Active Learning Technique for Multimodal Brain Tumor Segmentation using Limited Labeled Images

Dhruv Sharma[1,2], Zahil Shanis[1], Chandan K. Reddy[2], Samuel Gerber[1], and Andinet Enquobahrie[1]

[1] Kitware, Inc. Carrboro, NC 27510, USA
[2] Virginia Polytechnic Institute and State University, Blacksburg, VA 24060, USA

Abstract. Image segmentation is an essential step in biomedical image analysis. In recent years, deep learning models have achieved significant success in segmentation. However, deep learning requires the availability of large annotated data to train these models, which can be challenging in biomedical imaging domain. In this paper, we aim to accomplish biomedical image segmentation with limited labeled data using active learning. We present a deep active learning framework that selects additional data points to be annotated by combining U-Net with an efficient and effective query strategy to capture the most uncertain and representative points. This algorithm decouples the representative part by first finding the core points in the unlabeled pool and then selecting the most uncertain points from the reduced pool, which are different from the labeled pool. In our experiment, only 13% of the dataset was required with active learning to outperform the model trained on the entire 2018 MICCAI Brain Tumor Segmentation (BraTS) dataset. Thus, active learning reduced the amount of labeled data required for image segmentation without a significant loss in the accuracy.

Keywords: Deep Learning, Active Learning, Segmentation

1 Introduction

Image segmentation is a critical task in computer vision as it helps us understand the image at a semantic level. While there have been various algorithms devised to carry out semantic segmentation[1], deep learning models have become the main choice because of their supreme performance and generalization[2, 3]. The recent advances in deep learning have motivated some encouraging works for medical image segmentation[4–6] as well. But the deep learning models require large volumes of training data, which can be difficult to collect. The problem becomes more challenging for medical image segmentation as the data needs to be labelled at a pixel level. Annotating the data can be tedious for the human expert, thus, affecting the labeling accuracy. The limited number of domain experts in medical imaging adds more challenge to the process of collecting labelled data.

In biomedical imaging, domain expertise is required for labeling the data. Thus, it is very important to get only those data points labeled which contribute the most in the learning of the model. Active learning[7] is a paradigm that helps in querying the labels of only the most informative data points. Two factors - uncertainty and representativeness, define the informativeness of a data point. When working with deep neural networks, a single query point is not enough to fine-tune the weights of the model. Thus, pool-based strategies such as ranked batch-mode sampling, exploration-exploitation[8] are commonly used with the deep learning models. In past couple of years, there have been several efforts in applying active learning for medical image segmentation. Suggestive Annotation[9] is one of the initial frameworks for biomedical image segmentation using deep active learning for the MICCAI 2015 gland challenge. It uses the Fully Convolutional Network(FCN) and formulates the representativeness as the maximum set-cover problem. In a similar way, Representative Annotation[10] reduces the computational complexity of finding the most representative points by first using agglomerative clustering and then applying the maximum set-cover over each cluster . Both of these works require 50% of the labeled data. In this work, we show that the amount of labelled data can be reduced further.

In our work, we focus on applying active learning for brain tumor lesion segmentation from MR images by using a very limited amount of labeled data and computation time. We hypothesize that the data required to train a machine learning model can be reduced by selecting the training points intelligently, at a cost of a reduced accuracy within a certain limit. To test this hypothesis, we used the 2018 Brain Tumor Segmentation(BraTS) MICCAI challenge dataset. We also devised our own query strategy: A Coreset based Ranked Batch Mode Sampling algorithm to reduce the computational cost of selecting the query points, yet selecting the best ones. We used U-Net as the base deep learning model to perform the lesion segmentation task. Figure 1 summarizes the entire proposed deep active learning framework.

2 Methodology

The two major components of our framework are the deep learning model and the query strategy. In this section, we will discuss them in detail.

