MULTISCALE DEEP CONVOLUTIONAL NETWORKS FOR CHARACTERIZATION AND DETECTION OF ALZHEIMER'S DISEASE USING MR IMAGES

Chenjie Ge[1,3], *Qixun Qu*[1], *Irene Yu-Hua Gu*[1], *Asgeir Store Jakola*[2]

[1]Dept. of Electrical Engineering, Chalmers University of Technology, Sweden
[2]Inst. of Neuroscience and Physiology, Sahlgrenska Academy, University of Gothenburg, Sweden
[3]Inst. of Image Processing and Pattern Recognition, Shanghai Jiao Tong University, China

ABSTRACT

This paper addresses the issues of Alzheimer's disease (AD) characterization and detection from Magnetic Resonance Images (MRIs). Many existing AD detection methods use single-scale feature learning from brain scans. In this paper, we propose a multiscale deep learning architecture for learning AD features. The main contributions of the paper include: (a) propose a novel 3D multiscale CNN architecture for the dedicated task of AD detection; (b) propose a feature fusion and enhancement strategy for multiscale features; (c) empirical study on the impact of several settings, including two dataset partitioning approaches, and the use of multiscale and feature enhancement. Experiments were conducted on an open ADNI dataset (1198 brain scans from 337 subjects), test results have shown the effectiveness of the proposed method with test accuracy of 93.53%, 87.24% (best, average) on subject-separated dataset, and 99.44%, 98.80% (best, average) on random brain scan-partitioned dataset. Comparison with eight existing methods has provided further support to the proposed method.

Index Terms— Alzheimer's disease detection, MR images, multiscale features, multiscale CNN, feature fusion and enhancement.

1. INTRODUCTION

Alzheimer's disease (AD) is a progressive brain disease affecting people in various ways. By destroying brain cells, AD gradually leads to memory and thinking skill losses. Eventually patients even fail to carry out the simplest tasks. Hence, detection and treatment of AD is very important. Some existing techniques for medical assessment of AD include physical and neurobiological exams, Mini-Mental State Examination (MMSE). Recently, detection of AD using brain MR images has drawn more attentions due to the possibility of early treatment or intervention. For detection of AD patients, MR images are often used in addition to other clinical symptoms. Characterization of AD from MR images are usually done by extracting distinctive features related to the disease, where a detection or classification method is then followed to distinguish between AD patients and normal control (NC) subjects.

Feature learning/extraction methods can be roughly divided into two paradigms, depending on the manner whether features are extracted by hand-crafted designing techniques or by automatic approaches. The former is often related to analysis or model-based methods, and the latter related to deep learning methods. Automatically learning features by deep learning methods has been gaining considerable interests lately, as hand-crafted features are heavily dependent on the knowledge of human experts. For AD detection, many promising works exist by using methods under these paradigms. Tong et al. [1] proposed a multiple instance learning strategy to classify dementia and extracted features using bags of MRI voxel patches and graph mapping. Liu et al. [2] proposed to extract multi-view features using selected templates. Encoding features were then obtained by clustering subjects in each view space, followed by an ensemble of SVMs to classify the subjects. For deep learning methods, Bäckström el al. [3] proposed an efficient and simple-structure 3D CNN for AD detection, where good results were obtained on a relatively large dataset containing about 340 subjects. Sarraf el al. [4] employed CNN architectures LeNet and GoogleNet to detect Alzheimer's disease using sMRI and fMRI brain scans. [5, 6] proposed MRI image segmentation using CNNs on filters with different patch sizes, or images with low and normal resolutions. By pre-selecting CNNs with three fixed patch sizes in [5], or, using low and normal resolution images in [6], multiscale-based segmentation was obtained. Although these methods are rather promising, further improvement is still needed before putting into clinical usage. One issue that might hamper the robust performance in the above methods is that they only use single-scale features to characterize ADs. It is observed that AD characteristics not only exist in the detailed voxel levels, but also in semantic level reflecting the tissue structural changes in AD (e.g. large CSF area, loss of details in GW and WM, and many more).

