LSTM-MA: A LSTM METHOD WITH MULTI-MODALITY AND ADJACENCY CONSTRAINT FOR BRAIN IMAGE SEGMENTATION

*Kai Xie and Ying Wen**

Shanghai Key Laboratory of Multidimensional Information Processing
Department of Computer Science and Technology
East China Normal University, Shanghai 200062, China

ABSTRACT

MR brain tissue segmentation is a significant problem in biomedical image processing. Inhomogeneous intensity and image noise influence the segmentation accuracy. In this paper, we propose a LSTM method with multi-modality and adjacency constraint for brain image segmentation, named LSTM-MA. Two feature sequence generation ways in our method are used, i.e., features with pixel-wise and superpixel-wise adjacency constraint. The LSTM model classifies the generated features into semantic labels to form the segmentation result. The evaluation experiments on BrainWeb and MRBrainS demonstrate that the proposed LSTM-MA with pixel-wise adjacency constraint achieves promising segmentation results, while LSTM-MA with superpixel-wise adjacency constraint shows its computational efficiency as well as robustness to noise.

***Index Terms*—** LSTM, Multi-modality, Superpixel, Brain Segmentation, Noise Robustness

1. INTRODUCTION

The segmentation of MR brain images is important to the analysis of neuroscience and the diagnosis of diseases. Many methods have been proposed to segment MR brain images into three tissues, namely white matter (WM), gray matter (GM), and cerebrospinal fluid (CSF). Clustering methods like k-means [1], fuzzy c-means (FCM) [2] have shown good pixel accuracy for certain entire and high-contrast brain slices. However, FCM is quite time-consuming while k-means is sensitive to initialization. Inhomogeneous intensity and image noise remains to be an obstacle to the robustness and generalization for different modalities of MR brain images. Additionally, previous supervised methods [3, 4] do not take multi-modality information into consideration, which is beneficial to deal with the problem of insufficient tissue contrast [5, 6].

The problem of inhomogeneous intensity could be substantially improved via a supervised learning pathway. Deep learning methods have shown its superiority in feature extraction, among which fully convolutional neural network (FCN) [7] is first proposed for the task of semantic segmentation. However, FCN-like models [8, 9] provide relatively coarse results in boundary adherence and are not optimized for the task of brain segmentation. Recurrent neural network (RNN), especially long short-term memory (LSTM) [10], is intrinsically suitable for processing sequence data due to its special design of memory gate and forget gate. Liang *et al.* [11, 12] used LSTM based model for the task of semantic object parsing and achieved promising result. Liu *et al.* [13] further combined superpixel with deep learning for image segmentation. Nevertheless, deep learning methods, especially LSTM and its variants are still not widely used in processing medical images.

In this paper, we propose a LSTM method with multi-modality and adjacency constraint (LSTM-MA) to improve the segmentation accuracy for brain images. We first use two ways of pixel-wise and superpixel-wise constraint to generate feature sequences, then LSTM based model is utilized for feature sequence classification. The prediction labels for each pixel or superpixel are combined together to form the final segmentation result. The contribution of our method is that LSTM-MA is used for spatial sequence data classification based on multi-modality information and adjacency constraint, and the experimental results show the efficiency of the proposed method.

2. METHOD

Our MR brain segmentation method incorporates multi-modality and local adjacency information to generate feature sequences, and utilizes LSTM for sequence classification to generate the segmentation result. Fig.1 illustrates our designed pipeline. For a given brain slice with multiple modalities (e.g., T1, T2, PD), we first consider two

This work was supported by the National Nature Science Foundation of China under Grant 61773166, the Natural Science Foundation of Shanghai under Grant 17ZR1408200 and the Science and Technology Commission of Shanghai Municipality under Grant 14DZ2260800.

