Iowa Initiative for Artificial Intelligence

Final Report

Project title:	Automated Identification and Segmentation of Intracranial Aneurysms	
Principal Investigator:	Edgar Samaniego	
Prepared by (IIAI):	Nam H. Le	
Other investigators:	Colin P. Derdeyn, Mohamed Elshikh	
Date:	June 27, 2021	
Were specific aims fulfilled:		YES
Readiness for extramural proposal?		In part
If yes … Planned submission date		
Funding agency		
Grant mechanism		
If no … Why not? What went wrong?		

Brief summary of accomplished results:

Three challenges of aneurysm segmentation are addressed in this work. First, the lack of quality strong labels is remedied by a self-training framework that utilizes weak labels that are easier to annotate. Second, the guarantee of LOGISMOS to produce a visually accurate mesh is leveraged to improve the quality of the pseudo labels. Third, the LOGISMOS module is adapted to further improve the quality of the final predictions.

We proposed a semi-supervised learning method to train an automatic segmentation network capable of detecting unruptured aneurysms in TOF-MRA images. The model was trained on two large datasets which were made available publicly. Our framework starts with a baseline U-Net model trained with strong labels selected from both datasets.

In our study, the weakly annotated data underwent an inference process utilizing the baseline model to produce pseudo labels. These pseudo labels were evaluated based on their dice loss, and those exceeding a predetermined threshold were reintroduced into the training set. This iterative process continued until the final model reached a point where it could no longer accurately detect true aneurysms within the remaining weakly annotated data. Because of the superior performance of the LOGISMOS graph-based surface optimization with geometric constraints in its capability to improve segmentation mesh, we introduced a tailored LOGISMOS module that further enhance the quality of (1) the pseudo labels and (2) the final predictions.

The proposed method was evaluated in terms of segmentation dice, Jaccard, detection sensitivity, and false positive count. The results were compared with the baseline nnU-Net model. We also investigated the effects of iterative concatenation of new training examples and dice selection thresholds on the model performance.

We trained thirteen deep learning models, the resulting DeepLOGISMOS_SSL model achieves a dice score of 0.5 and a detection rate of

Research report:

- Aims: Develop an automatic method for the segmentation of unruptured intracranial aneurysms.

- Data:

The data were collected from three sources:

(1) our in-house the University of Iowa Hospitals and Clinics (UIHC),

(2) the public dataset of the Lausanne University Hospital \cite{dinotoAutomatedBrainAneurysm2023},

(3) the publicly available dataset provided in the ADAM challenge by UMC Utrecht, the Netherlands.

The UIHC dataset was used as the held-out test set unseen during the training process. Healthy individuals with no presence of UIAs were excluded from this study. Other exclusion criteria include patients with ruptured aneurysms and fusiform aneurysms. Treated aneurysms were not of interest and thus were coded as background in the reference standard. Table 1 summarizes the number of scans and UIAs present in the used data. In total, we retrieved 292 deidentified scans with 361 aneurysms present. The majority of the scans have only one aneurysm.

Table 1. Dataset summary

			Cases per aneurysm				Label counts		
	Case count	Aneurysm count	1	2	3	4	Strong	Weak	Train, test plan
ADAM	93	126	67	21	4	1	126	0	train, validation
Lausanne	162	193	137	21	2	2	45	148	train, validation
UIHC	37	42	32	5	0	0	0	0	held-out test
Total	292	361	236	47	6	3	171	148	

AI/ML Approach:

Strong labels or voxel-wised labels are the most common form of annotations in medical image analysis. They are usually obtained by manually drawing the region of interest with supporting software such as ITK-SNAP or Paraview. Weak labels, on the other hand, are generated with non-traditional processes such as manual drawing or point-annotating anatomical landmarks. We utilized both types of labeled data in our model development.

Landmark annotations are a unique form of weak labels that rely on a single voxel, identified by its three-dimensional coordinate, to indicate the approximate location of a UIA. The purpose of these annotations is to pinpoint the aneurysm by locating its sac. Landmark annotations, being the most expedient to generate, are employed as user inputs in the inference process to assist the bounding-box model in pinpointing the aneurysms.