Motivated by the above, we propose a novel 3D multiscale CNN architecture to detect Alzheimer's disease using MR

Emails: {chenjie, irenegu}@chalmers.se, quqixun@gmail.com, asgeir.jakola@vgregion.se. The project is partially financed by STINT in Sweden and CSC scholarship in China

Authorized licensed use limited to: Tianjin University of Technology. Downloaded on September 01,2020 at 05:07:39 UTC from IEEE Xplore. Restrictions apply.

images. The proposed method is based on the observation that tissues in a brain scan may vary in scales and structures. Therefore, incorporating the multiscale feature learning into a deep convolutional network may lead to learning detailed tissue features (in several scales) as well as semantic features, which are effective for the characterization of Alzheimer's disease. To make the full use of multiscale features, a fusion and enhancement strategy is applied for refining and combining features over different scales. Features from these scales are pooled separately to form multiscale feature maps.

2. PROPOSED AD DETECTION METHOD

The basic idea behind the proposed AD detection method is to incorporate multiscale feature learning into a deep convolutional network architecture for the characterization of Alzheimer's disease from MRIs. This is motivated by the idea that multiscale features are rich in both semantic as well as voxel level information, and hence may lead to more discriminative feature representation of AD. In the proposed method, 3D MRIs are first pre-processed, followed by a novel 3D multiscale convolutional network (3D MSCNN) to extract features in different scales. These features are then fused through concatenation and enhanced through boosting, followed by fully-connected layers for the classification. Fig.1 shows the block diagram of the proposed method, where the detail in the last three blocks is further sketched in Fig.2.

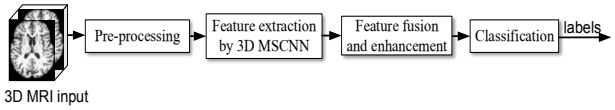

Fig. 1. The pipeline of the proposed method.

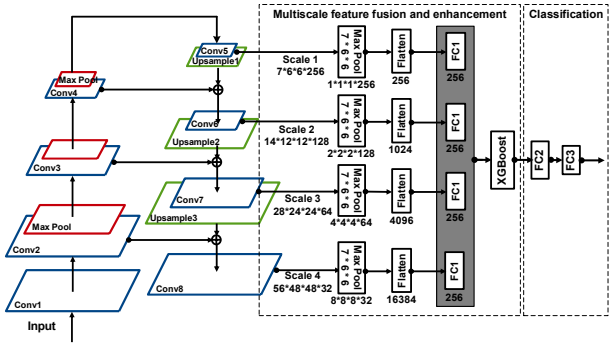

Fig. 2. Detail of feature extraction, fusion and enhancement, and classification in Fig.1 (blocks 2-4).

In the following, we describe the proposed method in details.

2.1. Data Pre-Processing

Pre-processing is an essential step that would significantly influence the performance of AD detection. The pre-processing consists of three steps: cortical reconstruction, image size normalization, and intensity normalization. The step *cortical reconstruction* is conducted by the dataset provider, though one can also perform this step by using *recon-all* from *FreeSurfer* software package [7] that contains a set of processing such as motion correction and conform, non-uniform intensity normalization, Talairach transform computation, intensity normalization, skull and neck removal. The step *image size normalization* includes removing the excessive background image volume outside the brain tissue by trimming the 3D image according to the largest brain size in the dataset, and then resizing each trimmed 3D images into a pre-selected size (112 x 96 x 96 pixels in our tests). Choosing the normalization size is a tradeoff between the GPU memory available, computational cost, and possible performance degradation using low resolution 3D images.

2.2. 3D MSCNN Architecture and Feature Selection

3D MultiScale Convolutional Neural Network (MSCNN) layers, and feature fusion and enhancement layers, as shown in Fig.2, are used for extracting multiscale features. During the training, these steps are used to learn the network model and coefficients related to feature extraction.

Multiscale convolutional network layers: The proposed MSCNN consists of 8 layers (Conv1-Conv8) where the last 3 layers (Conv6-Conv8) are obtained from "skip connection" as shown in Fig.2. The first 5 convolutional layers generate feature maps in coarse-to-fine scales. The outputs of bottom-up pathway of convolutional layers 2,3,4 are added to the upsampled (by a factor of 2) top-down pathway of convolutional layers 7,6,5, respectively.

