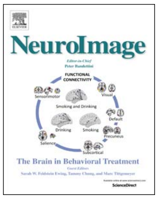
Contents lists available at ScienceDirect

NeuroImage

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



VoxResNet: Deep voxelwise residual networks for brain segmentation from 3D MR images

Hao Chen[a,*], Qi Dou[a,*], Lequan Yu[a], Jing Qin[b], Pheng-Ann Heng[a,c]

[a] *Department of Computer Science and Engineering, The Chinese University of Hong Kong, Hong Kong, China*
[b] *School of Nursing, The Hong Kong Polytechnic University, Hong Kong, China*
[c] *Guangdong Provincial Key Laboratory of Computer Vision and Virtual Reality Technology, Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences, Shenzhen, China*

ARTICLE INFO

Keywords:
Convolutional neural network
3D deep learning
Residual learning
Brain segmentation
Auto-context
Multi-modality
Multi-level contextual information

ABSTRACT

Segmentation of key brain tissues from 3D medical images is of great significance for brain disease diagnosis, progression assessment and monitoring of neurologic conditions. While manual segmentation is time-consuming, laborious, and subjective, automated segmentation is quite challenging due to the complicated anatomical environment of brain and the large variations of brain tissues. We propose a novel voxelwise residual network (*VoxResNet*) with a set of effective training schemes to cope with this challenging problem. The main merit of residual learning is that it can alleviate the degradation problem when training a deep network so that the performance gains achieved by increasing the network depth can be fully leveraged. With this technique, our *VoxResNet* is built with 25 layers, and hence can generate more representative features to deal with the large variations of brain tissues than its rivals using hand-crafted features or shallower networks. In order to effectively train such a deep network with limited training data for brain segmentation, we seamlessly integrate multi-modality and multi-level contextual information into our network, so that the complementary information of different modalities can be harnessed and features of different scales can be exploited. Furthermore, an auto-context version of the *VoxResNet* is proposed by combining the low-level image appearance features, implicit shape information, and high-level context together for further improving the segmentation performance. Extensive experiments on the well-known benchmark (i.e., *MRBrainS*) of brain segmentation from 3D magnetic resonance (MR) images corroborated the efficacy of the proposed *VoxResNet*. Our method achieved the first place in the challenge out of 37 competitors including several state-of-the-art brain segmentation methods. Our method is inherently general and can be readily applied as a powerful tool to many brain-related studies, where accurate segmentation of brain structures is critical.

1. Introduction

Brain parcellation from volumetric medical images, especially 3D magnetic resonance (MR) images, is a prerequisite for quantifying the structural volumes. It is of great significance on diagnosis, progression assessment, and treatment of a wide range of neurodegenerative diseases such as dementia and Alzheimer's disease (Petrella et al., 2003; Giorgio and De Stefano, 2013). Particularly, the segmentation of brain tissues into white matter (WM), gray matter (GM), and cerebrospinal fluid (CSF) is essential for measuring and visualizing anatomical structures (Wright et al., 2014), analyzing the brain changes (Zhang et al., 2015), conducting large-scale studies with images acquired at all ages (Moeskops et al., 2016; Thambisetty et al., 2010), making surgical planning and performing image-guided interventions (Despotović et al., 2015).

However, manual segmentation of brain structures from 3D images is an extremely laborious and time-consuming task, which requires a sophisticated knowledge base of brain anatomy, and is difficult, if not impossible, to be performed at a large scale. Furthermore, manual segmentation suffers from low reproducibility, which is easily prone to errors due to inter- or intra-operator variabilities. In this regard, automated segmentation methods are highly desired in practice for providing consistent measurements and quantitative analyses. However, the automatic brain segmentation is quite challenging due to the low contrast of anatomical structures in some modalities, the large intra-class variations of these structures among different subjects (Moeskops et al., 2016) or caused by various lesions (Menze et al., 2015; Maier et al., 2017), the confounding appearance of different

[*] Corresponding authors.
E-mail addresses: hchen@cse.cuhk.edu.hk (H. Chen), qdou@cse.cuhk.edu.hk (Q. Dou).

http://dx.doi.org/10.1016/j.neuroimage.2017.04.041
Accepted 18 April 2017

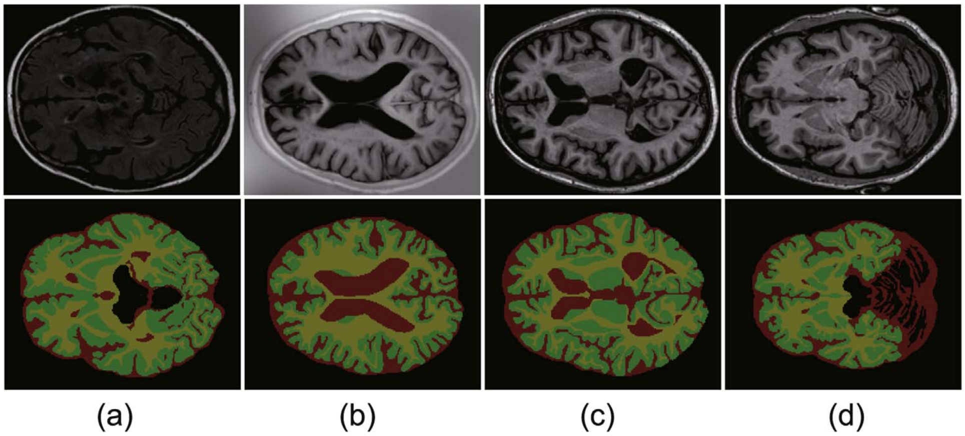


Fig. 1. Examples of brain images from different image modalities (the first row) and the segmentation annotations by experts (the second row): (a) T2-FLAIR modality of subject2, (b) T1-IR modality of subject2, (c) T1 modality of subject2, (d) T1 modality of subject4 (yellow, green, and red colors represent the WM, GM, and CSF, respectively). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article).

inter-class anatomical regions, etc. In Fig. 1, we illustrate the appearance of key structures in brain, including WM, GM, and CSF, from different image modalities. From these images, we can observe the challenges of brain segmentation mentioned above. In addition, the imaging acquisition protocol can affect the imaging quality significantly, which further poses challenges for automated segmentation methods (Despotović et al., 2015).

In the past decade, a lot of automated methods have been developed for brain segmentation. Broadly speaking, they can be categorized into three classes. **(1) Machine learning based methods with hand-crafted features.** This kind of method employs different classifiers with various hand-crafted features, such as support vector machines (SVM) with spatial and intensity features (van Opbroek et al., 2013; Moeskops and Benders, 2015a, 2015b), Gaussian mixture models (GMM) with intensity features (Ashburner and Friston, 2005; Rajchl et al., 2013; Prakash and Kumari), random forests (RF) with 3D Haar like features (Wang et al., 2015) or appearance as well as spatial features (Pereira et al., 2016). However, the main limitation of these methods is that hand-crafted features usually suffer from limited representation capability for accurate recognition, considering the large variations of brain structures. **(2) Deep learning based methods with automatically end-to-end learned features.** These methods learn the feature representations in a data-driven way, such as 3D convolutional neural network (Cicek et al., 2016), parallelized long short-term memory (LSTM) (Stollenga et al., 2015), convolutional neural networks with multiple pathways (de Brebisson and Montana, 2015) or multiple patch and kernel sizes (Moeskops et al., 2016), and 2D fully convolutional networks (FCN) (Nie et al., 2016). These methods can achieve more accurate segmentation results without manually designing sophisticated input features explicitly. Nevertheless, more elegant architectures such as residual networks are required to further advance the performance. In addition, the complementary information of different modalities and multi-level contextual features should be sufficiently considered to enhance the discrimination capability of the generated features. **(3) Multi-atlas registration based methods** (Klein and Hirsch, 2005; Aljabar et al., 2009; Artaechevarria et al., 2009; Sarikaya et al., 2013; Habas et al., 2010; Shi et al., 2010). For example, multi-atlas label fusion (MALF) made use of multiple reference atlases and achieved good performance in brain segmentation tasks (Aljabar et al., 2009; Lötjönen et al., 2010; Wang et al., 2013; Heckemann et al., 2006). However, current MALF methods often employ single image modality for segmentation or treat each modality equally when employing multiple image modalities. Furthermore, these methods are usually computationally expensive, making them infeasible to be used in applications requiring fast processing speed. In addition, the errors originated from the registration process can decrease the accuracy of fusion results from multiple atlases.

