Deep Learning and Image Segmentation of 3D MRI of Multiple Brain Metastases for Stereotactic Radiosurgery

Denise S. Davis, MS., University of South Carolina.

UNIVERSITY OF South Carolina

Introduction

Brain metastases are complication associated with cancer. Stereotactic Radiosurgery, SRS, is a specialized radiation treatment modality that focuses on small treatment volumes to spare damage to normal tissue. Patients are simulated, planned and treated in the same day. Manually delineating multiple tumor volumes within 3D MRI images is time consuming and an arduous task performed by specialists that include radiologists, neurologists and radiation oncologists. Automatic contouring, the radiation oncology term, is synonymous with instance segmentation, the computer vision term. In this survey, we have reviewed and compared six different Deep Learning solutions for solving automatic image segmentation of multiple brain metastases found in 3D MRI images. At the time of this paper, no one technique had been adopted for SRS and appears to be evolving with technological advances in computer vision. These deep learning neural networks ranged in depth from 11 layers to 101 layers. We compared reported sensitivity and false positive rates ranged from 0.77 to 0.98 and reported dice similarity coefficients ranged from 0.70 to 0.93. Furthermore, researchers agreed that brain metastases smaller than 0.40ml were a challenge to consistently detect and delineate automatically. Additionally, each study independently built training data from retrospective patient chart review, as there is no benchmark datasets available for this highly specialized radiation therapy field. As such, we also compared the size of the brain metastases in these studies and found them to be different as they are unique to patients sampled. Data augmentation plays an important role in the training process for this treatment site. Finally, a comparative analysis of our findings and a proposal for further research of instance segmentation of brain metastases have been presented.

Literature Researched

[1] Charron, O., Lallement, A., Jarnet, D., Noblet, V., Clavier, J.B.and Meyer, P., 2018. Automatic detection and segmentation of brain metastases on multimodal MR images with a deep convolutional neural network. *Computers in biology and medicine*, *95*, pp.43-54.

[2] Liu Y, Stojadinovic S, Hrycushko B, Wardak Z, Lau S, Lu W, et al. (2017) A deep convolutional neural network based automatic delineation strategy for multiple brain metastases stereotactic radiosurgery.PLoSONE2(10):e0185844. https://doi.org/10.1371/journal.pone.0185844.

[3] Dikici, Engin, John L. Ryu, Mutlu Demirer, Matthew Bigelow, Richard D. White, Wayne Slone, Barbaros Selnur Erdal, and Luciano M. Prevedello. "Automated Brain Metastases Detection Framework for T1-Weighted Contrast-Enhanced 3D MRI." arXiv preprint arXiv:1908.04701 (2019).

[4] Bousabarah K, Ruge M, Brand JS, et al. Deep convolutional neural networks for automated segmentation of brain metastases trained on clinical data. Radiat Oncol. 2020;15(1):87. Published 2020 Apr 20. doi:10.1186/s13014-020-01514-6.

[5] Grøvik, E., Yi, D., Iv, M., Tong, E., Rubin, D. and Zaharchuk, G., 2020. Deep learning enables automatic detection and segmentation of brain metastases on multisequence MRI. *Journal of Magnetic Resonance Imaging*, *51*(1), pp.175-182.

[6] Lei, Yang, Zhen Tian, Shannon Kahn, Walter J. Curran, Tian Liu, and Xiaofeng Yang. "Automatic detection of brain metastases using 3D mask R-CNN for stereotactic radiosurgery." In *Medical Imaging 2020: Computer-Aided Diagnosis*, vol. 11314, p. 113142X. International Society for Optics and Photonics, 2020.

[7] He, Kaiming, Georgia Gkioxari, Piotr Dollár, and Ross Girshick. "Mask r-cnn." In *Proceedings of the IEEE international conference on computer vision*, pp. 2961-2969. 2017.

Methods

The dice similarity coefficient, DSC, is used for geometric analysis, and sensitivity and false positive rates, FPR, are used for statistical analysis of classification function. DSC is a measure of the similarity between two datasets, in this case, predicted mask and ground truth contours. Sensitivity measures probability that a positive prediction represents a positive ground truth value. FPR measures the number of false positives compared to the total number of negatives.

Eq. 1 $DSC = \frac{2TP}{2TP+FP+FN}$

Eq. 2 $Sensitivity = \frac{TP}{TP+FN}$

Eq. 3 $FPR = \frac{FP}{N} = \frac{FP}{FP+TN}$

$TP = \#\ of\ true\ positives, FP = \#\ of\ false\ positives, FN = \#\ of\ false\ negatives,\ N = total\ number\ of\ negatives$

Discussion

- The percentage of the images that contain positive instances is relatively small. This means in 28 slices of 256 x 256 MRI images, the number of positive pixels or voxels is less than 25 percent of each image.
- For multiple lesions found in the same image(slice or plane) there is no correlation between the spatial location of each instance.
- One researcher [5] reported a decrease in performance of the DNN with increasing number of lesions in the same slice(2.5D) or plane(3D).
- The workflow order impacts the speed of training, segment then classify or classify then segment. Which approach is optimal for speed and performance of the DNN?
- Details for statistics reported in each paper were not consistent.
- It was not clear the true assessment of instance segmentation tasks. More often than not, the results were reported referring to pixel classification versus contouring as a whole.
- In all of the six papers, the DNN's implemented detection using bounding boxes. The cost, optimization and loss functions varied with each DNN.

