M-NET: A CONVOLUTIONAL NEURAL NETWORK FOR DEEP BRAIN STRUCTURE SEGMENTATION

Raghav Mehta, *Jayanthi Sivaswamy*

Center for Visual Information Technology (CVIT), IIIT-Hyderabad, India

ABSTRACT

In this paper, we propose an end-to-end trainable Convolutional Neural Network (CNN) architecture called the *M-net*, for segmenting deep (human) brain structures from Magnetic Resonance Images (MRI). A novel scheme is used to learn to combine and represent 3D context information of a given slice in a 2D slice. Consequently, the *M-net* utilizes only 2D convolution though it operates on 3D data, which makes *M-net* memory efficient. The segmentation method is evaluated on two publicly available datasets and is compared against publicly available model based segmentation algorithms as well as other classification based algorithms such as Random Forrest and 2D CNN based approaches. Experiment results show that the *M-net* outperforms all these methods in terms of dice coefficient and is at least 3 times faster than other methods in segmenting a new volume which is attractive for clinical use.

Index Terms— Magnetic Resonance Images, Convolutional Neural Networks, Segmentation, Deep Brain Structures

1. INTRODUCTION

Diagnosis of neuro-degenerative diseases, analysis of development of neonatal brain etc., rely heavily on quantitative and qualitative measurements of different human brain structures, especially deep brain structures [1]. For instance, morphometry and volumetry of hippocampus plays an important role in Alzheimer's disease assessment [2]. Thus, segmentation of these deep brain structures, *in less time*, is critical.

This problem is generally solved either using Non-rigid Registration [3] or Model based techniques [4], with both relying on training atlases (with manual segmentation) to segment new volumes, albeit in different ways. The former class of techniques label a new volume by registering (non-rigid) training atlases to it and then fusing the propagated labels whereas, the latter techniques predict the labels from a mathematical model learnt from the training atlases. Model based techniques typically require 15-20 *minutes* to label new volumes as against 20-25 *hours* required by registration based methods [5].

This work was partly funded by the Department of Science and Technology, Govt. of India, under Grant SR/CSRI/194/2013(G).

In this paper, we focus on model based segmentation techniques given their computational efficiency and hence potential for clinical use. FSL-FIRST is a popular technique based on statistical shape modeling [4]. Here, the relationship between shape and intensity are modeled via the conditional distribution of intensity given shape aiding in producing smooth shapes.

Voxel-level classification models have also been used for deep brain structure segmentation. A set of random forests are learnt in [6] for each training atlases separately using intensity and contextual features, majority voting of the predicted output of these set of random forests is used to label voxels in a new volume. Deep learning techniques, especially Convolution Neural Networks (CNN), have also been explored for many medical image analysis tasks including deep brain segmentation, given the network's ability to learn features for the given task from training data. A Multi-Scale CNN (MS-CNN) architecture is proposed in [7] using intensity as well as contextual information as inputs. Final voxel labeling is achieved by postprocessing the CNN output using Random Walker (RW) based graphical model. A Fully Convolutional Network (FCN) based on 2D-CNN architecture is proposed in [8] to label voxels with a single slice of MRI volume as input, here label consistency across slices is achieved using Markov Random Field (MRF). Both [7] and [8] perform well in terms of labeling accuracy but need separate graphical models-based post-processing for final segmentation which precludes end-to-end CNN training in addition to increasing the run time.

In this paper, we present a novel CNN architecture, which eliminates the need of any post-processing step making it end-to end trainable. It is designed to leverage 3D information around a slice at the input level and yet operate only on 2D information beyond the first stage to produce a labeled slice as output. This ensures labels are consistent and accurate across slices without using any post-processing steps, which in-term reduces run-time and memory requirement.

The paper is organized in the following way, in Sec:2 we introduces as well as give detail description about the proposed architecture. In Sec:3, we analyze the performance of our proposed architecture on two publicly available datasets and compare its performance with other methods. In Sec:4, we conclude the paper.