Recent years, deep learning especially deep convolutional neural networks (CNNs) have emerged as one of the most prominent approaches for image recognition problems in both natural image processing (Krizhevsky et al., 2012; Simonyan and Zisserman, 2014; Long et al., 2015; Szegedy et al., 2015; Chen et al., 2015c; Ji et al., 2013) and medical image analysis (Prasoon et al., 2013; Ronneberger et al., 2015; Chen et al., 2015a; Shin et al., 2016; Nogues et al., 2016; Li et al., 2014; Zheng et al., 2015; Chen et al., 2017). Although significant improvements have been achieved in many applications compared to previous methods employing hand-crafted features, most of these studies focused on the 2D images. However, in the field of medical image computing, volumetric data accounts for a large portion of medical image modalities, such as 3D computed tomography (CT), 3D MR images, 3D ultrasound, etc. Note that developing an effective 3D neural network is quite challenging due to not only the higher dimensionality but also the more complicated anatomical environment along with volumetric data than in 2D images.

To our best knowledge, nowadays there are two main types of CNNs developed for volumetric image processing. The first type employed modified variants of 2D CNNs by taking single slice (Lee et al., 2011), aggregated adjacent slices (Chen et al., 2015b; Zhang et al., 2015) or orthogonal planes (i.e., axial, coronal and sagittal) (Prasoon et al., 2013; Roth et al., 2014) as input to make up three dimensional spatial information. Although preliminary good performance has been validated, these methods cannot sufficiently make use of the 3D contextual information, which greatly limited their capability to segment objects from volumetric data more accurately. The other type of methods employed real 3D CNNs to detect or segment objects from volumetric data and demonstrated compelling performance (Dou et al.; Dou et al., 2016a, 2016b; Cicek et al., 2016; Yu et al., 2017; Merkow et al., 2015; Milletari et al., 2016; Kamnitsas et al., 2017; Chen et al., 2016a). Nevertheless, these methods may suffer from limited representation capability using a relatively shallow network. On the other hand, when we want to train deep neural networks to capture more representative features, we may confront the *degradation problem*, where the performance of the network gets saturated and then degrades rapidly if simply increasing the depth of network without any effective training schemes (He et al., 2016a).

Recently, deep residual learning with substantially enlarged depth advanced the state-of-the-art performance on 2D image recognition tasks (He et al., 2016a, 2016b; Lequan et al., 2016). Instead of simply stacking layers, it alleviated the optimization degradation issue by approximating the objective function with residual functions, which are *skip connections* between layers of the network. Such a technique allows a network to pass derivatives backwards through the network sometimes skipping layers (i.e., not passing through all non-linearities). In this paper, we propose a novel voxelwise residual network (*VoxResNet*) to cope with the challenging problem of segmentation of key brain tissues from 3D MR images by introducing residual learning to volumetric data processing. As mentioned previously, the main merit of residual learning is that it can alleviate the degradation problem when training a deeper network so that the performance gains achieved by increasing network depth can be fully leveraged. With this technique, our *VoxResNet* is built with 25 layers, and hence can generate more powerful features to deal with the large variations of brain tissues than its competitors either using hand-crafted features or applying shallower networks. In order to effectively train such a deep network for brain segmentation, we seamlessly integrate multi-modality and multi-level contextual information into our network, so that the complementary information of different modalities can be harnessed and features of different scales can be exploited. Furthermore, an auto-context version of *VoxResNet* is proposed by combining the low-level image appearance features, implicit shape information, and high-level context together for further improving the segmentation performance. The auto-context is a well-known and effective algorithm for image segmentation by integrating low-level and context information through fusing a large number of low-level appearance features with context and implicit shape information (Tu, 2008; Tu and Bai, 2010). Extensive experiments on the well-known benchmark (i.e., *MRBrainS*) of brain segmentation from 3D MR images corroborated the efficacy of the proposed *VoxResNet*. Our method achieved the first place in the challenge out of 37 competitors including several state-of-the-art brain segmentation methods. Our main contributions can be summarised as follows:

1) We propose a novel deep voxelwise residual network, referred as *VoxResNet*, which borrows the spirit of deep residual learning from 2D image recognition tasks and extends it into a 3D variant to fully explore the volumetric spatial information for accurate segmentation of brain structures from 3D MR images.

2) To tackle the large variation of brain structures, we validate the efficacy and necessity of complementary information from multiple imaging modalities and multi-level contextual feature representations by integrating them within our unified deep learning framework.

3) An auto-context version of *VoxResNet* is proposed by seamlessly integrating the low-level image appearance features, implicit shape information, and high-level context together for further improving the volumetric segmentation performance. Extensive experiments on a well-known benchmark dataset corroborated the efficacy of our method, outperforming other state-of-the-art methods by a great margin.

The remainder of paper is organized as follows. In Section 2, we first describe the experimental datasets, then elaborate deep residual learning for effective feature representations and detail the proposed *VoxResNet* for volumetric brain segmentation. We report the experiments and results in Section 3. We further discuss and analyze our study in Section 4. Finally, conclusions are drawn in Section 5.

2. Material and methods

2.1. Data acquisition and pre-processing

We validated our method on the 2013 MICCAI *MRBrainS* challenge, which is a well-known benchmark for evaluating algorithms on brain segmentation. The target of *MRBrainS* challenge is to segment the brain into four-class structures, i.e., WM, GM, CSF, and background. The datasets were acquired at the UMC Utrecht of patients with diabetes and matched controls with varying degrees of atrophy and white matter lesions (Mendrik et al., 2015). Multi-sequence 3 T MRI brain scans, including T1, T1-IR, and T2-FLAIR, were acquired from each subject. All scans were bias corrected using SPM8 (Penny et al., 2011) and three 3D images with different sequences (modalities) for each patient were aligned by rigid registration using Elastix (Klein et al., 2010). The voxel spacing of all provided sequences (i.e., T1, T1-IR, and T2-FLAIR) was $0.958\ \text{mm} \times 0.958\ \text{mm} \times 3.00\ \text{mm}$ after bias correction and registration. The training dataset consisted of five representative subjects (2 male and 3 female, varying degrees of atrophy and white matter lesions) with manual segmentations provided. The test data included 15 subjects with ground truth held out by the challenge organizers for independent and fair evaluation.

Following the method in Stollenga et al. (2015), we pre-processed the images by subtracting Gaussian smoothed image and applying Contrast-Limited Adaptive Histogram Equalization (CLAHE) to enhance the local contrast. Then six input volumes of each subject, including three original images and three pre-processed ones, were used as input data in our experiments. To reduce the variations of input data, we normalized the intensities of each slice with zero mean and unit variance before feeding them into our network.

2.2. Deep residual learning

Deep neural networks hierarchically stack multiple layers of neurons, forming a low-middle-high feature representation and classifier in an end-to-end way (LeCun et al., 2015; Schmidhuber, 2015; Bengio, 2009). Previous studies have evidenced that the network depth is of crucial importance on the feature representations (Simonyan and Zisserman, 2014; Szegedy et al., 2015). It is postulated that more stacked layers may improve the discrimination capability of a network. However, deeper neural networks are usually more difficult to train than their shallower counterparts, because simply increasing the depth of a network does not always lead to improvements. This phenomenon is known as the degradation problem and it becomes more severe with accuracy degrading rapidly when the network goes very deeper.

Recently, deep residual networks with residual units have shown compelling accuracy and nice convergence behaviors on several large-scale image recognition tasks, such as ImageNet (He et al., 2016a, 2016b) and MS COCO (Dai et al., 2016) competitions. These studies reveal that the residual learning framework is effective to ease the degradation problem to train a deeper network, as the residual learning mechanism makes the network easy to be optimized (He et al., 2016a, 2016b; Srivastava et al., 2015). By using identity mappings as the skip connections and after-addition activation, residual units allow signals to be directly propagated from one block to other blocks, as shown in Fig. 2(b). Therefore, the information encoded in the training data can be sufficiently and effectively exploited to increase the performance. Another merit of such skip connections is that they do not add extra parameters or computation complexity.

