Fully Convolutional Network with Hypercolumn Features for Brain Lesion Segmentation

Mobarakol Islam[1] and Hongliang Ren[2]

[1] Department of NUS Graduate School for Integrative Sciences and Engineering (NGS), National University of Singapore, Singapore
mobarakol@u.nus.edu
[2] Department of Biomedical Engineering, National University of Singapore, Singapore
ren@u.nus.sg

Abstract. The segmentation of stroke lesion is very necessary for diagnosis, planning treatment strategies and monitoring disease progression. We propose a fully convolutional network (FCN) with hypercolumns features and sparse pixel predictions (e.g. PixelNet) for automatic brain lesion segmentation. PixelNet extracts feature from multiple layers that correspond to the same pixel and samples a modest number of pixels across a small number of images for each SGD (Stochastic gradient descent) batch update. Deep Learning (DL) models like Convolutional Neural network (CNN) requires large training data to generalize the model where most of the biomedical problems have small available dataset. Moreover, the problem of label imbalance leads the CNN often converge to the certain labels. PixelNet deals these problems by utilizing sparse pixel prediction on a modest number of pixels. We utilize PixelNet in ISLES (Ischemic Stroke Lesion Segmentation) challenge 2017 and achieves 68% Dice accuracy as preliminary result.

1 Copyright Statement

2

3

4

5

6

7

8

9

10

11

12 Introduction

It is very important to segment the gliomas and its intra-tumoral structures as well as estimate relative volume to monitor the progression, assessment, treatment planning and follow-up studies. Generally, the segmentation of gliomas observes in various regions such as active tumorous tissue, necrotic tissue, and the peritumoral edematous which defined through intensity changes relative to the surrounding normal tissue. However, gliomas or glioblastomas are usually spread out, poorly contrasted and intensity information being disseminated across various modalities that make them difficult to segment [1]. The tumor intensity also differs across the patients like HGG patients the tumor consists of enhancing, non-enhancing and necrotic parts, while in the LGG patients it is not necessarily to include an enhancing part [2]. Due to inconsistency and diversity of MRI acquisition parameters [3] and hardware variations, there are large difference in appearance, shape and intensity ranges among the same sequences and acquisition scanners [5], which make the segmentation more challenging. Thus, physicians conventionally use rough evaluation or manual segmentation; however, manual

segmentation is time-consuming and laborious task that is inclined to misinterpretation and observer bias [4].

In recent years, Deep Learning (DL) have drawn increasing attention medical applications such as in object detection [6, 7], semantic segmentation [8] and classification [9]. DL models like convolutional neural networks (CNN) are capable of learning high level and task adaptive hierarchical features from training data and take part as an effective approach. Havae et al. [10] build a CNN based two-pathway cascade network which performs a two-phase training using both local and global contextual features and tackle difficulties related to the imbalance of tumor labels in data. Another similar approach DeepMedic [11] uses two convolutional parallel pathways and 3D CNN architecture with 11-layers for brain lesion segmentation. Later, modified version of DeepMedic with residual connection utilize for brain tumor segmentation [12]. On the other hand, Pandian et al. [13] and Casamitjana et al. [14] use 3D volumetric CNN to train sub- volume of multi-modal MRIs and show that 3D CNN performs well for segmentation as MRI acquires 3D information. The benefit of these architectures is that they performed well with a comparatively smaller dataset. However, they are computationally expensive as it needs 3D kernels and a large number of trainable parameters. Alex et al. [15] uses 5 layers deep Stacked Denoising Auto-Encoder (SDAE) and Randhawa et al [16] uses 8 layers CNN and Pereira et al. [17] uses deeper CNN architecture with small kernel for segmenting gliomas from MRI.