2.1 Model Architecture

U-Net[12] is one of the first successful models used for medical image segmentation. It is a classic encoder-decoder model, where the encoder captures the spatial information into a reduced form using Convolutional Neural Network(CNN). This is similar to any CNN model being used for classification, which first encodes the important features and then uses them for classification. But U-Net uses these encoded spatial features to reconstruct an output of same size as of input. This captures the semantic information by combining the high-level features from the encoder feature maps with the decoder using skip connections. We

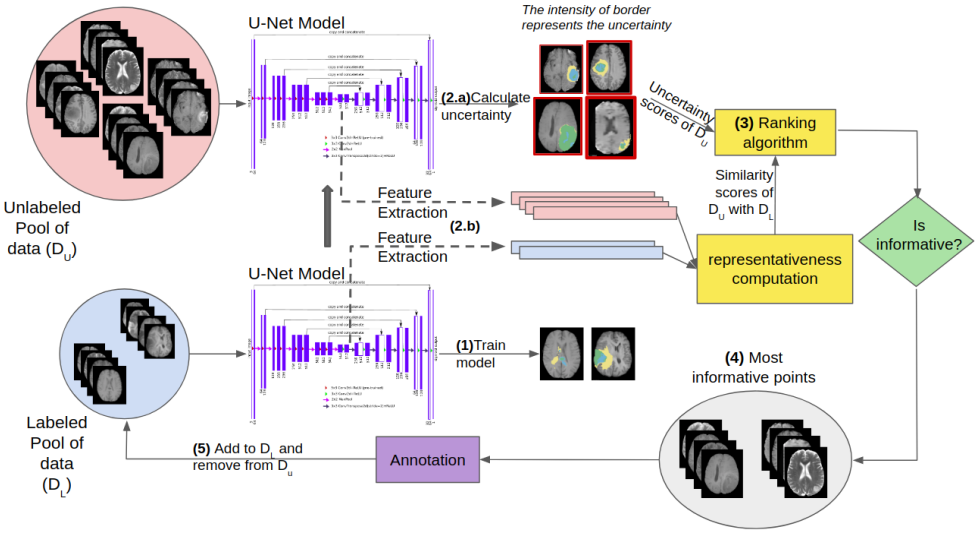

Fig. 1. Our active learning framework for MRI image segmentation. (1.) The model is trained on the labeled pool. (2.) Find the (a)uncertainty and (b)representative scores for the unlabeled pool using the trained model weights. (3.) The query strategy combines these scores (4.) and yields the query points, (5.) which are the annotated and included in labeled pool. This process is repeated until required.

used a modified version of U-Net as explained in[13]. This model uses a dice loss function. The dice loss function directly aims at maximizing the dice coefficient metric, thus performing better for data with class imbalance.

2.2 Query Strategies

Query strategy is an essential part of active learning and is used to find the most informative data points in the unlabeled pool. The two major factors that determine the informativeness of a data point are *uncertainty* - the confidence of the model in predicting the correct output of the unlabeled data point, and *representativeness* - there should be minimal similarity between the query pool and the labeled pool, as well as diversity within the query pool. There have been various approaches used to estimate the uncertainty of the model, like classification uncertainty and entropy based uncertainty. Exploration-exploitation[14], batch-mode sampling[15] are some query strategies that also capture the representativeness. We used uncertainty sampling, ranked batch mode sampling, and proposed coreset based ranked batch mode sampling, as described below.

Uncertainty Sampling. Uncertainty sampling[16] is the most common and one of the first query strategies. It is a classic stream-based selective sampling strategy that selects one query point at a time. Hence, it is more famous with

traditional ML models that can be fine-tuned by a single data point. But we need a batch of data points with deep learning models. For this, all the points in unlabeled pool are ranked according to their uncertainty and top n data points with the highest uncertainty are sampled.

This algorithm uses the prediction of the model to capture the uncertainty. There are three ways of doing this - least confidence, margin sampling, and entropy. We used the least confidence approach for the experiments as it is computationally cheap and yields good results empirically.

Ranked Batch-Mode Sampling(RBM). Uncertainty sampling is not the best strategy to mine the most informative query points as it fails to capture the representativeness. Batch Mode Active Learning (BMAL)[15] is a query strategy that incorporates reprsentativeness into its selection strategy. BMAL has its own drawbacks as discussed in[17]. Ranked batch-mode sampling algorithm describes a novel way of sampling points, keeping the major factors of uncertainty and representativeness in mind. It calculates the scores of the two factors and take the weighted mean by maintaining a running ratio. The points in the unlabeled pool are ranked based on this combined score. To increase the diversity in the labeled pool, it selects those points from the unlabeled pool which dont have a high similarity with the labeled pool and the query pool, thus capturing the intra- and the inter-diversity of the unlabeled pool as explained in Figure 2. Initially, the algorithm gives higher preference to the diversity factor, but as the sampling continues, the weighting scheme (α) shifts to the uncertainty scores.

