CONVOLUITIONAL NEURAL NETWORKS FOR MS LESION SEGMENTATION, METHOD DESCRIPTION OF DIAG TEAM

Mohsen Ghafoorian

Institute for Computing and Information Sciences
Radboud University
Nijmegen, the Netherlands

Bram Platel

Diagnostic Image Analysis Group
Radboud University Medical Center
Nijmegen, the Netherlands

ABSTRACT

Recently deep neural networks have shown many successful applications in different domains. For this lesion segmentation task, we utilize a deep convolutional neural network with 5 layers in a sliding window fashion to create a voxel-based classifier. We evaluate our system with Dice similarity, misclassification rate and area under the ROC curve. Based on experimental results our proposed CAD system reaches average Dice similarity, misclassification rate and area under the ROC curve of 0.976, 0.565, 0.073 respectively.

Index Terms— Multiple sclerosis, automated segmentation, convolutional neural networks, deep learning

1. INTRODUCTION

Since the past several years there have been many attempts to incorporate automated feature representation learning methods in our machine learning tasks, resulting in a broad range of approaches [1]. Among all these methods, convolutional neural networks (CNN) [2] have been extremely popular due to several interesting properties such as spatial invariance, hierarchical feature learning and scalability [1]. CNNs are a special form of artificial neural networks that significantly reduce the dimensionality of the weights to be learned by means of weight sharing among different connections, that can be interpreted as convolutions of filters with the input images. CNNs consist of multiple layers, each containing a bank of filters, that are convolved with the input, a non-linearity and a pooling module. If CNNs are trained with a large set of training samples, they will automatically learn a hierarchical feature representation, starting with various edge structures in the first layer and more complex feature representations in the deeper layers.
In this description, we elaborate a voxel classification approach considering a fixed-sized neighborhood of the voxel of interest that uses CNNs to learn and soft-classify all of the candidate voxels in the query images.

2. METHODS

We establish a method that learns to label small $w \times w$ patches, indicating if the central voxel is representing a lesion part or not. Although we classify patches what we offer can be seen as a voxel classification approach.

2.1. Preprocessing

Since the images were already brain extracted and co-registered to the other modalities of the same patient, we only normalized the image intensities in all of the modalities to be in range of [0,1]. To prevent over suppression of image intensities due to possible hyper-intense abnormalities, we calculate 95th intensity percentile and normalize the intensities with that value as the maximum intensity that maps to one. We skip other time consuming preprocessing steps such as bias-field correction in favor of time efficiency.

2.2. Sampling

We use a leave one patient out cross validation method and in the ith iteration, we pick samples from all patients except patient i, which is used to extract the validation patches. While sampling from each patient, we use all four time points and from each we pick all possible positive samples. The negative patches are accordingly sampled to keep the balance between the two classes in the dataset. The approximate final sizes of the five created training datasets are 430k, 320k, 540k, 570k and 560k respectively. No data augmentation methods have been applied to artificially increment the size of data. Since human experts are usually better in specificity than in sensitivity, we use the logical OR operation to create a better reference standard from the two provided human expert annotations.

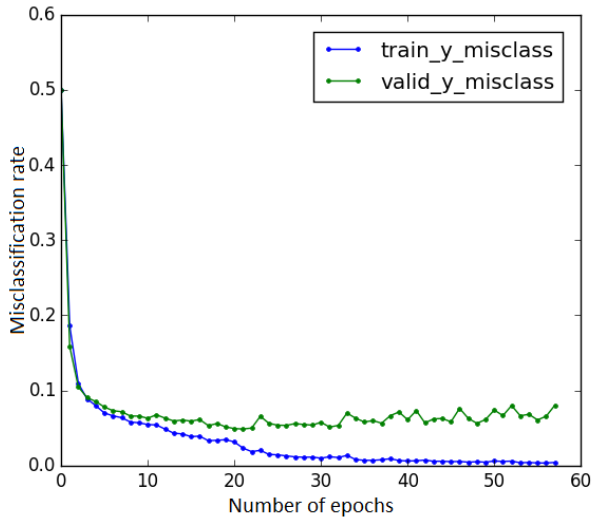

Fig. 1: Performance of the proposed CAD system evaluated on patient 5

2.3. CNN Architecture and Parameters

For the classification of the patches, we train a 5 layers CNN that takes 32×32 patches in four channels as its input samples. There are four convolutional layers with rectified linear non-linearities that have respectively 15 filters of size (13×13), 25 filters of size (9×9) , 60 filters of size (7×7) and finally 130 (3×3) filters. We do not use pooling since it results in a sort of translation invariance that is not desirable for a classifier that assigns the label of the whole patch to its central voxel. A final logistic regression model will classify the resulting responses to the filters in the last convolutional layer.

We use a stochastic gradient descend for the optimization with batch size of 64 and 0.0001 as the learning rate. We run the optimization for 50 epochs and pick the best classifier based on validation set misclassification rate.

3. RESULTS

Figure 1 shows the error rate curve of our system on subsequent epochs evaluated on patient 5 and compares these on training and validation set. The best obtained misclassification rate and area under the ROC curve for each of the 5 trained CNNs are shown in table 1.

	P1	P2	P3	P4	P5
Error rate	0.079	0.062	0.059	0.12	0.048
AUC	0.975	0.980	0.991	0.947	0.990

Table 1: Error rate and area under the ROC curve for the 5 trained CNNs on each five patients

Since the most commonly used measure for the performance of segmentation tasks is Dice similarity measure, we also calculate that to assess the performance of the system. Table 2 demonstrates the Dice similarities for each of the classifier-patient pair segmentation with the reference standard and compares it to the Dice similarity of the two human experts.

	P1	P2	P3	P4	P5
CAD - (exp1 or exp2)	0.63	0.66	0.61	0.37	0.56
exp1 - exp2	0.73	0.84	0.77	0.61	0.65

Table 2: A comparison of Dice similarity of the two human experts and the CAD system

4. REFERENCES

[1] Yoshua Bengio, Aaron Courville, and Pascal Vincent, "Representation learning: A review and new perspectives," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 35, no. 8, pp. 1798–1828, 2013.

[2] Yann LeCun, Bernhard Boser, John S Denker, Donnie Henderson, Richard E Howard, Wayne Hubbard, and Lawrence D Jackel, "Backpropagation applied to handwritten zip code recognition," *Neural computation*, vol. 1, no. 4, pp. 541–551, 1989.