In spatially-invariant label prediction problem like semantic segmentation, every separate label per pixel predicts using a convolutional architecture. As a result, gradient based learning like Stochastic gradient descent (SGD) treats training data as sampled independently and form an identical distribution [18]. Hyvärinen et al. [19] demonstrate that pixel in a given image is highly correlated and neighbouring pixels are not independent. To capture the high-level global context and minimize the loss of the contextual information in higher convolutional layers, there have built many predictors based on multiscale feature extraction from multiple layers of a CNN [20]. Hariharan et al. [21] extracted features of the same pixels from multiple layers and accumulate in a feature vector called “Hypercolumns”. To extract feature, FCNs [22] efficiently implemented linear prediction in a coarse to fine manner. To reduce memory footprint DeepLab [23] incorporate filter dilation and linear-weighted fusion in fully connected layers. ParseNet [24] averages the pooling feature by normalization and concatenation to add spatial context for a layer response. PixelNet [25] adopt both Hariharan et al. [21] and ParseNet [24] to build hypercolumn and concatenate spatial context in the layer where the tradeoff between statistical and computational efficiency for convolutional learning. PixelNet shows state of art performance for in BRATS 2017 [27,28,29,30] training and validation dataset.

13 Methodology

PixelNet extract multi-scale convolution and feature and concatenate them as hypercolumn to ensure all local and global contextual information in the learning phase (Fig. 1). A hyper descriptor can be written as:

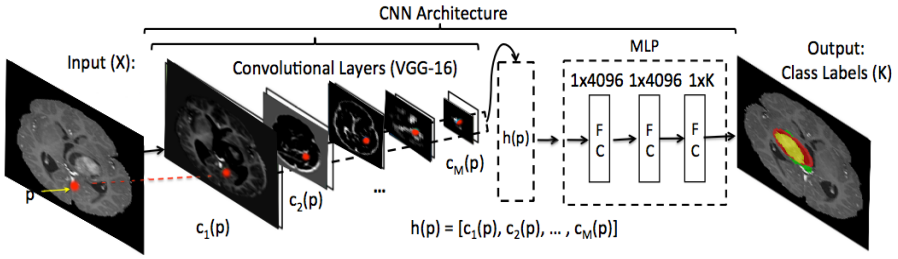

Fig. 1. Three modalitities (Flair, T1, T1c) input in a CNN and extract hypercolumn descriptor for a sample pixel from multiple convolutional layers. The hypercolumn descriptor is then fed to a multi-layer perception for non-linear optimization.

$$h_p = [c_1(p), c_2(p), ..., c_M(p)] \tag{1}$$

where h_p denote the multiscale hypercolumn features for the pixel p, and $c_i(p)$ denote the feature vector from layer i. PixelNet considers pixel wise prediction as operating over hypercolumn features. For example, the final prediction for pixel p,

$$f_{\theta,p}(X) = g(h_p(X)) \tag{2}$$

where θ represent both hypercolumn features h and pixel wise predictor g. θ updates by using SGD training. We use a series of fully connected layers followed by ReLU activation function similar to VGG-16 [26] to implement non-linear predictor. We adopt sparse pixel prediction at training time for efficient mini-batch generation. In sparse prediction, hypercolumn features h_p choose from dense convolutional responses at all layers by computing the 4 discrete locations in the feature map c_i (for i^{th} layer) closest to sampled pixel $p \in P$ and finally apply bilinear interpolation to get i^{th} layer response in hypercolumn.

14 Experiment

14.1 Dataset

BRATS 2017 (Brain Tumor Image Segmentation Benchmark) [27, 28, 29, 30] training database consists in total 285 cases of patients. It is a multi-modal MRI scans of 210 high-grade glioma (HGG) and 75 low-grade glioma (LGG) and 4 different modalities including T1 (spin- lattice relaxation), T1c (T1-contrasted), T2 (spin-spin relaxation) and FLAIR (fluid attenuation inversion recovery). Each scan is a continuous 3D volume of 155 2D slices of size 240x240. The volume of the various modalities is already skull-stripped, aligned with T1c and interpolated to 1 mm voxel resolution. The provided ground truth with manual segmentation includes three labels: GD-enhancing tumor (ET — label 4), the peritumoral edema (ED — label 2), and the necrotic and non-enhancing

tumor (NCR/NET — label 1). The predicted labels are evaluated by merging three regions: whole tumor (WT: all four labels), tumor core (TC: 1,2) and enhancing tumor (ET: 4).