Generally, the residual unit can be expressed as following:

$$x_{l+1} = x_l + \mathcal{F}_l(x_l,\ W_l) \tag{1}$$

here the $\mathcal{F}_l$ denotes the residual function, i.e., a stack of two convolutional layers with batch normalization (BN) in our implementation as shown in Fig. 2(b), x_l is the input feature to the l-th residual unit and W_l is a set of weights correspondingly associated with the residual unit. The key idea of deep residual learning is to learn the additive residual function $\mathcal{F}_l$ with respect to the input feature x_l. Hence by unfolding above equation recursively, the $x_L\,(L > l \geq 1)$ can be derived as:

$$x_L = x_l + \sum_{i=l}^{L-1} \mathcal{F}_i(x_i,\ W_i). \tag{2}$$

Therefore, the feature x_L of any deeper layers can be represented as the feature x_l of shallow unit l plus summarized residual functions $\sum_{i=l}^{L-1} \mathcal{F}_i(x_i,\ W_i)$. According to the chain rule of backpropagation (LeCun et al., 1989), we get the derivatives as:

$$\frac{\partial \mathcal{L}}{\partial x_l} = \frac{\partial \mathcal{L}}{\partial x_L}\frac{\partial x_L}{\partial x_l} = \frac{\partial \mathcal{L}}{\partial x_L}(1 + \frac{\partial}{\partial x_l}\sum_{i=l}^{L-1} \mathcal{F}_i(x_i,\ W_i)) \tag{3}$$

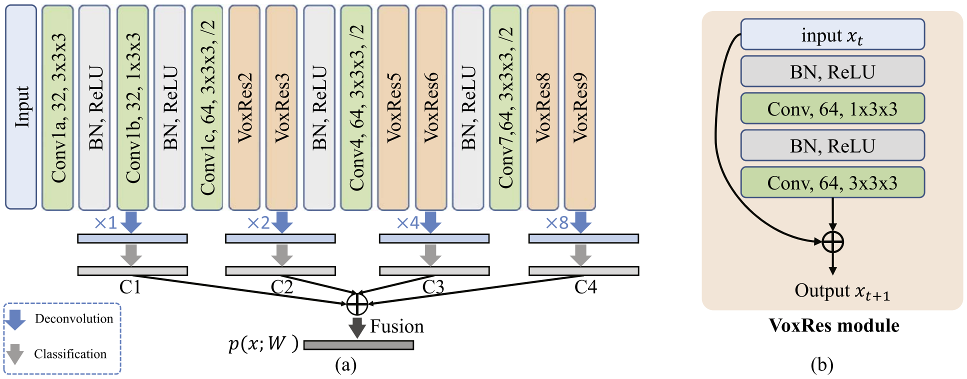


Fig. 2. (a) The architecture of proposed *VoxResNet* for volumetric image segmentation, consisting of batch normalization layers (BN), rectified linear units (ReLU), and convolutional layers N (ConvN) wtih number of channels, filter size and downsampling stride; (b) The illustration of VoxRes module.

where $\mathcal{L}$ denotes the loss function of deep residual networks. These derivations reveal that residual unit mechanism can make information propagate through the entire network smoothly in both forward and backward passes (He et al., 2016b).

2.3. VoxResNet for volumetric image segmentation

Although 2D deep residual networks have been employed in many 2D natural image processing tasks (He et al., 2016a, 2016b; Shen et al., 2016; Zagoruyko and Komodakis, 2016; Lequan et al., 2016), to the best of our knowledge, seldom studies of residual learning have been dedicated to processing volumetric data. However, in the medical image analysis domain, including neuroimaging, more and more volumetric data are employed in clinical practice as well as pathology research. In this work, we, for the first time, extend 2D deep residual networks to 3D residual networks in order to take the advantage of residual learning to extract more representative features to meet the challenges of brain segmentation from 3D MR data. The architecture of our neural network is shown in Fig. 2(a).

Our architecture consists of stacked residual modules (i.e., VoxRes module) with a total of 25 volumetric convolutional/deconvolutional layers (Long et al., 2015). In each VoxRes module, the input feature x_l and transformed feature $\mathcal{F}_l(x_l, W_l)$ are added together with skip connection shown in Fig. 2(b), and hence the information can be directly propagated in the forward and backward passes (He et al., 2016b). Note that in our network, all the operations are implemented in a 3D way to strengthen the volumetric feature representation learning. Following the principle from VGG network (Simonyan and Zisserman, 2014) and deep residual networks (He et al., 2016b), we employ small convolutional kernels (i.e., $1 \times 3 \times 3$ or $3 \times 3 \times 3$) in the convolutional layers, which have demonstrated evident advantages on computation efficiency and representation capability. Three convolutional layers are along with a stride 2, which reduced the resolution size of input volume by a factor of 8. This enables the network to have a large receptive field size and hence enclose more contextual information for improving the discrimination capability. Batch normalization layers are inserted into the architecture intermediately for reducing internal covariance shift in order to accelerate the training process and improve the performance (Ioffe and Szegedy, 2015). In our network, the rectified linear units, i.e., $f(x) = \max(0, x)$, are utilized as the activation function for non-linear transformation (Krizhevsky et al., 2012).

There is a huge variation on the size of 3D brain anatomical structures, which demands including a range of different receptive field sizes in our network for better recognition performance. In order to handle the large variation of size, we fuse multi-level contextual information (i.e., 4 auxiliary classifiers C1-C4 in Fig. 2(a)) with deep supervision (Lee et al., 2015; Chen et al., 2016b) in our network. To the end, the whole network is trained by minimizing following objective function with standard back-propagation:

$$\mathcal{L}(x, y; \theta) = \lambda\psi(\theta) - \sum_{\alpha}\sum_{x\in\mathcal{V}}\sum_{c} w_\alpha y_c^x \log p_c^\alpha(x; \theta) - \sum_{x\in\mathcal{V}}\sum_{c} y_c^x \log p_c(x; \theta) \quad (4)$$

where the first part is the regularization term (L_2 norm in our experiments) and latter one is the fidelity term consisting of auxiliary classifiers and final target classifier. The tradeoff of these terms is controlled by the hyperparameter λ. The w_α (where α indicates the index of auxiliary classifiers) is the weights of auxiliary classifiers, which were set as 1 initially and decreased till marginal values (i.e., 10^{-3}) in our experiments. The weights and biases of network are denoted as $\theta = \{W, b\}$, $p_c(x; \theta)$ or $p_c^\alpha(x; \theta)$ denotes the predicted probability of cth class after softmax classification layer for voxel x in volume space $\mathcal{V}$, and $y_c^x \in \{0, 1\}$ is the corresponding ground truth, i.e., $y_c^x = 1$ if voxel x belongs to the cth class, otherwise 0.

2.4. Multi-modality and auto-context information fusion

In the field of medical image computing, the volumetric data is usually acquired with multiple imaging modalities for robustly examining different tissue structures. For example, in our application, three imaging modalities including T1, T1-IR, and T2-FLAIR are available in the brain structure segmentation task (Mendrik et al., 2015). In other examples, four imaging modalities were used in brain tumor segmentation including T1, T1 contrast-enhanced, T2, and T2-FLAIR MRI (Menze et al., 2015) and four imaging modalities including T1-weighted, T2-weighted, diffusion weighted imaging (DWI), and FLAIR MRI were employed in brain lesion analysis (Maier et al., 2017). The main reason for acquiring multi-modality images is that the information from multi-modality dataset can be complementary, which provides more robust diagnosis results. Inspired by this clinical observation, we concatenate these multi-modality data as input channels into neural network, with weights from each modality into the first layer of feature maps. Hence the complementary information is jointly harnessed and fused during the training of network in an implicit way, which demonstrated consistent improvements compared to any single modality in our experiments.

Furthermore, in order to harness the integration of high-level context information, implicit shape information, and original low-level image appearance for improving recognition performance, we formulate the learning process as a version of auto-context algorithm (Tu, 2008). Auto-context is an effective algorithm for image segmentation by integrating the image appearances (i.e., volume data in our case) together with the context information by learning a series of classifiers such as random forests (Tu and Bai, 2010). Given a set of training data and their corresponding label maps, it first learns a classifier on local image patches. Then the probability maps generated by the learned classifier are used as high-level context information, in addition to the original image patches, to train a new classifier. The algorithm refines the output results in an iterative way. It integrates low-level and context information by fusing a large number of low-level appearance features with context and implicit shape information. Compared with the recognition tasks in natural image processing, the role of auto-context information can be more important in the medical domain as the anatomical structures are roughly positioned and constrained (Tu and Bai, 2010). Different from Tu and Bai (2010), which utilized the probabilistic boosting tree as the classifier, we employ the powerful deep residual networks as the classifier. Specifically, as shown in Fig. 3, given the training volumes, we first train a *VoxResNet* classifier on the original training sub-volumes with image appearance information. Then, the discriminative probability maps generated from *VoxResNet* are used as the context information, together with the original volumes (i.e., appearance information) as input, to train a new classifier *Auto-context VoxResNet*, which further refines the semantic segmentation results and removes the outliers. Different from the original auto-context algorithm, which performed in an iterative way (Tu and Bai, 2010), our empirical study showed that following iterative refinements bring few improvements. In this regard, we chosen the output of *Auto-context VoxResNet* as the final segmentation results.