Given the dataset containing strong labels and weak labels we aim to train a model that can produce accurate segmentation masks for both types of labels. The process starts by establishing a baseline model, which is trained using data samples and their corresponding strong labels. These labels are associated with the training set. Each subsequent iteration of the base model has its training set expanded to include new training examples. The process was repeated until the final model was no longer able to detect true aneurysms in the remaining weakly annotated data,

Thirteen variations of the DeepLOGISMOS models were trained and evaluated. The baseline model nnUNet_bl was used as a starting model that would generate the first set of pseudo labels for subsequent training iterations. The names of these models are suffixed with a number dr indicating $10 \times d$ with d being the dice threshold used to select new training examples from the current test set. DeepLOGISMOS_SSL_iter$i$ is the i-th iteration of DeepLOGISMOS_SSL which is trained on additional images with pseudo label predicted by the previous iteration $i-1$ of DeepLOGISMOS_SSL. For example, DeepLOGISMOS_SSL_iter1 was trained with a set of new pseudo labels, which were inferred by the baseline nnUNet_bl and whose dice score is greater than 0.1.

Experimental Methods, Validation Approach

Description of the trained 13 deep learning models:

- nnUNet_bl: baseline nnU-Net, trained on images with strong labels.
- Bbox: nnU-Net model trained on a series of smaller images cropped around the known aneurysm coordinates in the index space.
- nnUNet_d5: a self-training model trained on additional images with pseudo label predicted by nnUNet_bl
- DeepLOGISMOS_SSL_iter1: a self-training model trained on an additional image with pseudo label predicted by nnUNet_bl with post-processing of predicted mask. The dice threshold for the selection of pseudo labels was set to d = 0.5.
- DeepLOGISMOS_SSL_iter2: 2nd iteration of DeepLOGISMOS_SSL, d = 0.5.
- DeepLOGISMOS_SSL_iter3: 3rd iteration of DeepLOGISMOS_SSL, d = 0.4.
- DeepLOGISMOS_SSL: final iteration of DeepLOGISMOS_SSL, d = 0.1.
- DeepLOGISMOS_SSL_iter$i$: i-th iteration of DeepLOGISMOS_SSL which trained on additional images with pseudo label predicted by the previous iteration $i-1$ of DeepLOGISMOS_SSL.
- DeepLOGISMOS_dk: a variation of DeepLOGISMOS_SSL_iter1 which the dice the threshold for a pseudo-label generation was set to $0.1 * k$.

Table 1 and 2 shows the segmentation and detection metrics in the 5-fold cross-validation sets. The DeepLOGISMOS_SSL achieve a balanced performance, with 0.5 in dice and 0.59 in successful detection rate.

Table 2. Segmentation dice of 13 deep learning models

	Segmentation				
	Dice		Jaccard		Missed
nnUNet_bl	0.46	±0.33	0.36	±0.28	30
Bbox	0.57	±0.25	0.44	±0.23	10
nnUNet_d5	0.48	±0.33	0.37	±0.28	28
DeepLOGISMOS_SSL_iter1	0.51	±0.32	0.40	±0.27	27
DeepLOGISMOS_SSL_iter2	0.49	±0.32	0.63	±0.27	25
DeepLOGISMOS_SSL_iter3	0.49	±0.34	0.38	±0.29	30
DeepLOGISMOS_SSL	0.50	±0.33	0.40	±0.28	28
DeepLOGISMOS_d1	0.51	±0.31	0.40	±0.27	23
DeepLOGISMOS_d2	0.51	±0.31	0.40	±0.27	24
DeepLOGISMOS_d3	0.51	±0.33	0.40	±0.28	29
DeepLOGISMOS_d4	0.50	±0.32	0.39	±0.28	28
DeepLOGISMOS_d5	0.51	±0.32	0.40	±0.27	27
DeepLOGISMOS_d6	0.48	±0.34	0.38	±0.29	32