14.2 Training and Evaluation for Segmentation Task

We use depth slicing images on axial orientation on the mixed HGG and LGG data (285 MRI cases in total). However, this dataset is highly imbalance where ground-truth contains 98% pixel are healthy tissue (label 0) and remaining are also unequal ratios of four ROI labels such as edema, enhancing, necrotic and non-enhancing. We deal this issue by ignoring all the blank slices in groundtruth (both background and healthy issue) and train PixelNet with corresponding 3 modalities such as flair, T1C and T2. So we ignore T1 scan to see the performance of the PixelNet. Though BRATS 2017 has in total 44175 (285x155) slices, we utilize only 18924 (43% data) slices corresponding to ground-truth with non-zero class (contains at least one class 1 or 2 or 4) in our first observation. We use Caffe deep learning platform to perform all of our experiments.

Table 1. Dice and Sensitivity for BRATS 2017 training dataset

Level	Dice			Sensitivity		
	ET	WT	TC	ET	WT	TC
Mean	0.711	0.909	0.866	0.771	0.897	0.831
StdDev	0.293	0.070	0.115	0.231	0.091	0.139
Median	0.830	0.929	0.902	0.849	0.924	0.872
25quantile	0.706	0.896	0.837	0.727	0.872	0.786
75quantile	0.882	0.946	0.930	0.916	0.951	0.916

Table 2. Specificity and Hausdorff95 for BRATS 2017 training dataset

Level	Specificity			Hausdorff95		
	ET	WT	TC	ET	WT	TC
Mean	0.998	0.995	0.998	6.946	7.275	6.103
StdDev	0.002	0.005	0.002	15.362	13.494	11.546
Median	0.999	0.996	0.999	2.000	3.000	3.162
25quantile	0.998	0.994	0.998	1.414	2.236	2.236
75quantile	1.000	0.998	0.999	3.606	4.899	4.583

After PixelNet prediction, we evaluate all the cases for training set (285 cases) and validation (46 cases) using online evaluation portal for BRATS 2017 challenge. Table 1 and 2 represent the training set evaluation results where whole tumor average Dice accuracy is 90% and Hausdorff distance is 7.3 which is quite promising. Table 3 and 4 shows the evaluation results of validation set where average 87% dice accuracy and 9.8 Hausdorff distance. Though enhance tumor and tumor core region have lower accuracy than whole tumor, the individual accuracy can be considered as state of art performance.

Fig. 2 shows some visualized examples of PixelNet prediction with comparing ground-truth.

Table 3. Dice and Sensitivity for BRATS 2017 validation dataset

Level	Dice			Sensitivity		
	ET	WT	TC	ET	WT	TC
Mean	0.689	0.876	0.761	0.720	0.861	0.710
StdDev	0.304	0.086	0.221	0.287	0.136	0.253
Median	0.829	0.902	0.849	0.840	0.906	0.795
25quantile	0.625	0.877	0.708	0.657	0.849	0.562
75quantile	0.881	0.929	0.912	0.883	0.949	0.910

Table 4. Specificity and Hausdorff95 for BRATS 2017 validation dataset

Level	Specificity			Hausdorff95		
	ET	WT	TC	ET	WT	TC
Mean	0.998	0.995	0.998	12.938	9.820	12.361
StdDev	0.002	0.005	0.003	26.453	13.516	20.826
Median	0.999	0.996	0.999	2.449	4.581	5.050
25quantile	0.998	0.993	0.998	1.799	2.828	3.041
75quantile	1.000	0.998	1.000	8.569	8.093	11.176

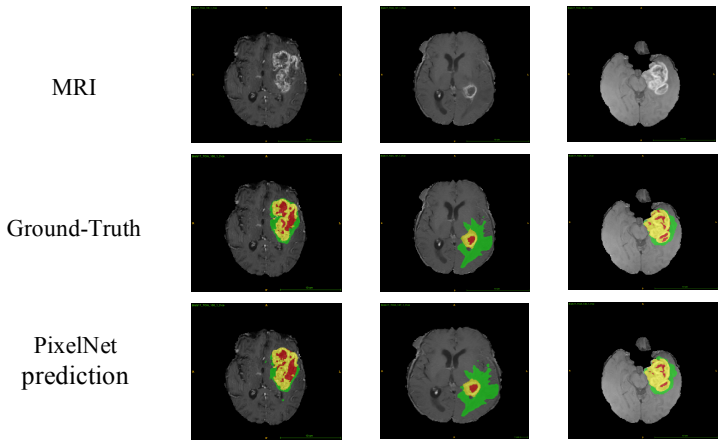

Fig. 2. PixelNet prediction for BRATS 2017 Training Dataset.