3. Experiments and results

3.1. Evaluation metrics

The evaluation metrics of *MRBrainS* challenge consist of three types of measures: Dice coefficient (DC), the 95th-percentile of the Hausdorff distance (HD) and absolute volume difference (AVD), which are calculated for each tissue type (i.e., GM, WM, and CSF), respectively (Mendrik et al., 2015). The Dice coefficient measures the spatial overlap between the segmentation result and ground truth, with a larger value denoting a higher segmentation accuracy. It is defined as

$$D(G, S) = \frac{2|G \cap S|}{|G| + |S|}\cdot 100\% \tag{5}$$

where S is the set of segmentation results, G is the set of reference standard (i.e., ground truth), and D is the Dice coefficient expressed as percentages. Another measure is the Hausdorff distance, which is employed to measure the distance between the segmentation results and the ground truth. Because the conventional Hausdorff distance is very sensitive to the outliers, the Kth ranked distance, i.e., $h_{95}(S, G) = {}^{95}K^{th}_{s\in S} \min_{g\in G} \parallel g - s \parallel$, is used as to suppress the outliers (Huttenlocher et al., 1993). It is defined as

$$\mathrm{HD}(G, S) = \max\{h_{95}(S, G), h_{95}(G, S)\} \tag{6}$$

A smaller value of $\mathrm{HD}(G, S)$ denotes a higher proximity between ground truth and segmentation results, i.e., a higher segmentation accuracy. The last metric is the absolute volume difference, defined as

$$\mathrm{AVD}(G, S) = \frac{|V_s - V_g|}{V_g}\cdot 100\% \tag{7}$$

where V_s is the volume of segmentation results and V_g is the volume of ground truth. A smaller value of $\mathrm{AVD}(G, S)$ denotes a better segmentation accuracy. For the overall performance evaluation, each method is assigned one ranking number for each type of brain tissue based on the three metrics mentioned above using a standard competition ranking mechanism.[1] The sum score of these numbers is used for the final ranking, i.e., a smaller score stands for better overall segmentation performance. More details of evaluation can be found in the challenge website.[2]

3.2. The efficacy of multi-modality and auto-context information

To investigate the efficacy of employing multi-modality and auto-context information, we performed extensive ablation studies on the validation data (using subject 5 as validation data). The results of cross-validation using different modalities are reported in Table 1. Among the results of using only single image modality, we can see that the T1 image modality achieves overall better segmentation performance than other two image modalities, which indicates that this type of modality possesses higher image quality for discrimination on most of anatomical structures. When combining the multi-modality information from all available image modalities, the segmentation performance is obviously improved for almost all the evaluation metrics compared with that of any single image modality, especially on the metric of DC. This highlights the complementary characteristics of different imaging modalities. The example results of validation data using different image modalities can be see in Fig. 4. It is observed that the results using all image modalities are visually more accurate than those of single image modality.

It is also observed in Table 1 that by integrating the auto-context information, the performance of DC can be further improved. The qualitative results of brain segmentation with or without auto-context information can be seen in Fig. 5, which shows that the results by fusing auto-context information can generate more accurate results than the network without integrating it.

3.3. Comparison with other methods

Regarding the evaluation of testing data, we compared our method with both state-of-the-art deep learning based methods, including MDGRU, 3D U-net (Cicek et al., 2016), and PyraMiD-LSTM (Stollenga et al., 2015), and hand-crafted feature based methods, such as Mahbod (2016), Moeskops et al. (2015b), Wang et al. (2015) and Pereira et al. (2016). The MDGRU applied a neural network with the main components being multi-dimensional gated recurrent units and achieved quite good performance. The 3D U-net extended previous 2D version (Ronneberger et al., 2015) into a 3D variant and highlighted the necessity for volumetric feature representation when applied to 3D recognition tasks. The PyraMiD-LSTM parallelised the multi-dimensional recurrent neural networks in a pyramidal fashion and achieved compelling performance. These methods based on hand-crafted feature utilized various hand-crafted features, including histogram based features (Mahbod, 2016), gradients (Pereira et al., 2016), 3D Haar-like features (Wang et al., 2015), etc.

The results of different methods on *MRBrainS* challenge are reported in Table 2. It is observed that, generally, the deep learning based methods can achieve much better performance than hand-crafted feature based methods, which validated the superiority of feature learning representation of deep neural networks. The results of our *VoxResNet* (CU_DL in Table 2) by fusing multi-modality information achieved better performance than other deep learning based methods, which demonstrated the efficacy of our proposed framework. By incorporating the auto-context information in *Auto-context VoxResNet* (CU_DL2 in Table 2), the performance of DC can be further improved. Overall, our methods achieved the first place in the challenge leader board out of 37 competitors, outperforming other methods on most of evaluation metrics.

3.4. Implementation details

Our method was implemented using Matlab and C++ based on Caffe library (Jia et al., 2014; Tran et al., 2016).[3] We trained our neural

[1] https://en.wikipedia.org/wiki/Ranking, Dec, 17, 2016.

[2] MICCAI MRBrainS Challenge: http://mrbrains13.isi.uu.nl/details.php.

[3] The prototxt of our network architecture was provided at http://www.cse.cuhk.edu.hk/hchen/research/seg_brain.html.

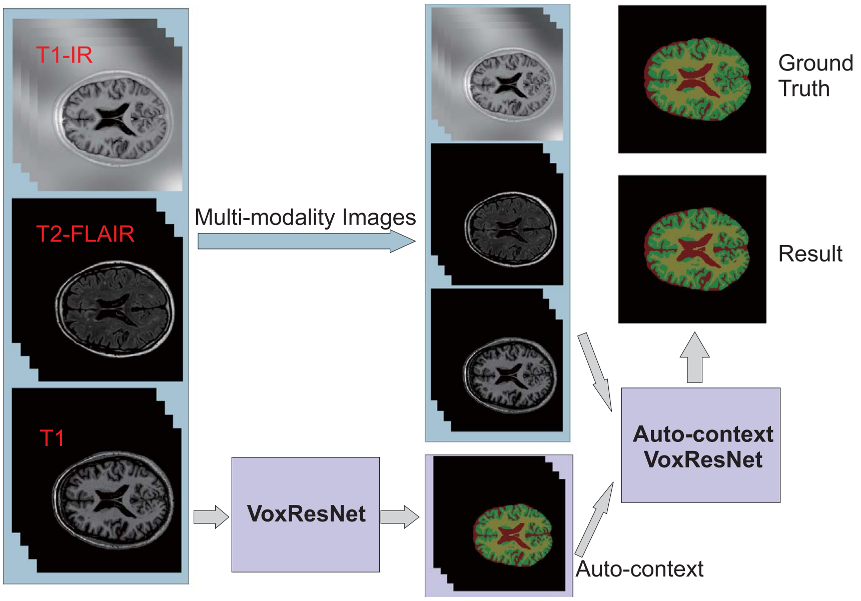


Fig. 3. An overview of our proposed framework for integrating auto-context with multi-modality information.

Table 1
Cross-validation results of MRI brain segmentation using different image modalities (DC: %, HD: mm, AVD: %).

Modality	GM			WM			CSF		
	DC	HD	AVD	DC	HD	AVD	DC	HD	AVD
T1	86.96	1.36	**4.67**	89.70	1.92	6.85	79.58	2.71	17.55
T1-IR	80.61	1.92	8.45	85.89	2.87	7.42	76.44	3.00	12.87
T2-FLAIR	81.13	1.92	9.15	83.21	3.00	4.99	75.34	3.03	3.77
All	88.08	**0.96**	5.82	90.93	**1.36**	2.05	82.51	**2.14**	3.90
All+auto-context	**88.50**	**0.96**	5.91	**91.06**	**1.36**	**1.05**	**82.70**	2.71	**2.50**

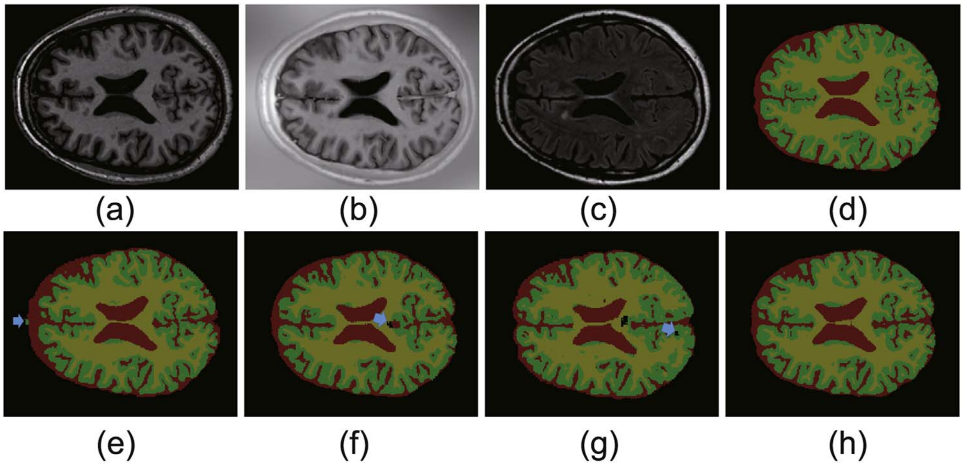


Fig. 4. Results of validation data using different modalities: (a)-(c) denote the original T1, T1-IR, and T2-FLAIR MR images and (e)-(g) show the corresponding segmentation result using single image modality, respectively; (d) is the ground truth label; (h) is the result using all image modalities without auto-context information (yellow, green, and red colors represent the WM, GM, and CSF, respectively). The blue arrows indicate some incorrect predictions with anatomical structure GM present in (e), WM absent in (f), and GM absent in (g). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article).

networks by using mini-batch gradient descent (the batch size was set as 5). It took about one day to train the network while less than 2 minutes for processing each test volume (size 240 × 240 × 48) using a standard workstation with one NVIDIA TITAN X GPU. Due to the limited capacity of GPU memory, we cropped sub-volume samples (size $80 \times 80 \times 48 \times m$, m is number of image modalities and set as 6 in our experiments) for the input into the network. In order to save the hard disk memory and augment the training data easily and flexibly, the extraction of training samples from the whole input volumes was implemented in an on-the-fly way during the training. A total of around 100,000 sub-volume samples were extracted for training the networks. In the test phase, the probability map of whole volume was generated in an overlap-tiling strategy for stitching the sub-volume results.