Feature fusion and enhancement layers: As shown in Fig.2, feature fusion and enhancement layers fuse multiscale feature maps into a vector. Before fusion, each feature map is separately pooled so that feature maps in different scales can remain different resolution. A simple feature fusion strategy is applied to fuse the multiscale features by concatenating feature vectors from all scale levels. For further enhancing the features and reducing the feature dimension, a gradient boosting-based machine learning method XGBoost [8] is then employed, where the features are ranked according to their importance in characterization of AD. Only feature components with high importance are chosen to form the final feature vector.

Classification layers: The classifier consists of 2 fully connected layers, where *softmax* is used as the activation function for the final layer.

Remarks: The MSCNN architecture is different from the U-nets [9], since the input of a top-down pathway layer is formed from the sum of the bottom-up pathway output and the upsampled top-down pathway output. Further, the architecture has some similarity to that in [10], however, it differs in terms of convolutional filter sizes and subsequent layers, e.g. using 3*3*3 convolution filter before merging layers, further, multiscale feature maps from upsampling and convolution at different scales are pooled with the same rates for obtaining different size feature maps in all scales. Furthermore, the feature fusion and enhancement step is added to obtain the final feature vector where the dimension of multiscale feature vec-

Authorized licensed use limited to: Tianjin University of Technology. Downloaded on September 01,2020 at 05:07:39 UTC from IEEE Xplore. Restrictions apply.

tor is significantly reduced without compromising the performance. The method also differs from the multiscale concept in [5, 6] used for MRI segmentation whose multiscale was obtained by selecting either fixed patch sizes in 3 different streams or low/normal resolution MRIs for training, whereas the multiscale concept in the proposed method is obtained in a similar fashion as that in the multiscale wavelet analysis however through a deep learning approach.

3. EXPERIMENTAL RESULTS

3.1. Setup

KERAS library [11] with TensorFlow [12] backend was used for network training. All experiments were done on a workstation with Intel-i7 3.40GHz CPU, 48G RAM and a NVIDIA Titan XP 12GB GPU. Network weights were learnt using *Adam* optimizer. Hyper-parameters were chosen through carefully tuning the network through experiments: the number of epochs for training was set to 150; Step-wise learning rate was used, where the learning rate was set to 0.001 for epochs $\in$ [1, 50], 0.0001 for epochs $\in$ [51, 100], and 0.00001 for epochs $\in$ [101, 150]; the dropout rate used in the flatten layers and concatenation layer was set to 0.5. For all layers except the output layer, L2 regularization was used with the regularization parameter 0.00005. Batch processing was used where the batch size was set to 8, and the batch normalization momentum value was set to 0.9. The initialization method was chosen as *glorot_uniform* for all layers.

Dataset: The ADNI dataset used in our experiments was obtained from Alzheimer's Disease Neuroimaging Initiative [13]. In our experiments, T1 MRI scans of 2 classes (AD and NC) were used for the classification. Detailed information of the dataset is given in Table 1. From Table 1, one can see that each subject may contain several scans (average 3.5) that were made from different time.

Table 1. ADNI dataset used for our experiments. Females/males: F/M, Four age groups were included: ([<60)/ [60,70) / [70,80) /[≥80])

Class	# Subjects (F/M)	Age group	#3D Scans
AD	198 (95/103)	(9/35/92/62)	600
NC	139 (66/73)	(0/5/108/26)	598
Total	337 (161/176)	(9/40/200/88)	1198

3.2. Dataset Partitioning and Approaches

The whole dataset was then partitioned to 70%, 15%, 15% to form the training, validation and testing sets, respectively. Two dataset partition strategies were applied, related to 2 case studies: Case-A and Case-B.

In Case-A study, we randomly partition the dataset into two subsets, i.e., training + validation, and testing, where brain scans are partitioned according to randomly selected subjects. This means that different scans from a same subject would only end up in the training, or the testing subset, but not in both. This partition strategy is more relevant to clinical scenarios and usefulness.

