CS482/682 Final Project Report Group 17

Unsupervised Learning in Medical Image Segmentation

Maxine Gong, xgong7; Ting He, the14; Muchen Li, mli166; Victor Wen, vwen2

May 4 2022

1 Introduction

Background Segmentation of lesions is a task of separating lesions from other objects for medical-based imaging. This task requires a high-quality and large annotated training dataset, which normally requires medical experts to manually label them. However, this supervised method not only requires a large amount of human effort but also suffers from less generalizability to apply in atypical or unseen situations than unsupervised methods (1). Hence, an unsupervised approach can greatly help the segmentation task with limited labels.

To achieve better unsupervised segmentation performance, reinforcement learning (RL) can be introduced to enable an agent to learn the optimal behavior through maximizing reward. Among RL algorithms, Deep Q-learning (DQN) has demonstrated improved performance on radiological images and outperformed the supervised approach with a very small training set on image classification and segmentation problems (2).

Related Work Mask-Regional-CNN and U-net are the two most common CNN architectures used in medical segmentation problems. Among the top performers in the BraTS challenge, DeepMedic (3) is a two 11-layer 3D-CNN with residual connections to improve the performance. It has achieved a dice score of 0.89 for whole tumor segmentation tasks. Another 3D-Unet model (4) generated similar dice scores for whole tumor segmentation and tumor core segmentation tasks.

2 Methods

Dataset Our dataset comes from BraTS brain tumor segmentation 2016 and 2017 challenges. We used 60 images for training, 20 for validation, and 20 for testing. Our team implemented 2 different unsupervised methods to compare the results to a supervised approach.

Setup, Training and Evaluation The goal is to find the approach that achieves the best performance for medical image segmentation with limited labels.

We used a supervised 3D-Unet as a baseline model to compare performance with two unsupervised approaches, 3D-Unet-KMeans method and 3D-Clustering-CNN-DQN. Our hypothesis is that the supervised 3D-Unet would have the best performance overall, while unsupervised 3D-Clustering-CNN-DQN outperforms unsupervised 3D-Unet-KMeans.

For the unsupervised 3D-Unet KMeans, KMeans clustering was applied on the outputs of 3D-Unet to generate pseudo-labels, which are used to compute a dice loss for backpropagation to update weights. For the 3D-Clustering-CNN-DQN, the pseudo-labels were generated by unsupervised 3D-Clustering-CNN and superpixel, where the CNN has 3 convolutional layers with 3x3x3 kernels, 1x1x1 padding, and 1x1x1 stride, and applied batch normalization after each layer with the final layer having output channel of 6. We selected the 60 masks as input for a DQN to predict the best cluster to serve as a lesion mask for the remaining test dataset.

The RL DQN-agent chooses the best mask from those 6 masks generated from the 3D-clustering-

CNN. The agent is trained by sampling from the experience buffer, calculating the mean-square error between optimal Q-value and the current Q-value and backpropating the gradients. We evaluated the performance of the models by 1) segmentation accuracy: testing dice scores, 2) efficiency: total running time.

3 Results

Supervised 3D-Unet The model was trained for 50 epochs using the Adam optimizer and dice loss with learning rate=1. The model had the best performance in terms of segmentation accuracy and efficiency out of all experimenting models (Table 1). The model successfully segmented the tumor from the rest of brain tissue and was the closest to the ground truth (Figure 1).

Unsupervised KMeans 3D-Unet The model was trained for 50 epochs using the Adam optimizer and dice loss with learning rate=0.00005. The training and validation dice loss was the smallest across all models. Yet, the test dice score was the worst. Visualization of the pseudo-labels suggests that the model learned to segment the whole brain from the background, but could not detect the tumor in the brain.

Unsupervised 3D-clustering-CNN and Reinforcement Learning The superpixel uses nearest-neighbors with features of pixel intensity to generate cluster regions of images that are similar in color and close in distance. (3) We trained our 3D-clustering-CNN for 10 epochs using stochastic gradient descent, cross-entropy loss, and learning-rate=0.1. Pseudolabels were generated as masks based on the superpixels. For the RL DQN was trained using the SGD optimizer (learning rate=0.001) and an experience replay buffer with buffer size=1000.

	3DUnet	3DUnet Kmeans	3DCNN-RL
Accuracy	0.65	0.10	0.73
Efficiency	24.89s	51.67s	85.00s

Table 1: Modeling Accuracy and Efficiency

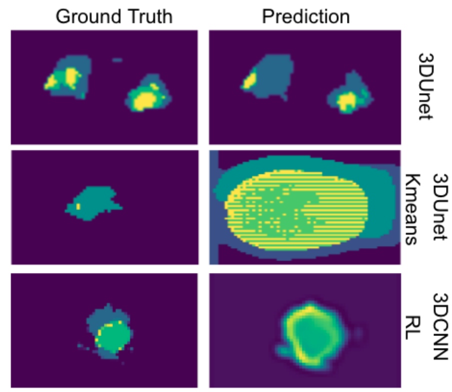

Figure 1: *Model Segmentation Results*

4 Discussion

Discussion The results show that the unsupervised 3D-clustering-CNN-RL outperformed the supervised 3D-Unet in terms of segmentation accuracy by a small margin, and achieved much higher accuracy than the unsupervised 3D-Unet KMeans. We proposed that as RL involves user-selected masks, it is guided towards choosing the mask closest to the ground truth through rewards and punishments. The pseudo-labels generated by this method also learned superpixel-wise rather than pixel-wise. In addition, supervised method often suffers from overfitting problem, which potentially explains why 3D-clustering-CNN-RL did better than supervised U-net. Moreover, there is concern of biased dataset as we observed the first few training samples have much lower segmentation accuracy than the other training samples and testing samples. On the other hand, since the pseudo-labels generated by KMeans clustering are significantly influenced by the features of the raw input images, the model might not be optimized towards ground truths, which is the case we observed. During training, the unsupervised 3D-Unet KMeans had extremely low dice loss, suggesting that the model is optimized. Yet, it did not have a high dice score with respect to the ground truths, as the model segmented the whole brain from the background, which were more distinguishable by KMeans than the tumor in the brain.

5 References

1. Zeng, C., Gu, L., Liu, Z., Zhao, S. (2020, November). Review of Deep Learning Approaches for the segmentation of multiple sclerosis lesions on Brain MRI. Frontiers. DOI: 10.3389/fninf.2020.610967

2. Stember, J., Shalu H. (2020, December). Unsupervised deep clustering and reinforcement learning can accurately segment MRI brain tumors with very small training sets. https://doi.org/10.48550/arXiv.2012.13321

3. Kamnitsas, K., Ferrante, E., Parisot, S., Ledig, C., Nori, A. V., Criminisi, A., ... Glocker, B. (2016, October). DeepMedic for brain tumor segmentation. In International workshop on Brainlesion: Glioma, multiple sclerosis, stroke and traumatic brain injuries (pp. 138-149). Springer, Cham.

4. Iftekharuddin, K. M., Jia, W., Marsh, R. (2000, July). A fractal analysis approach to identification of tumor in brain MR images. In Proceedings of the 22nd Annual International Conference of the IEEE Engineering in Medicine and Biology Society (Cat. No. 00CH37143) (Vol. 4, pp. 3064-3066). IEEE.

5. Ghaffari, M., Sowmya, A., Oliver, R. (2019). Automated brain tumor segmentation using multimodal brain scans: a survey based on models submitted to the BraTS 2012–2018 challenges. IEEE reviews in biomedical engineering, 13, 156-168.