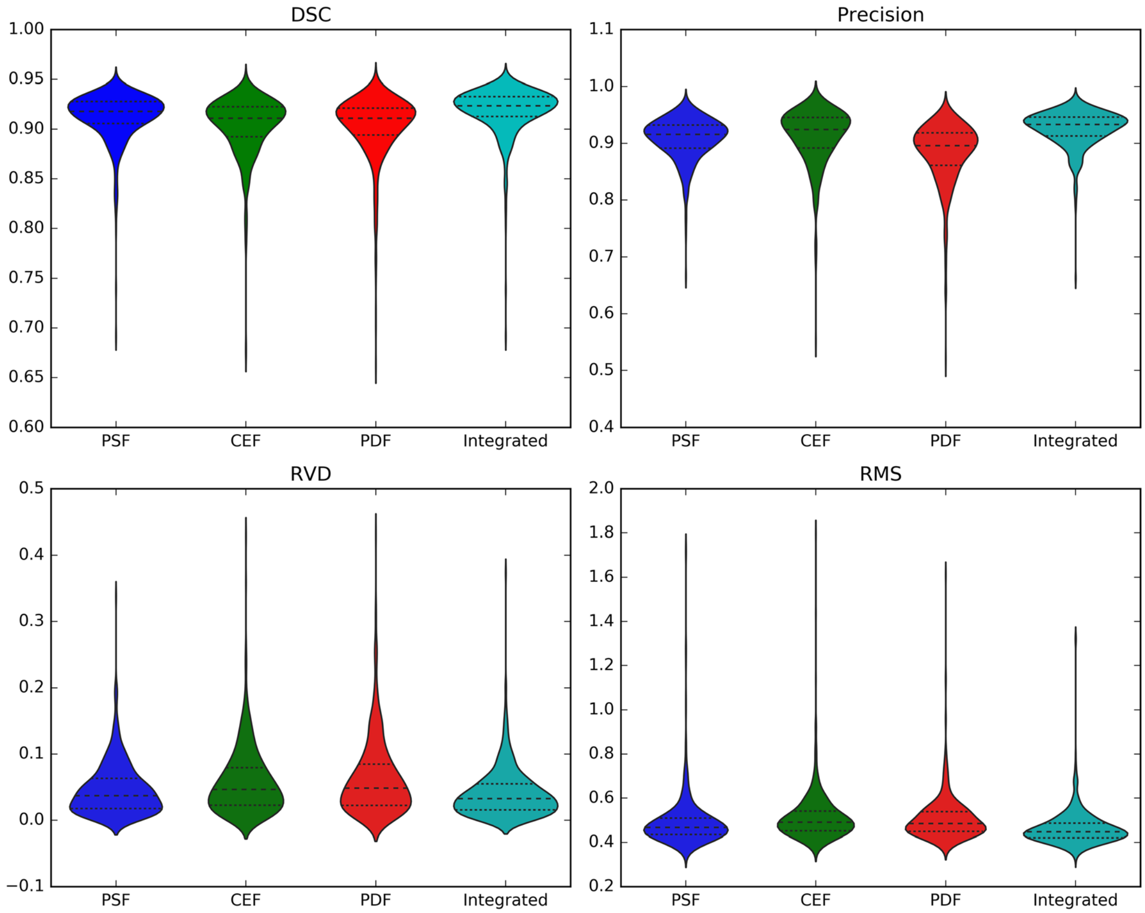

Fig. 8 Graphical violin depiction of models using DSC (top left corner), precision (top right corner), RVD (bottom left corner), and RMS (bottom right corner) on 550 cases by quartile

combines all the advantages of objective functions and is guaranteed both in accuracy and robustness.