In Case-B study, the second partition strategy was used where training, validation and testing subsets were partitioned according to randomly selected brain scans regardless of subjects. Case-B study is clinically less relevant, however, this study enables us to obtain some indication of the proposed method as comparing with other existing methods.

3.3. Results

To test the effectiveness of the proposed method, we conducted experiments on both Case-A and Case-B studies.

Results from Case-A study: To test the effectiveness of the proposed method, Case-A study was conducted on the dataset using subject-separated dataset partition approach. Fig.3 shows the overall accuracy and loss vs. epochs in the training and validation.

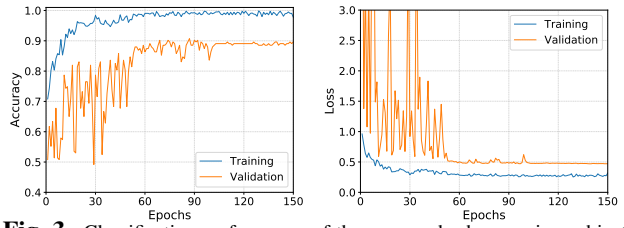

Fig. 3. Classification performance of the proposed scheme using subject-separated dataset partition in Case-A study. Left: accuracy vs. epochs; right: loss vs. epochs.

Observing Fig.3, one can see that training and validation have converged more or less after 100 epochs. There is a performance drop in the validation as comparing with that in the training, indicating possible overfitting probably due to the relatively moderate size of the dataset used.

For evaluating the performance of the proposed method, Table 2(a) shows the performance on testing sets in 5 runs. In each run, the dataset was re-partitioned randomly again according to subjects. To further examine the performance of individual classes, Table 2(b) shows the confusion matrix on the test set according to one selected run.

Table 2. Test performance from Case-A study, where dataset was partitioned according to randomly separated subjects.

Run	1	2	3	4	5	Average ($\|\sigma\|$)
Accuracy (%)	91.28	82.95	85.64	**93.53**	82.78	**87.24** (4.40)

(a) Overall accuracy (on AD and NC) on testing sets and averaged result.

True \ classified	AD (%)	NC (%)
AD (%)	95.51	4.49
NC (%)	8.64	91.36

(b) Confusion matrix for individual classes on the testing set (from the best result in run 4).

Observing Table 2(a), the proposed scheme has resulted in rather good test accuracy, highest 93.53% and average 87.24% in 5 runs. The standard deviation is relatively large probably due to the moderate size of dataset used, hence relatively poor distribution of feature statistics. Observing Table 2(b), the performance in two classes (AD and NC) are relatively balanced, where AD has a slightly high detection rate 95.51% with a relatively high false alarm 8.64%.

Authorized licensed use limited to: Tianjin University of Technology. Downloaded on September 01,2020 at 05:07:39 UTC from IEEE Xplore. Restrictions apply.

Results from Case-B study: This case study is designed to further examine the effectiveness of the proposed method through comparing with a baseline method and several existing methods, even though Case-B study is clinically less important. Table 3(a) shows the testing performance of the proposed method when dataset was randomly partitioned according to brain scans. In each of the 5 runs, the dataset was re-partitioned randomly again. To examine the performance on individual classes, Table 3(b) shows the confusion matrix on the test set from the best run (run 5).

Table 3. Experimental results in Case-B study, where dataset was partitioned randomly according to brain scans.

Run	1	2	3	4	5	Average ($\|\sigma\|$)
Accuracy	99.00	98.33	98.33	98.89	**99.44**	**98.80** (0.42)

(a) overall accuracy (on AD and NC) on testing sets.

True \ classified	AD	NC
AD	98.89 %	1.11%
NC	0.01%	99.99 %

(b) confusion matrix for individual classes on the test set (in run 5).

Observing Table 3(a), the proposed scheme has generated very high accuracy on the test set (highest 99.44% and average 98.80% in 5 runs with a small standard deviation 0.42). Also one can observe from Table 3(b) that results from 2 class are well balanced, where the false alarm is very low (0.01%). Compared with Case-A results, this is not a surprise as different scans from same subjects, which could be highly correlated though different, might be ended up in both the training and testing subsets.