*Corresponding author: ywen@cs.ecnu.edu.cn

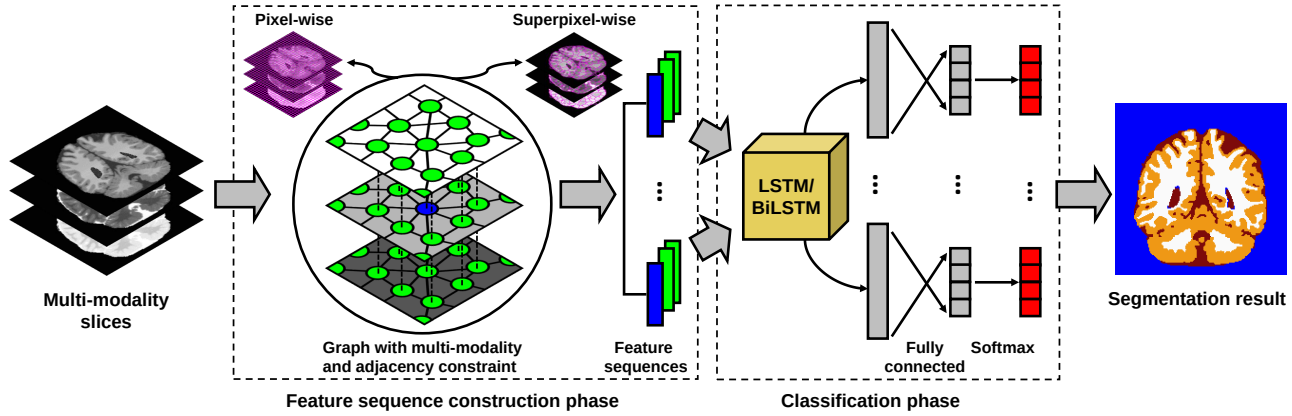


Fig. 1. Illustration of our proposed segmentation pipeline. Given the input of multi-modality slices, two phases are followed to obtain the final segmentation result. First is the sequence construction phase, feature sequences are generated in two ways, namely pixel-wise constraint and superpixel-wise constraint. Second is the classification phase, feature sequences are fed into a LSTM or BiLSTM layer separately followed by fully connected and softmax layer.

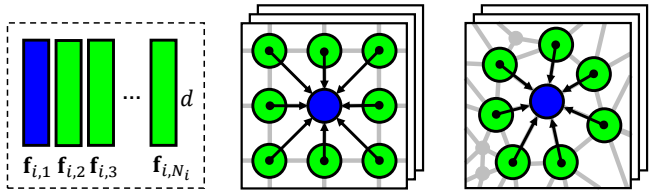
approaches to generate a set of feature sequences with different adjacency constraints, namely pixel-wise constraint and superpixel-wise constraint. Each feature sequence contains several spatial-related feature vectors corresponding to a pixel or a superpixel feature of the input slice. Then, each feature sequence is fed into a LSTM or BiLSTM layer, followed by a fully connected and softmax layer to learn its belonging class. Finally, the prediction labels of feature sequences are combined to form the segmentation image.

2.1. Image as Undirected Graph

In feature construction phase, we consider an input slice as an undirected graph $\mathcal{G}$ with adjacency constraint. It is built by connecting a set of graph nodes $\{v_i\}_{i=1}^{M}$ via the graph edges $\{\mathcal{E}_{ij}\}$. Each node v_i represents a single pixel or a superpixel region and is connected with several adjacent nodes $\{v_{i,j}\}, j \in \mathcal{N}_{\mathcal{G}}(i)$. Each edge $\mathcal{E}_{ij}$ indicates the spatial relationship between the two corresponding nodes. A feature sequence $\mathbf{F}_i = \{\mathbf{f}_{i,1}, \mathbf{f}_{i,2}, ..., \mathbf{f}_{i,N_i}\}, \mathbf{F}_i \in \mathbb{R}^{d \times N_i}$ is generated for each node v_i as the input of the following classification phase, where N_i is the length of sequence, d denotes the feature dimension. Here we use multiple modalities which can be taken as stacked feature maps. As is illustrated in Fig.2 (a), $\mathbf{F}_i$ is the combination of vectors $\{\mathbf{f}_{i,k}\}_{k=2}^{N_i}$ where $\mathbf{f}_{i,1}$ indicates the features of v_i and others indicate the features of its adjacent nodes.

2.2. Features with Adjacency Constraint

Our method uses the gray intensity of each node to extract feature vectors $\{\mathbf{f}_{i,k}\}_{k=1}^{N_i}$ with its multiple modalities and two ways of adjacency constraint considered as follows.



(a) Feature Sequence $\mathbf{F}_i$ (b) Pixel-wise Constraint (c) Superpixel-wise Constraint

Fig. 2. (a) The components of feature sequence $\mathbf{F}_i$. (b) 8 neighbors are considered for each pixel node. (c) The adjacent nodes are considered for each superpixel node.