The proposed automatic hippocampal segmentation architecture has great significance in clinical practice. As is known to all, operational complexity and running time are rudimentary elements in clinical application. In view of operational complexity, the framework proposed in this paper is based on deep learning, which is built upon the characterization of

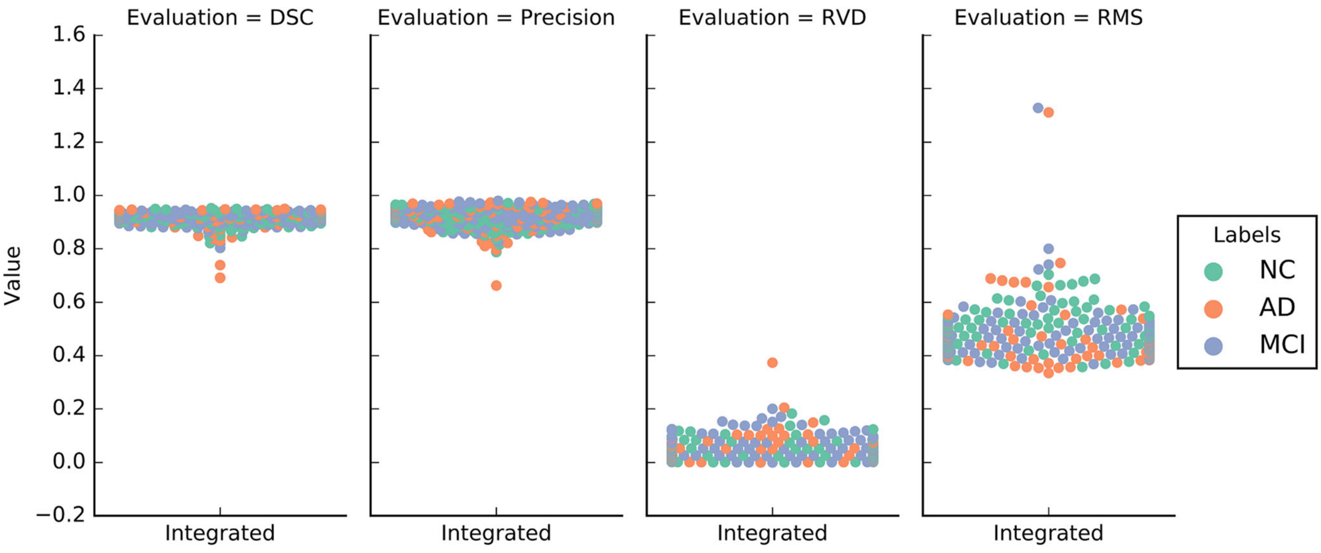

Fig. 9 Scatter density map by diagnosis group in each evaluation indicator

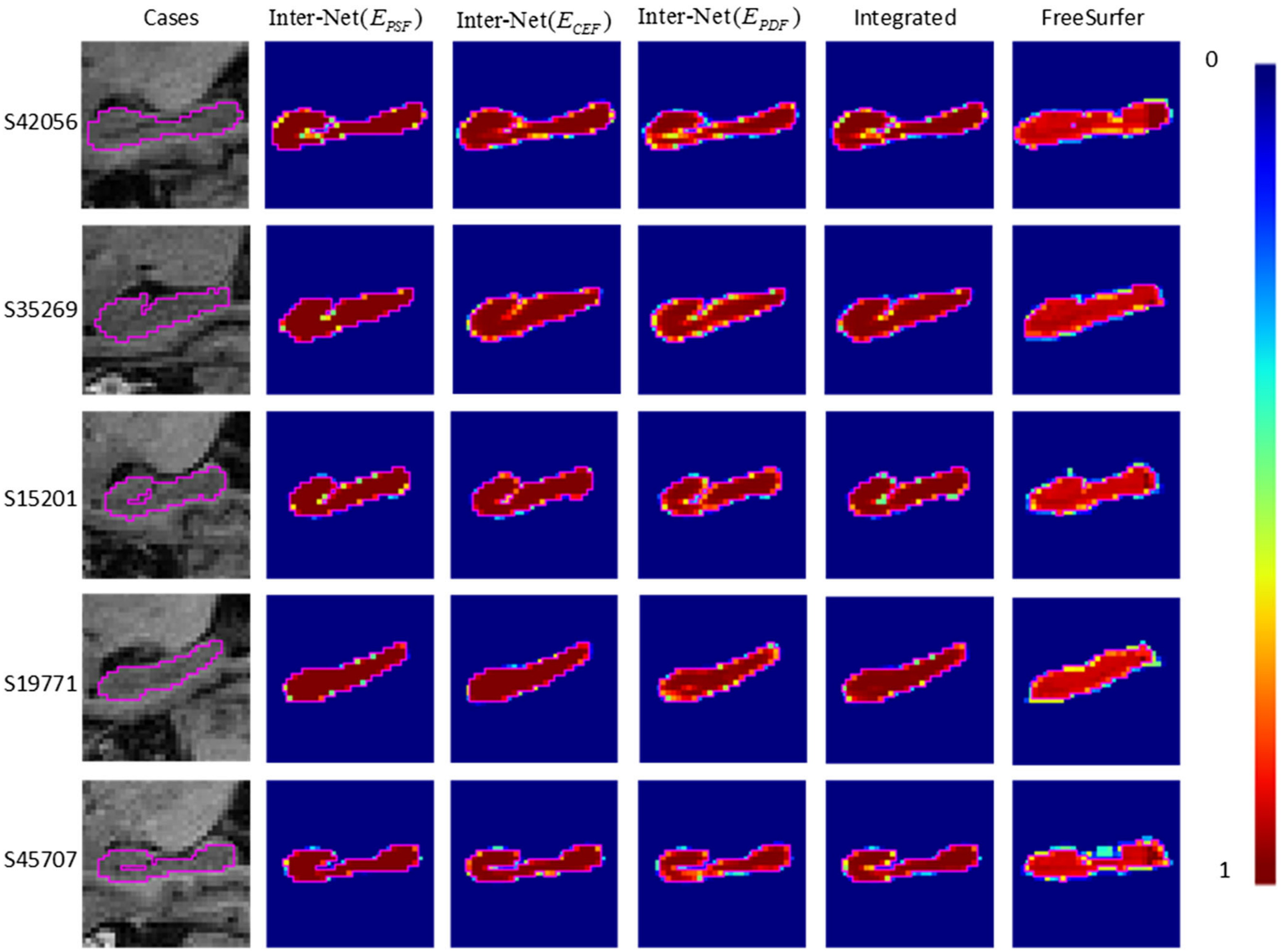

Fig. 10 Comparison of segmentation results for the left hippocampus on a transverse slice in sagittal view. The second, third, and fourth correspond to different objective function models, respectively. The final column represents FreeSurfer results

data and study convolutional filters through forward propagation and backward feedback. Once the convolutional kernels are determined, the mapping from input image I to the segmentation result A is also established and this entire process occurs without human intervention. In regard to the running time, the tested architecture running time is 42.1 s from the image reading to the segmentation result output. Moreover, the running time can be shortened to 29.4 s when only one objective network is used instead of three parallel architectures. There is a noticeable improvement compared to 2 min per volume reported in [53]. In brief, this work is derived from the clinical consideration, taking into account the operational complexity, the running time of the program, the accuracy of hippocampal segmentation, and so on.

