Alzheimer Diagnosis by a Brain MR-Image

Deep Learning Course (CS230) Project

Soheil Esmaeilzadeh (soes@stanford.edu), Dimitrios Belivanis (dbelivan@stanford.edu)

Introduction

Can a brain MR-Image tell us whether someone has an Alzheimer disease?

In this project, we develop a 3-dimensional Convolutional Neural Network (3D-CNN) that learns the correlation between a brain MR-Image and existence of Alzheimer disease (AD) in a patient. The resulting trained model will be used further to discover regions in a brain that are critical for AD diagnosis.

Data Set & Features

A public dataset is used in this study, Alzheimer's Disease Neuroimaging Initiative-1 (ADNI-1) that includes 816 brain MR-Images. The age-class distribution of this data set is shown in figure 1. As can be seen in figure 1 a good distribution between the 3 AD classes exists.

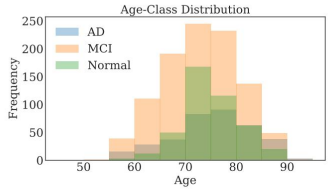

Figure 1: Age-class distribution of brain MR-Images

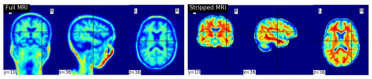

Figure 2: Brain MR-Image skull stripping - from *left* to *right*: Coronal, Axial, and Sagittal views

In this work, we use the Brain Extraction Technique (BET) together with an Statistical Parametric Mapping (SPM) as a voxel based approach for brain image segmentation and extraction and choose the stripped version with higher brain tissue intensity. Performing skull stripping on brain MRI images reduces the size of images by a factor of 2 and reduces the amount of time that is needed for training. Figure 2 shows the raw MR-Images together with the skull stripped versions which have dimension of $116 \times 130 \times 83$.

Training Process

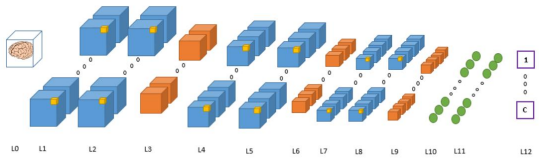

Figure 3: 3D-convolutional neural network (L_0: MR-Image ($116 \times 130 \times 83$); $L_{1,2}$: Conv. ($3^3 \times 16$); $L_{4,5}$: Conv. ($3^3 \times 64$); $L_{7,8}$: Conv. ($3^3 \times 128$); L_3, L_6, L_9: Max-pool (2^3 ,A^3 ,A^3); L_{10}, L_{11}: F.C.(512 ,128); L_{12}: Output (2))

We have built a 3D-CNN model in tensorflow, two architectures are considered, (I). a complex architecture as described in figure 3 and (II). a simplified version which has only one convolution before each max-pooling and also has only one F.C. layer, for both architectures *ReLu* is used as the activation function. Our cost function of choice is cross entropy and is minimized with the *Adam* optimizer. The hyperparameters we experiment on are the β coefficient of the L_2-regularization, the dropout probability and the size of batches, in addition to learning rate, number of filters in the convolutional layers and number of neurons in the F.C. layers.

We use F_2 score (given by equations (1) and (2)) which depends on precision and recall to evaluate the performance of our model with true positive, true negative, false positive, and false negative being as TP, TN, FP, and FN respectively.

$$\text{Precision} = \frac{TP}{TP + FP}, \quad \text{Recall} = \frac{TP}{TP + FN} \tag{1}$$

$$F_2 = \frac{5 \text{ precision} \cdot \text{recall}}{4 \text{ precision} + \text{recall}} \tag{2}$$

Results

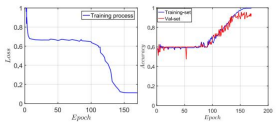

Figure 4: Train-set and validation-set accuracy and loss function's value based on the number of Epochs

Figure 4 shows the training process accuracy and the value of loss for the model with the best performance in Table.1.

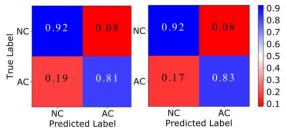

Figure 5: Normalized confusion matrix - *left*: dev-set, *right*: test-set

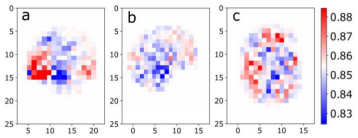

Figure 6: Probability of correct class for Alzheimer detection - (a). Coronal (b). Axial (c). and Sagittal views

Furthermore, We evaluated the confidence of our model for correctly predicting AD by introducing a cube of $5 \times 5 \times 5$ voxels in an MR-Image. As shown in figure 6 the alteration of the model's confidence is small, meaning that the prediction of AD is based on global information. Furthermore, the area which affects the prediction the most coincides with the *Hippocampus*, a region of the brain which is related to memory and is reported to play roles in occurrence of AD.

Experiments

Our results are given by the following table.

Experiments	F2-Train	F2-Val
Cmp	0.987	0.691
Smp	0.983	0.713
Cmp+Reg	0.981	0.751
Smp+Reg	0.978	0.765
Cmp+Drp+Reg	0.967	0.821
Smp+Drp+Reg	0.973	0.811
Aug.-Cmp+Drp+Reg	0.991	0.891
Aug.-Smp+Drp+Reg	0.998	0.933

Table 1: Performance of experiments (Cmp: complex CNN; Smp: simple CNN; Reg: with L_2-regularization; Drp: with dropout; Aug: augmented data)

Learning Transfer

We use the best model that is found in the binary classification of AD in Table.1. to further subclassify Alzheimer in another category called Mild Cognitive Impairment, which is the stage before a full AD.

Experiments	F2-Train	F2-Val
Aug.-Cmp+Drp+Reg	0.983	0.572
Aug.-Smp+Drp+Reg	0.991	0.611

Table 2: Performance of learning transfer for 3-class AD diagnosis (Cmp: complex CNN; Smp: simple CNN; Reg: with L_2-regularization; Drp: with dropout; Aug: augmented data)

Conclusion & Future Work

In this work, we diagnosed Alzheimer disease by MR-Images. We found critical parts (*hippocampus*) in three anatomical planes contributing to AD. With an extensive hyperparameter tuning and finding the best model architecture for binary classification, we fine-tuned it for further subclassifying AD into i.e. Mild Cognitive Impairment.

- Improving the 3-class AD diagnosis model
- Getting access to ADNI2 dataset and further training