4. Discussion

We proposed a deep voxelwise residual network for brain segmentation from 3D MR images. The deep residual learning technique was originally developed for recognition in 2D images. In this paper, we generalize it with 3D convolutions and develop a set of effective training schemes for handling the segmentation of 3D brain MR images. The proposed *VoxResNet* can fully explore the spatial contextual information and generate more distinctive features to achieve

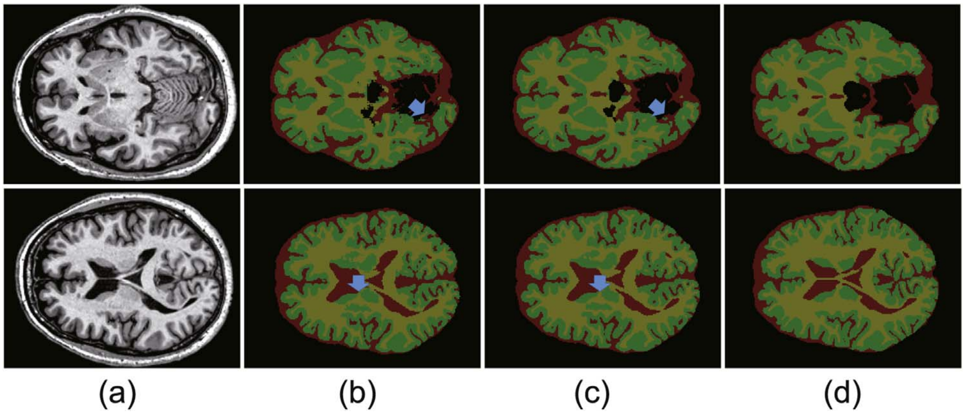


Fig. 5. Results of validation data using all image modalities: (a) original T1 MR images, (b) results of *VoxResNet*, (c) results of *Auto-context VoxResNet*, (d) ground truth labels (yellow, green, and red colors represent the WM, GM, and CSF, respectively). The blue arrows indicate some predictions with anatomical structures mistakenly classified in *VoxResNet* but correct in *Auto-context VoxResNet*. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article).

Table 2
Results of MICCAI MRBrainS challenge of different methods (DC: %, HD: mm, AVD: %. only top 11 teams are shown here).

Methods	GM			WM			CSF			Score*
	DC	HD	AVD	DC	HD	AVD	DC	HD	AVD	
CU_DL (ours)	86.12	1.47	6.42	89.39	**1.94**	**5.84**	83.96	2.28	7.44	**39**
CU_DL2 (ours)	**86.15**	**1.45**	6.60	**89.46**	**1.94**	6.05	**84.25**	2.19	7.69	**39**
MDGRU	85.40	1.55	6.09	88.98	2.02	7.69	84.13	2.17	7.44	57
PyraMiD-LSTM2	84.89	1.67	6.35	88.53	2.07	5.93	83.05	2.30	7.17	59
FBI/LMB Freiburg (Cicek et al., 2016)	85.44	1.58	6.60	88.86	1.95	6.47	83.47	2.22	8.63	61
IDSIA (Stollenga et al., 2015)	84.82	1.70	6.77	88.33	2.08	7.05	83.72	**2.14**	7.09	77
STH (Mahbod, 2016)	84.77	1.71	6.02	88.45	2.34	7.67	82.77	2.31	**6.73**	86
ISI-Neonatology (Moeskops et al., 2015b)	85.77	1.62	6.62	88.66	2.07	6.96	81.08	2.65	9.77	87
UNC-IDEA (Wang et al., 2015)	84.36	1.62	7.04	88.68	2.06	6.46	82.81	2.35	10.5	90
MNAB2 (Pereira et al., 2016)	84.50	1.70	7.10	88.04	2.12	7.74	82.30	2.27	8.73	109
KSOM GHMF (Rajchl et al., 2013)	84.13	1.92	**5.44**	87.96	2.49	6.59	82.10	2.71	12.78	112

* Score=Rank DC+Rank HD+Rank AVD; a smaller score means better performance.

much better performance compared to the competitors using either variants of 2D CNNs or relatively shallower networks. This work demonstrates that (1) increasing network depth can lead to performance improvements and (2) residual connections can effectively ease the degradation problems when training a deep network. Actually, the auxiliary classifiers for integrating multi-level contextual information can also be considered as long-range residual connections by regarding the intermediate layers as transformation functions. Extensive experiments on the well-known benchmark dataset demonstrated the superiority of our method, outperforming all rivals by a large margin.

Our method offers neuroimaging and neuroscience researchers a powerful tool for automated and accurate brain structure segmentation, which plays a significant role in quantification of brain structures for diagnosis, assessment, and treatment of various neurologic diseases. One of the main challenges for automated and accurate brain structure segmentation is that the contrast of various tissues may vary significantly in different imaging modalities. For example, in T1 images, tissues with high fat content, e.g., white matter, appear bright, which is contrary to their appearance in T2-weighted images. Therefore, combining the multi-modality information can improve the recognition performance. Our neural network can effectively harness the complementary multi-modal information for more accurate segmentation of multiple tissues in the brain. In our experiments, we validated that fusing multi-modality information can dramatically improve the segmentation performance than any single image modality. For example, as shown in Table 1, the Dice coefficients of CSF from T1, T1-IR, and T2-FLAIR on validation data are 79.58%, 76.44%, and 75.34%, respectively, while fusing the multi-modality information can improve the Dice coefficient to 82.51%. Integration of the auto-contextual information can further improve the segmentation results.

There is a large variation of brain structures, either in shape or size. Hence, it is necessary to generate multi-scale features to cover such a large variation to achieve more accurate results. In the architecture of convolutional neural network, the size of receptive field enclosing contextual information is becoming larger along with going deeper into the network. To probe the influence of multi-level contextual information, we report the results of different auxiliary classifiers encoding features with different levels in Table 3 (C1–C4 is illustrated in Fig. 2(a)). It is observed that the classifier C2 generated better performance than classifier C1 and C3 on GM segmentation while worse performance than C1 on WM and CSF, indicating different anatomical structures indeed require different receptive field sizes. The performance of C3 was dropped significantly compared to C1 and C2 due to the inclusion of several max-pooling layers, which leads to the spatial information loss. Integrating the multi-level information can give much better performance than single level contextual information on most of evaluation metrics. By fusing the information from C1-C4, the network generated the best segmentation results regarding the overall performance. Typical segmentation results of validation data using different levels of contextual information are shown in Fig. 6. We can see that by aggregating the multi-level information from C1 to C4, the segmentation results gradually become more and more accurate compared with the ground truth.

Regarding the depth of deep neural networks, the situation is different in our application compared to the original application of residual networks. The underlying problem of brain segmentation is inherently a segmentation task with voxel-wise predictions while previous deep residual learning is employed in an image-level classi-

Table 3
Results of MRI brain segmentation using different levels of contextual information (DC: %, HD: mm, AVD: %).