978-1-5090-1172-8/17/$31.00 ©2017 IEEE

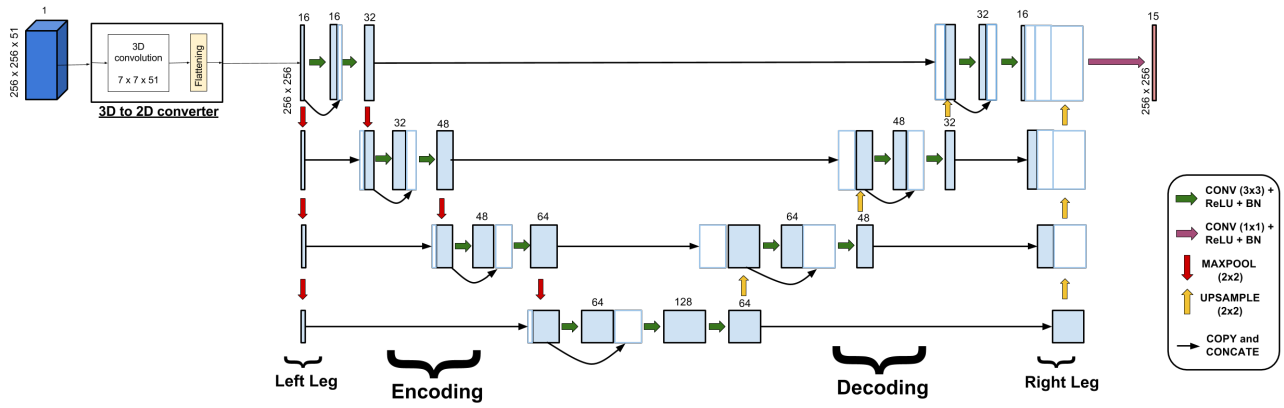

Fig. 1: Schematic representation of the *M-net* CNN architecture. Solid blue boxes represent multi-channel feature maps. Blue framed boxes represent copied feature maps. Number of channels is denoted on the top of the box.

2. METHODOLOGY

CNN is a deep learning architecture inspired by the biological networks akin to the multilayer perceptron. It has been widely used for the segmentation and recognition tasks. Basic blocks of a CNN are Convolutional Layer, Maxpooling layer, Dropout layer [9] and Activation Functions. For a detailed description of these blocks readers are refered to [10]. The proposed architecture is shown in Fig:1

Our architecture is inspired by the U-net [11][12]. In [11], a 2D U-net is used for segmentation of neuronal structures in electron microscopic stacks while in [12], the same architecture with 3D filters is used for segmentation of the Xenopus kidney from 3D volumes of confocal microscopy. Although, both give good performance for their respective tasks, they are not appropriate for segmentation of MRI volume of size 256x256x256, as 2D U-net does not utilize any 3D information, while 3D U-net does it but at the cost of a high memory ($\sim$10 GB) requirement for a small input of 200x200x50. The latter is a bottleneck when working with MRI volume of size 256x256x256 as maximum memory available in most advanced GPU is only 12 GB.

We address this issue of memory constraint, while still utilizing necessary 3D information for MRI segmentation, in a novel way in our proposed CNN architecture, henceforth known as *M-net*. As shown in Fig:1, a slice s and its neighbors are used to form a stack s-n:s+n which serves as an input. The value of n is determined empirically. This allows us to utilize 3D information. The stack of slices is passed through a *3D-to-2D* converter block, which learns a 3D convolution filter of size 7x7x(2n+1), to combine the stack of 2D slices into one single 2D slice $\bar{s}$. This $\bar{s}$ is then processed through the *M-net* architecture to obtain the desired segmentation. Thus, segmentation of a whole volume is done slice by slice.