Pixel-wise Constraint. One way to generate feature sequences is to take each pixel as a graph node, and extract the features of each node in terms of its adjacent nodes. Since each pixel shares the similar gray intensity with its adjacent pixels in MR brain images, the generated feature sequences are more robust to noise compared with that without adjacency constraint. As is shown in Fig.2 (b), for each node v_i, we choose its 8-neighbors and set $N_i = 9$, $\{\mathbf{f}_i\}_{i=1}^{N_i}$ is the gray intensity of each modality.

Superpixel-wise Constraint. There exists two advantages when generating features with superpixel-wise constraint. Firstly, it can achieve dimensionality reduction as well as improve time efficiency. Secondly, since each superpixel node has internal region consistency, the generated feature sequences are more robust to noise. SLIC [14] is a good superpixel method and thus it is first used on T1-weighted slice in our method since T1 has better discrimination in different classes compared with other modalities like T2 or PD. Then, we set a threshold T to remove most of the superpixel nodes that belong to the background according to [3]. If

the average value of pixels within the node is smaller than T, then the node is classified as background. This will help in reducing training sequences in the classification phase for time acceleration. Next, we generate feature sequences $\mathbf{F}_i$ for the remaining nodes. As is shown in Fig.2 (c), the adjacent superpixel nodes of v_i are considered. We perform pooling operation on each node to compute $\mathbf{F}_i$. Global average pooling is performed on v_i to generate the feature vector $\mathbf{f}_{i,1}$ while global max pooling is performed on its adjacent nodes to generate $\{\mathbf{f}_{i,k}\}_{k=2}^{N_i}$.

2.3. Classification with LSTM/BiLSTM

The generated feature sequence $\mathbf{F}_i$ of node v_i is then fed into the LSTM or BiLSTM units to train the classification model. BiLSTM is a variant of LSTM which combines the forward and backward LSTM. During updating, the k-th LSTM layer determines the current states of each node v_i that comprises the hidden states $\mathbf{h}_{i,k} \in \mathbb{R}^d$ and the memory states $\mathbf{c}_{i,k} \in \mathbb{R}^d$ of each node. Each node is affected by its previous states and the states of adjacent graph nodes as well in order to propagate information to the whole image. Thus, given the current state $\mathbf{f}_{i,k}$ of $\mathbf{F}_i$, the updating rules are as follows:

$$\begin{aligned}
\mathbf{g}^u &= \delta(\mathbf{W}_{fu}\mathbf{f}_{i,k} + \mathbf{W}_{hu}\mathbf{h}_{i,k-1} + \mathbf{b}_u) \\
\mathbf{g}^f &= \delta(\mathbf{W}_{ff}\mathbf{f}_{i,k} + \mathbf{W}_{hf}\mathbf{h}_{i,k-1} + \mathbf{b}_f) \\
\mathbf{g}^o &= \delta(\mathbf{W}_{fo}\mathbf{f}_{i,k} + \mathbf{W}_{ho}\mathbf{h}_{i,k-1} + \mathbf{b}_o) \\
\mathbf{g}^c &= \tanh(\mathbf{W}_{fc}\mathbf{f}_{i,k} + \mathbf{W}_{hc}\mathbf{h}_{i,k-1} + \mathbf{b}_c) \\
\mathbf{c}_{i,k} &= \mathbf{g}^f \odot \mathbf{c}_{i,k-1} + \mathbf{g}^u \odot \mathbf{g}^c \\
\mathbf{h}_{i,k} &= \mathbf{g}^o \odot \tanh(\mathbf{c}_k)
\end{aligned} \tag{1}$$

where $\mathbf{g}^u, \mathbf{g}^f, \mathbf{g}^o, \mathbf{g}^c$ indicate the four gates of LSTM. $\mathbf{W}$ are the weight matrices learned from the network and $\mathbf{b}$ are the bias vectors. $\delta(\cdot)$ is the logistic sigmoid function and $\odot$ denotes the point-wise product. Back propagation is utilized to update the weights during the training phase. After that, the LSTM last hidden output will be used to predict the final class label followed by a fully connected layer and softmax layer. Finally, the segmentation image can be obtained by combining all the classified nodes of a slice.

3. EXPERIMENTS

In this section, we evaluate the performance of LSTM-MA on two datasets: BrainWeb [15] and MRBrainS [16]. BrainWeb contains simulated MRI volumes for normal brain with three modalities: T1, T2 and PD. We use 399 slices from one volume among which 239 slices for training, and 160 for testing. MRBrainS contains T1, T1 inversion recovery and FLAIR sequences and we choose 104 slices for training and 70 slices for testing from transversal view. The quantitative evaluation metrics used are Pixel Accuracy (PA) and Dice Similarity Coefficient (DSC). When training our LSTM-MA model, Adam [17] is used as the optimization technique. The number of hidden units in LSTM/BiLSTM is set to 40 for a balance of time efficiency and segmentation performance [10].