Table 5. Dice, Sensitivity and Hausdorff95 for BRATS 2017 validation dataset with reduced training data.

	Dice			Sensitivity			Hausdorff95		
	ET	WT	TC	ET	WT	TC	ET	WT	TC
PixelNet (43% data)	0.689	0.876	0.761	0.720	0.861	0.710	12.938	9.820	12.361
PixelNet (29% data)	0.677	0.861	0.775	0.722	0.818	0.776	14.675	11.808	23.726

To see the small dataset effect to PixelNet we train our model with only 12930 (29% data) slices corresponding to ground-truth consisting all three classes (1, 2, 4) and compare the performance with trained model of all non-zero slices or at least one class slices (43% data: 18924 slices). Table 5 shows the Dice and Hausdorff95 evaluation for validation dataset by varying amount of training data. PixelNet achieves almost same performance for less dataset.

15 Conclusion

We presented an automatic brain tumor segmentation method based on pixel level semantic segmentation. We choose PixelNet which extracts multi layers convolutional feature and form hypercolumn. Hypercolumn contains useful contextual information and use sparse pixel prediction to generate efficient mini-batch which produces promising results for Brain tumor segmentation. Though the preliminary results of the ET and TC are not good as WT however we are still working on this model to achieve better accuracy in all the regions of the tumor.

Acknowledgments: I would like to thank Aayush Bansal, the author of PixelNet [25], for the assistance to implement PixelNet for this project.

References

1. Nyul, L.G., Udupa, J.K., Zhang, X.: New variants of a method of MRI scale standardization. IEEE Transactions on Medical Imaging 19(2),143-50 (2000).
2. Ellwaa, A., et al: Brain Tumor Segmantation using Random Forest trained on iterative selected patients. In proc. of BRATS-MICCAI, (2016).
3. Omuro, A., DeAngelis, L. M.: Glioblastoma and Other Malignant Gliomas: A Clinical Review. Jama, vol. 310, pp. 1842-1850, (2013).
4. Bauer, S., et al: A survey of MRI-based medical image analysis for brain tumor studies. Physics in Medicine and Biology, vol. 58, pp. R97, (2013).
5. Inda, Maria-del-Mar, Bonavia, R., Seoane, J.: Glioblastoma multiforme: A look inside its heterogeneous nature. Cancers 6.1, pp. 226-239 , (2014).
6. Baumgartner, C. F., Kamnitsas, K., Matthew, J., Smith, S., Kainz, B., Rueckert, D.: Real-time standard scan plane detection and localisation in fetal ultrasound using fully convolutional neural networks. In: Med Image Comput Comput Assist Interv. Vol. 9901 of Lect Notes Comput Sci. pp. 203–211, (2016).