In principle, our segmentation architecture is general and can be easily extended to other tissue segmentation in 3D medical images, such as the thalamus, corpus callosum, cerebellum segmentation, and so on. Since the location and shape of brain tissues are relatively fixed, prior knowledge concerning target tissue can be roughly used to select a VOI, which alleviates current computer memory limitation effectively. However, we must take the whole image as input when dealing with specific issues following little prior knowledge regarding location, size, and shape, take multiple sclerosis as an example, how to design an end-to-end architecture without human intervention to realize the precise segmentation of lesions. In our future work, we shall consider cascade interactive neural network with the entire image as an input, to reduce the dependence on prior knowledge and achieve automatic segmentation of various areas of interests.

Conclusion

In this paper, an effective and robust architecture is presented for hippocampal segmentation based on deep convolutional neural networks. On the one hand, we design an interactive neural network to achieve end-to-end segmentation of the hippocampus, which solves gradient dispersion by adding shortcut connections in intra- and inter-pathway. On the other hand, the concept of multi-target which combines the advantages of multiple objective functions is proposed to increase the robustness of the segmentation. Experimental results demonstrate that the proposed approach is superior to the traditional methods dramatically following with higher DSC. Furthermore, the proposed algorithm goes beyond hippocampal segmentation and it can be employed to other volumetric image segmentation tasks.