In our experiments, we first report an ablation study to learn the effect of using the adjacent nodes and multi-modality slices. Then, we compared our LSTM-MA/BiLSTM-MA with some popular segmentation methods. Finally, we evaluate the robustness of our method on MR images corrupted by noise.

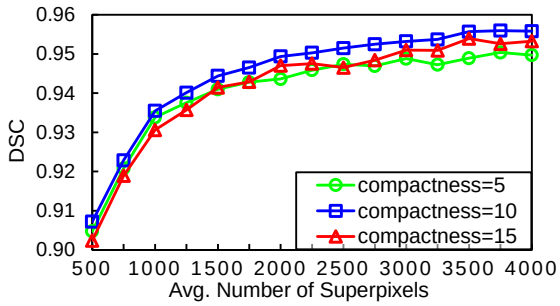


Fig. 3. The impact of SLIC superpixel parameters (average number of superpixels and compactness) on the DSC.

Table 1. Ablation study.

Constraint	DSC		Test time / Slice(s)	
	BrainWeb	*MRBrainS*	*BrainWeb*	*MRBrainS*
None	0.9802	0.8405	3.4153	6.3972
Mod	0.9841	0.8525	3.2418	6.4694
Adj	0.9850	0.8601	6.0525	15.6678
Mod+Adj	**0.9866**	**0.8677**	6.0799	15.4801
SLIC	0.9467	0.8115	0.1883	0.1825
SLIC+Mod	0.9489	0.8243	0.1956	0.1746
SLIC+Adj	0.9476	0.8239	0.2718	0.2843
SLIC+Mod+Adj	**0.9494**	**0.8330**	0.2761	0.2947

3.1. Ablation Study

In order to verify the effectiveness of multi-modality and adjacency constraint, we test our pixel-wise and superpixel-wise feature generation methods on two datasets. It is worth noticing that when using SLIC, we should choose a suitable number of superpixels and the compactness. As is shown in Fig.3, a large number of superpixel brings more accurate boundary adherence but costs more time in classification phase. Thus, we set the suitable number of superpixels to 2000 and compactness to 10 in the following experiments. We choose BiLSTM in the classification phase.

The experiments on the proposed method with different information constraint are shown in Table 1. Note that "Mod" refers to the use of multi-modality information and "Adj" refers to the use of adjacency information. It is obvious that employing multi-modality and adjacency constraint are benefit to the improvement of segmentation accuracy and the introduction of superpixel method SLIC speeds up at a large margin compared with pixel-wise ways.

Table 2. Comparison of normal brain image segmentation results on BrainWeb and MRBrianS.

Method	PA		DSC	
	BrainWeb	*MRBrainS*	*BrainWeb*	*MRBrainS*
k-means [1]	0.9866	0.9378	0.9731	0.7873
FCM [2]	0.9902	0.9388	0.9850	0.7909
FCN-8s [7]	0.9030	0.9261	0.7889	0.7570
U-Net [8]	0.9326	0.9183	0.8089	0.7402
SegNet [9]	0.9336	0.9244	0.8613	0.7560
SVM [18]	0.9936	**0.9678**	0.9859	0.8472
KNN [19]	0.9920	0.9460	0.9824	0.8021
Tree [20]	0.9904	0.9502	0.9802	0.7745
LSTM-MA	0.9919	0.9632	0.9827	0.8700
BiLSTM-MA	**0.9938**	0.9633	**0.9866**	**0.8702**

3.2. Experiments on Normal Brains

We evaluate our method compared with the leading methods in brain segmentation. K-means and FCM are unsupervised clustering algorithms and we set the number of clusters K to 4, which corresponds to four classes in our task (including background). SVM, KNN, and Decision Tree are frequently used classifiers in supervised learning. The same features are fed into the five classifiers in classification phase. FCN, U-Net and SegNet are encoder-decoder models based on CNN, and we use the default settings in [7, 8, 9]. The testing results on two datasets are shown in Table 2. We can find that LSTM-MA and BiLSTM-MA achieve competitive results compared with other methods on both datasets, and it is helpful to use multi-modality and adjacency information for better segmentation accuracy. Fig.4 shows a visualization of different methods on BrainWeb. We only choose FCM, SVM, SegNet and BiLSTM-MA due to limited space. Although it is not obvious to differ the differences visually, LSTM-MA or BiLSTM-MA still has its advantage in quantitative performance.

