Segmentation of Brain Tumor Tissues with Convolutional Neural Networks

Conference Paper · September 2014

CITATIONS
2

READS
368

4 authors, including:

Yani Ioannou
University of Cambridge
15 PUBLICATIONS **80** CITATIONS
SEE PROFILE

Antonio Criminisi
Microsoft
144 PUBLICATIONS **8,224** CITATIONS
SEE PROFILE

Segmentation of Brain Tumor Tissues with Convolutional Neural Networks

Darko Zikic[1], Yani Ioannou[1,2], Matthew Brown[2], and Antonio Criminisi[1]

[1] Microsoft Research, Cambridge, UK
[2] University of Bath, Bath, UK

Abstract In this work, we investigate the possibility to directly apply convolutional neural networks (CNN) to segmentation of brain tumor tissues. As input to the network, we use multi-channel intensity information from a small patch around each point to be labeled. Only standard intensity pre-processing is applied to the input data to account for scanner differences. No post-processing is applied to the output of the CNN. We report promising preliminary results on the high-grade training data from the BraTS 2013 challenge. Work for the final submission will include architecture modifications, parameter tuning and training on the BraTS 2014 training corpus.

1 Introduction

In this work, we apply convolutional neural networks (CNNs) to the problem of brain tumor segmentation. The work is motivated by the recent success of CNNs for object recognitionion 2D images [1], and the availability of efficient off-the-shelf implementations such as Caffe [2].

CNNs are currently primarily used for object recognition, i.e. if an image contains an object, the complete image is assigned the corresponding label. Two exceptions are [3,4], where CNNs are used inside more complex frameworks in order to perform the segmentation. In the domain of medical image analysis, CNNs have been very successfully applied for mitosis detection in 2D histology images [5]. The intermediate step of [5] can be seen as a binary segmentation of mitotic cells, and the use of CNNs in that work as a per-pixel classifier is similar to the one we use here.

In this work, we explore the possibility of applying CNNs to segmentation of brain tumors *directly*. The CNNs operate on standardly pre-processed intensity information, and we apply no further post-processing to their output.

2 Method

For the segmentation task, we use a standard CNN implementation based on multi-channel 2D convolutions, and adapt it such that it operates on multi-channel 3D data usually available for the brain tumor segmentation task.

We apply the CNN in a sliding-window fashion in the 3D space, for each point inside the brain masks. At each point x, the CNN takes as input a multi-channel 3D patch around this point $P(x)$. Given $P(x)$, the CNN is trained to make a class prediction for the central patch point x.

2.1 Input Data Representation

For each case in the BraTS database, the multi-channel 3D data consists of 4 different 3D MR contrast images: contrast enhanced T1 (T1c), T1, T2 and FLAIR. While T1c usually has an isotropic resolution, the other channels originally have a slice distance which is larger than the in-slice element spacing. In the BraTS challenge, all data is resampled to fit the T1c resolution. For each point x to be labeled, we extract a multi-channel patch $P(x)$ around it, which has spatial dimensions d_1, d_2, d_3. Here, $d1$ and $d2$ are taken to be in-slice dimensions corresponding to high resolution, and d_3 is the lower-resolution axial direction.

Having 4 channels in our task, each 4-channel 3D patch $P(x)$ of size $(d_1 \times d_2 \times d_3 \times 4)$ can also be interpreted as a $(4 \cdot d_3)$-channel 2D patch of size $(d_1 \times d_2 \times 4d_3)$, where the 2D space d_1-d_2 corresponds to original MR-slices, in which the original data generally has the highest resolution. We use this interpretation to apply a standard 2D-CNN convolutional architecture to our 3D problem. Thusly, in the first convolutional layer, we use convolutional filters of size $5 \times 5 \times 4d_3$, and perform a 2D convolution with this filter along the dimensions d_1 and d_2 within each patch $P(x)$ of size $19 \times 19 \times 4d_3$.

This approach is taken for two reasons. First, we can use existing efficient off-the-shelf CNN implementations for 2D convolutions without large modifications. Second, performing 2D instead of 3D convolution is computationally more efficient. The justification for this step is that due to lower resolution in d_3 dimensions, we expect that omitting the convolution in this direction will have a minor impact on accuracy.

2.1.1 Pre-processing As additional pre-processing for the BraTS data, we perform inhomogeneity correction in each channel by [6], set the median of each channel to a fixed value of 0, and downsample the images by a factor of two with nearest-neighbor interpolation. Testing is also performed on down-sampled images, and the results are correspondingly upsampled before quantitative evaluation.