Funding This work was supported by National Natural Science Foundation of China (61471064), National Science and Technology Major Project of China (No.2017ZX03001022), and BUPT Excellent Ph.D. Students Foundation (No.CX2019309).

Compliance with Ethical Standards

Conflict of Interest The authors declare that they have no conflict of interest.

Ethical Approval This article does not contain any studies with human participants or animals performed by any of the authors.

References

1. Czepielewski LS, Wang L, Gama CS, et al. The relationship of intellectual functioning and cognitive performance to brain structure in schizophrenia. Schizophr Bull. 2017;43(2):355–64.
2. Steiger VR, Brühl AB, Weidt S, Delsignore A, Rufer M, Jäncke L, et al. Pattern of structural brain changes in social anxiety disorder after cognitive behavioral group therapy: a longitudinal multimodal MRI study. Mol Psychiatry. 2017;22(8):1164–71.
3. den Heijer T, van der Lijn F, Vernooij MW, et al. Structural and diffusion MRI measures of the hippocampus and memory performance. Neuroimage. 2012;63(4):1782–9.
4. Wixted JT, Squire LR. The medial temporal lobe and the attributes of memory. Trends Cogn Sci. 2011;15(5):210–7.
5. Jeneson A, Squire LR. Working memory, long-term memory, and medial temporal lobe function. Learn Mem. 2012;19(1):15–25.
6. Bobinski M, Wegiel J, Wisniewski HM, Tarnawski M, Bobinski M, Reisberg B, et al. Neurofibrillary pathology—correlation with hippocampal formation atrophy in Alzheimer disease. Neurobiol Aging. 1996;17(6):909–19.
7. Geuze E, Vermetten E, Bremner JD. MR-based in vivo hippocampal volumetrics: 1. Review of methodologies currently employed. Mol Psychiatry. 2005;10(2):147–59.
8. Knickmeyer RC, Gouttard S, Kang C, Evans D, Wilber K, Smith JK, et al. A structural MRI study of human brain development from birth to 2 years. J Neurosci. 2008;28(47):12176–82.
9. Filippi M, Rocca MA, Ciccarelli O, De Stefano N, Evangelou N, Kappos L, et al. MRI criteria for the diagnosis of multiple sclerosis: MAGNIMS consensus guidelines. Lancet Neurol. 2016;15(3): 292–303.
10. Jacobsen C, Hagemeier J, Myhr KM, Nyland H, Lode K, Bergsland N, et al. Brain atrophy and disability progression in multiple sclerosis patients: a 10-year follow-up study. J Neurol Neurosurg Psychiatry. 2014;85(10):1109–15.
11. Andreasen NC, Liu D, Ziebell S, Vora A, Ho BC. Relapse duration, treatment intensity, and brain tissue loss in schizophrenia: a prospective longitudinal MRI study. Am J Psychiatr. 2013;170(6): 609–15.
12. Scheenstra AEH, van de Ven RCG, van der Weerd L, van den Maagdenberg AM, Dijkstra J, Reiber JH. Automated segmentation of in vivo and ex vivo mouse brain magnetic resonance images. Mol Imaging. 2009;8(1):35–44.
13. Carmichael OT, Aizenstein HA, Davis SW, Becker JT, Thompson PM, Meltzer CC, et al. Atlas-based hippocampus segmentation in Alzheimer's disease and mild cognitive impairment. Neuroimage. 2005;27(4):979–90.