Information level	GM			WM			CSF		
	DC	HD	AVD	DC	HD	AVD	DC	HD	AVD
C1	80.75	13.86	5.36	86.76	5.23	**1.78**	76.99	4.26	19.42
C2	80.96	2.14	**3.40**	83.40	3.03	3.54	75.24	2.87	19.97
C3	60.41	4.16	9.93	69.13	7.37	6.87	53.11	11.70	8.73
C1–C2	86.62	**0.96**	8.25	90.55	**0.96**	2.41	81.35	3.00	19.55
C1–C3	87.75	**0.96**	6.22	90.80	1.36	3.12	**82.74**	**1.92**	12.61
C1–C4	**88.08**	**0.96**	5.82	**90.93**	1.36	2.05	82.51	2.14	**3.90**

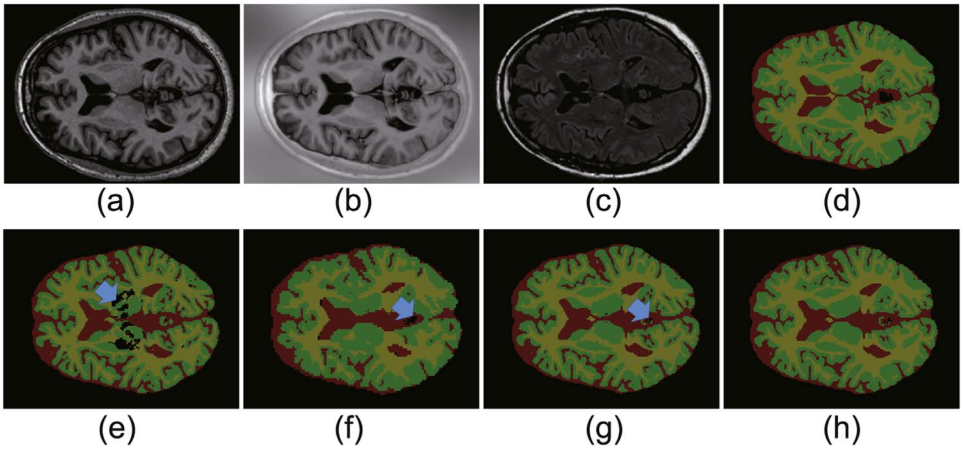


Fig. 6. Typical segmentation results of validation data using different levels of contextual information: (a)-(c) denote the original T1, T1-IR, and T2-FLAIR MR images, respectively; (d) ground truth label; (e)-(h) show the results using C1, C2, C1-C2, and C1-C4 contextual information, respectively (yellow, green, and red colors represent the WM, GM, and CSF, respectively). The blue arrows indicate some incorrect predictions of anatomical structures with fewer levels contextual information. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article).

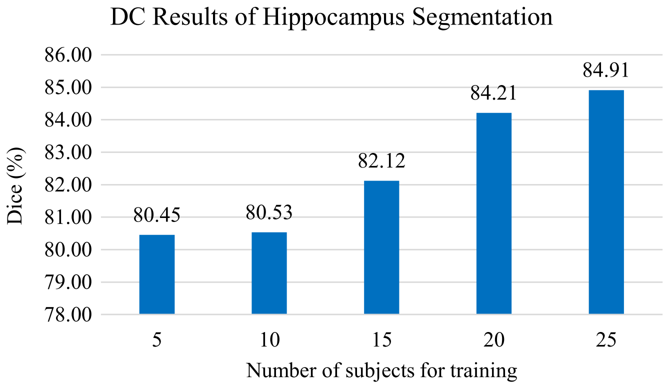


Fig. 7. Dice results of hippocampus segmentation regarding the number of subjects for training.

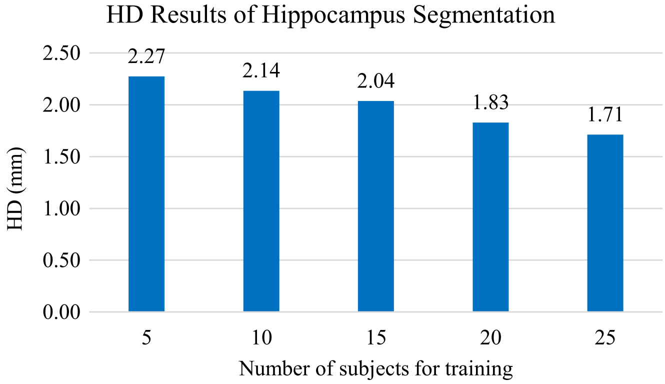


Fig. 8. HD results of hippocampus segmentation regarding the number of subjects for training.

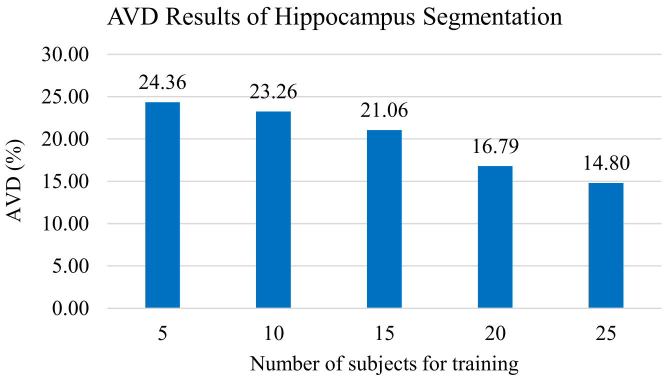


Fig. 9. AVD results of hippocampus segmentation regarding the number of subjects for training.

fication task. We cannot construct a very deep network with much more layers as the original residual network for our application mainly for the following two reasons. First, increasing the number of layers will decrease the resolution of volumetric feature maps, leading to the degradation of the segmentation accuracy due to spatial information loss, as demonstrated in the results using different levels of contextual information (Table 3). Second, the volumetric data processing with 3D feature maps typically consumes much larger GPU memories (might be hundreds of times) compared to the 2D images. Therefore, it is not practical to continually increase the network depth and construct a network with hundreds of layers for segmentation from volumetric data.

One of the main potential hurdles for applying deep learning to brain images is that training deep neural networks typically requires a large number of training data. However, in brain image analysis tasks, similar to many other medical image processing applications, it is usually difficult to obtain a large amount of data as a lot of time and effort have to be paid in accurately annotating the anatomical structures from medical images by experts. Although our method was trained on a relatively small dataset at the subject-level (five subjects), it still achieves quite competitive segmentation results with Dice coefficients on GM, WM, and CSF as 86.15%, 89.46%, and 84.25%, respectively. This is because the proposed *VoxResNet* can be trained in an end-to-end way, which takes the sub-volume as input and output the segmentation results with the same size as input directly. In such a way, the loss function is defined in a voxel-wise fashion, and hence the number of training samples can be extensively enlarged, with the number of training samples being considered in the voxel-level instead of subject-level. In addition, we randomly cropped the sub-volume training samples from the whole volume, which augmented the training samples as well. In the future work, we will investigate the performance of our method on more datasets.

Our method is inherently generic and can be retrained with a few manual segmentations on new real-world tasks. Considering the relatively small training data (5 subjects with each providing 3 image modalities) on MBrainS challenge (Mendrik et al., 2015), we tried our method on another larger dataset from OASIS project (Marcus et al., 2007) with 35 annotated MRIs on the task of hippocampus segmentation. In order to investigate the performance variation regarding the number of training MRIs, we split the data as 25 MRIs for training and 10 MRIs for testing. The number of training MRIs is increased from 5 to 25 with an interval of 5. The evaluation performance results including DC, HD, and AVD can be seen in Figs. 7–9, respectively. From these results, we can see the performance of hippocampus segmentation is gradually improved (80.45–84.91% on DC, 2.27–1.71 mm on HD, and 24.36–14.80% on AVD) with the increase of number of training MRIs. We also observe that the room of performance improvement is becoming smaller with the increasement of training data. With tens of (e.g., 20) manual annotations from clinical experts, the metric of DC can reach to around 85%, which is superior compared to other methods (Wang et al., 2014). This highlights the potential of generalization capability of our method on new clinical tasks.

It is well known that the infant brain MR images are significantly different with adult-like images with rapid tissue growth and developments of a wide range of cognitive as well as motor functions (Wang et al., 2015; Išgum et al., 2015; Shi et al., 2010). The reduced tissue contrast, increased noise, severe partial volume effect, and on-going white matter myelination pose great challenges for automatic infant brain MR images segmentation (Wang et al., 2015). In the future, we also plan to apply our network to infant brain segmentation.

5. Conclusions

In this paper, we developed a novel 3D residual network, named *VoxResNet*, and analyzed its capability in automatically segmenting brain structures from 3D MR images. Our method extended the 2D residual learning into a 3D variant for solving challenging segmentation tasks from volumetric data with a deeper network than our previous competitors. Both multi-modality and multi-level contextual information were elegantly integrated into our end-to-end network to improve the segmentation performance. Furthermore, an auto-context version of *VoxResNet* was proposed to further boost the performance under an integration of low-level appearance information, implicit shape information, and high-level context. Extensive experiments on a well-known challenging segmentation benchmark corroborated the efficacy of our method for brain structure segmentation, ranking first out of 37 teams including several state-of-the-art methods. The proposed method can advance the research on automated brain structure segmentation as well as offer a powerful and effective tool for more neuroimaging and neuroscience studies, where accurate segmentation of brain structures is essential.

