Jenish Maharjan **Week 7**

Paper: *Interactive Medical Image Segmentation Using Deep Learning with Image-Specific Fine Tuning*

Summary:

Wang et al, in this paper propose a deep learning framework to address problems and limitations of using convolutional neural networks (CNNs) for the task of medical image segmentation. While CNNs are widely used for automatic medical image segmentation, they haven't been as accurate and robust as it is required for use in the clinical context. CNNs are also limited by the lack of image-specific adaptation and generalizability to previously unseen object classes. In their approach, they introduce an interactive segmentation framework that incorporates CNNs into a bounding box and scribble-based segmentation pipeline.

In the interactive framework with Bounding box and Image-specific Fine-tuning-based Segmentation (BIFSeg) that the authors have propose, they use a CNN that takes the content of a bounding box of one instance and produces a binary segmentation for that instance. The user provides the bounding box in the testing stage after and then the BIFSeg extracts the region inside the bounding box. The extracted region is fed into the pre-trained CNN to obtain an initial segmentation. They designed their CNNs to learn some common features across different objects which allows for generalizability to unseen objects. The segmentation is then further refined by using unsupervised or supervised image-specific fine-tuning. Fine-tuning is used to leverage image-specific features which makes their CNNs adaptive to a specific test image for better segmentation. Their framework can handle both 2D and 3D segmentations. For 2D images, they use the P-Net for bounding box-based binary segmentation. They use dilated convolution to preserve resolution. For 3D images, they use a network extended from P-Net which has an anisotropic receptive field. During training, each instance is cropped with its bounding box and the CNN is trained for binary segmentation. In the testing stage, they use image-specific fine-tuning with optional scribbles and a weighted loss function.

In the presented work, they validate their framework with two applications – 2D segmentation of multiple organs from fetal MRI and 3D segmentation of brain tumors from contrast enhanced T1-weighted and Fluid-attenuated Inversion Recovery (FLAIR) images. For comparison they also investigate the performance of FCN and U-Net for 2D images and PC-Net with DeepMedic and HighRes3DNet for 3D images. The performance of BIFSeg was compared to the initial output of P-Net/PC-Net, post processing the initial output with a CRF and image-specific fine-tuning. Comparisons were also made with other interactive methods like GrabCut, Slic-Seg and Random Walks for 2D segmentation and GeoS, GrowCut and 3DGrabCut for 3D segmentation.

For the 2D segmentation application, data was obtained from 18 patients. Data splitting was performed at patient level and images from three patients were used for training, validation and testing respectively. As for the ground truth, manual segmentations by a Radiologist were used. The input of P-Net were resized to deal with organs at different scales. In this application, FCN and U-Net are shown to over-fit in the training set while P-Net generalizes better to previously unseen inputs. P-Net also had the best accuracy with a low average machine time. The initial segments obtained by P-Net was refined by CRF, BIFSeg(-w) and BIFSeg without additional scribbles for unsupervised refinement. The same initial segmentation and scribbles were used for supervised refinement. Similarly, for the 3D segmentation task, the ground truths were based on manual labels given by experts. The dataset had scans using multiple

MR sequences with different contrasts that came from 198 patients. The performance of their framework in both of these applications show that the user time and final accuracy are similar to human experts. With these results they demonstrate a novel approach to interactive medical image segmentation which is more robust to segment previously unseen objects than the existing systems that is accurate even with less user interactions under less time which is very critical in terms of clinical use.

Strengths:

- The images showing various stages in the framework were really helpful.
- The two applications of their framework have been explained in detail. The details about data acquisition and use of various other systems to validate their framework is great as well.

Weaknesses:

- Since the target audience for this paper are scholar in the biomedical field, they have used a lot of medical jargons. As a person from a different background, it was difficult to comprehend some of the more domain specific concepts. A brief background of some of the jargons would have been helpful.
- The organization of the content could have been modified to increase readability of the paper.

Confusions:

- How and why is data splitting done at patient level?
- What dataset is used for the pre-trained CNN?

Discussion Questions:

- What kind of similarities between the training and testing object/image classes are desired for zero-shot learning?
- What sort of feature representation can be expected from the outputs of each block?
- What are some key issues that systems like these need to address so that they’re ready for real world implementation?