M-net has mainly 4 pathways of 2D filters: two main encoding and decoding paths, and two side paths which gives our architecture functionality of deep-supervision [13]. Each pathway has 4 steps. In the encoding path, each step has a cascade of 2D convolution filters of size 3x3 and maxpooling by 2x2, which reduces the size of input by half and allows network to learn contextual information. In the cascade of convolution filters, skip connection is introduced to enable the network to learn better features [14]. The decoding layer is identical to encoding layers with one exception: maxpooling is replaced by upsampling layer to double the size of input and recover an output image of original size. Similarly, skip connections are also implemented between corresponding encoding and decoding layers to ensures that the network has sufficient information to derive fine grain labeling of an image without the need for any post-processing [11]. The left leg operates on $\bar{s}$ with 4 maxpooling layers of size 2x2 and the outputs are given as input to the corresponding encoding layers. The right leg upsamples the output of each of the decoding layers to the original size of $\bar{s}$. Finally, the output of the decoding layer and the right leg is processed by a 1x1 convolution layer with L channels, where L is the number of structures of interest including background. Dropout (with probability 0.3) [9] and batch normalization (BN) [15] are applied after each step and each convolution layers respectively, to reduce overfitting. For all the layers except the last, a ReLU activation is applied after every convolution layer. For the last layer, a softmax activation is applied, which gives the probability of each voxel belonging to different structures. The final label for any voxel is the structure with maximum probability.

A weighted Categorical Cross Entropy function was used to tackle the class imbalance problem. This loss function and weights are defined such that the weight increases whenever there are fewer voxels in a particular class.

The advantage of the *M-net* is that barring one 3D convolution filter, all other filters are 2D filters which allows end-to-end training of the network with considerably low memory requirement ($\sim$5GB).

Table 1: Quantitative comparison of performance on the IBSR dataset. Reported Dice coefficient values for a structure are averaged over the values for the 2 hemispheres.

	Freesurfer	FSL-FIRST	RF + MRF	FCN + MRF	MS-CNN + RW	*M-net*
Amygdala	0.69	0.70	0.62	0.64	0.67	**0.73**
Caudate	0.82	0.83	0.78	0.78	**0.87**	**0.87**
Hippocampus	0.77	0.81	0.59	0.71	**0.82**	**0.82**
Pallidum	0.71	0.76	0.62	0.75	0.80	**0.82**
Putamen	0.81	0.84	0.77	0.83	0.88	**0.90**
Thalamus	0.86	0.88	0.80	0.87	**0.90**	**0.90**
Accumbens Area	0.69	0.73	0.60	0.63	0.69	**0.75**
Overall	0.76	0.79	0.69	0.75	0.80	**0.83**

3. EXPERIMENTS AND RESULTS

The proposed architecture was used for the task of segmenting deep brain structures like Thalamus, Putamen, Pallidum, Hippocampus, Amygdala, Caudate and Accumbens area. Data of varying size is drawn from two publicly available datasets.

First dataset considered is the International Brain Segmentation Repository (IBSR) dataset[1] which has 18, 3D T1 MR images of 1.5 mm thick cortical slices. The size of the MRI volume is 256x256x128. Manual Segmentation of 32 structures is available, however we restrict our attention to only the deep brain structures.

Second dataset considered is a Diencephalon dataset released as a part of the MICCAI 2013 SATA Challenge[2]. [16]. The data consists of 35 training and 12 testing T1 MR images of 1 mm thick cortical slices from the OASIS project with corresponding 14 sub-cortical label maps as provided by Neuromorphometrics Inc. Input volume size is 256x256x300. This dataset is more challenging than the IBSR dataset, as the test volumes are of patients with different abnormalities. Manual segmentation is only available for the training set and evaluation on testing set requires submitting the segmentation results to the challenge organizers.

3.1. Implementation Details

The proposed CNN was trained on a NVIDIA K40 GPU, with 12GB of RAM for 30 epochs. Approximate training time was 3 days. The CNN was trained using Adam Optimizer [17] with following hyper parameters: learning rate =0.001, beta1=0.9, beta2=0.999 and epsilon=10^{-08}. Learning rate was reduced by a factor of 10 after 20 epochs. Code was written in Keras Library using Python. A hyperparameter of *M-net* is *n*, which denotes the number of neighbor slices given as additional input to CNN. This was empirically finalized to be 25. Thus, segmentation of any given slice *s* is done by taking that as a central slice and (25+25=) 50 of its neighboring slices as input (total 51 slices).

