Multi-Scale 3D Convolutional Neural Networks for Lesion Segmentation in Brain MRI

Konstantinos Kamnitsas*, Liang Chen, Christian Ledig, Daniel Rueckert, and Ben Glocker

Biomedical Image Analysis Group, Imperial College London, UK

Abstract. We present our 11-layers deep, double-pathway, 3D Convolutional Neural Network, developed for the segmentation of brain lesions. The developed system segments pathology voxel-wise after processing a corresponding multi-modal 3D patch at multiple scales. We demonstrate that it is possible to train such a deep and wide 3D CNN on a small dataset of 28 cases. Our network yields promising results on the task of segmenting ischemic stroke lesions, accomplishing a mean Dice of 64% (66% after postprocessing) on the ISLES 2015 training dataset, ranking among the top entries. Regardless its size, our network is capable of processing a 3D brain volume in 3 minutes, making it applicable to the automated analysis of larger study cohorts.

1 Introduction

The blockage of an artery during a stroke may disrupt the supply of oxygen and other required substances to brain regions, leading to neuronal death. If the blood flow is restored quickly enough, parts of the affected brain tissue may survive and gradually recover [1]. Automatic detection of the affected but salvageable penumbral tissue can accelerate the decision making and treatment of the patient in the acute clinical setting, increasing the likelihood of a more favourable outcome. Development of robust and accurate segmentation techniques could also facilitate the longitudinal monitoring and analysis of stroke lesions which evolve over time [2] and enable larger-scale studies that can further our understanding of the relations between tissue damage and functional deficits.

Following their success on challenging tasks in the field of Computer Vision [3,4,5], Convolutional Neural Networks (CNNs) have been subsequently applied successfully on a variety of biomedical segmentation problems. Most developed approaches relied on the adaptation of 2D CNNs for processing 3D volumes [6,7,8], with difficulties being reported when training of 3D CNNs was attempted. While these architectures of 2D networks might be successful in some problems, they are suboptimal in their use of available 3D information. The first pure 3D CNN reviewed was presented in [9], where it was employed for brain tumour

* The first author was funded by an Imperial College Phd Scholarship.
Email correspondence to: konstantinos.kamnitsas12@imperial.ac.uk

segmentation. Regardless of its small size, the system demonstrated excellent performance and formed the starting point for our approach.

2 Method

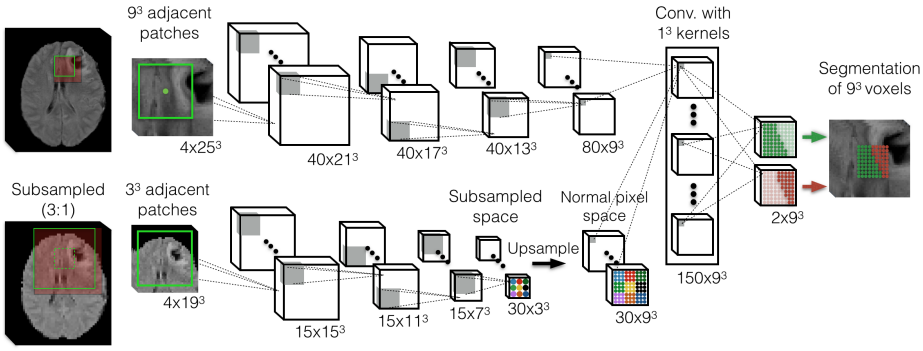


Fig. 1: The architecture of one of our earlier networks. Number and dimensions of the feature maps (FMs) are depicted in the format $(Number \times Dimension)$. Please consult the text for more details on our latest 11-layer network. Multi-modal input is not depicted, in order to avoid cluttering the figure.

2.1 Network Architecture

Although inherently classifiers, CNNs can tackle segmentation tasks by casting them to voxel-wise classification. The network processes a 3D patch around each voxel of an image. It is trained to predict whether the central voxel is pathology or normal brain tissue, depending on the content of the surrounding 3D patch. During training, the parameters of the kernels are optimized using gradient descent, with the target of minimizing the error between the predictions and the true labels.

One of the limitations in the above setting is that the segmentation of each voxel is performed solely by processing the contents of a small patch around it. It is intuitive that greater context is likely to lead to better results. However, a straight-forward increase in the size of the 3D input patch would prohibitively increase the memory requirement and computational burden. Our proposed solution is to perform parallel processing of the image at multiple scales. Our network architecture consists of two parallel convolutional pathways, where both have receptive fields of the same size. The input to the second pathway, however, is a patch extracted from a subsampled-version of an image, thus allowing it to efficiently process greater area around each voxel. This architectural design is presented in figure 1.

Another important feature of our architecture is its full convolutional nature, which allows its efficient application on larger parts of the image. By providing as input segments of an image larger than the receptive field of the final layer's neurons, the network can efficiently process the larger input and provide as output

predictions for multiple neighbouring voxels. Following [9,5], we also exploit this feature during training, constructing our training batches by extracting image segments of size larger than the network's receptive field.