14. Chupin M, Mukuna-Bantumbakulu AR, Hasboun D, Bardinet E, Baillet S, Kinkingnéhun S, et al. Anatomically constrained region deformation for the automated segmentation of the hippocampus and the amygdala: method and validation on controls and patients with Alzheimer's disease. Neuroimage. 2007;34(3):996–1019.
15. Fischl B, Salat DH, Busa E, Albert M, Dieterich M, Haselgrove C, et al. Whole brain segmentation: automated labeling of neuroanatomical structures in the human brain. Neuron. 2002;33(3):341–55.
16. Zandifar A, Fonov V, Coupé P, Pruessner J, Collins DL, Alzheimer's Disease Neuroimaging Initiative. A comparison of accurate automatic hippocampal segmentation methods. NeuroImage. 2017;155:383–93.
17. Hosseini MP, Nazem Zadeh MR, Pompili D, Jafari-Khouzani K, Elisevich K, Soleanian-Zadeh H. Comparative performance evaluation of automated segmentation methods of hippocampus from magnetic resonance images of temporal lobe epilepsy patients. Med Phys. 2016;43(1):538–53.
18. Dill V, Franco AR, Pinho MS. Automated methods for hippocampus segmentation: the evolution and a review of the state of the art. Neuroinformatics. 2015;13(2):133–50.
19. Birenbaum A, Greenspan H. Multi-view longitudinal CNN for multiple sclerosis lesion segmentation. Eng Appl Artif Intell. 2017;65: 111–8.
20. Kwak K, Yoon U, Lee DK, Kim GH, Seo SW, Na DL. Fully-automated approach to hippocampus segmentation using a graph-cuts algorithm combined with atlas-based segmentation and morphological opening. Magn Reson Imaging. 2013;31(7):1190–6.
21. Pipitone J, Park MTM, Winterburn J, Lett TA, Lerch JP, Pruessner JC, et al. Multi-atlas segmentation of the whole hippocampus and subfields using multiple automatically generated templates. Neuroimage. 2014;101:494–512.
22. Sabuncu MR, Yeo BTT, Van Leemput K, Fischl B, Golland P. A generative model for image segmentation based on label fusion. IEEE Trans Med Imaging. 2010;29(10):1714–29.
23. Van der Lijn F, De Bruijne M, Klein S, Den Heijer T, Hoogendam YY, Van der Lugt A, et al. Automated brain structure segmentation based on atlas registration and appearance models. IEEE Trans Med Imaging. 2012;31(2):276–86.
24. Kim M, Wu G, Li W, Wang L, Son YD, Cho ZH, et al. Automatic hippocampus segmentation of 7.0 Tesla MR images by combining multiple atlases and auto-context models. NeuroImage. 2013;83: 335–45.
25. Hao Y, Wang T, Zhang X, Duan Y, Yu C, Jiang T, et al. Local label learning (LLL) for subcortical structure segmentation: application to hippocampus segmentation. Hum Brain Mapp. 2014;35(6): 2674–97.
26. Moghaddam MJ, Soltanian-Zadeh H. Automatic segmentation of brain structures using geometric moment invariants and artificial neural networks//International conference on Information Processing in Medical Imaging. Berlin: Springer; 2009. p. 326–37.
27. Long J, Shelhamer E, Darrell T. Fully convolutional networks for semantic segmentation. Proceedings of the IEEE conference on computer vision and pattern recognition, 2015. pp. 3431–3440.
28. Ronneberger O, Fischer P, Brox T. U-net: convolutional networks for biomedical image segmentation. International conference on medical image computing and computer-assisted intervention. Cham: Springer; 2015. p. 234–41.