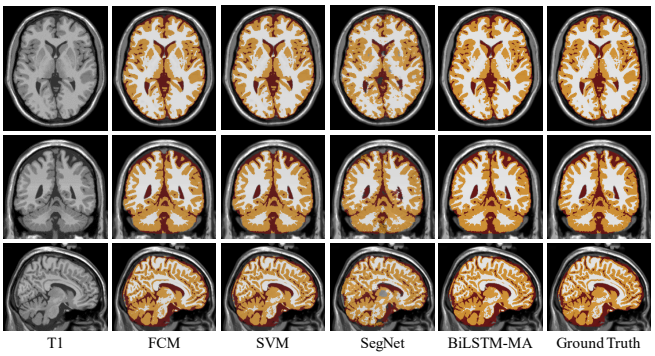


Fig. 4. Segmentation results for WM (white), GM (yellow), and CSF (red) in three orthogonal views on BrainWeb.

3.3. Experiments on Images with Noise

MR brain images are often corrupted by white noise in real environment. In order to evaluate the robustness of our method, we test LSTM-MA and BiLSTM-MA with pixel-wise and superpixel-wise constraint on different levels of noise in comparison with other five segmentation methods, i.e., k-means, FCM, SVM, KNN, Decision Tree. BrainWeb provides image noise with Rayleigh noise. As is shown in Fig.5 (a)(b), we test the seven methods on two slices. The segmentation results are quite closed when the noise ratio is small, yet LSTM-MA/BiLSTM-MA performs much better results when noise ratio becomes larger. This proves that LSTM/BiLSTM is more robust classifiers to noise due to its advantage for sequence processing. Fig.5 (c)(d) illustrates two test results on MRBrainS. The slices on MRBrainS are eroded by Gaussian noise with mean μ set to 0 and variance σ^2 set from 0.01 to 0.09. Our LSTM-MA/BiLSTM-MA with pixel-wise constraint still achieves the best results compared with other methods. Note that PAs of our SLIC+LSTM-MA/BiLSTM-MA do not drop much when σ^2 becomes larger, which suggests that superpixel-wise constraint further facilitates the robustness of the classifier with a little sacrifice in segmentation performance.

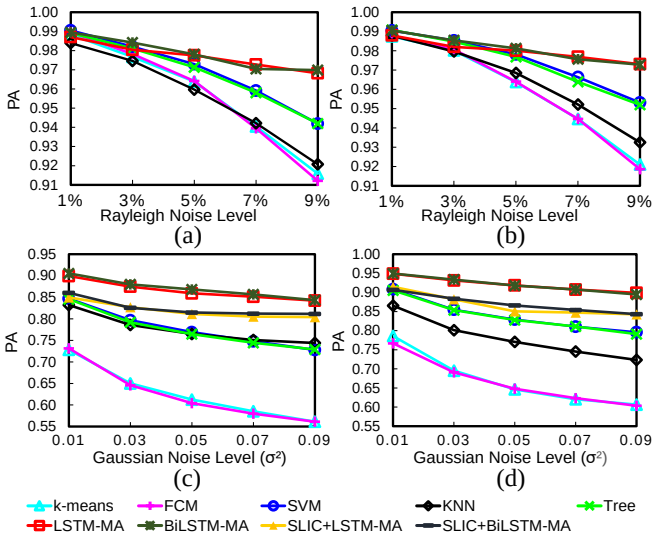


Fig. 5. Experimental results on images corrupted by noise.

4. CONCLUSION

In this paper, we propose a novel LSTM based method with multi-modality and adjacency constraint named LSTM-MA for brain image segmentation. By considering the image as an undirected graph, we utilize two feature sequence generation ways with pixel-wise and superpixel-wise constraint. The final prediction results are obtained by combining the classification result of each graph node. The experimental results demonstrate that our method with pixel-wise constraint can obtain competitive segmentation accuracy, and the one with superpixel-wise further speeds up with a little sacrifice in accuracy. Better segmentation performance can be obtained if we choose more robust superpixel algorithms for future study.