[1] http://www.nitrc.org/projects/ibsr
[2] http://tinyurl.com/SATAchallenge

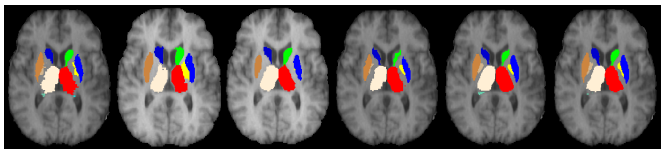

Fig. 2: Qualitative comparison of segmentation results for a sample slice from IBSR dataset. Left to Right: Ground truth, RF+MRF, FCN+MRF, Freesurfer, FSL-FIRST and *M-net*

The IBSR dataset was randomly divided into two equal sets with 9 volumes each. CNN was trained 2 times on the two sets separately, considering the other set as a testing set. For SATA challenge dataset, CNN was trained on all the 35 training volumes. Performance on testing set was evaluated by uploading the segmented volumes on the challenge server.

3.2. Results and Comparison with other methods

The segmentation performance is quantitatively evaluated using the mean Dice Coefficient (DC) across a dataset. Let A and B denote the binary segmentation labels generated manually and computationally, respectively. DC is defined as

$$DC(A,B) = \frac{2|AB|}{|A| + |B|}$$

where $|A|$ denotes the number of positive elements in the binary segmentation A, and $|AB|$ is the number of shared positive elements by A and B. DC$\in[0,1]$.

On IBSR dataset, we compare the output of *M-net* with 5 different model based methods: FSL-FIRST [4], MS-CNN+RW [7], FCN+MRF [8], RF+MRF [6] and Freesurfer [18]. The approximate CPU run time, for segmenting a new volume, for all of these methods are: FSL-FIRST ($\sim$15 min), Freesurfer ($\sim$90 min), MS-CNN+RW ($\sim$20 min), FCN+MRF ($\sim$15 min) and *M-net* ($\sim$5 min).

The mean DC values for each structure as well as for the whole volume is listed in Table:1. Segmentation result for a sample slice is shown in Fig:2. Based on these results, we can observe that *M-net* is able to outperform all the other segmentation methods on all the deep brain structures. It should be

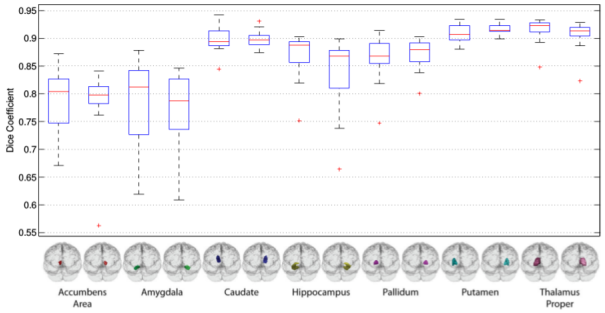

Fig. 3: DC values for different structures obtained with *M-net* on the Diencephalon dataset

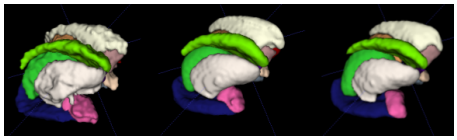

Fig. 4: Qualitative comparison of segmentation results, using 3D rendering, for the Diencephalon dataset. From left to right: Freesurfer,FSL-FIRST and *M-net*

noted that there is major boost in DC (0.06) for small structures like amygdala and Accumbens area. Results in Fig:2 indicate tendency to undersegment by RF+MRF, FCN+MRF and Freesurfer methods, which is not the case with *M-net* despite not using any post-processing.

Next, we experimented with Diencephalon dataset. As per the evaluation results displayed in the website hosted by SATA challenge organizers (Diencephalon Challenge Midbrain - Free Competition) , the mean dice coefficient for *M-net* was *0.85780*. This is considerably better than the values for other model based methods: FSL-FIRST (0.82437), Atlas Forrest based method (0.82819), Freesurfer (0.75761).

A bar plot of DC values for all the deep brain structures across the dataset is shown in Fig:3. We can see that, as is the case with all the other methods, the performance of *M-net* is marginally lower for smaller structures compared to bigger structures. Qualitative comparison of segmentation using 3D rendering for *M-net*, FSL-FIRST and Freesurfer is shown in Fig:4. The results for FSL-FIRST and *M-net* are smooth for all the structures, compared to Freesurfer.