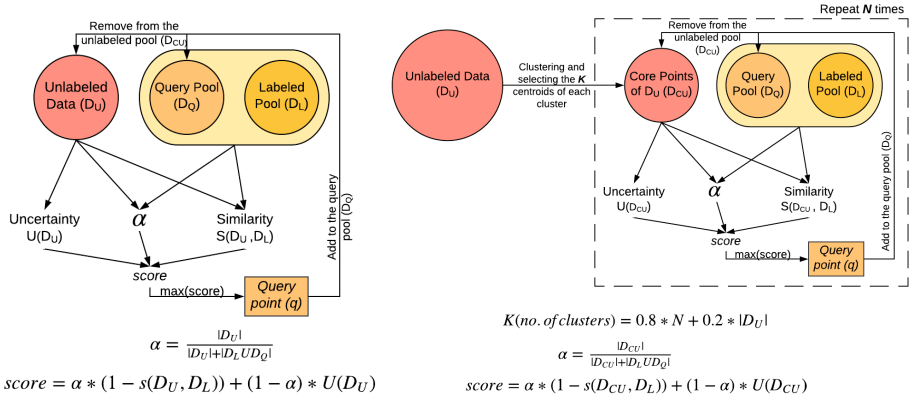

Fig. 2. (left)Ranked batch-mode sampling algorithm and (right)coreset based ranked batch-mode sampling algorithm

Coreset based Ranked Batch-Mode Sampling. The ranked batch-mode algorithm[17] performs well because it captures both, uncertainty and repre-

sentativeness. But because of the exhaustive similarity computation between the two pools, labeled and unlabeled, for selecting each query point, it slows down the process and also requires a large amount of memory. We propose a refined version of the ranked batch-mode algorithm by decoupling the step of finding the representative points to capture intra-pool diversity and inter-pool diversity. First, K-means is used to reduce the size of the unlabeled pool and keep the most diverse points. The K is decided by using the following formula - $K = 0.8N_q + 0.2N_u$, where N_q is the number of data points in the query pool and N_u is the number of data points in the unlabeled pool. Then, we choose the data points closest to the centroid of each cluster. The reduced pool is then used to select the data points most dissimilar to the labeled pool and capture the inter-diversity as shown in Figure 2.

2.3 Training process

We divided the available annotated data into two pools - labeled data and unlabeled data to emulate the active learning process. The entire process was then run as explained in Figure 1. The model was trained on the labeled pool of data. Predictions from the trained model are then used to determine the uncertainty scores of the of the data points in the unlabeled pool. For the representativeness computation, the encoder output of the same trained model was used to extract the features of the data points of the two pools. The uncertainty scores and the representativeness scores were then fed into the query strategy to find the most informative points. The model was trained from scratch after including annotated query points to the labeled pool. Also, the number of training epochs were increased by 2 after each query to compensate for the increased training dataset size. The batch-mode sampling techniques used Euclidean distance to determine representativeness.

3 Data and Experiments

We used the 2018 Brain Tumor Segmentation (BraTS) MICCAI challenge dataset to train and evaluate our algorithm. It consists of 210 cases of High Grade Gliomas (HGG) and 75 cases of Low Grade Gliomas (LGG) along with the ground-truth markings for the tumor. Each slice has been manually annotated into 4 categories - enhancing tumor, tumor core, whole tumor, and the background and normal brain pixels. Each case has 4 modalities - T1, T1 contrast enhanced (T1ce), T2 and FLAIR. The dataset is skull-stripped, interpolated to the same resolution of $1mm^3$, and registered to the same dimension of 240 x 240 x 155.

The data is randomly split into the train-validation-test parts in the ratio of 80:10:10 on case level. The training data has 166 HGG and 62 LGG cases, validation data with 22 HGG and 6 LGG cases, and testing data with 22 HGG and 7 LGG cases after the split. Each slice of the four modalities for every case is normalized to have zero mean and unit variance. The tumor is present only in

a small part of the brain. The healthy voxels comprise 98% of total voxels,0.18% belongs to necrosis, 1.1% to edema and non-enhanced, and 0.38% to enhanced tumor. Patches of size $128 * 128 * 4$ are randomly sampled from each slice after eliminating the zero-intensity pixels to tackle this class-imbalance problem. This populates the **training data** with **99,864 patches**, **validation data** with **12,264 patches**, and **testing data** with **12,702 patches**.