29. Kamnitsas K, Ledig C, Newcombe VF, Simpson JP, Kane AD, Menon DK, et al. Efficient multi-scale 3D CNN with fully connected CRF for accurate brain lesion segmentation. Med Image Anal. 2017;36:61–78.
30. Liu X, Deng Z. Segmentation of drivable road using deep fully convolutional residual network with pyramid pooling. Cogn Comput. 2017:1–10.
31. Liu W, Tao D. Multiview Hessian regularization for image annotation. IEEE Trans Image Process. 2013;22(7):2676–87.
32. Liu W, Yang X, Tao D, Cheng J, Tang Y. Multiview dimension reduction via Hessian multiset canonical correlations. Information Fusion. 2018;41:119–28.
33. Yuan Y, Xun G, Ma F, et al. Muvan: a multi-view attention network for multivariate temporal data. 2018 IEEE International Conference on Data Mining (ICDM). Piscataway: IEEE; 2018. p. 717–26.
34. Kang G, Liu K, Hou B, Zhang N. 3D multi-view convolutional neural networks for lung nodule classification. PloS one, Public Library of Science. 2017;12(11):e0188290.
35. Setio AAA, Ciompi F, Litjens G, Gerke P, Jacobs C, van Riel SJ, et al. Pulmonary nodule detection in CT images: false positive reduction using multi-view convolutional networks. IEEE Trans Med Imaging. 2016;35(5):1160–9.
36. Fischl B. FreeSurfer. Neuroimage. 2012;62(2):774–81.
37. Chen Y, Shi B, Wang Z, Zhang P, Smith CD, Liu J. Hippocampus segmentation through multi-view ensemble ConvNets[C]// Biomedical Imaging (ISBI 2017), 2017 IEEE 14th International Symposium on. IEEE, 2017. pp. 192–196.
38. Jack CR Jr, Bernstein MA, Fox NC, et al. The Alzheimer's disease neuroimaging initiative (ADNI): MRI methods. J Magn Reson Imaging. 2008;27(4):685–91.
39. Wen G, Hou Z, Li H, Li D, Jiang L, Xun E. Ensemble of deep neural networks with probability-based fusion for facial expression recognition. Cogn Comput. 2017;9(5):597–610.
40. Brosch T, Tang LY, Yoo Y, Li DK. Deep 3D convolutional encoder networks with shortcuts for multiscale feature integration applied to multiple sclerosis lesion segmentation. IEEE Trans Med Imaging. 2016;35(5):1229–39.
41. Veit A, Wilber M, Belongie S. Residual networks are exponential ensembles of relatively shallow networks. arXiv preprint. arXiv preprint arXiv:1605.06431. 2016;1(2):3.
42. He K, Zhang X, Ren S, Sun J. Identity mappings in deep residual networks. European Conference on Computer Vision. Cham: Springer; 2016. p. 630–45.
43. Nair V, Hinton G E. Rectified linear units improve restricted boltzmann machines. Proceedings of the 27th international conference on machine learning (ICML-10). 2010. pp. 807–814.
44. Zeiler MD. ADADELTA: an adaptive learning rate method. arXiv preprint arXiv:1212.5701. 2012.
45. Kingma DP, Ba J. Adam: a method for stochastic optimization. arXiv preprint arXiv:1412.6980. 2014.
46. Dauphin Y, de Vries H, Bengio Y. Equilibrated adaptive learning rates for non-convex optimization[C]. Adv Neural Inf Proces Syst. 2015:1504–12.
47. Srivastava N, Hinton G, Krizhevsky A, Stuskever I, Salakhutdinov R. Dropout: a simple way to prevent neural networks from overfitting. J Mach Learn Res. 2014;15(1):1929–58.
48. Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556. 2014.
49. Zeng D, Zhao F, Shen W, Ge S. Compressing and accelerating neural network for facial point localization. Cogn Comput. 2017: 1–9.
50. Dice LR. Measures of the amount of ecologic association between species. Ecology. 1945;26(3):297–302.
51. Cabezas M, Oliver A, Lladó X, Freixenet J, Cuadra MB. A review of atlas-based segmentation for magnetic resonance brain images. Comput Methods Prog Biomed. 2011;104(3):e158–77.
52. Ghanei A, Soltanian-Zadeh H, Windham JP. A 3D deformable surface model for segmentation of objects from volumetric data in medical images. Comput Biol Med. 1998;28(3):239–2.
53. Lötjönen JMP, Wolz R, Koikkalainen JR, Thurfjell L, Waldemar G, Soininen H, et al. Fast and robust multi-atlas segmentation of brain magnetic resonance images. Neuroimage. 2010;49(3):2352–65.
54. 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. 2016. pp. 770–778.
55. Wolz R, Aljabar P, Hajnal JV, Hammers A, Rueckert D. Alzheimer's Disease Neuroimaging Initiative. LEAP: learning embeddings for atlas propagation. NeuroImage. 2010;49(2):1316–25.