Table 3. Detection results

	Mean detection metric per case (N=37)							
	FP count				Sensitivity			
	Baseline		+ LOGISMOS		Baseline		+ LOGISMOS	
nnUNet_bl	0.65	±1.32	0.43	±0.89	0.46	±0.50	0.54	±0.50
Bbox	0.19	±0.39	0.86	±5.33	0.74	±0.44	0.88	±0.32
nnUNet_d5	0.78	±1.42	0.54	±0.83	0.46	±0.50	0.54	±0.50
DeepLOGISMOS_SSL_iter1	0.73	±1.31	0.62	±0.97	0.51	±0.50	0.54	±0.50
DeepLOGISMOS_SSL	0.86	±1.32	0.57	±0.89	0.46	±0.50	0.59	±0.48
DeepLOGISMOS_d2	0.76	±1.50	0.57	±1.00	0.51	±0.50	0.57	±0.48

Limitations:

None

Publications resulting from project:

A journal entitled is being prepared.

N. Le, H. Zhang, E. Mohamed, E. A. Samaniego, C. P. Derdeyn, and M. Sonka, "Deep LOGISMOS: A Self-learning Semi-supervised Framework for Aneurysm Segmentation with Graph Label Refinement."

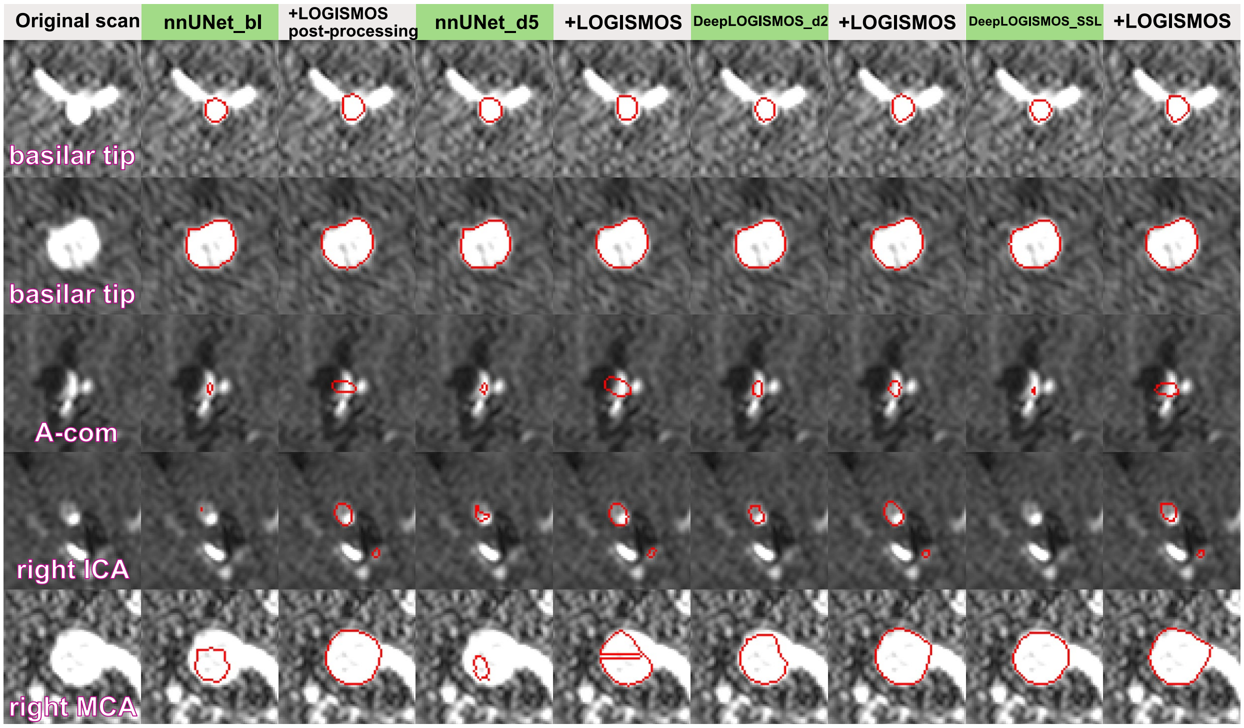

Figure 1. Segmentation results of some representative examples from the in-house UIHC dataset.