3.4. Analysis through Empirical Tests

Two tests were conducted to examine the impact of different parts in the proposed method, described as follows.

Impact of feature enhancement: To examine the effectiveness of feature fusion and enhancement part, Table 4 shows the dimension of feature vector before and after the feature enhancement. Observing Table 4, there is a significant reduction of the feature vector dimension (reduced to 19.0% of its original size), without compromising the test performance.

Table 4. Feature fusion and enhancement: dimension reduction.

Feature vector	After fusion	After XGBoost
Dimension	1024	195

Impact of multiscale features: To examine the impact of using multiscale features, we compare the proposed method with a baseline 3D CNN method [3], where the results were on the test set, using subject-separated partition approach on the same dataset (see Table 1), and the best results were used in both case. Observing Table 5, applying the proposed method has improved test performance by 3.42%.

Table 5. Comparison between the proposed scheme and a baseline 3D CNN method in [3].

Method	3D CNN	Proposed
Test accuracy (%)	90.11	**93.53**

3.5. Comparisons with State-of-the-Art Methods

We include performance comparisons with eight state-of-the-art methods, noting that such comparison may only give an *indication of the relative performance* since the size and scans in the datasets, as well as the number of brain scans per subject were different in these methods. Table 6 shows the results from eight existing methods, where only the best run result is shown in the proposed method, as most existing methods did not show multiple test results/statistics.

Table 6. Comparison of the proposed method with 8 existing state-of-the-art methods, as an indication of the relative performance.

Method	# Subjects AD/NC/MCI	# Scans AD/NC/MCI	Accuracy (%) AD vs. NC
Deep AD [4]	211/91/-	5:1 AD/NC	98.84
3DAECNN [14]	70/70/70	-	97.60
AE$^+$ [15]	65/77/169	-	87.76
SAE [16]	200/232/411	755/1278/2282	94.74
ICA [17]	202/236/410	-	85.70
MIL [1]	198/231/405	-	88.80
3DCNN [3]	198/139/-	600/598	98.74
Proposed	198/139/-	600/598	**99.44**

(a) using random partition according to MRI scans.

Method	# Subjects AD/NC/MCI	# Scans AD/NC/MCI	Accuracy (%) AD vs. NC
SAE-CNN [18]	755/755/755	755/755/755	**95.39**
3DCNN [3]	198/139/-	600/598	90.11
Proposed	198/139/-	600/598	93.53

(b) using random partition according to subjects.

Results in Table 6 indicate that proposed method achieved relatively good performance, also SAE-CNN [18] has the best performance on a large dataset with one scan per subject.

Discussion: The proposed method was tested on a moderate dataset, downloadable for public use. We did not use data augmentation, as the simple augmentation by left-right brain image flipping did not lead to performance improvement in our tests. Further performance improvement using subject-separated dataset partition is needed, e.g., by using a large dataset, or robust augmentation methods to increase the training dataset.

4. CONCLUSION

The proposed Alzheimer's disease detection method, using 3D multiscale CNN followed by feature fusion and enhancement, has been tested on a moderate ADNI dataset (1198 scans from 337 subjects). Our test results have shown that the proposed method is effective for learning and characterizing the discriminant features of Alzheimer's disease. The proposed scheme has obtained good classification performance (with best 93.53% and average 87.24% in 5 runs) on AD and NC classes using subject-separated dataset. Comparison with 8 existing methods has indicated that the proposed method is promising for AD detection. Future work will be on using large datasets either from augmented or measured data to further improve the performance, and on adding mild cognitive impairment (MCI) subject class in the experiments, as MCI can indicate some early stage AD and hence is also clinically useful information.

Authorized licensed use limited to: Tianjin University of Technology. Downloaded on September 01,2020 at 05:07:39 UTC from IEEE Xplore. Restrictions apply.

5. REFERENCES

[1] Tong Tong, Robin Wolz, Qinquan Gao, Ricardo Guerrero, Joseph V Hajnal, and Daniel Rueckert, "Multiple instance learning for classification of dementia in brain mri," *Medical image analysis*, vol. 18, no. 5, pp. 808–818, 2014.