5. REFERENCES

[1] James MacQueen et al., "Some methods for classification and analysis of multivariate observations," in *Proceedings of the fifth Berkeley symposium on mathematical statistics and probability*. Oakland, CA, USA, 1967, vol. 1, pp. 281–297.

[2] Joseph C Dunn, "A fuzzy relative of the isodata process and its use in detecting compact well-separated clusters," 1973.

[3] Hayat Al-Dmour and Ahmed Al-Ani, "A clustering fusion technique for mr brain tissue segmentation," *Neurocomputing*, vol. 275, pp. 546–559, 2018.

[4] Brian Patenaude, Stephen M Smith, David N Kennedy, and Mark Jenkinson, "A bayesian model of shape and appearance for subcortical brain segmentation," *Neuroimage*, vol. 56, no. 3, pp. 907–922, 2011.

[5] Li Wang, Feng Shi, Yaozong Gao, Gang Li, John H. Gilmore, Weili Lin, and Dinggang Shen, "Integration of sparse multi-modality representation and anatomical constraint for isointense infant brain mr image segmentation," *NeuroImage*, vol. 89, pp. 152–164, 2014.

[6] Kuan-Lun Tseng, Yen-Liang Lin, Winston H. Hsu, and Chung-Yang Huang, "Joint sequence learning and cross-modality convolution for 3d biomedical segmentation," *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 3739–3746, 2017.

[7] Evan Shelhamer, Jonathan Long, and Trevor Darrell, "Fully convolutional networks for semantic segmentation," *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 3431–3440, 2015.

[8] Olaf Ronneberger, Philipp Fischer, and Thomas Brox, "U-net: Convolutional networks for biomedical image segmentation," in *MICCAI*, 2015.

[9] Vijay Badrinarayanan, Alex Kendall, and Roberto Cipolla, "Segnet: A deep convolutional encoder-decoder architecture for image segmentation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 39, pp. 2481–2495, 2017.

[10] Sepp Hochreiter and Jürgen Schmidhuber, "Long short-term memory," *Neural computation*, vol. 9, no. 8, pp. 1735–1780, 1997.

[11] Xiaodan Liang, Xiaohui Shen, Jiashi Feng, Liang Lin, and Shuicheng Yan, "Semantic object parsing with graph lstm," in *European Conference on Computer Vision*. Springer, 2016, pp. 125–143.

[12] Xiaodan Liang, Xiaohui Shen, Donglai Xiang, Jiashi Feng, Liang Lin, and Shuicheng Yan, "Semantic object parsing with local-global long short-term memory," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2016, pp. 3185–3193.

[13] Yun Liu, Peng-Tao Jiang, Vahan Petrosyan, Shi-Jie Li, Jiawang Bian, Le Zhang, and Ming-Ming Cheng, "Del: Deep embedding learning for efficient image segmentation.," in *IJCAI*, 2018, vol. 864, p. 870.

[14] Radhakrishna Achanta, Appu Shaji, Kevin Smith, Aurelien Lucchi, Pascal Fua, Sabine Süsstrunk, et al., "Slic superpixels compared to state-of-the-art superpixel methods," *IEEE transactions on pattern analysis and machine intelligence*, vol. 34, no. 11, pp. 2274–2282, 2012.

[15] Chris A Cocosco, Vasken Kollokian, Remi K-S Kwan, G Bruce Pike, and Alan C Evans, "Brainweb: Online interface to a 3d mri simulated brain database," in *NeuroImage*. Citeseer, 1997.

[16] Adriënne M Mendrik, Koen L Vincken, Hugo J Kuijf, Marcel Breeuwer, Willem H Bouvy, Jeroen De Bresser, Amir Alansary, Marleen De Bruijne, Aaron Carass, Ayman El-Baz, et al., "Mrbrains challenge: online evaluation framework for brain image segmentation in 3t mri scans," *Computational intelligence and neuroscience*, vol. 2015, pp. 1, 2015.

[17] Diederik P. Kingma and Jimmy Ba, "Adam: A method for stochastic optimization," *CoRR*, vol. abs/1412.6980, 2014.

[18] Corinna Cortes and Vladimir Vapnik, "Support-vector networks," *Machine learning*, vol. 20, no. 3, pp. 273–297, 1995.

[19] Trevor J. Hastie and Robert Tibshirani, "Discriminant adaptive nearest neighbor classification," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 18, pp. 607–616, 1995.

[20] Leo Breiman, Joseph H Friedman, R. A. Olshen, and C. J. Stone, "Classification and regression trees," 1984.