7. Azizi, S., Imani, F., Ghavidel, S., Tahmasebi, A., Kwak, J. T., Xu, S., Turkbey, B., Choyke, P., Pinto, P., Wood, B., Mousavi, P., Abolmaesumi, P.: Detection of prostate cancer using temporal sequences of ultrasound data: a large clinical feasibility study. Int J Comput Assist Radiol Surg 11 (6), 947–956, (2016).
8. Ronneberger, O., Fischer, P., Brox, T.: U-net: Convolutional networks for biomedical image segmentation. arXiv:1505.04597v1, (2015).
9. Milletari, F., Navab, N., Ahmadi, S. A.: V-net: Fully convolutional neural networks for volumetric medical image segmentation. arXiv:1606.04797v1, (2016).
10. Havaei, M., Davy, A., Warde-Farley, D.: Brain tumor segmentation with Deep Neural Networks. Medical Image Analysis, vol. 35, pp. 18-31, (2017).
11. Kamnitsas, K., 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).
12. Kamnitsas, K., et al: DeepMedic on Brain Tumor Segmentation. In proc. of BRATS-MICCAI, (2016).
13. Pandian1, B., Boyle1, J., Orringer, D. A.: Multimodal Tumor Segmentation with 3D Volumetric Convolutional Neural Networks. In proc. of BRATS-MICCAI, (2016).
14. Casamitjana, A., et al: 3D Convolutional Networks for Brain Tumor Segmentation. In proc. of BRATS-MICCAI, (2016).
15. Alex, V., Krishnamurthi, G.: Brain Tumor Segmentation from Multi Modal MR images using Stacked Denoising Autoencoders. In proc. of BRATS-MICCAI, (2016).
16. Randhawa, R., Modi, A., Jain, P., Warier, P.: Improving segment boundary classification for Brain Tumor Segmentation and longitudinal disease progression. In proc. of BRATS-MICCAI, (2016).
17. Pereira, S., et al: Brain Tumor Segmentation Using Convolutional Neural Networks in MRI Images. IEEE Transactions on Medical Imaging, vol. 35, pp. 1240-1251, (2016).
18. Bottou, L.: Large-scale machine learning with stochastic gradient descent," in Proceedings of COMPSTAT' 2010, pp. 177–186, Springer, (2010).
19. Hyvärinen, A., Hurri, J., Hoyer, P. O.: Natural Image Statistics: A Probabilistic Approach to Early Computational Vision. (1. Aufl. ed.) 200939. DOI: 10.1007/978-1-84882-491-1, (2009).
20. Denton, E. L., Chintala, S., Fergus, R., et al: Deep generative image models using a laplacian pyramid of adversarial networks. In NIPS, (2015).
21. Hariharan, B., Arbelaez, P., Girshick, R., Malik, J.: Hypercolumns for object segmentation and fine-grained localization. In CVPR, (2015).
22. Long, J., Shelhamer, E., Darrell, T.: Fully convolutional models for semantic segmentation. In CVPR, (2015).
23. Chen, L.-C., Papandreou, G., Kokkinos, I., Murphy, K., Yuille, A. L.: Semantic image segmentation with deep convolutional nets and fully connected CRFs. In ICLR, (2015).
24. Liu, W., Rabinovich, A., Berg, A. C.: Parsenet: Looking wider to see better. arXiv preprint arXiv:1506.04579, (2015).
25. Bansal, A., et al: PixelNet: Representation of the pixels, by the pixels, and for the pixels. arXiv:1702.06506v1, (2017).
26. Simonyan, K., Zisserman, A.: Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556, (2014).
27. Menze, B., et al: The multimodal brain tumor image segmentation benchmark (brats). IEEE Transactions on Medical Imaging 34 (2015).
28. Bakas, S., Akbari, H., Sotiras, A., Bilello, M., Rozycki, M., Kirby J. S., Freymann, J. B., Farahani, K., Davatzikos, C.: Advancing the Cancer Genome Atlas glioma MRI collections

with expert segmentation labels and radiomic features. Nature Scientific Data, (2017) [In Press].
29. Bakas, S., Akbari, H., Sotiras, A., Bilello, M., Rozycki, M., Kirby, J., Freymann, J., Farahani, K., Davatzikos, C.: Segmentation Labels and Radiomic Features for the Pre-operative Scans of the TCGA-GBM collection. The Cancer Imaging Archive, (2017). DOI: 10.7937/K9/TCIA.2017.KLXWJJ1Q.
30. Bakas, S., Akbari, H., Sotiras, A., Bilello, M., Rozycki, M., Kirby, J., Freymann, J., Farahani, K., Davatzikos, C.: Segmentation Labels and Radiomic Features for the Pre-operative Scans of the TCGA-LGG collection. The Cancer Imaging Archive, (2017). DOI: 10.7937/K9/TCIA.2017.GJQ7R0EF.