Exp. No.	Model	Training Data Used	Hyperparameters
1.	Vanilla U-Net	99k+ patches	Epochs = 40
2.	U-Net with Uncertainty Sampling	7k patches (∼7% data)	Initial Epochs = 10 Initial pool size = 2k patches Number of queries = 10 Patches labeled/query = 500
3.	U-Net with RBM Sampling	15k patches (∼15% data)	Initial Epochs = 10 Initial pool size = 9k patches **Number of queries = 10** Patches labeled/query = 600
4.	U-Net with Coreset Based RBM Sampling	13k patches (∼13% data)	Initial Epochs = 10 Initial pool size = 10k patches **Number of queries = 5** Patches labeled/query = 600

Table 1. Experiments conducted and the hyperparameters setting for each experiment

We conducted four experiments as in table 1. First, we used the entire data for training and trained the model for 40 epochs. In the second experiment, we used uncertainty sampling as the query strategy, 7% of the training data, and conducted 10 queries. In third experiment, ranked batch-mode sampling was used as the query strategy using 15% of the data and required 10 queries. Finally, we used the coreset based ranked batch mode sampling which used only 13% of the entire data and only 5 queries.

4 Results and Discussion

Table 2 presents the results for the experiments conducted. The Coreset based Ranked Batch Mode sampling outperformed all the other methods with the dice coefficient scores of 0.844 for whole tumor, 0.83 for tumor core, and 0.799 for enhancing tumor with a reduced average query time of just 43 minutes. The Ranked Batch Mode sampling gives somewhat closer results, but at a cost of greater computation time and memory.

With our coreset based ranked batch-mode algorithm, we were able to achieve better results with the limited labeled data (only ∼13%) and query computation time. We have also tackled the much prevalent issue of class imbalance using active learning: Our algorithm selects the under-represented data points to be included in the labeled pool, thus reducing the imbalance as much as possible.

Exp. No.	Model	Whole Tumor Dice Coefficient	Tumor Core Dice Coefficient	Enhancing Tumor Dice Coefficient	Avg Query Time
1.	Vanilla U-Net	0.815	0.689	0.608	-
2.	U-Net with Uncertainty Sampling	0.802	0.724	0.727	-
3.	U-Net with RBM Sampling	0.829	0.812	0.788	1hr 50mins
4.	U-Net with Coreset Based RBM Sampling	**0.844**	**0.83**	**0.799**	**43 mins**

Table 2. The dice coefficients for the three tumors for each experiment

This is evident from the comparable dice coefficient scores of the three tumors for our algorithm, versus the large difference in them when no active learning is used. Also, the increased scores of the coreset based approach empirically validates the fact that the points being selected are more informative as compared to using the previous strategies. Furthermore, decoupling the selection steps of the intra-pool and inter-pool diverse points yields a faster convergence as we require only 5 queries to reach the final accuracy scores, with average query time also reduced by one hour.

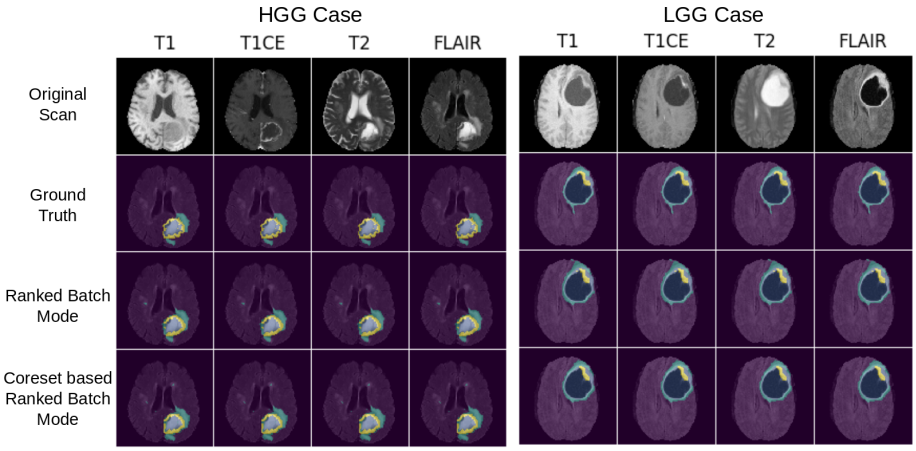

Fig. 3. The test results for HGG(left) and LGG(right) cases.