Acknowledgments

The work described in this paper was supported by Hong Kong Research Grants Council, General Research Fund (Project number-14203115).

References

Aljabar, P., Heckemann, R.A., Hammers, A., Hajnal, J.V., Rueckert, D., 2009. Multi-atlas based segmentation of brain images: atlas selection and its effect on accuracy. NeuroImage 46, 726–738.

Artaechevarria, X., Munoz-Barrutia, A., Ortiz-de Solórzano, C., 2009. Combination strategies in multi-atlas image segmentation: application to brain mr data. IEEE Trans. Med. Imaging 28, 1266–1277.

Ashburner, J., Friston, K.J., 2005. Unified segmentation. NeuroImage 26, 839–851.

Bengio, Y., 2009. Learning deep architectures for ai. Found. Rrends® Mach. Learn. 2, 1–127.

Chen, H., Ni, D., Qin, J., Li, S., Yang, X., Wang, T., Heng, P.A., 2015a. Standard plane localization in fetal ultrasound via domain transferred deep neural networks. IEEE J. Biomed. Health Inform. 19, 1627–1636.

Chen, H., Qi, X., Yu, L., Dou, Q., Qin, J., Heng, P.A., 2017. Dcan: deep contour-aware networks for object instance segmentation from histology images. Med. Image Anal. 36, 135–146.

Chen, H., Dou, Q., Wang, X., Qin, J., Cheng, J.C., Heng, P.A., 2016a. 3d fully convolutional networks for intervertebral disc localization and segmentation. In: Proceedings of the International Conference on Medical Imaging and Virtual Reality. Springer. pp. 375–382.

Chen, H., Qi, X.J., Cheng, J.Z., Heng, P.A., 2016b. Deep contextual networks for neuronal structure segmentation. In: Proceedings of the Thirtieth AAAI Conference on Artificial Intelligence.

Chen, H., Yu, L., Dou, Q., Shi, L., Mok, V.C., Heng, P.A., 2015b. Automatic detection of cerebral microbleeds via deep learning based 3d feature representation. In: Proceedings of the 2015 IEEE 12th International Symposium on Biomedical Imaging (ISBI). IEEE. pp. 764–767.

Chen, L.C., Papandreou, G., Kokkinos, I., Murphy, K., Yuille, A.L., 2015c. Semantic image segmentation with deep convolutional nets and fully connected crfs. In: Proceedings of the ICLR. URL ⟨http://arxiv.org/abs/1412.7062⟩.

Çiçek, Ö., Abdulkadir, A., Lienkamp, S.S., Brox, T., Ronneberger, O., 2016. 3d u-net: Learning dense volumetric segmentation from sparse annotation. In: Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer. pp. 424–432.

Dai, J., He, K., Sun, J., 2016. Instance-aware semantic segmentation via multi-task network cascades. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 3150–3158.

de Brebisson, A., Montana, G., 2015. Deep neural networks for anatomical brain segmentation. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops. pp. 20–28.

Despotović, I., Goossens, B., Philips, W., 2015. MRI segmentation of the human brain: challenges, methods, and applications. Comput. Math. Methods Med., 2015.

Dou, Q., Chen, H., Yu, L., Zhao, L., Qin, J., Wang, D., Mok, V.C., Shi, L., Heng, P.A., 2016. Automatic detection of cerebral microbleeds from mr images via 3d convolutional neural networks. IEEE Trans. Med. Imaging 35, 1182–1195.

Dou, Q., Chen, H., Jin, Y., Yu, L., Qin, J., Heng, P.A., 2016a. 3d deeply supervised network for automatic liver segmentation from ct volumes. In: Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer. pp. 149–157.

Dou, Q., Chen, H., Yu, L., Qin, J., Heng, P.A. Multi-level contextual 3d CNNs for false positive reduction in pulmonary nodule detection. IEEE Transactions on Biomedical Engineering, http://dx.doi.org/10.1109/TBME.2016.2613502.

Giorgio, A., De Stefano, N., 2013. Clinical use of brain volumetry. J. Magn. Reson. Imaging 37, 1–14.

Habas, P.A., Kim, K., Rousseau, F., Glenn, O.A., Barkovich, A.J., Studholme, C., 2010. Atlas-based segmentation of developing tissues in the human brain with quantitative validation in young fetuses. Human. Brain Mapp. 31, 1348–1358.

He, K., Zhang, X., Ren, S., Sun, J., 2016a. Deep residual learning for image recognition. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 770–778.

He, K., Zhang, X., Ren, S., Sun, J., 2016b. Identity mappings in deep residual networks. In: European Conference on Computer Vision. Springer. pp. 630–645.

Heckemann, R.A., Hajnal, J.V., Aljabar, P., Rueckert, D., Hammers, A., 2006. Automatic anatomical brain MRI segmentation combining label propagation and decision fusion. NeuroImage 33, 115–126.

Huttenlocher, D.P., Klanderman, G.A., Rucklidge, W.J., 1993. Comparing images using the Hausdorff distance. IEEE Trans. Pattern Anal. Mach. Intell. 15, 850–863.

Ioffe, S., Szegedy, C., 2015. Batch Normalization: Accelerating Deep Network Training By Reducing Internal Covariate Shift. arXiv preprint arXiv:1502.03167.

Išgum, I., Benders, M.J., Avants, B., Cardoso, M.J., Counsell, S.J., Gomez, E.F., Gui, L., Hűppi, P.S., Kersbergen, K.J., Makropoulos, A., et al., 2015. Evaluation of automatic neonatal brain segmentation algorithms: the neobrains 12 challenge. Med. Image Anal. 20, 135–151.

Ji, S., Xu, W., Yang, M., Yu, K., 2013. 3d convolutional neural networks for human action recognition. IEEE Trans. Pattern Anal. Mach. Intell. 35, 221–231.

Jia, Y., Shelhamer, E., Donahue, J., Karayev, S., Long, J., Girshick, R., Guadarrama, S., Darrell, T., 2014. Caffe: convolutional architecture for fast feature embedding. In: Proceedings of the 22nd ACM International Conference on Multimedia, ACM. pp. 675–678.

Kamnitsas, K., Ledig, C., Newcombe, V.F., Simpson, J.P., Kane, A.D., Menon, D.K., Rueckert, D., Glocker, B., 2017. Efficient multi-scale 3d CNN with fully connected CRF for accurate brain lesion segmentation. Med. Image Anal. 36, 61–78.

Klein, A., Hirsch, J., 2005. Mindboggle: a scatterbrained approach to automate brain labeling. NeuroImage 24, 261–280.

Klein, S., Staring, M., Murphy, K., Viergever, M.A., Pluim, J.P., 2010. Elastix: a toolbox for intensity-based medical image registration. IEEE Trans. Med. Imaging 29, 196–205.

Krizhevsky, A., Sutskever, I., Hinton, G.E., 2012. Imagenet classification with deep convolutional neural networks. In: Proceedings of the Advances in Neural Information Processing Systems. pp. 1097–1105.

LeCun, Y., Boser, B., Denker, J.S., Henderson, D., Howard, R.E., Hubbard, W., Jackel, L.D., 1989. Backpropagation applied to handwritten zip code recognition. Neural Comput. 1, 541–551.

LeCun, Y., Bengio, Y., Hinton, G., 2015. Deep learning. Nature 521, 436–444.

Lee, C.Y., Xie, S., Gallagher, P., Zhang, Z., Tu, Z., 2015. Deeply-supervised nets. In: Proceedings of the AISTATS. p. 6.

Lee, N., Laine, A.F., Klein, A., 2011. Towards a deep learning approach to brain parcellation. In: Proceedings of the 2011 IEEE International Symposium on Biomedical Imaging: From Nano to Macro, IEEE. pp. 321–324.

Lequan, Y., Chen, H., Dou, Q., Qin, J., Heng, P.A., 2016. Automated melanoma recognition in dermoscopy images via very deep residual networks. IEEE Trans. Med. Imaging. http://dx.doi.org/10.1109/TMI.2016.2642839.

Li, R., Zhang, W., Suk, H.I., Wang, L., Li, J., Shen, D., Ji, S., 2014. Deep learning based imaging data completion for improved brain disease diagnosis. In: Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention, Springer. pp. 305–312.

Long, J., Shelhamer, E., Darrell, T., 2015. Fully convolutional networks for semantic segmentation. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 3431–3440.

Lötjönen, J.M., Wolz, R., Koikkalainen, J.R., Thurfjell, L., Waldemar, G., Soininen, H., Rueckert, D., Initiative, A.D.N., et al., 2010. Fast and robust multi-atlas segmentation of brain magnetic resonance images. NeuroImage 49, 2352–2365.

Mahbod, A., 2016. Structural Brain MRI Segmentation Using Machine Learning Technique. (Master's thesis). KTH, School of Technology and Health (STH).