[2] Mingxia Liu, Daoqiang Zhang, Ehsan Adeli, and Dinggang Shen, "Inherent structure-based multiview learning with multitemplate feature representation for alzheimer's disease diagnosis," *IEEE Transactions on Biomedical Engineering*, vol. 63, no. 7, pp. 1473–1482, 2016.

[3] Karl Bäckström, Mahmood Nazari, Irene Yu-Hua Gu, and Asgeir Store Jakola, "An efficient 3d deep convolutional network for alzheimers disease diagnosis using mr images," in *Biomedical Imaging, IEEE 15th International Symposium on*. IEEE, 2018, pp. 149–153.

[4] Saman Sarraf, Ghassem Tofighi, et al., "Deepad: Alzheimer s disease classification via deep convolutional neural networks using mri and fmri," *bioRxiv*, p. 070441, 2016.

[5] Pim Moeskops, Max A Viergever, et al., "Automatic segmentation of mr brain images with a convolutional neural network," *IEEE transactions on medical imaging*, vol. 35, no. 5, pp. 1252–1262, 2016.

[6] Konstantinos Kamnitsas, Christian Ledig, Virginia FJ Newcombe, et al., "Efficient multi-scale 3d cnn with fully connected crf for accurate brain lesion segmentation," *Medical Image Analysis*, vol. 36, pp. 61–78, 2017.

[7] "Freesurfer," https://surfer.nmr.mgh.harvard.edu/fswiki.

[8] Tianqi Chen and Carlos Guestrin, "Xgboost: A scalable tree boosting system," in *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining*. ACM, 2016, pp. 785–794.

[9] Olaf Ronneberger, Philipp Fischer, and Thomas Brox, "U-net: Convolutional networks for biomedical image segmentation," in *International Conference on Medical image computing and computer-assisted intervention*. Springer, 2015, pp. 234–241.

[10] Tsung-Yi Lin, Piotr Dollár, Ross Girshick, Kaiming He, Bharath Hariharan, and Serge Belongie, "Feature pyramid networks for object detection," in *CVPR*, 2017, vol. 1, p. 4.

[11] F Chollet et al., "Keras," https://github.com/fchollet/keras, 2015.

[12] Abadi M and Agarwal A, "TensorFlow: Large-scale machine learning on heterogeneous systems," 2015.

[13] "Alzheimer's disease neuroimaging initiative," https://adni.loni.usc.edu/about/.

[14] Ehsan Hosseini-Asl, Robert Keynton, and Ayman El-Baz, "Alzheimer's disease diagnostics by adaptation of 3d convolutional network," in *Image Processing (ICIP), 2016 IEEE International Conference on*. IEEE, 2016, pp. 126–130.

[15] Siqi Liu, Sidong Liu, Weidong Cai, Sonia Pujol, Ron Kikinis, and Dagan Feng, "Early diagnosis of alzheimer's disease with deep learning," in *Biomedical Imaging (ISBI), 2014 IEEE 11th International Symposium on*. IEEE, 2014, pp. 1015–1018.

[16] Ashish Gupta, Murat Ayhan, and Anthony Maida, "Natural image bases to represent neuroimaging data," in *International Conference on Machine Learning*, 2013, pp. 987–994.

[17] Wenlu Yang, Ronald LM Lui, Jia-Hong Gao, Tony F Chan, Shing-Tung Yau, Reisa A Sperling, and Xudong Huang, "Independent component analysis-based classification of alzheimer's disease mri data," *Journal of Alzheimer's disease*, vol. 24, no. 4, pp. 775–783, 2011.

[18] Adrien Payan and Giovanni Montana, "Predicting alzheimer's disease: a neuroimaging study with 3d convolutional neural networks," *arXiv preprint arXiv:1502.02506*, 2015.

Authorized licensed use limited to: Tianjin University of Technology. Downloaded on September 01,2020 at 05:07:39 UTC from IEEE Xplore. Restrictions apply.