5 Conclusion and Future Work

In this paper, we presented a deep active learning based solution for MRI brain tumor image segmentation. Our contributions are (1) a more efficient and effective query strategy, and (2) method to tackle the class-imbalance problem using

active learning. This improved the results as shown in Figure 3 and Table 2 as our algorithm is able to capture even the minute details of the enhancing tumor. In future work, we will evaluate our framework with other datasets and explore bayesian networks[18] which can provide a better estimate of uncertainty of the unlabeled data points.

References

1. Zaitoun, N. et al.:Survey on Image Segmentation Techniques. In. Procedia Computer Science, vol. 65, pp. 797-806, Springer (2015)
2. Long, J. et al.: Fully Convolutional Networks for Semantic Segmentation. In: The IEEE Conference on Computer Vision and Pattern Recognition (CVPR). (2015)
3. Girshick, R.B. et al.: Rich feature hierarchies for accurate object detection and semantic segmentation. CoRR. abs/1311.2524, (2013)
4. Chen, H., Qi, X., Cheng, J.Z., Heng, P.A.: Deep contextual networks for neuronal structure segmentation. In: AAAI. pp. 11671173 (2016)
5. Chen, H., Qi, X., Yu, L., Heng, P.A.: Dcan: Deep contour-aware networks for accurate gland segmentation. In: CVPR. pp. 24872496 (2016)
6. Ronneberger, O., Fischer, P., Brox, T.: U-net: Convolutional networks for biomedical image segmentation. In: MICCAI. pp. 234241 (2015)
7. Settles, B.: Active Learning Literature Survey. University of Wisconsin–Madison (2009)
8. Yin, C. et al.: Deep Similarity-Based Batch Mode Active Learning with Exploration-Exploitation. In: Raghavan, V. et al. (eds.) ICDM. pp. 575584 IEEE Computer Society (2017)
9. Yang, L. et al.: Suggestive Annotation: A Deep Active Learning Framework for Biomedical Image Segmentation. CoRR. abs/1706.04737, (2017)
10. Zheng, H. et al.: Biomedical Image Segmentation via Representative Annotation (2019)
11. Multimodal Brain Tumor Segmentation Challenge 2018, `https://www.med.upenn.edu/sbia/brats2018/data.html`
12. Ronneberger, O., Fischer, P., Brox, T.: U-Net: convolutional networks for biomedical image segmentation. In: Navab, N., Hornegger, J., Wells, W.M., Frangi, A.F. (eds.) MICCAI 2015. LNCS, vol. 9351, pp. 234241. Springer, Cham (2015)
13. Isensee, F., Kickingereder, P., Wick, W., Bendszus, M., Maier-Hein, K.H.: Brain tumor segmentation and radiomics survival prediction: contribution to the BRATS 2017 challenge. In: Crimi, A., Bakas, S., Kuijf, H., Menze, B., Reyes, M. (eds.) BrainLes 2017. LNCS, vol. 10670, pp. 287297. Springer, Cham (2018)
14. Loy, C.C. et al.: Stream-based joint exploration-exploitation active learning. In: CVPR. pp. 15601567 IEEE Computer Society (2012)
15. Guo, Y., Schuurmans, D.: Discriminative Batch Mode Active Learning. In: Platt, J.C. et al. (eds.) NIPS. pp. 593600 Curran Associates, Inc. (2007)
16. H. Xu, X. Wang, Y. Liao and C. Zheng.: An Uncertainty Sampling-Based Active Learning Approach for Support Vector Machines. In: International Conference on Artificial Intelligence and Computational Intelligence, Shanghai 2009 pp. 208-213 (2009)
17. Cardoso, T.N.C. et al.: Ranked batch-mode active learning. Inf. Sci. 379, 313337 (2017)
18. Gal, Y. et al.: Deep Bayesian Active Learning with Image Data. CoRR. abs/1703.02910, (2017)