Imaging Datasets

- There is a need for MRI image benchmark dataset of brain metastases available for commissioning or accepting a SRS image segmentation workflow.
- Each researcher provided their own MRI images from their own clinical service, except in the case of [3] who also included BRATS 2015 imaging dataset.

	[1]	[2]	[3]	[4]	[5]	[6]
1.5T MRI Unit	✓					
3.0T MRI Unit		✓	✓	✓	✓	✓
2D/3D T1-weighted T1c	✓	✓	✓	✓	✓	✓
T2-weighted FLAIR			✓	✓	✓	
stripped skull from image	✓	✓				
data augmentation		✓	✓	✓	✓	✓
contouring by:						
radiation oncologist	✓			✓		
neurosurgeon		✓		✓		
radiologist/neuro radiologist			✓		✓	
unspecified						✓

DNN Performance

- Sensitivity and False Positive Rates were presented for [1,3,4,5] to indicate classification /detection accuracies accompanied by details of specific problems associated with detecting small BM's.
- [1,5] both reported higher sensitivity, 0.98, and lower AFPR, 0.04. Dice Similarity Coefficients, DSC, were reported in [1,2,4.5].
- Although researchers imply that DSC value is dependent upon BM size, we did not find a complete study for BM size versus DSC value. Further investigation is needed to confirm the usefulness to optimize instance segmentation.
-
- [2] reported the best DSC 0.80-0.93. [3,4,5] all reported the importance of stride value and max pooling size with regards to the possibility of misclassification of voxels without adequate overlap between segments.
- [6] approach with MASK R-CNN is influenced by segmentation prediction, voxel classification, and bounding box accuracy. Although [6] dataset was small, their overall sensitivity aligned within the median results found in the other studies.

Results

STUDY	Segmentation Technique	DNN Activation function	CNN	Training Time	Image Segmenting Time	Dataset #BM (#patients)	Train/Test Sets	Median BM volume	Sensitivity	DSC	AFPR	Output
Charron et al.[1]	DeepMedic	Sigmoid	2 paths 11 layers	30 hours	20 mins	374(164)	374/38	0.1-37ml	0.93-0.98	0.77-0.79	0.04-.144	Detection only
Liu et. al[2]	En-DeepMedic	PReLU	3 paths 11 layers	2 days per fold	2 mins	N/A (490)	5-fold CV	0.129-6.918ml	N/A	0.80-0.93	N/A	Contours DicomRT
Dikici et al[3]	CropNet	ReLU	Variable	3.5 hours per fold	30.6 sec	932(158)	5-fold CV	<0.16ml	0.90	N/A	.0912	Detection only
Bousabarah[4]	Ensemble U-Net	Leaky ReLU	23 layers	N/A	4-5 mins	1223(509)	(469)/(40)	0.01-10ml (524 lesions <0.4ml)	0.77-0.82	0.70-0.74	0.08-0.35	Contours
Grovik et al[5]	2.5 GoogLeNet	Sigmoid	22 layers	15 hours	1 minute	N/A(156)	N/A(100)/856(51)	N/A	0.92-0.98	0.79+-0.12	0.034-0.083	Contours
Lei et al[6]	Mask R-CNN ResNet101 & RPN	ReLU	101 layers	N/A	N/A	N/A(20)	RPN 5-fold CV	N/A	0.865±0.032	N/A	N/A	Contours

DSC-Dice Similarity Coefficient; AFPR-Average False Positive Rate; [5]Test set was not selected randomly. Selection based upon #lesions per patient and then separated into 4 groups

Conclusion

We have identified three areas to investigate next that may help improve automatic delineation of the smaller tumors.

1. Deeper DNN's are employed to achieve higher resolution on the MRI image. What advances in computer vision can be leveraged to achieve higher resolution in 3D MRI images similar to the advantages of residual blocking in ResNet.
2. Mask R-CNN was implemented with ResNet101 on the backend[6]. Can RoIAlign be beneficial for DNN with less layers? MASK R-CNN was a fix implemented by [7].
3. One researcher reported that the probability a segment is positive, contains metastatic mass, is smaller than 1:30.
4. The datasets in this paper were considerably small. We should request access to the larger datasets from [3,4,5] to conduct independent comparisons with Mask R-CNN[7] and any future proposal we investigate.