2.2 CNN Architecture and Optimization

We use a standard CNN framework following [1], with the following per layer characteristics of the architecture:

- layer 0: input patch of size $19 \times 19 \times 4$,
  (i.e. we currently only use a single slice from each of the 4 channels)
- layer 1: 64 filters of size $5 \times 5 \times 4$,
  (resulting in $15 \times 15 \times 64$ nodes)

- layer 2: max-pooling with kernel size 3 and stride of 3,
  (resulting in $5 \times 5 \times 64$ nodes)
- layer 3: 64 filters of size $3 \times 3 \times 64$,
  (resulting in $3 \times 3 \times 64$ nodes)
- layer 4: fully connected with 512 nodes
- layer 5: soft-max (fully-connected) with 5 output nodes (for the 5 classes)

All inner nodes in the network use a rectified linear unit (ReLU) as a non-linearity term.

We use log-loss as the energy function for training, and optimization is performed with a stochastic gradient descent with momentum.

3 Preliminary Evaluation

Since we did not have access to the BraTS evaluation platform at the time of this submission, we perform the preliminary evaluation on the training data set from the BraTS 2013 challenge. We focus on the 20 high-grade cases from training set. To provide some context, we relate to results of our previous method from [7], which is based on randomized forests (RF).

We perform the evaluation of the CNN approach with a 2-fold validation where, based on the ascending ordering of the test cases IDs, the first fold contains the odd cases, and the second fold contains the even ones. Thus each fold contains 10 cases. Results for each fold are computed by a CNN which was trained on the other fold. For training, we use all samples available for the tumor classes, and we randomly subsample the number of background/brain samples to correspond to the total of the tumor samples for each case.

The results for the RF approach are computed in a leave-1-out manner, where for each case, the RF method was trained on the remaining 19 high-grade cases. For RF training, the background is randomly subsampled by a factor of 0.1 which is very similar to the ones used for the CNN training. Thus, the RF approach has access to almost double the amount of training data compared to the CNN approach, which seems like an advantage.

The results are summarized in Table 1 and Figure 1, and show a promising performance of the CNN-based approach.

4 Discussion and Future Work

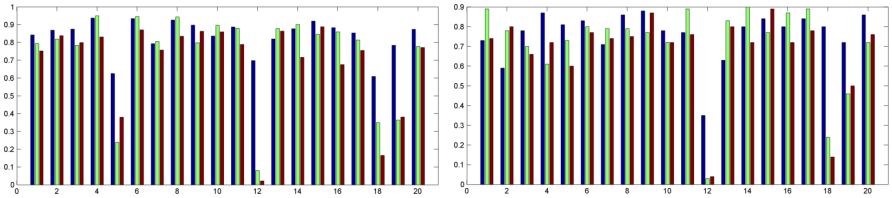
The preliminary results indicate that the unoptimized CNN architecture is already capable of achieving acceptable results. Our work for the final submission will include training on the large BraTS 2014 training corpus, improvements of the network architecture, and parameter tuning.

References

1. Krizhevsky, A., Sutskever, I., Hinton, G.E.: Imagenet classification with deep convolutional neural networks. In: Advances in Neural Information Processing Systems 25. (2012)

Method	Training HG (BraTS 2013)		
	complete	core	enhancing
RF	76.3±12.4	70.9±22.5	67.4±21.7
CNN	83.7±9.4	73.6±25.6	69.0±24.9

Table 1: Quantitative summary of results on the high-grade training data from the BraTS 2013 challenge (10 cases). The results for CNN are obtained by a 2-fold data split for training and testing. The results for RF are obtained with a leave-one-out experiment. This means that for CNN each prediction is based on a classifier trained on 10 cases, while for RF, each classifier is trained on 19 cases, i.e. nearly the double the amount of data.



(a) RF (leave-1-out) (b) CNN (2-fold)

Figure 1: Visualization of the results on the training data from BraTS 2013, and relation to results of a randomized forest from [7]. Results are shown for the complete tumor (blue), core tumor (green) and enhancing tumor (red).

2. Jia, Y.: Caffe: An open source convolutional architecture for fast feature embedding. `http://caffe.berkeleyvision.org/` (2013)
3. Farabet, C., Couprie, C., Najman, L., LeCun, Y.: Learning hierarchical features for scene labeling. Pattern Analysis and Machine Intelligence, IEEE Transactions on **35**(8) (2013) 1915–1929
4. Girshick, R., Donahue, J., Darrell, T., Malik, J.: Rich feature hierarchies for accurate object detection and semantic segmentation. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). (2014)
5. Ciresan, D.C., Giusti, A., Gambardella, L.M., Schmidhuber, J.: Mitosis detection in breast cancer histology images with deep neural networks. In: MICCAI. Volume 2. (2013) 411–418
6. Tustison, N., Avants, B., Cook, P., Zheng, Y., Egan, A., Yushkevich, P., Gee, J.: N4ITK: Improved N3 Bias Correction. Medical Imaging, IEEE Transactions on (2010)
7. Zikic, D., Glocker, B., Konukoglu, E., Shotton, J., Criminisi, A., Ye, D., Demiralp, C., Thomas, O.M., Das, T., Jena, R., Price, S.J.: Context-sensitive classification forests for segmentation of brain tumor tissues. In: MICCAI 2012 Challenge on Multimodal Brain Tumor Segmentation (BraTS). (2012)