Maier, O., Menze, B.H., von der Gablentz, J., Häni, L., Heinrich, M.P., Liebrand, M., Winzeck, S., Basit, A., Bentley, P., Chen, L., et al., 2017. ISLES 2015-a public evaluation benchmark for ischemic stroke lesion segmentation from multispectral MRI. Med. Image Anal. 35, 250–269.

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.

Mendrik, A.M., Vincken, K.L., Kuijf, H.J., Breeuwer, M., Bouvy, W.H., De Bresser, J., Alansary, A., De Bruijne, M., Carass, A., El-Baz, A., et al., 2015. Mrbrains challenge: Online evaluation framework for brain image segmentation in 3T MRI scans. Computational intelligence and neuroscience.

Menze, B.H., Jakab, A., Bauer, S., Kalpathy-Cramer, J., Farahani, K., Kirby, J., Burren, Y., Porz, N., Slotboom, J., Wiest, R., et al., 2015. The multimodal brain tumor image segmentation benchmark (brats). IEEE Trans. Med. Imaging 34, 1993–2024.

Merkow, J., Kriegman, D., Marsden, A., Tu, Z., 2015. Dense volume-to-volume vascular boundary detection. In: Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention, Springer. pp. 565–572.

Milletari, F., Navab, N., Ahmadi, S.A., 2016. V-net: fully convolutional neural networks for volumetric medical image segmentation. In: 2016 Proceedings of the Fourth International Conference on 3D Vision (3DV), IEEE. pp. 565–571.

Moeskops, P., Benders, M.J., Chiţ, S.M., Kersbergen, K.J., Groenendaal, F., de Vries, L.S., Viergever, M.A., Išgum, I., 2015. Automatic segmentation of mr brain images of preterm infants using supervised classification. NeuroImage 118, 628–641.

Moeskops, P., Viergever, M.A., Mendrik, A.M., de Vries, L.S., Benders, M.J., Išgum, I., 2016. Automatic segmentation of mr brain images with a convolutional neural network. IEEE Trans. Med. Imaging 35, 1252–1261.

Moeskops, P., Viergever, M.A., Benders, M.J., Išgum, I., 2015b. Evaluation of an automatic brain segmentation method developed for neonates on adult mr brain images. In: Proceedings of the SPIE Medical Imaging, International Society for Optics and Photonics. pp. 941315–941315.

Nie, D., Wang, L., Gao, Y., Sken, D., 2016. Fully convolutional networks for multi-modality isointense infant brain image segmentation. In: 2016 IEEE Proceedings of the 13th International Symposium on Biomedical Imaging (ISBI), IEEE. pp. 1342–1345.

Nogues, I., Lu, L., Wang, X., Roth, H., Bertasius, G., Lay, N., Shi, J., Tsehay, Y., Summers, R.M., 2016. Automatic lymph node cluster segmentation using holistically-nested neural networks and structured optimization in ct images. In: Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention, Springer. pp. 388–397.

Penny, W.D., Friston, K.J., Ashburner, J.T., Kiebel, S.J., Nichols, T.E., 2011. Statistical Parametric Mapping: The Analysis of Functional Brain Images. Academic Press.

Pereira, S., Pinto, A., Oliveira, J., Mendrik, A.M., Correia, J.H., Silva, C.A., 2016. Automatic brain tissue segmentation in mr images using random forests and conditional random fields. J. Neurosci. Methods 270, 111–123.

Petrella, J.R., Coleman, R.E., Doraiswamy, P.M., 2003. Neuroimaging and early diagnosis of alzheimer disease: a look to the future 1. Radiology 226, 315–336.

Prakash, R.M., Kumari, R.S.S. Modified Expectation Maximization Method for Automatic Segmentation of MR Brain Images.

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. In: Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention, Springer. pp. 246–253.

Rajchl, M., Baxter, J., McLeod, A.J., Yuan, J., Qiu, W., Peters, T.M., White, J., Khan, A., 2013. Asets: Map-based Brain Tissue Segmentation Using Manifold Learning and Hierarchical Max-Flow Regularization.

Ronneberger, O., Fischer, P., Brox, T., 2015. U-net: convolutional networks for biomedical image segmentation. In: Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention, Springer. pp. 234–241.

Roth, H.R., Lu, L., Seff, A., Cherry, K.M., Hoffman, J., Wang, S., Liu, J., Turkbey, E., Summers, R.M., 2014. A new 2.5 d representation for lymph node detection using random sets of deep convolutional neural network observations. In: Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention, Springer. pp. 520–527.

Sarikaya, D., Zhao, L., Corso, J.J., 2013. Multi-atlas brain MRI segmentation with multiway cut. In: Proceedings of the MICCAI WorkshopsThe MICCAI Grand Challenge on MR Brain Image Segmentation (MRBrainS13).

Schmidhuber, J., 2015. Deep learning in neural networks: an overview. Neural Netw. 61, 85–117.

Shen, F., Gan, R., Zeng, G., 2016. Weighted residuals for very deep networks. In: 2016 Proceedings of the 3rd International Conference on Systems and Informatics (ICSAI), IEEE. pp. 936–941.

Shi, F., Fan, Y., Tang, S., Gilmore, J.H., Lin, W., Shen, D., 2010. Neonatal brain image segmentation in longitudinal MRI studies. NeuroImage 49, 391–400.

Shin, H.C., Roth, H.R., Gao, M., Lu, L., Xu, Z., Nogues, I., Yao, J., Mollura, D., Summers, R.M., 2016. Deep convolutional neural networks for computer-aided detection: CNN architectures, dataset characteristics and transfer learning. IEEE Trans. Med. Imaging 35, 1285–1298.

Simonyan, K., Zisserman, A., 2014. Very Deep Convolutional Networks for Large-scale Image Recognition. arXiv preprint arXiv:1409.1556.

Srivastava, R.K., Greff, K., Schmidhuber, J., 2015. Highway Networks. arXiv preprint arXiv:1505.00387.

Stollenga, M.F., Byeon, W., Liwicki, M., Schmidhuber, J., 2015. Parallel multi-dimensional lstm, with application to fast biomedical volumetric image segmentation. Adv. Neural Inf. Process. Syst., 2998–3006.

Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhoucke, V., Rabinovich, A., 2015. Going deeper with convolutions. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 1–9.

Thambisetty, M., Wan, J., Carass, A., An, Y., Prince, J.L., Resnick, S.M., 2010. Longitudinal changes in cortical thickness associated with normal aging. NeuroImage 52, 1215–1223.

Tran, D., Bourdev, L., Fergus, R., Torresani, L., Paluri, M., 2016. Deep end2end voxel2voxel prediction. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops. pp. 17–24.

Tu, Z., Bai, X., 2010. Auto-context and its application to high-level vision tasks and 3d brain image segmentation. IEEE Trans. Pattern Anal. Mach. Intell. 32, 1744–1757.

Tu, Z., 2008. Auto-context and its application to high-level vision tasks. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2008, IEEE. pp. 1–8.

van Opbroek, A., van der Lijn, F., de Bruijne, M., 2013. Automated Brain-Tissue Segmentation by Multi-feature SVM Classification.

Wang, H., Suh, J.W., Das, S.R., Pluta, J.B., Craige, C., Yushkevich, P.A., 2013. Multi-atlas segmentation with joint label fusion. IEEE Trans. Pattern Anal. Mach. Intell. 35, 611–623.

Wang, J., Vachet, C., Rumple, A., Gouttard, S., Ouziel, C., Perrot, E., Du, G., Huang, X., Gerig, G., Styner, M.A., 2014. Multi-atlas segmentation of subcortical brain structures via the autoseg software pipeline. Front. Neuroinform. 8, 7.

Wang, L., Gao, Y., Shi, F., Li, G., Gilmore, J.H., Lin, W., Shen, D., 2015. Links: learning-based multi-source integration framework for segmentation of infant brain images. NeuroImage 108, 160–172.

Wright, R., Kyriakopoulou, V., Ledig, C., Rutherford, M.A., Hajnal, J.V., Rueckert, D., Aljabar, P., 2014. Automatic quantification of normal cortical folding patterns from fetal brain MRI. NeuroImage 91, 21–32.

Yu, L., Chen, H., Dou, Q., Qin, J., Heng, P.A., 2017. Integrating online and offline three-dimensional deep learning for automated polyp detection in colonoscopy videos. IEEE J. Biomed. Health Inform. 21, 65–75.

Zagoruyko, S., Komodakis, N., 2016. Wide Residual Networks. arXiv preprint arXiv:1605.07146.

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.

Zheng, Y., Liu, D., Georgescu, B., Nguyen, H., Comaniciu, D., 2015. 3d deep learning for efficient and robust landmark detection in volumetric data. In: Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention, Springer. pp. 565–572.