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

Nirajan Koirala

CSC 9010_001

Summary

Authors investigate combining CNNs with user interactions for medical image segmentation to achieve higher segmentation accuracy and robustness with fewer user interactions and less user time in the given paper. These kind of studies using CNNs for interactive segmentation have been very few so far mainly due to the requirement of large amount training images, lack of image specific adaptation and demanding balance between model complexity, inference time and memory space efficiency. The paper presents a new framework to address these challenges. Authors propose a bounding box based segmentation pipeline to generalize to unseen objects and image-specific fine-tuning to make CNNs adaptive to different test images. Also, their models have a balance among receptive field, inference time and memory efficiency so that it can be more responsive and run on devices with low GPU power.

In section II, they describe their method of using CNN which takes an input as the content of bounding box of one instance and gives a binary segmentation which helps in dealing with previously unseen objects in a unified framework. During testing phase, bounding box is provided by qualified individuals and their model BIFSeg extracts the region inside the box and feeds it into the pre-trained CNN with a forward pass to obtain an initial segmentation. They have used both supervised and unsupervised image-specific fine-tuning methods for models to make CNNs adaptive to specific test image for better segmentation. Also, this framework can handle both 2D and 3D segmentations. For the 2D images, they adopt the P-Net architecture for bounding box-based binary segmentation and provide further details about it through appropriate figures. To ensure efficient fine tuning and fast response to user interactions, they only fine-tune parameters of the classifier. For 3D images, they use a network extended from P-Net (PC-Net), however it considers a trade-off among receptive field, inference time and memory efficiency.

Training description for the 2D/3D segmentation is provided where models are trained to extract the target form the bounding box, which is a binary segmentation problem and a cross-entropy loss function is used for training. Later, they describe the methods of unsupervised and supervised image-specific fine-tuning where a user may provide a set of scribbles to guide the update of labels. Formulas used and optimization techniques are described later in the sub-section. Caffe library is used to implement the P-Net and PC-Net and for testing they use a machine having medium specs. In section III, experiments and results are discussed after comparing these models with other state-of-art networks. Separate comparisons are done for 2D and 3D models using different type of networks for each. For 2D-images, the BiFSeg model performs better that others and also takes less time in average for unsupervised image-specific fine-tuning. When the scribbles provided by Obstetrician and Radiologist are used for supervised image-specific fine-tuning, segmentation is significantly improved. BIFSeg model is again found to outperform other models mainly for previously unseen organs in terms of accuracy and time. After using segmentation in 3D-images, PC-Net is able to obtain similar results as other models by using less complexity and lower memory consumption. Using unsupervised fine-tuning for brain tumor segmentation without additional user interactions, BIFSeg model is found to perform better than others and a similar situation is observed for segmentation of previously unseen whole tumor. With additional scribbles provided by the user, in supervised image specific fine-tuning, it can be observed that BIFSeg again achieves better refined results for both tumor core and whole tumor. It is also found that the user time is significantly lower for BIFSeg models. Different figures and tables further aid the quantitative comparison of the results of different models discussed in paper.

In conclusion, authors discuss how their models deal with unseen objects and it can be a major advantage compared with traditional CNNs or even transfer learning since they reduce the efforts needed for gathering and annotating training data and they can be applied to some unseen organs directly. Their experimental results also show that the image-specific fine-tuning can improve the segmentation performance.

Strengths

- The paper is fairly easy to follow and the methods and techniques discussed are intuitive.
- Good amount of figures and tables are provided for the user to follow results and experiments.
- Various statistical techniques are used to justify the validity of their results.

Weaknesses

- I think the organization of figures and tables presented in this paper could be improved to help readability.
- Some sections like conclusion repeat the information already provided in other sections and are thus lengthier than necessary.

Points of Confusion

- What is a dilated convolution and how does it helps the network in preserving resolution?
- What is an anisotropic receptive field and why is it beneficial for 3D image classification?
- Section II C is confusing as well due to the presence of many Mathematical terms.

Discussion Questions

- How does uncertainty plays an important role for the image-specific fine tuning?
- Would the networks be more accurate if the scribbles provided by the users are more specific instead of just a box?

- I think accuracy is a far more important factor than time while using these type of systems for medical use. So, would it be better to make the network more deeper and accurate at the cost of time?