An earlier version of our system is depicted in figure 1. Our final network resulted from the replacement of each convolutional layer with a kernel of size 5^3 with two layers with 3^3 kernels and the addition of another layer before the final classification. The final 11-layers deep network exhibited significantly more accurate segmentation performance. The network is regularised using Dropout [10] at the 9^{th} and 10^{th} layers with a rate of 50%, on top of L1 (10^{-6}) and L2 (10^{-4}) regularisation. Initial learning rate was set at 0.01 and was gradually reduced during training, along with constant momentum equal to 0.6. ReLu activation functions [11] and batch-normalisation [4] were used for the acceleration of convergence that they provide. The training time required for convergence of the final system is roughly one day using a NVIDIA GTX Titan Black GPU. Segmentation of a 3D scan of a brain with four modalities requires 3 minutes.

2.2 Data Preprocessing, Augmentation and Postprocessing

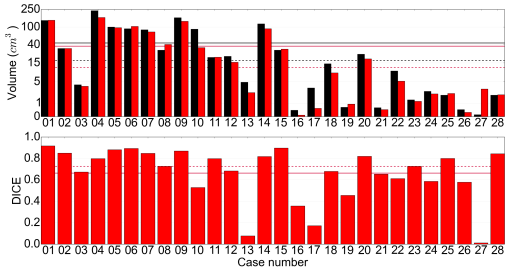
The modalities of each patient were individually normalised to have zero mean and unary standard deviation, as preliminary experiments showed the networks to behave better on input in this intensity range. In order to regularise the network, we augment the dataset by reflecting the images with respect to the sagittal axes. This processing, along with the subsampling of the images, is performed in parallel with training, when the image segments for the next training iteration are extracted, thus effectively adds no computational time. The output of the network was postprocessed by our version of the CRF presented in [12], extended in order to be able to process 3D biomedical images.

3 Evaluation

The system was evaluated on the training dataset of the ISLES 2015 Challenge. Four modalities (Flair, DWI, T1, T2) were available for 28 cases of patients with ischemic stroke lesions. We performed 5-fold validation for this evaluation. Our system achieved a mean Dice coefficient of 66%, with our network alone achieving 64% and an additional increase of 2% achieved through post-processing using the CRF. In figure 2 we present our results for each case and the current results on the evaluation platform, where we rank among the top entries.

4 Conclusion and Discussion

Our architecture exhibits promising performance, with capabilities for delicate segmentations. Difficulties are observed in the segmentation of lesions of particularly small size. The separation of lesions into different categories, for instance according to their size, and their treatment by separate classifiers could simplify the task for each learner and help alleviating the problem.



(a)

User	ASSD average	Dice average	Hausdorff Distance average
dutif1	8.92	0.69	31.75
halmh1	6.77	0.63	36.16
jessa1	11.59	0.45	39.23
kamnk1	5.00	0.66	55.93
mahmq2	10.30	0.54	82.78
maieo1	12.36	0.36	56.94
muscj1	56.77	0.48	76.88
pinta1	12.18	0.50	43.21
robbd1	9.36	0.57	53.88

(b)

Fig. 2: (a) Top: Volume of lesions for each case according to manual (black) and our system's segmentation (red). Bottom: DICE coefficient for our system's segmentation for each case, along with mean (continuous) and median (dashed) values. (b) Online evaluation on training data, with our system ranking 1st and 2nd with respect to mean ASSD and Dice metrics at the time of writing.

References

1. Rekik, I., Allassonnière, S., Carpenter, T.K., Wardlaw, J.M.: Medical image analysis methods in MR/CT-imaged acute-subacute ischemic stroke lesion: Segmentation, prediction and insights into dynamic evolution simulation models. A critical appraisal. NeuroImage. Clinical **1**(1) (January 2012) 164–78
2. Mitra, J., Bourgeat, P., Fripp, J., Ghose, S., Rose, S., Salvado, O., Connelly, A., Campbell, B., Palmer, S., Sharma, G., Christensen, S., Carey, L.: Lesion segmentation from multimodal MRI using random forest following ischemic stroke. NeuroImage **98** (September 2014) 324–35
3. Krizhevsky, A., Sutskever, I., Hinton, G.: Imagenet classification with deep convolutional neural networks. Advances in neural ... (2012) 1–9
4. Szegedy, C., Ioffe, S.: Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift. arXiv:1502.03167 (February 2015)
5. Long, J., Shelhamer, E., Darrell, T.: Fully convolutional networks for semantic segmentation. arXiv preprint arXiv:1411.4038 (2014)
6. Ciresan, D., Giusti, A.: Deep neural networks segment neuronal membranes in electron microscopy images. Advances in neural ... (2012) 1–9
7. Prasoon, A., Petersen, K., Igel, C., Lauze, F.: Deep feature learning for knee cartilage segmentation using a triplanar convolutional neural network. ... Image Computing and ... (09) (2013) 246–253
8. Roth, H., Lu, L., Seff, A., Cherry, K.: A new 2.5 D representation for lymph node detection using random sets of deep convolutional neural network observations. In: ... Image Computing and .... (2014)
9. Urban, G., Bendszus, M., Hamprecht, F., Kleesiek, J.: Multi-modal brain tumor segmentation using deep convolutional neural networks. in proc of BRATS-MICCAI (2014)
10. Hinton, G., Srivastava, N.: Improving neural networks by preventing co-adaptation of feature detectors. arXiv preprint arXiv:1207.0580 (2012) 1–18
11. Nair, V., Hinton, G.: Rectified linear units improve restricted boltzmann machines. Proceedings of the 27th International ... (3) (2010)
12. Krähenbühl, P., Koltun, V.: Efficient inference in fully connected crfs with gaussian edge potentials. arXiv:1210.5644 (2012) 1–9