4. CONCLUSION

In this paper, we proposed a novel CNN architecture which utilizes 3D information with the help of comparatively less expensive 2D convolution filter, this is achieved by using a single 3D convolution filter to combine a slice and its neighboring slices into one slice as a first step. We also introduced skip connections between convolution filters and deep supervision functionality in our network which allows it to learn better features. Experimental results on two publicly available datasets, with different volume dimensions, shows that proposed network outperforms the current state of the art model based segmentation techniques and at considerably ($\sim\frac{1}{3^{rd}}$) less processing time. These aspects paves way for its use in clinical scenario. The *M-net* segments a volume slice by slice and hence it can potentially be used to segment any 3D dataset, which is to be explored in the future.

5. REFERENCES

[1] Y Chudasama and TW Robbins, "Functions of frontostriatal systems in cognition: comparative neuropsychopharmacological studies in rats, monkeys and humans," *Biological psychology*, vol. 73, no. 1, pp. 19–38, 2006.

[2] Liana Apostolova et al., "Subregional hippocampal atrophy predicts alzheimer's dementia in the cognitively normal," *Neurobiology of aging*, vol. 31, no. 7, pp. 1077–1088, 2010.

[3] Rolf Heckemann et al., "Automatic anatomical brain mri segmentation combining label propagation and decision fusion," *NeuroImage*, vol. 33, no. 1, pp. 115–126, 2006.

[4] Brian Patenaude et al., "A bayesian model of shape and appearance for subcortical brain segmentation," *Neuroimage*, vol. 56, no. 3, pp. 907–922, 2011.

[5] Juan Eugenio Iglesias and Mert R Sabuncu, "Multi-atlas segmentation of biomedical images: a survey," *Medical image analysis*, vol. 24, no. 1, pp. 205–219, 2015.

[6] Stavros Alchatzidis et al., "Discrete multi atlas segmentation using agreement constraints," in *British Machine Vision Conference*, 2014.

[7] Siqi Bao and Albert CS Chung, "Multi-scale structured cnn with label consistency for brain mr image segmentation," *Computer Methods in Biomechanics and Biomedical Engineering: Imaging & Visualization*, pp. 1–5, 2016.

[8] Mahsa Shakeri et al., "Sub-cortical brain structure segmentation using f-cnn's," *arXiv preprint arXiv:1602.02130*, 2016.

[9] Nitish Srivastava et al., "Dropout: a simple way to prevent neural networks from overfitting.," *Journal of Machine Learning Research*, vol. 15, no. 1, pp. 1929–1958, 2014.

[10] Alex Krizhevsky et al., "Imagenet classification with deep convolutional neural networks," in *Advances in neural information processing systems*, 2012, pp. 1097–1105.

[11] Olaf Ronneberger et al., "U-net: Convolutional networks for biomedical image segmentation," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2015, pp. 234–241.

[12] Özgün Çiçek et al., "3d u-net: Learning dense volumetric segmentation from sparse annotation," *arXiv preprint arXiv:1606.06650*, 2016.

[13] Chen-Yu Lee et al., "Deeply-supervised nets.," in *AISTATS*, 2015, vol. 2, p. 6.

[14] Rupesh Kumar Srivastava et al., "Highway networks," *arXiv preprint arXiv:1505.00387*, 2015.

[15] Sergey Ioffe and Christian Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," *arXiv preprint arXiv:1502.03167*, 2015.

[16] Andrew Asman et al., "Miccai 2013 segmentation algorithms, theory and applications (sata) challenge results summary," in *MICCAI Challenge Workshop on Segmentation: Algorithms, Theory and Applications (SATA)*, 2013.

[17] Diederik Kingma and Jimmy Ba, "Adam: A method for stochastic optimization," *arXiv preprint arXiv:1412.6980*, 2014.

[18] Bruce Fischl et al., "Whole brain segmentation: automated labeling of neuroanatomical structures in the human brain," *Neuron*, vol. 33, no. 3, pp. 341–355, 2002.